RESEARCH

An ultrasound image segmentation method for thyroid nodules based on dual-path attention mechanism-enhanced UNet++

Peizhen Dong[1], Ronghua Zhang[1], Jun Li[2], Changzheng Liu[1*], Wen Liu[2], Jiale Hu[1], Yongqiang Yang[1] and Xiang Li[1]

Abstract

Purpose This study aims to design an auxiliary segmentation model for thyroid nodules to increase diagnostic accuracy and efficiency, thereby reducing the workload of medical personnel.

Methods This study proposes a Dual-Path Attention Mechanism (DPAM)-UNet++ model, which can automatically segment thyroid nodules in ultrasound images. Specifically, the model incorporates dual-path attention modules into the skip connections of the UNet++ network to capture global contextual information in feature maps. The model's performance was evaluated using Intersection over Union (IoU), F1_score, accuracy, etc. Additionally, a new integrated loss function was designed for the DPAM-UNet++ network.

Results Comparative experiments with classical segmentation models revealed that the DPAM-UNet++ model achieved an IoU of 0.7451, an F1_score of 0.8310, an accuracy of 0.9718, a precision of 0.8443, a recall of 0.8702, an Area Under Curve (AUC) of 0.9213, and an HD95 of 35.31. Except for the precision metric, this model outperformed the other models on all the indicators and achieved a segmentation effect that was more similar to that of the ground truth labels. Additionally, ablation experiments verified the effectiveness and necessity of the dual-path attention mechanism and the integrated loss function.

Conclusion The segmentation model proposed in this study can effectively capture global contextual information in ultrasound images and accurately identify the locations of nodule areas. The model yields excellent segmentation results, especially for small and multiple nodules. Additionally, the integrated loss function improves the segmentation of nodule edges, enhancing the model's accuracy in segmenting edge details.

Keywords Ultrasound image segmentation, Thyroid nodule segmentation, Dual path attention module, Computer aided diagnosis, Deep learning

Introduction

Thyroid nodules are joint endocrine dysfunctions caused by the localized irregular proliferation of thyroid cells [1]. While most thyroid nodules are benign, some can progress to thyroid cancer. Globally, thyroid cancer ranks as the seventh most common cancer [2], with a notably higher incidence in women. Statistics indicate that the incidence rate of thyroid cancer is 4.6% in men and 13.6% in women, making it three times more prevalent in

*Correspondence:
Changzheng Liu
liucz@shzu.edu.cn
[1] College of Information Science and Technology, Shihezi University, Shihezi 832003, Xinjiang, China
[2] Department of Medical Ultrasound, The First Affiliated Hospital of Medical College, Shihezi University, Shihezi 832003, Xinjiang, China

Dong *et al. BMC Medical Imaging*   (2024) 24:341

Page 2 of 15

women [2]. Furthermore, the incidence of thyroid cancer is increasing annually due to the increasing presence of radioactive substances in the environment and unhealthy dietary habits. Therefore, early detection and treatment of thyroid nodules are essential for reducing the potential risk of cancer and avoiding unnecessary biopsies and surgeries [3].

Ultrasound examination is widely utilized for detecting thyroid diseases because of its simplicity, real-time efficiency, and low cost. In ultrasound images, clinicians diagnose thyroid nodules' benign or malignant nature based on solidity, echogenicity, calcification, shape, and aspect ratio. Ultrasound images typically suffer from low resolution and contrast, often plagued by speckle noise and echo interference [4]. Moreover, the thyroid is surrounded by structures such as the trachea, blood vessels, and muscles, which further blur the boundaries between the thyroid and these adjacent tissues due to the influence of artifacts and speckle [5]. The diverse appearances, shapes, and sizes of thyroid nodules add to the complexity of ultrasound image segmentation [6]. The annotation of thyroid nodules is time-consuming and requires the expertise of experienced clinicians, leading to increased time costs. Consequently, it is crucial to develop an automated segmentation approach that not only improves segmentation accuracy [7] but also significantly lessens the workload of medical staff.

The development of modern image processing technology promotes the application of computer-aided diagnosis in the field of thyroid nodule ultrasound diagnosis, alleviating doctors' workloads and improving diagnostic objectivity [8]. Traditional medical image segmentation methods mostly use image processing techniques such as thresholding, edge detection, and morphology. Senthilkumaran et al. [9] proposed a local adaptive thresholding method to segment medical images. Szenasi et al. [10] employed a distributed region growing method to segment major images in colon tissue samples. Singh et al. [11] cropped the regions of interest from the input color fundus images and subsequently applied a series of steps, including adaptive median filtering, histogram equalization, and morphological processing. Finally, the segmentation of the optic disc and cup was achieved using the maximally stable extremal regions (MSER) method. Deep learning has been widely applied in the field of ultrasound image segmentation due to its significant ability to enhance diagnostic accuracy and efficiency [12–18]. Yang et al. [19] combined the Res-U-net framework with a priori level set methods to increase the accuracy of thyroid tumor segmentation and then trained a classification network by sharing the segmentation network's shallow layers, ultimately achieving the classification of different thyroid tumors. Nie et al. [20] proposed a fully

convolutional neural network that combines a multiscale input layer and an attention guidance module. They used the attention guiding module to filter features, eliminate noise, and reduce the negative impact of the background. Yadav et al. [21] applied a speckle reduction filter to process thyroid ultrasound images, followed by segmentation using Hybrid-UNet. Experimental results demonstrated that this method effectively segmented thyroid nodules and cystic components. Chen et al. [22] used deep convolutional neural networks to encode local view features. They devised unique graph convolution components that dynamically capture the interplay between high- and low-level characteristics by establishing graphs across numerous layers. Furthermore, to enhance feature fusion, they employed a channel-aware graph attention mechanism that efficiently fuses multisource feature information to improve the accuracy of thyroid nodule segmentation.

However, traditional medical image segmentation methods are susceptible to image noise and intensity variations [23], which can result in either over-segmentation or under-segmentation. Some deep learning segmentation techniques focus primarily on local feature extraction, lacking the ability to capture and utilize the contextual information of the entire input image, making it challenging to effectively integrate global context [24]. Additionally, capturing long-range dependencies from an overall image perspective is difficult, leading to the neglect of important dispersed features in medical images.

In response to the above issues, this paper proposes a DPAM-UNet++ network. By incorporating a dual-path attention mechanism into the skip connections of the UNet++ architecture, the model effectively captures critical information from the feature maps, integrates global contextual information, and focuses on the areas where nodules are located while minimizing the influence of image noise and other irrelevant factors. Additionally, an integrated loss function is introduced to enhance the model's accuracy in segmenting nodule edges. In summary, the main contributions of this paper are as follows:

(1) A dual-path attention mechanism has been proposed and integrated into the skip connections of the UNet++ network, enhancing the overall segmentation performance of the model. This dual-path attention mechanism consists of two pathways: the first utilizes a global attention mechanism to capture information from the global context, while the second combines spatial and channel attention to extract both channel and spatial information from the feature maps. The collaboration of these two pathways allows the model to effectively

Dong *et al. BMC Medical Imaging*   (2024) 24:341

Page 3 of 15

focus on critical regions, thereby improving segmentation accuracy.

(2) An integrated loss function has been designed to enhance the effectiveness of the segmentation. This loss function combines multiple loss components, strengthening the model's optimization of edge details in important feature maps by taking similarity into account.

The structure of the remainder of this paper is as follows: the "Materials and methods" section describes the dataset, data processing techniques, and model architecture; the "Results" section presents the experimental findings; the "Discussion" section analyzes these results; and the "Conclusion" section summarizes the main conclusions.

Materials and methods

Dataset

We conducted this research using the TN3K public dataset [4], which includes 3,493 ultrasound images sourced from 2,421 patients. Based on the publicly available division of the TN3K dataset, the training set contains 2,879 ultrasound images, and the test set contains 614 ultrasound images. All data have undergone ethical review.

