Attention gate and dilation U-shaped network (GDUNet): an efficient breast ultrasound image segmentation network with multiscale information extraction

Jiadong Chen[1], Xiaoyan Shen[2], Yu Zhao[1], Wei Qian[1], He Ma[1,3]^, Liang Sang[4]

[1]College of Medicine and Biological Information Engineering, Northeastern University, Shenyang, China; [2]School of Life and Health Technology, Dongguan University of Technology, Dongguan, China; [3]Key Laboratory of Intelligent Computing in Medical Image, Ministry of Education, Northeastern University, Shenyang, China; [4]Department of Ultrasound, The First Hospital of China Medical University, Shenyang, China

Contributions: (I) Conception and design: J Chen, X Shen, H Ma; (II) Administrative support: H Ma, W Qian; (III) Provision of study materials or patients: H Ma, W Qian, L Sang; (IV) Collection and assembly of data: J Chen, X Shen, Y Zhao; (V) Data analysis and interpretation: J Chen, X Shen; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: He Ma, PhD. College of Medicine and Biological Information Engineering, Northeastern University, 195 Chuangxin Road, Hunnan District, Shenyang 110819, China; Key Laboratory of Intelligent Computing in Medical Image, Ministry of Education, Northeastern University, Shenyang 110819, China. Email: mahe@bmie.neu.edu.cn; Liang Sang, MD. Department of Ultrasound, The First Hospital of China Medical University, 155 Nanjing North Street, Heping District, Shenyang 110001, China. Email: 13889167622@163.com.

Background: In recent years, computer-aided diagnosis (CAD) systems have played an important role in breast cancer screening and diagnosis. The image segmentation task is the key step in a CAD system for the rapid identification of lesions. Therefore, an efficient breast image segmentation network is necessary for improving the diagnostic accuracy in breast cancer screening. However, due to the characteristics of blurred boundaries, low contrast, and speckle noise in breast ultrasound images, breast lesion segmentation is challenging. In addition, many of the proposed breast tumor segmentation networks are too complex to be applied in practice.

Methods: We developed the attention gate and dilation U-shaped network (GDUNet), a lightweight, breast lesion segmentation model. This model improves the inverted bottleneck, integrating it with tokenized multilayer perceptron (MLP) to construct the encoder. Additionally, we introduce the lightweight attention gate (AG) within the skip connection, which effectively filters noise in low-level semantic information across spatial and channel dimensions, thus attenuating irrelevant features. To further improve performance, we innovated the AG dilation (AGDT) block and embedded it between the encoder and decoder in order to capture critical multiscale contextual information.

Results: We conducted experiments on two breast cancer datasets. The experiment's results show that compared to UNet, GDUNet could reduce the number of parameters by 10 times and the computational complexity by 58 times while providing a double of the inference speed. Moreover, the GDUNet achieved a better segmentation performance than did the state-of-the-art medical image segmentation architecture.

Conclusions: Our proposed GDUNet method can achieve advanced segmentation performance on different breast ultrasound image datasets with high efficiency.

---

^ ORCID: 0000-0002-5054-3586.

Introduction

Breast cancer is one of the most common malignancies affecting women worldwide (1). Early detection and precise diagnosis are critical for successful treatment and improved patient prognosis. Clinically, breast ultrasound has become an important tool in the screening and diagnosis of breast lesions due to its noninvasive nature, cost-effectiveness, easy operation, and lack of ionizing radiation (2). In recent years, with the advancement of computer vision technology, Computer-aided diagnosis (CAD) systems have been widely used in clinical practice, especially in the early screening and diagnosis of breast cancer via ultrasound (3-5). CAD can facilitate the automated analysis and interpretation of breast ultrasound images, helping to detect and localize potential breast lesions. Image segmentation is a crucial step in CAD systems, and an efficient segmentation method can improve their accuracy in diagnosing diseases. However, accurate segmentation of breast ultrasound images is challenging due to the complexity of breast tissue, the presence of noise and artifacts in the images, and other factors. Therefore, the development of an accurate and efficient automatic breast ultrasound segmentation network has considerable clinical importance.

In recent years, researchers have proposed various methods for breast ultrasound image segmentation tasks based on convolutional neural networks (CNNs). Among them, the U-shaped network (UNet) (6) is a landmark network model in medical image segmentation, and many mainstream medical image segmentation methods have been derived from UNet. Almajalid *et al.* (7) improved UNet by using contrast enhancement and speckle-reduction preprocessing techniques. However, the 3×3 convolution kernel of UNet fixes the receptive field and only captures local information, ignoring the connection of global contextual information. To obtain a global view, Irfan *et al.* (8) obtained a larger receptive field by introducing dilated convolutions (9) with different dilation rates. However, considering the characteristics of blurred boundaries and shadowing effects of breast ultrasound images, the use of dilated convolution alone cannot accurately segment breast tumors. To further optimize the performance of breast tumor segmentation, an attention-enhanced UNet with hybrid dilation convolution was proposed by Yan *et al.* (10). Zhuang *et al.* (11) segmented breast lesions by introducing residual units, dilated convolution, and attention gate (AG). Although the performance of these segmentation networks is optimized, the lower layers of the networks still use small convolution kernels, which results in the extraction of shallow features that are too localized to fully cope with the perturbation of the fuzzy boundaries of breast ultrasound images. With the development of deep learning, many transformer-based medical image segmentation methods have emerged, which have improved the segmentation performance through the learning of images' global information. Among these methods is TransUNet (12), whose overall architecture is UNet, but the encoder structure is combined with the encoder structure of the transformer so that the network can better capture local information and global information. MedT (13) extends the structure by introducing additional control mechanisms in the self-attentive module and further improves the performance of the model by applying a local-global training strategy. However, most transformer-based networks need to be trained on large datasets, but breast ultrasound images are rare and typically included in small datasets, and thus the segmentation results of these transformer-based medical image segmentation networks are not satisfactory.

