RESEARCH ARTICLE

Automatic delineation of clinical target volumes and organs at risk in cervical cancer radiotherapy using ResAU-Net

Xueping Liu[1], Guiping Zhao[1, *], Silu Ding[2], Shi Bai[3], Jiahui Zhang[1], Yukai Lu[1], Jingjing Du[1], Xingya Liu[1]

[1]College of Artificial Intelligence, Shenyang Aerospace University, Shenyang, Liaoning, China. [2]Department of Radiation Oncology, The First Hospital of China Medical University, Shenyang, Liaoning, China. [3]School of Information Science and Engineering, Shenyang University of Technology, Shenyang, Liaoning, China.

Received: April 11, 2024; **accepted:** April 25, 2024.

Cervical cancer is the most prevalent gynecological malignancy worldwide, and radiotherapy is a primary clinical treatment approach. Segmentation of clinical target volumes (CTVs) and organs at risk (OARs) in planning cervical cancer radiotherapy is critical and time-consuming. This study aimed to achieve accurate and efficient automatic delineation of CTVs and OARs in cervical cancer, providing robust support for clinical diagnosis and treatment. A comprehensive dataset comprising 156 patient cases was constructed by employing advanced image augmentation techniques to enhance dataset variability. ResAU-Net, a novel image segmentation model that integrates residual and attention modules into the U-Net framework, was introduced to enhance segmentation accuracy and performance. A comparative analysis of four commonly used loss functions including CE, SCE, Dice, and Lovász was performed for cervical cancer image segmentation to identify the most suitable one for the task. In addition, a comprehensive evaluation approach was employed by utilizing the Dice similarity coefficient (DSC) and 95th percentile Hausdorff distance (95HD) as quantitative metrics, and expert evaluation by experienced radiation oncologists to ensure the clinical relevance and robustness of the findings. The results showed that, by comparing four commonly used loss functions, the Lovász loss function was selected for this study, which exhibited clear advantages in cervical cancer image segmentation. Comparative analysis with other models indicated that the approach developed in this study performed well in cervical cancer segmentation tasks. The DSC values of the proposed method were 0.907, 0.727, 0.827, and 0.828 for urinary bladder, rectum, lymphatic drainage target area (CTV-N), and uterine cervix target area (CTV-UV), respectively. The 95HD values were 2.253, 1.985, 9.588, and 4.967 mm for urinary bladder, rectum, CTV-N, and CTV-UV, respectively. Radiation oncologists evaluated the automated delineation results and found 87.53% of the results to be highly acceptable. This study utilized deep learning technology to provide a more accurate and efficient method for automatically delineating CTVs and OARs in cervical cancer, offering strong support for clinical diagnosis and treatment planning.

Keywords: artificial intelligence; deep learning; automatic delineation; cervical cancer; radiation therapy.

[*]**Corresponding author:** Guiping Zhao, College of Artificial Intelligence, Shenyang Aerospace University, Shenyang, Liaoning, China. Email: 1252415034@qq.com.

Introduction

Cervical cancer was ranked the second most prevalent malignancy among women, posing significant threats to female health and life [1]. Radiotherapy constitutes a primary clinical approach for treating cervical cancer, wherein the core principle involves delivering a sufficient

prescribed dose to the clinical target volumes (CTVs) while minimizing radiation exposure to surrounding organs at risk (OARs) [2, 3]. Precise segmentation of CTVs and OARs is crucial for achieving accurate cancer cell eradication and reducing radiation-induced damage to typical structures during radiotherapy. Manual delineation of CTVs and OARs by radiation oncologists remains the clinical standard, yet its accuracy hinges on their interpretation of clinical guidelines, consensus, and experience [4, 5]. Inter-observer variations in delineation quality, efficiency, and time-consuming manual delineation procedures affect physician workload. Disparities in tumor target delineation among physicians or institutions can lead to heterogeneous dose distributions and complicate the standardization of assessment [6, 7].

Automated delineation methods based on atlas libraries have gained prominence in clinical practice to address this repeated, time-consuming task [8], which leverage model libraries and image registration techniques to optimally match anatomical features of reference and target images, enabling automated delineation of target regions and OARs. However, these methods often require patient-specific models, more universality, and time-consuming elastic registration processes, limiting their clinical applicability. Recent advances in artificial intelligence, particularly deep learning, have provided a broader scope for image segmentation in the medical field, offering valuable applications in medical imaging diagnosis, surgical navigation, and precision medicine [9]. Automatic segmentation of CTVs and OARs in radiotherapy patient images using deep learning methods driven by large datasets has garnered attention for its potential to substantially reduce manual delineation time and enhance segmentation consistency and accuracy. Various models and algorithms have been discussed for this purpose. Nonetheless, the diverse morphologies, sizes, and lesions positions of cervical cancer necessitate deep learning segmentation methods to exhibit robustness and accurately delineating lesion regions across varying tumor shapes. Irregular edges are typical in cervical cancer lesions, often displaying branching, serration, and other complex features. Segmentation algorithms must effectively address such irregular edges while accurately outlining lesion regions. Image quality in cervical cancer images can be influenced by factors like scanning equipment, parameters, patient positioning, and muscle tension, demanding the segmentation algorithm to possess good generalization capabilities. In recent years, deep learning methods based on convolutional neural networks (CNN) have emerged as the predominant solution for biomedical image segmentation [10]. Notable CNN architectures include LeNet [11], AlexNet [12], ResNet [13], and U-Net. Researchers have applied CNN to traditionally challenging tumor target segmentation tasks. Lin *et al*. employed 3D CNN to segment nasopharyngeal cancer lesions from magnetic resonance imaging (MRI) images, achieving a Dice similarity coefficient (DSC) of 79% [14]. Men *et al*. utilized deep deconvolutional neural networks to segment nasopharyngeal primary lesions, lymph nodes, and CTVs from computed tomography (CT) images with DSC values of 80.9%, 62.3%, and 82.6%, respectively [15]. Trebeschi *et al*. utilized CNNs for rectal tumor segmentation, achieving a DSC of 69% [16]. Chen *et al*. improved the U-Net architecture for esophagus and esophageal cancer segmentation, yielding a DSC of 79% [17]. These studies have initially demonstrated the effectiveness of CNN-based models for automatic tumor target segmentation [18]. Among CNN-based models, U-Net is a widely favored choice, which features an end-to-end encoder-decoder architecture known for its "U" shape [19]. It addresses the limitations of the fully convolutional network (FCN) in preserving pixel spatial information and contextual details, preventing the loss of local and global features. U-Net has achieved a high intersection over union (IoU) of 92% and performed remarkably well in the ISBI 2014 glioblastoma segmentation challenge [20]. Due to its outstanding performance and favorable architecture, U-Net

