Technical Article

Ultrasonic Imaging
2023, Vol. 45(4) 159–174

DOI: 10.1177/01617346231169789
journals.sagepub.com/home/uix
SAGE

A Multiscale Attentional Unet Model for Automatic Segmentation in Medical Ultrasound Images

Rui Wang[1], Haoyuan Zhou[1], Peng Fu[2], Hui Shen[1], and Yang Bai[3]

Abstract

Ultrasonography has become an essential part of clinical diagnosis owing to its noninvasive, and real-time nature. To assist diagnosis, automatically segmenting a region of interest (ROI) in ultrasound images is becoming a vital part of computer-aided diagnosis (CAD). However, segmenting ROIs on medical images with relatively low contrast is a challenging task. To better achieve medical ROI segmentation, we propose an efficient module denoted as multiscale attentional convolution (MSAC), utilizing cascaded convolutions and a self-attention approach to concatenate features from various receptive field scales. Then, MSAC-Unet is constructed based on Unet, employing MSAC instead of the standard convolution in each encoder and decoder for segmentation. In this study, two representative types of ultrasound images, one of the thyroid nodules and the other of the brachial plexus nerves, were used to assess the effectiveness of the proposed approach. The best segmentation results from MSAC-Unet were achieved on two thyroid nodule datasets (TND-PUH3 and DDTI) and a brachial plexus nerve dataset (NSD) with Dice coefficients of 0.822, 0.792, and 0.746, respectively. The analysis of segmentation results shows that our MSAC-Unet greatly improves the segmentation accuracy with more reliable ROI edges and boundaries, decreasing the number of erroneously segmented ROIs in ultrasound images.

Keywords

ultrasound images, automatic ROI segmentation model, thyroid nodule segmentation, brachial plexus nerve segmentation, multiscale attentional convolution (MSAC), computer-aided diagnosis

Introduction

Ultrasonography is characterized by its low cost, safety, real-time imaging capability and has become an indispensable imaging modality in medical diagnosis, regional anesthesia, and intraoperative navigation. Moreover, it is commonly utilized in conjunction with other medical imaging methods, such as magnetic resonance imaging (MRI) and computed tomography (CT) in clinical diagnosis and therapy. However, the image quality of ultrasound images is relatively poor due to various limiting factors, such as noise and speckle, making it more difficult to utilize for subsequent diagnosis processes.[1] Specifically, the clinical applications of medical ultrasound images, such as fast and accurate diagnosis of malignant tumors or regional anesthesia, rely on the delineation of a target's boundaries. Therefore, although the acquisition of the region of interest (ROI), such as suspicious lesion areas, and targeted nerve areas, is not the end goal by itself, only when the ROI is obtained can a subsequent diagnosis be made. However, manually acquiring ROI is time-consuming and, because it often

[1]Laboratory of Precision Opto-mechatronics Technology, Ministry of Education, Institute of Instrumentation Science and Opto-electronics Engineering, Beihang University, Beijing, China
[2]Department of Ultrasound, Peking University Third Hospital, Beijing, China
[3]Department of General Surgery, Peking University Third Hospital, Beijing, China

Corresponding Author:
Rui Wang, Laboratory of Precision Opto-mechatronics Technology, Ministry of Education, Institute of Instrumentation Science and Opto-electronics Engineering, Beihang University, No.37, Xueyuan Road, Haidian District, 100191, China.
Email: wangr@buaa.edu.cn

requires the expertise of an operator, is also costly. Therefore, to assist in clinical diagnosis and therapy, it is crucial to employ computer-aided diagnosis (CAD) technology to detect and segment desired ROIs. CAD provides sonographers with an accurate and objective second opinion quickly while also decreasing their heavy workload, reducing the risk of misdiagnosis due to exhaustion. To date, according to human intervention, the ultrasound image segmentation methods applied in CAD can be categorized as semiautomatic methods[2–4] and fully automatic methods.[5–7]

The semiautomatic methods require operator interaction to mark the ROI either using seeds or assigning corresponding features manually, and the knowledge and experience of the clinician must generally be relied upon. Semiautomatic segmentation methods can be subdivided into image information-based models and learning-based models. In the former models, low-level image information and data-distribution features are utilized to complete segmentation, such as regional growth,[8] graph-cut and Gaussian process-based,[4,9,10] and active contour.[11] Gonzalez et al.,[4] for instance, utilized the graph cut and wavelet transform method to segment brachial plexus nerves, and the dice coefficient reached 60.57%. In the latter models, traditional machine learning, as a typical technology employed in the learning-based segmentation approaches, classifies pixels or image blocks by manually extracting features based on statistical knowledge. Moreover, the classifiers are constructed on various algorithms, such as decision tree (DT), and support vector machine (SVM). Chang et al.,[12] for example, manually extracted 41 features and employed the DT algorithm to segment thyroid nodules in six ultrasound images, and the segmentation accuracy was 97.5%. In brief, although semiautomatic segmentation methods can contribute to the development of CAD to some extent, manually extracting effective features is a complex process, and there are many subjective factors involved in determining the conditions for initial settings. Simultaneously, the initialization process has a massive effect on the segmentation results. In other words, migrating from one segmentation task to another is difficult.

On the other hand, fully automatic ROI segmentation approaches that can extract image features without manual intervention are drawing increasing attention as a solution to overcome the disadvantages of semiautomatic methods. Undoubtedly, the currently popular deep learning is a typical data-driven learning technology to achieve automatic ROI segmentation, which utilizes deep convolutional neural networks (CNNs) to automatically extract and exploit more abstract and high-level nonlinear features. Ma et al.[5] suggested a cascade CNN model that combines two CNNs and a new splitting approach to achieve thyroid nodule segmentation. The receiver operating characteristic area under the curve was 98.51%. In an ordinary way, the traditional CNN segmentation methods usually feed blocks of pixels into a CNN through a sliding window and employ fully connected layers for ROI segmentation, leading to significant repeated calculations and low computational efficiency.

In recent years, with the rise of the full convolution network (FCN),[13] which effectively avoids the redundant computation problem of traditional CNNs in image segmentation by skipping connections, end-to-end architectures without fully connected layers have attracted a great deal of attention in image-segmentation methods.[14] Ronneberger et al.,[15] for instance, proposed the Unet architecture for medical image segmentation, based on the encoder and decoder from FCN. In this architecture, context information is captured from a contracting path, while a symmetric expanding path is employed to ensure exact localization. Unet is becoming a prominent basic architecture for medical ROI segmentation tasks.[6,16]Chu et al.[16] performed thyroid nodule segmentation using a marker-guided Unet model, and this interactive segmentation method utilized four manually labeled endpoints of the long and short axes of the nodules to guide segmentation. The experiments were tested on 510 ultrasound images with a Dice coefficient of 95.76%. Kakade and Dumbali[6] proposed a brachial plexus segmentation study based on Unet. By employing Unet and a postprocessing algorithm based on principal component analysis, the segmentation Dice coefficient reached 68.83%. Certainly, there have been multiple instances of Unet and its modified CNNs being employed to segment ROIs in medical ultrasound images.[6,16] The most widely utilized ones are Res-Unet[17] as well as Unet++.[18] Res-Unet is constructed by adding a residual connection based on the Unet architecture.[17] The well-known residual module can effectively tackle the problem of training challenges caused by network depth and may improve CNNs for acquiring deeper medical ROI characteristics, and the insertion of residual connections might help to accelerate convergence while also avoiding CNN deterioration. On the other hand, from improving the connection method between convolutional layers, the densely connected convolutional networks (DenseNet) utilized shortcuts to expand the depth of CNNs for gradient propagation.[19] Combining the principle of DenseNet, Zhou et al. presented Unet++, connecting the first to the fourth layers in the Unet architecture,[18] and this architecture has the benefit of allowing the CNN to learn more features associations between various layers to some extent.

As everyone knows, semantic segmentation is a pixel-level classification task, and how to efficiently couple more dimensional receptive field features when adopting CNNs is the most vital hints for obtaining precise ROIs. It seems to be that small receptive fields are better at

capturing detailed features, such as edges, boundaries and texture features, while large receptive fields are more capable of extracting features from the entire ROI. Therefore, in order to more effectively combine multi-scale features to achieve ROI segmentation, Chen et al. proposed Deeplab V3,[20] which employ atrous convolution in cascade or in parallel to capture multi-scale context by adopting multiple atrous rates. Furthermore, they proposed to augment Atrous Spatial Pyramid Pooling (ASPP) module, which probes convolutional features at multiple scales. Specically, Chen et al. proposed Deeplab V3+,[21] extends Deeplab V3 by adding a simple yet effective decoder module to refine the segmentation results. They applied the depthwise separable convolution to both ASPP and decoder modules, resulting in a faster and stronger encoder-decoder network. Deeplab V3 + was validated on PASCAL VOC 2012 and cityscape datasets, with performance of 89.0% and 82.1%, respectively.

However, since most of the convolution kernels utilized in contemporary end-to-end CNNs (for instance, Unet,[15] Res-Unet,[17] and Unet++[18]) are of a single size, the size of the perceptive field reflected by the extracted features remains constant, leading to insufficient ability to obtain and utilize multiscale features.

