



Automatic Segmentation of Cervical Precancerous Lesions in Colposcopy Image Using Pyramid Scene Parsing Network and Transfer Learning

Jiahui Yang¹, Ying Zhang¹, Ying Liu², Shuang Liu^{1*}, Tetiana Chaikovska³, Chunhui Liu⁴

¹ College of Quality and Technical Supervision, Hebei University, Baoding 071002, China

² Lecturer in Strategy, Bournemouth University, Dorset House Williston Road Poole, BH12 5BB, UK

³ Department of Medical Imaging, Clinical Infectious Diseases Hospital N1, Kryvyi Rih 02154, Ukraine

⁴ Department of gynecology, Affiliated Hospital of Hebei University, Baoding 071000, China

Corresponding Author Email: whlius@hbu.edu.cn

<https://doi.org/10.18280/rces.100202>

ABSTRACT

Received: 27 April 2023

Revised: 4 May 2023

Accepted: 19 May 2023

Available online: 30 June 2023

Keywords:

cervical cancer, colposcopy examination, cervical precancerous lesions, automated segmentation, Pyramid Scene Parsing Network, transfer learning, PSPNet-ResNet50 Network

Cervical cancer is the second most common cancer among women worldwide. According to the 2020 estimates by GLOBOCAN in 185 countries, there were 604,000 new cases of cervical cancer and 342,000 deaths. In clinical practice, the segmentation of LSIL+ (cervical intraepithelial neoplasia+cervical cancer) lesions in colposcopic images (cervical imaging) is essential for assisting gynecologists in diagnosing cervical intraepithelial neoplasia grading and cervical cancer. It can also aid gynecologists in identifying the precise lesion area for further pathological examination. Existing computer-aided diagnosis algorithms exhibit poor segmentation performance due to insufficient training data that fail to focus on semantically meaningful lesion parts. In this study, we employed the improved Pyramid Scene Parsing Network (PSPNet-ResNet50) computer-aided diagnosis algorithm to automatically segment LSIL+ lesion areas in colposcopic images. We collected 971 images containing low-grade cervical intraepithelial neoplasia LSIL (CIN 1), high-grade cervical intraepithelial neoplasia HSIL (CIN 2/CIN 3), and cervical cancer from the Department of Obstetrics and Gynecology at Hebei University Affiliated Hospital. Two experienced gynecologists annotated the LSIL+ lesion areas to create a dataset for cervical lesion segmentation. We designed a lesion-aware convolution neural network transfer learning strategy to accomplish the lesion segmentation task. Comprehensive experiments were conducted to evaluate the proposed method's segmentation performance on clinical cervical images. Our research findings indicate that the PSPNet-ResNet50 network used in this study achieved the best segmentation results for automated (LSIL+ area) segmentation, with pixel accuracy (PA), mean pixel accuracy (MPA), precision (Pre), recall (Re), F1 score (F1), and mean intersection over union (MIoU) values of 95.21%, 89.83%, 84.58%, 82.22%, 83.38%, and 83.83%, respectively.

1. INTRODUCTION

Cervical cancer (CCA) is a malignant tumor caused by human papillomavirus (HPV) infection [1] and is one of the most common cancers among women. According to the latest data from the World Health Organization (WHO), the global incidence of cervical cancer reached 600,000 cases in 2020, with the disease causing approximately 340,000 deaths. Cervical cancer ranks fourth in terms of the number of new cases and deaths among all female cancers worldwide, accounting for 6.5% and 7.7% of the total incidence and mortality rates, respectively [2]. The cure rate for early-stage cervical cancer is relatively high, with different reports indicating a cure rate of over 95%, while the cure rate drops to around 20%-50% for advanced stages [3-7]. Although cervical cancer is one of the most preventable cancers, it remains the second leading cause of cancer-related death in women aged 20 to 39 [8]. This is primarily due to the lack of signs and symptoms in the early stages, hindering early diagnosis. Therefore, regular cervical cancer screening for women is crucial as it enables early detection and diagnosis of precancerous cervical lesions, preventing the progression to

