Supplementary Material for CAMEL: Confidence-Aware Multi-task Ensemble Learning with Spatial Information for Retina OCT Image Classification and Segmentation

Juho Jung[1*] Migyeong Yang[1*] Hyunseon Won[1] Jiwon Kim[1]
Jeong Mo Han[2] Joon Seo Hwang[3] Daniel Duck-Jin Hwang[4,5†] Jinyoung Han[1†]
[1]Sungkyunkwan University, Seoul, South Korea
[2]Seoul Bombit Clinic, Seoul, South Korea [3]Seoul Plus Eye Clinic, Seoul, South Korea
[4]Hangil Eye Hospital, Incheon, South Korea [5]Lux Mind, Incheon, South Korea
{jhjeon9, mgyang, dprth1014, jjeon416}@g.skku.edu,
{joehan712, poppn78, hallelu7}@gmail.com, jinyounghan@skku.edu

A. Data

A.1. Data Pre-processing

A.1.1 Pixel Label Mapping

The *Pixel Label Mapping* compares pixels between the resized and original masks, filtering out blurred pixels or pixels that are incorrectly mapped to colors of a different class. In particular, as the segmentation task involves pixel-level classification, the pixels of the train and test mask images can be denoted as $P = \{x_n, y_n\}_{n=1}^{M \times N}$, where $x \in \mathbb{R}^{M \times N}$ represents the shape of $M \times N$ input image and y can belong to one of the classes $L = \{l_1, l_2, l_3, ..., l_n\}$. Note that, if the dataset has n classes, there are n colors representing the classes. We first resized an original mask of size $M \times N$ into the target size of $k \times k$. If there are pixels in the resized $k \times k$ mask that do not belong to the class set L, including blurred pixels, their coordinates are stored separately in a dictionary format for modification in the following steps.

A.1.2 Pixel Remapping

The pixels, which are not mapped in *Pixel Label Mapping* process, are assigned with new values considering *Color Distance* and *Lesion Distance* that are determined by medical relationships with other lesions. Specifically, for each unmapped pixel $P = (x, y)$, where x and y represent the pixel coordinates of an $M \times N$ size input image, we calculate the *Color Distance*, $CD = \{d_1, d_2, d_3, ..., d_n\}$, between the pixel's color and the colors of the lesion classes $L = \{l_1, l_2, l_3, ..., l_n\}$. The reason for calculating the color distance from the labels of lesions rather than calculating the distance between adjacent pixels is that there may be inaccurate pixels among the adjacent pixels that are blurry or not properly mapped. Subsequently, we determine the nearest class l as follows:

$$\text{Nearest Class } (l) = \arg \min_{\text{Adjacent Pixels}} \|CD\| \tag{1}$$

The predefined *Lesion Distance* (LD) is then utilized to determine if the pixel value l can be assigned to its corresponding pixel location $P = (x, y)$ as follows:

$$\text{Lesion Distance } (LD) = \{(i, j) : \|RGB_i - RGB_j\|\} \tag{2}$$

For all i, j where i≠j. Here, the dictionary LD contains precomputed *CD* between pairs of lesions (i, j) based on medical knowledge, which identifies adjacent and non-adjacent lesions. Specifically, we calculate the Euclidean distances between the nearest class l and adjacent pixels of the target pixel P, excluding blurred or inaccurately mapped pixels that do not belong to class L. Based on the predefined LD, pixels representing adjacent lesions are retained, while those representing non-adjacent lesions are left unassigned unless the *CD* meets a specific *LD* threshold. By calculating lesion relationships using *LD*, this approach can be universally applied to other medical images, regardless of the medical domain.

To exemplify, in the case shown in the first column of Figure 1, *ERM* (Orange), which stands for '*Epiretinal Membrane*,' located '*above*' the Retina (Sky Blue), while *RPE* (Purple line), which represents '*Retinal Pigment Epithelium*,' exists '*beneath*' the Retina. Medical standards indicate that *ERM* and *RPE* cannot be adjacent [7]. We evaluate the *Color Distance* and assign pixel values according to the predefined *Lesion Distance*. If the calculated *Color Dis-*

*Equal contribution.
†Corresponding authors.