In our study, to overcome the aforementioned difficulties, a novel model is developed for automatically segmenting ROIs in medical ultrasound images, which not only maintains the end-to-end basic architecture but also effectively addresses the problem that state-of-the-art CNNs are inadequate for multiscale feature fusion. In particular, two representative ROI in ultrasound images, thyroid nodules and brachial plexus nerve, are selected for segmentation in our work. The main contributions of our study are as follows:

1. The multiscale attentional convolution (MSAC) module is proposed as a solution to the problem of inadequate extraction and exploitation of multi-scale features by typical CNNs. MSAC acquires features with various receptive fields by performing cascade convolution operations and deploys a self-attention approach to effectively merge these spatial multiscale features, simulating the cognitive capacity of sonographers.
2. The MSAC-Unet, which applies MSAC instead of standard convolution in each encoder and decoder, is developed on the backbone of Unet. As a lightweight architecture, it achieves competitive performance on both thyroid nodule and brachial plexus nerve segmentation tasks compared with several state-of-the-art CNNs.

The remainder of this paper is arranged as follows. Section 2 introduces the justification for segmenting thyroid nodules and brachial plexus nerves, as well as the datasets employed in our study. Section 3 describes the architecture and principle of MSAC and MSAC-Unet, focusing on an automatic ROI segmentation model. The experimental results and implementation details are presented in Section 4. In Section 5, we discuss the results, performance, and comparison with the results of other methods. Finally, the conclusion is given in Section 6.

Data and Material

Types of ROI Segmentation

In our work, to pursue higher accuracy in automatically accomplishing ROI segmentation in clinical competencies (i.e., diagnosis and therapeutic targeting, anatomy identification) in ultrasound images, for example, thyroid nodules,[3,5,22] brachial plexus nerves,[4,6] breast lesions,[7] hearts and lungs,[23] and so on, we focus on tackling two typical types of tasks, that is, thyroid nodule and brachial plexus nerve automatic segmentation.

It is well known that a thyroid nodule, a common clinical ailment, is defined as a lump inside the thyroid gland and occurs at high frequencies in the adult population. The incidence of thyroid cancer accounted for approximately 567,233 cases worldwide in 2019, ranking ninth in incidence.[24] In clinical practice, unnecessary surgical therapy due to undetermined thyroid nodules leads to the waste of medical resources and increases patient suffering. However, the application of CAD for the automatic segmentation of thyroid nodules can not only help doctors find the specific location of nodules but also provide the margin, shape and aspect ratio of nodules which are important for the subsequent diagnosis of thyroid nodules. Hence, driven by clinical needs, automatic segmentation should be developed to ensure subsequent classification and treatment.

The brachial plexus nerve, on the other hand, is the anterior branch of the fifth cervical spine (C5) to the first thoracic spinal nerve (T1) and is responsible for controlling the sensation and movement of the upper limbs, shoulders, back and chest. To supply anesthetic in the right place, in other words, to correctly locate the nerve structures, one of the effective measures is ultrasound imaging, which enables a noninvasive visualization of the nerve and peripheral structure, as well as reducing the risk of block failure, nerve trauma and local anesthesia toxicity. Furthermore, preoperative local anesthesia is guaranteed by brachial plexus nerve segmentation in ultrasound images, which mainly aims to minimize surgical damage, decrease patient suffering, and accelerate postoperative recovery. Therefore, developing and improving methods for automatically segmenting ultrasound images of nerves has become a research hot spot

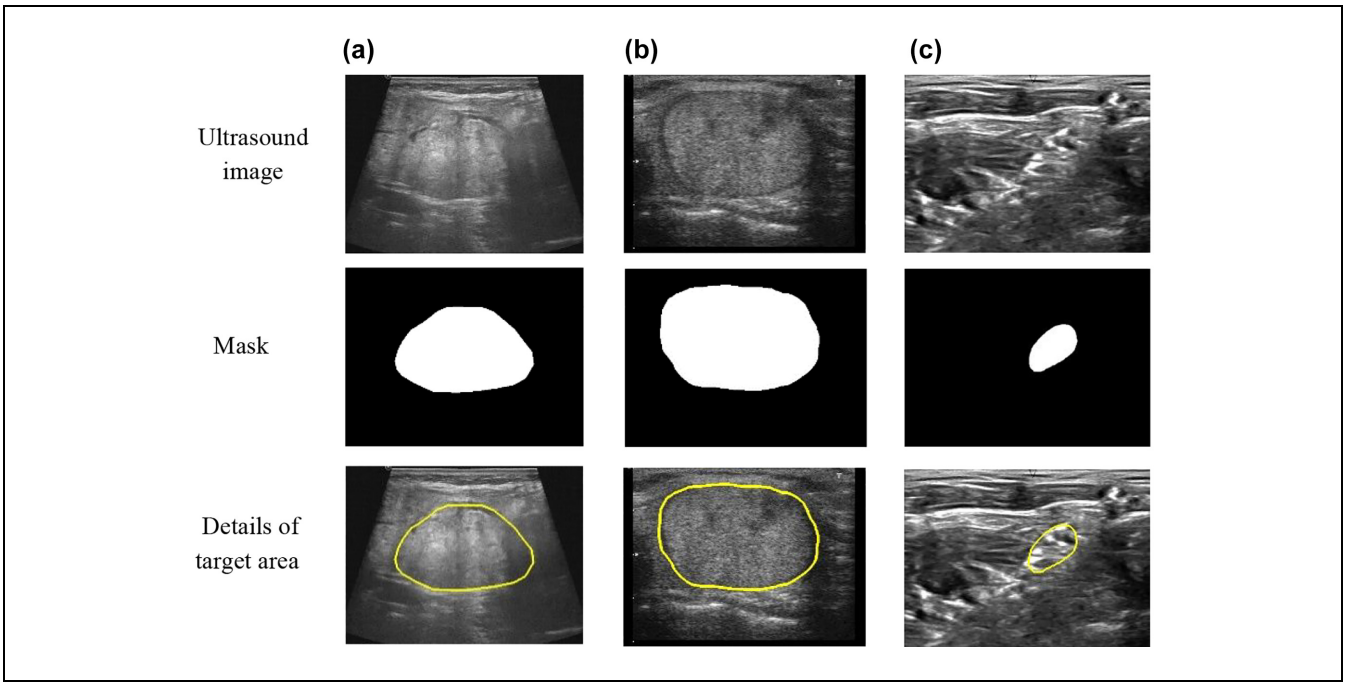

Figure 1. ROI segmentation samples. In the mask image, the white area represents the target area, and the black area represents the background area: (a) TND-PUH3, (b) DDTI, and (c) NSD.

to correctly place the needle when employing regional anesthesia or intraoperative navigation in tumor cutting.

In summary, the acquisition of reliable ROIs in ultrasound images can ensure better quality to meet the requirements of subsequent clinical applications. Specifically, ultrasound images of nodules and nerves not only have their own clinical application function but also reflect two representative types of ultrasound imaging patterns: nodule images are blurry and low contrast, whereas nerve images are relatively more textural and have higher contrast. Therefore, thyroid nodules and brachial plexus nerves were chosen as the main types of ROI segmentation in our study.

Image Acquisition and Simple Pre-processing

In this research, we utilize three datasets for ROI segmentation in medical ultrasound images, including two thyroid nodule datasets and a brachial plexus nerve dataset.

For the thyroid nodule datasets, the first dataset is denoted as TND-PUH3 (Thyroid Nodule Database-Peking University Third Hospital, TND-PUH3). All images were acquired using an ultrasound machine (Phillips, HITACHI, GE) with the probe frequency set as 5 to 17 MHz and labeled by the sonographers of Peking University Third Hospital. A total of 3771 ultrasound images were gathered from 2360 patients after surgery or fine needle aspiration (FNA), among which 1316 are benign and 2455 were malignant. The data set contains 1 to 2 ultrasound images per patient, with the probe in either transverse or longitudinal directions. It contained 639 male cases and 1721 female cases, with an overall average age of 38.47 years. In terms of different age groups, 563 cases were less than or equal to 30 years old, 944 cases were between 30 and 50 years old, 746 cases were between 50 and 70 years old, and 107 cases were older than 70 years old. All the thyroid instances involved in this dataset were examined via pathological examination and covered nodules of different sizes. The second dataset is the public database DDTI (Digital Database Thyroid Image, DDTI) proposed by Lina Pedraza et al. of the Faculty of Medicine of the National University of Colombia in Bogotá.[25] The database contains 480 ultrasound images of 299 patients, and each ultrasound image is annotated by the operator. The annotation file is saved as an .xml file, which contains the outline of the nodule, the TI-RADS information, and so on. A total of 463 ultrasound images with corresponding nodule contour annotations were employed in our ROI segmentation research. As shown in Figure 1(a) and (b), the ultrasound image we utilized in segmentation experiments is presented above, and the nodule contour labeling is displayed below. It is necessary to note that before performing the segmentation experiments of thyroid nodules, we cropped nonessential information from the original ultrasound images, including device name, acquisition time, image source, and so on.

