

Prior Activation Map Guided Cervical OCT Image Classification

Qingbin Wang¹[0009-0002-8623-597X], Wai Chon Wong¹, Mi Yin², and Yutao Ma^{2,*}[0000-0003-4239-2009]

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

Abstract. Cervical cancer poses a severe threat to women’s health globally. As a non-invasive imaging modality, cervical optical coherence tomography (OCT) rapidly generates micrometer-resolution images from the cervix, comparable nearly to histopathology. However, the scarcity of high-quality labeled OCT images and the inevitable speckle noise impede deep-learning models from extracting discriminative features of high-risk lesion images. This study utilizes segmentation masks and bounding boxes to construct prior activation maps (PAMs) that encode pathologists’ diagnostic insights into different cervical disease categories in OCT images. These PAMs guide the classification model in producing reasonable class activation maps during training, enhancing interpretability and performance to meet gynecologists’ needs. Experiments using five-fold cross-validation demonstrate that the PAM-guided classification model boosts the classification of high-risk lesions on three datasets. Besides, our method enhances histopathology-based interpretability to assist gynecologists in analyzing cervical OCT images efficiently, advancing the integration of deep learning in clinical practice.

Keywords: Cervical Cancer · Optical Coherence Tomography · Image Classification · Visualization · Interpretability.

1 Introduction

Cervical cancer (CC), attributed to human papillomavirus infection, is the fourth most common cancer among women globally [13], with higher incidence and mortality in countries with low human development index [12]. Compared with routine CC screening techniques, such as the thinprep cytology test and colposcopy, optical coherence tomography (OCT) [4] has recently been proven to have obvious advantages in real-time, high-resolution imaging of human cervical tissue *in vivo* and *ex vivo*, comparable nearly to histopathology [2, 9, 10, 17]. These advantages promoted the clinical application of OCT in the early and accurate detection of cervical lesions [16].

However, most gynecologists are not yet well-versed in diagnostic features in cervical OCT images. There is a pressing need to develop computer-aided

diagnosis systems to help gynecologists accurately and efficiently interpret cervical OCT images [8]. Developing an effective deep-learning model for cervical OCT image analysis currently faces two key challenges. **(1) Scarcity of labeled data.** Labeling cervical OCT images requires a precise, one-to-one correlation with the pathology results of biopsy samples. This makes the labeling process extremely time-consuming. As a result, amassing a substantial dataset of high-quality, labeled cervical OCT images to facilitate model training has proven highly difficult [1,15]. **(2) Intrinsic speckle noise.** Cervical OCT images inherently contain coherent noise (speckle noise), similar to OCT for ophthalmology imaging. This noise obstructs deep-learning models from identifying distinctive features in high-risk cervical lesion images. It also causes models to overfit more readily to the noise, resulting in poor generalizability on different datasets.

The above challenges also hamper the interpretability of deep-learning models. For example, we found that GradCAM [11] visualizations of classification results tend to be unstable and occasionally inaccurate for the cervical OCT image classification. This makes it difficult for gynecologists to comprehend and trust the deep-learning model’s reasoning. To address these issues, in this study, we leverage the knowledge and skills of pathologists in analyzing cervical OCT images to construct prior activation maps (PAMs). PAMs, generated via segmentation masks and manually annotated bounding boxes, indicate the significance of each pixel in the input image during classification. By leveraging these custom activation maps to constrain those generated by the class activation mapping (CAM) series [5] during training, along with the standard classification loss, our model can maintain strong classification performance and better interpretability. Overall, our contributions can be summarized in three aspects:

- This is the first work to construct custom activation maps that encode pathologists’ prior knowledge and skills in cervical OCT image analysis. PAMs can produce more reasonable heat maps to provide better visual interpretability for gynecologists when overlaid on the original image.
- We constrain the classification model’s CAM-generated activation maps with PAMs during training. This ensures the model focuses on key features for classification and mitigates the impact of speckle noise, thereby achieving better classification performance and interpretability for high-risk lesions.
- Through experiments using five-fold cross-validation, the PAM-guided classification model demonstrates improved classification performance on three datasets. Additionally, the model achieves more stable and accurate interpretability based on histopathology, promoting its potential clinical use.