In addition, many breast ultrasound image segmentation methods improve segmentation performance while ignoring model complexity. When segmentation methods are too complex, it is difficult to apply them in practical situations such as clinical diagnosis, so reducing the model complexity is a key consideration. Recently, researchers have explored using networks based on multilayer perceptron (MLP) to improve the performance of computer vision tasks. Tolstikhin *et al.* (14) proposed the MLP mixer, which is based entirely on MLP. Valanarasu *et al.* (15) combined convolution and MLP to develop UNeXt and used it for medical image segmentation. UNeXt mainly uses the tokenized MLP module, which is an improvement of the standard MLP based on the Swin transformer (16). First, the
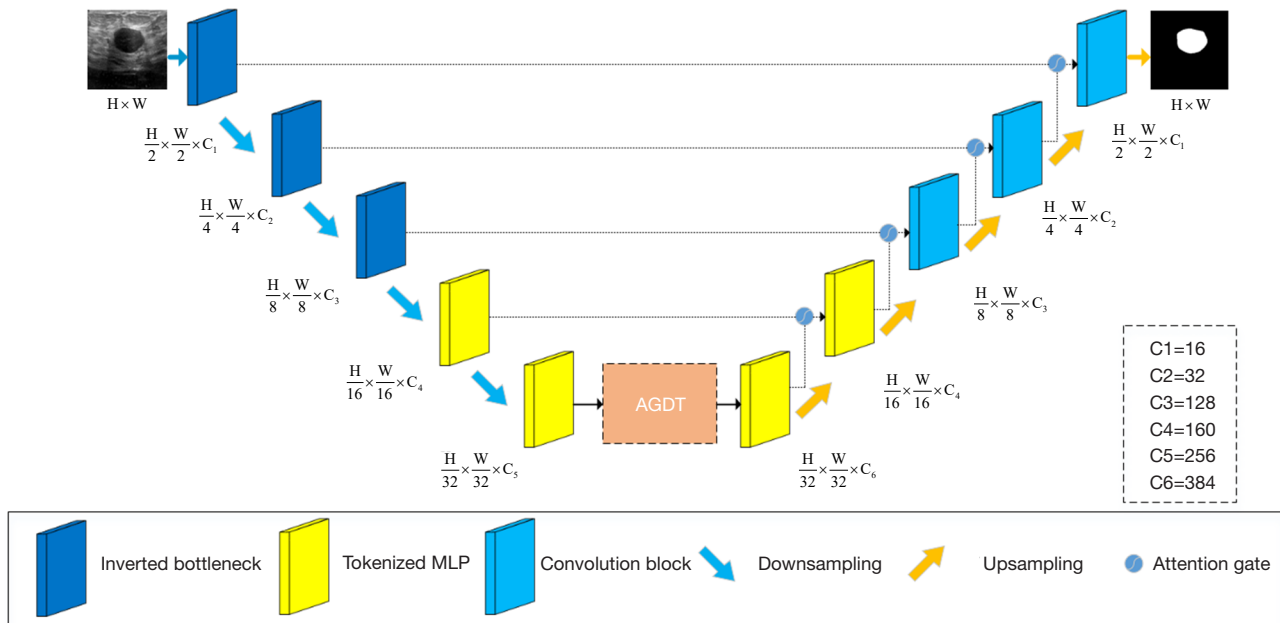
Figure 1 Overview of the proposed GDUNet architecture. H, height; W, width; C, number of channels; GDUNet, attention gate and dilation U-shaped network; MLP, multilayer perceptron; AGDT, attention gate dilation.

shifted MLP is used to move the axis of a specific channel in the feature map via a shift operation and then to focus on some specific locations. Two shifted MLPs are involved, one shifted in width and the other in height, similar to axial attention. Second, to encode the location information of the MLP features, a depth-wise separable convolution is performed between the two shifted MLPs, which involves fewer parameters. Third, a residual connection is made by adding the original features as residual information. The introduction of shift operations in MLP allows for the extraction of local information corresponding to different axial shifts. These MLP-based methods achieve provide a low number of parameters and a fast inference speed, yet being too lightweight may underrepresent task-specific features and thus affect segmentation performance. Nonetheless, these methods offer new ideas for designing lightweight networks.

To overcome the above-mentioned problems, we designed a lightweight breast tumor segmentation method, attention gate and dilation U-shaped network (GDUNet), by drawing on the modules of inverted bottleneck (17), dilated convolution, conditionally parameterized convolutions (CondConv) (18), AG (19), and tokenized MLP. The main contributions of our study are the following: (I) we propose a lightweight breast tumor segmentation method

with attention gating, which captures multiscale contextual information. (II) We conducted extensive experiments on two breast ultrasound datasets, and the results showed that our segmentation method has better a segmentation performance than do the state-of-the-art segmentation methods, with fewer parameters, faster inference speed, and low computational complexity.

Methods

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the Biological and Medical Ethics Committee of Northeastern University (No. NEU-EC-2021B019S). Informed consent was obtained from all patients.

GDUNet has an encoder-decoder architecture, as shown in *Figure 1*. The encoder consists of three inverted bottlenecks and two tokenized MLPs. Since the max pooling for downsampling causes some information loss, we use CondConv with kernel size of 2 and step size of 2 for downsampling. Between the decoder and encoder, we embed an AG dilation (AGDT) block. In the decoder stage, we use tokenized MLP blocks for the first two layers and normal convolutional layers for the last three layers. Upsampling involves bilinear interpolation, which
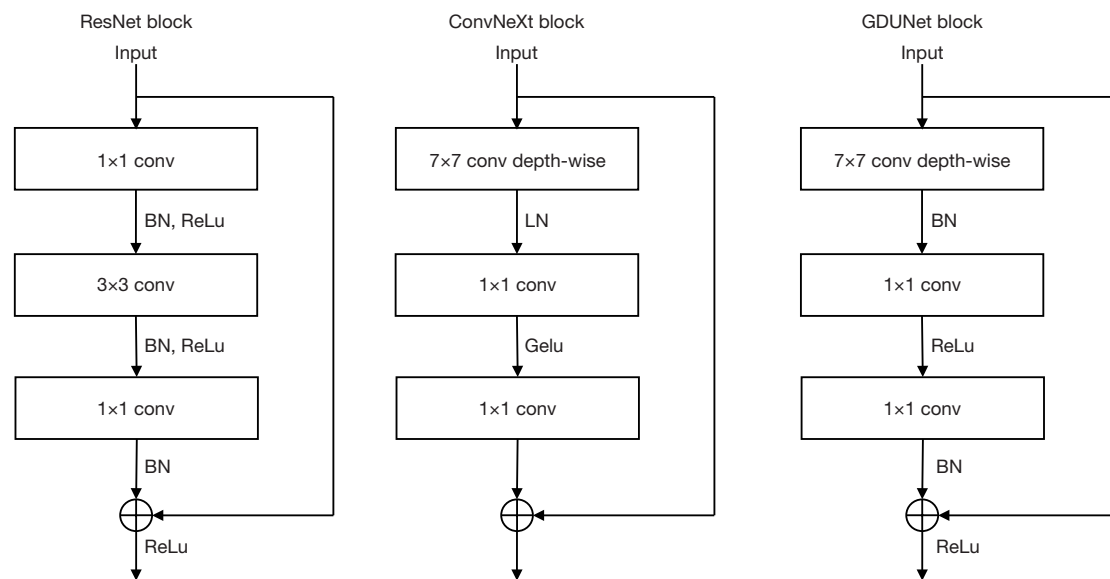
Figure 2 Inverted bottleneck. ResNet, residual network; ConvNeXt, a pure convolutional neural networks model; GDUNet, attention gate and dilation U-shaped network; conv, convolution; BN, batch normalization; ReLU, rectified linear unit; LN, layer normalization; Gelu, Gaussian error linear unit.

reduces the number of parameters while maintaining model performance. We reduce the number of channels per layer compared to UNet. In downsampling, the number of channels per layer of GDUNet is 16, 32, 128, 160 and 256, respectively. Due to the insertion of the AGDT block, in upsampling, the number of channels in each layer is set to 384, 160, 128, 32 and 16, respectively. In the skip connection, we introduce the improved AG.

CondConv

Typically, increasing the size of traditional convolutional kernels and the number of channels can increase the capacity of the model, thereby improving its performance. However, this method can also greatly increase the complexity of the model. Yang *et al.* proposed CondConv (18), which parameterizes the convolutional kernel as a linear combination of n experts, which can increase the model capacity by setting higher n values. This improves performance with almost no increase in model complexity. Inspired by the principle of CondConv, in GDUNet, the max-pooling is replaced by CondConv with a kernel size of 2 and a stride of 2 in the downsampling stage. The experimental results in the Module performance experiment section of this paper showed that the performance was improved after using CondConv. The formula of

CondConv is as follows:

$$Output(x) = \sigma\big((a_1 W_1 + \ldots + a_n W_n) * x\big) \qquad [1]$$

where $\alpha 1 \ldots \alpha n$ are functions of the input learned through gradient descent, σ is an activation function, and $W1 \ldots Wn$ are designed with different convolution kernels.