swiftly becomes a pivotal network model for medical image segmentation. It has been found to be applicable to segmenting lesions of organs like the eyes, heart, liver, brain, skin, prostate, and breast [21]. U-Net has been employed by researchers for male pelvic CT image segmentation involving the prostate, bladder, and rectum [22]. Balagopal *et al*. introduced a 3D U-Net algorithm based on ResNeXt for prostate, bladder, rectum, and femoral head segmentation [23]. Wang *et al*. proposed a multi-channel U-Net for automatic pelvic segmentation [24]. Further, Li *et al*. proposed an attention-based nested U-Net (ANUNet) for organ cancer segmentation, incorporating redesigned dense skip connections and attention mechanisms and yielding promising results on liver tumor segmentation datasets (LiTS) and combined healthy abdominal organ segmentation datasets (CHAOS) [19, 25]. Zhang *et al*. introduced the DIU-Net network for segmentation in retinal vessel images, lung CT images, and brain tumor MRI images, integrating inception and dense connection modules into the U-Net structure to enhance feature learning [26].

Distinct from natural images, medical images are challenging to gather and label, even for experienced radiation oncologists. Furthermore, due to patient privacy concerns, extensive data sharing with the research community is limited, leaving a gap in publicly available cervical cancer CTVs datasets. Therefore, the feasibility of deep learning-based automatic segmentation of cervical cancer CTVs and OARs, and strategies to enhance segmentation accuracy effectively remain research-worthy topics. The present study initially investigated the efficacy of our in-house dataset and the feasibility of deep learning-based methods for cervical cancer CTVs delineation by employing U-Net and its enhanced models. U-Net with its simple contraction path of two standard convolutional and pooling operations is prone to feature loss. While increasing network depth can capture more features, deeper networks face issues like gradient vanishing and decreased accuracy. This study proposed a novel model to effectively tackle these challenges, thereby enhancing the accuracy of segmentation. This research introduced a cervical cancer CTVs and OARs segmentation network based on ResAU-Net, which was built upon U-Net with optimization primarily focused on two aspects including (1) incorporating ResNet as the feature extraction frontend, leveraging ResNet networks to accelerate convergence and overcome gradient vanishing and degradation issues from deep network structures, effectively capturing deeper semantic information, and (2) replacing skip connection components with attention block, which filters out noise during low-to-high level connection processing with minimal computational cost, enhancing useful low-level features, and avoiding excessive use, thereby boosting model sensitivity and prediction accuracy, leading to improved precision in cervical cancer CTVs and OARs segmentation. This approach would not only significantly reduce the time and labor costs of manual delineation but also enhance the consistency and accuracy of the delineation.

Materials and methods

CT image acquisition

The anonymous image dataset consists of CT scans from 156 cervical cancer patients aged from 20 to 75 years old and collected from May 2022 to June 2023 at the First Affiliated Hospital of China Medical University in Shenyang, Liaoning Province, China. All patients were pathologically diagnosed. To ensure consistency across the dataset, a series of measures were implemented during the CT image collection process, which included the implementation of strict data collection standards to ensure that all CT scans adhered to the DICOM protocol for image standardization and consistency, and careful attention being paid to patient sample selection and distribution to avoid potential biases within the dataset. The dataset was partitioned into mutually exclusive training, validation, and test sets in a case ratio of

111:30:15 [27]. The design of this segmentation scheme considered the broad variability in patient characteristics, ensuring comprehensive coverage of diverse features during different stages of model training and evaluation. During the delineation of patient anatomical structures, experienced radiation oncologists from the radiation oncology department of the hospital were involved in meticulously delineating contours following clinical guidelines to ensure the consistency in delineation across identical anatomical structures. The delineated regions covered four common categories found in cervical cancer CT images including lymphatic drainage target area (CTV-N), the uterine cervix target area (CTV-UV) and the organs at risk, urinary bladder, and rectum (Figure 1). Furthermore, to ensure the delineation quality, all contours were jointly reviewed and revised by three radiation oncologists from the First Affiliated Hospital of China Medical University with their practice years in the specialty of 33, 13, and 7 years, respectively. The emphasis was eventually placed on monitoring and managing the overall quality of the dataset. Regular internal reviews and updates were conducted to identify and promptly rectify inconsistencies or errors. This systematic data management process contributed to ensuring the high quality and consistency of the dataset, laying a solid foundation for robust training of deep learning models.

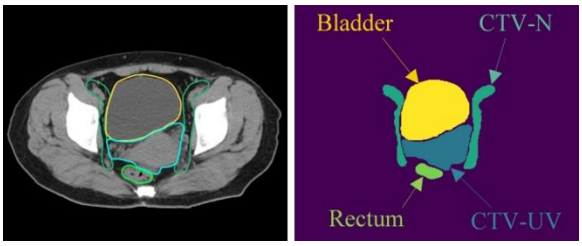

Figure 1. Regions delineated in cervical cancer CT images.

Dataset preprocessing

The DICOM files containing patient anatomical structures and delineation contours were processed using an in-house program, converting them into respective two-dimensional matrices. DICOM values were first transformed into a Hounsfiled unit (HU) to process the data, reflecting the absorption characteristics of various tissues to X-rays. The image window width and window level were adjusted to allow the optimal extraction of the HU range of target tissues in the image, facilitating optimal tissue display. The CT information outside this range was removed to reduce the computational burden on the deep learning model. A window width of 300 and a window level of 50 were applied in this study. The CT matrices underwent normalization to eliminate convergence issues arising from singular samples during training. To minimize computational load, the CT matrices were cropped from 512 × 512 to 256 × 256 pixels based on patient anatomical structure position information, ensuring that the patient's anatomical structures were centered within the matrix. Data augmentation was performed on the training set to prevent overfitting, which included flipping, rotation, random color augmentation, and random affine elastic distortion to diversify the dataset and enhance the model's robustness.

