IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 8 February 2024, accepted 12 March 2024, date of publication 18 March 2024, date of current version 22 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3378097

RESEARCH ARTICLE

Colposcopic Image Segmentation Based on Feature Refinement and Attention

YUXI HE[1], LIPING LIU[2], JINLIANG WANG[1], NANNAN ZHAO[1], AND HANGYU HE[1]
[1]College of Artificial Intelligence, North China University of Science and Technology, Tangshan, Hebei 063210, China
[2]College of Mechanical and Energy Engineering, Shanghai Electronic Information Vocational and Technical College, Shanghai 201411, China

Corresponding author: Liping Liu (11745430@qq.com)

This work was supported in part by Hebei Province University Student Science and Technology Innovation Capability Cultivation Program under Grant 22E50098D, and in part by Shanghai Electronic Information Vocational and Technical College High-Level Talent Initiation Project under Grant GCC2023001.

ABSTRACT The current computer-aided diagnosis for cervical cancer screening encounters issues with missing detailed information during colposcopic image segmentation and incomplete edge delineation. To overcome these challenges, this study introduces the RUC-U^{2}Net architecture, which enhances image segmentation through feature refinement and upsampling connections. Two variants are developed: RUC-U^{2}Net and the lightweight RUC$^+$-U^{2}Net. Initially, a feature refinement module that leverages an attention mechanism is proposed to improve detail capture by the model's fundamental unit during downsampling. Subsequently, the integration of diagonal attention in connecting peer-level encoders and decoders supplements finer semantic details to the decoder's feature maps, addressing the problem of incomplete edge segmentation. Finally, the application of the Focal Tversky loss function allows the model to concentrate on difficult samples, mitigating the challenges posed by imbalanced distributions of positive and negative samples in training datasets. Experimental evaluations on three publicly available datasets demonstrate that the proposed models significantly outperform existing methods across seven performance metrics, evidencing their superior segmentation accuracy.

INDEX TERMS Image segmentation, colposcopy image, feature refinement, lightweight upsampling, loss function.

I. INTRODUCTION

Colposcopic examination is a critical technique for identifying precancerous conditions of cervical cancer. The progression from HPV infection to cell differentiation and subsequent carcinogenesis in a normal cervix typically spans 5 to 10 years [1]. Hence, early detection through colposcopy is vital for cervical cancer prevention [2]. However, economic constraints in underdeveloped regions have hindered the widespread adoption of cervical cancer screening [3], resulting in notably higher incidence rates there compared to developed areas [4], [5], [6]. Introducing Computer-Aided Diagnosis (CAD) technology to in screening could enhance efficiency [7], [8] and extend screening efforts to these underprivileged regions [9], benefiting a broader patient population.

The associate editor coordinating the review of this manuscript and approving it for publication was Zhan-Li Sun.

Compared to other cervical cancer detection methods, colposcopic examination offers non-invasiveness, ease of operation, low cost, and superior image quality. Currently, the detection of the Region of Interest (ROI) in colposcopic images involves manual efforts by trained gynecologists. Due to the intricate edges and detailed information prevalent in the cervical region, such as instrument obstruction, pseudo shadows from hair, edge lesions, and capture blurriness, the manual segmentation process is both tedious and time-consuming, requiring substantial medical resources. Although manual annotation by doctors using traditional methods is dependable for assessing regions, the scarcity and energy of doctors, alongside environmental factors and individual circumstances, may introduce biases during the annotation process [10]. Therefore, we aim to explore a computer-assisted approach that could alleviate the burden on doctors in annotating lesion areas, thereby enhancing screening efficiency.

Presently, deep learning has demonstrated considerable potential in CAD, with numerous studies utilizing deep learning methods to support clinical doctors in colposcopic examinations. Most studies focus on the classification of cervical intraepithelial neoplasia (CIN) [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], with nearly all incorporating a common preprocessing step: extracting ROI from colposcopic examination images. Segmentation is a fundamental step in training deep learning-based models, as more accurate segmentation directly improves the precision of using deep learning for cervical cancer lesion classification [3]. Hence, this study proposes a deep learning-based segmentation model for enhanced segmentation results, aiming to direct clinical doctors' attention to specific areas when analyzing colposcopic images. This facilitates diagnosis and lays a robust foundation for further research on lesion grading classification and biopsy site selection. Additionally, generalization experiments show that this method can be a supplementary approach for other medical diagnoses, boosting diagnostic performance.

Deep learning has garnered widespread attention for its capacity to automatically learn and extract meaningful features from input data [22], proving effective and feasible in the medical field [23], [24]. For instance, UTNet [25] achieves a state-of-the-art level in cardiac MRI segmentation; AUNet [26] is applied in breast nodule segmentation; PSPNet [27] excelled in the 2016 ImageNet Scene Parsing Challenge, the 2012 PASCAL VOC benchmark, finding applications in tasks like segmenting coronary angiography images [28] and prostate magnetic resonance imaging [29]. These deep learning models and their variants have become central to contemporary medical image segmentation.

Research in the field of image ROI edge segmentation using deep learning has yielded various innovations. Wu et al. [30] developed the Channel Groupwise Drop Network (CGDNet), which achieves the fusion of fine-grained and global features through a channel grouping dropout strategy, thereby enhancing the accuracy of palmprint features. Furthermore, Zhang et al. [31] modified the stride of Region of Interest Align (ROIAlign) and incorporated the FPN structure, making the mask prediction more responsive to edge information. In fine-grained research, a study [32] introduces an enhanced deconvolution module, applied successfully in the network. Through a response alignment strategy, it minimizes inconsistent responses and mitigates undesirable predictions, offering a novel approach to the alignment issues in edge fine-grained segmentation features. Additionally, the Pulse-Coupled Neural Network (PCNN) is known for edge detection and image segmentation tasks; however, its performance depends heavily on parameter settings. To overcome this challenge, a study [33] introduces GSAPCNN (Gravitational Search Algorithm Pulse-Coupled Neural Network), employing the Gravitational Search Algorithm (GSA) to optimize PCNN model parameters. This strategy has proven successful in image segmentation, providing an effective means to boost PCNN performance. These studies offer valuable insights and methods, contributing to advancements in deep learning for image ROI edge segmentation.

In medical image analysis, especially in tackling challenges related to detailed feature extraction and fuzzy edge segmentation, attention mechanisms have gained prominence. Researchers have enhanced the sensitivity of networks to both global and local features by implementing various attention modules. Examples include the dual-branch geometric attention network for 3D tooth segmentation in DBGANet [34] and the Hybrid Adaptive Attention Module (HAAM) by Chen et al. [35], which are instrumental in this improvement. Additionally, Li et al. [36] utilized dual local attention (DLA) in the GT-DLA-dsHFF network to extract local vascular information, presenting an effective solution to complex challenges in medical image analysis. These approaches enrich the toolbox of medical image analysis, supporting more accurate diagnosis and segmentation.

Currently, research in cervical cancer using deep learning methods focuses primarily on the classification of lesion grades [37], [38], [39] and the segmentation of cervical cells [40], [41], [42]. This study highlights a gap in literature on end-to-end models for segmenting colposcopic images, noting that existing studies lack critical segmentation details and complete edge segmentation. Kim et al. [10] noted that more precise colposcopic image segmentation can significantly enhance the accuracy of cervical cancer lesion classification using deep learning. Thus, the aim of this study is to introduce a segmentation model that achieves improved accuracy with enhanced segmentation effects, laying a foundation for further research on lesion grading and biopsy site selection. Most current colposcopic image segmentation methods rely on the U-Net model [43], a specialized neural network model for medical images, based on Fully Convolutional Networks (FCN). It incorporates shortcut connections between feature maps of the same scale to fully integrate information, enabling superior segmentation performance in colposcopic image recognition. Subsequent research has extensively explored various models for medical image segmentation based on U-net. For example, Bai et al. [44] segmented colposcopic images using the U-Net model by replacing the fully connected layer with a convolutional layer and employing a deconvolution structure for data upsampling; Zhang et al. [45] substituted the convolutional layer in U-Net with a pooling layer and added a dropout layer for colposcopic image segmentation; Yuliana et al. [46] employed U-Net to segment ROI in colposcopic images. Additionally, Liu et al. [8] compared U-Net, FCN, and SEGNet [47] in CIN image segmentation tasks, finding that U-Net delivers precise edge segmentation. These methods underscore the superior performance of U-Net in colposcopic image segmentation.

Contrary to most previous studies that conducted pre-processing on colposcopic images [44], [45], [46], this study advocates for end-to-end segmentation of the

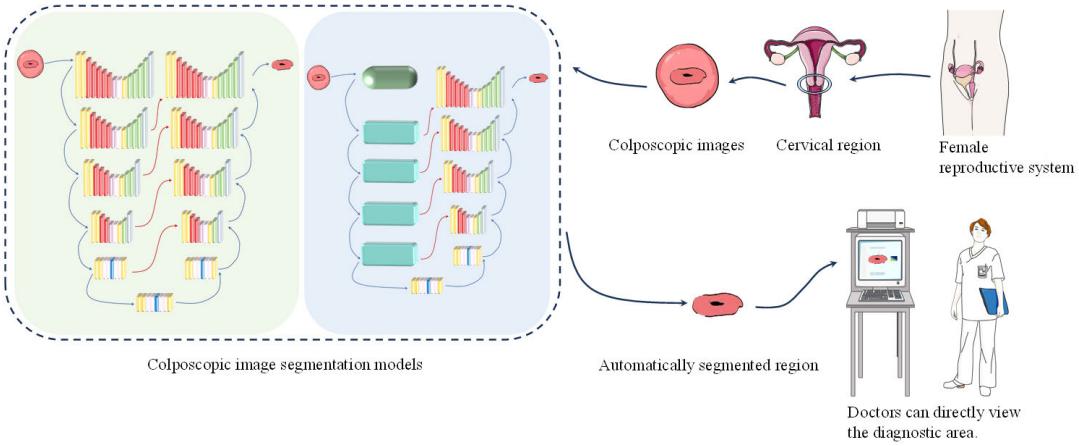

FIGURE 1. RUC-U^{2}Net network architecture working simulation diagram.

region of interest on unaltered colposcopic images, enabling clinicians to make diagnoses based on authentic image data. Given the often blurred borders and complex gradients in raw colposcopic images, high-resolution information is essential for segmenting lesion contours and refining edge details. Inspired by the U^2-Net model proposed by Qin et al. [48], which builds on U-Net and introduces ReSidual U-blocks (RSU) containing pooling operations, this approach deepens the network depth and uses different receptive field sizes to capture multiscale information, showing superior performance in segmenting colposcopic images.