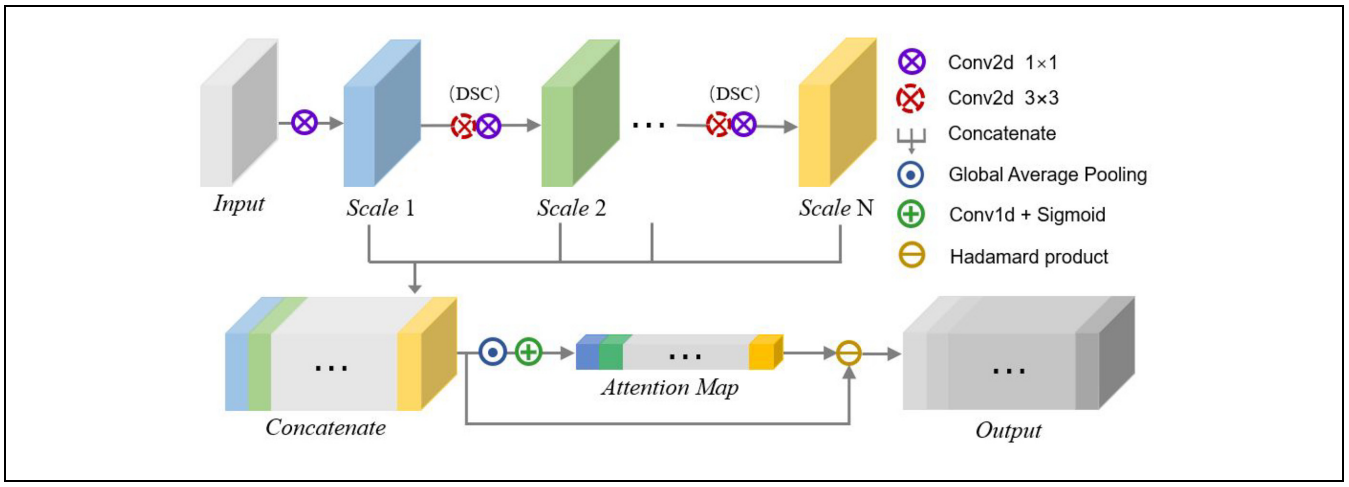

Figure 2. The architecture of multiscale attentional convolution (MSAC).

Meanwhile, the ultrasound brachial plexus nerve segmentate on public dataset NSD (Nerve Segment Database, NSD) from Kaggle Competition is employed in our research. The NSD database includes 5633 samples, each of which contains an original ultrasound image and its corresponding mask image. It is worth emphasizing that only 2322 samples contain brachial plexus nerves in the NSD. As illustrated in Figure 1(c), the top image is the original ultrasound image, while the bottom image is the corresponding mask image. It is important to note that none of the data sets used in this study violated Health Insurance Portability and Accountability Act (HIPAA) and Institutional Review Board (IRB) requirements.

Proposed Method

As is well known, image segmentation is the technology and process of dividing an image into several specific regions with unique properties to propose the object of interest. In segmentation, feature extraction is a special dimensionality reduction process, the main purpose of which is to obtain the relevant information from the low-dimensional spatial information expression of the original data.

Multiscale Attentional Convolution

Combined with clinical knowledge, multiple receptive field feature fusion is essential for medical ROI segmentation, but the complexity and the overfitting risk of CNN will increase if multiple filters with different receptive fields are combined straightforwardly. The multiscale attentional convolution (MSAC) module, as an innovative algorithm, is proposed in our work to efficiently handle the issue of spatial multiscale features combinations, mainly consisting of cascaded convolution operations and a self-attention approach of feature map channels, as illustrated in Figure 2.

MSAC utilizes cascade convolutions to generate features with multiple receptive field scales. $C^{(i)}(\cdot)$ represents the i-th convolution operation. If the input of the i-th convolution operation is x, its output is expressed as:

$$C^{(i)}(x) = W_i x \tag{1}$$

where W_i represents the weights of the convolution operation, and $C^{(1)}$, $C^{(2)}$, $\cdots$, and $C^{(n)}$ indicate the n-th convolution operation. The feature map of the n-th convolution operation is denoted as $I^{(n)}$:

$$I^{(n)}(x) = C^{(n)} \circ C^{(n-1)} \circ \cdots \circ C^{(1)}(x) = \left(\prod_{i=1}^{n} W_i \right) x \tag{2}$$

where $\left(\prod_{i=1}^{n} W_i \right)$ is the convolution weight expressed as a decomposition. It is essential to emphasize that $C^{(i)}$ is a linear layer and that there is no nonlinear activation or normalization operation. The feature map $I^{(n)}$ can reflect the features of the receptive field at various scales. Assume that the spatial dimension of the i-th convolution kernel is $S_i \times S_i$, where S_i is the height and width of convolution kernel W_i. $I^{(n)}(x)$ is the feature map obtained from the cascade convolution operation, and its size of the effective receptive field is expressed as:

$$\left(\sum_{i=1}^{n} S_i - n + 1 \right) \times \left(\sum_{i=1}^{n} S_i - n + 1 \right) \tag{3}$$

where $1 \leqslant n \leqslant N$, N represents the number of cascade convolutions performed by the MSAC module, that is, the number of various scales of receptive field features merged.

The feature maps $I^{(i)}(\cdot)$ are concatenated at the channel axis to form the cascaded feature map Z:

$$Z = [I^{(1)}, I^{(2)}, \cdots, I^{(N)}] \tag{4}$$

To make $I^{(i)}(\cdot)$ have the same spatial size, all the above convolution operations require appropriate padding operations. The height and width of $I^{(i)}(\cdot)$ are expressed by H and W, respectively, and the number of channels for each scale is denoted as P. Therefore, the spatial size of $I^{(i)}(\cdot)$ is $H \times W \times P$, and the spatial size of cascaded feature map Z is $H \times W \times NP$. Assuming $Q = NP$, then the channel dimension of Z can be represented by Q. Although the cascaded feature map Z reflects features from multiple scales and channels, if Z is employed directly as the ultimate convolution operation output, it might increase the risk of overfitting. Our proposed MSAC can effectively integrate features from N different receptive field scales by utilizing a channel self-attention approach, imitating the process of sonographers to obtain ROI. In clinical practice, the sonographers will weigh the overall ROI features as well as detailed features such as edges, boundaries and texture features to identify the location of ROI based on their clinical experience.[26] In our MSAC module, the channel self-attention approach can be applied to the cascaded feature map Z as follows:

$$Y = E(Z) \otimes Z \tag{5}$$

where E is the attention map, and $\otimes$ represents the Hadamard product. The spatial dimension of Y is the same as that of Z, which is $H \times W \times Q$. For the attention map $E(\cdot)$, the ECA (Efficient Channel Attention, ECA) method is deployed,[27] which is defined as:

$$E(Z) = \sigma(C1D_k(G(Z))) \tag{6}$$

where $C1D_k$ represents a one-dimensional convolution with kernel size k, σ is the sigmoid function, and $G(Z)$ represents the global average pooling operation of the cascaded feature map Z in the channel axis.

$$G(Z) = \frac{1}{WH} \sum_{i=1,j=1}^{W,H} Z_{ij} \tag{7}$$

ECA is suitable for MSAC because it focuses on module efficiency while also being able to capture information across channels to ensure performance. The coefficients of the one-dimensional convolutional kernel are the learnable parameters for the channel attention weights.[27] ECA is also easier to tune and apply, and the only hyperparameter is the one-dimensional convolutional kernel size, which can be altered adaptively by the number of channels in the CNN. It is worth pointing out that the MSAC is a separate linear convolutional layer, so both normalization and nonlinear activation blocks can be inserted after it. It is because of its universality that we have the idea of integrating it into typical segmentation networks to optimize the segmentation performance. In summary, our MSAC module addresses the problem of inadequate extraction and utilization of multi-scale features by typical CNNs by performing cascaded convolutional operations to obtain features with various receptive fields and deploying a self-attentive approach to efficiently merge these spatial multi-scale features.

MSAC-Unet

As mentioned before, the encoder-decoder architecture of Unet consists of a downsampling part and an upsampling part, allowing for end-to-end training with a small number of images. Furthermore, Unet effectively integrates deep and shallow features through skip connections, and the architecture is stable, making it ideal for medical ROI segmentation where the number of samples is insufficient. However, Unet only applies the same size filters in each encoder and decoder, limiting the potential to extract multiscale information. Consequently, MSAC-Unet is proposed here to execute automatic medical ROI segmentation, which both retains the great basic architecture of Unet and substitutes traditional convolution with MSAC module in all encoders and decoders to achieve better multiscale feature fusion. The architecture of MSAC-Unet is shown in Figure 3, where the MSAC module applied to Unet architecture is illustrated. Compared with the original Unet, our proposed MSAC-Unet replaces the standard convolutional layers and batch normalization (BN) layers are inserted in each encoder and decoder.

In MSAC-Unet, the convolution operation $C^{(i)}(\cdot)$ is set as follows. For the first convolution operation $C^{(1)}$, we utilize a 1×1 convolution to reduce the computational cost. For the remaining convolution operations $C^{(i)}$, we employ a 3×3 convolution. The MSAC module utilizing cascade convolutions to generate multiscale features necessitates more convolutional operations, and it is employed in each encoder and decoder, as shown in Figure 3. Thus, the computing load will greatly increase if standard convolutional operations are employed. Instead of the standard ones, the depthwise separable convolution (DSC)[28] is adopted to reduce the number of parameters, which dramatically improves the efficiency of the MSAC module. DSC splits the standard convolution operation into depthwise convolution and pointwise convolution, and the parameters are approximately one-third of the standard convolution. Thus, the 3×3 convolution of $C^{(i)}$ includes depthwise convolution C_d and pointwise convolution C_b.