Experimental setup

The machine specifications used in this experiment are detailed in Table 1. Additionally, the model was trained using PyTorch version 2.0.1 and CUDA 11.2. The versions of the libraries used in this experiment are as follows: Numpy 1.23.5, Torchvision 0.15.2, and OpenCV 4.8.1.78. The learning rate for the model was set to 0.00001, with a batch size of 6. The input size for the network was 512× 512, and the model was trained for 100 epochs. The input convolutional kernel size was 7×7, with a stride of 2. Furthermore, a padding of (3,3) was employed to preserve important edge information. The activation function used was the ReLU function. The RMSprop optimizer was utilized, with a weight decay coefficient of 1e-8 and a momentum coefficient of 0.9. During the training phase, we implemented data augmentation strategies using the Torchvision library to enhance the model's generalization and robustness. Specifically, data augmentation was applied with a 75% probability, including random horizontal flips, random vertical flips, and random combinations of horizontal and vertical flips. The hyperparameters involved in the experiment are listed in Table 2.

Model

As illustrated in the Fig. 1, the DPAM-UNet++ network is composed of four main components: the encoder, decoder, skip connections, and the DPAM module integrated into specific skip connections. The encoder extracts features, the decoder progressively restores the spatial resolution of feature maps, and the skip connections transmit high-resolution features from the encoder to the decoder. The DPAM module further refines the network's capability to capture detailed features by focusing on global contextual information.

The encoder leverages ResNet34 as its backbone (highlighted by the dashed box in the figure). The input image is first processed through a 7×7 convolutional layer with a stride of 2, followed by normalization and ReLU activation, generating the initial feature map x0,0. This feature map undergoes max pooling and is successively passed through the four layers of ResNet34, resulting in downsampled feature maps $x^{1,0}$, $x^{2,0}$, $x^{3,0}$, and $x^{4,0}$, thus completing the downsampling process. Down-sampling is represented by black solid arrows pointing downward in the figure.

Following feature extraction by the encoder, the decoder restores the spatial resolution of the feature maps. This is achieved by upsampling (marked by black solid arrows pointing upward in the figure) the downsampled feature maps using nearest-neighbor interpolation, followed by concatenation with high-resolution feature maps from the encoder through skip connections. The concatenated feature maps are then processed by the

Table 1 System environment and specifications

Environment Details	Specifications
Device Processor	Intel Xeon Silver 4210
Device GPU	NVIDIA Tesla T4
Device Memory	32G DDR4
Device Storage	8T HDD
Operating System	Ubuntu
Coding Language	Python
Libraries Used	PyTorch/Numpy/ OpenCV/Torchvision

Table 2 All hyperparameters of the experiment

Parameter	Value
Learning Rate	0.00001
Batch Size	6
Epochs	100
Optimizer	RMSprop
Weight Decay	1e-8
Momentum	0.9
Input Convolution Kernel Size	7x7
Stride	2

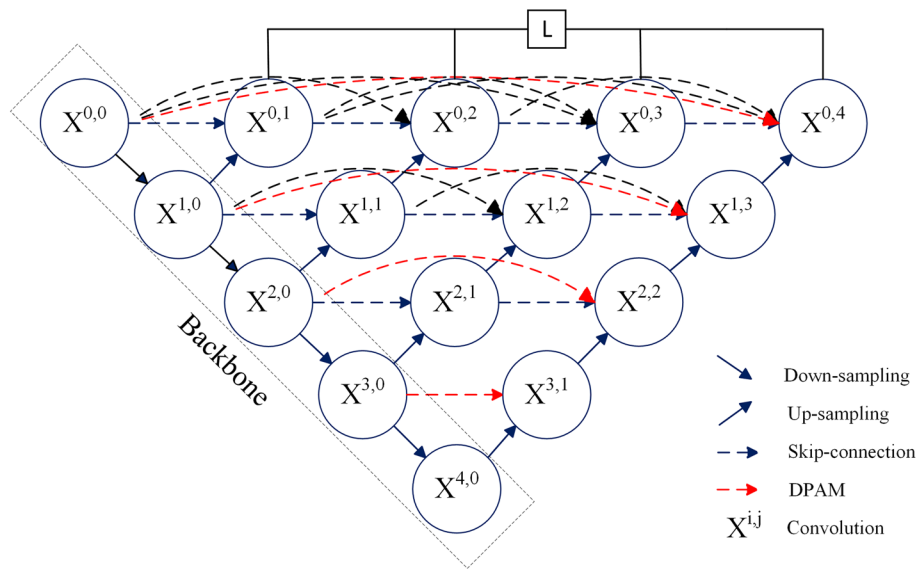Dong *et al. BMC Medical Imaging*   (2024) 24:341

Page 4 of 15

Fig. 1 Structure of DPAM-UNet++

DecoderBlock, which applies a sequence of convolutional layers, normalization, and ReLU activation, yielding the decoded output feature maps.

After decoding, feature maps $x^{0,1}$, $x^{0,2}$, $x^{0,3}$, and $x^{0,4}$ undergo convolution at different scales to produce the corresponding output feature maps (logit1, logit2, logit3, and logit4). These maps capture multi-scale semantic information, allowing the network to effectively balance both local and global features. The network then combines the output maps through a weighted summation to generate the final fused feature map, logit. The logit is subsequently upsampled to the input image size using bilinear interpolation, ensuring that the spatial resolution of the output feature map matches that of the input image.

Additionally, the black dashed arrows in the figure represent skip connections, which transfer high-resolution features between the encoder and decoder, facilitating the accurate restoration of spatial information. The red dashed arrows denote the DPAM module, which integrates global attention, channel attention, and spatial attention mechanisms. By doing so, DPAM enhances the network's ability to focus on global contextual information, enabling more precise feature extraction and improving the model's capacity to accurately identify nodule regions.

DPAM

The structure of the DPAM module is illustrated in Fig. 2. This module takes the feature maps extracted by the encoder as input and processes them through the Global Attention Module (GAM) [25] and the Global Context (GC) Module [26]. The outputs of these modules are concatenated with the original feature maps. The GC module captures global contextual information, while the GAM extracts channel and spatial features. Together, they generate an enhanced feature map that incorporates more critical information.

To optimize computational efficiency and enhance the model's adaptability to different applications, the Efficient Channel Attention (ECA) [27] replaces the traditional Channel Attention (CA) in the GAM. The
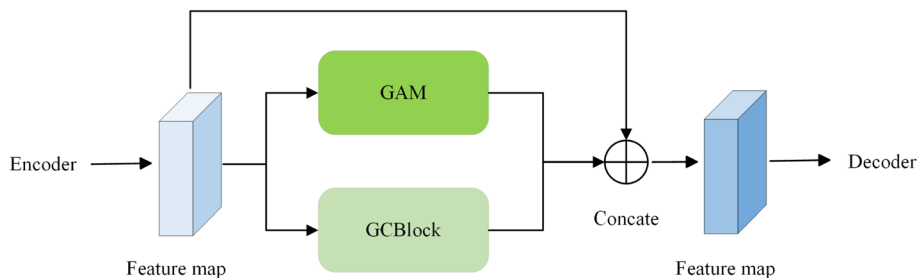

**Fig. 2** DPAM structure, where Concate denotes feature concatenation

overall architecture of the GAM is depicted in Fig. 3a, which revolves around the ECA and Spatial Attention (SA) submodules. Specifically, the input feature F1 first passes through the ECA submodule to obtain Mc(F1). Then, F1 is elementwise multiplied with Mc(F1), and the result is input to the SA submodule to obtain Ms(F2). Finally, Mc(F1) and Ms(F2) are elementwise multiplied, resulting in the final output feature F3. This design allows the GAM module to capture and utilize key image features more efficiently, enhancing the overall image segmentation performance.

The ECA submodule focuses on refining the channel feature representation, as illustrated in Fig. 3b. The input feature F1 undergoes global average pooling (GAP) to generate a channel descriptor. This descriptor is passed through a neural network with a single hidden layer, producing the efficient channel attention map Mc(F1).