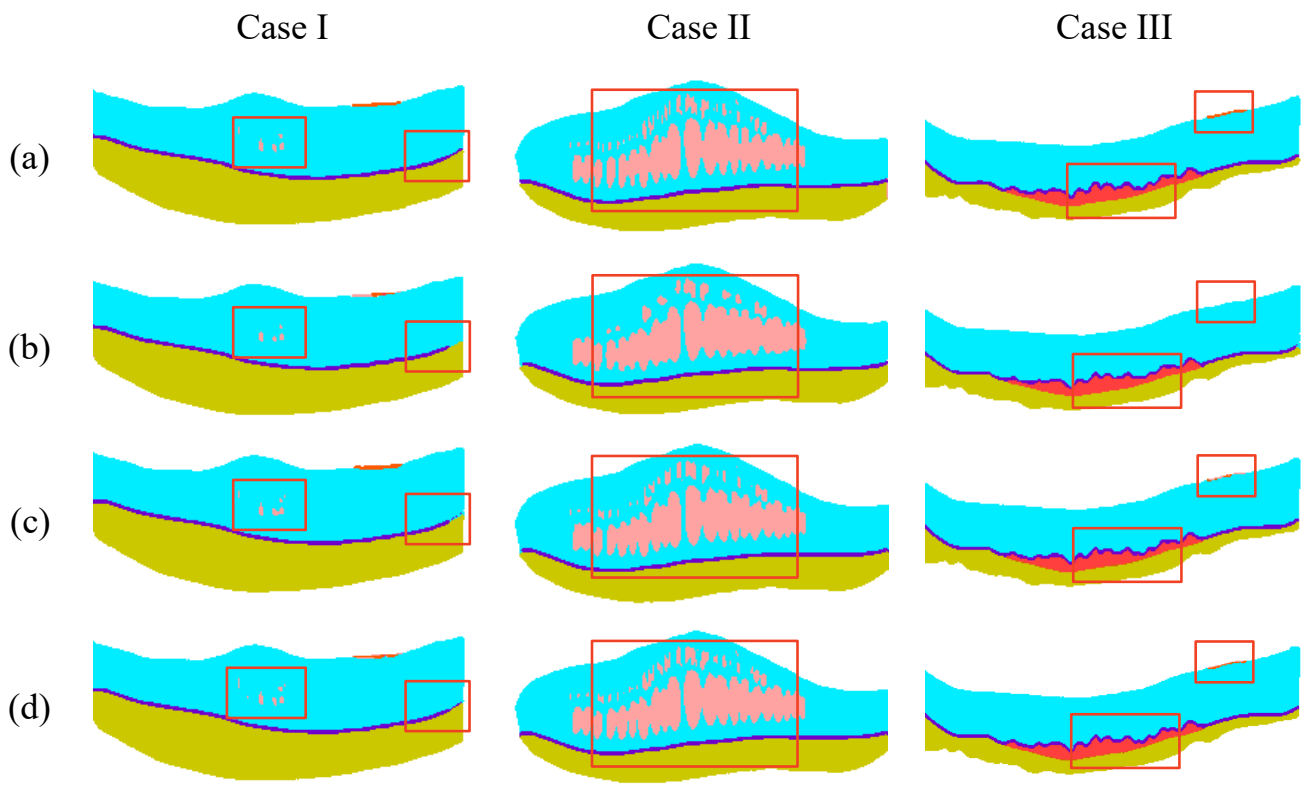

Figure 1. Data pre-processing results in the following scenarios: **(a)** Manual annotation by doctors (Ground Truth), **(b)** Only *K-Neighbor Post-Processing* applied, **(c)** Incorporating *Color Distance* and *Lesion Distance*, and **(d)** The proposed data pre-processing method. Combined utilization of *Color Distance* and *Lesion Distance* with *K-Neighbor Post-Processing* yields finer and more accurate pixel resizing.

tance between lesion pairs corresponds to the expected *Lesion Distance*, pixel values are assigned; otherwise, they remain unassigned and the process proceeds to the final step.