However, the issues of missing detailed information and incomplete edge segmentation persist in computer-aided diagnostic cervical cancer screening that utilizes U^2-Net. Therefore, this study introduces the colposcopic image segmentation architecture, RUC-U^{2}Net, which enhances feature extraction and incorporates upsampling concatenation. It experimentally confirms the efficacy and advancement of two models developed through this architecture: RUC-U^{2}Net and its lightweight counterpart, RUC$^+$-U^{2}Net, in colposcopic image segmentation. Moreover, these models demonstrate improved performance on the Drive fundus vascular segmentation public dataset and the DUT-OMRON saliency target detection public dataset, showing their strong generalization capabilities. The specific contributions are as follows:

1) We introduced a Refining Bottleneck (RB) module based on Shuffle Attention (SA) [49] for feature refinement, embedded within the R^{2}SU module, the core unit of the RUC-U^{2}Net architecture. This module merges larger-scale and smaller-scale feature maps, introducing richer details and semantic features into subsequent feature maps. It improves the feature refinement capability of the R^{2}SU module during downsampling and enhances the accuracy of the model's output features, addressing the issue of roughness in colposcopic image segmentation.
2) We introduced an upsampling Oblique Attention Connection Module (OACM) based on SA with Context-Aware ReAssembly of Features (CARAFE) [50], replacing the simple horizontal connection between the original encoder and decoder with a diagonal attention connection. This allows for the fusion of complementary refined information from different scales in the decoder stage, effectively addressing the issue of incomplete edge segmentation in colposcopic images.
3) We replaced the BCE loss function with the Focal Tversky loss function [51], enabling the model to adequately learn complex indistinguishable difficult samples and mitigate the problem of model degradation caused by the imbalanced distribution of positive and negative samples during training, thereby enhancing the model's detection accuracy of the lesion area.

II. CONSTRUCTION OF SEGMENTATION MODEL

A. RUC-U^{2}NET ARCHITECTURE

The precise segmentation of the cervix is crucial for the diagnostic accuracy of colposcopic analysis [10]. Existing methods face challenges, such as the loss of detail information and incomplete edge segmentation. Therefore, a colposcopic image segmentation architecture based on feature refinement and upsampling connection, RUC-U^{2}Net, is proposed (Figure 1). This architecture is realized in two models: the RUC-U^{2}Net and the lightweight RUC$^+$-U^{2}Net, as illustrated below:

1) RUC-U^{2}NET MODEL

The RUC-U^{2}Net model's structure for colposcopic image segmentation, emphasizing feature refinement and attention, retains a U-shaped profile but incorporates a nested structure for more precise feature merging across different scales, as illustrated in Figure 2. Specifically, RUC-U^{2}Net consists of three parts: the encoder on the left, the decoder on the right, and the mapping fusion module below. Both the encoder and decoder comprise basic units Refining called Residual u-blocks (R^{2}SU), with the RB module replacing the downsampling convolution module in the original RSU

to further refine feature extraction. Unlike U^2-Net, which connects the encoder and decoder horizontally, this study introduces OACM oblique attention connections to extract and fuse information at different scales. The mapping fusion module integrates features from each decoder stage for the final prediction map. To address data imbalance, the Focal Tversky loss function replaces the original BCE loss function during training. The RB module, OACM connection, and Focal Tversky loss function constitute the innovative elements of this study.

2) LIGHTWEIGHT MODEL RUC$^+$-U^{2}NET

In this paper, we introduce a lightweight model, RUC$^+$-U^{2}Net, derived from the RUC-U^{2}Net structure, as illustrated in Figure 3. This model utilizes a lightweight network structure to minimize model training time, thereby enhancing the model's real-time performance in practical applications. RUC$^+$-U^{2}Net maintains the U-shaped profile, but employs the Inverted ReSidual Block (IRB) extensively in the decoder stage. The IRB module's structure, shown in Figure 3, aims to lower the model's parameter count and training duration during semantic feature extraction.The IRB consists of a 1 × 1 convolution, a 3 × 3 Depthwise Separable Convolution, and the ReLu6 activation function. This strategy of increasing dimensionality before reducing it ensures efficient high-level semantic information extraction with fewer parameters. The residual connection is applied where the stride is 1, as depicted in Figure 3(a); in other cases, due to dimensional changes post-convolution, the residual connection is omitted, as detailed in Figure 3(b).

B. REFINING BOTTLENECK

The bottleneck [52], a residual module, captures semantic information in the form of residuals through convolution, Batch Normalisation (BN), and Rectified Linear Unit (ReLU) functions. This paper introduces SA and Sigmoid Linear Unit (Silu) [53], creating the RB module and applying it from the R^{2}SU-7 to R^{2}SU-4 downsampling stages to boost the model's feature extraction capability. The RB's structure, presented in Figure 4, shows the input feature map on the main path with dimensions h × w × c. The process begins by reducing the number of output channels to c/2 through a 3 × 3 convolution, followed by BN, SiLU, and Dropblock treatment. Subsequent operations include another 3 × 3 convolution to restore the channel count to c, with further processing by BN, SiLU, and Dropblock. The feature map then enters the SA module via a jump connection branch. The outputs from both primary and branch circuits are merged to produce features of size h × w × 2*c, with the input channel count returning to c through a 1 × 1 convolution, yielding the final output feature map after BN and SiLU processing.

The RB module innovatively incorporates the SA within the residual connection branch, enhancing the accuracy of output features by modulating the weight parameter of input features. This approach emphasizes critical features and integrates detailed features on a large scale with semantic features on a smaller scale, thereby refining the output features' accuracy.

SA (as shown in the SA structure in Figure 6) first divides the feature map into group G along the channel, denoted as $X = [X_1, X_2, \cdots, X_G]$, where each sub-feature X_k is divided into two branches, one processed using channel attention and the other one using spatial attention.Subsequently, channel shuffle(CA) is employed to interactively amalgamate the sub-features from each group, culminating in comprehensive feature fusion. The processed feature, denoted as X_sa, post-SA treatment is articulated as follows

$$X_{sa} = CA\left(X_c + X_s\right) \tag{1}$$

where X_{sa} denotes CA, X_c and X_s denote channel attention processing and spatial attention processing,the operational metrics are defined as

$$X_c = \sigma\left(W_1 s + b_1\right) \cdot X_{k1} \tag{2}$$

$$X_s = \sigma\left(W_2 \cdot GN\left(X_{k2}\right) + b_2\right) \cdot X_{k2} \tag{3}$$

In formula 2, X_{k1} is the feature information of input channel attention, W_1 and b_1 denote the parameters of the output feature map that will be rescaled in the channel attention, σ represents the sigmoid activation function, s is the generated statistic about the number of channels, and X_c is the feature information output by the channel attention. In formula 3, X_{k2} is the characteristic information of input spatial attention, GN is the use of group normalization to obtain spatial features, W_2 and b_2 denote the parameters of the output feature map that will be rescaled in the spatial attention, and X_s is the feature information output by the channel attention.

Furthermore, this paper applies Silu to the RB module. Compared to the original ReLu activation function, Silu offers the advantage of possessing a lower bound without an upper bound, while demonstrating more potent regularization effects. It can mitigate the overfitting problem caused by the monotonically increasing characteristics of the ReLu function and enhance the stability of model training. The metric is defined as

$$Silu = x \cdot \text{Sigmoid}(x) \tag{4}$$

$$\text{Sigmoid}(x) = \frac{1}{\left(1 + e^{-x}\right)} \tag{5}$$

$$Silu' = Silu + (1 - Silu) \cdot Sigmoid(x) \tag{6}$$

where x is the feature information of the input activation function.After calculating the derivative of the Silu activation function, a minimum with a derivative of 0 at a global level prevents the updating of significant weights within the network, thus effectively addressing the gradient explosion issue that arises from the constant multiplication of large weights.

Furthermore, RB integrates the Dropblock module following the convolution operation correlated with SA, as illustrated in Figure 5(a). This approach not only addresses

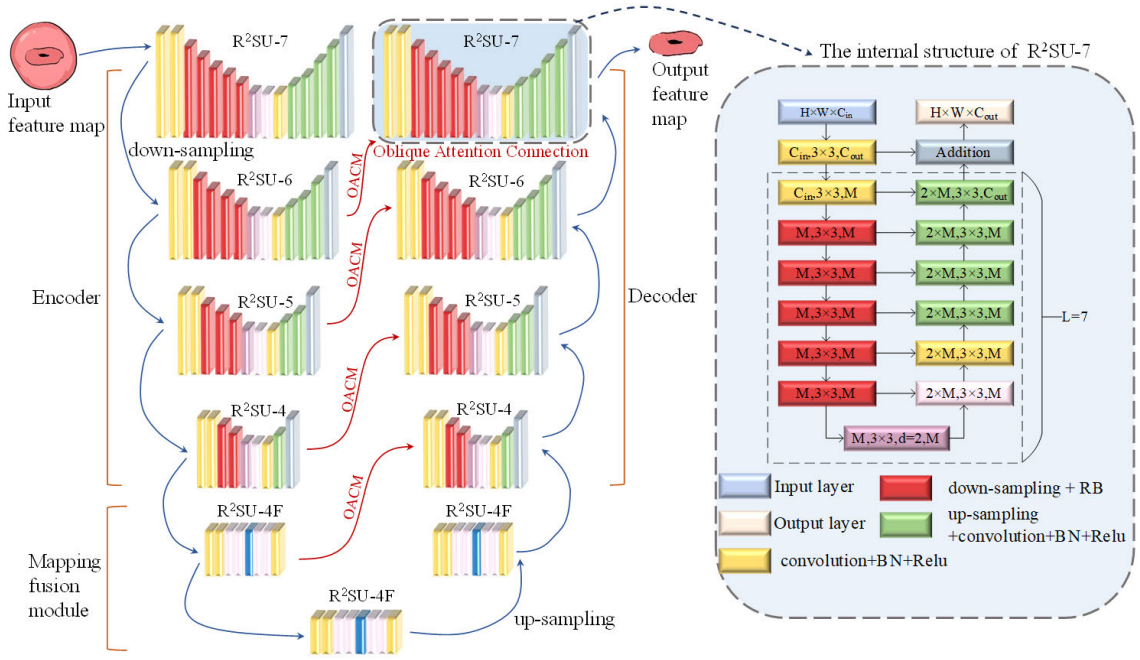

FIGURE 2. RUC-U²Net model and R²SU module structure.