The SA submodule, depicted in Fig. 3c, is designed to capture spatial dependencies within the feature map. The input feature F2 is first convolved with a 7×7 kernel, reducing the number of channels from C to C/r. These intermediate features are then convolved with another 7×7 kernel, restoring the number of channels
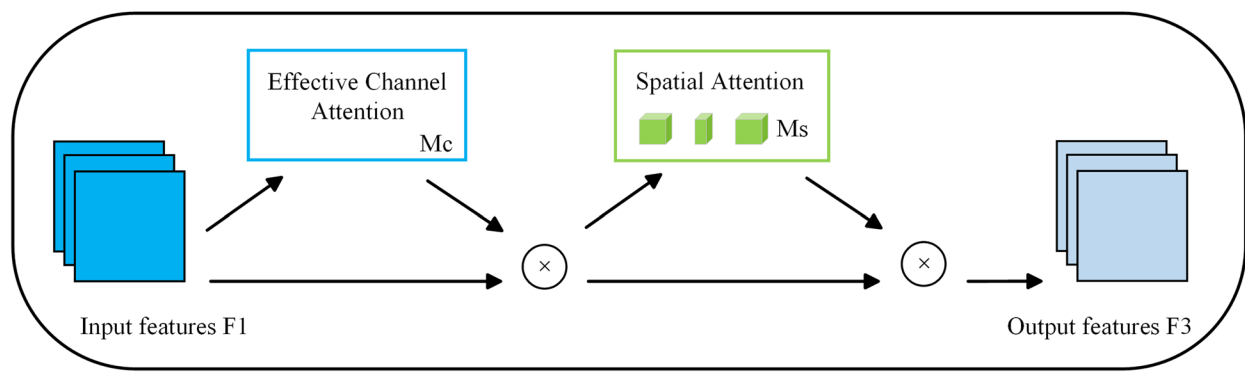
(a) The overview of GAM
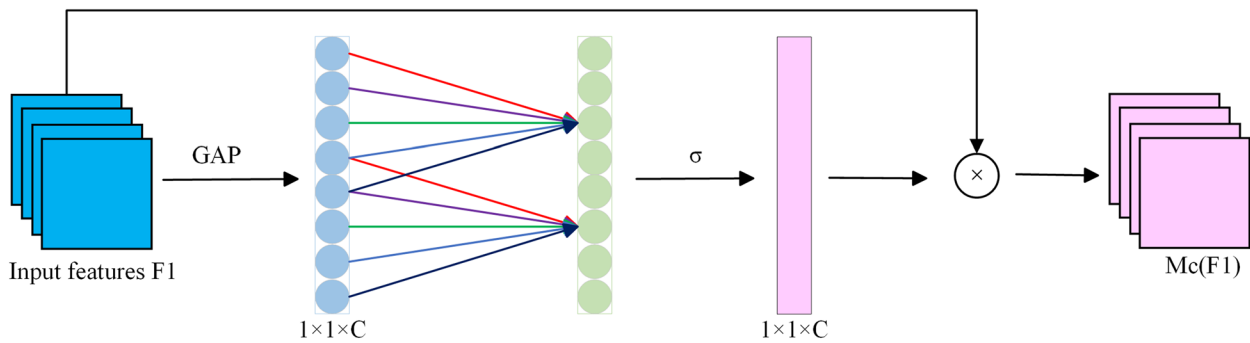
(b) Efficient Channel Attention submodule
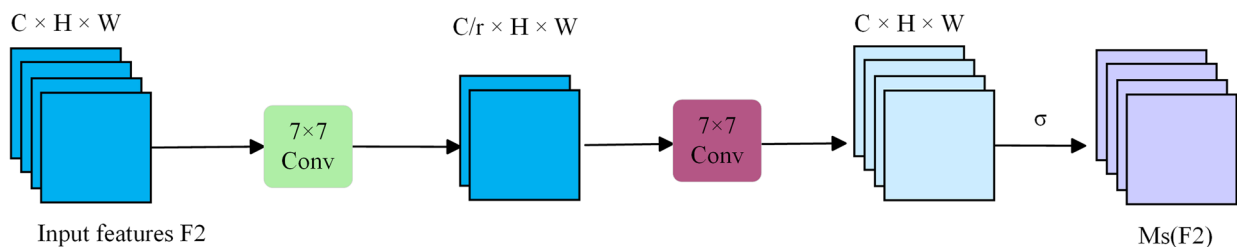
(c) Spatial attention submodule

Fig. 3 Structure of GAM

Dong *et al. BMC Medical Imaging*   (2024) 24:341

Page 6 of 15

to C. Finally, the feature map passes through a sigmoid activation function, generating the spatial attention map Ms(F2). This activation function normalizes weights and enhances stability within the attention mechanism, thereby ensuring the effectiveness of the attention module.

Figure 4 illustrates the structure of the GCBlock module. This module first uses a 1x1 convolution layer and Softmax function to model the context of the input features. The context information is then elementwise multiplied with the input features and then transformed through another 1x1 convolution layer, LayerNorm, and ReLU functions. The transformed features are added elementwise to the input features, producing the final output feature. The GCBlock module enhances the feature map's representation capability by incorporating the global context, enabling the model to perform well in complex scenarios.

By combining the outputs of the GAM and GCBlock with the original feature maps, the DPAM module creates a new feature map rich in contextual and attention information. This method improves segmentation accuracy and enhances feature map representational capacity, making them more suitable for various complex segmentation tasks. In conclusion, the DPAM module is crucial for optimizing feature map representation and improving segmentation performance. The pseudocodes for the GAM and GC attention modules are presented in Algorithms 1 and 2, respectively.

Algorithm 1 Global Attention Module (GAM)

```
1:  Inputs: F1: Input feature map, shape (C × H × W)
2:  Outputs: F3: Output feature map, shape (C × H × W)
3:  Step 1: Efficient Channel Attention Submodule
4:     GAP(F1) = 1/(H×W) Σ_{i=1}^{H} Σ_{j=1}^{W} F1_{ij}    // GAP: Global average pooling
5:     Fc1 = FC(GAP(F1)) = ReLU(W1 · GAP(F1) + b1)
6:     Fc2 = FC(Fc1) = W2 · Fc1 + b2
7:     Mc(F1) = σ(Fc2)    // σ: Sigmoid activation function
8:     F2 = F1 ⊗ Mc(F1)    // ⊗: element-wise multiplication, Mc: ECA
9:  Step 2: Spatial Attention Submodule
10:    Conv1 = Conv7 × 7(F2)
11:    Conv2 = Conv7 × 7(Conv1)
12:    Ms(F2) = σ(Conv2)
13:    F3 = F2 ⊗ Ms(F2)    // Ms: SA
14: return F3
```

Algorithm 2 *Global Context (GC) Block*

```
1:  Inputs: X: Input feature map, shape (C × H × W)
2:  Outputs: Y: Output feature map, shape (C × H × W)
3:  Step 1: Context Modeling
4:     U = Conv1 × 1(X)
5:     W = Softmax(U)
6:     V = X ⊗ W    // ⊗: element-wise multiplication
7:  Step 2: Feature Transformation
8:     T = Conv1 × 1(V)
9:     A = ReLU(LayerNorm(T))
10:    B = Conv1 × 1(A)
11:    Y = B ⊕ X    // ⊕: element-wise add
12: return Y
```

Loss function

The integrated loss function combines the advantages of various loss functions to provide a more comprehensive evaluation of the model's performance. This approach enables the model to capture finer details in the images, thereby enhancing segmentation accuracy [28–30]. We propose the following loss function:

$$L_{\text{total}} = L_{\text{bce}} + L_{\text{dice}} + L_{\text{lovasz}} + L_{\text{ssim}} \tag{1}$$

L_{bce}, L_{dice}, L_{lovasz}, and L_{ssim} in this equation denote BCELoss, DiceLoss [31], LovaszLoss [32] and SSIMLoss [33], respectively. BCELoss is often used in binary segmentation tasks. The formula for BCELoss is as follows:

$$L_{\text{bce}} = -\sum_{i}^{N} [g_i \log s_i + (1 - g_i) \log(1 - s_i)] \tag{2}$$

where g_i and s_i represented the ground truth and model predictions, respectively. DiceLoss measures the degree of similarity between ground truth and model predictions. The formula for DiceLoss is:

$$L_{\text{dice}} = 1 - 2|X \cap Y|/|X + Y| \tag{3}$$

where X and Y represented the number of elements in the ground truth and predicted images, respectively, LovaszLoss, explicitly designed for image segmentation tasks, is particularly effective in handling class imbalance. Its main feature optimizes the Jaccard index (Intersection over Union, IoU), a key metric for evaluating image segmentation performance. The formula for LovaszLoss is as follows:
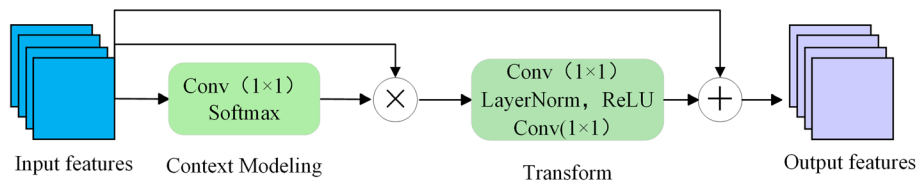