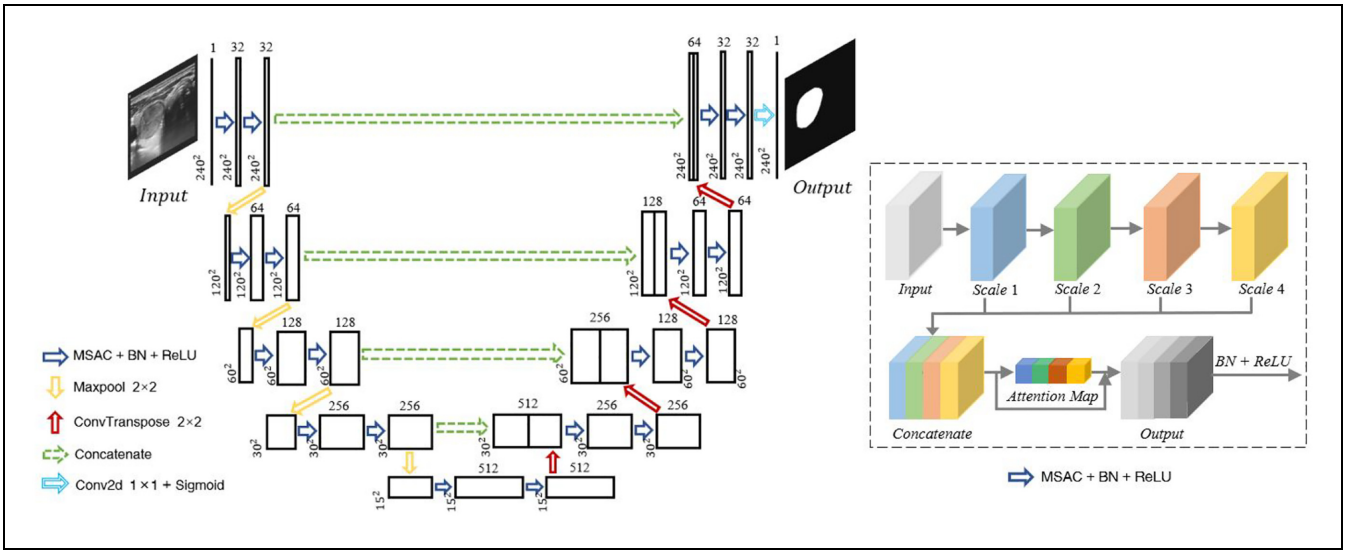

Figure 3. The architecture of MSAC-Unet.

$$C^{(i)}(\cdot) = \begin{cases} C_b(i) & i = 1 \\ C_b \circ C_d(i) & i = 2, 3, \cdots, N \end{cases} \tag{8}$$

where C_b represents a 1×1 convolution and C_d represents a 3×3 convolution. N refers to the number of scales in the MSAC module. Therefore, the height and width S_i of the convolution kernel W_i are given as:

$$S_i = \begin{cases} 1 & i = 1 \\ 3 & i = 2, 3, \cdots, N \end{cases} \tag{9}$$

Through equations (3) and (9), the receptive field sizes at different scales can be determined as 1×1, 3×3, and $(2N\text{-}1) \times (2N\text{-}1)$, respectively. Combined with the resolution of the ultrasound image utilized for segmentation, in our proposed MSAC-Unet, we set the scale number N to 4.

The number of channels in MSAC-Unet follows the design of Unet, with cascaded feature maps of 32, 64, 128, 256, and 512. Therefore, the number of output channels of the MSAC module is $Q \in \{32, 64, 128, 256, 512\}$. The size of the one-dimensional convolution kernel in ECA, that is, the hyperparameter k, can be adaptively determined as a nonlinear function related to the number of channels,[27] which is defined as:

$$k = \left| \frac{log_2(Q)}{\gamma} + \frac{b}{2} \right|_{odd} \tag{10}$$

where Q is the number of input feature map channels and $|a|_{odd}$ represents the odd number closest to a. Meanwhile, γ and b are set to 2 and 1, respectively.[27] According to the equation (10), k corresponding to the number of channels $Q \in \{32, 64, 128, 256, 512\}$ is given as:

Table 1. Parameter Count and Computational Costs.

Architecture	Parameters	FLOPs
Unet	7,759,521	84.69G
Res-Unet	8,214,881	89.17G
Unet++	9,041,601	209.36G
Deeplab V3+	2,752,881	9.87G
MSAC-Unet	1,091,197	8.70G
MSAC-Res-Unet	1,534,781	13.05G

$$k = \begin{cases} 3 & Q = 32, 64, 128 \\ 5 & Q = 256, 512 \end{cases} \tag{11}$$

Through a self-attention approach, the MSAC effectively captures information from multiple scales of the receptive fields and combines these multiscale features. MSAC is adopted to substitute the standard convolution in the Unet architecture, forming MSAC-Unet with parameters less than 15% of that of the original Unet, and this is the result of MSAC adopting DSC. Table 1 shows the parameter count and computational costs for the CNNs we employed in this study. To assess effectiveness of our MSAC module and the superiority of the lightweight CNN MSAC-Unet, we compare MSAC-Unet model with current representative CNNs, as well as MSAC-Res-Unet.

Evaluation Indicator

In this study, we perform 10-fold cross-validation on each dataset. Each dataset is separated into nonoverlapping parts, with a 9:1 ratio between the training and validation sets and the test set being the remaining portion

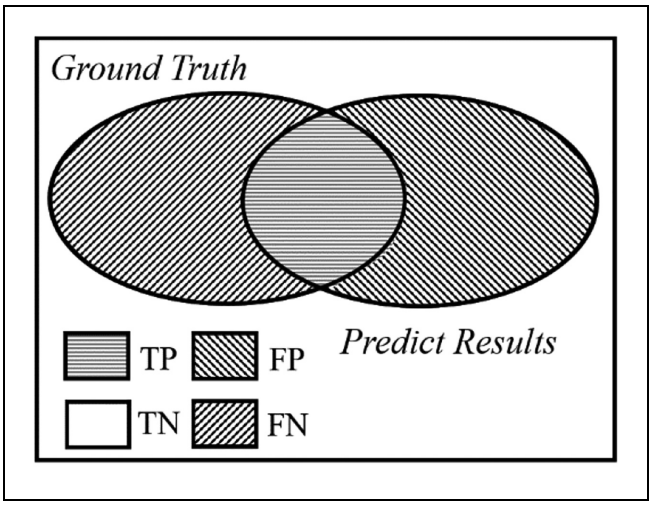

Figure 4. The fundamental definitions of evaluation metrics.

of the dataset. There are 500, 50, and 909 ultrasound images utilized to assess the performance of CNNs on the test sets of TND-PUH3, DDTI, and NSD, respectively. The test set in each database remains constant to compare the segmentation results of different CNNs for the same ROI. Moreover, different metrics should be employed to analyze the performance of the CNN. Thus, we adopt the Dice coefficient, sensitivity and IoU to assess segmentation performance.

$$Dice = \frac{2TP}{2TP + FP + FN} \tag{12}$$

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

$$IoU = \frac{TP}{TP + FP + FN} \tag{14}$$

The Dice coefficient is a function of the similarity measure, which is usually employed to calculate the similarity of two samples. The values of the above evaluation indicators are between 0 and 1. As shown in Figure 4, TP is the true positive area, TN is the true negative area, FP is the false-positive area, and FN is the false negative area.

In our study, the loss function adopted in the model training process is *Dice loss*, which is defined as:

$$Dice\ loss = 1 - \frac{2|\hat{y} \cap y| + 1}{|\hat{y}| + |y| + 1} \tag{15}$$

where $\hat{y}$ is the output of the CNN and y is the ground truth. The "+1" in the equation prevents the denominator from being zero. The *Dice loss* function has great performance when positive and negative samples are unbalanced.

Experimental Results

Implementation Details

The experiments based on deep learning are accomplished on the Ubuntu 18.04.1 system with an NVIDIA GeForce 3090 GPU, and the deep learning framework is TensorFlow 2.4. The size of the input image in CNNs is $240 \times 240 \times 1$, and all input images are pre-processed into data with a mean of 0 and variance of 1. During the training of CNNs, the batch size is set to 4, the number of epochs is 150, the optimizer is Adam, and the learning rate is 5×10^{-4}, learning decay rate is 0.8, learning decay step is 10. The model of the validation set with the greatest Dice coefficient is saved to evaluate the performance of the CNN.

Segmentation of Thyroid Nodules

In this part, we focus on evaluating the accuracy of our proposed MSAC-Unet for thyroid nodule segmentation. Therefore, we perform comparative experiments of MSAC-Unet with Unet, ResUnet, Unet++, Deeplab V3+, and MSAC-Res-Unet on the TND-PUH3 and DDTI datasets. The Dice coefficients, sensitivities, and IoUs of these CNNs, with 95% confidence intervals, are reported in Tables 2 and 3. The Dice coefficient of MSAC-Unet is 3.9% higher than that of Unet on the TND-PUH3 test set; on the DDTI test set, the Dice coefficient of MSAC-Unet is 4.4% higher than that of Unet, confirming the effectiveness of MSAC. Compared with other CNNs, MSAC-Unet achieves the best segmentation performance on both the TND-PUH3 and DDTI datasets.

As illustrated in Figures 5(a) and 6(a), the MSAC-Unet converges at a faster rate during the training process. Although the dice coefficient of its training set is lower than other CNNs, it still exceeds 0.9, which is approximately 0.05 lower than others. In addition, the MSAC-Unet can achieve the lowest loss on the validation set, as low as 0.183 in TND-PUH3 and 0.259 in DDTI. Unet, Unet++, Res-Unet, and Deeplab V3+ have higher Dice coefficients during the training process, while the overfitting of these CNNs is more serious. Owing to the limited number of samples in the medical image database, the performance of these CNNs in both validation and test sets is inferior to that of MSAC-Unet, which has fewer parameters.

Figure 7 displays the segmentation results of thyroid nodules on the test set, and these segmentation results are obtained from the same fold in 10-fold cross-validation. In the TND-PUH3 database, there is only one thyroid nodule in each ultrasound image. However, in the DDTI database, some of the ultrasound images are stitched from different angles of the nodule images so

Table 2. Performance and Corresponding 95% Confidence Intervals (CIs) of Different Models Used in Our Study on the TND-PUH3 dataset.

