A segmentation model to detect cevical lesions based on machine learning of colposcopic images

Zhen Li [a,1], Chu-Mei Zeng [b,1], Yan-Gang Dong [c,1], Ying Cao [d,1], Li-Yao Yu [d,1], Hui-Ying Liu [b,1], Xun Tian [d], Rui Tian [e], Chao-Yue Zhong [e], Ting-Ting Zhao [e], Jia-Shuo Liu [b], Ye Chen [b], Li-Fang Li [b], Zhe-Ying Huang [b], Yu-Yan Wang [b], Zheng Hu [a], Jingjing Zhang [a,****], Jiu-Xing Liang [c,***], Ping Zhou [f,**], Yi-Qin Lu [g,*]

[a] Department of Gynecological Oncology, Zhongnan Hospital of Wuhan University, Wuhan, Hubei, 430071, China
[b] Department of Obstetrics and gynecology, the First Affiliated Hospital, Sun Yat-sen University, Guangzhou, Guangdong, 510062, China
[c] Institute for Brain Research and Rehabilitation, the South China Normal University, Guangzhou, Guangdong, 510631, China
[d] Department of Obstetrics and Gynecology, Academician expert workstation, The Central Hospital of Wuhan, Tongji Medical College Huazhong University of Science and Technology, Wuhan, Hubei, 430014, China
[e] the Generulor Company Bio-X Lab, Zhuhai, Guangdong, 519060, China
[f] Department of Gynecology, Dongguan Maternal and Child Hospital, Dongguan, Guangdong, 523057, China
[g] Department of Gynecology, Dongzhimen Hospital, Beijing University of Chinese Medicine, Beijing, 101121, China

A R T I C L E   I N F O

A B S T R A C T

Background: Semantic segmentation is crucial in medical image diagnosis. Traditional deep convolutional neural networks excel in image classification and object detection but fall short in segmentation tasks. Enhancing the accuracy and efficiency of detecting high-level cervical lesions and invasive cancer poses a primary challenge in segmentation model development.
Methods: Between 2018 and 2022, we retrospectively studied a total of 777 patients, comprising 339 patients with high-level cervical lesions and 313 patients with microinvasive or invasive cervical cancer. Overall, 1554 colposcopic images were put into the DeepLabv3+ model for learning. Accuracy, Precision, Specificity, and mIoU were employed to evaluate the performance of the model in the prediction of cervical high-level lesions and cancer.
Results: Experiments showed that our segmentation model had better diagnosis efficiency than colposcopic experts and other artificial intelligence models, and reached Accuracy of 93.29 %, Precision of 87.2 %, Specificity of 90.1 %, and mIoU of 80.27 %, respectively.
Conclusion: The DeepLabv3+ model had good performance in the segmentation of cervical lesions in colposcopic post-acetic-acid images and can better assist colposcopists in improving the diagnosis.

 * Corresponding authors.
 ** Corresponding author.
 *** Corresponding authors.
 **** Corresponding author.
   E-mail addresses: zhangjingjinggangda@126.com (J. Zhang), liangjiuxing@m.scnu.edu.cn (J.-X. Liang), 2045214580@qq.com (P. Zhou), k_lyq@sina.com (Y.-Q. Lu).
   [1] These authors contributed equally to this work and share first authorship.

1. Introduction

Cervical cancer, the most common gynecological malignant tumor, exhibits a well-defined and gradual precancerous progression from human papillomavirus (HPV) infection status, low-grade squamous intraepithelial lesion (LSIL) to high-grade squamous intra-epithelial lesion (HSIL) [1]. Through decades of progression and early cervical cancer screening, it has become known as one of the most preventable and treatable cancers [2]. The final critical step in the screening is the colposcopic biopsy, which involves optical imaging to magnify the lesions and tests such as acetic-acid and iodine staining to highlight them. Biopsy, being the gold standard, confirms suspicious HSIL or microscopic invasive cancer and determines subsequent therapeutic options [3]. However, the accuracy of colposcopic diagnosis relies on colposcopists' subjective assessment and requires years of practical experience. Identifying minor or multifocal lesions can be challenging, especially in remote areas where experienced colposcopists are scarce [4]. Consequently, numerous researchers are attempting to use artificial intelligence (AI) to analyze typical colposcopic pictures of different cervical disorders to achieve automated diagnoses.

In recent years, the significant advancements in Deep Convolutional Neural Network (DCNN) have propelled artificial intelligence colposcopy diagnosis [4]. Popular DCNNs such as Visual Geometry Group Net (VGGNet) [5], Residual Network (ResNet) [6], InceptionNet [7] have initially been tested for image classification and object detection purposes [8], for example, distinguishing LSIL and HSIL or predicting real-time binary results of cervical images as normal or abnormal [9,10]. However, these models lack the ability to predict specific biopsy areas due to a lack of detailed spatial information [4]. Semantic segmentation models are the improved version of previous DCNNs with fully connected layers replaced by convolution layers [10]. They assign semantic labels to each pixel in an image, capturing contextual information at multiple scales and accurately predicting the biopsy sites [11]. Several challenges exist in colposcopic images segmentation: low contrast, nonuniformity, and unique characteristics in colposcopic images; complex and variable shapes of the cervix, sometimes obscured by blood and leucorrhea; and inconspicuous early cervical lesions [12]. Therefore, researchers have proposed various modified models to get more thorough segmentation results.