cervical cancer and reducing the incidence and persistence of the disease [9]. Women with lower socioeconomic status are approximately twice as likely to die from cervical cancer compared to affluent women [10, 11]. Developing countries face a severe shortage of cervical cancer screening facilities due to a lack of experienced healthcare personnel and insufficient funding for screening systems [12]. Therefore, implementing effective cervical cancer screening programs can significantly reduce the disease incidence and prevent a substantial number of cervical cancer-related deaths. Deep learning has achieved remarkable success in the field of medical image segmentation, and computer-aided diagnosis plays an increasingly important role in the diagnosis of malignant tumors. In recent years, researchers have proposed various methods for diagnosing lesions from colposcopic images, mainly focusing on the extraction and segmentation of cervical regions or transformation zones. The current research primarily involves the segmentation of the acetic acid area and the segmentation of LSIL+ lesions.

Shi et al. [13] segmented the acetic acid area using the gray-level co-occurrence feature and the level set algorithm. Yue et al. [14] generated attention maps based on CICN combined

with UNet and CAM blocks to segment the acetic acid area using the proposed AWL-CNN network. Liu et al. [15] proposed a method for automatically segmenting the acetic acid area from colposcopic images. The method utilized the k-means clustering algorithm to extract the cervical region from the original colposcopic images. Then, DeepLab V3+ was employed to segment the acetic acid area from the cervical images, achieving an average accuracy of 91.2%. However, the acetic acid area includes not only lesion areas but also inflammation, partially normal metaplastic squamous epithelium areas, and other non-lesion areas. Thus, segmenting only the acetic acid area does not provide more accurate lesion-assisted diagnosis for doctors.

Currently, the segmentation of cervical precancerous lesions mainly involves region-based methods and pixel-based methods. Park et al. [16] generated anatomical maps based on color and texture in addition to region segmentation. They defined adult areas of the tissue region based on the anatomical feature maps using K-means clustering. Then, they combined the classification results of adjacent regions using the probability from the CRF classifier and determined the final classification result through KNN and LDA integration. Viñals et al. [17] proposed reducing the dimensionality of RGB vectors using PCA and generated probability maps for precancerous lesions in each pixel using ANN. They then connected points exceeding the threshold to segment the segment regions using seed point region growing and determined whether they were HSIL+ based on the size of the lesion area. However, HSIL+ is not related to the size of the region. Moreover, this method is susceptible to noise, resulting in cavities and over-segmentation, and requires high accuracy in extracting results from neural networks. Zhang [18] performed general lesion localization based on CAM on the region segmentation but did not provide specific contours. Zhang [19] proposed an improved U-Net by adding two convolutional blocks at the input and output to better extract image feature information. Yuan et al. [20] used ResNet instead of U-Net for CIN 1+ segmentation. In this paper, we propose a deep learning-based cervical lesion segmentation method. The main contributions of this study are as follows:

(1) This paper investigates a deep learning-based method for LSIL+ area segmentation and collects and annotates a dataset for cervical lesion segmentation.

(2) We employ the PSPNet-ResNet50 image segmentation model based on an encoder-decoder architecture, which efficiently segments lesions (LSIL+) and normal areas under colposcopy, accurately delineating the HSIL+ lesion area.

(3) The PSPNet-ResNet50 network-based model can segment cervical lesion areas and achieves the best segmentation results in terms of evaluation metrics such as pixel accuracy (PA), mean pixel accuracy (MPA), precision (Pre), recall (Re), F1 score (F1), and mean intersection over union (MIoU) when compared to other segmentation models. This demonstrates the effectiveness of the proposed segmentation algorithm for cervical lesion segmentation and its potential for assisting clinical diagnosis.

2. METHOD DESCRIPTION

The overall workflow of this method consists of image preprocessing, network architecture, and model evaluation. Specifically, the five-class labels of the training set are first processed as single-channel representations. Then, the training samples are proportionally scaled. Next, the improved network

is trained using the augmented training images with various transformations. Finally, to evaluate the classification performance, the test samples are fed into the pre-trained model, and the network's performance is analyzed and evaluated through the comparison of evaluation indices. The methods for each step will be comprehensively discussed in the following subsections.

