BMC Medical Imaging

RESEARCH

Open Access

Enhanced pediatric thyroid ultrasound image segmentation using DC-Contrast U-Net

Bo Peng[1,2*], Wu Lin[2], Wenjun Zhou[1,2], Yan Bai[3], Anguo Luo[1], Shenghua Xie[1] and Lixue Yin[1*]

Abstract

Early screening methods for the thyroid gland include palpation and imaging. Although palpation is relatively simple, its effectiveness in detecting early clinical signs of the thyroid gland may be limited, especially in children, due to the shorter thyroid growth time. Therefore, this constitutes a crucial foundational work. However, accurately determining the location and size of the thyroid gland in children is a challenging task. Accuracy depends on the experience of the ultrasound operator in current clinical practice, leading to subjective results. Even among experts, there is poor agreement on thyroid identification. In addition, the effective use of ultrasound machines also relies on the experience of the ultrasound operator in current clinical practice. In order to extract sufficient texture information from pediatric thyroid ultrasound images while reducing the computational complexity and number of parameters, this paper designs a novel U-Net-based network called DC-Contrast U-Net, which aims to achieve better segmentation performance with lower complexity in medical image segmentation. The results show that compared with other U-Net-related segmentation models, the proposed DC-Contrast U-Net model achieves higher segmentation accuracy while improving the inference speed, making it a promising candidate for deployment in medical edge devices in clinical applications in the future.

Keywords Pediatrics, Thyroid, Ultrasound images, Segmentation, DC-Contrast U-Net

Introduction

The thyroid is a crucial gland in the human body, located in the front of the neck in the shape of an "H" surrounding the trachea, and composed of left and right lobes, the isthmus, and the pyramidal lobe [1]. Thyroid follicles constitute the primary substance of the thyroid, and the epithelial cells within the follicles can secrete thyroid hormones. These hormones have diverse effects on the human body, making the normal functioning of the thyroid an essential foundation for development and health [1]. Particularly in the case of pediatric thyroid issues, the volume of the thyroid gland is a key factor in analyzing the secretion of thyroid hormones in children. Therefore, the location and size of the thyroid are also important features in diagnosing thyroid diseases in clinical settings. Thus, delineating the boundaries of the thyroid gland is necessary for evaluating diseases [2]. It is important to note that children (up to 14 years old) are not miniature versions of adults. Children exhibit significant differences from adult patients in biological characteristics, clinical manifestations, diagnosis, and prevention. Studies have shown that treatment measures for adult thyroid disease patients may not be entirely applicable to children. For example, radioactive iodine therapy, a common

*Correspondence:
Bo Peng
bpeng1@vip.163.com
Lixue Yin
yinlixue_cardiac@163.com
[1] Ultrasound in Cardiac Electrophysiology and Biomechanics Key Laboratory of Sichuan Province, Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, Chengdu 611731, China
[2] School of Computer Science and Software Engineering, Southwest Petroleum University, Chengdu 610500, China
[3] Sichuan Provincial Woman's and Children's Hospital / The Affiliated Women's and Children's Hospital of Chengdu Medical College, Chengdu 610000, China

treatment for hyperthyroidism in adults, is often not recommended for children due to its potential long-term effects on growth and development. Instead, antithyroid medications are preferred in pediatric cases to avoid the risks associated with radiation exposure [3]. Similarly, thyroidectomy, which involves the surgical removal of the thyroid gland and is frequently used in adults with thyroid cancer or nodules, presents significant challenges for children. The risks of complications and the impact on growth and hormone regulation make this approach less favorable for pediatric patients. Research suggests that less invasive options and careful monitoring are generally recommended for children, with surgery considered only in specific cases where other treatments have failed or the condition is severe [4].

Improper treatment in children can result in long-term damage to their growth. Adult thyroid development is generally mature, and the assessment of the severity of the disease often revolves around the benign or malignant nature of nodules. For the detection of pediatric thyroid diseases in the context of this study, it is noteworthy that nodules in children have a shorter growth period, and cases of tumors in children are relatively rare. Moreover, the incidence of nodules in children is much lower than in adults. Most thyroid problems in children include developmental abnormalities, ectopia, or absence of the thyroid. Therefore, the size and location of the thyroid are crucial features in diagnosing thyroid diseases in children. The early screening methods for the thyroid are divided into palpation and imaging examinations. Palpation is relatively simple but may have limited effectiveness in detecting early clinical symptoms of the thyroid, especially when the thyroid's growth time is short in children. In recent years, with the vigorous development of imaging examinations, they have become widely popular. Unlike palpation, imaging examinations can directly visualize internal organ tissues, providing a more intuitive display and enabling early diagnosis and treatment of thyroid diseases [5]. Common imaging examinations include ultrasound diagnosis, computer tomography imaging (CT), magnetic resonance imaging (MRI), and nuclear imaging. Among them, ultrasound medical imaging, with its advantages of safety, non-invasiveness, real-time imaging, multi-sectional views, no radiation, and cost-effectiveness, plays a significant role in preoperative diagnosis and planning, treatment, and postoperative monitoring. Currently, ultrasound (US) is the primary imaging technology for diagnosing thyroid diseases in clinical practice. However, accurately identifying the position and size of the thyroid in children is a challenging task. The accuracy depends on the experience of the ultrasound operator in current clinical practice, leading to subjective results. Even among experts, the consistency in thyroid identification is poor. Additionally, the use of ultrasound machines also relies on the experience of the ultrasound operator in current clinical practice. In many primary hospitals, blood tests are used for the diagnosis of pediatric thyroid diseases, which is an invasive examination. This study aims to use ultrasound detection as a non-invasive method for preliminary screening of pediatric thyroid diseases to assist in subsequent treatment. Moreover, the imbalance in medical resources can result in significant differences in diagnostic accuracy between different regions and hospitals. Therefore, the use of computer-aided diagnosis (CAD) systems can provide an objective, quantitative description of the problem, eliminate the subjectivity of doctors, and offer useful reference information and data [6].

Figure 1 displays clinical ultrasound images of pediatric thyroid collected from the ultrasound department of a certain hospital. The first column shows transverse section (cross-section) images of the thyroid, providing a clear and complete observation of the left and right thyroid lobes, isthmus, trachea, and esophagus, among other positional relationships. The second column shows longitudinal scan (sagittal section) images, offering a complete view of a single lobe of the thyroid. The illustrated clinical pediatric thyroid data contain various irrelevant information, such as patient personal details, image scan dates, scan location information, etc. The central part of the image represents the outline of the thyroid gland, surrounded by similar organs and tissues like the esophagus, trachea, nerve tissues, muscles, etc. There are no distinctly clear boundaries between these organ tissues. In such a complex situation, traditional segmentation methods struggle to achieve the segmentation task for pediatric thyroid. Additionally, the current study builds on existing models and methodologies in medical image segmentation. For instance, recent works such as the 2MGAS-Net, a multi-level multi-scale gated attentional squeezed network for polyp segmentation, provide valuable insights into advanced segmentation techniques [7]. Similarly, convolutional neural-adaptive networks for melanoma recognition highlight the use of adaptive techniques in medical imaging [8]. These approaches offer valuable context and comparison for evaluating the performance of the proposed DC-Contrast U-Net model in this study.

Related works

Thyroid segmentation is a hotspot in the field of ultrasound image segmentation. Before deep learning became the mainstream segmentation approach, image segmentation typically relied on traditional image processing techniques, which can be categorized into several types. The first type is threshold-based methods, which

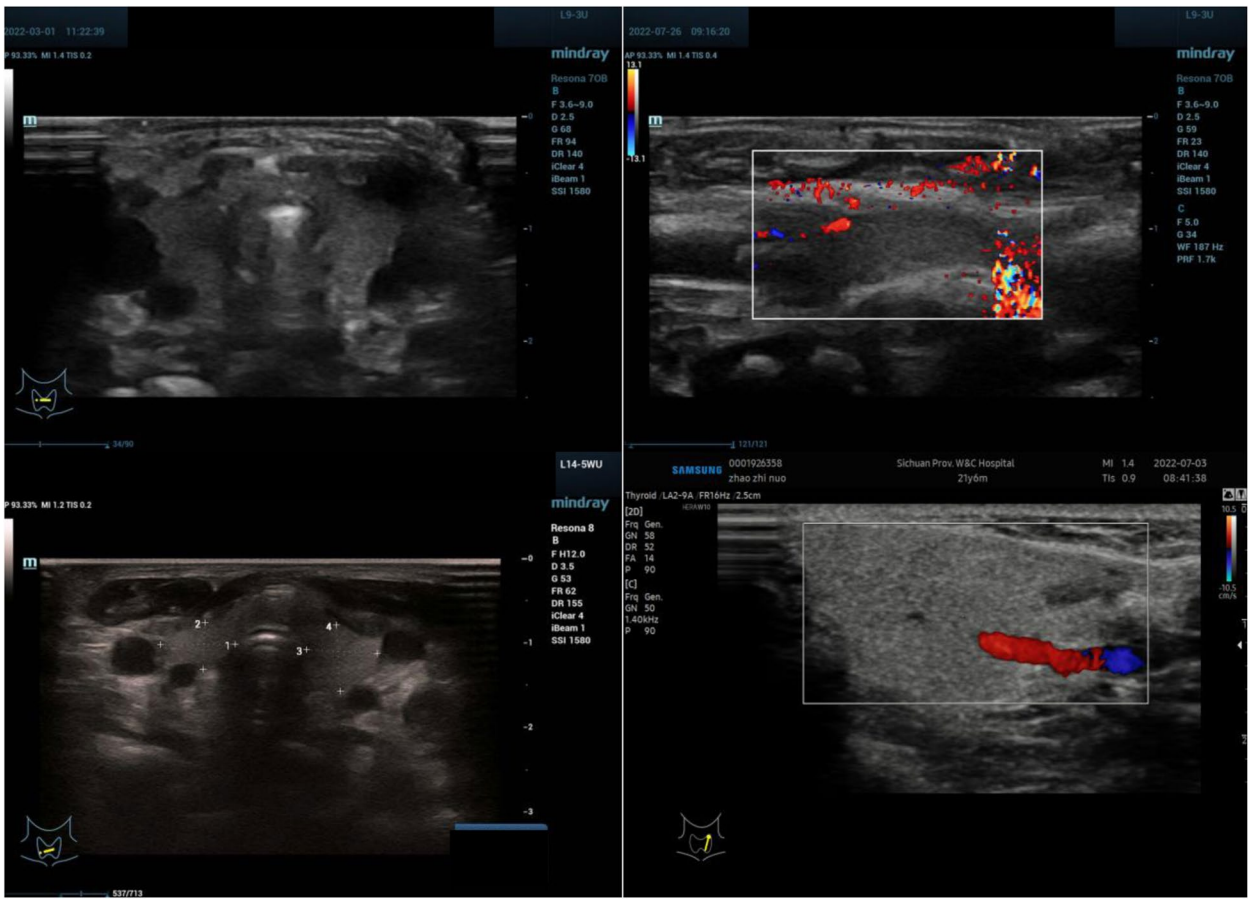


Fig. 1 The pediatric thyroid ultrasound images collected by the hospital

compare pixel values with predefined thresholds to achieve binary segmentation. The second type is edge detection-based methods, which achieve segmentation by detecting edges in the image. The third type is region-based methods, which divide pixels with high brightness similarity into a region for segmentation. The last type is contour-based methods, which achieve image segmentation by minimizing the driving energy function. Traditional methods are not entirely automated and are mostly composed of various algorithms and auxiliary processing steps. In contrast, deep learning, especially Convolutional Neural Networks (CNNs), addresses end-to-end problems, applying gradient-based learning to the entire system, from general features like edges and spots to advanced features like shapes.