Random affine elastic distortion

This study employed random affine elastic deformation for data augmentation, aiming to enhance the adaptability of models to various morphologies and poses of medical images by introducing randomness and elastic deformation. Briefly, random affine transformations were applied to images with transformation matrices determined using the three-point method. Then, displacement values were randomly generated for each pixel in the transformed images and a convolution operation with a Gaussian distribution was performed to generate a random displacement field. A control factor was introduced to adjust the deformation intensity of the displacement field, thereby controlling the smoothness of the images before the random displacement field was applied to the original images, and the final deformed images were obtained using nearest neighbor

Figure 2. Random affine elastic deformation. The leftmost image was the original image, while the two images on the right depicted the images after random affine elastic deformation.

interpolation (Figure 2). Random affine elastic deformation could introduce various forms of deformation including rotation, scaling, translation, and shearing, thereby generating images with greater diversity. In medical imaging, patients may exhibit significant differences in body posture, position, organ motion, *etc*. By introducing diverse deformations, it was possible to cover the variations in different anatomical structures better. Additionally, in medical imaging, models are prone to overfitting specific morphologies or structures due to significant inter-individual differences among patients and limited sample size. Introducing randomness in affine elastic deformation helped alleviate the sensitivity of models to specific morphologies, thereby improving the generalization performance of the models.

ResAU-Net

Cervical cancer CTVs and OARs segmentation are challenging due to the varying sizes, shapes, and positions of these organs with some having isolated regions and unclear boundaries. Compared to designed convolution combinations, U-Net employs primary two-layer convolution and pooling operations in the contracting path for feature extraction, which can lead to partial feature loss. In this study, ResNet34 was utilized as the front-end backbone network for feature extraction to address this issue, which effectively preserved information and reduced information loss, enhancing the model's performance and generalization while avoiding issues such as gradient divergence and accuracy decline associated with deeper networks. ResNet34 comprised five downsampling stages including a 7 × 7 convolutional layer and four convolutional blocks. The four upsampling stages of U-Net were retained, while the traditional skip connections were replaced by attention blocks. These blocks selectively focused on essential regions within the image, enhanced the feature representation of these regions, and ultimately achieved channel-level feature fusion. One upsampling operation was involved to restore the original image size as the final step. The specific design structure of the ResAU-Net model was depicted in Figure 3.

Attention block

The inspiration for attention mechanisms came from human visual attention, allowing models to dynamically focus on essential parts of input data by learning dynamic weights. This selective focus enhanced the model's performance and representation capabilities, enabling it to handle complex tasks and data better. Mnih *et al*. were the first to introduce attention mechanisms into RNN models for image classification, achieving good performance [28]. Subsequently, attention mechanisms were applied to specific tasks such as text processing and machine translation. Vaswani *et al*. proposed a transformer network structure entirely based on attention mechanisms, which exhibited substantial superiority in quality and parallelism, propelling the application of attention mechanisms to new heights [29]. Ozan *et al*. introduced a novel attention module for medical images that

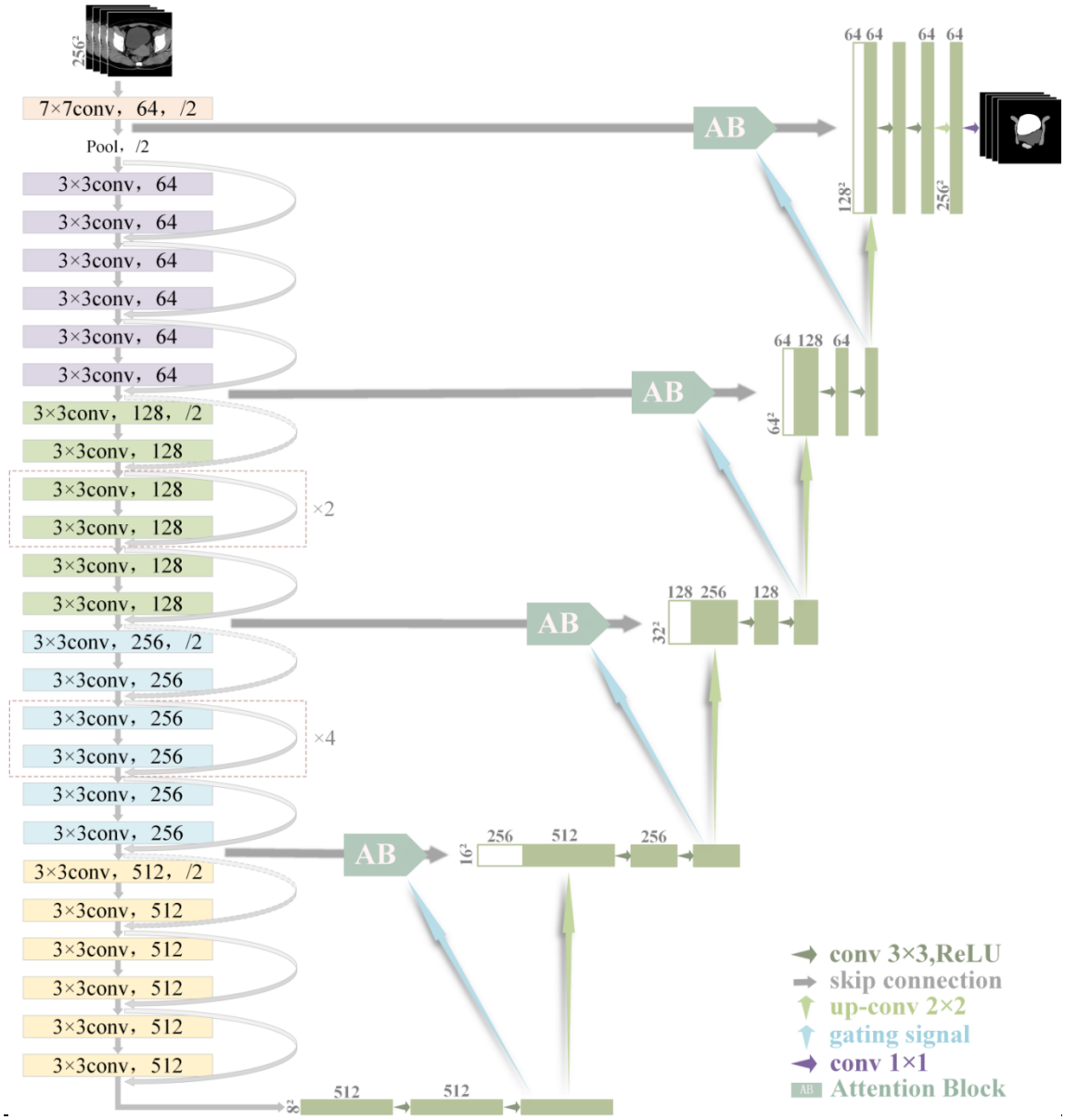

Figure 3. ResAU-Net architecture.