2.1 Pyramid pooling module in PSPNet

PSPNet is a deep learning network proposed by Zhao et al. [21] in 2017. The core module is the pyramid pooling module, which aggregates contextual information from different scales and enhances the capability of capturing multi-scale features. The specific model structure is shown in Figure 1. First, classic neural networks such as VGG and ResNet are utilized to extract features from the input image. This step is represented by the CNN box in Figure 1. Then, the obtained feature maps undergo pooling at different scales to enhance the actual receptive field of each pixel, referred to as the pyramid pooling module in this paper. The dashed rectangular boxes in Figure 1 represent the pyramid pooling module. The red section consolidates the entire feature map into a single pixel, followed by a 1×1 convolution with a kernel that reduces the depth to one-fourth of the original depth. The yellow module divides the entire feature map into 4 grids and merges them into 1 pixel, resulting in a feature map of size 2×2 pixels, which is then subjected to a 1×1 kernel convolution to reduce the depth to one-fourth. The blue module divides the entire feature map into 9 grids and pools them into 1 pixel, obtaining a feature map of size 3×3 pixels, which is then convolved with a 1×1 kernel to reduce the depth to one-fourth of the original. The green module divides the entire feature map into 36 grids and merges them into 1 pixel, yielding a 6×6 pixel feature map, which is then convolved with a 1×1 kernel to reduce the depth to one-fourth of the original. Finally, the merged results of the four different scales are upsampled to the same size as the input feature map, concatenated with the input feature map, and then fused through convolution to obtain the prediction result.

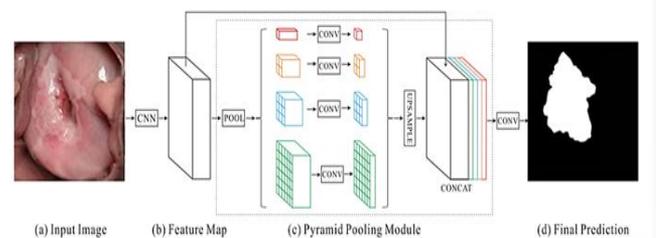


Figure 1. The PSPNet model architecture

2.2 PSPNet-ResNet50 model architecture

PSPNet utilizes a pre-trained CNN and dilated network technique to extract feature maps from input images. The size of the obtained feature maps is $1/8$ of the input image. Finally, the collection of these features is used to generate the output binary mask. The ResNet50 [22] model, which incorporates "skip connections" and "batch normalization" in a sequential manner, is employed to train deep networks without sacrificing efficiency. This allows gradients to bypass a certain

number of layers. The detailed diagram of PSPNet-ResNet50 is illustrated in Figure 2.

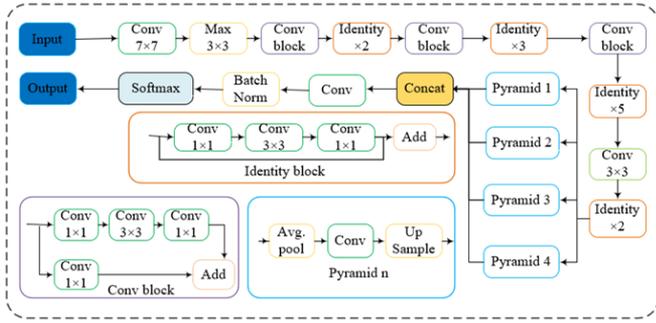


Figure 2. Detailed explanation of the PSPNet-ResNet50 network structure

2.3 Transfer learning

In the field of computer vision, transfer learning [23, 24] is commonly employed by leveraging pre-trained models. Pre-trained weights essentially refer to model parameters that have already been trained. In deep learning, model parameters are typically represented by weight matrices and bias vectors, which are learned through the backpropagation algorithm from a large volume of training data. Pre-trained models are trained on large benchmark datasets that usually contain millions or billions of images or textual data, such as ImageNet, COCO, Wikipedia, and others. Training models on these datasets enables the learning of general features and patterns, which can be transferred to other tasks, such as object detection, image segmentation, natural language processing, etc. For instance, in the task of image segmentation, the backbone neural network, such as VGG, ResNet, MobileNet, etc., is first utilized for feature extraction. When training a target segmentation model, the pre-trained weights of these neural networks can be used to initialize the network parameters of the backbone. By adjusting the parameters to learn features and patterns from the data, the model can better fit the data, thereby accelerating the training process and improving model performance.