Inverted bottleneck

GDUNet use large convolution kernels when extracting shallow features, and larger receptive fields can capture more contextual information, which can allow GDUNet to better distinguish the blurred boundaries of breast tumors. Large convolution kernels will increase the number of parameters, but inverted bottleneck can be a good solution to this problem. We refer to ConvNeXt's (17) inverted bottleneck design. To better integrate this module into our method, we made some changes to this module, as shown in *Figure 2*. The inverted bottleneck in ConvNeXt adopts the transformer (20) style, using fewer activation layers and normalization layers. On this basis, we added a normalization layer and an activation layer, similar to the design of the residual network (ResNet) (21). This change could provide an improved network performance. As shown in the GDUNet block in *Figure 2*, the block first uses a 7×7 depth-wise convolution (22) that is followed

2038

Chen et al. GDUNet for breast ultrasound image segmentation
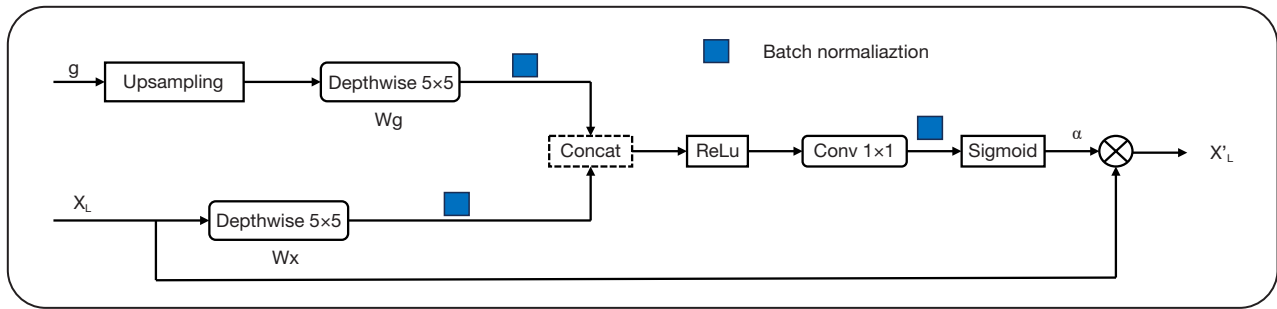
Figure 3 Attention gate. Circles represent multiplication, and arrows indicate the direction of data flow. ReLu, rectified linear unit; Conv, convolution; Concat, concatenation.

by a normalization layer and then uses a 1×1 convolution to amplify the number of channels by 4 times. After an activation layer, a 1×1 convolution kernel is used to reduce to the original number of channels. Subsequently, another normalization layer is added. Finally, residual connections and activation layers are added. Batch normalization (23) is the first choice for most vision tasks, and we also use batch normalization. Through experiments, it was found that batch normalization outperforms layer normalization (24) in our network, as the statistical data in the convolutional layer may vary greatly in space, so normalization using statistical data from the entire layer is usually not ideal. For the activation function, we used the rectified linear unit (ReLU) (25) function. Since the Gaussian error linear unit (Gelu) (26) function requires Gaussian error function calculation, the computational complexity is higher than that of the ReLU function. Through experiments, we found that by using these two functions respectively in our method, the final performance result of the ReLU function was slightly better.

AG

The AG (19) can help the focus attention on salient features useful for a specific task and suppress irrelevant regions in the input image, which is important for improving segmentation performance. For GDNet, instead of the traditional AG, we used a larger convolution kernel. Since breast ultrasound images have the characteristics of noise dispersion and shadow effects, the use of large convolution kernels can better capture the contextual information and thus avoid mistaking noise as salient features. However, using a large convolution kernel will increase the complexity of the model. To avoid increased complexity, channel information and spatial information are extracted separately.

In this way, we improve the segmentation performance while reducing the complexity.

In *Figure 3*, X_L denotes the feature map corresponding to the Lth layer of the encoder, while *g* denotes the feature map of the next layer corresponding to X_L in the decoder and is the gating signal. Since deeper *g* indicates that more knowledge would be learned by the model, the information contained in *g* can be used as a direction of attention for the model to subsequently learn. The space size of *g* is one-half that of X_L, so *g* is first upsampled to make the space size equal to X_L. After upsampling *g*, we perform a 5×5 depth-wise convolution of X_L and *g*, respectively. In this step, we extract only the spatial information of the features, which corresponds to the first step of the depth-wise separable convolution. We then concatenate the features after *g* and X_L convolution and then use 1×1 convolution to extract the channel information. This design can greatly reduce the complexity of the AG. Finally, X_L is multiplied by the calculated attention coefficient α. This is done to superimpose the information in *g* on X_L, and the attention can be directed to the target area. We applied the modified AG to the model, and the performance of the model was greatly improved. The computation of the AG can be summarized as follows:

$$y_1 = Depthwise_{5\times5}\big(Upsample(g)\big) \quad\quad [2]$$

$$y_2 = Depthwise_{5\times5}(x_l) \quad\quad [3]$$

$$y = Conv_{1\times1}\big(Relu\big(Cat(y_1, y_2)\big)\big) \quad\quad [4]$$

$$a = Sigmoid\big(BN(y)\big) \quad\quad [5]$$

$$x_l' = a * x_l \quad\quad [6]$$

where x_l denotes the feature map corresponding to the L layer of the encoder, g denotes the feature map of the next layer
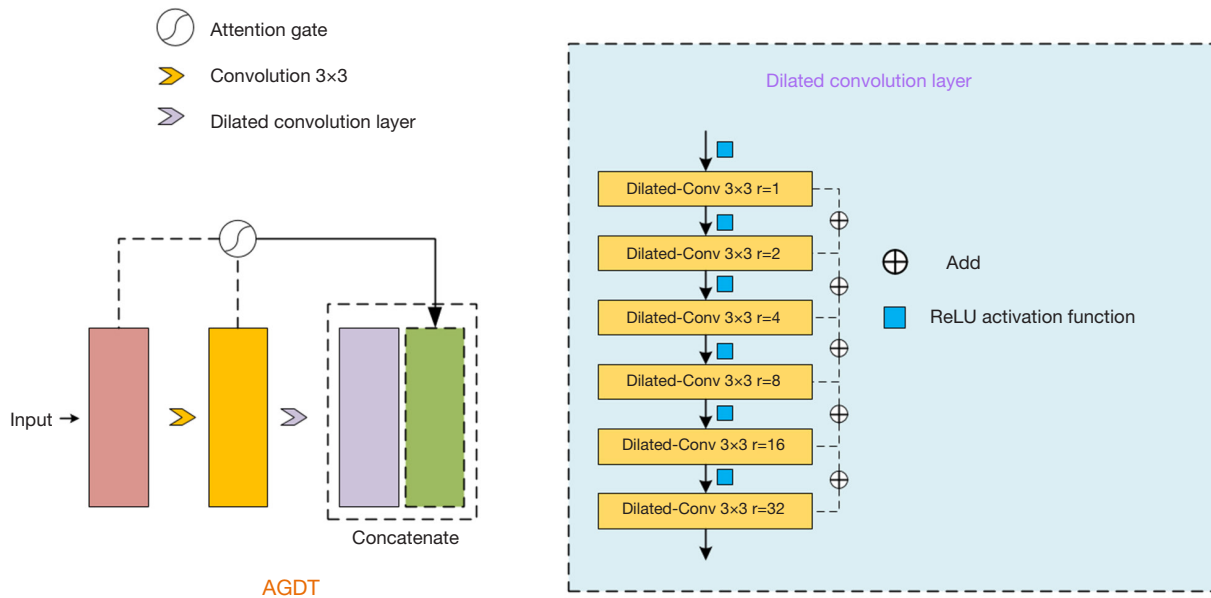
Figure 4 AGDT block. The rectangles of different colors in the left figure represent the different convolutional features. AGDT, attention gate dilation; Dilated-Conv, dilated convolution; ReLU, rectified linear unit.

corresponding to x_l in the decoder, Depthwise denotes the depth-wise convolution, Cat denotes concatenation, BN denotes Batch normalization, and α denotes the attention coefficient.