automatically learns to focus on structures of different shapes and sizes [30, 31]. These modules implicitly generated soft region proposals and highlight salient features useful for specific tasks. Moreover, they did not introduce significant computational overhead and did not require many model parameters like multi-model frameworks. Instead, the module enhanced sensitivity and accuracy in dense label prediction by suppressing feature activations from irrelevant regions [32]. In this study, certain dataset regions exhibited unclear contrast between target and non-target tissues in pixel values. Connecting the output features of encoding blocks directly with features generated after deconvolutional upsampling could introduce noise to the primary features, influencing the final output and decreasing prediction accuracy. Therefore, skip connections with attention mechanisms were introduced in this study to suppress interference from non-target tissues in each layer's feature extraction

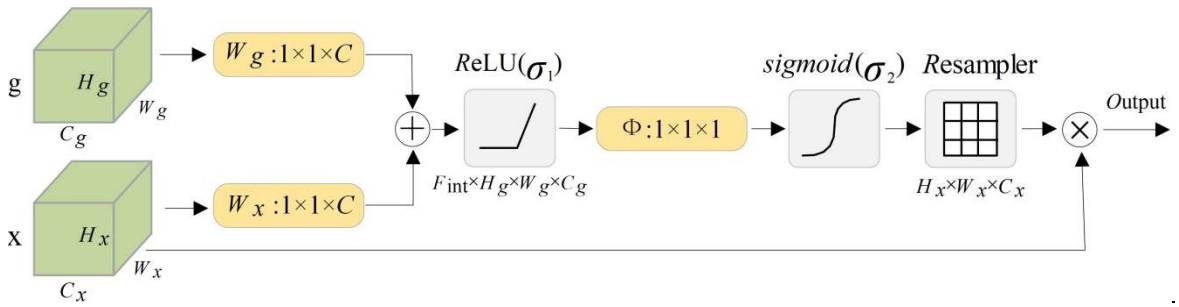

Figure 4. Attention block architecture.

results. By assigning different weights to target and non-target features, emphasis was placed on learning target features, reducing random interference noise from non-target features like background and other tissues, thereby enhancing the accuracy of cervical cancer CTVs and OARs segmentation. The specific structure and principles of the attention block were illustrated in Figure 4 and can be divided into feature extraction, feature weight update, and feature map update three parts. In the feature extraction part, the matrices *g* and *x* were the encoding and decoding feature maps. *H* and *W* were the height and width of the feature map. *C* was the number of channels. W_g and W_x were the feature weight matrix and were calculated through mean pooling by using Equations 1 and 2, respectively, which yielded weight matrices containing information from the *C* channels of the feature map.

$$W_g = \frac{1}{H \times W} \sum_{i=0}^{H} \sum_{j=0}^{W} g(i,j) \quad (1)$$

$$W_x = \frac{1}{H \times W} \sum_{i=0}^{H} \sum_{j=0}^{W} x(i,j) \quad (2)$$

Feature weight update was calculated as follows.

$$q_{att} = \Phi^T \left(\sigma_1 \left(W_g^T g + W_x^T x \right) \right) \quad (3)$$

where σ_1 was the ReLU activation function. W_g dot multiplication g and W_x dot multiplication *x* realized the fully connected operation of encoding and decoding feature maps. The result was summed and passed through a ReLU activation function followed by another fully connected operation of the result with Φ dot multiplication, yielding the intermediate matrix of attention, denoted as q_{att}. The intermediate result passing through the sigmoid activation function σ_2 was calculated using Equation 4.

$$\alpha = \sigma_2 \left(q_{att} \left(x; \theta_{att} \right) \right) \quad (4)$$

where ϑ_{att} was a set of parameters containing W_g, W_x, and Φ. W_g and W_x were learned through weight backpropagation using Equations 3 and 4, updating the feature weights of *g* and *x*. The updated weights were then mapped back to the feature maps. The updated weight matrix α was dot multiplied with the feature map *x* as shown in Equation 5. This process amplified the channel weights of the desired CTVs or OARs target pixels and reduced the weights of other background tissues, resulting in the attention mechanism feature map output. This map was then connected to the decoding network for fusion learning.

$$output = x \times \alpha \quad (5)$$

Loss function

This study employed four commonly used loss functions to train the model to comprehensively compare the performance of these four loss functions in the cervical cancer image

segmentation task and identify the most suitable loss function for this specific task, which included Cross-entropy loss (CE) that was applicable to multi-class classification tasks and exhibited suboptimal performance when dealing with pixel-level tasks such as image segmentation, especially in class imbalance or ambiguous boundaries; Symmetric-cross-entropy loss (SCE) that was introduced to better handle class imbalance issues and amplified the penalty for misclassifications, directing the model's focus toward learning from minority classes; Dice loss (Dice) that could effectively address class imbalance issues, especially in situations with ambiguous boundaries. However, it may have some challenges in optimization problems, particularly when dealing with many background pixels; and Lovász-softmax loss (Lovász), a loss function designed based on the Lovász extension theorem, that performed exceptionally well in addressing non-convex problems in image segmentation tasks, particularly in cases with sharp boundaries but existing class imbalance. However, it might be sensitive to noise in specific scenarios, and its computation is relatively complex [33]. Different loss functions might yield diverse effects on various tasks and datasets. By comparing the performance through multiple loss functions, researchers can gain insights into the strengths and weaknesses of each loss function in specific contexts, guiding the selection of the most appropriate loss function for the task. In some image segmentation tasks, there might be an imbalance in the number of samples across different classes. Some loss functions were specifically designed to address such imbalance. Using different loss functions allowed for better handling of this situation, enhancing the model's sensitivity to minority classes. Employing multiple loss functions contributed to improving the robustness of the model. If a particular loss function performed poorly under certain conditions, other loss functions could compensate for its shortcomings, thereby enhancing the model's performance across different scenarios.