The model adopted in this paper, DeepLabv3+, is the fourth generation of the DeepLab series proposed by Google [13–15]. DeepLabv3+ incorporates a streamlined and efficient Decoder module to achieve precise boundaries and faster computation [16]. Additionally, it replaces the original ResNet-101 backbone with Xception, resulting in a reduction of parameters by one-third [17]. Given these performance enhancements, it can significantly aid colposcopists in accurately delineating biopsy regions through semantic segmentation of colposcopic images, thereby enhancing the detection rate of HSIL and invasive cancer.

2. Results

2.1. Evaluation value

The pretrained model using only saline images yielded unsatisfactory results. The low specificity (56.60 %) and mean intersection over union (mIoU) (52.22 %) indicated limited overlap between the ground truth and the predicted segmentation. To avoid generating excessive errors in distinguishing between lesions and normal metaplastic squamous epithelium, we exclusively utilized post-acetic acid images to feed the model in the formal training and validation, given its higher specificity (90.10 %) and mIoU (80.27 %). Table 1 presents the distribution of Accuracy, Precision, Specificity, and mIoU for both saline images and post-acetic-acid images in the DeepLabv3+ model. The specific calculation methods for Accuracy, Precision, Specificity, and mIoU are provided in the numbered equations (1)–(4) in the 'Evaluation criteria' section of the Materials and Methods.

2.2. The best epoch

All networks underwent training for 50 epochs to enhance the model's ability to generalize. Fig. 1A&B depicted the training and validation loss and score curve of DeepLabv3+ as epochs progressed. Around the 30th epoch, the loss function curves for the training and verification sets began to converge. The calculation methods of loss and score are shown in (5) and (8). In the 35th epoch, the training set achieved a maximum score of 0.842, while the validation set reached its highest score of 0.8 in the 42nd epoch. As a result, the model selected as the best one was the 42nd epoch model, with the highest score in the validation set, and it underwent evaluation

Table 1
Distribution of Accuracy, Precision, Specificity and mIoU of saline images and post-acetic-acid images in DeepLabv3+.

	epochs	Accuracy(%)	Precision(%)	Specificity(%)	mIoU(%)
Saline Images	41st	90.75	72.71	68.70	60.13
	42nd	**91.35**	**81.87**	**56.60**	**52.22**
	43rd	91.14	76.02	58.9	53.99
	44th	91.24	76.37	60.00	54.96
	45th	91.23	75.76	61.00	55.71
Post-acetic-acid Images	41st	91.40	83.40	90.50	76.86
	42nd	**93.29**	**87.20**	**90.10**	**80.27**
	43rd	92.58	85.75	89.80	78.78
	44th	92.93	86.35	90.20	79.59
	45th	92.85	86.14	90.30	79.46

on the test set, yielding a score of 0.8.

At the 42nd epoch, the model attained the highest levels of Accuracy of 93.29 %, Precision of 87.2 %, Specificity of 90.1 %, and mIoU of 80.27 %, respectively, as shown in Table 1.

2.3. Feature visualization

The objective of the proposed approach was to generate accurate heat maps for lesion segmentation and distinguish between negative and positive cervigrams. Fig. 2 displayed four cervigrams along with the corresponding heat maps generated by DeepLabv3+ and the ground truth for the most severe lesions in these images. The colors in the heat maps represent the probability of the cervical skin belonging to the lesion. Red areas indicated a high probability, while blue areas indicated a low probability. The heat maps provide clear evidence that the regions identified as the most probable biopsy locations in the ground truth exhibit a high degree of alignment with the areas of focus detected by the DeepLabv3+ model. The observed alignment between the ground truth and the model's attention demonstrates the accurate concentration of the DeepLabv3+ model on relevant regions.

2.4. Comparison with experts

Table 2 demonstrated that DeepLabv3+ outperformed colposcopy experts with a significantly higher Accuracy of 93.29 % compared to the experts' Accuracy of 78.02 %. Additionally, the model achieved a remarkable Sensitivity of 97.92 % while the experts' Sensitivity was 73.39 %. However, it should be noted that the model occasionally misclassified LSIL or severe inflammation regions as HSIL/Cancer, leading to a higher false positive rate.

3. Discussion

In clinical experience, HSIL is the dividing point for clinical treatment. The detection rate of HSIL must be greater than 65 % [2] in order for colposcopy to meet the quality control requirement. Our artificial intelligence colposcopic diagnostic model achieved an accuracy rate of 93.29 %, enabling more precise localization of biopsy sites and benefiting colposcopists in detecting HSIL or invasive cancer, especially lesions in overlooked or challenging areas. Also, by effectively discerning non-lesion areas, our machine facilitates the reduction of minor yet "troublesome" side effects associated with excessive cervical treatment, including cervical adhesion and recurring cervicitis.

When comparing colposcopic learning models published in the past two years, it becomes apparent that recent model updates have resulted in significant improvements in accuracy (Table 3). Early models such as VGGNet [5], ResNet [6], InceptionNet [7], and EfficientNet [18] were traditional DCNN encoders, primarily designed for object classification. The semantic segmentation network consists not only of a front-end DCNN encoder but also a back-end decoder [10], and the decoder facilitates the reconstruction of high-resolution output, incorporates skip connections, and consequently helps the network gain the ability to capture both local and global context [16]. Its superiority over classification models lies in its ability to distinguish between normal and pathological regions within an image by learning from each individual pixel. With the advantage of the segmentation model, we were spared from collecting images from normal population, significantly reducing the workload without compromising the accuracy of the machine learning.