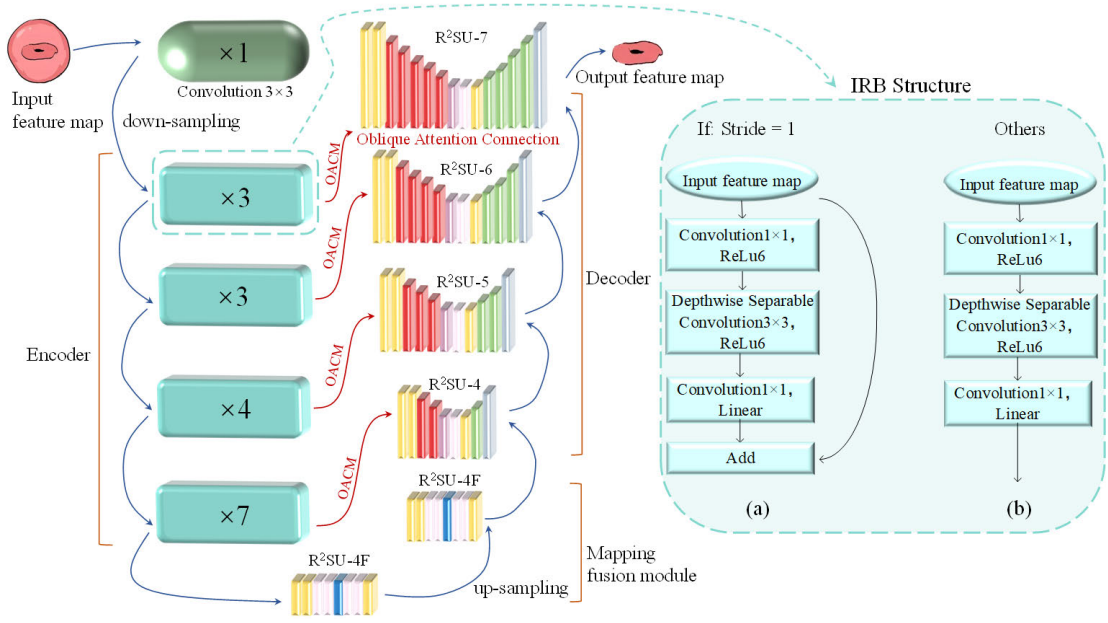

FIGURE 3. RUC⁺-U²Net model and IRB module structure.

overfitting but also overcomes the limitations of random feature elimination found in the Dropout prototype module (Figure 5(b)). The module prefers excluding semantic information considered irrelevant to its context, as demonstrated by the circles in the figure. This architecture enables the network to capture features from proximate information sources, thereby reducing the neuron count and simplifying the network structure.

The Dropblock metric is defined as

$$\kappa = \frac{1 - keep\_prob}{block\_size\, e^2} \frac{feat\_size^{\,2}}{(feat\_size - block\_size + 1)^2} \quad (7)$$

where κ is the probability of critical information loss and the probability generated by the Bernoulli function, $block\_size$ denotes the size of the dropped block and switches to Dropout when $block\_size$ equals 1, $keep\_prob$ reflects the likelihood

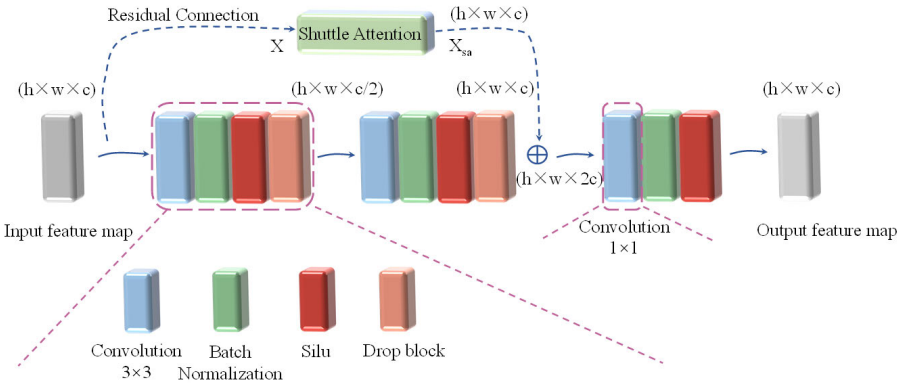

FIGURE 4. RB module structure.

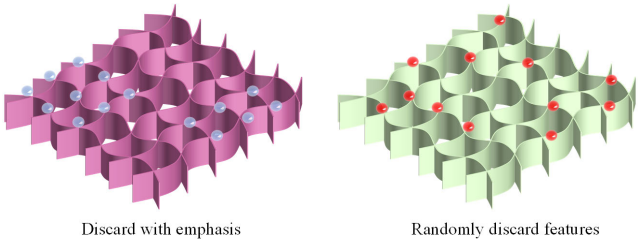

FIGURE 5. Comparison between dropout and dropblock modules. (a) Schematic diagram of Dropout module; (b) Schematic diagram of Dropblock module.

of retaining the cells containing feature details in the feature map and *feat_size* specifies the size of the feature map.

C. OBLIQUE ATTENTION CONNECTION MODULE

To optimize the use of complementary information across different scales, this study replaces the horizontal hopping connection of the U^2-Net encoder and decoder with an oblique attention connection. We introduce the Oblique Attention Connection Module (OACM), which integrates SA and CARAFE, to improve jump connections.

The OACM module consists of SA and CARAFE upsampling modules connected in series. Initially, irrelevant features are filtered out from the encoder-generated feature maps through SA grouping and double-branching processing, as depicted in Figure 6. Subsequently, these feature maps are upscaled using CARAFE.

SA is elaborated upon in Section B, along with a concise overview of the CARAFE upsampling module. CARAFE, a lightweight, plug-and-play model, is specifically engineered for feature upsampling. It begins with a compression of the input feature map's channel number via a 1×1 convolution to reduce computational demands. The upsampling kernels, as predicted by this convolutional layer, are normalized through the Softmax function, ensuring the summed weights of these kernels equal one. The output feature map from the upsampling prediction module is then processed by the feature recombination module, which enhances the semantic information of the feature map by extracting surrounding regions of feature points and performing dot product operations. The metric for the upsampling kernel prediction module is defined as

$$W_{l'} = \psi\left(N\left(X_l, k_{\text{encoder}}\right)\right) \tag{8}$$

where $N\left(X_l, k_{\text{encoder}}\right)$ is a subregion of size $k_{\text{encoder}} \times k_{\text{encoder}}$ of centred at position l, ψ is a prediction module and $W_{l'}$ is the output of upsampling kernels prediction module. The metric for the reorganization module is defined as:

$$X_{l'} = \phi\left(N\left(X_l, k_{up}\right), W_{l'}\right) \tag{9}$$

where ϕ is the content-aware reorganisation module, k_{up} is a sub-region with a range size of $k_{\text{up}} \times k_{\text{up}}$ and $X_{l'}$ is the output value of CARAFE.

Employing the OACM module for upsampling in the encoder-decoder connection significantly improves the RUC-U^{2}Net model's performance. The experimental findings are detailed in Section RESULTS. The OACM module is noted for its light weight and computational efficiency, markedly extending the receptive field range beyond that of models using only subpixel neighborhoods, such as bilinear interpolation models.

D. LOSS FUNCTION

The presence of non-correlated regions within the ROI in the background of positive colposcopy samples can markedly affect diagnostic accuracy. Misclassification of positive and negative samples during model training leads to increased loss values, compromising the model's segmentation capability. To address this issue, this study employs the Focal Tversky loss function instead of the BCE loss function used in U^2-Net. This adjustment allows the model to better understand and classify complex samples, thereby improving its accuracy in detecting lesion areas. The Focal Tversky loss metric is defined as

$$FTL_C = \sum_c \left(1 - TI_c\right)^{1/\gamma} \tag{10}$$

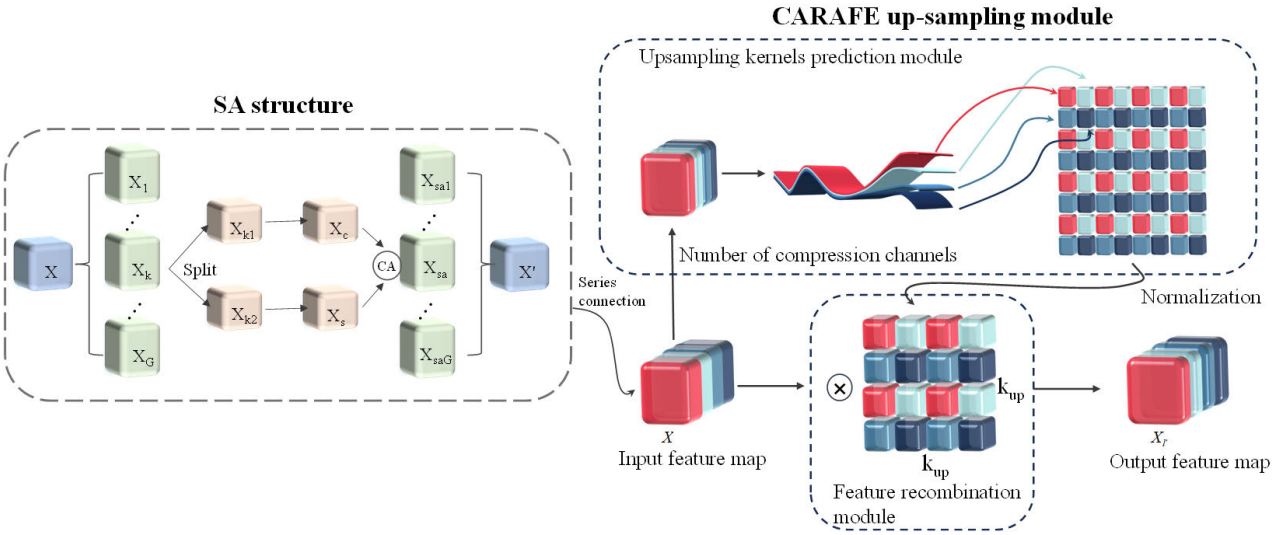

FIGURE 6. The OACM module structure.

TABLE 1. Experimental environment's parameters.