A.1.3 K-Neighbor Post-Processing

In this step, pixels that have not been mapped are processed in the final *K-Neighbor Post-Processing* step for mapping, where K denotes the number of neighboring pixels considered. This involves examining the surrounding K pixels of the target pixel and performing a majority voting to determine the final label assignment. The majority voting operation can be expressed as follows:

$$y_i = \arg\max_{l_j \in L} \sum_{k=1}^{K} I(y_k = l_j) \tag{3}$$

where y_i is the final label assigned to the pixel in question, $l_j \in L$ represents the possible classes, $I(\cdot)$ is the indicator function, and y_k denotes the label of the k-th neighboring pixel.

B. Experimental Settings

B.1. Baselines

To validate the performance of our proposed model, we compare recent baseline models that showed state-of-the-art performance in retinal imaging. Our baselines include multi-task learning, segmentation, and classification models.

- **UML** [5]: Uncertainty-informed Mutual Learning (UML) framework performs joint classification and segmentation tasks in medical image analysis. By generating image-level and pixel-wise confidence scores, UML employs an uncertainty navigator to enhance the reliability and interpretability of classification and segmentation output.
- **TCCT-BP** [6]: Tightly combined Cross-Convolution and Transformer with Boundary regression and feature Polarization (TCCT-BP) performs segmentation task by combining CNN and lightweight Transformer to enhance the perception of retinal layers. The model incorporates a feature grouping and polarization loss function to differentiate feature vectors, along with a