The published segmentation model achieved accuracy of 87.7–93.04 % and sensitivity of 78.02–95.60 %, surpassing colposcopists'
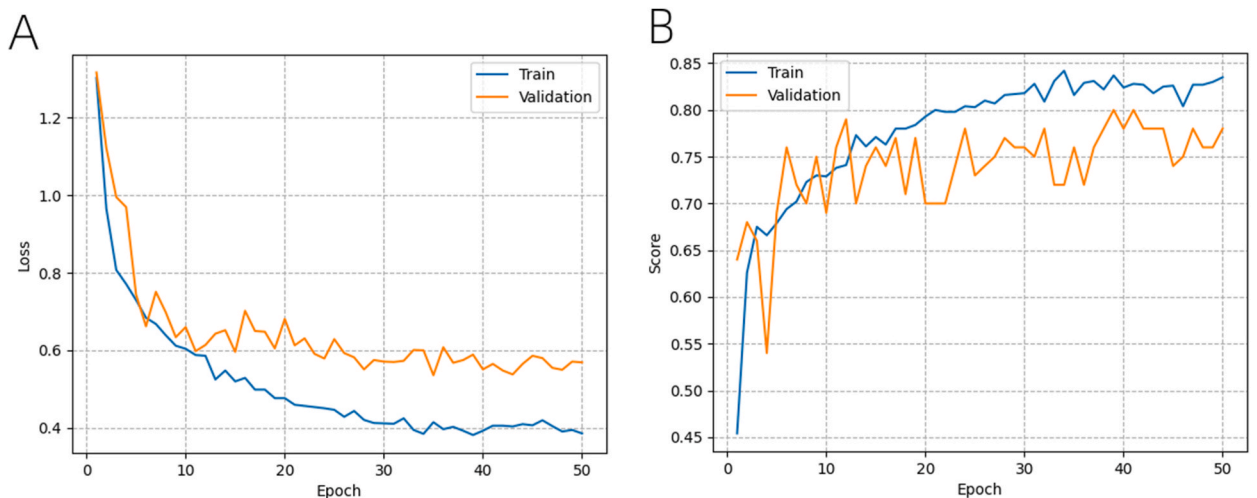

Fig. 1. Training and validation loss and score curve in DeepLabv3+.
(Fig. 1A. The loss curve of training and validation; Fig. 1B. The score curve of training and validation.).
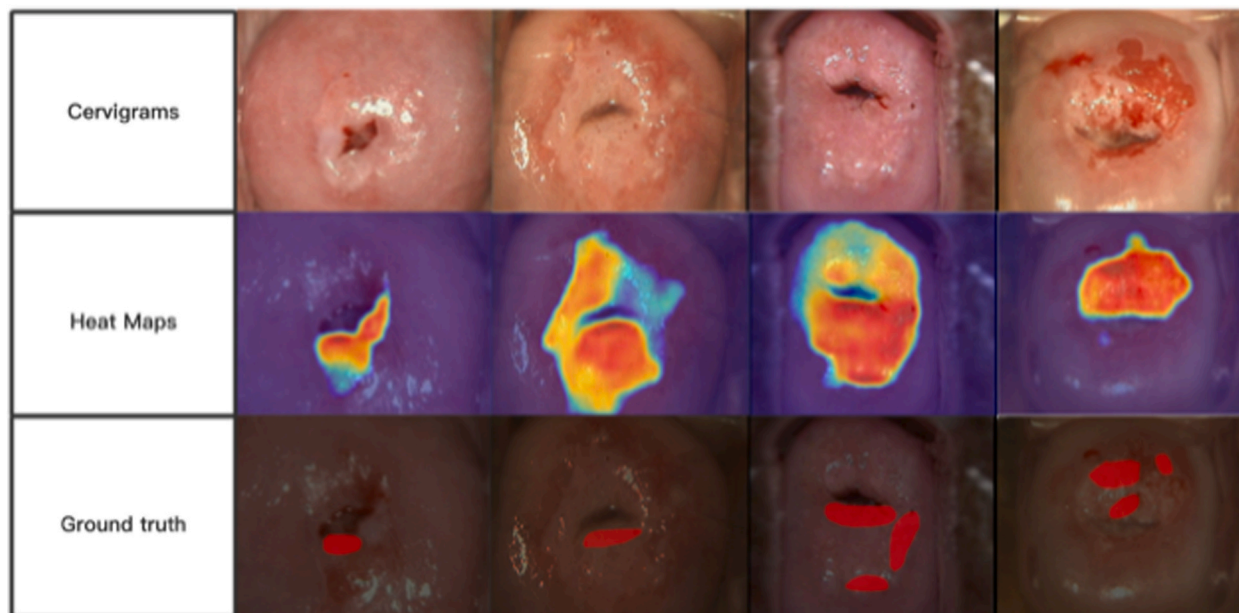
Fig. 2. Typical cervigrams and their commensurable heat maps as well as the most serious lesion ground truths..
(Cervigrams showed the post-acetic-acid images of typical HSIL/Cancer; Heat maps showed the prediction of lesions by DeepLabv3+; Ground truths showed the most serious lesions and most likely biopsy areas. The deeper red color in the heat maps indicated a higher likelihood of being recognized as HSIL/Cancer by the machine and was highlighted in red in the Ground truths.)

Table 2
Comparison in Sensitivity (equal the values of Recall), Spesiticy and Accuracy of Experts and DeepLabv3+.

	Sensitivity(%)	Specificity (%)	Accuracy (%)
Expert1	70.85	95.12	74.62
Expert2	72.39	95.00	79.69
Expert3	76.92	93.18	79.76
Average of experts	73.39	94.43	78.02
Deeplabv3+	**97.92**	**90.1**	**93.29**

Table 3
Comparison of previously published models based on post-acetic-acid images.