AGDT block

We designed an AGDT block as shown in *Figure 4*. Specifically, in the AGDT block, the input features are subjected to a dilated convolution layer after a 3×3 convolution operation. An enhanced AG is added between the 3×3 convolutional features and the input features to remove a few irrelevant features in the high-level semantic features. Subsequently, this feature is combined with the multiscale information obtained with the dilated convolution layer. The dilated convolution can obtain a larger receptive field according to the setting of different dilation rates, with the number of parameters remaining unchanged. This dilated convolution layer is superimposed by 6 dilated convolutions, and the dilation rates are 1, 2, 4, 8, 16, and 32. The derived features are salient multiscale contextual information. In addition, a ReLU activation function is added after each dilated convolution.

Loss function

The combination of loss functions used by our method is

a combination of binary cross-entropy (BCE) and Dice loss. This not only enhances boundary sensitivity but also improves overall pixel classification accuracy. The steps for calculating the loss function are as follows:

$$BCE_{loss} = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \log\left(p\left(y_i\right)\right) + \left(1-y_i\right)\cdot \log\left(1-p\left(y_i\right)\right) \quad [7]$$

where y is the label, and $p(y)$ is the predicted probability of the point being positive for all N points.

$$Dice_{loss} = 1 - \frac{2\left|\hat{y}\cap y\right|}{\left|\hat{y}\right|+\left|y\right|} \quad [8]$$

where y represents the pixel label of the real segmented image, and $\hat{y}$ represents the pixel category of the segmented image predicted by the model.

$$L_{BCE_{loss}-Dice_{loss}} = 0.5BCE_{loss}\left(\hat{y},y\right) + Dice_{loss}\left(\hat{y},y\right) \quad [9]$$

where y represents the true value, and $\hat{y}$ is the predicted value.

Results

Datasets

Two datasets of breast ultrasound images were used. Specifically, dataset A is a private dataset consisting of

2040

Chen et al. GDUNet for breast ultrasound image segmentation

Table 1 Statistical differences between the two datasets

Texture feature	P value
Differential entropy	3.27e–04
Correlation	4.02e–04
Sum average	2.59e–84
Difference average	1.35e–08
Difference variance	2.05e–12
Sum variance	9.13e–62
Sum entropy	9.14e–62
Angular second moment	2.70e–34
Contrast	6.54e–17
Entropy	5.17e–47
Homogeneity	5.22e–07
Variance	4.28e–90

878 breast ultrasound images of women aged 25–76 years, with a resolution of 775×580. The images contain tumors of different sizes, and each image contains only one tumor. The ultrasound machines used were the Logiq E9 (GE HealthCare, Chicago, USA) and the Epiq 5 (Philips, Amsterdam, Netherlands). Two experienced radiologists participated in the delineation of the ground truth. Since the noise area of the original dataset is too large and the background area and region of interest (ROI) are unbalanced, we cropped dataset A between the background area and the target area according to a ratio close to 1:1. In all experiments, we used preprocessed dataset A.

Dataset B is a public breast ultrasound dataset published by Zhang *et al.* (27) The dataset consists of 562 breast ultrasound images of women aged 26–78 years, with resolutions of 550×357, 555×490, 546×360, and 600×480 and each image containing only one tumor. The images were acquired and processed by the Second Affiliated Hospital of Harbin Medical University, Qingdao University Hospital, and the Second Hospital of Hebei Medical University using a variety of ultrasound devices: Vivid 7, EUB-6500 (Hitachi, Tokyo, Japan), iU22 (Philips), and Acuson S2000 (Siemens Healthineers, Erlangen, Germany). Four experienced radiologists participated in the ground truth delineation. More details on dataset B and how the final ground truth was obtained can be found in the literature (27).

To analyze the variability between dataset A and dataset B, we used the gray-level co-occurrence matrix method to extract the statistical features of difference entropy, sum entropy, correlation, sum average, difference average, difference variance, sum variance, angular second moments, entropy, contrast, homogeneity, and variance from each image as the statistics to be analyzed. The Mann-Whitney test was then used to obtain the P value of each statistical feature. As can be seen from *Table 1*, the obtained P values were all less than 0.05, indicating a significant statistical difference between dataset A and dataset B. Our segmentation method showed excellent segmentation results on both datasets, indicating that our segmentation method has a strong generalization ability.

Implementation details

The GDUNet was trained on the two datasets using the Adam optimizer. The two datasets were randomly split into training, validation, and test sets in a ratio of 3:1:1. The learning rate, batch size, and training epoch were set to 0.001, 10, and 300, respectively. All input images were resized to 256×256, and a cosine annealing learning rate scheduler with a minimum learning rate of 0.0001 was used. To make the experimental results more convincing, a fivefold cross-validation experimental method was used on the two datasets and was carried out a GTX 1080 Ti GPU (Nvidia, Santa Clara, USA).

Module performance experiment

To better explore the impact of each module on network performance, we designed experiments to test the performance of each module. In the experiments, the modules were filled with regular convolutional layers when removing the tokenized MLP or inverted bottleneck. When removing the CondConv, the a max-pooling layer was used. AG and AGDT could be removed directly without filling. We measured the segmentation results of each method using the Dice similarity coefficient (DSC), as shown in the *Table 2*. It was found that if each module was removed individually, the performance of the GDUNet was reduced. Thus, each module played an important role in the segmentation results of the model. Specifically, after the inverted was replaced with a bottleneck with a conventional convolutional layer alone, the segmentation performance decreased the most, and the DSC decreased by 2.7% and 1.9% on dataset A and

Table 2 Module performance experiment

Network	Dataset A	Dataset B
GDUNet w/o token	0.803	0.920
GDUNet w/o AGDT	0.799	0.918
GDUNet w/o inverted	0.779	0.906
GDUNet w/o AG	0.791	0.919
GDUNet w/o CondConv	0.795	0.918
GDUNet	0.806*	0.925*

*, best result in the table. GDUNet, attention gate and dilation U-shaped network; w/o, without; token, tokenized multilayer perceptron; AGDT, attention gate dilation; inverted, inverted bottleneck; AG, attention gate; CondConv, conditionally parameterized convolutions.

dataset B, respectively. When the tokenized MLP alone was removed, the segmentation performance of the model decreased the least, by 0.3% and 0.5% on the two datasets, respectively. It can be seen from the values in *Table 2* that the decrease in DSC in dataset A was basically greater than that in dataset B. This may be due to significant differences between dataset A and dataset B. The breast cancer ultrasound images included in dataset A are relatively complex, with different tumor sizes and locations, making target segmentation difficult. Each module we introduced plays an important role in the complex dataset, so when a module is missing, there will be a significant difference in the DSC. The images in dataset B have relatively similar feature distributions, the tumor location is often located in the center of the image, and the tumor size changes are relatively symmetrical. Therefore, dataset B was found to be more conducive to target segmentation, with a higher probability of obtaining a high DSC, so the improvement was relatively limited.