Fig. 4 Structure of GCBlock

Dong *et al. BMC Medical Imaging*   (2024) 24:341

Page 7 of 15

$$L_{\text{lovasz}} = \sum_{i=1}^{N} \Delta J_i \max(0, 1 - g_i s_i) \tag{4}$$

where ΔJ_i is the marginal contribution of pixel i to the Jaccard index, and g_i and s_i are the ground truth and predicted values, respectively. SSIMLoss can help the segmentation model preserve the structural information of the image better. The formula for SSIMLoss is as follows:

$$L_{\text{ssim}} = (2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)/\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right) \tag{5}$$

In this equation, c_1 and c_2 are constants, x and y denote the actual and predicted images, respectively, and μ, σ, and σ_{xy} denote the mean, variance, and covariance of these two images, respectively. The new composite loss function consists of multiple loss functions. BCELoss evaluates the difference between the predicted and actual values pixel by pixel. DiceLoss measures the similarity between the exact labels and the predicted images. LovaszLoss can solve the class imbalance problem in image segmentation. SSIMLoss improves the accuracy of boundary segmentation. Combining these loss functions can enhance segmentation from multiple perspectives, such as pixel level, sample similarity, image structure, and segmentation boundaries.

Evaluation metrics

We use metrics such as Intersection over Union (IoU), F1_score, Accuracy, Precision, HD95, Recall, to measure the segmentation performance of the model. These evaluation metrics are defined below:

$$IoU = TP/(FP + FN + TP) \tag{6}$$

$$Accuracy = (TN + TP)/(TN + TP + FN + FP) \tag{7}$$

$$Precision = TP/(TP + FP) \tag{8}$$

$$Recall = TP/(TP + FN) \tag{9}$$

$$F1\_score = 2TP/(FP + FN + 2TP) \tag{10}$$

HD95 measures the quality of segmentation boundaries by calculating the 95th percentile of the maximum distance between the predicted values and the ground truth. In addition, we evaluated the performance of the model using a confusion matrix. The detailed explanations of *TP* (True Positive), *TN* (True Negative), *FP* (False Positive), and *FN* (False Negative) in the formula are as follows: *TP* indicates that the model correctly classified pixels as part of the thyroid nodule region; *TN* indicates that the model correctly classified pixels as part of the background region; *FP* refers to the model incorrectly classifying background pixels as part of the nodule region; and *FN* refers to the model incorrectly classifying nodule pixels as part of the background region.

Results

Model segmentation performance

We compared the DPAM-UNet++ network model with a series of advanced models widely used in image segmentation tasks, including UNet [34], CMUNeXt [35], TRFE+ [4], DeepLabV3+ [36], TransUNet [37], SGUNet [38], and TransAttUNet [39]. As shown in Table 3, apart from the accuracy metric where TransUNet scored higher, the DPAM-UNet++ model outperformed the other models in the remaining metrics.

In Fig. 5, we provide the segmentation results for different models, with each row depicting the same ultrasound image segmented by various models. Original and GT in Fig. 5 denote original image and ground truth, respectively. The DPAM-UNet++ model's segmentation results are visibly the closest to the ground truth labels, maintaining nodule boundary information more accurately. Figure 6 shows the receiver operating characteristic (ROC) curves for the various segmentation models on the TN3K dataset, revealing that DPAM-UNet++ achieved an AUC value of 0.9213, outperforming the other models.

Table 3 Segmentation results of different models

Model	F1_score	IoU	Accuracy	Precision	Recall	HD95
SGUNet [38]	0.7323	0.6298	0.9540	0.8522	0.7120	60.06
UNet [34]	0.7486	0.6507	0.9583	0.8184	0.7668	62.27
CMUNeXt [35]	0.7812	0.6908	0.9643	0.8194	0.8192	50.47
TransAttUNet [39]	0.7992	0.7054	0.9645	0.8374	0.8271	39.56
TRFE+ [4]	0.8003	0.7109	0.9663	0.8536	0.8172	40.44
DeepLabV3+ [36]	0.8058	0.7159	0.9681	0.8525	0.8208	36.50
TransUNet [37]	0.8067	0.7188	0.9667	**0.8574**	0.8161	37.73
DPAM-UNet++(ours)	**0.8310**	**0.7451**	**0.9718**	0.8443	**0.8702**	**35.31**

Dong *et al. BMC Medical Imaging*   (2024) 24:341
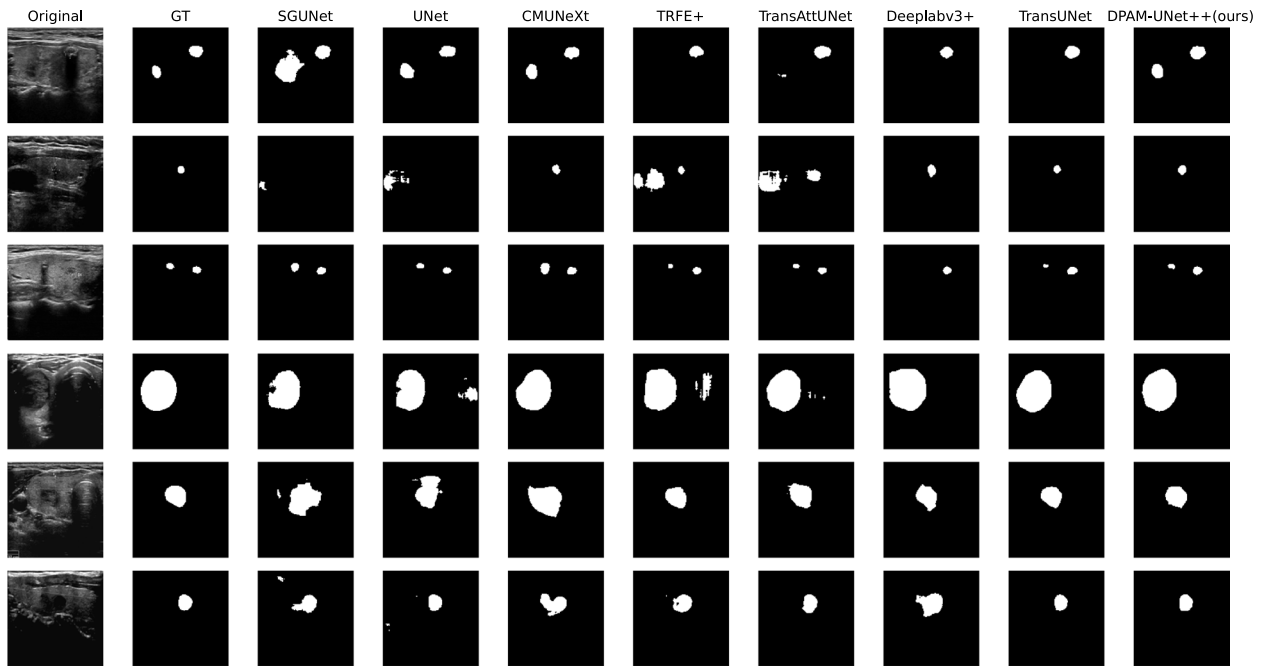
Page 8 of 15