Dataset	TND-PUH3					
Metric / model	Dice		Sensitivity		IoU	
	Mean (S.D.)	CI	Mean (S.D.)	CI	Mean (S.D.)	CI
Unet	0.783 (0.010)	[0.775, 0.790]	0.819 (0.006)	[0.814, 0.823]	0.682 (0.009)	[0.675, 0.689]
Res-Unet	0.775 (0.007)	[0.769, 0.780]	0.813 (0.015)	[0.802, 0.823]	0.674 (0.008)	[0.669, 0.680]
Unet + +	0.760 (0.010)	[0.753, 0.767]	0.800 (0.011)	[0.792, 0.808]	0.657 (0.010)	[0.649, 0.664]
Deeplab V3 +	0.761 (0.008)	[0.754, 0.766]	0.803 (0.015)	[0.792, 0.814]	0.653 (0.009)	[0.646, 0.659]
MSAC-Unet	0.822 (0.005)	[0.818, 0.826]	0.847 (0.012)	[0.839, 0.856]	0.718 (0.007)	[0.713, 0.723]
MSAC-Res-Unet	0.796 (0.011)	[0.788, 0.804]	0.832 (0.010)	[0.825, 0.840]	0.700 (0.011)	[0.692, 0.708]

Number Format of Performance: Mean (Standard Deviation, S.D.).

Table 3. Performance and Corresponding 95% Confidence Intervals (CIs) of Different Models Used in Our Study on the DDTI Dataset.

Dataset	DDTI					
Metric / model	Dice		Sensitivity		IoU	
	Mean (S.D.)	CI	Mean (S.D.)	CI	Mean (S.D.)	CI
Unet	0.748 (0.008)	[0.742, 0.754]	0.759 (0.014)	[0.749, 0.768]	0.637 (0.013)	[0.628, 0.646]
Res-Unet	0.755 (0.007)	[0.751, 0.760]	0.752 (0.017)	[0.740, 0.764]	0.629 (0.008)	[0.623, 0.635]
Unet + +	0.732 (0.010)	[0.725, 0.739]	0.746 (0.015)	[0.736, 0.757]	0.607 (0.017)	[0.589, 0.620]
Deeplab V3 +	0.743 (0.010)	[0.735, 0.751]	0.751 (0.011)	[0.743, 0.759]	0.613 (0.010)	[0.606, 0.619]
MSAC-Unet	0.792 (0.008)	[0.787, 0.798]	0.826 (0.014)	[0.816, 0.837]	0.673 (0.013)	[0.663, 0.682]
MSAC-Res-Unet	0.769 (0.010)	[0.762, 0.776]	0.777 (0.013)	[0.768, 0.786]	0.640 (0.010)	[0.630, 0.649]

Number format of performance: Mean (Standard Deviation, S.D.).

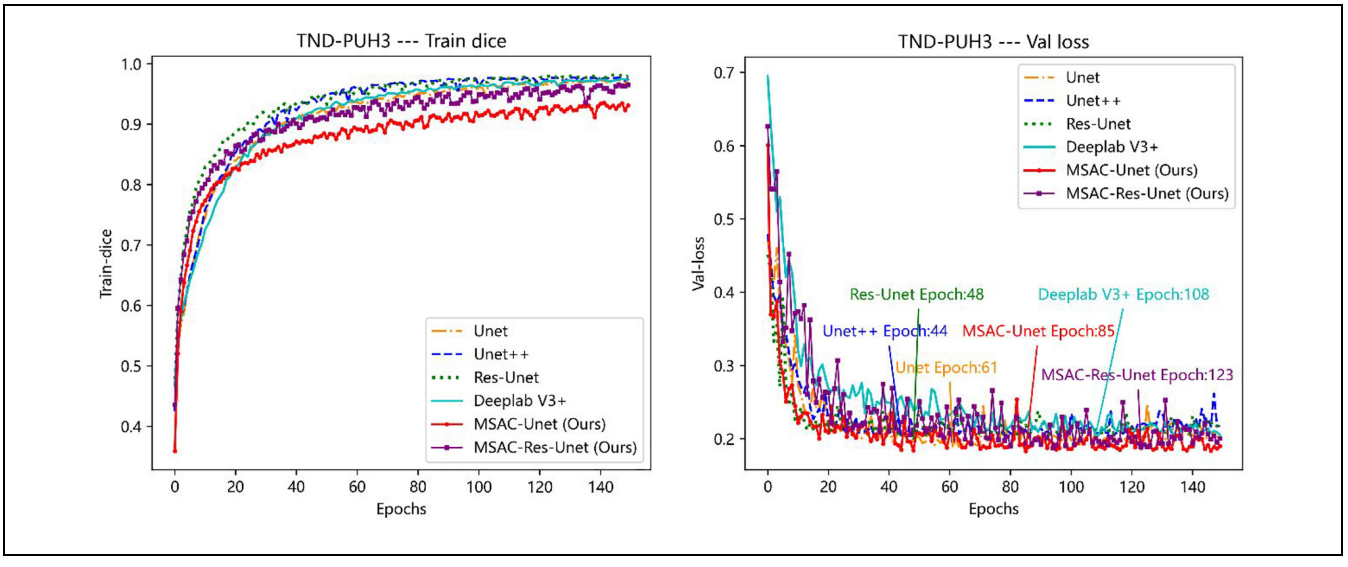

Figure 5. The training dice curve and val loss curve on the TND-PUH3 dataset. On the left: (a) Training dice curve, On the right: (b) Val loss curve. Both the train dice curve and the validation loss curve are for the same training and validation sets. In addition, the epoch of each network model used for testing is indicate in the validation loss curve.

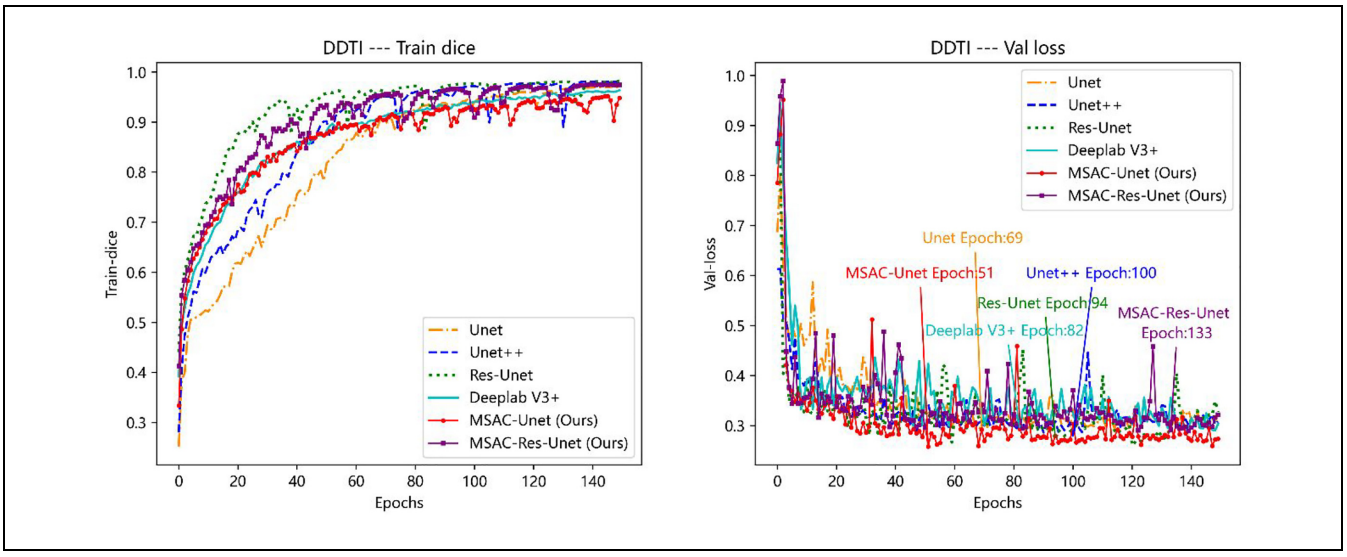

Figure 6. The training dice curve and val loss curve on the DDTI dataset. Both the train dice curve and the validation loss curve are for the same training and validation sets. On the left: (a) Training dice curve, On the right: (b) Val loss curve. Both the train dice curve and the validation loss curve are for the same training and validation sets. In addition, the epoch of each network model used for testing is indicate in the validation loss curve.

that two nodules appear in one image. As shown in Figure 7, the segmentation results of MSAC-Unet are good regardless of the presence of several nodules in the ultrasound images. For thyroid nodules with ambiguous boundaries, there are many isolated nonconnected regions in the segmentation results of Unet, Res-Unet, Unet++, and Deeplab V3+. Replacing the standard convolution with MSAC in CNNs will effectively ameliorate this problem. MSAC-Unet employs cascaded convolution in each codec and decoder to capture features with various receptive field sizes, whereas the original Unet only deploys 3×3 convolution kernels, implying that the extracted receptive field size of the convolution operation is 3×3 in each codec and decoder. In comparison to the features extracted by Unet, MSAC-Unet concatenates individual features and incorporates information from multiple scales, resulting in better edge and boundary features. In addition, although Deeplab *V3+* combines features of different receptive fields by ASPP, the weight set for different features is the same when realizing feature fusion In contrast, our MSAC-Unet utilizes the ECA self-attentiveness mechanism in the MSAC module, which can adaptively assign different weights to the features of different receptive fields, achieving better ROI segmentation performance. The segmentation dice coefficients of Deeplab V3+ is 6.1% and 4.9% lower than our MSAC-Unet on the TND-PUH3 and DDTI datasets, respectively.