Assessment of proposed model

1. Performance metrics

Performance evaluation of the proposed model was objectively quantified using the dice similarity coefficient (DSC) and the 95th percentile Hausdorff distance (95HD). The DSC measured the overlap between the manual and the automatic segmentation and was defined as:

$$\mathrm{DSC}(A,B)=\frac{2|A\cap B|}{|A|+|B|} \quad (6)$$

where A was the volume of manually generated contours. B was the volume of contours automatically generated by the model. A∩B was the volume of the intersection between A and B. The DSC value was between 0 and 1 (0 = no overlap, 1 = complete overlap). The 95HD was defined as follows:

$$95\mathrm{HD}(A,B)=\mathrm{percentile}(h(A,B)\,Y\,h(B,A),95\mathrm{th}) \quad (7)$$

$$h(A,B)=\max_{a\in A,\,b\in B}(\min(\|a-b\|)) \quad (8)$$

Where $\|\cdot\|$ was the Euclidean norm of the points of A and B. As the 95HD decreased, the overlap between A and B increased. mDSC and mHD were evaluation metrics for overall segmentation results, considering all segmentation targets.

2. Model evaluation by radiation oncologists

Fifteen randomly selected patients comprising 678 image slices were chosen for evaluation by radiation oncologists. Each slice featured both manually and automatically delineated contours. Two radiation oncologists with the practice experiences of 33 and 13 years, respectively, were assigned to evaluate the contours. Evaluation results for the automatically delineated contours were categorized into four levels including acceptable that the segmentation was perfect and completely acceptable for treatment, minor

revision that the segmentation needed a few minor editing but no significant clinical impact without correction, major revision that the segmentation needed significant revision before proceeding to the treatment planning, and unacceptable that the segmentation was unacceptable and needed to be redrawn.

Implementation of model

Most deep learning model training was executed on the NVIDIA GeForce RTX 4080 (16GB NVIDIA). The study employed the Windows 10 operating system. The related experimental code was developed using Python 3.8 (https://www.python.org/) and edited using PyCharm (JetBrains, Prague, Czech). The deep learning framework employed PyTorch (https://pytorch.org/) and computations related to deep learning were executed on GPU using the CUDA computing platform (Nvidia, Santa Clara, CA, USA). This study utilized the Adam optimization algorithm for model training. The decay coefficients were set to 0.9 and 0.999, respectively. A grid search was employed to fine-tune the learning rate and batch size. Grid search exhaustively explored predefined hyperparameter combinations, forming a grid where each node represented a specific hyperparameter combination. Potential learning rates were initially defined, typically on a logarithmic scale, such as 0.1, 0.01, 0.001, 0.0001, and a series of potential batch sizes, such as 16, 32, 64. Based on practical considerations, further subdivisions might be applied. All possible combinations of learning rates and batch sizes were arranged into a grid. For instance, the combinations for the aforementioned learning rates and batch sizes would be (0.1, 16), (0.1, 32), (0.1, 64), (0.01, 16), and so on. For each hyperparameter combination, the model was trained using the corresponding learning rate and batch size from the grid, and the performance was evaluated on the validation set. The optimal combination of learning rate and batch size was chosen based on the performance on the validation set. During the model training process, an early stopping mechanism was introduced. The model's loss on the validation set was continuously monitored, and if the loss did not decrease for three consecutive epochs, the training process was terminated, which helped to prevent overfitting on the training set, enhance generalization performance, and effectively avoid potential local optima traps. The clever aspect of this strategy laid in its ability to promptly identify turning points in the model's performance during training, ensuring efficient and effective training.

Results and discussion

Quantitative performance metrics

The quantitative experiments using U-Net, Attention U-Net, U-Net++, and proposed model were performed. The performances of different models based on the DSC and mDSC metrics were shown in Figure 5a. The results showed that ResAU-Net achieved the best results in both DSC and mDSC, demonstrating superior segmentation accuracy in different regions compared to the other models. The 95HD and mHD metrics reflected the model's ability to capture edge details with smaller values indicating better performance (Figure 5b). Similarly, ResAU-Net exceled in both metrics, indicating more accurately handling of edge details in segmented regions. The results of different models' performances in various regions demonstrated that ResAU-Net exhibited better overall performance than the other models in terms of DSC, mDSC, 95HD, and mHD (Table 1). Specifically, the DSC values for CTV-UV, CTV-N, rectum, and bladder were 0.828, 0.827, 0.727, and 0.907, respectively, with an mDSC value of 0.856. The 95HD values for CTV-UV, CTV-N, rectum, and bladder were 4.967, 9.588, 1.985, and 2.253, respectively, with an mHD value of 4.698. The results indicated that the proposed ResAU-Net model achieved superior performance in cervical cancer image segmentation tasks, especially in handling regions with irregular shapes and blurred edges, demonstrating a more stable and reliable performance.