Preference	Number of Subject	Algorithm	Task	Accuracy (%)
2020, Li [19]	2049 CIN2/3 and 510 cancer cases	VGG-16 [5]	classification	75.31
2021, Liu [20]	4458 HSIL and 469 cancer cases	ResNet50 [6]	classification	79.7
2020, Cho [21]	511 HSIL and 43 cancer cases	InceptioNet-v2 [17]	classification	63.2
2022, Fang [22]	no specific sample sizes mentioned	ShuffleNet [22]	classification	81.38
2023, Chen [23]	612 HSIL, 1101 LSIL , and 4289 normal cases	EfficientNet-B0[18] and GRU-based [23]	classification	91.18
2021, Liu [24]	70 CIN3 cases	DeepLab V3+[16] with ResNet-101 [6]	segmentation	87.7
2023, Shinohara [25]	no specific sample sizes mentioned	U-net [26]	segmentation	89.4
2022, Fan [27]	3093 normal, 2794 LSIL, and 1219 HSIL+ cases	CMF-CNN(EfficientNet-B3 [18]+ASPP [15])	segmentation	92.70
2022, Yu [28]	1668 cancer cases	R–CNN + CLS-Net(EfficientNet-B3 [18]+ASPP [15])	segmentation	93.04

LSIL, low-grade squamous intraepithelial lesion; HSIL, high-grade squamous intraepithelial lesion; CIN, cervical intraepithelial neoplasm.

global average accuracy (75.9–81.3 %) and sensitivity (45–85 %) [25,27–31]. Due to differences in the training and experience levels of colposcopists, the utilization of the segmentation model can effectively enhance their diagnostic efficiency. Wu's prospective study [32] further highlighted that AI colposcopy consistently maintained stable sensitivity and significantly improved that of young colposcopists. Our model significantly improved accuracy (93.29 %) and sensitivity (97.92 %) compared to experts, with slightly lower specificity. Similarly, Wei's study [29] reported enhanced sensitivity with a small decrease in specificity compared to experts. These phenomena were explained in studies by Yu et al. [28] and Li et al. [19], where they explained that LSILs' misdiagnosis risk arises from losing lesion edges, heavy colors, and subtle texture features.

DeepLabv3+ selected Xception as its encoder, which is an improvement over InceptionNet, adopts depthwise separable convolution to decouple channel and spatial dimensions, simplifying the definition and modification of the architecture [17]. Downscaling parameters with 1x1 convolution reduces dimensionality, accelerates convergence, improves accuracy, while residual connections enhance DeepLabv3+'s performance [16]. Our colposcopic diagnosis accuracy using Xception was significantly higher than that achieved using ResNet, although both ours and Liu's [24] models were based on the DeepLabv3+ architecture. U-net [26], another segmentation model, also incorporates skip connections analogous to the residual connections in DeepLabv3+. However, DeepLabv3+ surpasses U-net by incorporating an Atrous Spatial Pyramid Pooling (ASPP) module with various atrous rates to capture multi-scale information [15].

All learning machines struggle with a few common issues. First, lesions in the type 2 and type 3 transformation zones are deeply embedded in the cervical canal. Machines cannot aid colposcopists in identifying these lesions unless they actively expose, inspect, and sample the inside of the cervical canal. Second, Models trained with data from a single hospital or colposcope may lack stability and suitability for complex data inputs [4]. In consequence, AI serves as a supportive tool for colposcopists, enhancing accuracy and efficiency without fostering complacency.

4. Conclusion

Colposcopy is crucial in guiding biopsies to the most suspect cervical region. With the help of our segmentation model, colposcopy can be more accurate in diagnosing HSIL or microinvasive cervical cancer.

5. Materials and Methods

5.1. Dataset

From November 2018 to September 2022, a total of 1042 women who had positive screening tests were assessed for eligibility at the Dongguan Maternal and Child Health Care Hospital's Cervical Disease Center. We selected cases that had essential information, including patient age, type of cervical transformation zone (TZ), clear colposcopic images (unobstructed view of the cervix, without a history of hysterectomy or cervical cancer surgery, and without significant bloodstains or leucorrhea), HPV detection results, cervical cytology results, biopsy pathology results. Table 4 showed the distribution of basic information for the patients. After excluding cases with incomplete data, there were 777 cases and 1554 colposcopic images (including saline images and post-acetic-acid images, Fig. 3 displayed the representative images). Iodine-stained images were excluded due to their inadequate resolution of epithelium and blood vessels. All patients signed informed consents for the use of their colposcopic images for academic research and publication. This retrospective study adhered to ethical standards for human trials and received approval from the Ethics Review Committee of Dongguan Maternal and Child Health Care Hospital.

5.2. Case prepared

Histopathology served as the gold standard for evaluating the diagnostic performance of colposcopy in detecting high-grade cervical abnormalities. Following the 2021 edition of The Lower Anogenital Squamous Terminology (LAST) [33], the included histopathological diagnoses were categorized as LSIL, HSIL and cervical cancer (including microscopic invasive cancer and invasive cancer). Among the cases, there were 125 LSIL cases, 339 HSIL cases and 313 cases of cancer. The colposcopic images from eligible cases were resized to 512 × 512 pixels.

Three experienced colposcopists from the Dongguan Maternal and Child Health Care Hospital's Cervical Disease Center were convened to review colposcopic images and mark the target areas. Each colposcopist had more than 5 years of experience in using colposcopes for diagnostic assessment of the uterine cervix. Utilizing the LabelMe software (https://github.com/wkentaro/labelme),

Table 4
The patients' Demographics and clinical information.