Component	Name/Value
Operating system	Ubuntu 20.0, PyTorch 1.11.0, Python 3.8, Cuda: 11.3
CPU	15vCPU Intel(R) Xeon(R) Platinum 8358P CPU @ 2.60GHz
GPU	NVIDIA RTX 3090 (24GB)
Input image size	224×224
Batch size	16
Epoch	300
Optimizer	Adam
Learning rate	0.001

where TI_c is the Tversky loss function, γ used to adjust loss weights, the larger its value, the smaller thevalue of $1/\gamma$ Increasing the weight of loss within the ROI region of the colposcopy image reduces the weight of loss within the non-ROI region, resulting in more attention to challenging samples. Implementing the Focal Tversky loss function enhances the model's segmentation accuracy, with experimental results and analyses presented in Section RESULTS.

III. RESULTS

A. EXPERIMENTAL ENVIRONMENT AND PARAMETER SETTINGS

The model discussed herein was implemented using the PyTorch 1.11.0 deep learning framework, with Python 3.8 as the programming language. Experiments were carried out on an Ubuntu 20.04 system equipped with an NVIDIA GeForce RTX 3090 graphics card. For the training phase, we configured the settings to include 300 training epochs, an initial learning rate of 0.001, and a batch size of 16, as detailed in Table 1.

B. DATA SETS AND EVALUATION INDICATORS

The dataset for this study consists of 1200 colposcopic images sourced from the publicly available Intel & MobileODT dataset on Kaggle. Annotation was guided by the chief physician at the Affiliated Hospital of North China University of Science and Technology, utilizing Labelme annotation software for precise markings. The dataset was organized into three categories: (1) images showing both inner and outer boundaries of the transformation zone; (2) images with the transformation zone partially visible, both outside and inside the cervix; and (3) images where the transformation zone is entirely within the cervix and not visible. A balanced distribution was maintained across these categories, adhering to a 1:1:1 ratio. The dataset was expanded to 2400 images through augmentation techniques such as cropping, rotating, and mirroring. It was then randomly divided into training and test subsets at a ratio of 2:8.

The study employs several evaluation metrics to gauge the model's segmentation accuracy:

1) PA (Pixel Accuracy): PA quantifies the ratio of accurately predicted pixels to the total pixel count in the image, with higher values denoting superior model performance in capturing detailed segmentation.

$$PA = \frac{\sum_{i=0}^{k} p_{ii}}{\sum_{i=0}^{k} \sum_{j=0}^{k} p_{ij}} \tag{11}$$

2) MPA (Mean Pixel Accuracy): MPA offers an overall evaluation of the model's segmentation accuracy by averaging the accuracy across various classes. Greater MPA values indicate enhanced pixel-level segmentation accuracy by the model.

$$MPA = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij}} \tag{12}$$

3) MIoU (Mean Intersection over Union): MIoU calculates the mean ratio of the intersection to the overlap to the combined area of model-generated segmentation

TABLE 2. Comparison of segmentation results in colposcopic images by different models.

Preference	Models	PA/%	MPA/%	MIoU/%	FMIoU/%	Dice/%	Hausdorff distance/pixel	Paramas/million
2020, Sun [26]	AUNet	84.1089	83.4768	73.2874	78.1245	83.8145	1732.75	75.50
2020, Qin [48]	U^2-Net	92.6113	86.7372	81.5994	86.6840	87.8698	1629.32	44.01
2021, Gao [25]	UTNet	84.7736	80.8763	70.3487	76.9460	83.3246	1856.47	57.45
2022, Kim [10]	SEGNet	86.9990	82.3327	72.9583	77.9793	83.6756	1761.53	29.44
2023, Shinohara [55]	U-Net	86.9426	81.6980	72.7544	77.9162	82.5721	1872.83	13.40
2023, Yang [56]	PSPNet	82.5100	83.0399	71.2477	77.0924	83.1232	1751.91	49.07
2023, Li [57]	DeepLabv3+	86.8120	83.0399	74.2477	78.0924	85.2432	1702.24	59.50
(proposed)	**RUC-U^{2}Net**	**94.5508**	**91.6182**	**84.6094**	**90.6879**	**90.4637**	**1546.75**	42.59
(proposed)	**RUC$^+$-U^{2}Net**	88.0183	83.8283	75.2606	80.2810	85.4625	1721.52	**13.19**

and the ground truth. It equally assesses each class's performance, making it apt for scenarios with class imbalance. Increased MIoU values suggest improved model segmentation in imbalanced distributions of positive and negative samples.

$$MIoU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ij}} \tag{13}$$

4) FWIoU (Frequency Weighted Intersection over Union): FWIoU applies weighting to counteract class imbalance, prioritizing the performance of infrequently occurring classes. A higher FWIoU reflects broader and more inclusive segmentation performance by the model.

$$FMIoU = \frac{1}{\sum_{i=0}^{k}\sum_{j=0}^{k}p_{ij}}\times\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ii}} \tag{14}$$

In equations 11 to 14,where p_{ij} is the probability of predicting category i as category j, p_{ii} represents the probability of predicting category i as category i, p_{ji} represents the probability of predicting category j as category i.

5) DICE (Dice coefficient): DICE is used to evaluate the congruence between binary segmentation outcomes from the model and the actual ground truth. Values approaching 1 suggest greater resemblance. An elevated DICE score indicates that the model's segmentation is closely aligned with the ground truth, highlighting the model's effectiveness in capturing detailed segmentation information.

$$DICE = \frac{2\times TP}{2\times TP+FP+FN} \tag{15}$$

In equation 15, TP (True Positive) stands for accurate positives, FP (False Positive) for incorrect positives, and FN (False Negative) for missed positives.

6) Hausdorff Distance: This metric assesses the congruity between segmentation boundaries produced by the model and the actual boundaries. Contrary to the preceding metrics, a lower Hausdorff Distance suggests a minimal maximum discrepancy between the model's and the actual segmentation boundaries, demonstrating better fidelity of the model's segmented edge integrity.

$$H(A,B) = \max\left(\sup_{a\in A}\inf_{b\in B}d(a,b), \sup_{b\in B}\inf_{a\in A}d(a,b)\right) \tag{16}$$

Here, A and B denote the point sets along the segmentation boundaries created by the model and the actual boundaries, respectively. $d(a,b)$ represents the distance between points a and b.

C. EXPERIMENTAL RESULTS AND ANALYSIS

1) QUANTITATIVE EXPERIMENTS

In this section, the proposed models, RUC-U^{2}Net and RUC$^+$-U^{2}Net, are quantitatively compared with seven other mainstream medical image segmentation models using the dataset introduced in this study. The results are presented in Table 2. The models developed in this research outperform the comparison group across all seven evaluation metrics.

Table 2 illustrates that the RUC-U^{2}Net model yields superior performance on all six evaluation metrics relative to competing models. The RUC$^+$-U^{2}Net model, despite having the least number of parameters and ranking third in segmentation accuracy, demonstrates significant efficiency in colposcopic image segmentation while maintaining a reduced model complexity. This underscores its potential applicability in real-time cervical cancer screening. The U^2-Net model also shows notable effectiveness, closely aligning with the innovations of this paper, especially due to their mutual implementation of an interactive encoder-decoder structure and the utilization of feature information from multilevel decoder outputs. Such approaches markedly improve segmentation accuracy over other networks. Unlike U^2-Net, however, the models described herein incorporate a lightweight attention module and an upsampling module, which collectively enhance feature refinement and enable multi-scale feature fusion, thereby allowing RUC-U^{2}Net to surpass U^2-Net in performance and reduce training time.

2) QUALITATIVE EXPERIMENT

Figure 7 presents the segmentation analysis of colposcopic images using the two models proposed in this study, compared against seven other models.

This figure demonstrates that the segmentation results from RUC-U^{2}Net closely align with the true map. Conversely,

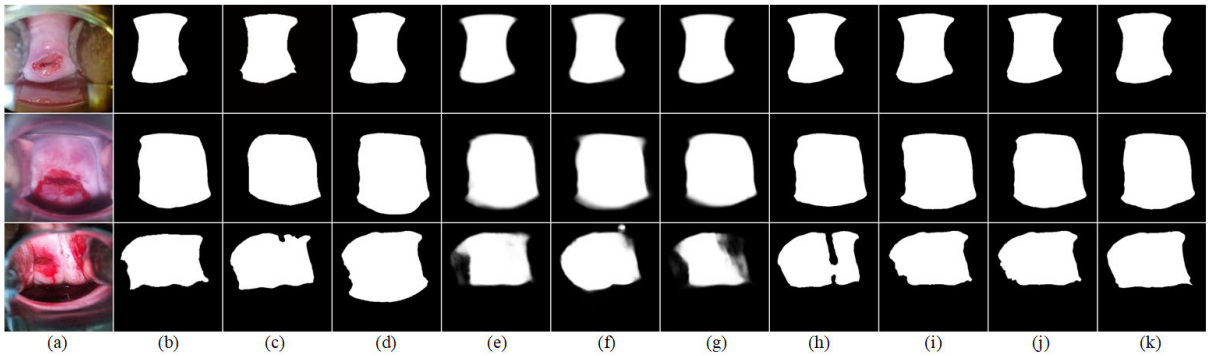

FIGURE 7. Comparison of different models for segmenting colposcopic Images (a) input image; (b) manual annotation of original drawings; (c) U-Net; (d) UTNet; (e) AUNet; (f) SEGNet; (g) PSPNet; (h) DeepLabv3+; (i) U^2-Net; (j) RUC-U^2Net; (k) RUC$^+$-U^2Net.

TABLE 3. Model improvement ablation experiment.

Models	PA/%	MPA/%	MIoU/%	FMIoU/%
U^2-Net	92.6113	86.7372	81.5994	86.6840
U^2-Net+ OACM	93.1367	88.3066	82.3518	87.9445
U^2-Net+RB	93.4186	88.2533	82.4212	88.9649
U^2-Net+ OACM+RB	94.0231	90.6835	83.5835	89.1618
RUC-U^2Net	**94.5508**	**91.6182**	**84.6094**	**90.6879**

SEGNe and PSPNet overlook critical details of the ROI in cervical images, leading to a significant missegmentation rate. While the U-Net model accurately segments the overall shape, its simplistic jump-connection method fails to sufficiently enhance feature information extraction, rendering the segmentation results of these models less refined. The performance of the U^2-Net model, as shown in Table 2 and Figure 7, closely matches that of the RUC-U^2Net model proposed in this paper. Both models benefit from the interactive encoder-decoder structure and the integration of feature map information from multilevel decoder outputs, setting them apart from other networks in terms of segmentation quality. Nevertheless, U^2-Net struggles with accurate edge segmentation, a problem evident in the lower left notch region in Figure 7, column i. The model introduced in this paper, depicted in Figure 7, column j, builds upon the U^2-Net foundation to improve segmentation accuracy. Compared to U^2-Net and the other models, RUC-U^2Net exhibits more precise and detailed edge processing in the ROI of the segmented cervical image, providing significantly enhanced detail and achieving the closest match to the manually annotated ROI.