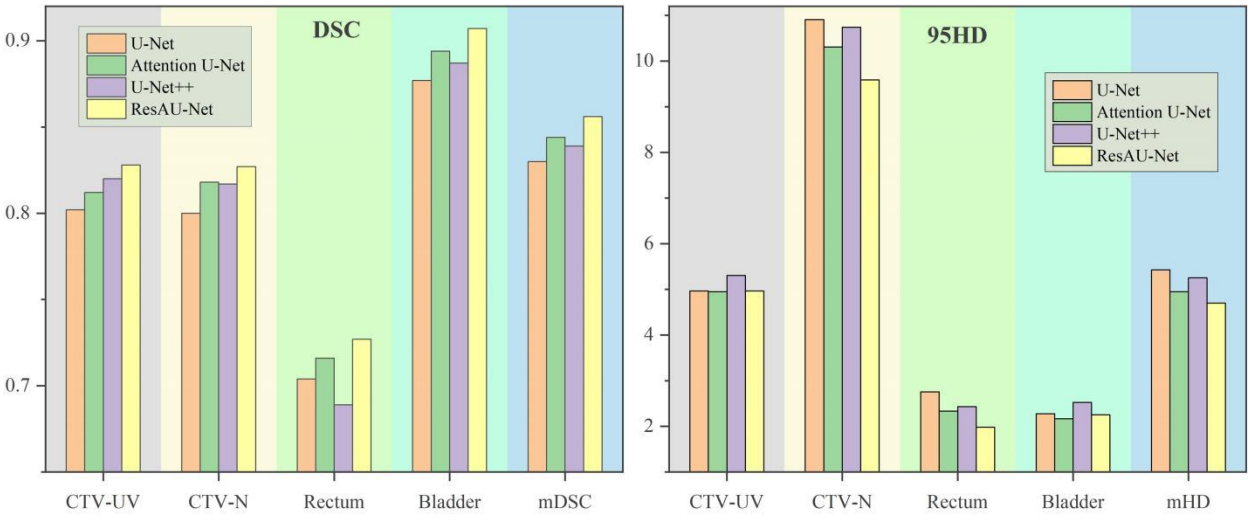

Figure 5. Performance comparison of different models.

Table 1. The comparison of DSC, 95HD, mDSC, mHD values of U-Net, Attention U-Net, U-Net++, and proposed model.

Methods	Anatomy	DSC	95HD	mDSC	mHD
U-Net	Bladder	0.877	2.277	0.830	5.427
	Rectum	0.704	2.757		
	CTV-N	0.800	10.910		
	CTV-UV	0.802	4.965		
Attention U-Net	Bladder	0.894	2.167#	0.844	4.950
	Rectum	0.716	2.336		
	CTV-N	0.818	10.310		
	CTV-UV	0.812	4.950#		
U-Net++	Bladder	0.887	2.527	0.839	5.252
	Rectum	0.689	2.433		
	CTV-N	0.817	10.740		
	CTV-UV	0.820	5.308		
ResAU-Net (Ours)	Bladder	0.907#	2.253	0.856*	4.698*
	Rectum	0.727#	1.985#		
	CTV-N	0.827#	9.588#		
	CTV-UV	0.828#	4.967		

Notes: # indicated the optimal performance within the same segmentation target. * indicated the model's optimal performance.

To provide a more intuitive comparison of the strengths and weaknesses of each algorithm, partial segmentation images from different algorithms were presented. The images demonstrated that the proposed ResAU-Net in this study achieved the best segmentation results for the cervical cancer target area and organs at risk (Figure 6).

Radiation oncologist evaluation

Table 2 demonstrated the assessment results by radiation oncologists with values indicating the number of slices and their corresponding percentages. The terms "Acceptable" and "Minor revision" were deemed acceptable by the radiation oncologists and suitable for clinical application. The results showed that the clinical acceptable percentages were 94.99%, 88.06%, 85.99%, and 81.12% for urinary bladder, rectum,

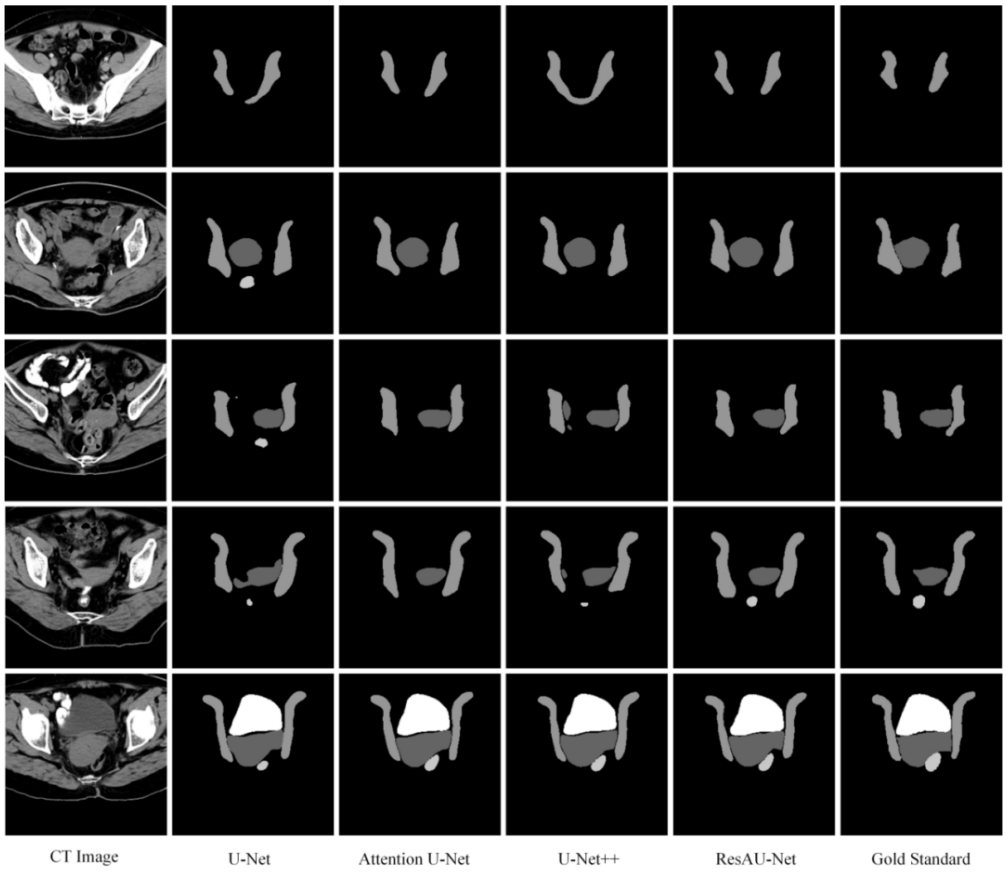

Figure 6. Visual comparison of cervical cancer image segmentation results.

Table 2. Evaluation of the prediction results from our proposed model by experienced oncologists.

