Feature Library: A Benchmark for Cervical Lesion Segmentation

Yuexiang Li, Jiawei Chen, Kai Ma and Yefeng Zheng

Tencent Jarvis Lab, Shenzhen, China

Abstract: Cervical cancer causes the fourth most cancer-related deaths of women worldwide. One of the most commonly-used clinical tools for the diagnosis of cervical intraepithelial neoplasia (CIN) and cervical cancer is colposcopy examination. However, due to the challenging imaging conditions such as light reflection on the cervix surface, the clinical accuracy of colposcopy examination is relatively low. In this paper, we propose a computer-aided diagnosis (CAD) system to accurately segment the lesion areas (i.e., CIN and cancer) from colposcopic images, which can not only assist colposcopists for clinical decision, but also provide the guideline for the location of biopsy sites. Furthermore, to well-train and evaluate our deep learning network, we collect a large-scale colposcopic image dataset for **C**erv**I**cal lesio**N** s**E**g**M**ent**A**tion (**CINEMA**), consisting of 34,337 images from 9,652 patients. The lesion areas in the colposcopic images are manually annotated by experienced colposcopists. Extensive experiments are conducted on the CINEMA dataset, which demonstrate the effectiveness of our feature library dealing with cervical lesions of varying sizes.

❑ Introduction:

Cervical cancer contributes the fourth highest number of deaths in female cancers, carrying high risks of morbidity and mortality \cite{PengL01}. Over 88\% of deaths from cervical cancer occur in low- and middle-income countries (LMICs), where gender discrimination and extreme poverty severely limit a woman's choice to seek care \cite{GinsburgOM01}. Due to the long period from precancerous cervical stage (i.e., cervical intraepithelial neoplasia (CIN)) to invasive cancer, the early identification of CIN can significantly decrease the number of deaths caused by cervical cancer.

In this paper, we propose a deep-learning-based CAD system (namely Feature Library) for the diagnosis of CIN and cervical cancer. The proposed system can not only identify the abnormal patients, but also segment the cervical lesion areas to guide the accurate location of biopsy sites for colposcopists. Furthermore, we notice that there are few of existing frameworks proposed for the task of cervical lesion segmentation, due to the difficulty of data collection and annotation. To this end, we collect a large-scale annotated colposcopic image dataset, namely CINEMA, which consists of 34,337 images captured from 9,652 patients, for the training and evaluation of deep learning networks.

❑ Feature Library:

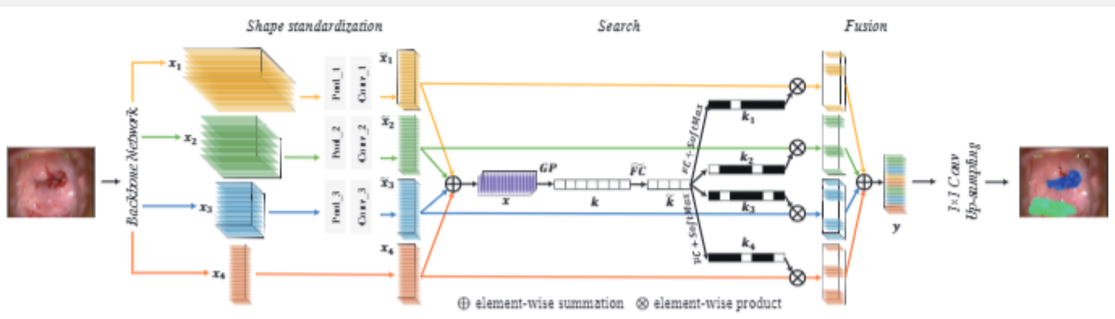

Fig. 2. The pipline of the library searching module (Conv_1: Conv 1×1, in_channel 256, out_channel 2048; Conv_2: Conv 1×1, in_channel 512, out_channel 2048; Conv_3: Conv 1×1, in_channel 1024, out_channel 2048; Pool_1: MaxPool 8×8; Pool_2: MaxPool 4×4; Pool_3: MaxPool 2×2).

