Supplementary Material for "Prior Activation Map Guided Cervical OCT Image Classification"

Qingbin Wang[1], Wai Chon Wong[1], Mi Yin[2], and Yutao Ma[2,*]

[1] School of Computer Science, Wuhan University, China
[2] School of Computer Science & Hubei Provincial Key Laboratory of Artificial Intelligence and Smart Learning, Central China Normal University, China
ytma@ccnu.edu.cn

1 Generating PAMs Based on Prior Knowledge

1.1 Cyst Image

Figure 2(b) in the main text illustrates how pixel activation values are set for a nabothian cyst. The activation values of pixels along the cyst contour are fixed at a maximum of 1.0, while pixels inside the cyst cavity are uniformly set to 0.8. Calculating activation values of pixels belonging to other regions follows the same method used for inflammation images, as described in the main text.

1.2 Ectropion Image

Figure 2(c) in the main text presents how papillary structures' pixel activation values are determined. The activation values of pixels along the contours of papillary structures are fixed at a maximum of 1.0. The method used for epithelial pixels in inflammation images is followed to calculate activation values of pixels inside the papillary structure, with two key differences. First, the reference pixels are those along the papilla contour (PC) rather than BM. Second, the lower bound of activation values for its interior pixels is $d_l^{a^{\mathrm{PS}}}$, instead of $d_l^{a^{\mathrm{Ep}}}$. Calculating the activation value for each pixel inside the papillary structure comprises the following four steps:

Step 1: Calculate the shortest distance between the current papillary pixel ($p^{\mathrm{PS}}(x_i, y_i)$) and N_{PC} pixels along the papilla contour ($p^{\mathrm{PC}}(x_j, y_j)$) as the distance between the interior and the contour of the papillary structure:

$$D\big(p^{\mathrm{PS}}(x_i, y_i)\big) = \min\left(Dis\left(p^{\mathrm{PS}}(x_i, y_i), \left\{p^{\mathrm{PC}}(x_j, y_j)\right\}_{j=1}^{N_{\mathrm{PC}}}\right)\right). \tag{1}$$

Step 2: Calculate the minimum and maximum distances from the interior of a papilla that has N_{PS} pixels to its contour:

$$d_{min}^{\mathrm{PS}} = \min\left(\left\{D\big(p^{\mathrm{PS}}(x_i, y_i)\big)\right\}_{i=1}^{N_{\mathrm{PS}}}\right), \tag{2}$$

$$d_{max}^{\mathrm{PS}} = \max\left(\left\{D\big(p^{\mathrm{PS}}(x_i, y_i)\big)\right\}_{i=1}^{N_{\mathrm{PS}}}\right). \tag{3}$$

Step 3: Determine the decay function of activation values regarding distance based on d_{min}^{PS}, d_{max}^{PS}, a^{PS}, and the lower bound of activation values for papillary pixels $d_l^{a^{\mathrm{PS}}}$:

$$A^{dis}\big(p^{\mathrm{PS}}(x_i,y_i)\big) = \frac{d_l^{a^{\mathrm{PS}}} - a^{\mathrm{PS}}}{d_{max}^{\mathrm{PS}} - d_{min}^{\mathrm{PS}}} D\big(p^{\mathrm{PS}}(x_i,y_i)\big) + a^{\mathrm{PS}} - \frac{d_l^{a^{\mathrm{PS}}} - a^{\mathrm{PS}}}{d_{max}^{\mathrm{PS}} - d_{min}^{\mathrm{PS}}} d_{min}^{\mathrm{PS}}. \quad (4)$$

Step 4: Integrate the pixel's grayscale value into the distance-based activation value $A^{dis}\big(p^{\mathrm{PS}}(x_i,y_i)\big)$ to get the final activation value $A\big(p^{\mathrm{PS}}(x_i,y_i)\big)$:

$$g_{avg}^{\mathrm{PS}} = \text{average}\left(\left\{G\big(p^{\mathrm{PS}}(x_i,y_i)\big)\right\}_{i=1}^{N_{\mathrm{PS}}}\right), \quad (5)$$

$$R_g\big(p^{\mathrm{PS}}(x_i,y_i)\big) = \min\left(\frac{G\big(p^{\mathrm{PS}}(x_i,y_i)\big)}{g_{avg}^{\mathrm{PS}}}, 1\right), \quad (6)$$

$$A\big(p^{\mathrm{PS}}(x_i,y_i)\big) = A^{dis}\big(p^{\mathrm{PS}}(x_i,y_i)\big) \times R_g\big(p^{\mathrm{PS}}(x_i,y_i)\big). \quad (7)$$

Pixels belonging to other regions in ectropion images, like background, condom, and the interspace between the condom and papillary structures, have their activation values set to 0.

2 Experiment Setups

The internal dataset was divided into five folds for cross-validation, ensuring that data from individual patients were exclusively allocated to either the training set or the test set to mitigate overfitting. The distribution of patients across different classes was approximately balanced in training and test sets for each fold. The five models trained on the internal data using five-fold cross-validation were also applied to two external datasets to further evaluate their generalizability.

All experiments were performed on an Inspur Server NF5280M5 workstation with an Intel Xeon Silver 4210 CPU (2.20 GHz) and a 48GB NVIDIA RTX A6000 GPU. The software environment comprised Ubuntu 18.04, Python 3.8.11, Pytorch 1.12.0, and scikit-learn 0.24.2.

The resolution of cervical OCT images was standardized to 512×1024 pixels. Data augmentation methods included random brightness and contrast adjustments, resize cropping, horizontal flipping, and Gaussian blurring. ResNet and ConvNeXt models were implemented using ResNet-18 with 11.18M parameters and ConvNeXt_Pico with 8.54M parameters for practical efficiency. The lower bounds for epithelial pixels' activation values $d_l^{a^{\mathrm{Ep}}}$, stromal pixels' activation values $d_l^{a^{\mathrm{St}}}$, and papillary pixels' activation values $d_l^{a^{\mathrm{PS}}}$ were set to 0.5, 0.3, and 0.1, respectively, according to pathologists' prior knowledge. The PAM alignment loss weight λ_a and the image classification loss weight λ_i were optimally set to 1. We trained the classification model for 35 epochs using the AdamW optimizer with cosine annealing weight decay, starting at 0.2 and decreasing to 0.05. The learning rate followed a cosine annealing schedule with a linear warm-up, beginning at $1e$-6, peaking at $1e$-4, and ending at $5e$-7, with a warm-up period over seven epochs and a batch size of 48.