For image segmentation tasks, a substantial amount of training data is typically required. To address the issue of insufficient training data, this study employs VGG [25], ResNet50, MobileNet [26], and Xception [27] as backbone networks in different segmentation networks for the feature extraction stage. The officially provided pre-trained weights are loaded, allowing for better initialization of neural network parameters and utilization of the general features learned from pre-trained models. This approach accelerates model training, enhances model performance, and strengthens generalization. Pre-trained models have been trained on large-scale datasets, capturing numerous useful general features applicable to various tasks. Therefore, the use of pre-trained weights enhances the model's generalization ability and enables better performance across different tasks.

3. DATA ACQUISITION AND PREPARATION

(1) Data Set Acquisition: In this study, the data were collected from 971 actual colposcopy cases provided by Hebei University Affiliated Hospital from June 2019 to February 2023. Each case includes a biopsy pathology report, a

colposcopy pathology report, as well as images taken before and after acetic acid application, an iodine test image, and a green light image. The data set consists of colposcopy images of normal cervix, different grades of cervical precancerous lesions (CIN 1 and CIN 2/3), and cervical cancer. Based on preliminary recommendations from physicians for computer-aided diagnosis, the research objective in this chapter focuses on a binary classification task: distinguishing between normal (non-lesion) and LSIL+ (Low-Grade Squamous Intraepithelial Lesion or higher) cases. This task is referred to as the LSIL+ diagnostic task. In other words, the aim of this study is to determine if the cervical image exhibits any abnormalities. When classifying the dataset, the pathology diagnosis report of each patient serves as the gold standard for data classification.

(2) Data Set Preparation: The collected data is anonymized by removing patient identity information. Subsequently, data cleansing is performed by filtering out inadequate images, selecting only those colposcopy images that meet the criteria of clear and complete cervix without obstructions, no severe bleeding, and lesions not heavily covered by discharge. This process is carried out by two colposcopy physicians with over five years of experience and one with over ten years of experience.

During the colposcopy examination, two images are captured separately, one before and one after the application of acetic acid, to record the surface changes of the cervix. The acetic acid is applied at a concentration of 3%-5%, and the second image is taken approximately two minutes after the application. The squamocolumnar junction is more clearly visible after two minutes of acetic acid application. Therefore, in this study, the image taken after two minutes of acetic acid application is used for annotation. Figure 3 illustrates the colposcopy images before and after acetic acid application.

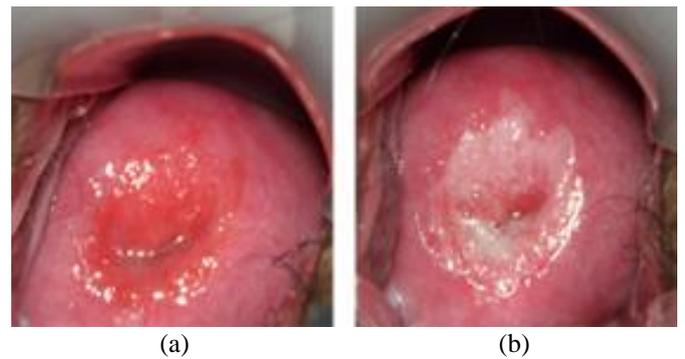


Figure 3. Raw colposcopy images. (a) image before the acetic acid test; (b) the 2-min image after the acetic acid test

A supervised learning approach is employed to segment the acetowhite area in colposcopy images. The acetowhite area is manually delineated by experienced experts in the field, serving as the standard for generating semantic segmentation labels. Deep learning-based image semantic segmentation methods require manual segmentation of target areas as references during the model training process. Semantic segmentation labels are typically grayscale or binary images, matching the size of the input original images. In this study, the Labelme software is used to create image semantic segmentation labels. The annotation of the target area is completed using points and line segments in the software, generating a JSON file (see Figure 4).