Medical image segmentation algorithms based on traditional imaging genomics

Traditional image segmentation methods refer to the use of conventional computer vision techniques to achieve image segmentation. The following are introductions to several common traditional image segmentation methods:

(1)Threshold-Based Segmentation

Threshold segmentation is a simple and widely used image segmentation method that partitions grayscale images based on different levels, labeling pixels within the same category as the same object or background. Prewitt et al. [9] proposed the global single-threshold segmentation method using histogram bimodal analysis. In 1979, Nobuyuki Otsu [10] introduced the OTSU algorithm, which determines the optimal threshold by maximizing the inter-class variance. Kapure et al. [11] later proposed the maximum entropy threshold method, which seeks the optimal threshold to maximize the sum of entropy for object and background parts. However, threshold-based methods struggle with complex backgrounds where target and background grayscale ranges overlap. Alternative methods such as edge-based segmentation, region-growing algorithms, and graph-based segmentation are often more effective in these scenarios.

(2)Edge-based Segmentation

Edge-based segmentation achieves image segmentation by detecting edges of different regions. Edge detection was first introduced by Julez [12] in 1959. In 2010, Li [13] proposed optimization criteria based on the OTSU algorithm to segment small target defects. In 2015, Du [14] introduced the DRLSE algorithm for segmenting thyroid nodules in ultrasound images, addressing noise and weak echoes but performing poorly with weak edges and being sensitive to manual contour initialization. In 2018, Kuchekar [15] used edge detection to classify rice grains based on features like shape and size. In 2020, Qu [16] applied homomorphic filtering and an improved Canny operator for edge detection in thyroid CT images, enhancing noise removal and detail accuracy. However, edge detection methods are more suitable for simple images, as they struggle with complex images, necessitating more advanced segmentation techniques.

(3)Region-based Segmentatio

Region-based image segmentation divides an image into distinct regions based on pixel brightness similarity. Key methods include region growing and region splitting and merging. In 1994, Adams et al. [17] introduced a seed region method using adjacent pixels, successfully segmenting diverse medical images. In 2012, Zhao et al. [18] proposed a thyroid nodule segmentation technique using Normalized Cut (NC), incorporating multiple filtering techniques to address noise suppression and detail preservation in ultrasound images, although it lacked comprehensive quantitative analysis. In 2016, Alrubaidi et al. [9] developed a 2D thyroid nodule segmentation method for ultrasound images, using variance reduction statistics and radial line edge points, improved with B-spline technology for better accuracy. Despite improvements, the method is time-consuming, resource-intensive, and sensitive to noise, which can lead to segmentation errors. Ongoing research seeks to address these challenges and improve region-based segmentation algorithms for medical imaging.

Medical image segmentation algorithms based on deep learning

With the rise of neural network models and the development of deep learning, many scholars have begun to adopt image segmentation techniques based on deep learning. These techniques have shown promising results in computer-aided diagnosis research, providing valuable assistance to medical professionals [10]. In this chapter, we will primarily discuss the current research status of deep learning image segmentation algorithms from two perspectives: research on networks based on the U-Net model and research on networks based on non-U-Net models.

(1)U-Net Related Models

State-of-the-art image segmentation models leverage unique information at different scales, such as Fully Convolutional Network (FCN), U-Net, SegNet, PSPNet [19], and various DeepLab versions [20]. U-Net, widely used in medical image segmentation, combines low-level detailed feature maps from the encoder with high-level semantic feature maps from the decoder using residual connections. Numerous variants of U-Net have been proposed. In 2020, Huang proposed UNet 3+, which uses comprehensive residual connections to aggregate feature maps at all scales, yielding good results with fewer parameters than U-Net. He [21] introduced a self-attention mechanism in a deep attention U-Net model for thyroid CT images, though its performance on ultrasound images remains untested. In 2021, Lou proposed DC-UNet, which uses dual-channel CNN blocks for higher performance with fewer parameters. Hu [22] incorporated a multi-scale dilated convolution pyramid and a spatial boundary attention module into U-Net to address morphological and size variations in thyroid images, but this model is parameter-heavy and time-consuming. Zhang [23] introduced a cascaded U-Net framework for thyroid nodule segmentation, using one U-Net for rough localization and another for fine segmentation. However, it struggles with nodules of varying sizes and positions, highlighting the challenge of achieving high precision in thyroid nodule segmentation due to their diverse and complex characteristics.

(2)Other Models

Deep Convolutional Neural Networks (DCNNs) have continuously evolved from traditional neural networks. In 1998, “LeNet” [24] formed a complete deep convolutional neural network, but its development was limited by computational capacity and data scale. Recent advancements in computer hardware and the rise of big data have significantly improved DCNN performance. In 2019, Gu [25] proposed CE-Net, which accelerates network convergence and avoids gradient disappearance using feature encoders, decoders, and context extraction modules. In 2021, Gong [26] introduced a thyroid region pre-guided feature enhancement model for thyroid nodule segmentation, which aids radiologists but focuses too much on global information, neglecting local details. Challenges remain in segmenting children’s thyroids due to unclear boundaries, limited data, high annotation costs, and variable morphological scales.

To address the unresolved issues in the aforementioned methods, we propose a new method: DC-Contrast U-Net. This method aims to improve the performance of existing models in handling medical image segmentation tasks, particularly in terms of key metrics such as accuracy and inference speed. Since our method is based on the U-Net model, we have compared it with several

classic and advanced U-Net models, including U-Net, UNet++, and Attention U-Net. These models were selected for their widespread application, representativeness in the field of medical image segmentation, and their performance in handling similar tasks. The experimental results will highlight the differences in key metrics such as accuracy and inference speed among these models, offering an in-depth understanding of the improvements and advantages of DC-Contrast U-Net relative to existing technologies.

Methodology

This section introduces a segmentation model called Deformable Convolutional Contrast U-Net (DC-Contrast U-Net) based on the features of pediatric thyroid ultrasound images. Built upon the U-Net architecture, this model incorporates Contrast Blocks designed to accurately extract texture information from pediatric thyroid glands. Additionally, to enhance the network's ability to learn texture information from each channel and better adapt to the image shape, Compression and Excitation Blocks, along with Deformable Convolution Layers, are introduced. The Compression and Excitation Blocks serve as auxiliary skip connections from the encoding to decoding process, while the Deformable Convolution Layers aim to better capture the shape information of irregular organs in the images.

Due to the continuous improvement in computer image processing capabilities, deep learning has been widely applied in various fields. The medical field is also exploring the application of deep learning methods to medical imaging. In this paper, taking the task of pediatric thyroid medical image segmentation as an example, we propose improvements to the semantic segmentation network model U-Net. The goal is to extract the thyroid target region more accurately from ultrasound images, especially in cases where pediatric ultrasound data is scarce. The specific improvements are as follows:

(1.) In this paper, we present a state-of-the-art Contrast Block design, showcasing its expertise in extracting abundant texture details from ultrasound images. An inherent feature of this design is its noteworthy ability to substantially reduce the number of parameters and computational intricacies associated with extracting information within the network from ultrasound images, thus contributing to a more streamlined and efficient image processing framework.

(2.) A new network named Contrast U-Net is proposed in the study. This network incorporates the Contrast Block into the encoder-decoder framework and introduces compression and excitation blocks (Squeeze-and-Excitation Block) in the skip connections [27]. These additions adaptively recalibrate the feature response intensity between channels. The proposed network shows effective improvements in segmentation accuracy, with a substantial reduction in parameters and computational workload compared to the U-Net network.

(3.) Ordinary convolutions in the base U-Net network are replaced with deformable convolutions [28] (abbreviated as DC) in this study, addressing spatial deformation issues in object space by introducing additional offsets to increase spatial sampling positions within the module.

Basic structure of the DC-Contrast U-Net network

In Fig. 2, (a) The overall network framework proposed in this paper, (b) the Contrast Block, and (c) the Compression and Excitation Block. The model consists of four main modules: encoder, decoder, Contrast Block, and Compression and Excitation Block. The encoder and decoder modules are like the encoder and decoder modules in the classical U-Net network, with ordinary convolutions replaced by deformable convolutions. DC-Contrast U-Net uses deformable convolutions and down sampling to extract multi-scale features. There are two different convolution blocks during down sampling, with one block identical to U-Net, stacking two 3×3 convolutions with an output channel number twice the input channel number. The second layer of deformable convolution in this block has the same properties as the preceding layer, consisting of a Contrast Block and a convolution kernel with a size of 3×3, and the output channel number is the same as the input channel number. The proposed Contrast Block and deformable convolution results are concatenated to extract texture information and extend the channel number. To adaptively recalibrate the feature response intensity between channels, Compression and Excitation Blocks are used at the skip connections between the encoder and decoder, replacing traditional skip connections.