Pathology	LSIL(n = 125)	HSIL(n = 339)	Invasive cancer(n = 313)
Age(y, mean)	34.68	36.69	40.52
TZ(%)			
I	44.36	51.20	5.62
II	21.05	20.00	0.00
III	34.59	28.80	84.38
HPV(%)			
positive	76.15	96.88	93.29
nagative	23.85	3.12	6.71
TCT(%)			
≥LSIL	80.36	35.91	33.57
≥HSIL	19.64	64.09	66.43

LSIL, low-grade squamous intraepithelial lesion; HSIL, high-grade squamous intraepithelial lesion; Invasive cancer, including microinvasive or invasive cervical cancer; TZ, type of cervical transformation zone; HPV, human papillomavirus; TCT, Thinprep cytologic test.
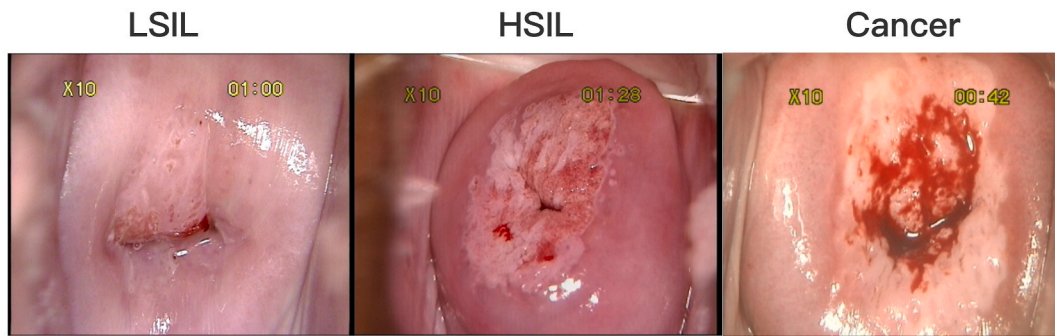
Fig. 3. Representative post-acetic-acid images of LSIL, HSIL and Cancer cases. (LSIL, low-grade squamous intraepithelial lesion; HSIL, high-grade squamous intraepithelial lesion; Cancer, including microinvasive or invasive cervical cancer. The figure displayed the magnification factor (10×) and the duration of the procedure.).

they identified and labeled the biopsy sites with the highest probability of being diagnosed as HSIL or cancer, corresponding to the colposcopy diagnoses based on histopathological findings. Each labeled site was assigned a unique diagnose to facilitate further analysis and comparison. Fig. 4A&B respectively showed schematics of the annotation processes for HSIL and cancer cases. The final labels were verified by two chief colposcopists with over 10 years of experience. Subsequently, the labeled data were divided into the training set, validation set and test set at ratio = 9:1:1.

6. Cervical lesion segmentation network

6.1. Experimental setup

The network was implemented using Python 3.8 and utilized the following libraries: PyTorch v1.10.1+cu113, torchvision v0.11.2+cu113, Matplotlib v3.6.1, NumPy v1.23.4, efficientnet-pytorch, and Cuda v11.3. The training was conducted on a system equipped with an NVIDIA GeForce RTX 3060 graphics card and 12-GB memory. All methods were measured on the same platform. The global optimizer employed was adaptive moment estimation (ADAM). The initial learning rate was set to 0.0005 and gradually reduced to 5e-6 after 50 epochs. The weight decay was set to 0 to avoid potential issues caused by ADAM's weight decay mechanism. The Xception module was pre-trained using VOC2012. All networks underwent training for 50 epochs and a batch size of 8.

6.2. Evaluation criteria

The four common criteria, namely, Accuracy, Precision, Specificity and mIoU were employed to evaluate the performance of the model [23]:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