3) ABLATION EXPERIMENT

In this section, we conducted ablation experiments to assess the effectiveness of RUC-U^2Net, the colposcopic image segmentation model proposed in this study. U^2-Net served as the fundamental network, into which the OACM and RB modules were integrated for experimentation.

To determine if the proposed modules could address the loss of detailed information and incomplete edge segmentation in colposcopic images, the OACM and RB modules were separately integrated into the baseline network to observe their effects. Data in Table 3 show significant improvements in the model's ability to segment detailed information following the integration of the RB module. Specifically, evaluation metrics for segmentation detail demonstrated enhancements: PA increased by 0.8073%, and DICE by 1.2861%. This improvement is attributed to the RB module's capability to refine features during the downsampling stages of the model's core units, indicating its effectiveness in resolving the issue of missing detailed information in colposcopic image segmentation. The inclusion of the OACM module led to a notable improvement in edge segmentation integrity, as evidenced by a significant reduction in the Hausdorff distance by 29.76. This suggests that the module adds finer-scale semantic information to the feature maps at the decoder stage, effectively addressing incomplete edge segmentation in colposcopic images.

Evaluating the combined performance of the OACM and RB modules, when added to the baseline network, revealed a substantial improvement. Table 3 data indicate increases in PA and DICE by 1.4118% and 1.8082%, respectively, while the Hausdorff distance decreased by 73.07. This enhancement is credited to the RB module providing more refined feature maps at the encoder stage, which deliver more essential and abstract semantic information to the OACM module, allowing the decoder to more efficiently and effectively process the input data. Consequently, this combination successfully addresses both the missing detailed information and incomplete edge segmentation issues observed in colposcopic image segmentation.

To lessen the impact of challenging samples on final segmentation accuracy, the loss function was switched to the Focal Tversky loss function. This change led to further improvements in the model's segmentation accuracy, particularly in metrics reflecting an imbalanced distribution of test samples. FWIoU increased by 1.5261%, and MIoU by 1.0259%. This indicates that the Focal Tversky loss function

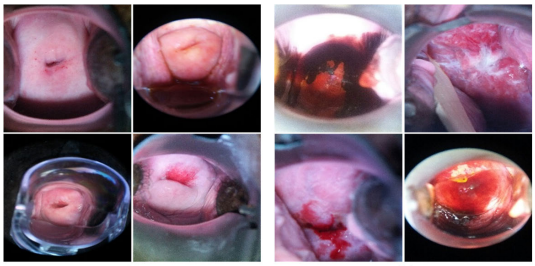

FIGURE 8. Sample experimental data for loss function (a) simple samples; (b) difficult samples.

TABLE 4. Comparison with other loss functions using simple samples.

Models	PA/%	MPA/%	MIoU/%	FMIoU/%
BCELoss	91.4289	86.9356	81.6243	87.2442
DiceLoss	91.0154	87.2546	80.8934	87.4376
Tversky Loss	85.8127	81.9024	68.8496	75.1523
Focal Tversky Loss	**94.5976**	**91.7728**	**84.7938**	**90.8773**

TABLE 5. Comparison with other loss functions using difficult samples.

Models	PA/%	MPA/%	MIoU/%	FMIoU/%
BCELoss	91.0013	86.2536	81.0187	86.2746
DiceLoss	90.1639	86.9854	80.1533	86.9253
Tversky Loss	82.4759	79.7796	67.4628	73.9153
Focal Tversky Loss	**94.5508**	**91.6182**	**84.6094**	**90.6879**

allows the model to concentrate on challenging samples, thereby enhancing the model's segmentation performance.

4) LOSS FUNCTION EXPERIMENTS

For this study, the dataset was manually curated. Images with clear and smooth boundaries within the ROI underwent filtration, yielding a refined sample of 250 images, as shown in Figure 8(a). Conversely, images with blurry and complex boundaries within the ROI, deemed challenging samples, amounted to 250 images, as illustrated in Figure 8(b).

Tables 4 and 5 depict the outcomes of applying the Focal Tversky loss function to the RUC-U^{2}Net segmentation model, in comparison with three other loss functions. Table 4 presents results from experiments on simple samples, while Table 5 focuses on challenging samples. Figure 9 visually represents the findings from the aforementioned tables.

The original loss function utilized in the U^2-Net model was the BCE loss function, which computes the discrepancy between the predicted values generated during model training and the actual values. However, it is prone to instability in training and a high risk of gradient explosion. In contrast, the Dice loss function offers smoother training progress but tends to penalize pixels with medium probability more harshly, leading to the neglect of a broader range of feature information pixels. According to experimental data in Tables 4 and 5, both loss functions exhibit comparable performance. The Tversky loss function, an extension of the Dice loss function, adjusts the balance between false negatives and false positives using hyperparameters. Nevertheless, its performance was less effective on challenging samples, resulting in suboptimal outcomes on this experimental dataset, which included a significant number of difficult samples. The Focal Tversky loss function, an advanced iteration of the Tversky loss function, prioritizes training on challenging samples while diminishing the impact of simpler ones. The model proposed in this study adopts the Focal Tversky loss function for two primary reasons:

1) Robustness to imbalanced data: The Focal Tversky loss function is designed to mitigate issues arising from class imbalance by focusing on learning from minority class samples. This approach prevents the model from favoring the majority class excessively during training and prediction, enhancing performance on minority classes. It is particularly effective in addressing the imbalance of positive and negative samples in the colposcopic image dataset of this study, thereby improving the model's performance on minority classes.
2) Ease of focusing on challenging samples: The Focal Tversky loss function is engineered to lessen the emphasis on easily classifiable samples, allowing the model to concentrate more on challenging samples. This feature is invaluable for overcoming issues of incomplete edge segmentation and the loss of detailed information in colposcopic image segmentation tasks.

Comparative analysis of the data in Tables 4 and 5 reveals that the Focal Tversky loss function outperforms other loss functions, especially as shown in Table 5, indicating its efficacy in learning from challenging samples and achieving more accurate segmentation results. Consequently, this study employs the Focal Tversky loss function for experimentation.

D. GENERALIZATION EXPERIMENTS

To assess the generalizability of the models proposed in this paper, RUC-U^{2}Net and RUC$^+$-U^{2}Net, they were applied to publicly available datasets for retinal vessel segmentation, the Drive dataset, and for salient object detection, the DUT-OMRON dataset. Comparative experiments were conducted against other mainstream medical segmentation models, with the results detailed below.

1) DRIVE RETINAL VESSEL SEGMENTATION PUBLIC DATASET

The Drive Retinal Vessel Segmentation Public Dataset was expanded to 400 images through random cropping, rotation, and mirroring techniques. The dataset was randomly divided into a test set and a training set at a 2:8 ratio.

Quantitative comparisons with other mainstream medical segmentation models demonstrate the superior performance of the model proposed in this study across all six evaluation metrics, as detailed in Table 6.

Figure 10 showcases the segmentation results of retinal vessel images using the proposed model and seven other models, highlighting RUC-U^{2}Net's ability to achieve precise

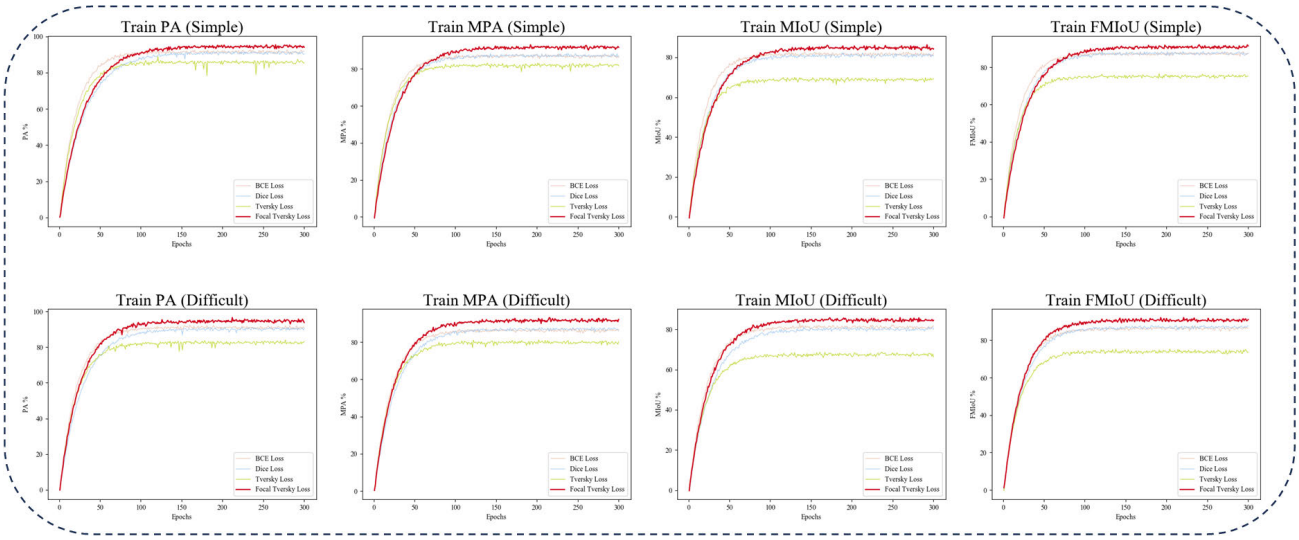

FIGURE 9. Experiment results using simple and difficult samples.

TABLE 6. Comparison of segmentation results in fundus vascular images by different models.