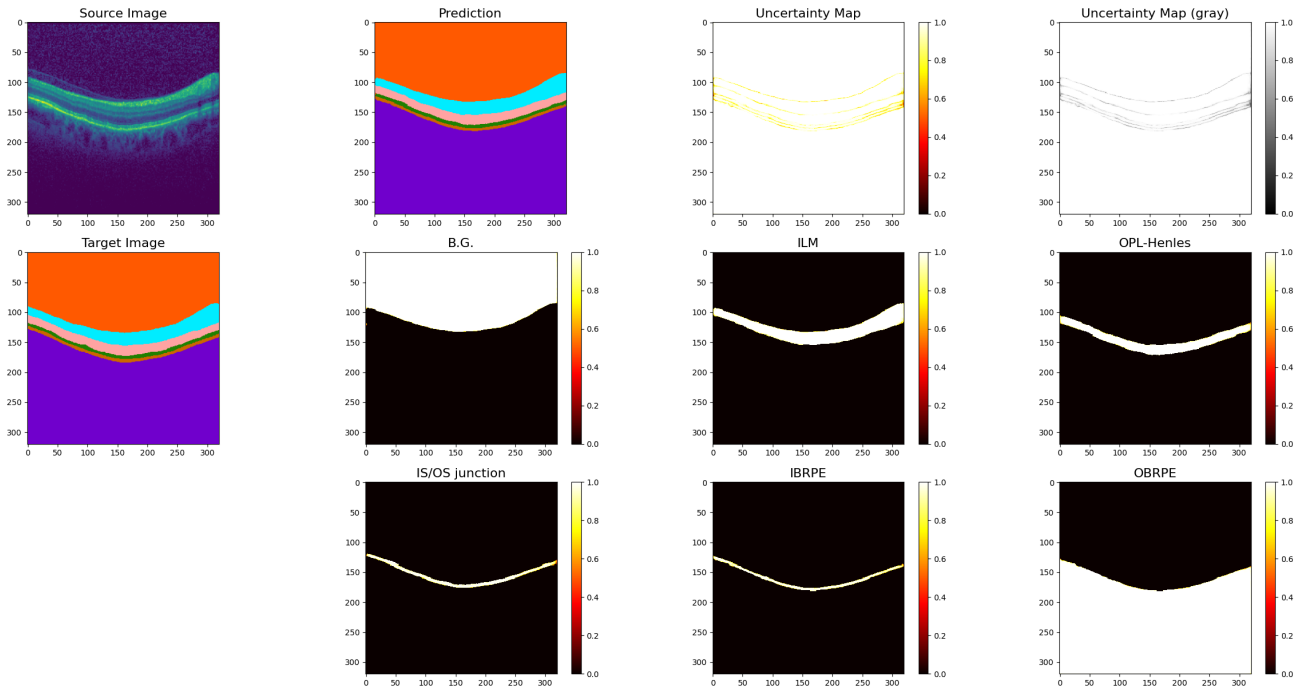

Figure 2. Visualization of region-specific uncertainty map through uncertainty estimation of *CAMEL* in OCT5k dataset.

boundary regression loss function to accurately align retinal boundaries with the ground truth.

- **Attention-based U-Net** [4]: Attention-based U-Net model integrates the soft attention mechanism into the U-Net architecture. The model demonstrates high performance in segmenting the fluids and the boundaries between layers in OCT images.
- **MedViT-S** [3]: MedViT is a robust and efficient CNN-Transformer hybrid model capable of performing various medical image classification tasks. The model demonstrates high generalization ability on the large-scale MedMNIST-2D dataset.
- **VGG-19-based model** [2]: The VGG-19-based model utilizes a well-known CNN architecture combined with transfer learning and mix-up-based data augmentation. The model's classification performance demonstrated the highest accuracy when compared to eight ophthalmologists.

C. Results

C.1. Validating CAMEL's Generalizability

We validated *CAMEL*'s generalizability on the OCT5k dataset [1], where the layers in the OCT images are labeled. Figure 2 presents the segmentation results of *CAMEL*. Despite variations in segmentation labels compared to our dataset, the model's outcomes align well with the target areas for each layer. The 'Uncertainty Map' also demonstrates that the model prevents overconfidence at the boundaries of each layer.

C.2. Comparison with Interpolation Methods in OCT Image Pre-processing

Figure 3 illustrates the comparison on a resized OCT mask image using *our pre-processing method* and other various image interpolation techniques. When observing the boundaries of each class, *our pre-processing method* exhibits clear pixel colors, while other interpolation methods show blurred boundaries.

To investigate whether *our pre-processing method* leads to performance improvements not only on our dataset but also on other datasets, we conducted a comparative study on the OCT5k dataset [1] with various image resizing methods. As shown in Table 1, utilizing resized mask images produced by *our pre-processing method* achieved the highest average Dice score, underscoring the effectiveness of the image resizing techniques for segmentation tasks in CAMEL. This confirms the general utility of *our pre-processing method* in medical segmentation tasks.

References

[1] Mustafa Arikan, James Willoughby, Sevim Ongun, Ferenc Sallo, Andrea Montesel, Hend Ahmed, Ahmed Hagag, Mar-