Figure 4. Annotation schematic diagram

The JSON file is then converted into a label map using a program. The label map contains two pixel values: 0 and 1. The corresponding regions for these two pixel values are the background area and the lesion area in the colposcopy image, as shown in Figure 5. From left to right, the first image is the original colposcopy image, the second image is the corresponding label map image, and the third image is the overlay of the original image and the mask image.



Figure 5. Colposcopy images and their labels

4. EXPERIMENTAL SETUP AND RESULT ANALYSIS

In this section, we conducted extensive experiments to test and evaluate the feature representation capability of the proposed method. Specifically, the evaluation metrics, implementation details, and comparative experiments are the main topics discussed in the following subsections.

4.1 Experimental environment and parameter configuration

Table 1. Experimental configuration table

Item	Configure
Operating system (OS)	Linux Ubuntu 18.04
System memory	128G DDR4
GPU	Nvidia GTX 2080Ti
CPU	Intel(R) Xeon(R) Gold 6240
Development language	Python 3.9
Deep learning framework	Pytorch 1.10.1; Cuda 11.2

After constructing the cervical lesion dataset, we trained the data using a U-Net-based network, performed testing, and made improvements to the model. The effectiveness of the network model was evaluated using evaluation metrics, and the model was fine-tuned for optimal performance. The hardware and software configurations used are shown in Table 1.

All network models were implemented on an Ubuntu 18.04 system with an Intel(R) Xeon(R) Gold 6240 processor and 128GB of DDR4 RAM. The GPU used was Nvidia GTX

2080Ti. The software stack included Python 3.9, PyTorch 1.10.1, and Cuda 11.2.

4.2 Model evaluation metrics

For the segmentation of cervical lesion regions, we employed six evaluation metrics to assess the segmentation performance: pixel accuracy (PA), mean pixel accuracy (MPA), precision (Pre), recall (Re), F1 score (F1), and mean intersection over union (MIoU).

$$PA = \frac{\sum_{i=1}^k p_{ij}}{\sum_{i=0}^k \sum_{j=0}^k p_{ij}} \quad (1)$$

$$MPA = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ij}}{\sum_{j=0}^k p_{ij}} \quad (2)$$

$$Pre = \frac{TP}{TP + FP} \quad (3)$$

$$Re = \frac{TP}{TP + FN} \quad (4)$$

$$F1 = \frac{2 \times Pre \times Re}{Pre + Re} \quad (5)$$

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ij}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}} \quad (6)$$

Among these metrics, true positive (TP) refers to the pixels that are actually lesions and correctly identified as lesions, true negative (TN) refers to the pixels that are actually non-lesions and correctly identified as non-lesions, false positive (FP) refers to the pixels that are actually non-lesions but mistakenly identified as lesions, and false negative (FN) refers to the pixels that are actually lesions but mistakenly identified as non-lesions.

Here, k represents the label results for different categories, where $k = 0$ represents the background category and $k = 1$ represents the cervical lesion category. i represents the ground truth, j represents the prediction, and p_{ij} represents the prediction of i as j . PA is the overall pixel accuracy. MPA is the average pixel accuracy for the lesion region and the background. Pre and Re represent the proportion of true lesions among the samples predicted as cervical lesions and the proportion of correctly predicted lesions among all cervical lesions, respectively. F1-score (F1) is a balanced measure determined by Pre and Re. Mean intersection over union (MIoU) is a standard measure for semantic segmentation, which evaluates the accuracy of the similarity between predicted and ground truth instances.