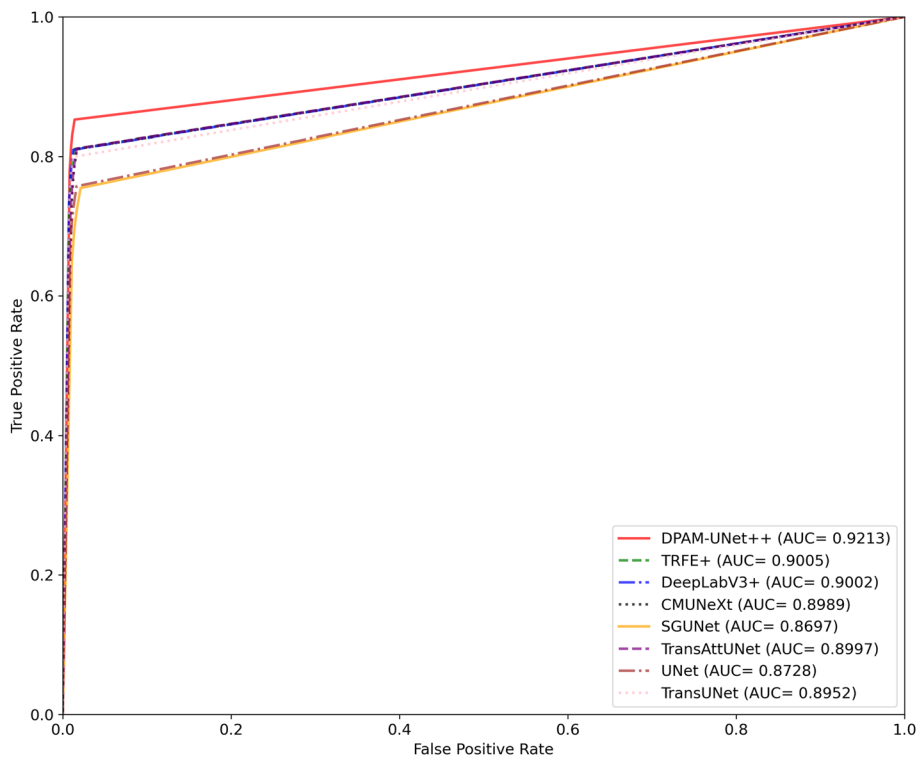**Fig. 5** The effect diagram of different model segmentation

Fig. 6 ROC curves for different segmentation models

Ablation study of the DPAM module

Table 4 presents the results of the ablation experiments on the DPAM attention module. The results indicate that the DPAM attention module increases the F1 score by 1.74% compared with the baseline. Additionally, the F1 score and IoU metric of the DPAM attention module outperform those of the other attention modules, significantly enhancing the segmentation accuracy. The segmentation effects of the ablation experiments on the attention modules are shown in Fig. 7. The figure clearly shows that the introduction of the DPAM attention module allows the segmentation model to more accurately capture and identify nodule regions, performing excellently in handling multiple and small nodules. Furthermore, to visually demonstrate the effectiveness of the proposed attention module, we employed the Grad-CAM visualization technique to generate attention heatmaps for different attention modules integrated into the UNet++ baseline, as shown in Fig. 8. The model with the DPAM attention module more effectively focuses on the significant nodule areas.

Ablation study of loss functions

We conducted ablation experiments to demonstrate the effectiveness and necessity of the proposed loss function. As shown in Table 5, the ablation study results indicate significant improvements in both the IoU and F1_score with the inclusion of LovaszLoss and SSIMLoss. Specifically, the addition of LovaszLoss increased the IoU by 1.48% and the F1_score by 1.28%. In comparison, the
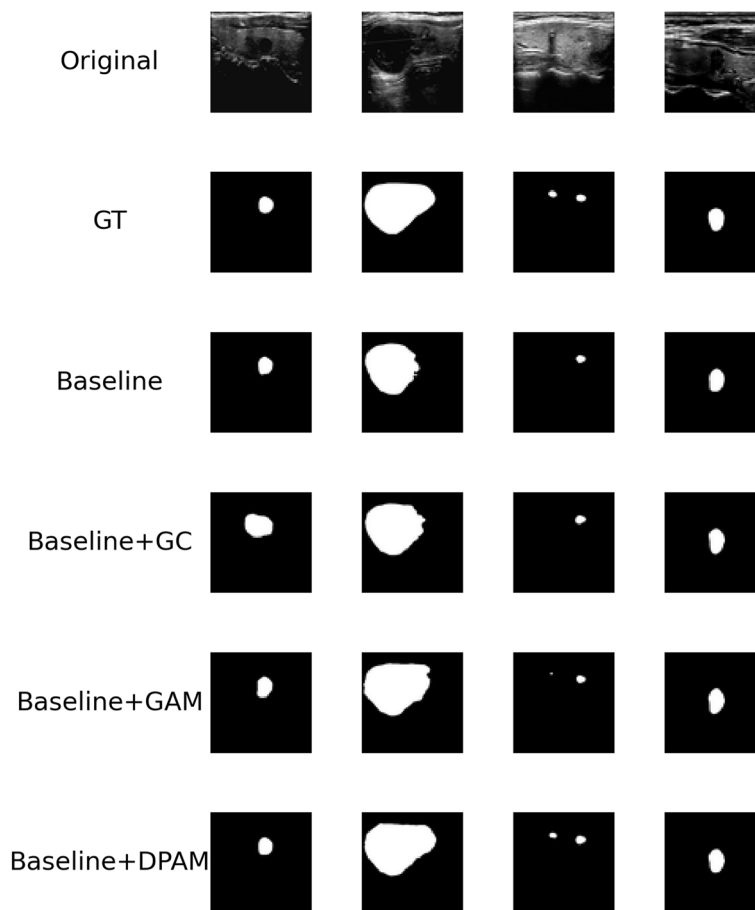
Table 4 Attention module ablation experiment

Model	ECA	GC	GAM	IoU	F1_score
Baseline				0.7311	0.8136
Baseline	✓			0.7244	0.8085
Baseline		✓		0.7370	0.8208
Baseline			✓	0.7286	0.8143
Baseline		✓	✓	0.7451	0.8310

Where Baseline is a UNet++ model

Fig. 7 Plot of segmentation results of the attention module ablation experiment
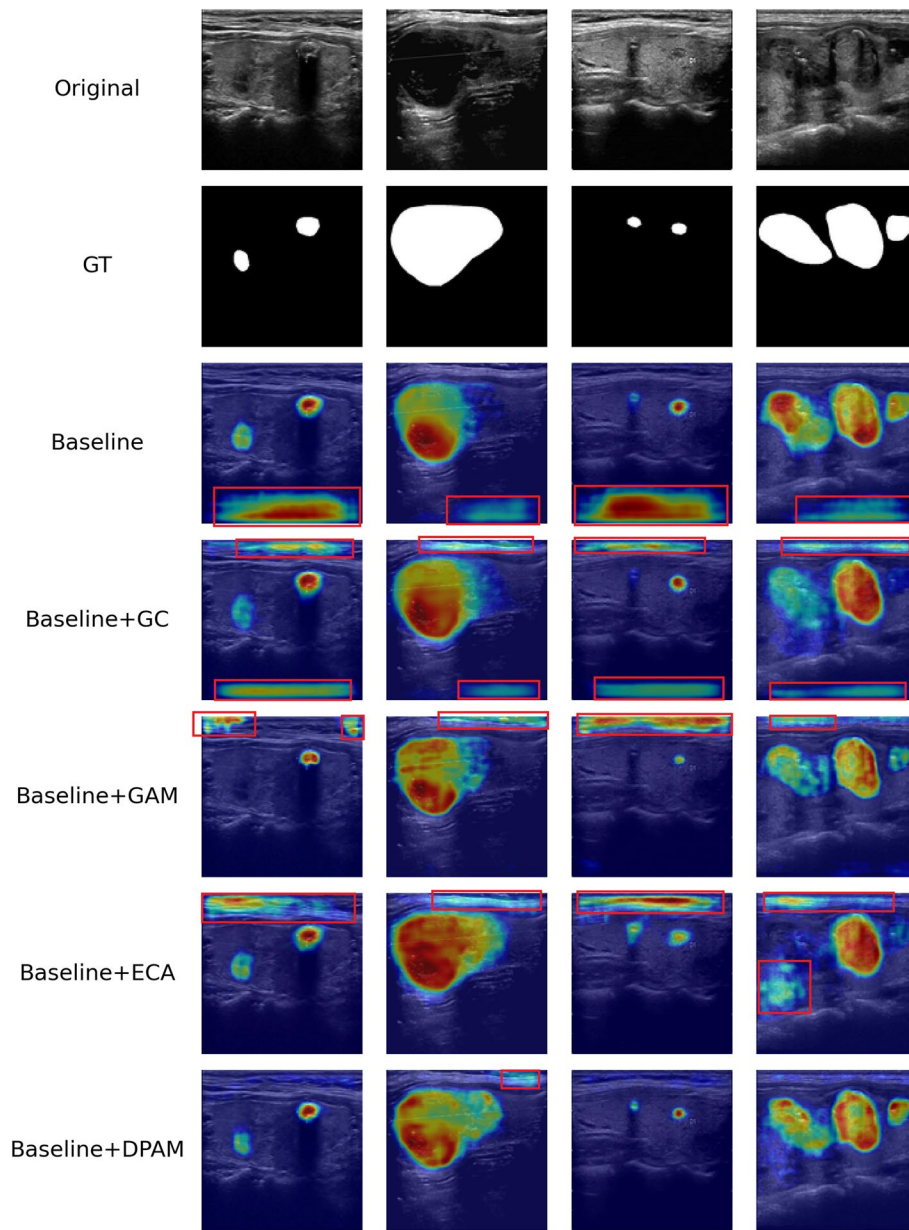
Fig. 8 Attention heatmap for different attention modules