2 Data Collection

A multi-center clinical study [10] was conducted from August 2017 to December 2019 to evaluate the efficacy of OCT for detecting cervical diseases. The study recruited 785 subjects from five hospitals and one examination center in China. Two external validation sets were also used to assess our method’s generalizabil-

ity, collected using the same protocol as the multi-center study. Ethical approval was obtained from the ethics committees at each participating institution.

Table 1 presents the statistics of the experimental datasets. The internal dataset has 1,683 cervical OCT volumes from 785 subjects labeled according to the pathology results of biopsy specimens. The Huaxi dataset includes 760 labeled OCT volumes from 228 subjects at West China Hospital, Sichuan University. The Xiangya dataset consists of 278 labeled OCT volumes from 120 subjects at Xiangya Second Hospital, Central South University. All the OCT images were categorized into mild inflammation (MI), cyst (CY), ectropion (EP), high-grade squamous intraepithelial lesion (HSIL), and CC. As with previous work for binary classification [1, 8, 10, 15, 17], MI, CY, and EP were classified as negative, while HSIL and CC were classified as positive.

Table 1. Statistics of experimental datasets.

Dataset	Size	MI	CY	EP	HSIL	CC	Total
Internal	#Subjects	273	173	106	176	57	785
	#Volumes	780	302	184	283	134	1,683
	#Images	8,810	3,950	2,360	4,285	3,140	22,545
External (Huaxi)	#Subjects	130	46	12	38	2	228
	#Volumes	545	106	28	74	7	760
	#Images	5,450	1,060	280	740	70	7,600
External (Xiangya)	#Subjects	22	39	24	26	9	120
	#Volumes	75	80	63	40	20	278
	#Images	750	800	630	400	200	2,780

3 Method

3.1 Generating PAMs Based on Prior Knowledge

Figure 1(a) illustrates the process of generating prior activation maps and heat maps based on pathologists’ prior knowledge. For the three negative categories (MI, CY, and EP), which exhibit apparent layered structures or papillary-shaped features, we train a segmentation model based on HRNet [14] using 4,200 annotated images to segment the three image types. For the two positive categories (HSIL and CC), which have entirely lost their layered structures, we frame the lesion region by manually annotating bounding boxes. Then, we briefly introduce our method for producing PAMs for each image category based on segmentation masks or bounding boxes derived from experts’ prior knowledge and skills.

Cervical inflammation image. As shown in Figure 2(a), the basement membrane (BM) is the most crucial feature for diagnosing cervical inflammation. Therefore, the activation value a^{BM} of all pixels corresponding to BM is set to the maximum value of 1.0. Given a cervical inflammation image, the calculation

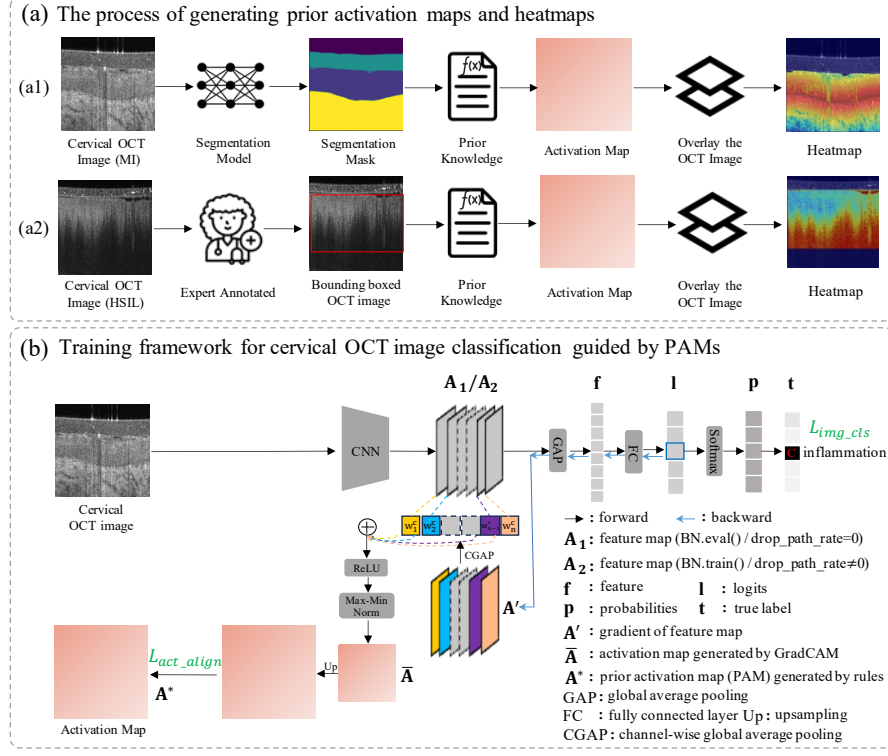


Fig. 1. The overall process of PAM-guided cervical OCT image classification.

of the activation value for each pixel corresponding to the epithelium comprises the following steps.