As shown in Table 4, we compared our models with other methods by investigating several typical studies of ultrasound thyroid nodules segmentation. On the open dataset DDTI, Nguyen et al. (2022)[29] employed Unet with attention module to achieve thyroid nodule segmentation, but this model only applied attention module in the encoders. However, our proposed MSAC adopted attention approach in each encoder and decoder, and the MSAC-Unet obtained higher Dice coefficient. In addition, compared with the marker-guided Unet (MGU-net) which was proposed by Chu et al. (2021),[16] although the segmentation performance was improved by MGU-net, manually labeling four points for each ultrasound image before the training process is a labor-intensive and time-consuming process that requires expert involvement. Analogously, despite traditional CNNs with fully connected layers (2017)[5] have high segmentation performance in thyroid nodule segmentation, these methods usually feed blocks of pixels into a CNN through a sliding window and employ fully connected layers for segmentation, resulting in serious repeated calculations and low computational efficiency. Meanwhile, each pixel can only be classified by local features in these medthods, which leads to insufficient reliability of the segmentation results. In conclusion, the proposed MSAC-Unet can better balance the segmentation efficiency and accuracy, which has the value of practical clinical application.

Segmentation of Brachial Plexus Nerve

The aim of this part is to assess the efficacy of our proposed MSAC-Unet for brachial plexus nerve segmentation in ultrasound images. In addition to comparative experiments on the NSD dataset, we also evaluate the

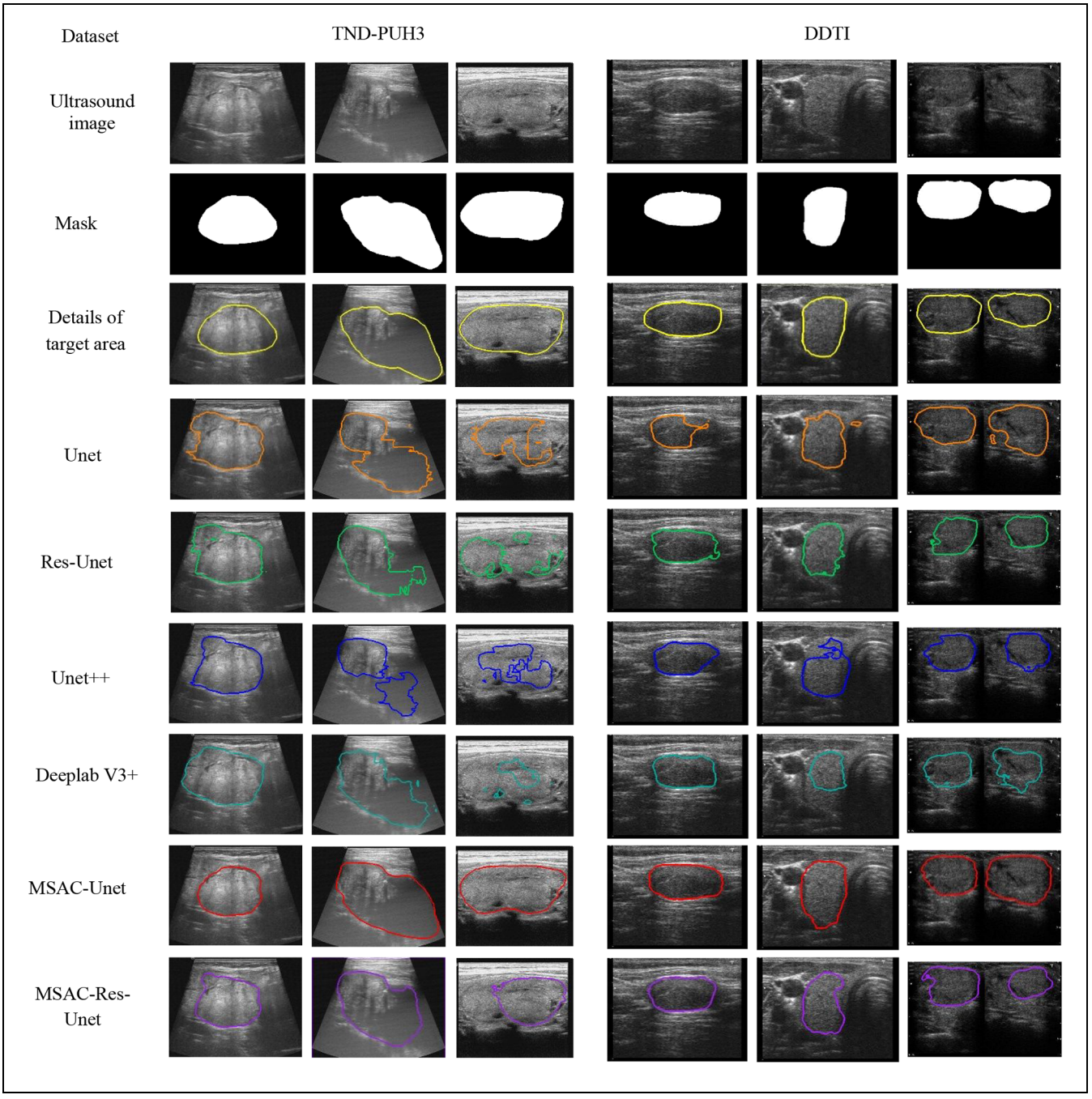

Figure 7. The segmentation results of thyroid nodules with different CNNs. From top to bottom: the original images, ground-truth masks, details of target area, and the segmentation results of Unet, Res-Unet, Unet + +, Deeplab V3 + , MSAC-Unet, MSAC-Res-Unet.

neural segmentation performance of MSAC-Unet with other methods in public research.

The evaluation metrics of different CNNs on the NSD test set, with 95% confidence intervals, are shown in Table 5. On the NSD test set, the Dice coefficient of MSAC-Unet is 2.7% higher than that of Unet. Compared with other models, similarly, the MSAC-Unet achieves the best segmentation performance.

Figure 8 shows the segmentation results of the brachial plexus nerve with different CNNs in test set. Although brachial plexus nerve ultrasound images have higher contrast and more textural features than thyroid nodule ultrasound images, there are 3311 ultrasound images excluding the brachial plexus in the NSD dataset with a total of 5633 samples. As a result, accurate segmentation of brachial plexus nerves with the NSD

Table 4. The Dice Coefficient of Different Models in the Segmentation of Thyroid Nodules.

Model	Database	Dice
Unet (ours)	DDTI	0.748
MSAC-Unet (ours)	DDTI	0.792
Unet[29]	DDTI	0.555
Unet + Attention block[29]	DDTI	0.596
DenseNet 161[16]	Non-public	0.917
MGU-net[16]	Non-public	0.958
DT algorithm[12]	Non-public	0.975
Cascade CNN[5]	Non-public	0.985

dataset is relatively more challenging than with thyroid nodules in our study. In the NSD test set, there were 545 ultrasound images without brachial plexus nerves, and the CNNs we utilized erroneously segment these ultrasound images to various extents. Summarizing the segmentation results, Unet, Res-Unet, Unet++, Deeplab V3+, MSAC-Unet, and MSAC-Res-Unet segmented on average 127, 103, 116, 109, 58, and 67 erroneous ROIs, respectively. As seen from Figure 8 and Table 5, MSAC-based CNNs effectively decrease the probability of segmenting incorrect ROIs and improve the segmentation performance since the MSAC module incorporates additional features with various sizes of receptive fields.

As shown in Table 6, we compared our models with other methods by investigating several representative researches of ultrasound nerve segmentation. To the best of our knowledge, the best Dice coefficient achieved by the Unet-related algorithm on the test set is 0.721 in the public report under the NSD database. To address the problem of inconsistent samples, that is, some ultrasound images in the NSD dataset contain no brachial plexus nerves, Van Boxtel et al.[30] used a CNN to classify the ultrasound images that contain brachial plexus nerves before performing ROI segmentation. However, the Dice coefficient obtained by this method is still 2.5% lower than our proposed MSAC-Unet. This demonstrates that our MSAC-Unet is superior for ROI segmentation in medical ultrasound images.

Calculation Speed Comparison

The deployment of the segmentation model to tomography devices is the ultimate goal of implementing ultrasound medical image segmentation, and since real time is crucial, the computational speed is a key metric to assess the effectiveness of the segmentation model. Specifically, 50,500, and 909 ultrasound images were randomly selected from the TND-PUH3, DDTI, and NSD datasets, respectively, as the test sets. The average test time comparison for segmenting one ultrasound image of a thyroid nodule or brachial plexus nerve is shown in Table 7. The testing time of MSAC-Unet is shorter than that of the other models, and segmenting an ultrasound image takes only 2.68 ms on average, which is approximately 10% shorter than that of the Unet. Consequently, the medical ultrasound image segmentation method based on the MSAC-Unet model proposed in this paper has the potential for clinical application, and it can quickly provide an accurate and objective second opinion on ROI segmentation.

Discussion

Though time-consuming and laborious, manual segmentation is often regarded as the gold standard in clinical medicine because of its accuracy, but the quality of the segmentation result completely depends on the experience and knowledge of the operator, thus the segmentation results are difficult to reproduce. In this study, we mainly focused on fully automatic methods for thyroid nodule and brachial plexus nerve ROI segmentation, it is expected to obtain accurate, reproducible and timesaving segmentation results on low contrast, ambiguous boundaries ultrasound medical images. Inspired by the fact that the combination of multiscale information is consistent

Table 5. Performance and Corresponding 95% Confidence Intervals (CIs) of Different Models Used in Our Study on the NSD dataset.