Contrast Block

A novel operator named Contrast Value Operator (CVO) is introduced in the Contrast Block. The Contrast Value Operator is an effective operator designed based on Pascal's Triangle. Due to the property of Pascal's Triangle, whose numbers can be used for computing differences and gradients, it enables smooth filtering and edge detection. As confirmed in previous research [29], the Contrast Value Operator is particularly sensitive to fine textures in images, making it well-suited for segmentation tasks. The Contrast Value Operator consists of two components: a difference component composed of 1 and -1, and a contrast component composed of 1 and 2. As shown in Fig. 3, convolution of the difference and contrast components in different directions yields Contrast

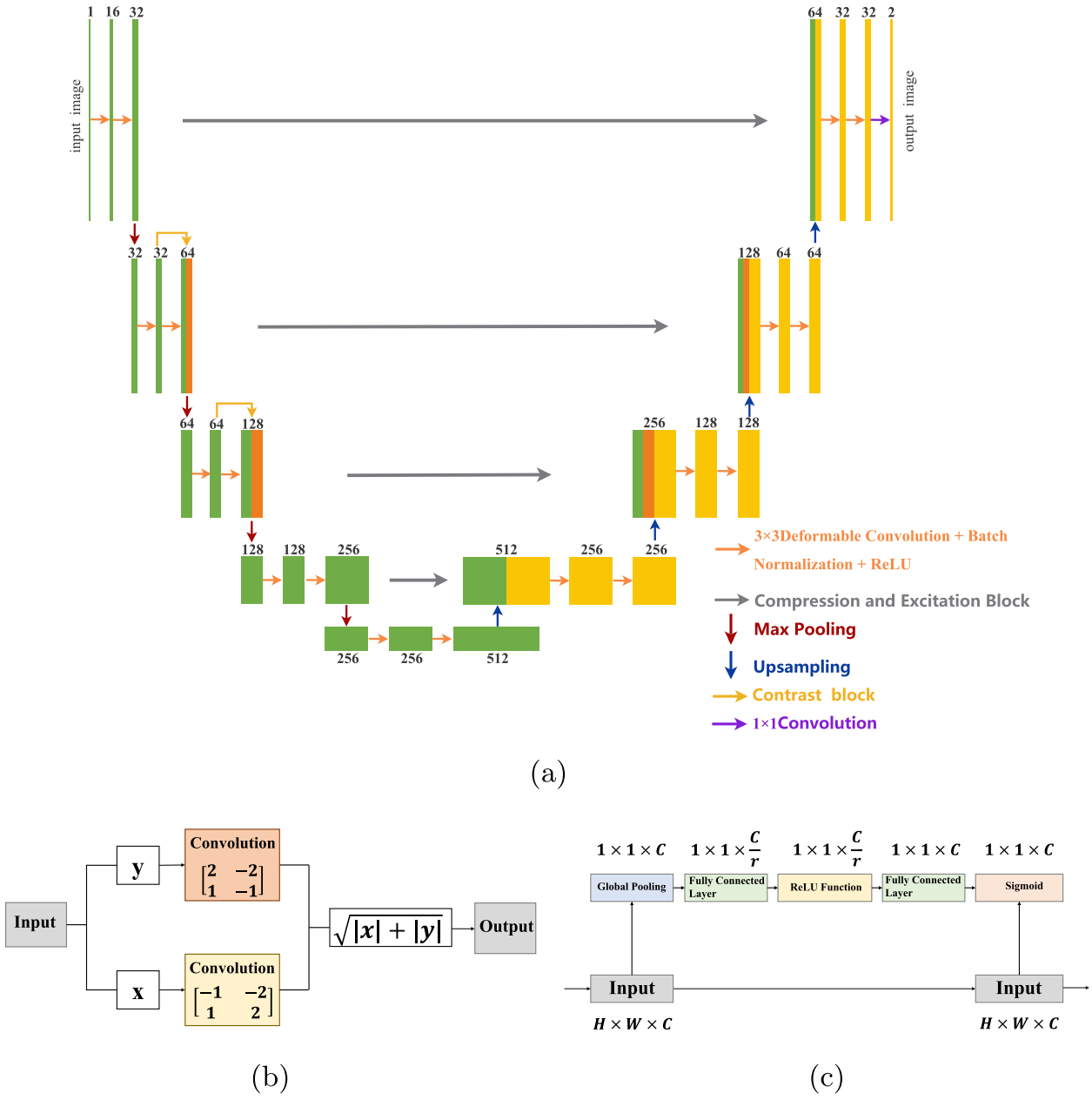


Fig. 2 DC-Contrast U-Net Architecture

Value Operators in different orientations. With the presence of the difference component, the Contrast Value Operator exhibits high sensitivity to fine textures, while the contrast component enhances image contrast.

The proposed Contrast Block is obtained by convolving the Contrast Value Operator in different directions and calculating gradients. The specific structure of the Contrast Block is illustrated in Fig. 3, where (a) shows the process of generating contrast value operator components using Pascal's Triangle; (b) displays the convolution kernel of the contrast value operator in the x-direction; and (c) shows the convolution kernel of the contrast value operator in the y-direction. The input image undergoes convolution with the contrast value operator from different directions, followed by the calculation of approximate gradients.

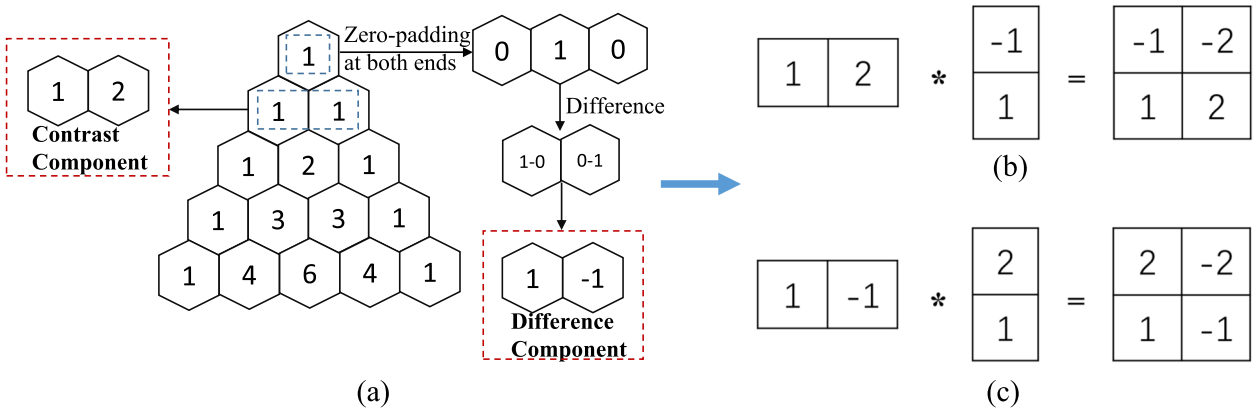


Fig. 3 Contrast Value Operator Generation Process Schematic

Deformable convolution

Traditional CNN modules in visual recognition have certain fixed geometric structure limitations. For example, convolutional units' sample and pool input feature maps at fixed positions, reduce spatial resolution and regions of interest at a fixed ratio, pool layers segment an area of interest into fixed spatial units, and lack internal mechanisms for handling geometric transformations. These limitations result in the same receptive field size for all activated units within the same layer. In practical visual recognition scenarios requiring precise localization, the same object at different positions should correspond to different scales or deformations. Therefore, the adaptive adjustment of scale and receptive field size holds great potential. To address these issues, deformable convolutions have been introduced, allowing the convolutional kernel to adaptively adjust its shape. See Fig. 4.

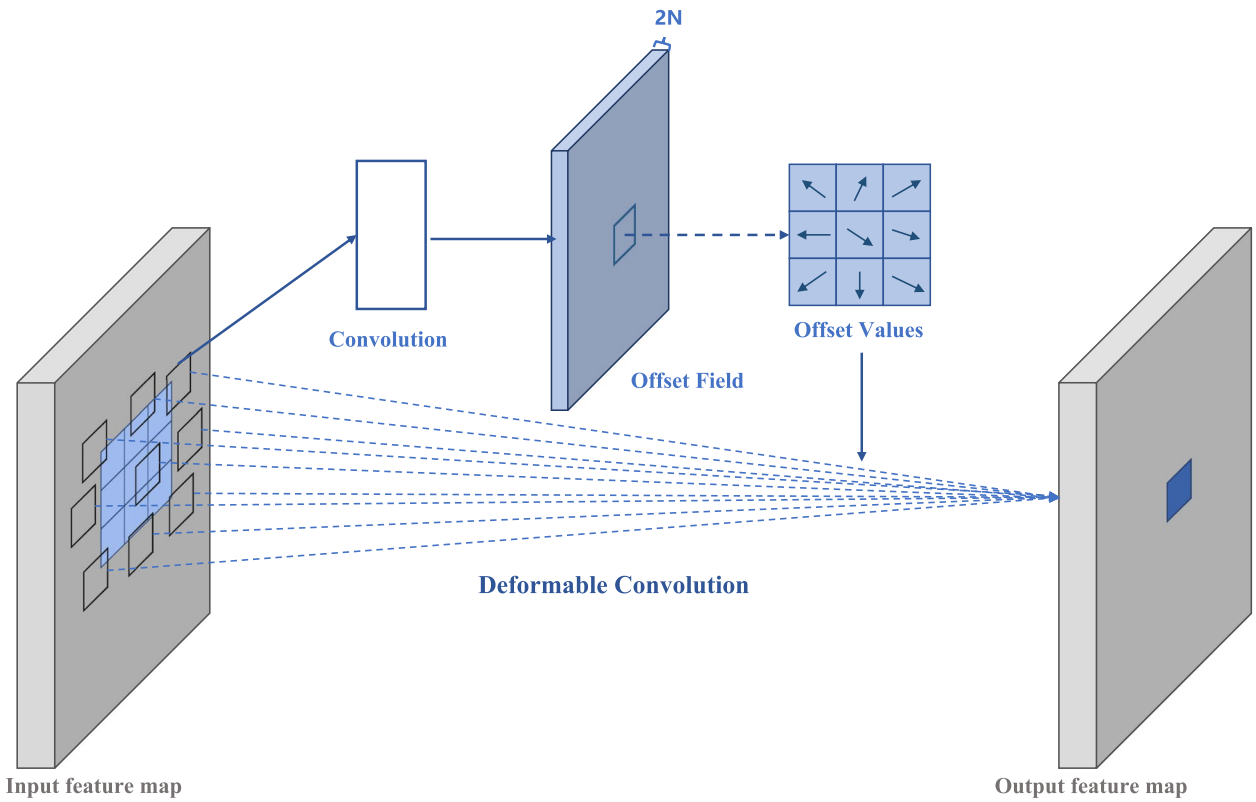


Fig. 4 3×3Illustration of Deformable Convolution

For a given input feature map, assuming the original convolution kernel size is 3×3, a new deformable convolution layer is defined to generate offset values. The kernel size of this convolution layer remains 3×3, and the output feature map has the same dimensions as the input map but with a channel number of 2N (representing offset values in the x and y directions). Deformable convolution can be viewed as interpolating the generated offset values in the upper part and then performing a regular convolution operation.

Traditional convolution can be defined by the formula (1):

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n) \tag{1}$$

Where p_0 is each position in the output feature map y, and p_n enumerates positions in the grid R. In deformable convolution, the regular grid R is enlarged by offsets $\{\delta_{p_n} = n = 1, \ldots, N\}$, where $N = |R|$. Formula (1) evolves into formula (2):

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n + \delta p_n) \tag{2}$$

Since in deformable convolution, offsets cause the convolution kernel positions to not correspond to integer pixel points on the feature map, interpolation methods are needed to obtain the pixel values after offset. Typically, bilinear interpolation is used for interpolation calculations, expressed by formula (3):

$$\begin{aligned} x(p) &= \sum_q G(q,p) \cdot x(q) \\ &= \sum_q g(q_x, p_x) \cdot g(q_y, p_y) \cdot x(q) \\ &= \sum_q \max(0, 1 - |q_x - p_x|) \cdot \max(0, 1 - |q_y - p_y|) \cdot x(q) \end{aligned} \tag{3}$$

Here, p represents any position ($p = p_0 + p_n + \delta p_n$), and q is chosen from all positions in the feature map x. $G(\cdot, \cdot)$ is the kernel function for bilinear interpolation, and since it is two-dimensional, $G(q,p) = g(q_x, p_x) \cdot g(q_y, p_y)$. The overall meaning of the formula is to set the pixel value of the interpolation point as the weighted sum of its neighboring four-pixel points. These four-pixel points are the closest and exist on the feature map. The weight of each pixel point is calculated based on the Euclidean distance between its coordinates and those of the interpolation point. The term $\max(0, 1 - \ldots)$ in the last line of formula (3) limits the distance between neighboring points and the interpolation point to be less than 1 pixel.