Comparison with state-of-the-art methods

To demonstrate the effectiveness of our method, we compared the performance of GDUNet with those of recent widely used medical image segmentation methods, including UNet (6), UNet++ (28), attention UNet (29), Residual-Dilated-Attention-Gate-UNet (RDAU) (11), an asymmetry encoder-decoder architecture using Ghost-Net and U-Net (ghost UNet) (30), semantic guided UNet (SGUNet) (31), medical transformer (MedT) (13), dilate transformer (DT) (32), MLP-based Rapid Medical Image Segmentation Network (UNeXt) (15), and ConvUNeXt (33).

The main performance index results included DSC, area error ratio (AER), Hausdorff error (HE), and mean absolute error (MAE) (*Table 3*). It was found that GDUNet outperformed UNet, attention UNet, and UNet++—which are classical medical image segmentation methods—in all indicators on both datasets. Furthermore, GDUNet demonstrated significant advantages over the transformer-based MedT and DT. ConvUNeXt achieved a good performance in breast lesion segmentation among the segmentation methods compared. Compared to ConvUNeXt, GDUNet had a 1.4% increase and 1.6% in DSC and an 8.17 and 9.97 reduction in HE for datasets A and B, respectively. This indicates that GDUNet better reduced boundary errors. Meanwhile, the AER of the GDUNet on datasets A and B was 0.4% and 1.9% lower than that of ConvUNeXt, respectively, while the MAE of the GDUNet on was 10.61 and 4.53 lower than that of ConvUNeXt, respectively. In *Table 4*, we provide the confidence interval results for both datasets, and all the findings reported in *Table 3* fall within the corresponding confidence intervals.

To better demonstrate the good performance of GDUNet, we extracted the visual segmentation results of lesions with different characteristics on the two datasets for display (*Figure 5*). As apparent in the first row of the two dataset images presented in *Figure 5*, many methods performed poorly in the segmentation of small breast tumors, and the segmentation results included missed detection and false detection. As can be seen in the first row of dataset B in *Figure 5*, ConvUNeXt and UNeXt could not detect small breast tumors; moreover, in the second row of both datasets, except for ConvUNeXt and GDUNet, the other methods produced results for large lesions that were undersegmented; the third row in both datasets shows that GDUNet also demonstrated a good performance in irregular breast ultrasound images. The proposed method can effectively improve the segmentation accuracy of tumors with blurred boundaries. As shown in the fourth row in both datasets in *Figure 5*, for lesions without clear contours, many methods failed to predict tumor margins well, showing both undersegmentation and oversegmentation. The segmentation results suggest that compared to the other methods, GDUNet possesses considerable advantages. According to the overall visualization results, GDUNet achieved the best segmentation results.

The paper by Zhang *et al.* (27) presents a benchmark for breast ultrasound image segmentation, dataset B, and compares the performance of 16 breast ultrasound

**Table 3** Performance comparison with convolutional and transformer baselines

Method	Dataset A				Dataset B			
	DSC	AER	HE	MAE	DSC	AER	HE	MAE
UNet	0.730±0.008	0.632±0.036	97.75±1.87	54.99±1.37	0.840±0.007	0.378±0.032	53.91±2.55	24.60±1.57
UNet++	0.704±0.008	0.702±0.037	102.48±1.93	65.18±1.49	0.849±0.006	0.310±0.019	83.89±3.07	42.83±1.94
Att UNet	0.703±0.008	0.654±0.032	92.25±1.99	44.23±1.35	0.872±0.005	0.245±0.010	72.05±2.93	33.30±1.71
RDAU	0.770±0.008	0.537±0.036	80.96±2.02	34.93±1.17	0.885±0.005	0.223±0.010	54.62±2.35	21.17±1.11
Ghost UNet	0.711±0.009	0.604±0.032	89.92±1.97	51.05±1.42	0.884±0.006	0.207±0.007	39.98±2.22	13.14±0.89
SGUNet	0.782±0.007	0.499±0.030	94.25±2.03	55.33±1.49	0.898±0.004	0.195±0.007	66.35±2.83	28.40±1.47
MedT	0.729±0.008	0.556±0.024	90.10±1.83	51.65±1.34	0.889±0.005	0.219±0.010	45.74±2.33	17.71±1.19
DT	0.769±0.007	0.588±0.036	83.50±1.99	35.74±1.18	0.889±0.004	0.224±0.009	38.89±2.03	14.48±1.05
UNeXt	0.769±0.008	0.546±0.037	80.67±2.09	35.11±1.23	0.902±0.004	0.189±0.007	29.58±1.76	10.71±0.91
ConvUNeXt	0.792±0.008	0.481±0.037	81.94±2.17	39.07±1.37	0.909±0.006	0.163±0.007	27.54±1.84	9.18±0.88
GDUNet	0.806±0.007*	0.477±0.036*	73.77±2.08*	28.46±1.05*	0.925±0.004*	0.144±0.005*	17.57±1.05*	4.65±0.41*

The data in the table are presented as the mean ± standard deviation. *, best result in the table. DSC, Dice similarity coefficient; AER, area error ratio; HE, Hausdorff error; MAE, mean absolute error; UNet, U-shaped network; Att UNet, attention UNet; RDAU, Residual-Dilated-Attention-Gate-UNet; Ghost UNet, Ghost-Net and U-Net; SGUNet, semantic-guided UNet; MedT, medical transformer; DT, dilate transformer; GDUNet, attention gate and dilation U-shaped network.

Table 4 Confidence intervals of the comparison experiment

Method	Dataset A				Dataset B			
	DSC	AER	HE	MAE	DSC	AER	HE	MAE
UNet	0.715, 0.747	0.562, 0.703	94.10, 101.41	52.30, 57.69	0.823, 0.850	0.322, 0.448	48.91, 58.91	21.53, 27.67
UNet++	0.688, 0.720	0.631, 0.774	101.72, 109.31	62.30, 68.16	0.839, 0.863	0.268, 0.342	77.87, 89.89	39.02, 46.62
Att UNet	0.686, 0.721	0.591, 0.719	88.49, 96.27	41.70, 46.99	0.861, 0.883	0.225, 0.264	66.40, 77.89	30.00, 36.71
RDAU	0.754, 0.785	0.467, 0.609	77.01, 84.91	32.64, 37.22	0.875, 0.895	0.205, 0.245	43.20, 52.39	15.87, 20.22
Ghost UNet	0.675, 0.713	0.584, 0.711	86.20, 93.92	48.17, 53.75	0.873, 0.895	0.194, 0.221	35.82, 44.16	11.39, 14.88
SGUNet	0.768, 0.795	0.439, 0.559	90.28, 98.23	52.40, 58.26	0.889, 0.906	0.182, 0.207	60.81, 71.88	25.51, 31.28
MedT	0.715, 0.744	0.508, 0.604	86.523, 93.68	49.02, 54.28	0.879, 0.898	0.200, 0.238	41.16, 50.31	15.38, 20.06
DT	0.755, 0.782	0.518, 0.659	79.61, 87.39	33.43, 38.06	0.884, 0.898	0.201, 0.237	34.96, 42.93	12.44, 16.57
UNeXt	0.752, 0.784	0.475, 0.619	76.59, 84.77	32.70, 37.50	0.894, 0.910	0.175, 0.203	26.15, 33.05	8.95, 12.49
ConvUNeXt	0.776, 0.807	0.409, 0.556	77.67, 86.16	36.36, 41.74	0.898, 0.920	0.148, 0.176	23.93, 31.13	7.46, 10.90
GDUNet	0.792, 0.821	0.406, 0.547	69.70, 77.86	26.41, 30.50	0.918, 0.931	0.135, 0.154	15.51, 19.63	3.86, 5.45