Dataset	NSD					
Metric	Dice		Sensitivity		IoU	
model	Mean (S.D.)	CI	Mean (S.D.)	CI	Mean (S.D.)	CI
Unet	0.719 (0.005)	[0.715, 0.722]	0.796 (0.010)	[0.788, 0.802]	0.673 (0.007)	[0.668, 0.678]
Res-Unet	0.725 (0.008)	[0.720, 0.731]	0.818 (0.009)	[0.812, 0.825]	0.692 (0.010)	[0.684, 0.699]
Unet++	0.703 (0.013)	[0.694, 0.712]	0.762 (0.011)	[0.794, 0.770]	0.682 (0.010)	[0.675, 0.689]
Deeplab V3+	0.718 (0.008)	[0.713, 0.725]	0.812 (0.011)	[0.803, 0.820]	0.688 (0.008)	[0.683, 0.694]
MSAC-Unet	0.746 (0.007)	[0.741, 0.751]	0.843 (0.008)	[0.837, 0.849]	0.714 (0.006)	[0.709, 0.718]
MSAC-Res-Unet	0.734 (0.009)	[0.728, 0.741]	0.830 (0.005)	[0.826, 0.833]	0.702 (0.009)	[0.696, 0.709]

Number format of performance: Mean (Standard Deviation, S.D.).

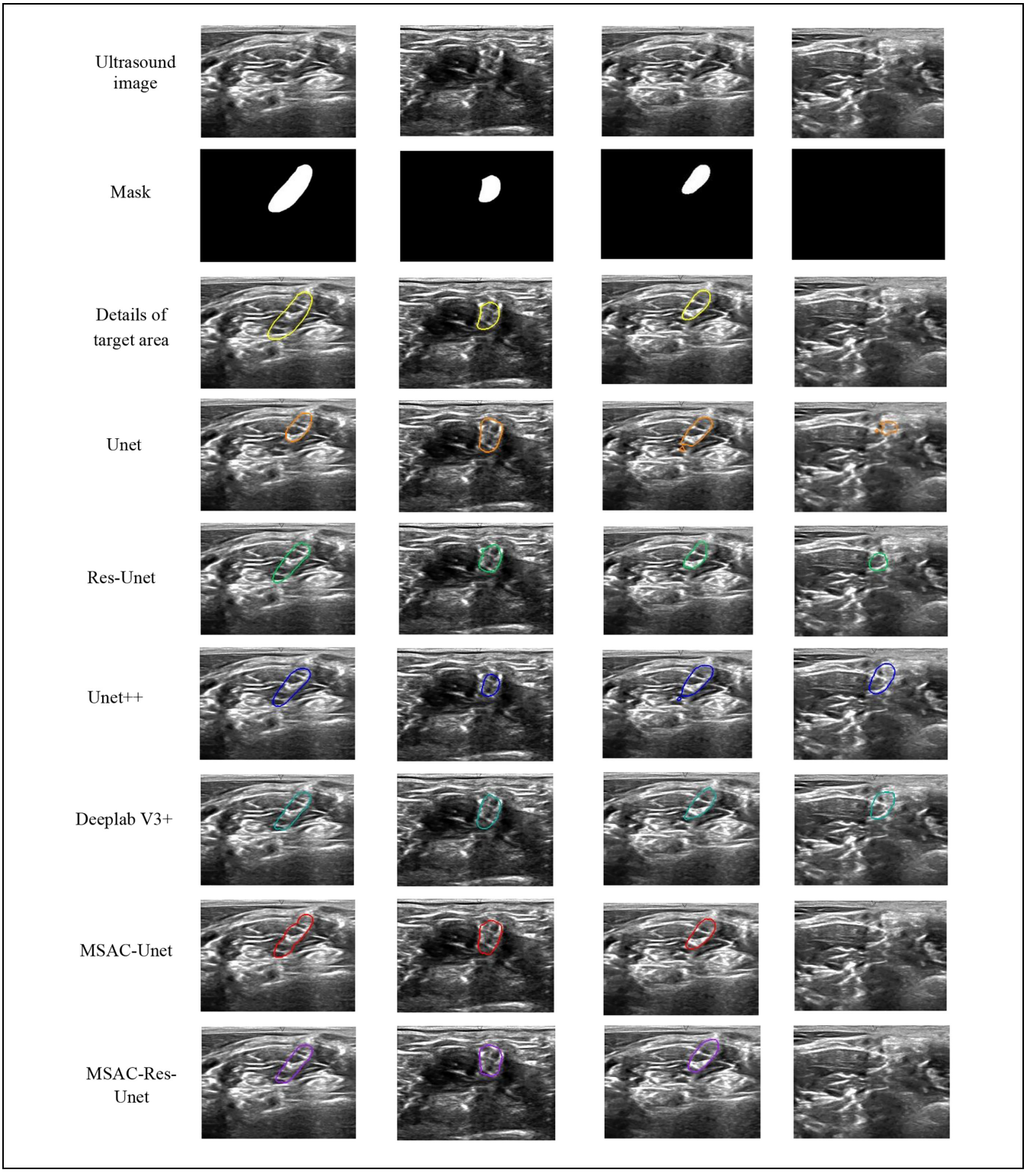

Figure 8. The segmentation results of brachial plexus nerves with different CNNs. From top to bottom: the original images, ground-truth masks, details of target area, and the segmentation results of Unet, Res-Unet, Unet + +, Deeplab V3 +, MSAC-Unet, MSAC-Res-Unet.

Table 6. The Dice Coefficient of Different Models in the Segmentation of Nerves.

Model	Database	Dice
Unet (ours)	NSD	0.719
MSAC-Unet (ours)	NSD	0.746
Graph Cut + Wavelet transform[4]	Non-public	0.606
SLIC + Gaussian process[10]	Non-public	0.652
Unet + PCA[6]	NSD	0.688
Unet without PCA[6]	NSD	0.670
Unet + hybrid model[30]	NSD	0.721

Table 7. The Comparison of Testing Time of Different Models.

Model	Average test time/ms
Unet	2.969
Res-Unet	3.171
Unet++	3.268
Deeplab V3+	2.863
MSAC-Unet	2.680
MSAC-Res-Unet	2.796

with clinical diagnostic experience, we propose the MSAC module, which utilizes cascaded convolutions and self-attention approaches to effectively aggregate spatial receptive field features and apply it to the Unet architecture.

To assess the superiority of the MSAC module, we compared the segmentation performance of Unet with MSAC-Unet, as well as ResUnet with MSAC-Res-Unet. According to Figure 7, MSAC-Unet and MSAC-Res-Unet can obtain better edge and boundary features of ROI in the segmentation task of thyroid nodules. This is certainly because the MSAC module can successfully combine spatial multiscale features. Moreover, we can tell that the ROIs obtained from the MSAC-based CNN are connected regions compared with those of other CNNs. Figure 8 illustrates the segmentation results of the brachial plexus nerve, which demonstrate that the MSAC-based CNNs can effectively decrease the possibility of incorrect ROI segmentation. Moreover, as shown in Figure 9, the segmentation performance of the MSAC-based CNNs is better than that of the original CNN for both segmentation tasks of thyroid nodules and brachial plexus nerves. It is hence not surprising that the MSAC module serves as a universal module that not only improves ROI segmentation performance but can also be applied to any task designed to combine spatial multi-scale information.

To evaluate the effectiveness of MSAC-Unet, we further compare MSAC-Unet with current representative segmentation CNNs, as well as the MSAC-Res-Unet. As shown in Figures 5 and 6, despite having a lower Dice coefficient than other CNNs during the training process, our lightweight MSAC-Unet outperforms other typical CNNs during both validation and testing, demonstrating the superiority of MSAC-Unet in mitigating the risk of overfitting. In addition, the accuracy results shown in Tables 2, 3 and 5 are encouraging. Specifically, the Dice coefficients of MSAC-Unet are 0.822, 0.792 and 0.746 in TND-PUH3, DDTI and NSD, respectively. Although the dilated convolution in DeeplabV3+ is able to acquire features of different receptive fields to some extent, there are some shortcomings in feature fusion. Our MSAC-Unet retains the structure of Unet and replaces the standard convolution with our proposed MSAC module. Using the self-attention mechanism in the MSAC module, our MSAC-Unet can better integrate features at different scales in the feature fusion process and achieve more precisely segmentation. Meanwhile, while MSAC-Unet converges more slowly during training than MSAC-Res-Unet (as shown in Figures 5 and 6), the architecture of MSAC-Unet is less complicated and contains fewer parameters (from Table 1). Therefore, MSAC-Unet can obtain superior segmentation performance compared with MSAC-Res-Unet. Moreover, our

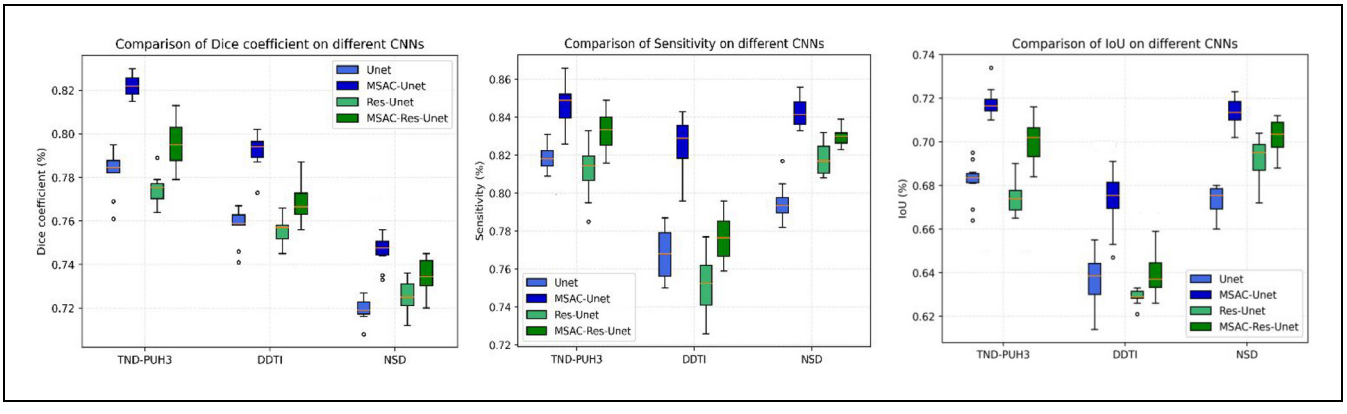

Figure 9. Comparison of segmentation performance of MSAC-based CNNs with original CNNs on the datasets we employed.