Shape standardization. The colored volumes (x1, x2, x3, x4) in Fig. 2 are the sets of feature maps yielded by different stages of ResNet-50 (i.e., conv2_3}, conv3_4, conv4_6 and conv5_3 according to ResNet-50). Since those feature maps have different sizes and numbers of channels, we first need to transform them to a uniform shape for the following search module, which consists of element-wise summation. The max pooling and 1 × 1 convolution are adopted for this purpose, as shown in Table 1.

Table 1. The detailed information of max pooling and convolutional layers used by the shape standardization module. The kernel size of max pooling and convolutional layers is listed. H, W and C are the height, width and number of channels, respectively.

Input Size	Max Pooling	Convolution	Output
$x_1 \in \mathbb{R}^{H_1 \times W_1 \times C_1}$	$(\frac{H_1}{H_4}, \frac{W_1}{W_4})$	$(1, 1, C_4)$	$\tilde{x}_1 \in \mathbb{R}^{H_4 \times W_4 \times C_4}$
$x_2 \in \mathbb{R}^{H_2 \times W_2 \times C_2}$	$(\frac{H_2}{H_4}, \frac{W_2}{W_4})$	$(1, 1, C_4)$	$\tilde{x}_2 \in \mathbb{R}^{H_4 \times W_4 \times C_4}$
$x_3 \in \mathbb{R}^{H_3 \times W_3 \times C_3}$	$(\frac{H_3}{H_4}, \frac{W_3}{W_4})$	$(1, 1, C_4)$	$\tilde{x}_3 \in \mathbb{R}^{H_4 \times W_4 \times C_4}$

Search and Fusion. Several studies have demonstrated that self-gating attention mechanism is an effective and lightweight solution to regulate the information flow of feature maps. In this regard, we integrate the self-gating approach into our search module to globally select the useful features among backbone network and adaptively construct the robust feature representation for cervical lesion segmentation. Please refer to our paper for the details of search and fusion step.

❑ CINEMA Dataset:

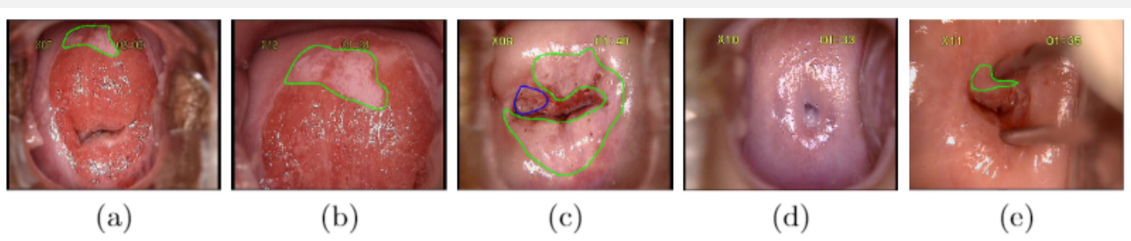

Fig. 1. Examplar colposcopic images in our CINEMA dataset. The lesion areas of CIN and cervical cancer are annotated with green and blue contours, respectively. (a) CIN patient under colposcopy. (b) Zoom-in view of the potential lesion areas of (a). (c) The patient may have the CIN and cancer areas at the same time. There are some difficulties for accurate segmentation of cervical lesion areas, such as the false-positive caused by the light reflection on normal cervix surface (d) and the occlusion caused by artifacts (e).

The CINEMA dataset consists of 34,337 colposcopic images from 9,652 patients (63 normal, 9,227 CIN, and 362 cancer cases), which are collected by the collaborative hospital. The colposcopic images have a uniform size 640 × 480 pixels. The cervical lesions, which can be categorized to CIN and cancer according to the pathological reports, are annotated by the experienced colposcopists. As shown in Fig. 1 (c), a colposcopic image may contain lesion areas of different categories. We randomly separate the dataset to training, validation and test sets, according to the ratio of 70:10:20. To our best knowledge, this is currently the largest dataset for cervical lesion segmentation.