The numbers are the 95% confidence intervals for qualitative segmentation results (lower limit, upper limit). DSC, Dice similarity coefficient; AER, area error ratio; HE, Hausdorff error; MAE, mean absolute error; UNet, U-shaped network; Att UNet, attention UNet; RDAU, Residual-Dilated-Attention-Gate-UNet; Ghost UNet, Ghost-Net and U-Net; SGUnet, semantic-guided UNet; MedT, medical transformer; DT, dilate transformer; GDUNet, attention gate and dilation U-shaped network.
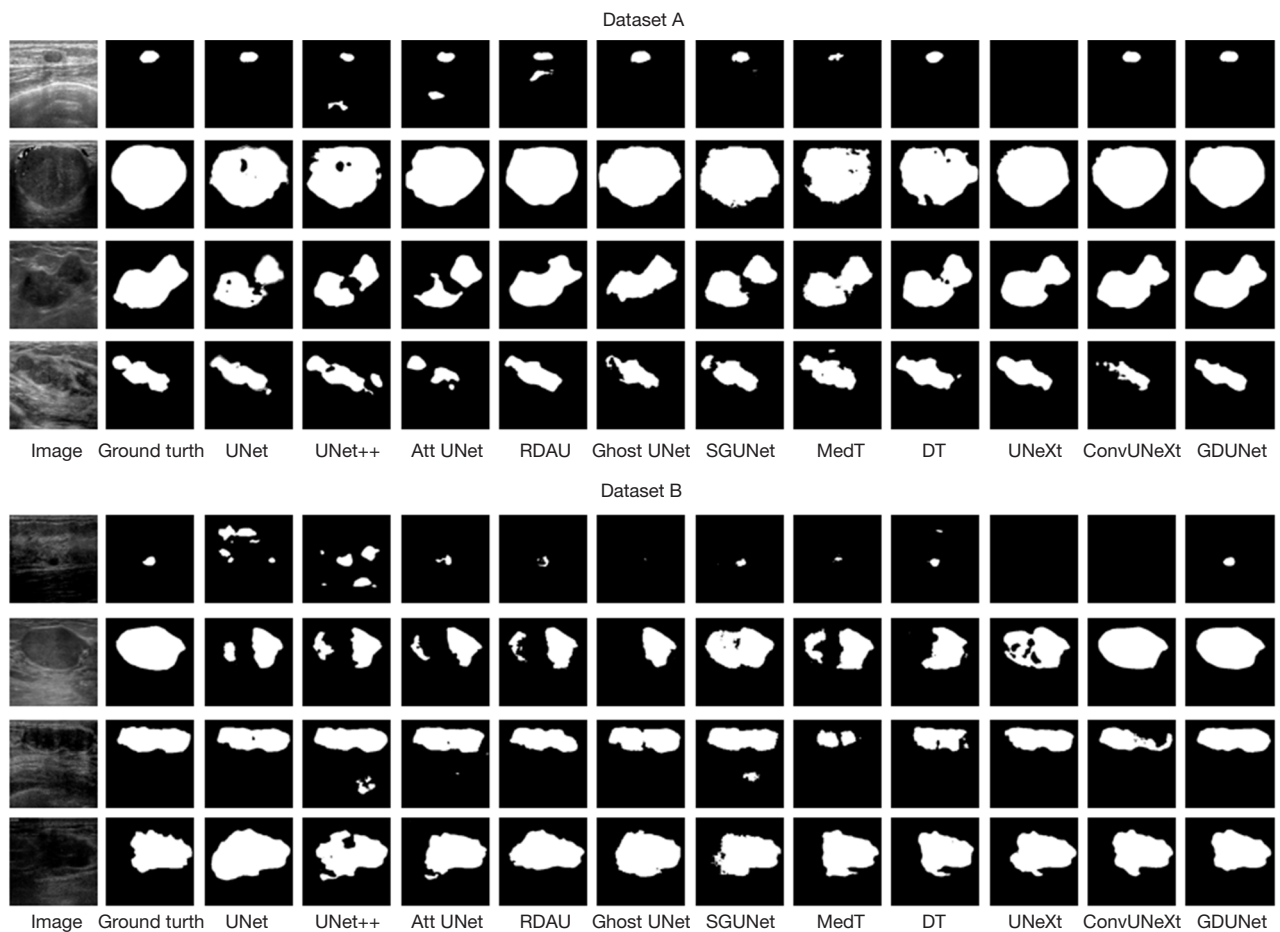
Figure 5 Visual segmentation comparison of the UNet, UNet++, Att UNet, RDAU, Ghost UNet, SGUNet, MedT, DT, UNeXt, ConvUNeXt and GDUNet models. UNet, U-shaped network; Att UNet, attention UNet; RDAU, Residual-Dilated-Attention-Gate-UNet; Ghost UNet, Ghost-Net and U-Net; SGUNet, semantic-guided UNet; MedT, medical transformer; DT, dilate transformer; GDUNet, attention gate and dilation U-shaped network.

segmentation methods on this dataset. In addition to the comparison with the reproduced segmentation methods in *Table 3*, GDUNet was also compared with the segmentation results of each segmentation method listed in Zhang *et al.*'s a paper (27) for public dataset B (*Table 5*). It can be concluded that GDUNet's DSC, AER, and HE were the highest, although the MAE was less distinguished, with a difference of only 0.8 from the lowest MAE. In addition, we extracted four classical models from *Table 5* and performed a fivefold cross-validation experiment on dataset A. The results in *Table 6* indicated these four models performed worse on dataset A than on dataset B, indicating that dataset A is more complex than is dataset B. Additionally, GDUNet had the best performance.

Efficiency analysis

To understand the segmentation efficiency of the models, we separately calculated the number of parameters, inference time, and computational complexity [in one billion floating-point operations per second (GFLOPs)] for each model. The results are shown in *Figure 6*. GDUNet's GFLOPs and inference time were 0.69 and 8.50, respectively, which, among those of all compared models, were only higher than those of UNeXt. The parameter set of GDUNet was 3.42, which is higher than those of MedT and UNeXt. This is mainly due to the ADGT block, which introduces multiple convolution operations to improve performance. Although MedT had a small number of parameters, its GFLOPs and inference time were much higher than

Table 5 Comparison with the state-of-the-art methods in dataset B

| Network	DSC	AER	HE	MAE
FCN-AlexNet (34)	0.84	0.39	25.1	7.1
SegNet (35)	0.89	0.22	21.7	4.5
CE-Net (36)	0.90	0.22	21.6	4.5
SCAN (37)	0.90	0.20	26.9	4.9
DenseU-net (38)	0.88	0.25	25.3	5.5
MultiResUNet (39)	0.91	0.19	18.8	4.1
STAN (40)	0.91	0.18	18.9	3.9*
Fuzzy FCN (41)	0.92	0.14*	19.8	4.2
Huang *et al.* (42)	0.93*	0.15	26.0	4.9
GDUNet	0.93*	0.14*	17.6*	4.7

*, best result in the table. DSC, Dice similarity coefficient; AER, area error ratio; HE, Hausdorff error; MAE, mean absolute error; FCN, fully convolutional network; SegNet, deep fully convolutional neural network architecture for semantic pixel-wise segmentation; CE-Net, context encoder network; SCAN, semantic context-aware network; MultiResUNet, the U-Net Architecture for Multimodal Biomedical Image Segmentation; STAN, small tumor-aware network; GDUNet, attention gate and dilation U-shaped network.