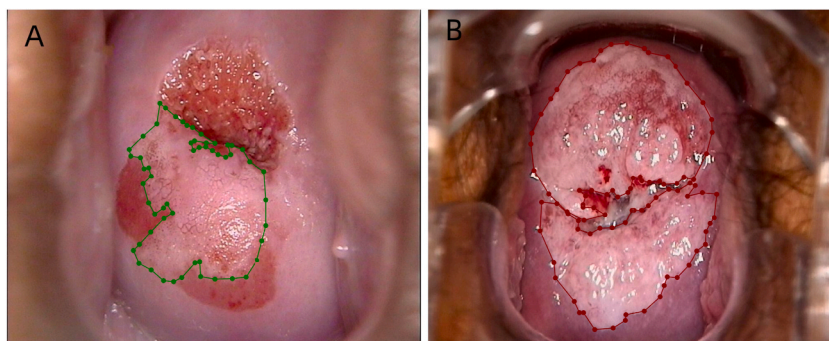$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Fig. 4. Schematic of the annotation.
(Fig. 4A. The typical JSON file in HSIL case. The green dotted line depicted the HSIL areas; Fig. 4B. The typical JSON file in cancer case. The red dotted line depicted the cancer areas.).
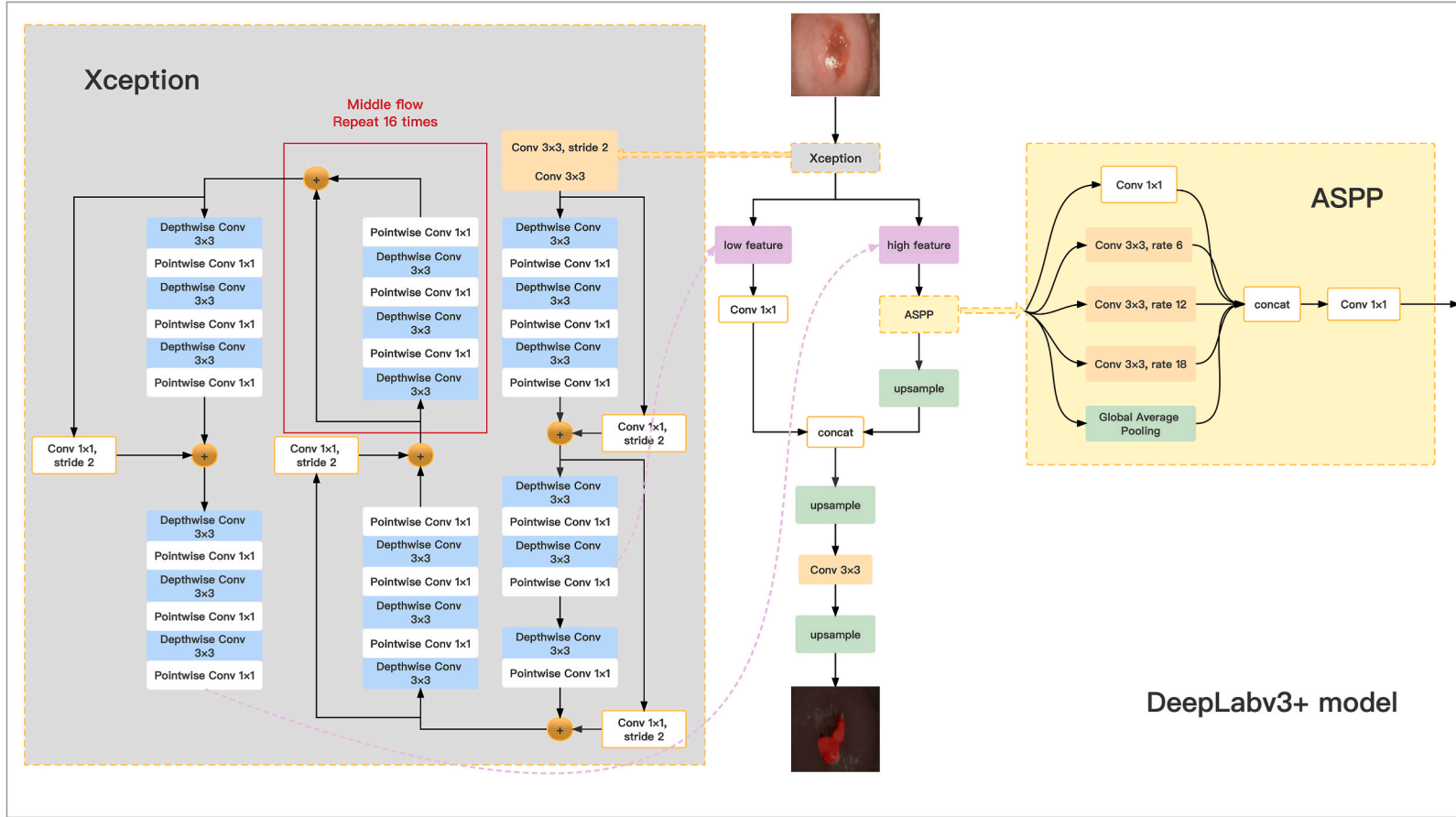
Fig. 5. Framework of the Deeplabv3+ model.

$$Specificity = \frac{TN}{TN + FP} \tag{3}$$

$$mIoU = \frac{1}{k+1}\sum\nolimits_{i=0}^{k} \frac{TP}{TP + FP + FN} \tag{4}$$

TP, TN, FP, and FN represent True Positive, True Negative, False Positive, and False Negative, respectively. In our study, regions within an image that were identified as high grade squamous intraepithelial lesion or cervical cancer were designated as positive, while the remaining areas were categorized as negative.

The loss function of the train set and validation set was Dice Loss [23].

$$Dice\ Loss = 1 - Dice \tag{5}$$

$$Dice = \frac{2 \times TP}{FP + FN + 2 \times TP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$Score = \frac{Dice + Recall}{2} \tag{8}$$

6.3. Framework of the Deeplabv3+ model

The DeepLabv3+ model utilized a dilated Fully Convolutional Network (FCN) architecture and employed an encoder-decoder structure. It incorporated depthwise separable convolutions and atrous convolutions to extract features at various resolutions. The overall framework was shown in Fig. 5. The encoder backbone of the model was Xception, which captured rich contextual features of cervical lesions. From the Xception model, the second residual module outputted low-level features with higher resolution but lower semantic information. Meanwhile, the Exit Flow of Xception provided high-level features with more semantic information but fewer details, which were then fed into the ASPP module to enhance spatial information. In the Decoder, the two types of extracted features were concatenated to refine the segmentation. More detailed descriptions of the Xception module, ASPP module, and Decoder are provided below.

6.4. Xception

Xception is a convolutional neural network architecture that relies entirely on depthwise separable convolutional layers. It consists of three main parts: Entry flow, Middle flow, and Exit flow. The network comprises 36 convolutional layers organized into 14 modules, with linear residual connections surrounding modules except the first and last ones, forming the foundation for feature extraction.

In the Entry flow, the first module examined each input channel using two regular 3×3 convolutions. Subsequently, in each of the following 12 residual modules, three 3×3 depthwise convolutions and three 1×1 pointwise convolutions, each followed by Batch Normalization (BN) and ReLU activation, are applied. This helped prevent gradient vanishing and improved convergence. The Middle flow consisted of these residual modules repeated 16 times. Lastly, the Exit flow's last modules consist of three 3×3 depthwise separable convolutions followed by BN and ReLU.