4.3 Analysis of experimental results

(1) Comparative experiments and result analysis

In this section, to fully validate the feasibility and effectiveness of the proposed model, we conducted comparative experiments between the PSPNet-ResNet50 model used in this study and seven other image semantic segmentation models: Unet-VGGNet, Unet-ResNet50,

DeeplabV3+-Xception, DeeplabV3+-MobileNet, HRNet-W18, HRNet-W32, and PSPNet-MobileNet. The segmentation performance was evaluated by comparing metrics such as pixel accuracy (PA), mean pixel accuracy (MPA), precision (Pre), recall (Re), F1 score (F1), and mean

intersection over union (MIoU). We comprehensively compared the performance of several algorithms. To ensure the credibility of the comparison results, all models used the dataset constructed in this study and were trained under the same environment and parameter settings.

Table 2. Network comparison of experimental data

Method	PA (%)	MP (%)	Pre (%)	Re (%)	F1 (%)	MIoU (%)
Unet-VGGNet	94.65	88.17	83.50	79.01	81.19	81.15
Unet-ResNet50	94.33	88.20	81.21	79.55	80.37	80.38
DeeplabV3+-Xception	94.29	88.29	80.84	79.81	80.32	80.32
DeeplabV3+-MobileNet	94.56	89.37	80.99	82.02	81.50	81.3
HRNet-W18	94.47	88.97	81.01	81.19	81.10	80.97
HRNet-W32	94.58	89.32	81.55	81.78	81.66	81.11
PSPNet- MobileNet	93.86	86.69	80.44	76.57	77.43	78.82
PSPNet- ResNet50	95.21	89.83	84.58	82.22	83.38	83.83

Table 2 presents a quantitative comparison of the cervical lesion segmentation method used in this study with other models. From the table, it can be observed that the proposed network achieved the best segmentation results with pixel accuracy (PA) of 95.21%, mean pixel accuracy (MPA) of 89.83%, precision (Pre) of 84.58%, recall (Re) of 82.22%, F1 score (F1) of 83.38%, and mean intersection over union (MIoU) of 83.83%. The effectiveness of the selected network in segmenting the LSIL+ region of the cervix can also be visually confirmed from Figure 6.

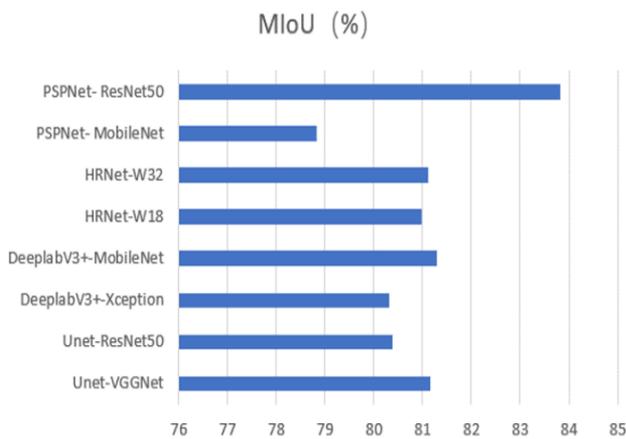
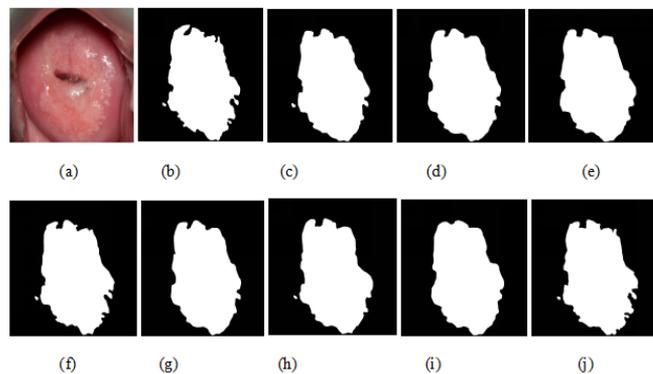


Figure 6. Mean intersection and ratio (MIoU) of different networks

To provide a more intuitive comparison of the training results from different networks, we randomly selected an image of the cervix and compared the segmentation results obtained by our model with several other models. Figure 7 illustrates the segmentation results of our model compared to other models. It can be observed that some models produced segmentation results that included areas of squamous metaplasia, which appear similar to the lesion region, leading to reduced segmentation accuracy. The PSPNet-ResNet50 model selected in this study effectively distinguished the lesion region from the squamous metaplasia region, as shown in Figure 7(j), which closely matches the manually annotated label in Figure 7(b). The integration of more low-level features in PSPNet-ResNet50 resulted in clearer detection of the lesion area and more accurate segmentation of the edge details, achieving finer results.