Compression and Excitation Block

The SE [27] module (Squeeze-and-Excitation Module) is a module designed to enhance the performance of convolutional neural networks. It was proposed by Hu et al. in 2018 and can be incorporated into existing convolutional neural networks. To simulate inter-channel dependencies, this paper introduces the Compression and Excitation Block (Squeeze-and-Excitation Block) to adaptively recalibrate the strength of feature responses between channels, explicitly establishing mutual dependencies between channels. The Compression and Excitation Block is illustrated in Fig. 3-1(c). As the name suggests, the Compression and Excitation Block is divided into two parts: Squeeze and Excitation.

The main idea of the SE module is to adjust the channel weights of the feature map by adaptively learning inter-channel correlations. The module consists of two steps: the first step involves compressing the features of each channel to a scalar through global average pooling, referred to as the "squeeze" operation; the second step entails using a fully connected layer to weight each scalar of each channel, known as the "excitation" operation. This process enables the network to pay more attention to important channels, thereby enhancing the expressiveness and discriminability of features.

In the SE module, the squeeze and excitation operations can be expressed as formulas (4) and (5):

$$z = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_{i,j} \tag{4}$$

$$s = \sigma(W_2 \delta(W_1 z)) \tag{5}$$

In this module, $x_{i,j}$ represents a pixel in the feature map, where H and W represent the height and width of the feature map, respectively. The compressed feature for each channel is denoted by z, and the weights of the fully connected layer are W_1 and W_2. The activation functions and sigmoid functions are represented by δ and σ, respectively.

The compression part involves a global average pooling operation to obtain a globally compressed feature vector, effectively compressing the original feature map from $H \times W \times C$ (height × width × number of channels) to $1 \times 1 \times C$, which is equivalent to compressing $H \times W$ into one dimension. This provides a globally panoramic view of the original $H \times W$ feature map, thereby expanding the receptive field. The excitation part involves two fully connected layers concatenated to obtain the weight values for each channel. The weighted input is then used as the output, resulting in a feature map of size $1 \times 1 \times C$ after the compression operation. A fully connected layer is added, predicting the importance of each channel. After determining the importance of different channels, the excitation operation is applied to the corresponding

channels in the original feature map. In the excitation operation, the parameter $1/r$ is a scaling parameter aimed at reducing the number of channels, thereby reducing computational complexity.

Loss function

In pediatric thyroid ultrasound images, there is a significant proportion of background regions, with only the central part being the region of interest for the pediatric thyroid. This leads to the issue of class imbalance in ultrasound images. To optimize the proposed network in this paper, Focal Loss (FL) [30] is used as the network's loss function to measure the relationship between predicted values and ground truth. Focal Loss is a dynamically scaled cross-entropy loss function that reduces the weights of easily distinguishable samples during training by dynamically adjusting the scaling factor, focusing on challenging samples. This approach allows the model to pay more attention to difficult-to-distinguish samples, thereby improving segmentation accuracy. Difficult-to-distinguish samples may be positive or negative samples but are helpful for training the network, expressed by the formula (6):

$$FL = (1 - p_t)^{\gamma} \log(p_t) \tag{6}$$

Among them, p_t reflects the degree of proximity between predicted values and actual values, and γ is a modifiable parameter.

In the field of image segmentation, imbalanced class distribution is often encountered, such as in tasks like defect detection in industrial products, road extraction, and segmentation of lesion areas. In scenarios with imbalanced class distribution, where the pixel counts for different classes vary significantly, models often perform poorly in predicting the minority class. Therefore, it is necessary to employ methods to address class imbalance issues and improve the segmentation accuracy of the model. One approach to handling such imbalanced class distribution is to use Lovasz Loss.

Lovasz Loss (LL) is a convex Lovasz extension loss function based on submodular losses, optimized for the mean intersection over union (mIoU) loss of neural networks. When using Lovasz Loss, different forms of loss functions can be chosen, including Lovasz Hinge Loss and Lovasz Softmax Loss. Lovasz Hinge Loss is suitable for binary classification problems, while Lovasz Softmax Loss is suitable for multi-class problems. As mIoU aligns more with the intuitive sense of image comparison, Lovasz Loss is selected to assist Focal Loss in fine-tuning the results. This can be expressed as formula (7):

$$LL = \overline{\Delta}_{J_1}(m) \tag{7}$$

Where $\overline{\Delta}_{J_1}$ represents the Jaccard Loss, and m is the optimization algorithm. In summary, the total loss function can be expressed as formula (8):

$$Loss = FL + \lambda LL \tag{8}$$

Experimental

Dataset creation and preparation

(1) Dataset Preprocessing

The pediatric thyroid ultrasound image dataset used in this study is sourced from the ultrasound department of the Sichuan Provincial Maternal and Child Health Hospital, totaling 380 original images. Due to incomplete or unclear contours of the pediatric thyroid in some data, a selection and cleaning process was applied to the raw data obtained from the hospital. A total of 310 usable pediatric thyroid ultrasound images remained after this process. Clinical data was collected from the hospital, and after consultation with doctors, professional ultrasound doctors used the LabelMe [30] annotation tool to annotate the thyroid contours, generating JSON files containing coordinate information for the thyroid contour positions.

Given the complexity of the thyroid itself and variations in human development, there may be some annotation errors, omissions, or duplications in the dataset during actual application. To ensure the quality of the training data, manual selection of well-annotated image data is necessary. Additionally, to protect patient privacy, sensitive information needs to be removed and renamed before using the image data.

The following preprocessing steps were applied to the acquired clinical data in this study, The detailed steps are shown in Fig. 5: Since the segmentation network in this chapter requires pediatric thyroid ultrasound images and corresponding label images, the annotation information was extracted from the JSON files and converted into corresponding annotated image data. This process involves parsing the JSON files using image processing libraries such as Python's PIL (Python Imaging Library) or OpenCV. It automatically retrieves contour coordinates in batches from the JSON files and draws corresponding-sized annotation label images based on the original image dimensions.

(2) Data Augmentation

Deep learning has achieved tremendous success in the task of image segmentation in computer vision, and one key factor contributing to this success is the support of powerful dataset resources. Improving the model's generalization ability, allowing it to perform well on unknown data, is currently a crucial task. Networks lacking

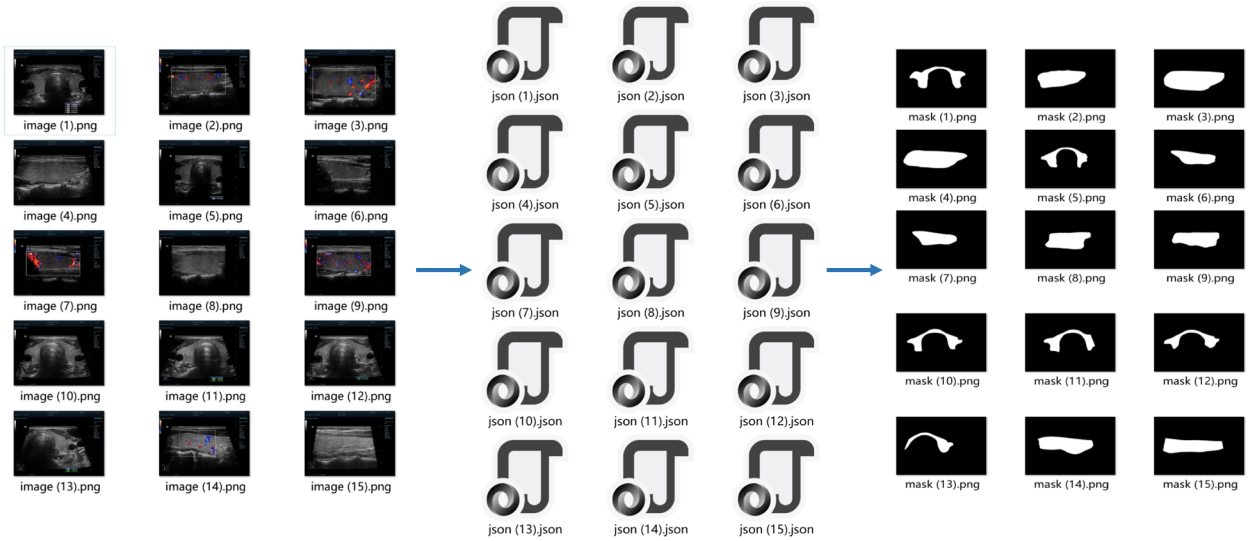


Fig. 5 The Process of Generating Labels

generalization ability are prone to overfitting, meaning they perform well on the training set but poorly on the test set.

To reduce the bias between the training and testing processes, data augmentation is a highly effective technique.

After data augmentation, the segmentation results represent a more comprehensive combination of information, aiming to reduce the discrepancy between training information and test data. The premise of implementing data augmentation techniques is to ensure translational invariance of the image data, meaning transformations applied to the images should not alter their intrinsic properties.

Due to the limited size of the pediatric thyroid ultrasound dataset, to prevent overfitting during network training, this study employed a data augmentation strategy to expand the original sample dataset, As shown in Fig. 6. Multiple-angle rotations (90°, 180°, 270°) were applied. The original set of 310 clinical pediatric thyroid ultrasound images was augmented to 1240 images.

Image segmentation evaluation metrics

A crucial foundation in medical image processing and analysis is medical image segmentation. Choosing a good segmentation algorithm for image segmentation can greatly assist in subsequent diagnostics. Therefore, determining the quality of an algorithm has become an important question. Typically, various evaluation metrics [31] are employed to quantitatively assess the segmentation results of a network, thereby examining the feasibility and accuracy of the network model.

Evaluation metrics include Intersection over Union (IoU), Mean Intersection over Union (MIoU), Accuracy, Precision, and Recall. All of these metrics can be represented by the four elements of a confusion matrix. The confusion matrix typically consists of four elements: TP (True Positive), FP (False Positive), TN (True Negative), and FN (False Negative).As shown in Fig. 7.