Xception provided two types of output features. The low-level features, originating from the second residual module, were adjusted to 48 channels through a 1×1 convolution in the Decoder. The high-level features, derived from the Exit flow, were fed into the ASPP network.

6.5. ASPP

The Atrous Spatial Pyramid Pooling module in the network probed a convolutional feature layer using filters at multiple sampling rates and effective fields-of-view. This allowed the module to capture objects and image context at multiple scales. After obtaining the high-level images from Xception, the modified ASPP module was applied to capture multi-scale information by adjusting the dilated rates of Atrous Convolution.

The modified ASPP module consisted of cascaded atrous convolutions. Specifically, it comprised four convolutional blocks: one 1×1 convolution and three 3×3 dilution convolutions at different atrous rates, along with global average pooling. The unit rates for the three dilation convolutional layers were defined as Multi Grid [15] = (r1, r2, r3). The final atrous rate was calculated as the multiplication of the unit rate and the corresponding rate. With an output stride of 16 and multigrid = (3, 6, 9), the three convolutions had atrous rates = 2 (3, 6, 9) = (6, 12, 18). To maintain a certain number of valid filter weights, the last block of ASPP employed global average pooling. The resulting image-level features were then fed into a 1×1 convolution and bilinearly upsampled to achieve the desired spatial dimension. The output features from all blocks at different scales were concatenated and passed through another 1×1 convolution. By avoiding downsampling, ASPP expanded the field of view and extracted rich semantic information while preserving feature resolution.

6.6. Decoder

DeepLabV3+ incorporated a Decoder module in its architecture to combine the low-level features extracted from Xception and the high-level features obtained from ASPP. Following the concatenation, a 3×3 convolution was applied to refine the combined representation. Subsequently, two simple bilinear upsamplings were performed, increasing the resolution by a factor of 4. Finally, bilinear interpolation was used to generate segmentation predictions that matched the original image's size. This approach ensured that the model produced accurate and detailed segmentation results.

Contribution to the field statement

Accurate identification of the appropriate biopsy location is essential for the precise diagnosis of high-level cervical lesions and cancer. However, in practical settings, the consistency of identifying these locations through colposcopic images varies, especially for colposcopists without comprehensive training. To address this issue and improve the effectiveness of cervical cancer screening in remote areas, artificial intelligence can play a vital role This work applied an efficient semantic segmentation model to learn the distinctive features of colposcopic pictures and predict regions corresponding to high-grade intraepithelial lesion and invasive cancer. In comparison to colposcopic experts, the Accuracy and Sensitivity of our machine were significantly higher; in comparison to other published models, the DeepLabv3+ model exhibited superior Accuracy and a more advanced structure.

Institutional review Board Statement

This retrospective study adhered to ethical standards for human trials and received approval from the Ethics Review Committee of Dongguan Maternal and Child Health Care Hospital (Approval Number 40).

Data availability statement

Due to the high memory footprint of high-resolution images, the data associated with the study has not been stored in a publicly accessible repository but can be provided upon request. For DeepLabv3+, you can find the running code at https://github.com/tensorflow/models/tree/master/research/deeplab.

CRediT authorship contribution statement

Zhen Li: Conceptualization. **Chu-Mei Zeng:** Writing – original draft, Formal analysis. **Yan-Gang Dong:** Software, Methodology. **Ying Cao:** Investigation. **Li-Yao Yu:** Investigation. **Hui-Ying Liu:** Data curation. **Xun Tian:** Writing – review & editing, Investigation. **Rui Tian:** Validation, Resources. **Chao-Yue Zhong:** Investigation. **Ting-Ting Zhao:** Investigation. **Jia-Shuo Liu:** Investigation. **Ye Chen:** Investigation. **Li-Fang Li:** Formal analysis. **Zhe-Ying Huang:** Formal analysis. **Yu-Yan Wang:** Investigation. **Zheng Hu:** Writing – review & editing, Supervision, Conceptualization. **Jingjing Zhang:** Writing – review & editing, Conceptualization. **Jiu-Xing Liang:** Software, Methodology. **Ping Zhou:** Data curation. **Yi-Qin Lu:** Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] H. Sung, J. Ferlay, R.L. Siegel, et al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J Clin 71 (3) (2021 May) 209–249.
[2] B.A. Wuerthner, M. Avila-Wallace, Cervical cancer: screening, management, and prevention, Nurse Pract 41 (9) (2016 Sep 22) 18–23.
[3] S. Nassiri, S. Aminimoghaddam, M.R. Sadaghian, et al., Evaluation of the diagnostic accuracy of the cervical biopsy under colposcopic vision, Eur J Transl Myol 32 (4) (2022 Oct 12), 10670.
[4] P. Xue, M.T.A. Ng, Y. Qiao, The challenges of colposcopy for cervical cancer screening in LMICs and solutions by artificial intelligence, BMC Med. 18 (1) (2020 Jun 3) 169.
[5] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv, 2014, p. 1409.
[6] K. He, X. Zhang, S. Ren, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[7] Z. Li, C. Li, L. Deng, Y. Fan, X. Xiao, H. Ma, J. Qin, L. Zhu, Improved AlexNet with inception-V4 for plant Disease diagnosis, Comput. Intell. Neurosci (2022 Sep 10), 5862600.
[8] P. Malhotra, S. Gupta, D. Koundal, , et al.W. Enbeyle, Deep neural networks for medical image segmentation, J Healthc Eng (2022), 9580991.
[9] Y. Miyagi, K. Takehara, T. Miyake, Application of deep learning to the classification of uterine cervical squamous epithelial lesion from colposcopy images, Mol Clin Oncol 11 (6) (2019) 583–589.
[10] A. Zimmer-Stelmach, J. Zak, A. Pawlosek, et al., The application of artificial intelligence-assisted colposcopy in a tertiary Care hospital within a cervical pathology diagnostic unit, Diagnostics 12 (1) (2022 Jan 4) 106.
[11] M. Krithika Alias AnbuDevi, K. Suganthi, Review of semantic segmentation of medical images using modified architectures of UNET, Diagnostics 12 (12) (2022 Dec 6) 3064.
[12] J.K. Bae, H.J. Roh, J.S. You, et al., Quantitative screening of cervical cancers for low-resource settings: pilot study of smartphone-based endoscopic visual inspection after acetic acid using machine learning techniques, JMIR Mhealth Uhealth 8 (3) (2020 Mar 11), e16467.