Step 1: A 7×7 average filter is applied to the image to reduce speckle noise.

Step 2: Calculate the shortest distance between the current epithelial pixel and all BM pixels as the distance between the epithelium and BM:

$$D(p^{\text{Ep}}(x_i, y_i)) = \min \left(Dis \left(p^{\text{Ep}}(x_i, y_i), \{p^{\text{BM}}(x_j, y_j)\}_{j=1}^{N_{\text{BM}}} \right) \right), \quad (1)$$

where $p^{\text{Ep}}(x_i, y_i)$ is an epithelial pixel with coordinates (x_i, y_i) , $\{p^{\text{BM}}(x_j, y_j)\}$ is the set of BM pixels, N_{BM} is the number of BM pixels, and $Dis(a, B)$ is the Euclidean distance between pixel a and all the pixels in set B .

Step 3: Calculate the minimum and maximum distances from the epithelium that has N_{Ep} pixels to BM:

$$d_{min}^{\text{Ep}} = \min \left(\{D(p^{\text{Ep}}(x_i, y_i))\}_{i=1}^{N_{\text{Ep}}} \right), \quad (2)$$

$$d_{max}^{\text{Ep}} = \max \left(\{D(p^{\text{Ep}}(x_i, y_i))\}_{i=1}^{N_{\text{Ep}}} \right). \quad (3)$$

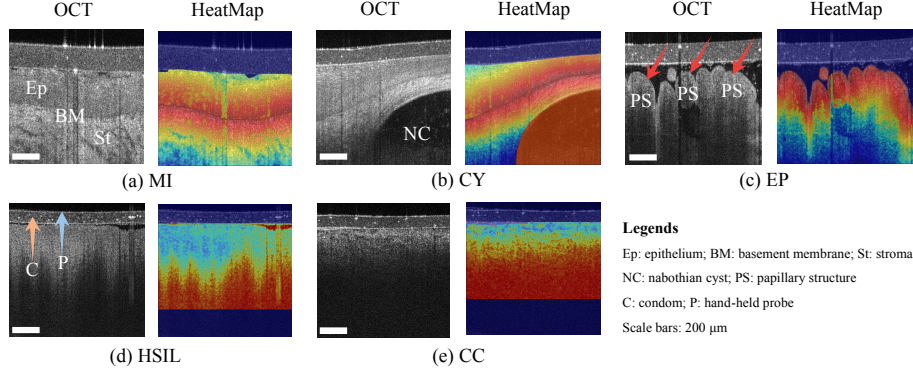


Fig. 2. Cervical OCT image examples of five categories and their corresponding heat maps generated based on prior knowledge.

Step 4: Calculate the activation value regarding decay distance based on d_{min}^{Ep} , d_{max}^{Ep} , a^{BM} , and the lower bound of epithelial activation values $d_l^{a^{\text{Ep}}}$:

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

Step 5: Integrate the pixel's grayscale value into the distance-based activation value $A^{dis}(p^{\text{Ep}}(x_i, y_i))$ to get the final activation value $A(p^{\text{Ep}}(x_i, y_i))$:

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

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

$$A(p^{\text{Ep}}(x_i, y_i)) = A^{dis}(p^{\text{Ep}}(x_i, y_i)) \times R_g(p^{\text{Ep}}(x_i, y_i)), \quad (7)$$

where $G(p^{\text{Ep}}(x_i, y_i))$ denotes the grayscale value of pixel $p^{\text{Ep}}(x_i, y_i)$, g_{avg}^{Ep} is the average grayscale value of epithelial pixels, and $R_g(p^{\text{Ep}}(x_i, y_i))$ denotes the scaling ratio of the pixel's activation value.