Where TP represents accurate prediction of true positive cases, where the true label is organ tissue or lesion area; TN represents accurate prediction of true negative cases, where the true label is background area; FP represents the misjudgment of background area as organ or lesion, indicating false positive cases (false alarm); FN represents the misjudgment of organ or lesion area as background, indicating false negative cases (missed detection). The above evaluation metrics can be expressed in terms of these four fundamental values:

$$IoU = \frac{TP}{TP + FP + FN} \tag{9}$$

$$mIoU = \frac{1}{2}\left(\frac{TP}{TP + FN + FP} + \frac{TN}{TP + FN + FP}\right) \tag{10}$$

$$Accuracy = \frac{TP}{TotalSample} \tag{11}$$

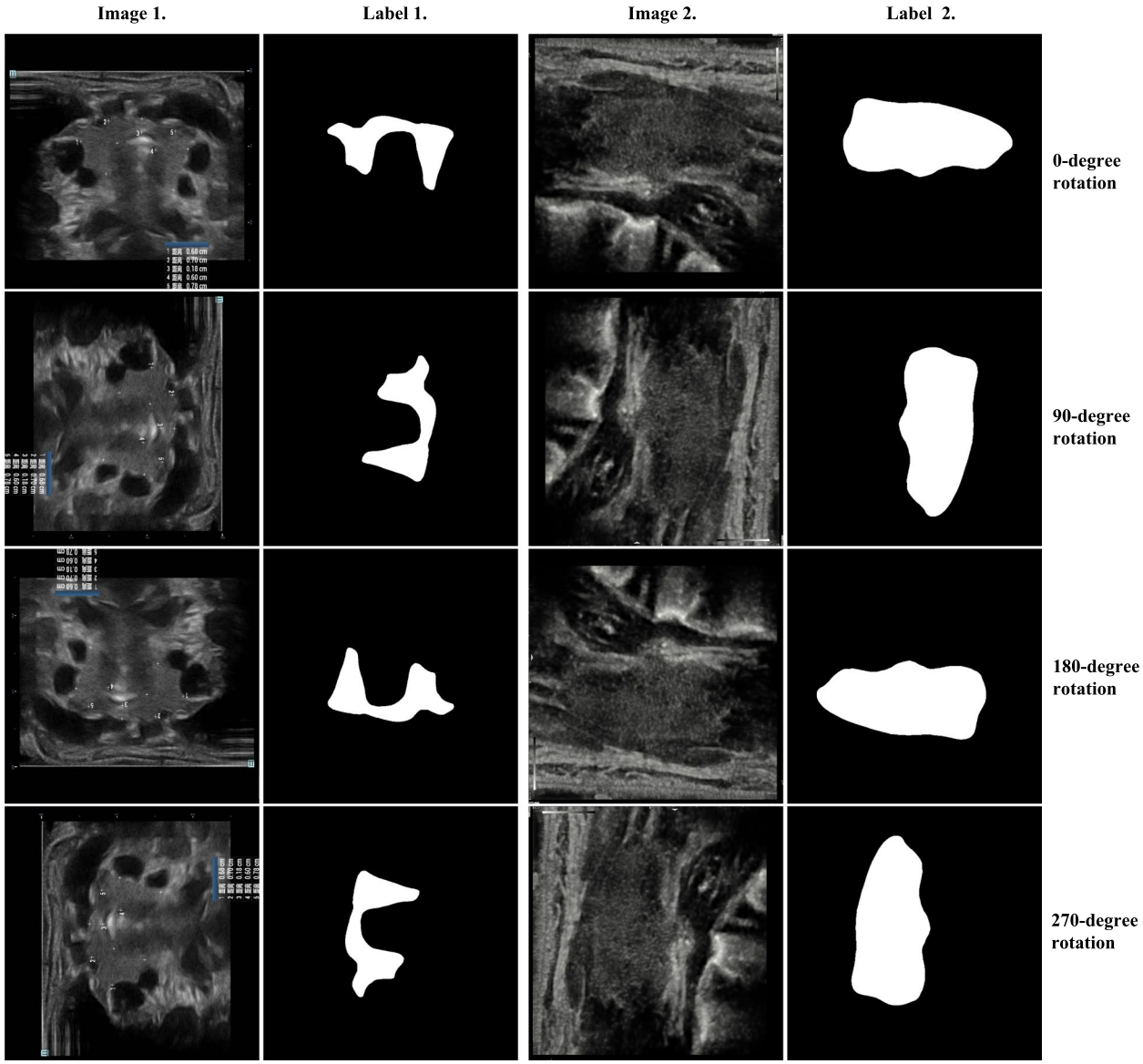


Fig. 6 The Process of Generating Labels

Predicted Value \ Actual Value	True	False
True	TP (True Positive)	FP (False Positive)
False	FN (False Negative)	TN (True Negative)

Fig. 7 The Process of Generating Labels

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

$$Recall = \frac{TP}{TP + FN} \tag{13}$$

IoU, mIoU, Accuracy, Precision, and Recall all have values ranging from 0 to 1, where a higher value within this interval indicates a higher degree of overlap between the predicted and true labels, and thus a more accurate model prediction [32]. These metrics are all used to evaluate the overall performance of the network in terms of segmentation accuracy. It is important to note that in the proposed DC-Contrast U-Net in this paper, the contrast block, while enhancing segmentation accuracy, also introduces higher complexity. This may limit the improvement of the segmentation

network in certain scenarios. Therefore, in addition to evaluating segmentation accuracy, experimental results also consider parameters and time consumption as evaluation metrics, providing a more comprehensive display of the computational complexity of various segmentation networks.

Experiment settings

(1) Experiment Environment

The experiment was conducted on a workstation running the Ubuntu 18.04 operating system. The workstation is equipped with two E5-2620 central processors with a clock frequency of 2.00 GHz, 128 GB RAM, and two GeForce RTX 3090 TURBO graphics cards, each with 24 GB RAM. The hardware configuration of the experiment environment is shown in Table 1. The final experiments were conducted using Visual Studio Code locally, connecting to the container through Remote SSH.

The experiment environment is set up with the Ubuntu operating system, and it comes pre-installed with the Conda 4.10.1 environment management system and open-source software package management system. Additionally, the necessary compilation software Pycharm and related frameworks required for the experiment are installed, as detailed in Table 2.

(2) Experimental Process

This study utilizes the PyTorch deep learning framework to establish the training environment for the entire experiment, which is primarily divided into the training and testing processes. To facilitate training and testing, the dataset needs to be centrally processed and placed in a specified directory. After data augmentation, ultrasound images and label data are passed as inputs to the model for training, and the model parameters are updated during the iterative process. After each iteration, the loss function value is computed using the label tensor and the tensor predicted by the network. The loss function is then used for backpropagation to update the parameters, continuously optimizing the loss function to reach its minimum. When the change in the loss function becomes small or almost negligible, the model is considered to have converged, and the model parameters are saved for future use. During the testing of the model's performance, the previously saved optimal model parameters are loaded. First, the saved model parameters are loaded into the model. Then, the test dataset is input into the model to obtain the model's predictions for the test data. Finally, by comparing the predicted results with the true label data, performance metrics such as mIoU, accuracy, precision, and recall are calculated.

(3) Experimental Parameter Settings

This study employs a total of 1240 images, with the dataset divided into training and testing sets in an 8:2 ratio. Consequently, the final experimental training set consists of 992 images, and the testing set consists of 248 images. The input image size for the model is set to 512×512×3, and the corresponding label data size is 512×512×1. The Batch Size is set to 4. The experiment utilizes the Adam optimizer with a well-optimized effect, and the initial learning rate is set to 0.0001. Focal Loss is combined with Lovasz Loss as the loss function, and the maximum number of iterations is set to 120.

To validate the proposed method in this chapter for thyroid ultrasound image segmentation networks in terms of performance and effectiveness, and to ensure the fairness and effectiveness of the experiments for comparison, various segmentation models mentioned in recent literature are compared. Ablation experiments demonstrate the impact of adding or removing different modules on segmentation performance. All other conditions remain the same; thus, all experiments are conducted under identical conditions, controlling for other variables.

The decision not to use a validation set was based on the following considerations: Firstly, we aimed to maximize the utilization of the limited dataset to enhance the model's generalization ability. Secondly, since we have controlled other variables through ablation studies and comparative experiments, ensuring the validity and reliability of the experimental results, the role of a validation set was deemed less critical in this study. Lastly, by directly evaluating the model using the test set, we simplified the experimental process and ensured a more direct and effective assessment of the model's performance.

Table 1 Experimental Hardware Environment Table

Hardware name	Model and size
CPU	Intel Xeon E5-2620
GPU	NVIDIA RTX3090 Turbo 24G
RAM	128G

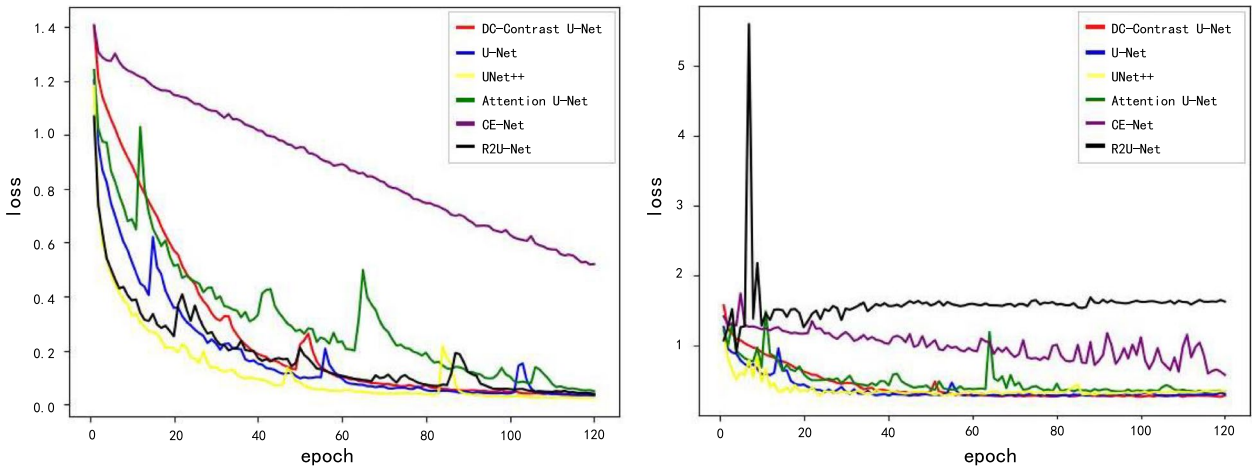
Table 2 Experimental Software Environment Table

Software name	Version
Host Operating System	Ubuntu System
Python	3.8
PyTorch	1.13.1
CUDA	11.8

Experimental results and analysis

Comparison experiments

Figure 8 illustrates the variation of the loss function for the training and testing sets during the training process of the proposed method and other networks. The red curve represents the loss function curve of the proposed



(a) Training Set Loss Change Curve (b) Test Set Loss Change Curve

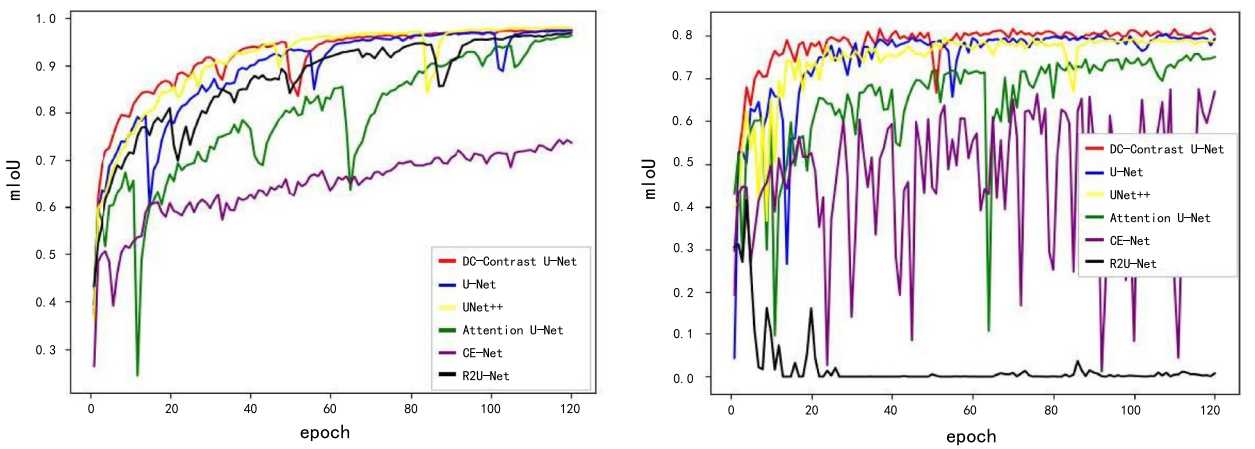
Fig. 8 Loss Function Variation Curves of Different Network Models

method, DC-Contrast U-Net, while the loss function curves of other networks are indicated by different colors, as shown in the legend in the upper right corner of the line chart.