MSAC-Unet enables faster automatic segmentation of ROI in ultrasound images, according to Table 7. Furthermore, once our proposed method has been assessed in a larger database, it can be deployed as a diagnostic tool in clinical practice, as the technique can obtain a more correct second objective opinion on ROI segmentation.

Additionally, compared with state-of-the-art segmentation network models in recent years, our MSAC-Unet can also achieve better comprehensive segmentation performance. Chen et al.,[16] for instance, proposed the marker-guided Unet model for the segmentation of thyroid nodules, which achieved great segmentation performance. However, manually labeling four points for each ultrasound image is a labor-intensive and time-consuming process. For brachial plexus nerves segmentation, Van Boxtel et al.[30] used a cascade CNN to classify the ultrasound images containing brachial plexus nerves and then perform ROI segmentation employing Unet. However, the Dice coefficient obtained by this method on the open dataset NSD is still 2.5% lower than that of our proposed MSAC-Unet. In summary, our MSAC-Unet can automatically achieve ROI segmentation without manually labeling points, and the segmentation performance, compared with the reported state-of-the-art segmentation methods of thyroid nodules and the brachial plexus[6,16,30] so far, is effective from the perspective of clinical application.

On the other hand, despite the fact that our proposed tactic of merging multiscale information from a single convolutional module is advantageous to obtain more features from a small number of samples, resulting in better segmentation performance, the ultrasound images used in our experiments were not improved for higher image quality but only cropped to remove extraneous information. However, ultrasound images are affected by speckle noise, which might influence the accuracy of segmentation to some extent. Therefore, further research into ways to pre-process ultrasound images for ROI segmentation to obtain images of greater quality is worthwhile.

Conclusion

Ultrasonography has irreplaceable benefits in clinical practice, but ultrasound images have the drawbacks of low contrast and noise. It is vital to automatically segment the ROI in ultrasound images to assist in subsequent applications. However, the existing typical CNN has limited ability to extract and exploit multiscale features. In our study, we integrated the designed MSAC with Unet to refine the ROI segmentation. In the MSAC module, we effectively combined spatial multiscale features, utilizing cascaded convolution operations to generate features with various receptive field scales, and employing not only an efficient self-attention approach to execute feature fusion but also adopting the depthwise separable convolution (DSC) with few parameters to improve operation efficiency. As a result, a lightweight CNN architecture denoted as MSAC-Unet, which employs MSAC module to replace standard convolution in Unet, can successfully detect richer features and improve the segmentation accuracy. Experiments, for the representative ROI segmentation tasks such as thyroid nodule and brachial plexus nerve in ultrasound images, show that the MSAC-Unet segmentation model not only facilitates the acquisition of more accurate edges and boundaries but also reduces the possibility of segmenting incorrect ROIs. The proposed tactic of automatically and quickly obtaining medical ROI through multiscale information fusion can provide sonographers with a more accurate objective second opinion while also relieving them of heavy workload, reducing the risk of misdiagnosis owing to exhaustion.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported in part by a grant from National Natural Science Foundation of China (61673039).

ORCID iD

Rui Wang https://orcid.org/0000-0002-4813-8514

Data Availability Statement

The DDTI database is available online at doi: 10.1117/12.2073532 (accessed on March 5, 2022) and NSD database is also available online at https://www.kaggle.com/c/ultrasound-nerve-segmentation/data (accessed on March 2, 2022). While the TND-PUH3 database is not shared.

References

1. Wu S, Zhu Q, Xie Y. Evaluation of various speckle reduction filters on medical ultrasound images. In: Proceedings of annual international conference of the IEEE engineering in medicine and biology society, IEEE, 2013, pp. 1148-1151.
2. Wang R, Li R, Lei Y, Zhu Q. Tuning to optimize SVM approach for assisting ovarian cancer diagnosis with photoacoustic imaging. Biomed Mater Eng. 2015;26 Suppl 1:S975.
3. Du W, Nong S. An effective method for ultrasound thyroid nodules segmentation. In: International symposium on bioelectronics and bioinformatics, IEEE, 2015, pp. 207-210.

4. Gonzalez JG, Alvarez MA, Orozco AA. Automatic segmentation of nerve structures in ultrasound images using graph cuts and gaussian processes. In: 37th annual international conference of the IEEE engineering in medicine and biology society, IEEE, 2015, pp. 3089-3092.
5. Ma J, Wu F, Jiang T, Zhu J, Kong D. Cascade convolutional neural networks for automatic detection of thyroid nodules in ultrasound images. Med Phys. 2017;44(5): 1678-91.
6. Kakade A, Dumbali J. Identification of nerve in ultrasound images using U-net architecture. In: International conference on communication information and computing technology, IEEE, 2018, pp. 1-6.
7. Chowdary J, Yogarajah P, Chaurasia P, Guruviah V. A multi-task learning framework for automated segmentation and classification of breast tumors from ultrasound images. Ultrason Imaging. 2022;44(1):3-12.
8. Wang R, Wang L, Yuan Y. Region-based statistical signal processing scheme for image fusion. J Beijing Univ Aeronaut Astronaut. 2010;36(2):140-4.
9. 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.
10. Gonzalez JG, Alvarez MA, Orozco AA. A probabilistic framework based on SLIC-superpixel and Gaussian processes for segmenting nerves in ultrasound images. In: 38th annual international conference of the IEEE engineering in medicine and biology society, IEEE, 2016, pp. 4133-4136.
11. Maroulis DE, Savelonas MA, Iakovidis DK, Karkanis SA, Dimitropoulos N. Variable background active contour model for computer-aided delineation of nodules in thyroid ultrasound images. IEEE Trans Inf Technol Biomed. 2007;11(5):537-43.
12. Chang CY, Huang HC, Chen SJ. Automatic thyroid nodule segmentation and component analysis in ultrasound images. Biomed Eng Appl Basis Commun. 2010;22(02): 81-9.
13. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. IEEE Trans Pattern Anal Mach Intell. 2015;39(4):640-51.
14. Liu S, Wang Y, Yang X, Lei B, Liu L, Li SX, et al. Deep learning in medical ultrasound analysis: A review. Engineering. 2019;5(2):261-75.
15. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention without volume, Springer, 2015,pp. 234–241.
16. Chu C, Zheng J, Zhou Y. Ultrasonic thyroid nodule detection method based on U-Net network. Comput Methods Programs Biomed. 2021;199(1):105906.
17. Drozdzal M, Vorontsov E, Chartrand G, Kadoury S, Pal C. The importance of skip connections in biomedical image segmentation. arXiv preprint arXiv: 1608.04117. 2016.
18. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. UNet++: redesigning skip connections to exploit multiscale features in image segmentation. IEEE Trans Med Imaging. 2020;39(6):1856-67.
19. Huang G, Liu Z, Laurens V, Weinberger K. Densely connected convolutional networks. In: IEEE conference on computer vision and pattern recognition, IEEE, 2017, pp. 2261-2269.
20. Chen LC, Papandreou G, Schroff F, Adam H. Rethinking Atrous Convolution for Semantic Image Segmentation, arXiv preprint arXiv:1706.05587. 2017.
21. Chen LC, Zhu YK, Papandreou G, Schro F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision, IEEE, 2018, pp. 801-818.
22. Keramidas EG, Maroulis D, Iakovidis DK. TND: a thyroid nodule detection system for analysis of ultrasound images and videos. J Med Syst. 2012;36(3):1271-81.
23. Xi J, Chen J, Wang Z, Ta D, Lu B, Deng X, et al. Simultaneous segmentation of fetal hearts and lungs for medical ultrasound images via an efficient multi-scale model integrated with attention mechanism. Ultrason Imaging. 2021;43(6):308-19.
24. Laetitia G, Sven S, Fabrice J. Combinatorial therapies in thyroid cancer: an overview of preclinical and clinical progresses. Cells. 2020;9(4):830.
25. Pedraza L, Vargas C, Narváez F, Durán O, Muoz EM, Romero E. An open access thyroid ultrasound image database. In: 10th international symposium on medical information processing & analysis, SPIE, 2015, p. 9287.
26. Zhou M, Wang R, Fu P, Bai Y, Cui L. Automatic malignant thyroid nodule recognition in ultrasound images based on Deep Learning. E3S Web Conf. 2020;185:03021.
27. Wang Q, Wu B, Zhu P, Li P, Hu Q. ECA-Net: efficient channel attention for deep convolutional neural networks. In: Conference on computer vision and pattern recognition, IEEE, 2020, pp. 11534-11542.
28. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. MobileNets: efficient convolutional neural networks for mobile vision applications, arXiv preprint arXiv: 1704.04861. 2017.
29. Nguyen DT, Choi J, Park KR. Thyroid nodule segmentation in ultrasound image based on information fusion of suggestion and enhancement networks. Mathematics. 2022;10(19):3484.
30. Van Boxtel JPA, Vousten VRJ, Pluim J, Rad NM. Hybrid deep neural network for brachial plexus nerve segmentation in ultrasound images, arXiv preprint arXiv: 2106.00373. 2021.