[13] L.C. Chen, G. Papandreou, I. Kokkinos, et al., Semantic image segmentation with deep convolutional nets and fully connected CRFs, IEEE Trans. Pattern Anal. Mach. Intell. (99) (2016) 834–848.

[14] L.C. Chen, G. Papandreou, I. Kokkinos, et al., DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs, IEEE Trans. Pattern Anal. Mach. Intell. 40 (4) (2018 Apr) 834–848.

[15] L.C. Chen, G. Papandreou, I. Kokkinos, Rethinking Atrous Convolution for Semantic Image Segmentation, 2017, https://doi.org/10.48550/arXiv.1706.05587.

[16] L.C. Chen, Y.K. Zhu, G. Papandreou, et al., Encoder-detector with atrous separate revolution for semantic image segmentation, in: Proceedings of the European Conference on Computer Vision, 2018, pp. 801–818.

[17] W. Muhammad, S. Aramvith, T. Onoye, Multi-scale Xception based depthwise separable convolution for single image super-resolution, PLoS One 16 (8) (2021), e0249278.

[18] M.X. Tan, Q.V. Le, EfficientNet: Rethinking model scaling for convolutional neural networks. International Conference on Machine Learning, PMLR, 2019, pp. 6105–6114.

[19] Y. Li, J. Chen, P. Xue, Computer-Aided cervical cancer diagnosis using time-lapsed colposcopic images, IEEE Trans Med Imaging 39 (11) (2020 Nov).

[20] Liu L, Wang Y, Liu X, Han S, Jia L, Meng L, Yang Z, Chen W, Zhang Y, Qiao X. Computer-aided diagnostic system based on deep learning for classifying colposcopy images. Ann. Transl. Med., 9(13):1045.

[21] B.J. Cho, Y.J. Choi, M.J. Lee, Classification of cervical neoplasms on colposcopic photography using deep learning, Sci. Rep. 10 (1) (2020), 13652.

[22] S. Fang, J. Yang, M. Wang, et al., An improved image classification method for cervical precancerous lesions based on ShuffleNet, Comput. Intell. Neurosci. 2022 (2022 Sep 13), 9675628.

[23] X. Chen, X. Pu, Z. Chen, et al., Application of EfficientNet-B0 and GRU-based deep learning on classifying the colposcopy diagnosis of precancerous cervical lesions, Cancer Med. 12 (7) (2023 Apr) 8690–8699.

[24] J. Liu, T. Liang, Y. Peng, Segmentation of acetowhite region in uterine cervical image based on deep learning, Technol. Health Care 30 (2) (2020) 469–482.

[25] T. Shinohara, K. Murakami, N. Matsumura, Diagnosis assistance in colposcopy by segmenting acetowhite epithelium using U-net with images before and after acetic acid solution application, Diagnostics 13 (9) (2023 Apr 29) 1596.

[26] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: 18th International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2015, pp. 234–241.

[27] Y. Fan, H. Ma, Y. Fu, et al., Colposcopic multimodal fusion for the classification of cervical lesions, Phys. Med. Biol. (13) (2022 Jun 22) 67.

[28] H. Yu, Y. Fan, H. Ma, Segmentation of the cervical lesion region in colposcopic images based on deep learning, Front. Oncol. 12 (2020), 952847.

[29] B. Wei, B. Zhang, P. Xue, et al., Improving colposcopic accuracy for cervical precancer detection: a retrospective multicenter study in China, BMC Cancer 22 (1) (2022 Apr 10) 388.

[30] X. Hou, G. Shen, L. Zhou, et al., Artificial intelligence in cervical cancer screening and diagnosis, Front. Oncol. 12 (2022 Mar 11), 851367.

[31] M. Hermens, R.M.F. Ebisch, K. Galaal, R.L.M. Bekkers, Alternative colposcopy techniques: a systematic review and meta-analysis, Obstet. Gynecol. 128 (4) (2016 Oct) 795–803.Citing datasets. [dataset]Everingham M, Van-Gool L, Williams C. K. I, Winn J, Zisserman A. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results, http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html.

[32] A. Wu, P. Xue, G. Abulizi, et al., Artificial intelligence in colposcopic examination: a promising tool to assist junior colposcopists, Front. Med. 10 (2023 Mar 15), 1060451.

[33] T.M. Darragh, T.J. Colgan, C.J. Thomas, The lower anogenital squamous terminology standardization project for HPV-associated lesions: background and consensus recommendations from the College of American Pathologists and the American Society for Colposcopy and Cervical Pathology, Int. J. Gynecol. Pathol. 32 (1) (2021) 76–115.