In Fig. 8, (a) represents the loss function variation curve on the training set, and (b) represents the loss function variation curve on the testing set. Combining (a) and (b), it can be observed that the proposed method in this chapter converges quickly and steadily on both the training and testing sets. After epoch=60 on the training set, the loss function curve shows no significant fluctuations. Similarly, throughout the entire training process on the testing set, there are no noticeable abrupt changes in the loss function curve. The proposed method achieves faster convergence to the minimum value, with the curve almost parallel to the x-axis. Compared to other networks, the proposed network exhibits the lowest training difficulty.

Figure 9 illustrates the variation curves of the mean Intersection over Union (mIoU) on the training and testing sets during the training process for the proposed method and other networks. The red curve represents the mIoU curve of the proposed method, DC-Contrast U-Net, while the mIoU curves of other networks are indicated by different colors, as shown in the legend in the upper right corner of the line chart.

In Fig. 9, (a) represents the mIoU variation curve on the training set, and (b) represents the mIoU variation



(c) Training Set Average Intersection over Union Change Curve (c) Test Set Average Intersection over Union Change Curve

Fig. 9 Loss Function Variation Curves of Different Network Models

curve on the testing set. After epoch=60 on the training set, the mIoU curve shows no significant fluctuations, and it remains relatively stable throughout the entire training process on the testing set. Combining (a) and (b), it can be observed that the proposed method in this chapter achieves the highest mIoU on both the training and testing sets, indicating optimal segmentation performance of the model.

Finally, it should be noted that from the graph, it can be observed that the CE-Net model and R2U-Net model perform poorly on the clinical dataset used in this chapter. They have limited relevance to the proposed method. Therefore, in the subsequent experiments, only the segmentation results of the other four methods are compared.

In determining the value of λ in the loss function, a sensitivity analysis was conducted, and the value of λ was chosen to be 0.4. The specific results are shown in Table 3: From Table 3, it can be observed that the experimental results with different values of λ follow a normal distribution. The segmentation performance is optimal when λ is set to 0.4 in the loss function. Therefore, for the subsequent experiments in this chapter, λ in the loss function is consistently set to 0.4.

To validate the effectiveness of the proposed DC-Contrast U-Net network in the clinical dataset experiment, comparative experiments were conducted using DC-Contrast U-Net, U-Net, UNet++, and Attention U-Net. The visual representation is shown in Fig. 10, which demonstrates the segmentation results of the proposed network compared to other well-performing segmentation networks on a clinical thyroid dataset. This figure illustrates the comparison of segmentation results between well-performing segmentation networks in medical images and the proposed DC-Contrast U-Net on a clinical thyroid dataset. The figure showcases five sets of pediatric thyroid images from different patients and cross-sections. Column (a) displays five randomly selected original pediatric thyroid images from the testing set, column (b) shows the corresponding label images, column (c) presents the segmentation results of the proposed method, and columns (d), (e), and (f) depict the segmentation results of U-Net, UNet++, and Attention U-Net, respectively. UNet++ can capture features at different hierarchical levels, resulting in better segmentation performance compared to the baseline network U-Net. Although Attention U-Net introduces an attention mechanism to enhance the network's focus on important features, UNet++ introduces multiple branches at each pooling stage, providing the network with better feature representation capabilities. The improved DC-Contrast U-Net can extract sufficient irregular image information.

From the segmentation results, it can be observed that DC-Contrast U-Net performs the best on pediatric thyroid clinical data. In the highlighted red box, the proposed method in this chapter tends to be closer to the label data compared to other networks. Due to the incomplete development of the pediatric thyroid, its shape is less regular than the adult thyroid, and the boundaries are not clearly separated from surrounding tissues. Addressing the specific challenges of pediatric data, DC-Contrast U-Net segments the thyroid boundaries closer to the label data than other methods. For uncertain thyroid regions, the segmentation by DC-Contrast U-Net is more accurate. Next, the effectiveness of the proposed method is quantitatively explained.

Table 4 presents the segmentation performance metrics of different networks on the clinical pediatric thyroid dataset collected by the Sichuan Provincial Maternal and Child Health Hospital. For the U-Net network, the values of IoU, mIoU, accuracy, precision, and recall are 0.8444, 0.8159, 0.9079, 0.8319, and 0.7944, respectively. In contrast, the proposed DC-Contrast U-Net achieves values of 0.8952, 0.8660, 0.9528, 0.9183, and 0.8311 for IoU, mIoU, accuracy, precision, and recall, respectively. The metrics of IoU, mIoU, accuracy, precision, and recall show improvements of 6.01%, 6.14%, 1.5%, 4.95%, and 4.62%, respectively. In Table 5, the first row displays two metrics: Parameters and Multiply-Accumulate Operations (MACs). Parameters represent the total number of parameters in the model, commonly used to measure the size of deep learning models. For example, a 3×3 convolutional layer has nine parameters for the convolution operation plus one parameter for the bias operation, totaling ten parameters. Another metric is Multiply-Accumulate Operations, where 1 MAC includes one multiplication operation and one addition operation.

Table 3 Sensitivity Analysis of the Parameter λ in the Loss Function

Parameter λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
mIoU	0.8432	0.8557	0.8567	0.8660	0.8634	0.8636	0.8559	0.8489	0.8511
Accuracy	0.9431	0.9511	0.9444	0.9528	0.9452	0.9519	0.9433	0.9418	0.9302
Precision	0.9007	0.9032	0.8927	0.9183	0.9103	0.9039	0.8999	0.9034	0.8839
Recall	0.8055	0.8261	0.8293	0.8311	0.8231	0.8136	0.8267	0.8302	0.8156

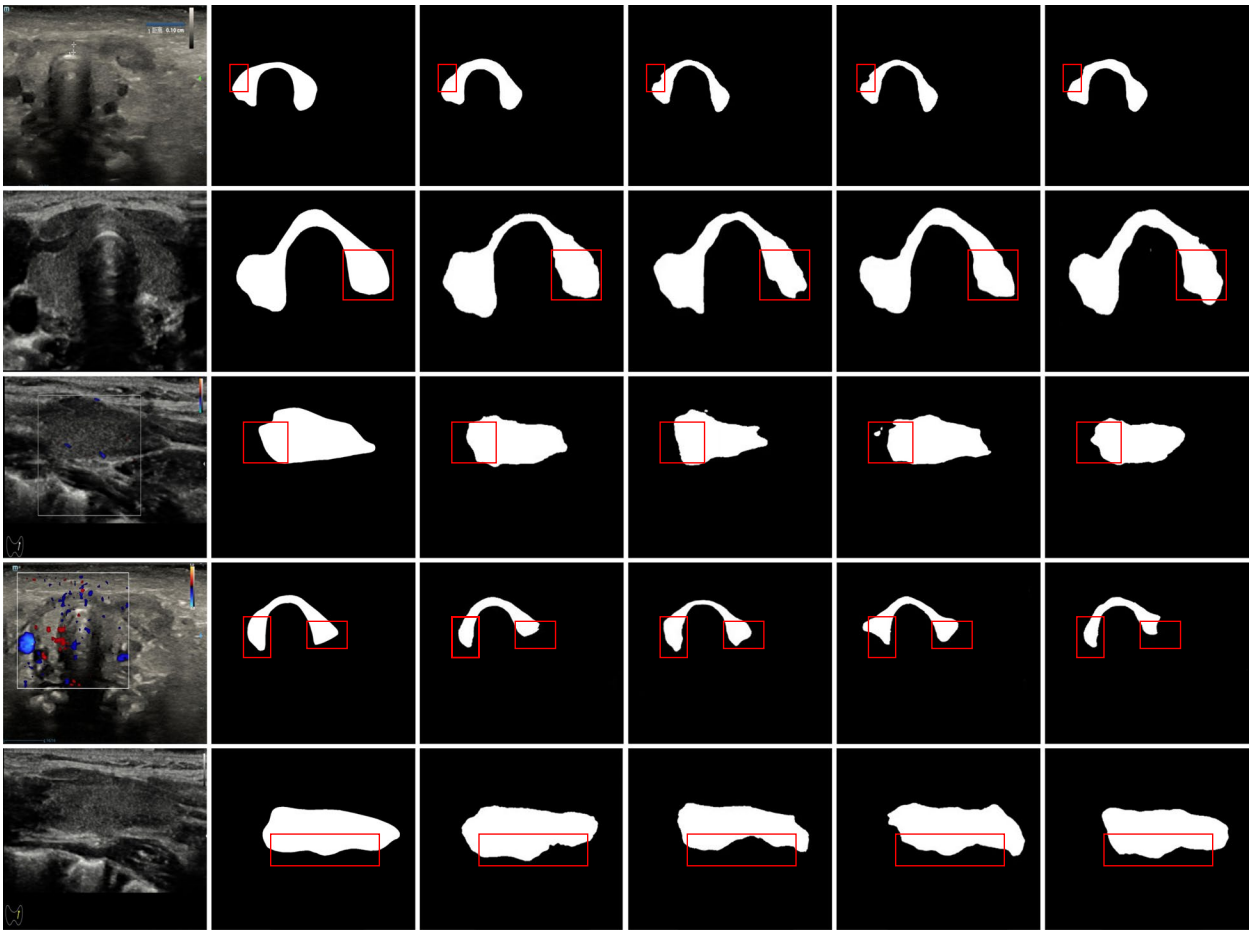


Fig. 10 Demonstrations of the segmentation results of DC-Contrast U-Net and other models on five sets of clinical data

Table 4 Network Performance on Pediatric Thyroid Dataset

Network	IoU	mIoU	Accuracy	Precision	Recall
U-Net [10]	0.8444	0.8159	0.9079	0.8319	0.7944
UNet++ [33]	0.8740	0.8386	0.9360	0.8758	0.8198
Attention U-Net [34]	0.8147	0.8064	0.9037	0.7758	0.7980
DC-Contrast U-Net	0.8952	0.8660	0.9528	0.9183	0.8311

Table 5 Network Performance on Pediatric Thyroid Dataset

Network	Number of parameters (M)	Multiply-accumulate operations (G)
U-Net [10]	7.3	4.1
UNet++ [33]	27.8	9.8
Attention U-Net [34]	19.3	16.7
DC-Contrast U-Net	2.5	1.2

This metric is also commonly used to evaluate the computational complexity and size of deep learning models.