Table 6 Extended experiments for dataset A

Network	DSC	AER	HE	MAE
SegNet	0.71	0.66	90.6	52.5
CE-Net	0.80	0.51	75.1	31.1
DenseU-net	0.76	0.62	91.1	45.3
MultiResUNet	0.74	0.79	114.2	70.8
GDUNet	0.81*	0.48*	73.8*	28.5*

*, best result in the table. DSC, Dice similarity coefficient; AER, area error ratio; HE, Hausdorff error; MAE, mean absolute error; SegNet, deep fully convolutional neural network architecture for semantic pixel-wise segmentation; CE-Net, context encoder network; MultiResUNet, the U-Net Architecture for Multimodal Biomedical Image Segmentation; GDUNet, attention gate and dilation U-shaped network.

those of GDUNet. Compared to ConvUNeXt, GDUNet reduced the number of parameters by 0.08, reduced the computational complexity by a factor of 10, and improved the inference speed by a factor of 4. In comparison with the classic UNet, the number of parameters was 10 times less, the computational complexity was 58 times less, and the inference speed was doubled.

Overall, GDUNet outperformed all other models in terms of segmentation performance. In terms of segmentation efficiency, GDUNet was second only to UNeXt. GDUNet had high efficiency and quality segmentation results, which are valuable features for CAD systems.

Weight analysis of BCE and Dice loss

The weights of BCE and Dice loss were set to 0.5 and 1, respectively, in UNeXt's loss function. We use this loss function setting as well. Since the boundaries of lesions in breast ultrasound images are blurred, giving more weight to Dice loss can help to improve the sensitivity to boundary pixels of the model. For this reason, we designed exploratory experiments for weighting coefficients on dataset A. The experimental results are shown in *Table 7*. It can be seen that when the BCE weight was 0.5 and the Dice loss weight was 1, the quantitative result of GDUNet reaches the best, although the MAE index is a bit higher.
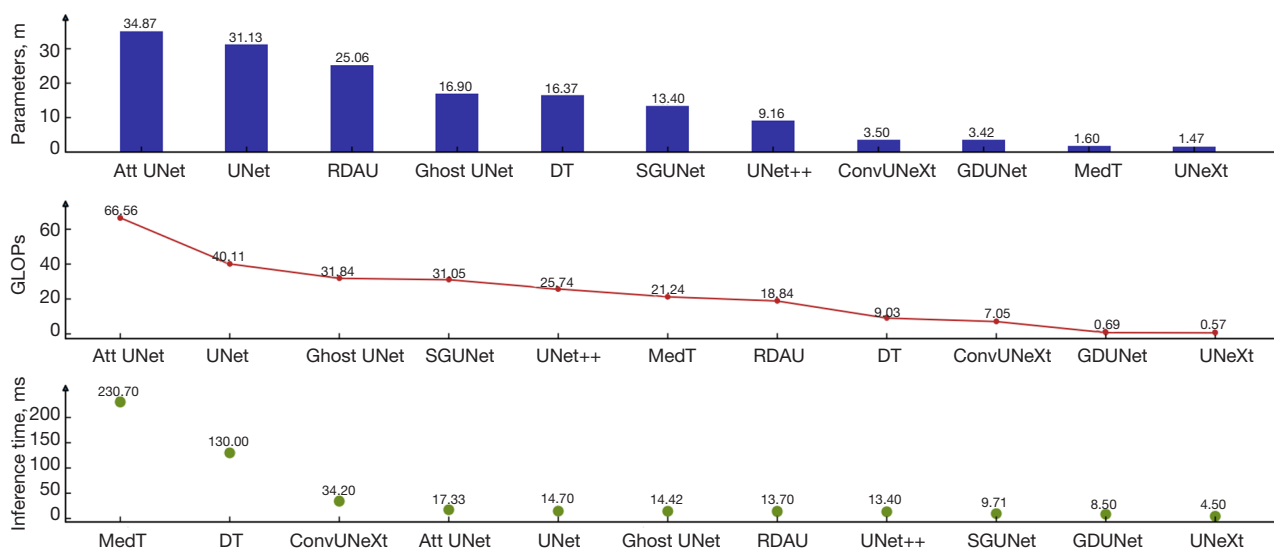
Figure 6 Comparison chart of the efficiency of GDUNet and the other methods. From left to right, the value of the ordinate runs from high to low. Parameters (M), number of parameters; UNet, U-shaped network; Att UNet, attention UNet; RDAU, Residual-Dilated-Attention-Gate-UNet; Ghost UNet, Ghost-Net and U-Net; SGUNet, semantic-guided UNet; MedT, medical transformer; DT, dilate transformer; GDUNet, attention gate and dilation U-shaped network; GLOPs, one billion floating-point operations per second.

Table 7 Results obtained from using varying weight coefficients in dataset A

Loss	DSC	AER	HE	MAE
0.5 × BCE + 0.5 × Dice loss	0.804	0.485	77.91	27.64
0.5 × BCE + Dice loss	0.814*	0.395*	75.73*	30.13
BCE + 0.5 × Dice loss	0.801	0.454	76.41	27.57*
1.5 × BCE + 0.5 × Dice loss	0.803	0.557	82.47	32.00
0.5 × BCE + 1.5 × Dice loss	0.805	0.509	79.76	30.46

*, best result in the table. BCE, binary cross entropy; DSC, Dice similarity coefficient; AER, area error ratio; HE, Hausdorff error; MAE, mean absolute error.

Channel number analysis

To reduce parameters, we attempted to decrease the number of channels. As shown in *Table 8*, we conducted experiments with different channel number settings on the single fold of dataset A. Finally, we used the settings of 16, 32, 128, 160, and 256 channels in UNeXt. The number of channels in the decoder stage changes to accommodate the dilated convolutional modules between the encoder and decoder. According to the experimental results, adding channels did not improve performance but instead increased computational overhead. Setting a lower number of channels will result in insufficient feature representation and therefore performance degradation.

Discussion

For segmentation networks, enlarging the receptive field and attention mechanism are two commonly used optimization strategies. We adopted a larger convolution kernel and used dilated convolution to enlarge the receptive field of the model. To improve attention gating, we used channel and spatial separation methods. In addition to improving model performance, we also adopted methods such as reducing the number of channels and introducing tokenized MLPs to reduce the number of parameters and computational complexity of the model in order to make the model more clinically applicable. The dataset used in this study only included two-dimensional breast ultrasound

Quant Imaging Med Surg 2024;14(2):2034-2048 | https://dx.doi.org/10.21037/qims-23-947

Table 8 Analysis of the number of channels

Method	Number of channels										Parameters (M)	GFLOPs	DSC	AER	HE	MAE
	Encoder					Decoder										
	Layer-1	Layer-2	Layer-3	Layer-4	Layer-5	Layer-5	Layer-4	Layer-3	Layer-2	Layer-1						
GDUNet-L	32	64	128	256	512	768	256	128	64	32	11.2	1.72	0.808	0.566	78.8	32.7
GDUNet	16	32	128	160	256	384	160	128	32	16	3.42	0.69	0.814	0.395	75.7	30.1
GDUNet-M	16	32	64	128	256	384	128	64	32	16	2.82	0.46	0.796	0.456	76.4	27.6
GDUNet-S	8	16	32	64	128	192	64	32	16	8	0.72	0.13	0.784	0.523	80.5	32.1

L, M, and S represent different the channel number settings. Parameters (M), number of parameters; GLOPs, one billion floating-point operations per second; DSC, Dice similarity coefficient; AER, area error ratio; HE, Hausdorff error; MAE, mean absolute error; GDUNet, attention gate and dilation U-shaped network.

images. As there are three-dimensional breast ultrasound images in the medical field, we will examine the application of GDUNet in three-dimensional image breast ultrasound segmentation in the future.