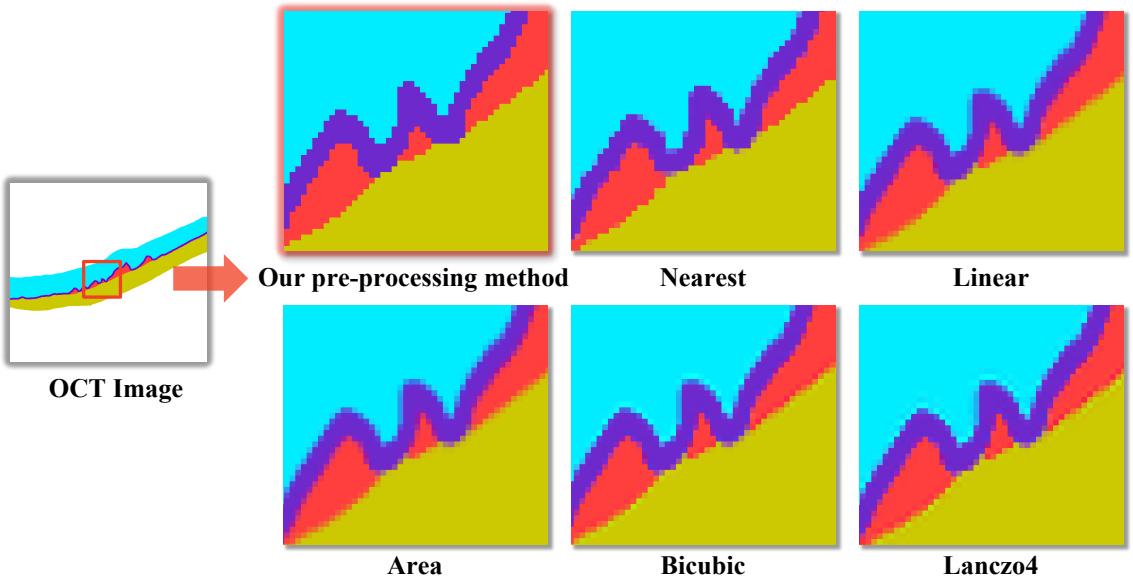

Figure 3. Visualization comparing the resized OCT (Optical Coherence Tomograph) image at 256 × 256 dimensions using *our pre-processing method* against other image interpolation methods.

Table 1. Performance comparison between using our data pre-processing method and other resizing methods on OCT-5k dataset.

Method	ILM	OPL-Henles	IS/OS junction	IBRPE	OBRPE	B.G.	Avg.
Bicubic	0.9880	0.8988	**0.8722**	0.7138	0.6802	0.9916	0.8574
Linear	**0.9881**	0.8978	0.8685	0.7059	0.6699	**0.9918**	0.8537
Lanczos4	0.8185	0.7594	0.6602	0.0249	0.0137	0.9812	0.5430
Nearest	0.9862	0.8981	0.8673	0.7174	0.6785	0.9905	0.8563
Our pre-processing method	0.9482	**0.9375**	0.8044	**0.7943**	**0.9980**	0.9876	**0.9117**

ius Book, Henrik Faatz, Maria Vittoria Cicinelli, et al. Oct5k: A dataset of multi-disease and multi-graded annotations for retinal layers. *bioRxiv*, pages 2023–03, 2023. 3

[2] Jinyoung Han, Seong Choi, Ji In Park, Joon Seo Hwang, Jeong Mo Han, Junseo Ko, Jeewoo Yoon, and Daniel Duck-Jin Hwang. Detecting macular disease based on optical coherence tomography using a deep convolutional network. *Journal of Clinical Medicine*, 12(3):1005, 2023. 3

[3] Omid Nejati Manzari, Hamid Ahmadabadi, Hossein Kashiani, Shahriar B Shokouhi, and Ahmad Ayatollahi. Medvit: a robust vision transformer for generalized medical image classification. *Computers in Biology and Medicine*, 157:106791, 2023. 3

[4] Martina Melinščak. Attention-based u-net: Joint segmentation of layers and fluids from retinal oct images. In *2023 46th MIPRO ICT and Electronics Convention (MIPRO)*, pages 391–396. IEEE, 2023. 3

[5] Kai Ren, Ke Zou, Xianjie Liu, Yidi Chen, Xuedong Yuan, Xiaojing Shen, Meng Wang, and Huazhu Fu. Uncertainty-informed mutual learning for joint medical image classification and segmentation. *arXiv preprint arXiv:2303.10049*, 2023. 2

[6] Yubo Tan, Wen-Da Shen, Ming-Yuan Wu, Gui-Na Liu, Shi-Xuan Zhao, Yang Chen, Kai-Fu Yang, and Yong-Jie Li. Retinal layer segmentation in oct images with boundary regression and feature polarization. *IEEE Transactions on Medical Imaging*, 2023. 2

[7] Jason R Wilkins, Carmen A Puliafito, Michael R Hee, Jay S Duker, Elias Reichel, Jeffery G Coker, Joel S Schuman, Eric A Swanson, and James G Fujimoto. Characterization of epiretinal membranes using optical coherence tomography. *Ophthalmology*, 103(12):2142–2151, 1996. 1