From the table, it can be observed that the proposed DC-Contrast U-Net has a parameter count of only 2.5M and MACs of only 1.2G, much smaller than the other networks. By introducing the contrast block into U-Net, the convolutional kernels responsible for extracting image texture information are fixed, eliminating the need to compute the convolution kernels for the contrast block during a significant portion of the learning process. This reduction results in a decrease in the number of parameters and model computation.

Differential experiments

In order to enhance model accuracy without introducing excessive computational complexity, DC-Contrast U-Net has been specifically improved based on the baseline U-Net. To validate the effectiveness of the proposed enhancements, a series of ablation experiments were

conducted using IoU, mIoU, accuracy, precision, and recall as performance and feasibility evaluation metrics.

To comprehensively evaluate the performance of DC-Contrast U-Net, we also conducted comparative experiments with other advanced segmentation models. The models selected for comparison include U-Net, UNet++, and Attention U-Net, as these models are widely used and representative in the field of segmentation technology. The comparison is based on key performance indicators such as accuracy and inference speed. The evaluation process includes a quantitative analysis of segmentation accuracy, involving metrics like Dice coefficient, Intersection over Union (IoU), and pixel accuracy. Additionally, inference speed was measured to assess the efficiency of each model in practical applications.An ablation study was also performed to investigate the impact of specific components in DC-Contrast U-Net. This includes evaluating the contribution of Contrast Blocks and Squeeze-and-Excitation Blocks to overall performance. By systematically removing or modifying these components, we analyzed the effect of each element on segmentation results and model efficiency. Detailed experimental setups and results of these comparisons are presented in the following sections. We provide both visual and quantitative assessments of DC-Contrast U-Net relative to other models, offering a deep understanding of its advantages and potential areas for improvement.

Firstly, the designed contrast block was added to the U-Net encoder-decoder framework, compressing and activating blocks were incorporated into skip connections, and ordinary convolutions were replaced with deformable convolutions. This approach is referred to as DC-Contrast U-Net. Another method, named Contrast U-Net, involves adding the contrast block to the U-Net encoder-decoder framework, introducing compressing and activating blocks into skip connections, but using ordinary convolutions in the convolutional part. Finally, DC U-Net involves adding compressing and activating blocks to skip connections in U-Net without introducing the contrast block and still replacing ordinary convolutions with deformable convolutions. To verify the effectiveness of the introduced contrast block, added compressing and activating blocks, and the substitution of deformable convolutions, ablation experiments were conducted on clinical pediatric thyroid ultrasound images obtained from the hospital. Table 6 presents the quantitative results obtained from the ablation study on clinical pediatric thyroid ultrasound images.

As shown in Table 6, different modules were successively added to the original U-Net network as a base, and the trained network model results were obtained accordingly. The proposed DC-Contrast U-Net improves image segmentation metrics, with mIoU, Dice, accuracy, precision, and recall increasing to 0.8660, 0.9528, 0.9183, 0.8311, respectively. Larger values in the range of 0 to 1 for these metrics indicate better model segmentation performance. Comparing Contrast U-Net with the baseline U-Net and DC-Contrast U-Net with DC U-Net in respective ablation experiments reveals performance improvements in mIoU, accuracy, precision, and recall. This suggests that the introduced contrast block enhances the network's sensitivity to fine textures in the image. Additionally, replacing ordinary convolutions with deformable convolutions enhances the segmentation performance of the network. Through ablative experiments, the metrics mIoU, accuracy, precision, and recall of the DC-Contrast U-Net model have seen improvements. The introduced contrast block, compression and excitation blocks, and deformable convolution components contribute to the enhanced network performance, allowing the model to focus on fine image textures and exhibit increased sensitivity to irregular regions, ultimately improving segmentation performance.

Table 6 Differential Contrast U-Net and ablative experiments with other networks

Network	Dataset	mIoU	Accuracy	Precision	Recall
U-Net [10]	Training Set	0.9547	0.9679	0.9501	0.9563
	Test Set	0.7947	0.9472	0.8945	0.8112
Contrast U-Net	Training Set	0.9342	0.9483	0.9419	0.9479
	Test Set	0.8050	0.9341	0.9012	0.8210
DC U-Net	Training Set	0.8900	0.9727	0.9046	0.9455
	Test Set	0.7930	0.9346	0.8503	0.8193
DC-Contrast U-Net	Training Set	0.9754	0.9786	0.9762	0.9524
	Test Set	0.8660	0.9528	0.9183	0.8311

To validate the effectiveness of the loss function, ablation experiments were conducted on the loss function used in the model, and the results are shown in Table 7. In Table 7, the proposed Focal Loss (FL) is introduced as the loss function to measure the relationship between predicted values and ground truth in the network. Considering the specificity of medical image segmentation in this study, the Binary Cross-Entropy (BCE) loss function is further utilized to assist FL with fine-tuning. The fine-tuning exponent is set to 0.4 and

Table 7 Loss Function Ablation Experiment

Method	mIoU	Accuracy	Precision	Recall
FL	0.7891	0.9090	0.8732	0.7964
LL	0.6259	0.6573	0.8599	0.7712
FL+LL	0.8537	0.9362	0.8813	0.8268
FL+0.4LL	0.8660	0.9528	0.9183	0.8311

1 for comparative experiments. The introduced loss function, Loss = FL + 0.4 BCE, shows improvements in mIoU, accuracy, precision, and recall compared to Loss = FL + BCE, with increases of 1.23%, 1.66%, 3.7%, and 0.43%, respectively. The ablation experiments demonstrate the effectiveness of the proposed loss function.

It is important to note that in the above experiments, separate training and testing sets were used. The results from the training set are primarily used to examine the fitting of the model during the training process, aiding the training phase and serving as a reference. However, they do not indicate the segmentation performance of the model. All evaluations of segmentation performance in this chapter are based on the segmentation results from the testing set.

Additionally, to validate the effectiveness of the proposed improvements, experiments were conducted for each individual improvement. First, each improvement was separately added to the base model for training and testing, with relevant performance metrics recorded. Subsequently, different combinations of improvements were added step by step, and performance metrics were recorded again. Finally, a comparison was made with the base model to assess the overall improvement effect. This approach ensures a thorough evaluation of each improvement, determining whether the proposed enhancements should be adopted.

Discussion

Performance analysis of DC-Contrast U-Net

In this paper, we provide a detailed overview of the data preparation, preprocessing pipeline, experimental setup, and parameter configuration for clinical pediatric thyroid ultrasound images. To address the specific characteristics of pediatric thyroid ultrasound tasks, we improved the widely used U-Net model and proposed the DC-Contrast U-Net model, aiming to achieve better segmentation performance with lower complexity in medical image segmentation. Compared to the traditional U-Net, DC-Contrast U-Net shows significant advantages in accurately extracting texture information from pediatric thyroid ultrasound images. Traditional U-Net often faces challenges such as blurred boundaries of the target and a limited number of pediatric thyroid ultrasound images, which hinder its ability to fully learn and adapt to image features, resulting in lower segmentation accuracy.

In contrast, DC-Contrast U-Net introduces the Contrast Block to more accurately extract texture information of the pediatric thyroid gland, improving segmentation precision and reliability. The Contrast Block excels at handling grayscale, brightness, and texture similarities, effectively overcoming the shortcomings of traditional U-Net in dealing with blurred boundaries. Additionally, DC-Contrast U-Net incorporates Squeeze-and-Excitation Blocks and Deformable Convolution Layers, further enhancing the model's ability to learn texture information for each channel and adapt to irregular organ shapes, significantly improving segmentation performance. The performance metrics of the DC-Contrast U-Net model show significant improvements, with IoU, mIoU, accuracy, precision, and recall metrics enhanced by 6.01%, 6.14%, 1.5%, 4.95%, and 4.62%, respectively. The results are as shown in Table 4. These improvements highlight the effectiveness of the proposed model in enhancing segmentation quality and efficiency in complex pediatric thyroid ultrasound tasks.

Effectiveness analysis of Squeeze-and-Excitation Block and Deformable Convolution Layer

The introduction of Squeeze-and-Excitation (SE) Blocks and Deformable Convolution Layers in DC-Contrast U-Net serves specific purposes and brings notable improvements, particularly in handling the shapes of irregular organs and adapting to the texture information of each channel. The structure of the SE block is shown in the Fig. 11.

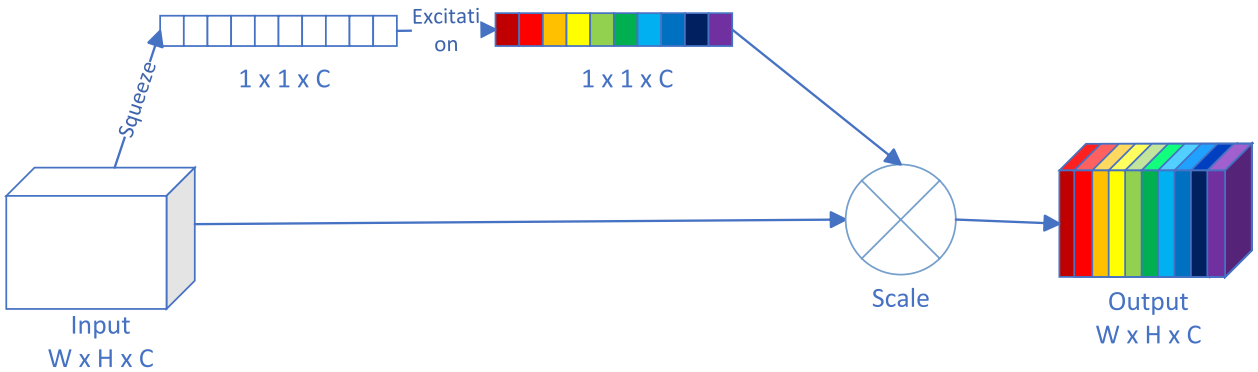


Fig. 11 The structure of the Squeeze-and-Excitation (SE) module

The primary motivation for introducing the SE Block is to enhance the network's ability to capture and emphasize important features. In medical image segmentation, each channel of the feature map carries different aspects of the image, such as intensity, texture, and structure. The SE Block is designed to dynamically recalibrate channel-wise feature responses, thereby improving the network's sensitivity to crucial information. The SE Block performs a squeeze operation to aggregate global information from each channel, followed by an excitation operation to reallocate weights, highlighting important features and suppressing less important ones. This process helps the network better understand and utilize key information in the images. By introducing the SE Block, DC-Contrast U-Net can better learn the texture information of each channel, improving segmentation accuracy, especially in complex and detailed image regions.