Table 5 Loss function ablation experiments

Model	BCELoss	DiceLoss	LovaszLoss	SSIMLoss	IoU	F1_score	HD95
DPAM-UNet++	✓				0.7270	0.8125	40.50
DPAM-UNet++		✓			0.7304	0.8135	39.73
DPAM-UNet++	✓	✓			0.7276	0.8138	39.58
DPAM-UNet++	✓	✓	✓		0.7424	0.8266	36.12
DPAM-UNet++	✓	✓		✓	0.7411	0.8244	33.93
DPAM-UNet++	✓	✓	✓	✓	0.7451	0.8310	35.31

inclusion of SSIMLoss resulted in a 1.35% increase in the IoU and a 1.06% increase in the F1_score. More importantly, when LovaszLoss and SSIMLoss were combined, the improvements in the IoU and F1_score were even more pronounced, with increases of 1.75% and 1.72%, respectively. These results validate the complementary nature of the two loss functions, indicating that their combination can more comprehensively enhance the model's segmentation performance. Additionally, the model trained with the integrated loss function achieved the lowest HD95 value, compared to models trained with individual loss functions. This demonstrates that the proposed integrated loss function can effectively optimize the segmentation of nodule boundaries.

Hyperparameters and confusion matrix

Figures 9 and 10 illustrate the training loss and training accuracy of the model, respectively. As shown in Fig. 9, the loss value of the model exhibits a clear downward trend as the number of iterations increases. In Fig. 10, the training accuracy gradually improves from 0.87 to 0.95. The reduction in the loss function indicates that the model is progressively learning the underlying patterns and features of the data during training. The increase in accuracy suggests that the discrepancy between the model's predictions and the actual outcomes is gradually decreasing. Figure 11 presents the confusion matrix, obtained by per-pixel classification and normalization of the segmentation results. From this matrix, it is evident that the probability of correctly identifying nodules (true label: nodule, predicted label: nodule) is 0.86, which

further demonstrates the model's effectiveness in recognizing nodule regions.

Discussion

The early detection and treatment of thyroid nodules are crucial for preventing thyroid cancer. Currently, the diagnosis of thyroid nodules relies primarily on radiologists segmenting ultrasound images, a process that is both time-consuming [40] and highly dependent on the physician's expertise [41]. Therefore, designing an automated segmentation method to assist physicians in diagnosis is highly important. In this study, we incorporated the DPAM attention module into the skip connections of the UNet++ network to capture global contextual information in feature maps. Additionally, we design a novel integrated loss function to increase the model's robustness and adaptability to complex environments.

Currently, dual-path attention networks are widely applied in medical image segmentation. Sun et al. [42] combined position and channel attention modules to propose a dual-attention module, which they integrated into the skip connections of the TransUNet network, enhancing the model's ability to capture segmentation details. However, spatial attention mechanisms were not incorporated into the skip connections. Cui et al. [43] introduced a dual dilated gated attention module by merging dilated spatial and channel gated attention mechanisms into the skip connections of the UNet network, achieving satisfactory segmentation results but without incorporating global context information. Similarly, Luo et al. [44] combined adaptive spatial and
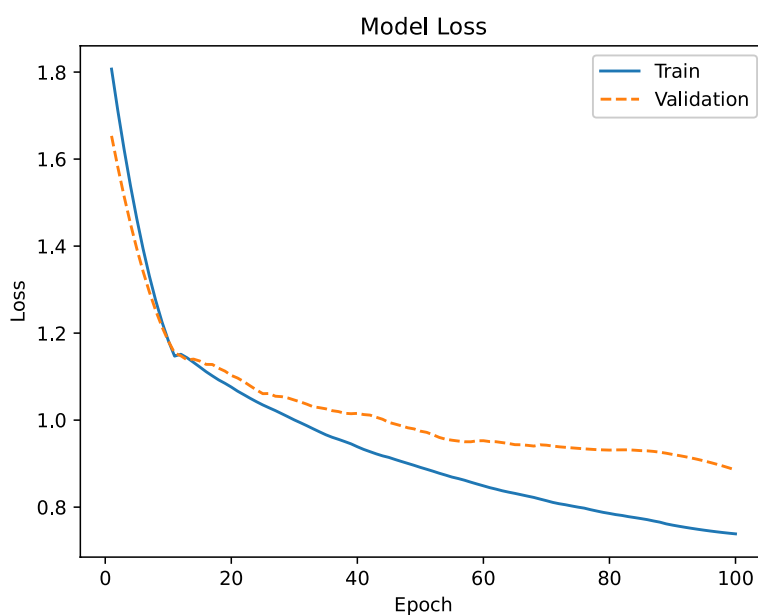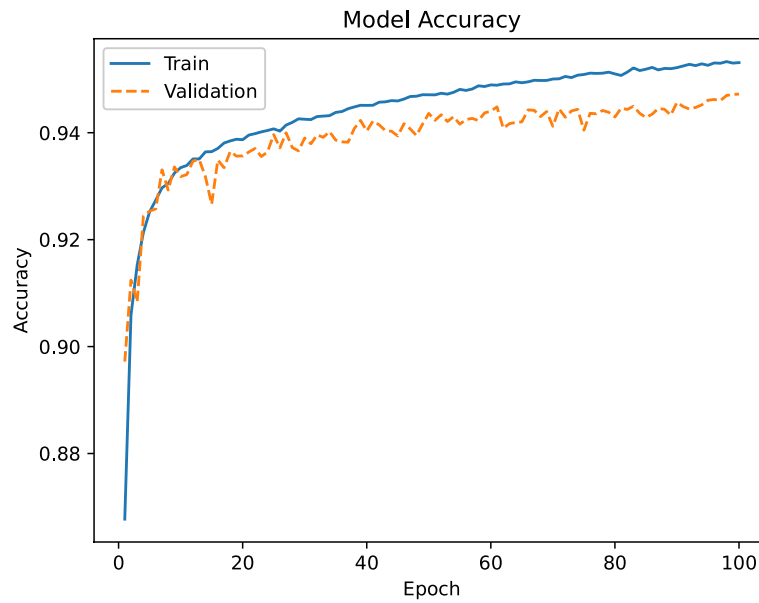