Figure 7. Effect diagram of different networks. (a) the original image; (b) the label; (c) Unet-VGGNet; (d) Unet-ResNet50; (e) DeeplabV3+-Xception; (f) DeeplabV3+-MobileNet; (g) HRNet-W18; (h) HRNet-W32; (i) PSPNet-MobileNet; (j) PSPNet-ResNet50

5. CONCLUSION

In this study, we presented a detailed introduction to the PSPNet-ResNet50 network structure, which can effectively detect and assist in the diagnosis of cervical lesion regions. The main advantage of this method is that the trained model enables rapid detection and diagnosis, significantly improving the speed and efficiency of diagnosis. In the experimental section, we provided specific details about the experimental environment, hardware parameter configuration, training parameters, data acquisition and processing, model evaluation metrics, and result analysis. The data acquisition and processing section described the acquisition and selection of original colposcopy images, along with the production of semantic segmentation label data. The manual annotation of the lesion region in the original colposcopy images using Labelme software and the conversion of the annotated data into binary image labels required for convolutional neural network training were explained. In the result analysis, we compared our network with other networks and observed improved segmentation performance in the cervical lesion region. The experimental results demonstrated that the optimized model achieved high accuracy in segmenting the cervical lesion region, providing valuable assistance in clinical diagnosis of cervical lesions.

FUNDING

This study was supported by Baoding Science and Technology Planning Project (2141ZF306, 2141ZF135), and supported by Youth Foundation of Affiliated Hospital of Hebei University (2022QC54).