The calculation process of the activation value for each stromal pixel is similar to that used for epithelial pixels, except that the lower bound of activation values applies to stromal pixels (d_l^{St}) instead. The difference in grayscale value between epithelial and stromal pixels is also factored in when computing stromal pixels' activation values. Pixels belonging to other regions, like background and condom, have their activation values set to 0.

For the calculation of activation maps for cervical cyst and ectropion images, please refer to **Supplementary Material**.

HSIL and CC images. As shown in Figures 2(d) and 2(e), the high-risk lesion region is first manually enclosed within a bounding box by pathologists. A

simple linear function is designed to calculate the activation value based solely on the pixel's grayscale value for all lesion pixels within the bounding box.

$$A(p^{\text{LR}}(x_i, y_i)) = \frac{a_l^{\text{LR}} - a_u^{\text{LR}}}{g_{\max} - g_{\min}} G(p^{\text{LR}}(x_i, y_i)) + a_u^{\text{LR}} - \frac{a_l^{\text{LR}} - a_u^{\text{LR}}}{g_{\max} - g_{\min}} g_{\min}, \quad (8)$$

where a_l^{LR} and a_u^{LR} denote the lower and upper bounds of activation values for lesion pixels, $g_{\max}$ is the maximum grayscale value that is equal to 255, and $g_{\min}$ is the minimum grayscale value that is equal to 0.

3.2 Training Process

We then train a classification model after obtaining cervical OCT images' PAMs. Figure 1(b) depicts the training framework, including two forward processes.

First forward process. Given an image $I \in \mathbb{R}^{H \times W \times 3}$, features are extracted by a convolutional neural network (CNN) to form a feature map $\mathbf{A}_1 \in \mathbb{R}^{\frac{H}{32} \times \frac{W}{32} \times K}$. To ensure the feature map is consistent with what the classification model generates during inference, different operations are applied based on the CNN architecture: ResNet [3]: *Set all batch normalization layers to the eval() mode*; ConvNeXt [6]: *Set the drop path rate to 0 for all blocks*. Following the same approach as the CAM series, an activation map $\bar{\mathbf{A}} \in \mathbb{R}^{\frac{H}{32} \times \frac{W}{32}}$ is then obtained as

$$w_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial l^c}{\partial \mathbf{A}_{1ij}^k}, \quad (9)$$

$$\bar{\mathbf{A}} = \text{norm}_{\max\text{-min}} \left(\text{ReLU} \left(\sum_k w_k^c \mathbf{A}_1^k \right) \right), \quad (10)$$

where l^c is the predicted logit for class c , $\mathbf{A}_{1ij}^k$ is the value for the (i, j) position in the k -th channel of the feature map, $Z = \frac{H}{32} \times \frac{W}{32}$ is the total number of elements contained in a channel, and w_k^c is the weight of class c for channel k .

Second forward process. *Reset the batch normalization layers back to the train() mode for ResNet; reactivate the drop path rate across all blocks for ConvNeXt.* Then, we extract a new feature map $\mathbf{A}_2 \in \mathbb{R}^{\frac{H}{32} \times \frac{W}{32} \times K}$ to generate the final class probability distribution $\mathbf{p} \in \mathbb{R}^{1 \times C}$, calculated as follows:

$$\mathbf{p} = \text{softmax}(\text{FC}(\text{GAP}(\mathbf{A}_2))), \quad (11)$$

where $\text{GAP}(\cdot)$ denotes the operation of global average pooling, and $\text{FC}(\cdot)$ denotes a fully-connected layer operation.

Loss function. When the two forward processes end, we design an activation map alignment loss $\mathcal{L}_{\text{act_align}}$ and an image classification loss $\mathcal{L}_{\text{img_cls}}$ as

$$\mathcal{L}_{\text{act_align}} = \text{MSE}(\text{Up}(\bar{\mathbf{A}}), \mathbf{A}^*), \quad (12)$$

where $\text{MSE}(\cdot)$ denotes the mean squared error function, $\text{Up}(\cdot)$ denotes an up-sampling operation (here we use bilinear interpolation for $32\times$ upsampling), and $\mathbf{A}^* \in \mathbb{R}^{H \times W}$ is a PAM obtained in Section 3.1.

$$\mathcal{L}_{img_cls} = \text{CE}(\mathbf{p}, t), \quad (13)$$

where $\text{CE}(\cdot)$ is the cross-entropy loss, and t is the class label of image I . The final loss is then defined as

$$\mathcal{L} = \lambda_a \mathcal{L}_{act_align} + \lambda_i \mathcal{L}_{img_cls}, \quad (14)$$

where λ_a and λ_i are the weights of the activation map alignment loss $\mathcal{L}_{act_align}$ and the image classification loss $\mathcal{L}_{img_cls}$, respectively.