❑ Experimental Results:

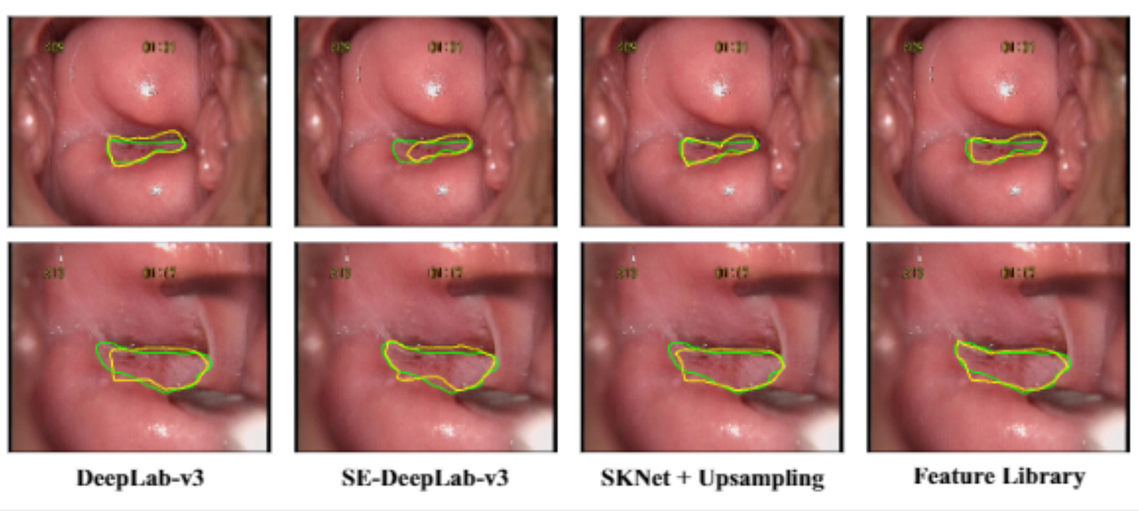

Table 2. Comparison of different frameworks in terms of DSC (%) and network parameters (Params). (F. L.—Feature Library)

	Validation			Test			Params
	CIN	Cancer	mDSC	CIN	Cancer	mDSC	(million)
ResNet-50 [7]	-	-	-	-	-	-	25
ResU-Net [7, 19]	70.79	84.26	77.53	68.12	82.92	75.52	48
SE-ResU-Net [7, 19, 8]	71.10	85.17	78.14	69.44	84.42	76.93	51
PSPNet [25]	67.37	83.51	75.44	64.48	80.43	72.46	89
SE-PSPNet [25, 8]	69.03	84.42	76.73	66.21	83.48	74.85	92
DeepLab-v3 [3]	68.65	83.99	76.32	65.36	81.50	73.43	64
SE-DeepLab-v3 [3, 8]	70.14	85.02	77.58	67.88	83.24	75.56	67
SKNet + Upsampling [14]	67.77	88.84	77.81	67.22	83.69	75.45	28
F. L. w/ (x_3, x_4) (Ours)	73.93	87.05	80.49	71.35	84.80	78.08	31
F. L. w/ (x_2, x_3, x_4) (Ours)	**74.82**	**88.75**	**81.79**	**72.97**	**86.12**	**79.55**	35
F. L. w/ (x_1, x_2, x_3, x_4) (Ours)	74.19	88.21	81.20	71.85	85.28	78.57	37
w/o Attention Mechanism*							
F. L. w/ (x_3, x_4)	70.69	84.19	77.44	68.49	81.86	75.18	25
F. L. w/ (x_2, x_3, x_4)	72.30	85.76	79.03	70.88	83.34	77.11	26
F. L. w/ (x_1, x_2, x_3, x_4)	71.02	84.86	77.94	69.52	82.07	75.80	27

* Using element-wise summation directly for feature fusion.