Models	PA/%	MPA/%	MIoU/%	FMIoU/%	Dice/%	Hausdorff distance/pixel
AUNet [26]	89.0378	77.0036	73.5819	83.7923	69.4224	2615.24
U^2-Net [48]	90.4772	77.8379	73.8774	84.0459	70.7621	2580.60
UTNet [25]	88.4024	74.5012	67.9064	78.8174	67.6842	2675.63
SEGNet [10]	86.2533	73.3358	66.1296	77.9503	66.4699	2827.24
U-Net [55]	88.2942	73.3524	66.1335	78.0557	66.7300	2784.17
PSPNet [56]	87.1976	73.9379	67.0504	78.8527	66.7866	2788.29
DeepLabv3+ [57]	90.1715	77.4371	73.6441	83.9412	69.9515	2594.58
RUC-U^2Net(proposed)	**92.5442**	**79.3584**	**76.0308**	**86.1619**	**71.3241**	**2520.34**
RUC^+-U^2Net(proposed)	89.2425	77.2863	73.7363	83.9801	69.8271	2604.61

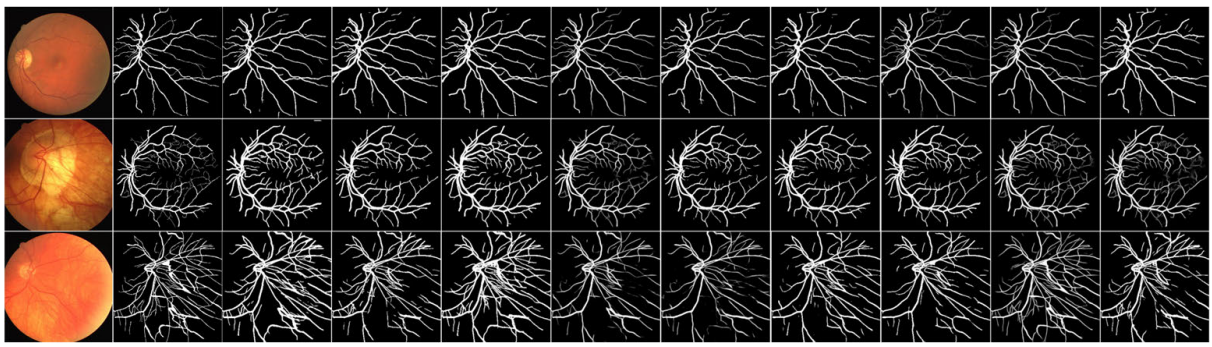

FIGURE 10. Comparison of different models for segmenting Fundus Vascular Images (a) input image; (b) manual annotation of original drawings; (c) U-Net; (d) UTNet; (e) AUNet; (f) SEGNet; (g) PSPNet; (h) DeepLabv3+; (i) U^2-Net; (j) RUC-U^2Net; (k) RUC^+-U^2Net.

segmentation of fine retinal vessels closely matching the manually annotated ROI. Notably, models f and g in the third row of Figure 10 experienced information loss during segmentation. The proposed model (column j) incorporates an adaptive attention mechanism that improves weight allocation between vessels and background in the learning process, effectively addressing the information loss issue encountered by other models. This experiment confirms the strong generalization capability of the proposed model, demonstrating its practical applicability.

2) DUT-OMRON DATASET

The DUT-OMRON public dataset, consisting of 5172 high-quality images, was augmented to 10,344 images through random cropping, rotation, and mirroring. The test and training sets were divided at a 2:8 ratio.

TABLE 7. Comparison of Segmentation results on DUT-OMRON dataset by different models.

Models	PA/%	MPA/%	MIoU/%	FMIoU/%	Dice/%	Hausdorff distance/pixel
AUNet [26]	87.6548	76.7888	69.3590	78.9342	72.5264	2372.55
U^2-Net [48]	92.6818	84.7322	79.8012	87.3034	76.2984	1985.83
UTNet [25]	87.9348	76.3278	69.2847	79.3813	73.2751	2327.46
SEGNet [10]	85.9900	74.6151	66.4167	76.1881	71.8942	2484.73
U-Net [55]	88.8129	77.9679	71.2979	80.8527	73.6428	2218.98
PSPNet [56]	86.8796	75.9686	68.2281	77.8506	72.1377	2414.71
DeepLabv3+ [57]	91.5471	81.9477	76.4816	84.7217	75.8447	2016.41
RUC-U^2Net(proposed)	**94.0080**	**86.5301**	**81.1306**	**89.9635**	**77.7824**	**1953.18**
RUC^+-U^2Net(proposed)	90.3337	79.0719	73.1683	83.2734	75.4812	2039.57

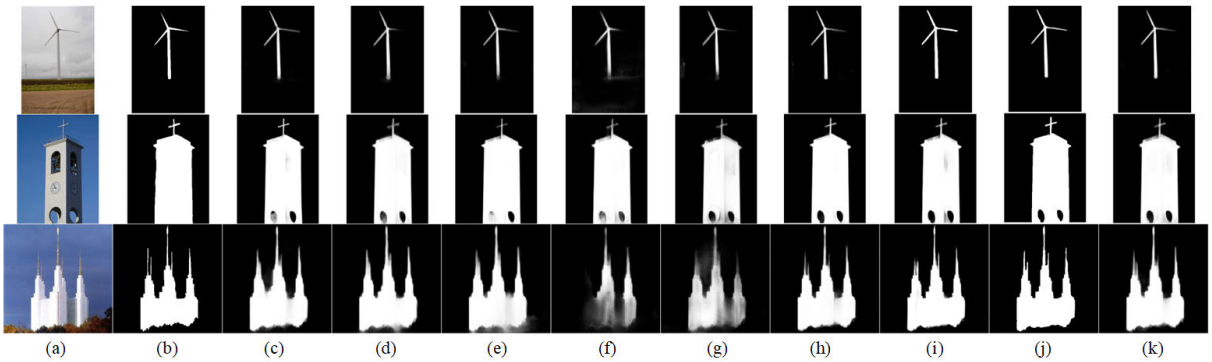

FIGURE 11. Comparison of different models for segmenting fundus vascular images (a) input image; (b) manual annotation of original drawings; (c) U-Net; (d) UTNet; (e) AUNet; (f) SEGNet; (g) PSPNet; (h) DeepLabv3+; (i) U^2-Net; (j) RUC-U^2Net; (k) RUC^+-U^2Net.

The model introduced in this study was quantitatively evaluated against other leading medical segmentation models, with the findings detailed in Table 7. Relative to competing models in the experiment, the model proposed herein exhibits enhanced performance across all six evaluation metrics.

Figure 11 displays the segmentation outcomes of semantic images using the model proposed in this study alongside seven other distinct models. The figure shows that the model proposed in this paper, RUC-U^2Net, achieves clearer segmentation of wind turbine blades and main architectural structures, with a minimal inclusion of ambiguous regions in the segmentation results. Its boundary distinction capability is superior to that of the other models. As illustrated in the third row, column j of Figure 11, in comparison to the segmentation results from the previous seven columns, the RUC-U^2Net model exhibits more accurate segmentation of church architectural images. Especially when compared to columns d and f, the proposed model not only avoids incorrectly segmenting stones at the bottom of the image as part of the church but also effectively captures the detailed contours and edges of the church, resulting in a segmentation that closely matches the manually annotated ROI. This experiment reaffirms the model's strong generalization capability and its high suitability for practical applications.

IV. DISCUSSION

Cervical colposcopy analysis typically involves three sequential stages: detection of the cervical region, extraction of lesion-related features, and diagnostic evaluation. Our study concentrates on the first stage, detecting the cervix region, with the aim of enhancing the efficiency of ROI detection in colposcopic images and reducing the reliance on manual operations by trained gynecologists:

The proposed RUC-U2^{2}Net model enables automatic detection of ROI in colposcopic images. By learning to identify complex edge and detail information, it tackles various challenges such as instrument occlusion, hair artifacts, edge lesions, and image blurriness. Introducing an automated segmentation network aids doctors in concentrating on potentially abnormal areas, thereby improving the efficiency of cervical cancer detection. It lightens doctors' workload, cuts down on the time needed for manual annotations, and reduces judgment biases caused by subjective factors. This method enhances the analysis of colposcopic images, making the process more efficient, accurate, and facilitating quicker utilization of medical resources.

Currently, in the field of CAD, deep learning has shown considerable promise. Numerous studies utilize deep learning techniques to support clinical doctors in colposcopic examinations, primarily focusing on the classification of CIN [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21]. Almost all these studies begin with the common preprocessing step of extracting the ROI from colposcopic examination images. Image segmentation, a critical step in training deep learning-based models, is of utmost importance. More precise segmentation of colposcopic images can greatly improve the accuracy of cervical cancer lesion classification using deep learning techniques [3]. Thus, this study proposes a

deep learning-based segmentation model of higher accuracy. Through this model, we offer insights into areas within colposcopic image analysis that merit clinical doctors' attention, aiding in the diagnostic process. This model not only lays a solid foundation for further research into lesion grade classification and biopsy site selection but also provides more accurate auxiliary information. Moreover, extensive generalization experiments indicate that this approach can also serve as an effective auxiliary tool for other medical diagnostic tasks, enhancing overall diagnostic performance. Additionally, this study adopts an end-to-end research approach, maximizing the retention of patient cervical image ROI feature information, thereby assisting doctors in making the most accurate judgments based on original data.

In the experimental analysis, the proposed RUC-U^{2}Net model for colposcopic image segmentation demonstrated optimizations of 1.9395%, 2.5939%, and 82.57 for the PA, DICE, and Hausdorff distance metrics, respectively. Applied to retinal vessel image segmentation, the model improved these evaluation metrics by 2.0670%, 0.562%, and 60.26. Moreover, for segmentation of the DUT-OMRON dataset, enhancements of 1.3262%, 1.484%, and 32.65 were observed across the three metrics. The PA and DICE metrics gauge segmentation detail accuracy, while the Hausdorff distance assesses the integrity of segmented edges. A comprehensive assessment reveals the model's exceptional performance in segmenting colposcopic images, characterized by intricate details and edge information. This success is ascribed to the RB module introduced in this study, which supplies more sophisticated feature maps during the encoding phase. These maps provide essential and abstract semantic information for further processing by the OACM module, facilitating a more efficient and effective learning process by the decoder, particularly in addressing the challenges of missing detailed information and incomplete edge segmentation identified in this study. In the final stage, the original loss function was substituted with the Focal Tversky loss function to lessen the impact of challenging samples on ultimate segmentation accuracy. This modification led to further accuracy improvements, especially in conditions of imbalanced test sample distribution, with FWIoU and MIoU increasing by 1.5261% and 1.0259%, respectively. This demonstrates the Focal Tversky loss function's effectiveness in focusing the model on challenging samples, thus boosting its segmentation capability. From these quantitative experiments, it is clear that the proposed model excels in segmenting colposcopic images with complex details, edge nuances, and uneven sample distribution, providing a solid basis for future research into lesion grade classification and biopsy site selection.