4 Experiment Setups and Results

4.1 Experiment Setups

ResNet and ConvNeXt models were implemented using ResNet-18 with 11.18M parameters and ConvNeXt_Pico with 8.54M parameters for practical efficiency. OCT image resolution was standardized to 512×1024 . The lower bounds for the epithelial distance decay activation $d_t^{a^{ep}}$, stromal distance decay activation $d_t^{a^{st}}$, and the distance decay activation for pixels within papillary structures $d_t^{a^{ps}}$ were set to 0.5, 0.3, and 0.1, respectively. Both the activation map alignment loss weight λ_a and the image classification loss weight λ_i were set to 1. We trained the models for 35 epochs using the AdamW [7] optimizer with a batch size 48. The source code is available at https://github.com/ssea-lab/AMGuided_Cervical_OCT_Classification. For more details on experiment setups, please refer to **Supplementary Material**.

4.2 Results

Classification Results on Three Datasets. Table 2 presents image classification results using five-fold cross-validation on three datasets. Adding a PAM alignment loss (denoted as “+PAM”) improved the overall classification performance regarding binary accuracy and AUC across ResNet and ConvNeXt models on three different datasets. Remarkably, the two models achieved notable increases in sensitivity across the three datasets, indicating improved detection of positive cases. The highest sensitivity gains ($> 4.70\%$) were seen on the Xiangya dataset. However, specificity decreased slightly ($< 0.80\%$) on the Huaxi dataset for both models. This rise in false positives suggests the need to refine PAMs, especially for HSIL and CC, to enhance sensitivity while mitigating decreases in specificity. Overall, our method shows promise for improving sensitivity while maintaining high specificity for CNN-based cervical OCT image classification.

Table 2. Classification results of the PAM-guided model on three datasets (mean $\pm$ std).

Dataset	Model	Five-class Accuracy (%)	Binary Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
Internal	ResNet	88.45 $\pm$ 2.00	96.21 $\pm$ 0.99	94.33 $\pm$ 2.91	96.88 $\pm$ 0.74	0.9907 $\pm$ 0.0135
	ResNet	89.60 $\pm$ 2.42	97.48 $\pm$ 1.20	94.70 $\pm$ 5.84	98.24 $\pm$ 0.98	0.9914 $\pm$ 0.0162
	+PAM	(1.15)	(1.27)	(0.37)	(1.36)	(0.07%)
	ConvNeXt	91.88 $\pm$ 2.29	98.04 $\pm$ 0.87	96.02 $\pm$ 4.21	98.53$\pm$0.99	0.9928 $\pm$ 0.0206
	ConvNeXt	92.65$\pm$2.00	98.19$\pm$0.63	97.26$\pm$2.73	98.29 $\pm$ 1.10	0.9942$\pm$0.0185
	+PAM	(0.77)	(0.15)	(1.24)	(-0.24)	(0.14%)
Huaxi	ResNet	91.60 $\pm$ 0.65	95.79 $\pm$ 0.35	74.50 $\pm$ 6.78	98.33$\pm$0.51	0.9807 $\pm$ 0.0137
	ResNet	91.89 $\pm$ 1.02	96.04$\pm$0.82	77.63 $\pm$ 2.93	98.23 $\pm$ 1.34	0.9821$\pm$0.0084
	+PAM	(0.29)	(0.25)	(3.13)	(-0.10)	(0.14%)
	ConvNeXt	92.09$\pm$0.35	95.25 $\pm$ 0.34	71.14 $\pm$ 5.90	98.13 $\pm$ 0.38	0.9683 $\pm$ 0.0258
	ConvNeXt	91.73 $\pm$ 1.54	95.35 $\pm$ 0.71	78.55$\pm$3.01	97.35 $\pm$ 0.76	0.9711 $\pm$ 0.0152
	+PAM	(-0.36)	(0.10)	(7.41)	(-0.78)	(0.28%)
Xiangya	ResNet	87.87 $\pm$ 0.24	92.81 $\pm$ 0.76	67.00 $\pm$ 3.60	99.92$\pm$0.10	0.9739 $\pm$ 0.0186
	ResNet	87.69 $\pm$ 2.05	94.09 $\pm$ 0.45	74.92 $\pm$ 2.28	99.36 $\pm$ 0.43	0.9863 $\pm$ 0.0197
	+PAM	(-0.18)	(1.28)	(7.92)	(-0.56)	(1.24%)
	ConvNeXt	88.18 $\pm$ 0.56	94.31 $\pm$ 0.49	75.00 $\pm$ 1.96	99.62 $\pm$ 0.12	0.9901 $\pm$ 0.0092
	ConvNeXt	88.96$\pm$0.27	95.42$\pm$0.26	79.71$\pm$0.75	99.75 $\pm$ 0.22	0.9912$\pm$0.0058
	+PAM	(0.78)	(1.11)	(4.71)	(0.13)	(0.11%)