Anatomy	Acceptable	Minor revision	Major revision	Not acceptable
Bladder	616 (90.86%)	28 (4.13%)	14 (2.06%)	20 (2.95%)
Rectum	536 (79.06%)	61 (9.00%)	23 (3.39%)	58 (8.55%)
CTV-N	466 (68.73%)	117 (17.26%)	69 (10.18%)	26 (3.83%)
CTV-UV	460 (67.85%)	90 (13.27%)	61 (9.00%)	67 (9.88%)
Mean Percentage	76.62%	10.91%	6.16%	6.31%

CTV-N, and CTV-UV, respectively, with the average acceptability for all segmented targets of 87.53%, which was deemed acceptable in the test set and implied that the proposed model could provide relatively accurate segmentation, demonstrating good concordance with manually delineated contours. Furthermore, the model was deterministic, ensuring consistent segmentation results and offering physicians an objective delineation reference. However, 6.31% of cases still involved oncologists, deeming the automatically generated contours unacceptable. A retrospective analysis of these exceptional cases revealed many unacceptable results from the model's difficulty determining when an organ or structure appeared or disappeared. Such instances limited the generalizability of the model, thus requiring more caution. Nevertheless, experienced radiation oncologists could promptly correct it based on clinical judgment.

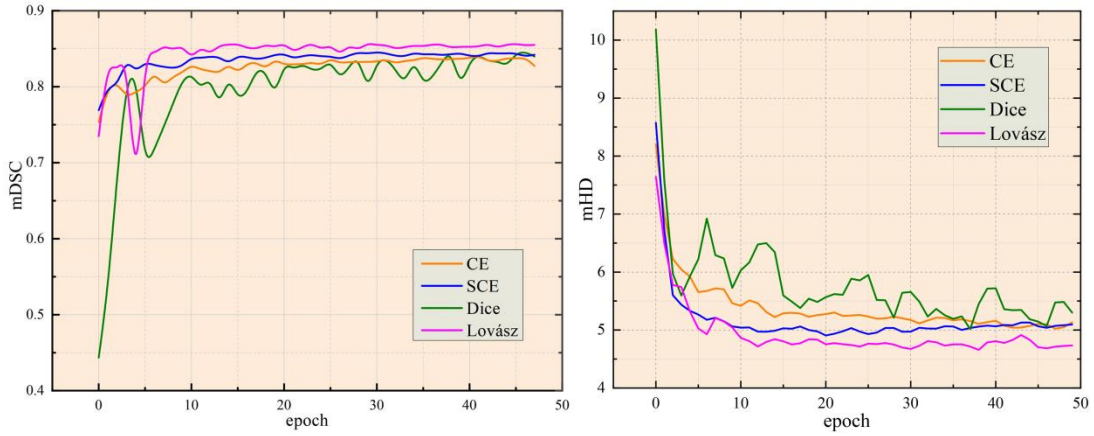

Figure 7. The mDSC and mHD results on the validation set achieved by proposed model when employing four loss functions of CE, SCE, Dice, and Lovász.

Loss function

The mDSC and mHD results obtained on the validation set using the four aforementioned loss functions were shown in Figure 7. The mDSC value of the Lovász loss function was 0.856, surpassing the second-place SCE of 1.2%. The mHD value of the Lovász loss function was 4.698, leading the second-place SCE by 0.379. The performance of Dice loss function was the worst one. The CE loss function was sensitive to imbalanced samples and might inaccurately predict minority classes. The SCE loss function required manual setting of class weights to address sample imbalance, making it less straightforward to determine. Moreover, it did not consider segmentation boundary continuity and smoothness, potentially leading to unclear boundaries. The Dice loss function faced gradient vanishing issues during training, complicating the training process. The Lovász loss function evaluated boundary continuity and accurately assessed boundary smoothness in image segmentation tasks. By minimizing the misclassification of segmentation boundaries, it better handled sample imbalance. Although the Lovász loss function had higher computational complexity, it yielded more accurate segmentation results, making it suitable for tasks requiring high precision. Therefore, the Lovász loss function might be more suitable for cervical cancer image segmentation tasks. In future studies, exploring the application of the Lovász loss function to other computer vision tasks and conducting more in-depth comparative analyses with other loss functions can further enhance model performance and accuracy.

Conclusion

This study addressed the critical need to accurately segment clinical target volumes (CTVs) and organs at risk (OARs) in planning cervical cancer radiotherapy. By establishing a comprehensive dataset and employing advanced image augmentation techniques, dataset variability and diversity were enhanced. The developed ResAU-Net, a novel image segmentation model integrating residual and attention modules, significantly improved segmentation accuracy and performance. Through comparative analysis of commonly used loss functions, the Lovász loss function emerged as the most suitable one for cervical cancer image segmentation tasks. The comprehensive evaluation approach including quantitative metrics and evaluation by experienced radiation oncologists ensured the clinical relevance and robustness of the proposed model. The proposed model achieved promising results with Dice similarity coefficient (DSC) value and 95th percentile Hausdorff distance (95HD) value.

Overall, this proposed model demonstrated significant potential for improving cervical cancer segmentation accuracy and efficiency in clinical practice as evidenced by the high acceptability rate of the automatic delineation results evaluated by radiation oncologists. This study created a cervical cancer dataset for model training and validation and applied appropriate image augmentation techniques to enhance the variability and diversity of the dataset. Through this study, a novel image segmentation model, ResAU-Net, was proposed based on U-Net with optimization. By comparing the performance of four commonly used loss functions in cervical cancer image segmentation tasks, the Lovász loss function was selected as the most suitable one to measure the continuity of predicted boundaries in image segmentation tasks and accurately assess the smoothness of segmentation boundaries. Minimizing misclassifications of segmentation boundaries optimized the model and better addressed class imbalance issues, making it suitable for high-precision image segmentation tasks. DSC and 95HD were employed as quantitative metrics. Additional evaluations of 15 patients' segmentation results by two experienced radiation oncologists enhanced the clinical relevance of the study findings.

Acknowledgements

This study was supported by part of the National Natural Science Foundation of China (Grant No. 62001313) and the Key Project of Liaoning Provincial Department of Science and Technology (Grant No. 2021JH2/10300134, 2022JH1/10500004), as well as the Natural Science Foundation of Liaoning Province (Grant No. 2023-MS-147).