Future research may benefit from adopting design principles from immune-based artificial intelligence models. For example, the LapEIKR [58] model linearly combines multiple kernels to address non-linear challenges effectively, enhancing feature extraction in datasets with complexities akin to colposcopic images. Addressing significant target area variations, a method [59] introduces an infrared target segmentation approach utilizing immune growth fields and clone thresholds, applying a clone selection algorithm for optimal thresholding. This strategy allows for the adaptive adjustment of growth parameters, achieving accurate target image segmentation. The Immune Coordinated Deep Network (ICDNet) [60] offers a unique method by segmenting the target area into multiple blocks for feature extraction and classification, incorporating prior information to refine target feature extraction, potentially enhancing medical image segmentation accuracy. Additionally, addressing colposcopic image challenges like insufficient lighting and occlusions, a study employs [61] multi-scale Gaussian functions for illumination correction and innate immune mechanisms at the segmentation onset to reduce environmental impact. These preprocessing methods could provide valuable insights for future research.

While the RUC-U^{2}Net model effectively addresses issues of missing fine-grained details and incomplete edge segmentation, it faces limitations such as a high parameter count and extended training durations. Future efforts will explore strategies to decrease the model's parameter count. For applications requiring a model with fewer parameters, the RUC$^+$-U^{2}Net model, also proposed in this study, offers a viable alternative, delivering superior colposcopic image segmentation with a reduced parameter footprint.

V. CONCLUSION

This study introduces an attention-based RUC-U^{2}Net structure for colposcopic image segmentation, aiming to overcome challenges associated with the loss of fine-grained information and incomplete edge segmentation. Initially, the RB module is incorporated within the R^{2}SU unit of the RUC-U^{2}Net framework, enhancing feature refinement during downsampling and increasing the accuracy of model outputs. This development addresses the issue of missing fine-grained information in colposcopic image segmentation. Additionally, the OACM module is developed, leveraging attention-based diagonal connections between the encoder and decoder to effectively merge detailed and semantic features. This integration of information from various scales enriches the feature extraction process in the decoder phase, addressing incomplete edge segmentation in colposcopic images. To tackle the issue of data imbalance and improve segmentation accuracy, the Focal Tversky loss function replaces the BCE loss function. The RUC-U^{2}Net architecture is materialized into two models: RUC-U^{2}Net and the streamlined RUC$^+$-U^{2}Net. Experimental evaluations on three public datasets confirm that both models surpass competing methods. Future applications of these models may significantly aid physicians in cervical cancer screening within cervical cancer recognition systems.

In conclusion, this paper emphasizes colposcopic image segmentation technology to reduce manual processing time and aid physicians in devising more efficient and precise diagnostic approaches. By enhancing diagnostic efficiency, it is possible to facilitate early diagnosis for potential cervical

cancer patients, ultimately aiming to decrease cervical cancer incidence. This approach bears considerable practical significance for the diagnostic process in gynecology and the advancement of intelligent healthcare.

REFERENCES

[1] M. F. Janicek and H. E. Averette, "Cervical cancer: Prevention, diagnosis, and therapeutics," *CA, A Cancer J. Clinicians*, vol. 51, no. 2, pp. 92–114, Mar. 2001.

[2] A. Buskwofie, G. David-West, and C. A. Clare, "A review of cervical cancer: Incidence and disparities," *J. Nat. Med. Assoc.*, vol. 112, no. 2, pp. 229–232, Apr. 2020.

[3] J. Park, H. Yang, H.-J. Roh, W. Jung, and G.-J. Jang, "Encoder-weighted W-Net for unsupervised segmentation of cervix region in colposcopy images," *Cancers*, vol. 14, no. 14, p. 3400, Jul. 2022.

[4] M. Silveira, J. C. Nascimento, J. S. Marques, A. R. S. Marcal, T. Mendonca, S. Yamauchi, J. Maeda, and J. Rozeira, "Comparison of segmentation methods for melanoma diagnosis in dermoscopy images," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 1, pp. 35–45, Feb. 2009.

[5] B. H. Brown and J. A. Tidy, "The diagnostic accuracy of colposcopy—A review of research methodology and impact on the outcomes of quality assurance," *Eur. J. Obstetrics Gynecol. Reproductive Biol.*, vol. 240, pp. 182–186, Sep. 2019.

[6] P. Xue, M. T. A. Ng, and Y. Qiao, "The challenges of colposcopy for cervical cancer screening in LMICs and solutions by artificial intelligence," *BMC Med.*, vol. 18, no. 1, p. 169, Jun. 2020.

[7] Z. Alyafeai and L. Ghouti, "A fully-automated deep learning pipeline for cervical cancer classification," *Exp. Syst. Appl.*, vol. 141, Mar. 2020, Art. no. 112951.

[8] J. Liu, Q. Chen, J. Fan, and Y. Wu, "HSIL colposcopy image segmentation using improved U-Net," in *Proc. 36th Youth Acad. Annu. Conf. Chin. Assoc. Autom. (YAC)*, May 2021, pp. 891–897.

[9] K. Canfell, J. J. Kim, and M. Brisson, "Mortality impact of achieving WHO cervical cancer elimination targets: A comparative modelling analysis in 78 low-income and lower-middle-income countries," *Lancet*, vol. 395, pp. 591–603, Feb. 2020.

[10] J. Kim, C. M. Park, S. Y. Kim, and A. Cho, "Convolutional neural network-based classification of cervical intraepithelial neoplasias using colposcopic image segmentation for acetowhite epithelium," *Sci. Rep.*, vol. 12, no. 1, p. 17228, Oct. 2022.

[11] K. T. Desai, B. Befano, Z. Xue, and H. Kelly, "The development of 'automated visual evaluation' for cervical cancer screening: The promise and challenges in adapting deep-learning for clinical testing," *Int. J. Cancer*, vol. 150, no. 5, pp. 741–752, 2022.

[12] X. Chen, X. Pu, Z. Chen, L. Li, K. Zhao, H. Liu, and H. Zhu, "Application of EfficientNet-B0 and GRU-based deep learning on classifying the colposcopy diagnosis of precancerous cervical lesions," *Cancer Med.*, vol. 12, no. 7, pp. 8690–8699, Apr. 2023.

[13] S. Angara, P. Guo, Z. Xue, and S. Antani, "Semi-supervised learning for cervical precancer detection," in *Proc. IEEE 34th Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jun. 2021, pp. 202–206.

[14] O. E. Aina, S. A. Adeshina, A. P. Adedigba, and A. M. Aibinu, "Classification of cervical intraepithelial neoplasia (CIN) using fine-tuned convolutional neural networks," *Intell.-Based Med.*, vol. 5, Jan. 2021, Art. no. 100031.

[15] V. Chandran, M. G. Sumithra, A. Karthick, T. George, M. Deivakani, B. Elakkiya, U. Subramaniam, and S. Manoharan, "Diagnosis of cervical cancer based on ensemble deep learning network using colposcopy images," *BioMed Res. Int.*, vol. 2021, pp. 1–15, May 2021.

[16] Y. Yu, J. Ma, W. Zhao, Z. Li, and S. Ding, "MSCI: A multistate dataset for colposcopy image classification of cervical cancer screening," *Int. J. Med. Informat.*, vol. 146, Feb. 2021, Art. no. 104352.

[17] R. Elakkiya, V. Subramaniyaswamy, V. Vijayakumar, and A. Mahanti, "Cervical cancer diagnostics (don't short) healthcare system using hybrid object detection adversarial networks," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 4, pp. 1464–1471, Apr. 2022.

[18] C. Yuan, Y. Yao, B. Cheng, Y. Cheng, Y. Li, Y. Li, X. Liu, X. Cheng, X. Xie, J. Wu, X. Wang, and W. Lu, "The application of deep learning based diagnostic system to cervical squamous intraepithelial lesions recognition in colposcopy images," *Sci. Rep.*, vol. 10, no. 1, p. 11639, Jul. 2020.

[19] B.-J. Cho, Y. J. Choi, M.-J. Lee, J. H. Kim, G.-H. Son, S.-H. Park, H.-B. Kim, Y.-J. Joo, H.-Y. Cho, M. S. Kyung, Y.-H. Park, B. S. Kang, S. Y. Hur, S. Lee, and S. T. Park, "Classification of cervical neoplasms on colposcopic photography using deep learning," *Sci. Rep.*, vol. 10, no. 1, p. 13652, Aug. 2020.

[20] S. K. Saini, V. Bansal, R. Kaur, and M. Juneja, "ColpoNet for automated cervical cancer screening using colposcopy images," *Mach. Vis. Appl.*, vol. 31, no. 3, pp. 1–15, Mar. 2020.

[21] T. Zhang, Y.-M. Luo, P. Li, P.-Z. Liu, Y.-Z. Du, P. Sun, B. Dong, and H. Xue, "Cervical precancerous lesions classification using pre-trained densely connected convolutional networks with colposcopy images," *Biomed. Signal Process. Control*, vol. 55, Jan. 2020, Art. no. 101566.

[22] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 7553.

[23] J. Wang, H. Zhu, S.-H. Wang, and Y.-D. Zhang, "A review of deep learning on medical image analysis," *Mobile Netw. Appl.*, vol. 26, no. 1, pp. 351–380, 2021.

[24] X. Chen, X. Wang, K. Zhang, K.-M. Fung, T. C. Thai, K. Moore, R. S. Mannel, H. Liu, B. Zheng, and Y. Qiu, "Recent advances and clinical applications of deep learning in medical image analysis," *Med. Image Anal.*, vol. 79, Jul. 2022, Art. no. 102444.

[25] Y. Gao, M. Zhou, and D. N. Metaxas, "UTNet: A hybrid transformer architecture for medical image segmentation," in *Proc. Med. Image Comput. Comput. Assist. Interv. (MICCAI)*, Sep. 2021, pp. 61–71.

[26] H. Sun, C. Li, B. Liu, Z. Liu, M. Wang, H. Zheng, D. D. Feng, and S. Wang, "AUNet: Attention-guided dense-upsampling networks for breast mass segmentation in whole mammograms," *Phys. Med. Biol.*, vol. 65, no. 5, Feb. 2020, Art. no. 055005.

[27] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Aug. 2017, pp. 2881–2890.