Visualization of Classification Results. Figure 3 demonstrates the improved visual interpretability of classification results after using the PAM alignment loss. Neither model identified BM’s exact location in the MI image. Instead, ResNet+PAM highlighted BM, and ConvNeXt+PAM accurately focused on the epithelium, stroma, and BM, resembling the ground truth heat map. Although the models roughly located the nabothian cysts in the CY image, they neglected BM. ConvNeXt+PAM generated a better, more ground truth-like visualization result than ResNet+PAM. For the EP image, it was tough for the noise-affected models to accurately detect the hyper-scattering papillary structures. ResNet+PAM improved the contour clarity of papillae, while ConvNeXt+PAM accurately located those papillae and was almost unaffected by speckle noise. For the HSIL and CC images, the models seem to wrongly focus on the condom while neglecting the lesion region where the light intensity decayed rapidly. Compared with ResNet+PAM, ConvNeXt+PAM achieved more accurate focus. Overall, adding the PAM alignment loss to ResNet and ConvNeXt models greatly enhanced their visualization accuracy and stability, with ConvNeXt+PAM’s results closely matching ground truth.

5 Conclusion

In this study, we leverage pathologists’ prior knowledge of diagnosing cervical OCT images to generate custom activation maps called PAM. By aligning the classification models’ activation maps to PAMs during training via the CAM series, CNN-based models are guided to focus on relevant regions for classification. The experimental results demonstrate that incorporating the PAM alignment

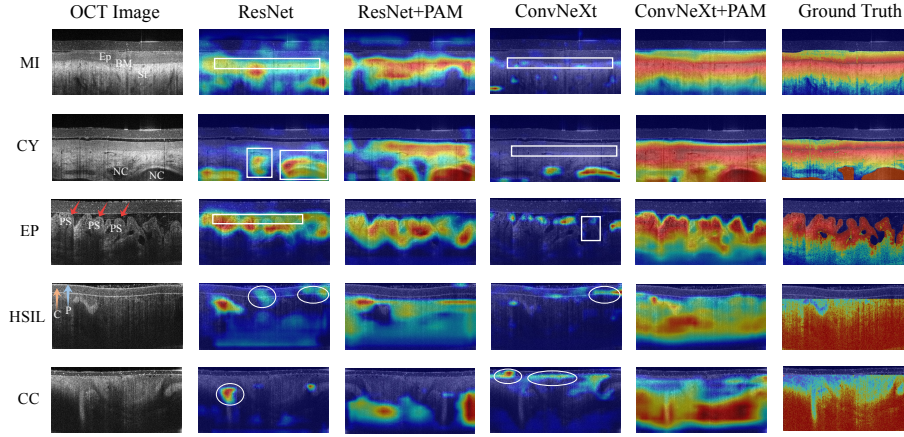


Fig. 3. Heat maps generated by GradCAM [11] for CNN models trained with different methods (“+PAM” denotes a classification model trained with the PAM alignment loss, and “Ground Truth” means the heat map generated based on prior knowledge).

loss improves model performance for detecting positive cases on three datasets from different sources. Additionally, ResNet and ConvNeXt models can generate more accurate and reasonable interpretability with heat maps to assist gynecologists in cervical OCT diagnosis. In the future, we will extend this method to Transformers and non-gradient-based visualization techniques to balance gains in sensitivity with the maintenance of high specificity.