DATA AVAILABILITY

The data used to support the research findings are available from the corresponding author upon request.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- [1] Wang, P., Xiao, Y.B. (2022). Research progress in human papillomavirus infection and treatment. *Journal of Surgery*, 31(12): 1107-1111.
- [2] Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 71(3): 209-249. <https://doi.org/10.3322/caac.21660>
- [3] Torre, L.A., Siegel, R.L., Ward, E.M., Jemal, A. (2016). Global cancer incidence and mortality rates and trends—An update global cancer rates and trends—An update. *Cancer Epidemiology, Biomarkers & Prevention*, 25(1): 16-27.
- [4] Peng, L., Yuan, X., Jiang, B., Tang, Z., Li, G.C. (2016). LncRNAs: Key players and novel insights into cervical cancer. *Tumor Biology*, 37: 2779-2788. <https://doi.org/10.1007/s13277-015-4663-9>
- [5] Pfaendler, K.S., Wenzel, L., Mechanic, M.B., Penner, K.R. (2015). Cervical cancer survivorship: Long-term quality of life and social support. *Clinical Therapeutics*, 37(1): 39-48. <https://doi.org/10.1016/j.clinthera.2014.11.013>
- [6] Kessler, T.A. (2017). Cervical cancer: Prevention and early detection. *Seminars in Oncology Nursing*, 33(2): 172-183. <https://doi.org/10.1016/j.soncn.2017.02.005>
- [7] Sabik, L.M., Tarazi, W.W., Hochhalter, S., Dahman, B., Bradley, C.J. (2018). Medicaid expansions and cervical cancer screening for low-income women. *Health Services Research*, 53: 2870-2891. <https://doi.org/10.1111/1475-6773.12732>
- [8] Siegel, R.L., Miller, K.D., Fuchs, H.E., Jemal, A. (2022). *Cancer statistics, 2022*. CA: A Cancer Journal for Clinicians, 72(1): 7-33. <https://doi.org/10.3322/caac.21708>
- [9] Hua, W., Xiao, T., Liu, Z., Wang, M., Zheng, H., Wang, S. (2020). Lymph-vascular space invasion prediction in cervical cancer: Exploring radiomics and deep learning multilevel features of tumor and peritumor tissue on multiparametric MRI. *Biomedical Signal Processing and Control*, 58: 101869. <https://doi.org/10.1016/j.bspc.2020.101869>
- [10] Singh, G.K., Jemal, A. (2017). Socioeconomic and racial/ethnic disparities in cancer mortality, incidence, and survival in the United States, 1950–2014: Over six decades of changing patterns and widening inequalities. *Journal of Environmental and Public Health*, 2017.
- [11] Siegel, R.L., Miller, K.D., Jemal, A. (2019). *Cancer statistics, 2019*. CA: A Cancer Journal for Clinicians, 69(1): 7-34. <https://doi.org/10.3322/caac.21551>
- [12] Yusufaly, T.I., Kallis, K., Simon, A., et al. (2020). A knowledge-based organ dose prediction tool for brachytherapy treatment planning of patients with cervical cancer. *Brachytherapy*, 19(5): 624-634. <https://doi.org/10.1016/j.brachy.2020.04.008>
- [13] Shi, H.J., Liu, J., Huang, H.Y., Du, H.W. (2018). Acetowhite region segmentation in cervix based on gray level co-occurrence characteristic and level set algorithm. *Journal of Nanchang Hangkong University*, 32(2): 8-16. <http://dx.doi.org/10.3969/j.issn.1001-4926.2018.02.002>
- [14] Yue, Z., Ding, S., Li, X., Yang, S., Zhang, Y. (2021). Automatic acetowhite lesion segmentation via specular reflection removal and deep attention network. *IEEE Journal of Biomedical and Health Informatics*, 25(9): 3529-3540. <https://doi.org/10.1109/JBHI.2021.3064366>
- [15] Liu, J., Liang, T., Peng, Y., Peng, G., Sun, L., Li, L., Dong, H. (2022). Segmentation of acetowhite region in uterine cervical image based on deep learning. *Technology and Health Care*, 30(2): 469-482. <https://doi.org/10.3233/THC-212890>
- [16] Park, S.Y., Sargent, D., Lieberman, R., Gustafsson, U. (2011). Domain-specific image analysis for cervical neoplasia detection based on conditional random fields. *IEEE Transactions on Medical Imaging*, 30(3): 867-878. <https://doi.org/10.1109/TMI.2011.2106796>
- [17] Viñals, R., Vassilakos, P., Rad, M.S., Undurraga, M., Petignat, P., Thiran, J.P. (2021). Using dynamic features for automatic cervical precancer detection. *Diagnostics*, 11(4): 716. <https://doi.org/10.3390/diagnostics11040716>
- [18] Zhang, D. (2018). Deep learning-based colposcopy HSIL detection. Zhejiang University.
- [19] Zhang, T. (2020). Assisted screening study for cervical cancer based on deep neural network. Huaqiao University.
- [20] Yuan, C.N., Yao, Y.L., Cheng, B., Cheng, Y.F., Li, Y., Li, Y., Liu, X.C., Cheng, X.D., Xie, X., Wu, J., Wang, X.Y., Lu, W.G. (2020). The application of deep learning based diagnostic system to cervical squamous intraepithelial lesions recognition in colposcopy images. *Scientific Reports*, 10(1): 1-12. <https://doi.org/10.1038/s41598-020-68252-3>
- [21] Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J. (2017). Pyramid scene parsing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2881-2890.
- [22] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778.
- [23] Fawaz, H.I., Forestier, G., Weber, J., Idoumghar, L., Muller, P.A. (2018). Transfer learning for time series classification. In *2018 IEEE international conference on big data (Big Data)*, 10-13 December 2018, USA: IEEE, 1367-1376.
- [24] Zhao, X.B., Wang, J.J. (2023). Bridge crack detection based on improved DeeplabV3+ and transfer learning. *Computer Engineering and Application*, 59(5): 262-269.

- <https://doi.org/10.3778/j.issn.1002-8331.2204-0503>
- [25] Simonyan, K., Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. <https://doi.org/10.48550/arXiv.1409.1556>
- [26] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4510-4520.
- [27] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1251-1258.