[28] X. Zhu, Z. Cheng, S. Wang, X. Chen, and G. Lu, "Coronary angiography image segmentation based on PSPNet," *Comput. Methods Programs Biomed.*, vol. 200, Mar. 2021, Art. no. 105897.

[29] L. Yan, D. Liu, Q. Xiang, Y. Luo, T. Wang, D. Wu, H. Chen, Y. Zhang, and Q. Li, "PSP net-based automatic segmentation network model for prostate magnetic resonance imaging," *Comput. Methods Programs Biomed.*, vol. 207, Aug. 2021, Art. no. 106211.

[30] W. Rong, Z. Yang, and L. Leng, "Channel group-wise drop network with global and fine-grained-aware representation learning for palm recognition," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Oct. 2022, pp. 1–9.

[31] Y. Zhang, J. Chu, L. Leng, and J. Miao, "Mask-refined R-CNN: A network for refining object details in instance segmentation," *Sensors*, vol. 20, no. 4, p. 1010, Feb. 2020.

[32] W. Lin, J. Chu, L. Leng, J. Miao, and L. Wang, "Feature disentanglement in one-stage object detection," *Pattern Recognit.*, vol. 145, Jan. 2024, Art. no. 109878.

[33] K. Jiao, P. Xu, and S. Zhao, "A novel automatic parameter setting method of PCNN for image segmentation," in *Proc. IEEE 3rd Int. Conf. Signal Image Process. (ICSIP)*, Jul. 2018, pp. 265–270.

[34] Z. Lin, Z. He, X. Wang, B. Zhang, C. Liu, W. Su, J. Tan, and S. Xie, "DBGANet: Dual-branch geometric attention network for accurate 3D tooth segmentation," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Nov. 8, 2024, doi: 10.1109/TCSVT.2023.3331589.

[35] G. Chen, L. Li, Y. Dai, J. Zhang, and M. H. Yap, "AAU-Net: An adaptive attention U-Net for breast lesions segmentation in ultrasound images," *IEEE Trans. Med. Imag.*, vol. 42, no. 5, pp. 1289–1300, May 2023.

[36] Y. Li, Y. Zhang, J.-Y. Liu, K. Wang, K. Zhang, G.-S. Zhang, X.-F. Liao, and G. Yang, "Global transformer and dual local attention network via deep-shallow hierarchical feature fusion for retinal vessel segmentation," *IEEE Trans. Cybern.*, vol. 53, no. 9, pp. 5826–5839, Sep. 2023.

[37] W. Liu, C. Li, N. Xu, T. Jiang, M. M. Rahaman, H. Sun, X. Wu, W. Hu, H. Chen, C. Sun, Y. Yao, and M. Grzegorzek, "CVM-cervix: A hybrid cervical pap-smear image classification framework using CNN, visual transformer and multilayer perceptron," *Pattern Recognit.*, vol. 130, Oct. 2022, Art. no. 108829.

[38] M. M. Rahaman, C. Li, Y. Yao, F. Kulwa, X. Wu, X. Li, and Q. Wang, "DeepCervix: A deep learning-based framework for the classification of cervical cells using hybrid deep feature fusion techniques," *Comput. Biol. Med.*, vol. 136, Sep. 2021, Art. no. 104649.

[39] J. Shi, R. Wang, Y. Zheng, Z. Jiang, H. Zhang, and L. Yu, "Cervical cell classification with graph convolutional network," *Comput. Methods Programs Biomed.*, vol. 198, Jan. 2021, Art. no. 105807.
[40] S. P. Oliveira, D. Montezuma, A. Moreira, D. Oliveira, P. C. Neto, A. Monteiro, J. Monteiro, L. Ribeiro, S. Gonçalves, I. M. Pinto, and J. S. Cardoso, "A CAD system for automatic dysplasia grading on H&E cervical whole-slide images," *Sci. Rep.*, vol. 13, no. 1, p. 3970, Mar. 2023.
[41] L. Cao, J. Yang, Z. Rong, L. Li, B. Xia, C. You, G. Lou, L. Jiang, C. Du, H. Meng, W. Wang, M. Wang, K. Li, and Y. Hou, "A novel attention-guided convolutional network for the detection of abnormal cervical cells in cervical cancer screening," *Med. Image Anal.*, vol. 73, Oct. 2021, Art. no. 102197.
[42] C.-W. Wang, Y.-A. Liou, Y.-J. Lin, C.-C. Chang, P.-H. Chu, Y.-C. Lee, C.-H. Wang, and T.-K. Chao, "Artificial intelligence-assisted fast screening cervical high grade squamous intraepithelial lesion and squamous cell carcinoma diagnosis and treatment planning," *Sci. Rep.*, vol. 11, no. 1, p. 16244, Aug. 2021.
[43] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III*. Berlin, Germany: Springer, 2015, pp. 234–241.
[44] Y. Liu, B. Bai, H.-C. Chen, P. Liu, and H.-M. Feng, "Cervical image segmentation using U-Net model," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst. (ISPACS)*, Dec. 2019, pp. 1–2.
[45] X. Zhang and S. Zhao, "Cervical image classification based on image segmentation preprocessing and a CapsNet network model," *Int. J. Imag. Syst. Technol.*, vol. 29, no. 1, pp. 19–28, Mar. 2019.
[46] Y. J. Gaona, D. C. Malla, B. V. Crespo, M. J. Vicuña, V. A. Neira, S. Dávila, and V. Verhoeven, "Radiomics diagnostic tool based on deep learning for colposcopy image classification," *Diagnostics*, vol. 12, no. 7, p. 1694, Jul. 2022.
[47] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
[48] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. R. Zaiane, and M. Jagersand, "U^2-Net: Going deeper with nested U-structure for salient object detection," *Pattern Recognit.*, vol. 106, Oct. 2020, Art. no. 107404.
[49] Q.-L. Zhang and Y.-B. Yang, "SA-Net: Shuffle attention for deep convolutional neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 2235–2239.
[50] J. Wang, K. Chen, R. Xu, Z. Liu, C. C. Loy, and D. Lin, "CARAFE: Content-aware reassembly of features," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3007–3016.
[51] N. Abraham and N. M. Khan, "A novel focal Tversky loss function with improved attention U-Net for lesion segmentation," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 683–687.
[52] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, vol. 16, 2016, pp. 770–778.
[53] C.-Y. Wang, A. Bochkovskiy, and H.-Y. Mark Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 2022, *arXiv:2207.02696*.
[54] G. Ghiasi, T. Y. Lin, and Q. V. Le, "DropBlock: A regularization method for convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–11.
[55] T. Shinohara, K. Murakami, and N. Matsumura, "Diagnosis assistance in colposcopy by segmenting acetowhite epithelium using U-Net with images before and after acetic acid solution application," *Diagnostics*, vol. 13, no. 9, p. 1596, Apr. 2023.
[56] J. Yang, Y. Zhang, Y. Liu, S. Liu, T. Chaikovska, and C. Liu, "Automatic segmentation of cervical precancerous lesions in colposcopy image using pyramid scene parsing network and transfer learning," *Rev. Comput. Eng. Stud.*, vol. 10, no. 2, pp. 28–34, Jun. 2023.
[57] Z. Li, C.-M. Zeng, Y.-G. Dong, Y. Cao, L.-Y. Yu, H.-Y. Liu, X. Tian, R. Tian, C.-Y. Zhong, T.-T. Zhao, J.-S. Liu, Y. Chen, L.-F. Li, Z.-Y. Huang, Y.-Y. Wang, Z. Hu, J. Zhang, J.-X. Liang, P. Zhou, and Y.-Q. Lu, "A segmentation model to detect cevical lesions based on machine learning of colposcopic images," *Heliyon*, vol. 9, no. 11, Nov. 2023, Art. no. e21043.
[58] T. Yang, D. Fu, and C. Wu, "Laplacian embedded infinite kernel model for semi-supervised classification," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 30, no. 10, Dec. 2016, Art. no. 1650022.
[59] X. Yu, Z. Zhou, Q. Gao, D. Li, and K. Ríha, "Infrared image segmentation using growing immune field and clone threshold," *Infr. Phys. Technol.*, vol. 88, pp. 184–193, Jan. 2018.
[60] Z. Zhou, B. Zhang, and X. Yu, "Immune coordination deep network for hand heat trace extraction," *Infr. Phys. Technol.*, vol. 127, Dec. 2022, Art. no. 104400.
[61] X. Yu, X. Ye, and S. Zhang, "Floating pollutant image target extraction algorithm based on immune extremum region," *Digit. Signal Process.*, vol. 123, Apr. 2022, Art. no. 103442.

YUXI HE was born in 1998. She is currently pursuing the master's degree with the North China University of Science and Technology. Her research interests include deep learning and image processing, with specialization in medical image segmentation tasks. She has participated in the Hebei Provincial Scientific Research Programme and the Shanghai Electronic Information Vocational and Technical College High-Level Talent Initiation Project.

LIPING LIU is currently a Professor with Shanghai Electronic Information Vocational and Technical College, China. She also heads different research projects at both the provincial and university levels, which include the Hebei Provincial Scientific Research Programme, Shanghai Electronic Information Vocational and Technical College High-Level Talent Initiation Project, and Hebei Provincial Cultivation Programme of Scientific and Technological Innovation Ability for University Students. She leads the Tangshan Sanitary Ceramics Quality Intelligent Monitoring Technology Basic Innovation Team. Additionally, she has multiple patented innovations. Her research interests include pattern recognition, intelligent systems, and mining engineering.

JINLIANG WANG was born in 1993. He is currently pursuing the master's degree with the North China University of Science and Technology. His research interests include machine vision, image processing, and pattern recognition. He has a particular proficiency in utilizing deep learning algorithms to extract data from pointer meters. He has participated in Hebei Provincial Scientific Research Programme and Hebei Provincial Cultivation Programme of Scientific and Technological Innovation Ability for University Students.

NANNAN ZHAO is currently an Associate Chief Physician of Obstetrics and Gynaecology with the Affiliated Hospital, North China University of Science and Technology, where she has demonstrated exceptional proficiency in clinical research pertaining to obstetrics and gynaecology. Her research interests include the various common and frequent ailments affecting obstetrics and gynaecology patients. Furthermore, she has several patents to her name in relation to medical inventions.

HANGYU HE was born in Nanchong, Sichuan, in 2002. He is currently pursuing the bachelor's degree. His main research interests include the application of artificial intelligence, big data, and data mining technologies in fields, such as medicine and aquaculture.