Experimental results and performance evaluation

In this study, we compared the performance of DC-Contrast U-Net with other segmentation models, including U-Net, UNet++, and Attention U-Net, focusing on segmentation accuracy and inference speed. The results are as shown in Table 4. The experimental results demonstrate that DC-Contrast U-Net excels in both metrics. Compared to the traditional U-Net, DC-Contrast U-Net shows a significant improvement in segmentation accuracy when dealing with pediatric thyroid ultrasound images, particularly in complex and detailed image regions, where it more precisely extracts the texture information of the target area. While UNet++ achieves better segmentation performance by leveraging features at different levels, DC-Contrast U-Net further enhances accuracy through the introduction of Contrast Blocks and Squeeze-and-Excitation Blocks, particularly excelling in handling irregular boundaries and fine details. In terms of inference speed, DC-Contrast U-Net also performs better. Its optimized network structure and efficient convolution operations significantly improve computational efficiency. Compared to the more computationally intensive Attention U-Net, DC-Contrast U-Net achieves faster inference while maintaining high accuracy. This indicates that DC-Contrast U-Net offers a more precise and efficient solution for medical image segmentation tasks, especially when dealing with complex and irregular targets.

Summary

In this paper, we provided a detailed overview of the data preparation, preprocessing pipeline, experimental setup, and parameter configuration for clinical pediatric thyroid ultrasound images. Addressing the specific challenges of pediatric thyroid ultrasound tasks, we improved the widely-used U-Net model and proposed the DC-Contrast U-Net model, aiming to achieve better performance in medical image segmentation while reducing complexity. The novel contrast block introduced effectively handles grayscale, brightness, and texture similarities, thereby enhancing segmentation precision. Additionally, the compression and excitation blocks added to the skip connections improve the adaptability of feature response between channels, further boosting segmentation accuracy. The use of deformable convolutions enhances the model's ability to adapt to the shapes of irregular organs. Comparative analysis with other U-Net-based improved models demonstrated that DC-Contrast U-Net outperforms other related models in both segmentation accuracy and inference speed. Specifically, the model showed improvements of 6.01% in IoU, 6.14% in mIoU, 1.5% in accuracy, 4.95% in precision, and 4.62% in recall. These enhancements highlight the effectiveness of DC-Contrast U-Net in improving segmentation quality and efficiency for complex pediatric thyroid ultrasound tasks. However, challenges remain, such as accurately segmenting regions with fuzzy boundaries and handling significant noise. Future research should focus on further optimizing the proposed medical image segmentation network to enhance its performance.

Acknowledgements

Not applicable.

Authors' contributions

Bo Peng: Undertook the writing of the manuscript and was responsible for the main research and experimental design. Wu Lin: Undertook the writing of the manuscript and was responsible for the main research and experimental design. Wenjun Zhou: Participated in the writing of the manuscript and provided key theoretical and experimental support. Yan Bai: Provided valuable advice and discussions for the research, and contributed to the interpretation and analysis of experimental results. Anguo Luo: Participated in the collection and processing of experimental data and conducted preliminary analysis of the results. Shenghua Xie: Provided guidance and support in the professional field and played an important role in the implementation of the study. Lixue Yin: Provided valuable suggestions and assistance in experimental design and data analysis.

Funding

No fund was received for this study.

Availability of data and materials

The original contributions presented in the study are included in the article. Further inquiries can be directed to the corresponding author.

Code availability

The original contributions presented in the study are included in the article. Further inquiries can be directed to the corresponding author.

Declarations

Ethics approval and consent to participate

We confirm that the experimental protocol was approved by the ethics committee of Sichuan Provincial Maternity and Child Health Care Hospital, located in Chengdu, China (Address: Street, Chengdu, 610000, China). Written informed consent was obtained from all participants in this study.We confirm that all methods were conducted in accordance with relevant guidelines and regulations.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 11 June 2024 Accepted: 29 August 2024
Published online: 11 October 2024

References

1. Maroulis DE, Savelonas MA, Karkanis SA, Iakovidis DK, Dimitropoulos N. Computer-aided thyroid nodule detection in ultrasound images. 2005:271–276. https://doi.org/10.1109/CBMS.2005.44.
2. Li M, Zhou H, Li X, Yan P, Jiang Y, Luo H, et al. SDA-Net: Self-distillation driven deformable attentive aggregation network for thyroid nodule identification in ultrasound images. Artif Intell Med. 2023;146:102699. https://doi.org/10.1016/j.artmed.2023.102699.
3. Hanley P, Lord K, Bauer AJ. Thyroid Disorders in Children and Adolescents: A Review. JAMA Pediatr. 2016 10;170(10):1008–19. https://doi.org/10.1001/jamapediatrics.2016.0486.
4. Francis GL, Waguespack SG, Bauer AJ, Angelos P, Benvenga S, Cerutti JM, et al. Management Guidelines for Children with Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid®. 2015;25(7):716–59. https://doi.org/10.1089/thy.2014.0460.
5. Chanchal R, Neha N, Shipra A, Prabhakar M, Akanksha S, Andrey B. Cytological evaluation of thyroid nodules in children and young adults: a multi-institutional experience. Endocrine. 2023;80:580–8. https://doi.org/10.1007/s12020-022-03297-0.
6. Angel Viji KS, Jayakumari J. Automatic detection of brain tumor based on magnetic resonance image using CAD System with watershed segmentation. 2011:145–150. https://doi.org/10.1109/ICSCCN.2011.6024532.
7. Bakkouri B I. 2MGAS-Net: Multi-level Multi-scale Gated Attentional Squeezed Network for Polyp Segmentation. SIViP. 2024;18(6):5377-86. https://doi.org/10.1007/s11760-024-03240-y.
8. Bakkouri I, Afdel K. Convolutional Neural-Adaptive Networks for Melanoma Recognition. In: Mansouri A, El Moataz A, Nouboud F, Mammass D. Image and Signal Processing. Cham: Springer International Publishing. 2018. p. 453–460.
9. Yu T, Muying L, Shaoyi W, Shunping J. An interactive method for bridging the gap between deep learning based building contour segmentation and manual annotation. Int J Remote Sens. 2024;45:2702–20.
10. Chen Y, Li D, Zhang X, Jin J, Shen Y. Computer aided diagnosis of thyroid nodules based on the devised small-datasets multi-view ensemble learning. Med Image Anal. 2021;67:101819. https://doi.org/10.1016/j.media.2020.101819.
11. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015. pp. 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965.
12. Julesz B. A method of coding television signals based on edge detection. Bell Syst Tech J. 1959;38(4):1001–20. https://doi.org/10.1002/j.1538-7305.1959.tb01586.x.
13. Li Z, Liu C, Liu G, Cheng Y, Yang X, Zhao C. A novel statistical image thresholding method. AEU Int J Electron Commun. 2010;64(12):1137–47. https://doi.org/10.1016/j.aeue.2009.11.011.
14. Du W, Sang N. An effective segmentation method of ultrasonic thyroid nodules. 2015;9814:98140F. https://doi.org/10.1117/12.2205406.
15. Parveen Z, Alam MA, Shakir H. Assessment of quality of rice grain using optical and image processing technique. In: 2017 International Conference on Communication, Computing and Digital Systems (C-CODE), 2017. pp. 265–270. https://doi.org/10.1109/C-CODE.2017.7918940.
16. Zohra BF, Leila D. Active Contour Extension Basing on Haralick Texture Features, Multi-gene Genetic Programming, and Block Matching to Segment Thyroid in 3D Ultrasound Images. Arab J Sci Eng. 2022;48:2429–40. https://doi.org/10.1007/s13369-022-07286-3.
17. Adams R, Bischof L. Seeded region growing. IEEE Trans Pattern Anal Mach Intel. 1994;16(6):641–7. https://doi.org/10.1109/34.295913.
18. Jie Z, Wei Z, Li Z, Hua T. Segmentation of ultrasound images of thyroid nodule for assisting fine needle aspiration cytology. Health Inf Sci Syst. 2013;1:5. https://doi.org/10.1186/2047-2501-1-5.
19. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid Scene Parsing Network. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. pp. 6230–6239. https://doi.org/10.1109/CVPR.2017.660.
20. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Trans Pattern Anal Mach Intel. 2018;40(4):834–48. https://doi.org/10.1109/TPAMI.2017.2699184.
21. Xiuxiu H, Jun GB, Yang L, Yingzi L, Tonghe W, J CW, et al. 3D Thyroid Segmentation in CT Using Self-attention Convolutional Neural Network. Medical imaging 2020: computer-aided diagnosis, vol. 11314. 2020. https://doi.org/10.1117/12.2549786.
22. Zhou S, Chen W. Prof. Young Jun Chai: artificial intelligence for thyroid ultrasound image analysis. Ann Thyroid. 2019;4(0). https://doi.org/10.21037/aot.2019.07.05.
23. Zhou R, Wang J, Xia G, Xing J, Shen H, Shen X. Cascade Residual Multiscale Convolution and Mamba-Structured UNet for Advanced Brain Tumor Image Segmentation. Entropy. 2024;26:385.
24. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE. 1998;86(11):2278–324. https://doi.org/10.1109/5.726791.
25. Gu Z, Cheng J, Fu H, Zhou K, Hao H, Zhao Y, et al. CE-Net: Context Encoder Network for 2D Medical Image Segmentation. IEEE Trans Med Imaging. 2019;38(10):2281–92. https://doi.org/10.1109/TMI.2019.2903562.
26. Gong H, Chen G, Wang R, Xie X, Mao M, Yu Y, et al. Multi-Task Learning For Thyroid Nodule Segmentation With Thyroid Region Prior. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), 2021. pp. 257–261. https://doi.org/10.1109/ISBI48211.2021.9434087.
27. Hu J, Shen L, Sun G. Squeeze-and-Excitation Networks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018. pp. 7132–7141. https://doi.org/10.1109/CVPR.2018.00745.
28. Dai J, Qi H, Xiong Y, Li Y, Zhang G, Hu H, et al. Deformable Convolutional Networks. In: 2017 IEEE International Conference on Computer Vision (ICCV), 2017. pp. 764–773. https://doi.org/10.1109/ICCV.2017.89.
29. He Y, Xiang S, Zhou W, Peng B, Wang R, Li L. A Novel Contrast Operator for Robust Object Searching. In: 2021 17th International Conference on Computational Intelligence and Security (CIS), 2021. pp. 309–313. https://doi.org/10.1109/CIS54983.2021.00071.
30. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal Loss for Dense Object Detection. In: 2017 IEEE International Conference on Computer Vision (ICCV), 2017. pp. 2999–3007. https://doi.org/10.1109/ICCV.2017.324.
31. Russell BC, Torralba A, Murphy KP, Freeman WT. LabelMe: a database and web-based tool for image annotation. Int J Comput Vis. 2008;77:157–73.
32. Aziz TA, Allan H. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med Imaging. 2015;15:29.

Publisher's Note