Fig. 9 Model loss graphs

Dong *et al. BMC Medical Imaging*　　(2024) 24:341

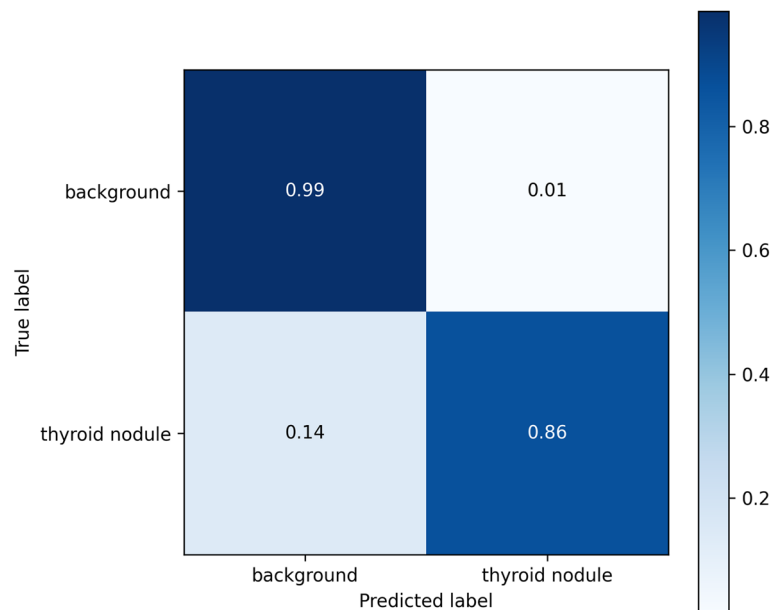Page 12 of 15

Fig. 10 Model accuracy graphs

Fig. 11 Pixel level binary classification confusion matrix

channel attention to create an adaptive dual-path attention module integrated with UNet, yet they also did not utilize global context. To capture critical information from the feature maps from global, channel, and spatial perspectives, we propose the DPAM attention module, which consists of two paths: one for capturing global context and the other for extracting spatial and channel features.

We utilized 3,493 thyroid ultrasound images from 2,421 patients and conducted comparative experiments with seven typical deep learning segmentation models. The experimental results demonstrate that the proposed segmentation model can accurately identify lesion locations by disregarding irrelevant information, thereby improving segmentation accuracy. Moreover, the model can finely segment the edges of nodules, making the

segmentation results closer to the ground truth and further enhancing precision.

Multiple nodules are a common clinical finding, and the probability of these nodules becoming malignant is similar to or even greater than that of a single nodule [45]. The DAPM module introduced in this study effectively utilizes global contextual information from feature maps to detect nodules accurately, offering significant improvements in identifying multiple nodules. As shown in Table 3, our model's recall rate is significantly higher than those of the other models, outperforming the second-best TransAttUNet model by 4.31%. This indicates the model's strong ability to detect multiple nodules, as evidenced by the ablation experiments in Fig. 7. The excellent performance is attributed to the incorporation of the DPAM module, which extracts features from different perspectives and scales, enhancing the feature representation for small and multiple nodules and thus improving the recall rate. Figure 8 shows that the model with the DPAM attention module accurately identifies significant nodule regions while ignoring irrelevant factors, increasing the segmentation accuracy. Where the irrelevant factors are marked with a red boxed line in the figure.

The use of an integrated loss function helps the model capture finer details in images, improving segmentation precision [30]. In this study, we innovatively introduced LovaszLoss and SSIMLoss on top of BCELoss and DiceLoss to form a new integrated loss function. By combining the advantages of each loss function, we improve the segmentation accuracy of the model. As shown in Fig. 5, the segmentation results of our proposed model closely align with the ground truth. Additionally, the high AUC values and well-formed ROC curves in Fig. 6 further validate the model's superior performance.

Conventional medical segmentation models are prone to noise interference [46], and many deep learning-based segmentation models focus primarily on local features, often ignoring global features [47]. This study introduces the DPAM attention module to capture global feature information within feature maps and utilizes an integrated loss function to optimize the segmentation of nodule edges. This method enhances the accuracy and efficiency of nodule segmentation from both global and local perspectives, helps reduce misdiagnosis due to limited physician experience and decreases the workload of doctors. This method has potential for use in the future clinical auxiliary diagnosis of thyroid nodules.

While our study produced promising results, there are still several limitations: the dataset size remains relatively small despite enhancing the studied image data through data augmentation. Therefore, in future research, we plan to use generative adversarial networks for ultrasound data augmentation and train the model with augmented data to further enhance segmentation performance.

Conclusion

This study innovatively introduces a dual-path attention module integrated into the skip connections of the UNet++ network, along with a novel integrated loss function, successfully enhancing the segmentation accuracy of thyroid nodules. Comparative experiments with other models demonstrate that our approach excels in detecting multiple nodules and achieving precise edge segmentation. These results contribute to improved efficiency and accuracy in thyroid nodule segmentation and provide new insights for clinical decision support, potentially advancing the precision and intelligence of thyroid nodule diagnosis. In future research, we will further explore this method and investigate additional optimization strategies to enhance the model's performance and generalization capabilities, ultimately contributing more significantly to the precise diagnosis and treatment of thyroid nodules.

Abbreviations

DPAM	Dual-path attention mechanism
IoU	Intersection over union
AUC	Area under curve
GAM	Global attention module
GC	Global context
CA	Channel attention
ECA	Efficient channel attention
ROC	Receiver operating characteristic curve
GAP	Global average pooling
GT	Ground truth
TP	True positive
FP	False positive
TN	True negative
FN	False negative

Dong *et al. BMC Medical Imaging*   (2024) 24:341

Page 14 of 15

Declarations

Clinical trial number
Not applicable.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References