References

1. Ali CI, Makata NE, Ezenduka PO. 2016. Cervical cancer: a health limiting condition. Gynecol Obstet. 6(378):2161-0932.
2. Cibula D, Pötter R, Planchamp F, Avall-Lundqvist E, Fischerova D, Haie-Meder C, *et al*. 2018. The European Society of Gynaecological Oncology/European Society for Radiotherapy and Oncology/European Society of Pathology guidelines for the management of patients with cervical cancer. Virchows Archiv. 472:919-936.
3. Marth C, Landoni F, Mahner S, McCormack M, Gonzalez-Martin A, Colombo N. 2017. Cervical cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 28:72-83.
4. Sleeman IV, Srinivasan S, Ghosh P, Palta J, Kapoor R. 2022. Big data applications in radiation oncology: challenges and opportunities. Artificial Intelligence in Cancer Diagnosis and Prognosis, Volume 3: Brain and prostate cancer. 8:1-28.
5. Vinod SK, Min M, Jameson MG, Holloway LC. 2016. A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology. J Med Imaging Radiat Oncol. 60(3):393-406.
6. Chang AT, Tan LT, Duke S, Ng WT. 2017. Challenges for quality assurance of target volume delineation in clinical trials. Front Oncol. 7:221.
7. Van der Veen J, Gulyban A, Nuyts S. 2019. Interobserver variability in delineation of target volumes in head and neck cancer. Radiother Oncol. 137:9-15.
8. Vrtovec T, Močnik D, Strojan P, Pernuš F, Ibragimov B. 2020. Auto-segmentation of organs at risk for head and neck radiotherapy planning: from atlas-based to deep learning methods. Med Phys. 47(9):929-950.
9. Cardenas CE, Yang J, Anderson BM, Court LE, Brock KB. 2019. Advances in auto-segmentation. Semin Radiat Oncol. 29(3):185-197.
10. Abdou MA. 2022. Literature review: Efficient deep neural networks techniques for medical image analysis. Neural Comput Appl. 34(8):5791-5812.
11. Islam MR, Matin A. 2020. Detection of COVID 19 from CT image by the novel LeNet-5 CNN architecture. In 23rd International Conference on Computer and Information Technology (ICCIT). 2020:1-5.
12. Alom MZ, Taha TM, Yakopcic C, Westberg S, Hasan M, van Esesn BC, *et al*. 2018. The history began from alexnet: A comprehensive survey on deep learning approaches. arXiv preprint arXiv. 1803:01164.
13. 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. 2016:770-778.
14. Lin L, Dou Q, Jin YM, Zhou GQ, Tang YQ, Chen WL, *et al*. 2019. Deep learning for automated contouring of primary tumor volumes by MRI for nasopharyngeal carcinoma. Radiology. 291(3):677-686.
15. Men K, Chen X, Zhang Y, Zhang T, Dai J, Yi J, *et al*. 2017. Deep deconvolutional neural network for target segmentation of nasopharyngeal cancer in planning computed tomography images. Front Oncol. 7:315.
16. Trebeschi S, van Griethuysen JJM, Lambregts DMJ, Lahaye MJ, Parmar C, Bakers FCH, *et al*. 2017. Deep learning for fully-automated localization and segmentation of rectal cancer on multiparametric MR. Sci Rep. 7(1):5301.

17. Chen S, Yang H, Fu J, Mei W, Ren S, Liu Y, *et al*. 2019. U-net plus: Deep semantic segmentation for esophagus and esophageal cancer in computed tomography images. IEEE Access. 7:82867-82877.
18. Liu Z, Liu X, Xiao B, Wang S, Miao Z, Sun Y, *et al*. 2020. Segmentation of organs-at-risk in cervical cancer CT images with a convolutional neural network. Phys Med. 69:184-191.
19. Li C, Tan Y, Chen W, Luo X, He Y, Gao Y, *et al*. 2020. ANU-Net: Attention-based nested U-Net to exploit full resolution features for medical image segmentation. Comput Graph. 90:11-20.
20. Liu L, Cheng J, Quan Q, Wu FX, Wang YP, Wang J. 2020. A survey on U-shaped networks in medical image segmentations. Neurocomputing. 409:244-258.
21. Krithika alias AnbuDevi M, Suganthi K. 2022. Review of Semantic Segmentation of Medical Images Using Modified Architectures of UNET. Diagnostics (Basel). 12(12):3064.
22. Kazemifar S, Balagopal A, Nguyen D, McGuire S, Hannan R, Jiang S, *et al*. 2018. Segmentation of the prostate and organs at risk in male pelvic CT images using deep learning. Biomed Phys Eng Express. 4(5):055003.
23. Balagopal A, Kazemifar S, Nguyen D, Lin MH, Hannan R, Owrangi A, *et al*. 2018. Fully automated organ segmentation in male pelvic CT images. Phys Med Biol. 63(24):245015.
24. Wang C, Connolly B, de Oliveira Lopes PF, Frangi AF, Smedby O. 2018. Pelvis segmentation using multi-pass U-Net and iterative shape estimation. Computational Methods and Clinical Applications in Musculoskeletal Imaging: 6th International Workshop. 49-57.
25. Cai Y, Yuan J. 2022. A review of U-Net network medical image segmentation applications. International Conference on Artificial Intelligence and Pattern Recognition. 457-461.
26. Zhang Z, Wu C, Coleman S, Kerr D. 2020. DENSE-INception U-net for medical image segmentation. Comput Methods Programs Biomed. 192:105395.
27. Wen J, Liu P. 2023. A classification method for English texts based on hybrid recurrent neural network and graph construction in social recommendation systems. IEEE Syst J. 17(4):5272-5279.
28. Mnih V, Heess N, Graves A. 2014. Recurrent models of visual attention. Advances in neural information processing systems. 27:2204-2212.
29. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, *et al*. 2017. Attention is all you need. Advances in neural information processing systems. 30:6000-6010.
30. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich MP, Misawa K, *et al*. 2018. Attention u-net: Learning where to look for the pancreas. ArXiv preprint arXiv. 1804:03999.
31. Schlemper J, Oktay O, Schaap M, Heinrich M, Kainz B, Glocker B, *et al*. 2019. Attention gated networks: Learning to leverage salient regions in medical images. Med Image Anal. 53:197-207.
32. Huang C, Yin C. 2022. A coronary artery CTA segmentation approach based on deep learning. J Xray Sci Technol. 30(2):245-259.
33. Berman M, Triki AR, Blaschko MB. 2018. The lovász-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition. 4413-4421.