Acknowledgments. This work was supported by the Key Research and Development Program of Hubei Province in China (Grant No. 2022BCA010).

Disclosure of Interests. The authors have no competing interests to declare relevant to this article’s content.

References

1. Chen, K., Wang, Q., Ma, Y.: Cervical optical coherence tomography image classification based on contrastive self-supervised texture learning. *Med. Phys.* **49**(6), 3638–3653 (2022)
2. Gallwas, J., Jalilova, A., Ladurner, R., Kolben, T.M., Kolben, T., Ditsch, N., Homann, C., Lankenau, E., Dannecker, C.: Detection of cervical intraepithelial neoplasia by using optical coherence tomography in combination with microscopy. *J. Biomed. Opt.* **22**(1), 016013 (2017)
3. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of IEEE CVPR 2016*. pp. 770–778. IEEE Computer Society (2016)
4. Huang, D., Swanson, E.A., Lin, C.P., Schuman, J.S., Stinson, W.G., Chang, W., Hee, M.R., Flotte, T., Gregory, K., Puliafito, C.A., et al.: Optical coherence tomography. *Science* **254**(5035), 1178–1181 (1991)

5. Jung, H., Oh, Y.: Towards better explanations of class activation mapping. In: Proceedings of IEEE/CVF ICCV 2021. pp. 1316–1324. IEEE (2021)
6. Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S.: A convnet for the 2020s. In: Proceedings of IEEE/CVF CVPR 2022. pp. 11966–11976. IEEE (2022)
7. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: Proceedings of ICLR 2019. p. 939. OpenReview.net (2019)
8. Ma, Y., Xu, T., Huang, X., Wang, X., Li, C., Jerwick, J., Ning, Y., Zeng, X., Wang, B., Wang, Y., et al.: Computer-aided diagnosis of label-free 3-D optical coherence microscopy images of human cervical tissue. *IEEE Trans. Biomed. Eng.* **66**(9), 2447–2456 (2019)
9. Paczos, T., Bonham, A., Canavesi, C., Rolland, J.P., O’Connell, R.: Near-histologic resolution images of cervical dysplasia obtained with gabor domain optical coherence microscopy. *J. Low. Genit. Tract. Dis.* **25**(2), 137 (2021)
10. Ren, C., Zeng, X., Shi, Z., Wang, C., Wang, H., Wang, X., Zhang, B., Jiang, Z., Ma, H., Hu, H., et al.: Multi-center clinical study using optical coherence tomography for evaluation of cervical lesions in-vivo. *Sci. Rep.* **11**(1), 7507 (2021)
11. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Int. J. Comput. Vis.* **128**(2), 336–359 (2020)
12. Singh, D., Vignat, J., Lorenzoni, V., Eslahi, M., Ginsburg, O., Lauby-Secretan, B., Arbyn, M., Basu, P., Bray, F., Vaccarella, S.: Global estimates of incidence and mortality of cervical cancer in 2020: a baseline analysis of the WHO global cervical cancer elimination initiative. *Lancet Glob. Health* **11**(2), e197–e206 (2023)
13. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F.: Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**(3), 209–249 (2021)
14. Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., et al.: Deep high-resolution representation learning for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **43**(10), 3349–3364 (2020)
15. Wang, Q., Chen, K., Dou, W., Ma, Y.: Cross-attention based multi-resolution feature fusion model for self-supervised cervical OCT image classification. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **20**(4), 2541–2554 (2023)
16. Xiao, X., Yan, L., Yang, X., Zhou, Z., Shi, L., Fu, C.: Optical coherence tomography can reduce colposcopic referral rates in patients with high-risk human papillomavirus. *J. Low. Genit. Tract. Dis.* **27**(4), 324–330 (2023)
17. Zeng, X., Zhang, X., Li, C., Wang, X., Jerwick, J., Xu, T., Ning, Y., Wang, Y., Zhang, L., Zhang, Z., et al.: Ultrahigh-resolution optical coherence microscopy accurately classifies precancerous and cancerous human cervix free of labeling. *Theranostics* **8**(11), 3099 (2018)