1. Dai R, Peng W. Research Progress of Deep Learning in Thyroid Nodule Imaging Examination. In: International Conference on Medical Imaging and Computer-Aided Diagnosis. Springer; 2023. pp. 307–17. https://doi.org/10.1007/978-981-97-1335-6_27.
2. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229–63. https://doi.org/10.3322/caac.21834.
3. Kang Q, Lao Q, Li Y, Jiang Z, Qiu Y, Zhang S, et al. Thyroid nodule segmentation and classification in ultrasound images through intra-and inter-task consistent learning. Med Image Anal. 2022;79:102443. https://doi.org/10.1016/j.media.2022.102443.
4. Gong H, Chen J, Chen G, Li H, Li G, Chen F. Thyroid region prior guided attention for ultrasound segmentation of thyroid nodules. Comput Biol Med. 2023;155:106389. https://doi.org/10.1016/j.compbiomed.2022.106389.
5. Chen J, You H, Li K. A review of thyroid gland segmentation and thyroid nodule segmentation methods for medical ultrasound images. Comput Methods Prog Biomed. 2020;185:105329. https://doi.org/10.1016/j.cmpb.2020.105329.
6. Shahroudnejad A, Vega R, Forouzandeh A, Balachandran S, Jaremko J, Noga M, et al. Thyroid nodule segmentation and classification using deep convolutional neural network and rule-based classifiers. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE; 2021. pp. 3118–21. https://doi.org/10.1109/embc46164.2021.9629557.
7. Chambara N, Ying M. The diagnostic efficiency of ultrasound computer-aided diagnosis in differentiating thyroid nodules: A systematic review and narrative synthesis. Cancers. 2019;11(11):1759. https://doi.org/10.3390/cancers11111759.
8. Wang S, Li Z, Liao L, Zhang C, Zhao J, Sang L, et al. DPAM-PSPNet: ultrasonic image segmentation of thyroid nodule based on dual-path attention mechanism. Phys Med Biol. 2023;68(16):165002. https://doi.org/10.1088/1361-6560/ace6f1.
9. Senthilkumaran N, Vaithegi S. Image segmentation by using thresholding techniques for medical images. Comput Sci Eng Int J. 2016;6(1):1–13. https://doi.org/10.5121/cseij.2016.6101.
10. Szénási S. Distributed region growing algorithm for medical image segmentation. Int J Circ Syst Signal Process. 2014;8(1):173–81.
11. Singh LK, Khanna M, Thawkar S. A novel hybrid robust architecture for automatic screening of glaucoma using fundus photos, built on feature selection and machine learning-nature driven computing. Expert Syst. 2022;39(10):e13069. https://doi.org/10.1111/exsy.13069.
12. Han Z, Jian M, Wang GG. ConvUNeXt: An efficient convolution neural network for medical image segmentation. Knowl Based Syst. 2022;253:109512. https://doi.org/10.1016/j.knosys.2022.109512.
13. Ansari MY, Mohanty S, Mathew SJ, Mishra S, Singh SS, Abinahed J, et al. Towards developing a lightweight neural network for liver CT segmentation. In: International Conference on Medical Imaging and Computer-Aided Diagnosis. Springer; 2022. pp. 27–35. https://doi.org/10.1007/978-981-16-6775-6_3.
14. Jafari M, Auer D, Francis S, Garibaldi J, Chen X. DRU-Net: an efficient deep convolutional neural network for medical image segmentation. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE; 2020. pp. 1144–8. https://doi.org/10.1109/isbi45749.2020.9098391.
15. Xie Y, Zhang J, Shen C, Xia Y. Cotr: Efficiently bridging cnn and transformer for 3d medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24. Springer; 2021. pp. 171–80. https://doi.org/10.1007/978-3-030-87199-4_16.
16. Ansari MY, Mangalote IAC, Meher PK, Aboumarzouk O, Al-Ansari A, Halabi O, et al. Advancements in Deep Learning for B-Mode Ultrasound Segmentation: A Comprehensive Review. IEEE Trans Emerg Top Comput Intell. 2024. https://doi.org/10.1109/tetci.2024.3377676.
17. Yadav N, Dass R, Virmani J. Deep learning-based CAD system design for thyroid tumor characterization using ultrasound images. Multimed Tools Appl. 2024;83(14):43071–113. https://doi.org/10.1007/s11042-023-17137-4.
18. Singh LK, Khanna M. Introduction to artificial intelligence and current trends. In: Innovations in Artificial Intelligence and Human-Computer Interaction in the Digital Era. Elsevier; 2023. pp. 31–66. https://doi.org/10.1016/b978-0-323-99891-8.00001-2.
19. Yang B, Yan M, Yan Z, Zhu C, Xu D, Dong F. Segmentation and classification of thyroid follicular neoplasm using cascaded convolutional neural network. Phys Med Biol. 2020;65(24):245040. https://doi.org/10.1088/1361-6560/abc6f2.
20. Nie X, Zhou X, Tong T, Lin X, Wang L, Zheng H, et al. N-Net: a novel dense fully convolutional neural network for thyroid nodule segmentation. Front Neurosci. 2022;16:872601. https://doi.org/10.3389/fnins.2022.872601.
21. Yadav N, Dass R, Virmani J. Objective assessment of segmentation models for thyroid ultrasound images. J Ultrasound. 2023;26(3):673–85. https://doi.org/10.1007/s40477-022-00726-8.
22. Chen G, Tan G, Duan M, Pu B, Luo H, Li S, et al. MLMSeg: a multi-view learning model for ultrasound thyroid nodule segmentation. Comput Biol Med. 2024;169:107898. https://doi.org/10.1016/j.compbiomed.2023.107898.
23. Xu H, Ye C, Zhang F, Li X, Zhang C. A medical image segmentation method with anti-noise and bias-field correction. IEEE Access. 2020;8:98548–61. https://doi.org/10.1109/access.2020.2996603.
24. Song J, Chen X, Zhu Q, Shi F, Xiang D, Chen Z, et al. Global and local feature reconstruction for medical image segmentation. IEEE Trans Med Imaging. 2022;41(9):2273–84. https://doi.org/10.1109/tmi.2022.3162111.
25. Liu Y, Shao Z, Hoffmann N. Global attention mechanism: Retain information to enhance channel-spatial interactions. arXiv preprint arXiv:211205561. 2021. https://doi.org/10.48550/arXiv.2112.05561.
26. Cao Y, Xu J, Lin S, Wei F, Hu H. Gcnet: Non-local networks meet squeeze-excitation networks and beyond. In: Proceedings of the IEEE/CVF international conference on computer vision workshops. 2019. https://doi.org/10.1109/iccvw.2019.00246.
27. Wang Q, Wu B, Zhu P, Li P, Zuo W, Hu Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. pp. 11534–42. https://doi.org/10.1109/cvpr42600.2020.01155.
28. Yang D, Li Y, Yu J. Multi-task thyroid tumor segmentation based on the joint loss function. Biomed Signal Process Control. 2023;79:104249. https://doi.org/10.1016/j.bspc.2022.104249.
29. Yeung M, Sala E, Schönlieb CB, Rundo L. Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. Comput Med Imaging Graph. 2022;95:102026. https://doi.org/10.1016/j.compmedimag.2021.102026.
30. Ma J, Chen J, Ng M, Huang R, Li Y, Li C, et al. Loss odyssey in medical image segmentation. Med Image Anal. 2021;71:102035. https://doi.org/10.1016/j.media.2021.102035.
31. Li X, Sun X, Meng Y, Liang J, Wu F, Li J. Dice loss for data-imbalanced NLP tasks. arXiv preprint arXiv:191102855. 2019. https://doi.org/10.48550/arXiv.1911.02855.
32. Berman M, Triki AR, Blaschko MB. The lovász-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure

in neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. pp. 4413–21. https://doi.org/10.48550/arXiv.1705.08790.

33. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process. 2004;13(4):600–12. https://doi.org/10.1109/tip.2003.819861.

34. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer; 2015. pp. 234–41. https://doi.org/10.1007/978-3-319-24574-4_28.

35. Tang F, Ding J, Quan Q, Wang L, Ning C, Zhou SK. Cmunext: An efficient medical image segmentation network based on large kernel and skip fusion. In: 2024 IEEE International Symposium on Biomedical Imaging (ISBI). IEEE; 2024. pp. 1–5. https://doi.org/10.1109/isbi56570.2024.10635609.

36. Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV). 2018. pp. 801–18. https://doi.org/10.1007/978-3-030-01234-2_49.

37. Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, et al. Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:210204306. 2021. https://doi.org/10.48550/arXiv.2102.04306.

38. Pan H, Zhou Q, Latecki LJ. Sgunet: Semantic guided unet for thyroid nodule segmentation. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). IEEE; 2021. pp. 630–4. https://doi.org/10.1109/isbi48211.2021.9434051.

39. Chen B, Liu Y, Zhang Z, Lu G, Kong AWK. Transattunet: Multi-level attention-guided u-net with transformer for medical image segmentation. IEEE Trans Emerg Top Comput Intell. 2023. https://doi.org/10.1109/tetci.2023.3309626.

40. Das D, Iyengar MS, Majdi MS, Rodriguez JJ, Alsayed M. Deep learning for thyroid nodule examination: a technical review. Artif Intell Rev. 2024;57(3):47. https://doi.org/10.1007/s10462-023-10635-9.

41. Tang L, Tian C, Yang H, Cui Z, Hui Y, Xu K, et al. TS-DSANN: Texture and shape focused dual-stream attention neural network for benign-malignant diagnosis of thyroid nodules in ultrasound images. Med Image Anal. 2023;89:102905. https://doi.org/10.1016/j.media.2023.102905.

42. Sun G, Pan Y, Kong W, Xu Z, Ma J, Racharak T, et al. DA-TransUNet: integrating spatial and channel dual attention with transformer U-net for medical image segmentation. Front Bioeng Biotechnol. 2024;12:1398237. https://doi.org/10.3389/fbioe.2024.1398237/full.

43. Cui R, Yang R, Liu F, Geng H. HD2A-Net: A novel dual gated attention network using comprehensive hybrid dilated convolutions for medical image segmentation. Comput Biol Med. 2023;152:106384. https://doi.org/10.1016/j.compbiomed.2022.106384.

44. Luo S, Zhang J, Xiao N, Qiang Y, Li K, Zhao J, et al. DAS-Net: A lung nodule segmentation method based on adaptive dual-branch attention and shadow mapping. Appl Intell. 2022;52(13):15617–31. https://doi.org/10.1007/s10489-021-03038-2.

45. Rehman AU, Ehsan M, Javed H, Ameer MZ, Mohsin A, Aemaz Ur Rehman M, et al. Solitary and multiple thyroid nodules as predictors of malignancy: a systematic review and meta-analysis. Thyroid Res. 2022;15(1):22. https://doi.org/10.1186/s13044-022-00140-6.

46. Jyothi P, Singh AR. Deep learning models and traditional automated techniques for brain tumor segmentation in MRI: a review. Artif Intell Rev. 2023;56(4):2923–69. https://doi.org/10.1007/s10462-022-10245-x.

47. Emek Soylu B, Guzel MS, Bostanci GE, Ekinci F, Asuroglu T, Acici K. Deep-learning-based approaches for semantic segmentation of natural scene images: A review. Electronics. 2023;12(12):2730. https://doi.org/10.3390/electronics12122730.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.