Conclusions

We propose the GDUNet, a lightweight and efficient CNN-based breast ultrasound image segmentation method. GDUNet uses an improved inverted bottleneck and tokenized MLP blocks to form an encoder and uses CondConv for downsampling. To further improve the performance, the AGDT block and the AG are included to enhance the segmentation performance of the model. The experimental results showed that GDUNet not only has high efficiency but also achieved a state-of-the-art segmentation performance on two breast cancer ultrasound datasets.

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims.amegroups.com/article/view/10.21037/qims-23-947/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are

appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the Biological and Medical Ethics Committee of Northeastern University (No. NEU-EC-2021B019S). Informed consent was obtained from all patients.

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin 2011;61:69-90.
2. Zhong L, Shi L, Zhou L, Liu X, Gu L, Bai W. Development of a nomogram-based model combining intra- and peritumoral ultrasound radiomics with clinical features for differentiating benign from malignant in Breast Imaging Reporting and Data System category 3-5 nodules. Quant Imaging Med Surg 2023;13:6899-910.
3. Gonçalves VM, Delamaro ME, Nunes FLS. A systematic review on the evaluation and characteristics of computer-aided diagnosis systems. Revista Brasileira de Engenharia Biomédica 2014;30:355-83.
4. Tsochatzidis L, Zagoris K, Arikidis N, Karahaliou A, Costaridou L, Pratikakis I. Computer-aided diagnosis of mammographic masses based on a supervised content-

based image retrieval approach. Pattern Recognition 2017;71:106-17.

5. Halalli B, Makandar A. Computer aided diagnosis-medical image analysis techniques. Breast imaging 2018;85:85-109.

6. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing; 2015:234-41.

7. Almajalid R, Shan J, Du Y, Zhang M. Development of a deep-learning-based method for breast ultrasound image segmentation. 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, 2018:1103-8.

8. Irfan R, Almazroi AA, Rauf HT, Damaševičius R, Nasr EA, Abdelgawad AE. Dilated Semantic Segmentation for Breast Ultrasonic Lesion Detection Using Parallel Feature Fusion. Diagnostics (Basel) 2021;11:1212.

9. Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122, 2015.

10. Yan Y, Liu Y, Wu Y, Zhang H, Zhang Y, Meng L. Accurate segmentation of breast tumors using AE U-net with HDC model in ultrasound images. Biomedical Signal Processing and Control 2022;72:103299.

11. Zhuang Z, Li N, Joseph Raj AN, Mahesh VGV, Qiu S. An RDAU-NET model for lesion segmentation in breast ultrasound images. PLoS One 2019;14:e0221535.

12. TChen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, Lu L, Yuile AL, Zhou Y. Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306, 2021.

13. Valanarasu JMJ, Oza P, Hacihaliloglu I, Patel VM. Medical transformer: Gated axial-attention for medical image segmentation. Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24. Springer International Publishing; 2021:36-46.

14. Tolstikhin I O, Houlsby N, Kolesnikov A, Beyer L, Zhai X, Unterthiner T, Yung J, Steiner A, Keysers D, Uszkoreit J, Lucic M, Dosovitskiy A. Mlp-mixer: An all-mlp architecture for vision. Advances in neural information processing systems 2021;34:24261-72.

15. Valanarasu JMJ, Patel VM. Unext: Mlp-based rapid medical image segmentation network. Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore,

September 18–22, 2022, Proceedings, Part V. Cham: Springer Nature Switzerland; 2022:23-33.

16. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B. Swin transformer: Hierarchical vision transformer using shifted windows. Proceedings of the IEEE/CVF International Conference on Computer Vision; 2021:10012-22.

17. Liu Z, Mao H, Wu CY, Feichtenhofer C, Darrell T, Xie S. A convnet for the 2020s. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2022:11976-86.

18. Yang B, Bender G, Le QV, Ngiam J. Condconv: Conditionally parameterized convolutions for efficient inference. Advances in Neural Information Processing Systems 2019:32.

19. Schlemper J, Oktay O, Schaap M, Heinrich M, Kainz B, Glocker B, Rueckert D. Attention gated networks: Learning to leverage salient regions in medical images. Med Image Anal 2019;53:197-207.

20. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention is all you need. Advances in Neural Information Processing Systems; 2017:30.

21. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016:770-8.

22. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861, 2017.

23. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. International Conference on Machine Learning. PMLR; 2015:448-56.

24. Ba JL, Kiros JR, Hinton GE. Layer normalization. arXiv preprint arXiv:1607.06450, 2016.

25. Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. Proceedings of the fourteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings; 2011:315-23.

26. Hendrycks D, Gimpel K. Gaussian error linear units (gelus). arXiv preprint arXiv:1606.08415, 2016.

27. Zhang Y, Xian M, Cheng HD, Shareef B, Ding J, Xu F, Huang K, Zhang B, Ning C, Wang Y. BUSIS: A Benchmark for Breast Ultrasound Image Segmentation. Healthcare (Basel) 2022;10:729.

28. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. UNet++:

A Nested U-Net Architecture for Medical Image Segmentation. Deep Learn Med Image Anal Multimodal Learn Clin Decis Support (2018) 2018;11045:3-11.

29. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B, Glocker B, Rueckert D. Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999, 2018.

30. Kazerouni IA, Dooly G, Toal D. Ghost-UNet: An asymmetric encoder-decoder architecture for semantic segmentation from scratch. IEEE Access 2021;9:97457-65.

31. Pan H, Zhou Q, Latecki LJ. Sgunet: Semantic guided unet for thyroid nodule segmentation. 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). IEEE, 2021:630-4.

32. Shen X, Wang L, Zhao Y, Liu R, Qian W, Ma H. Dilated transformer: residual axial attention for breast ultrasound image segmentation. Quant Imaging Med Surg 2022;12:4512-28.

33. Han Z, Jian M, Wang GG. ConvUNeXt: An efficient convolution neural network for medical image segmentation. Knowledge-Based Systems 2022;253:109512.

34. Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation. IEEE Trans Pattern Anal Mach Intell 2017;39:640-51.

35. Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Trans Pattern Anal Mach Intell 2017;39:2481-95.

36. Gu Z, Cheng J, Fu H, Zhou K, Hao H, Zhao Y, Zhang T, Gao S, Liu J. CE-Net: Context Encoder Network for 2D Medical Image Segmentation. IEEE Trans Med Imaging 2019;38:2281-92.

37. Guan L, Wu Y, Zhao J. Scan: Semantic context aware network for accurate small object detection. International Journal of Computational Intelligence Systems 2018;11:951-61.

38. Dong R, Pan X, Li F. DenseU-net-based semantic segmentation of small objects in urban remote sensing images. IEEE Access 2019;7:65347-56.

39. Ibtehaz N, Rahman MS. MultiResUNet : Rethinking the U-Net architecture for multimodal biomedical image segmentation. Neural Netw 2020;121:74-87.

40. Shareef B, Xian M, Vakanski A. STAN: Small tumor-aware network for breast ultrasound image segmentation. Proc IEEE Int Symp Biomed Imaging 2020;2020:1469-73.

41. Huang K, Zhang Y, Cheng HD, Xing P, Zhang B. Semantic segmentation of breast ultrasound image with fuzzy deep learning network and breast anatomy constraints. Neurocomputing 2021;450:319-35.

42. Huang K, Zhang Y, Cheng H D, Xing P, Zhang B. Semantic Segmentation of Breast Ultrasound Image with Pyramid Fuzzy Uncertainty Reduction and Direction Connectedness Feature. 2020 25th International Conference on Pattern Recognition (ICPR). IEEE, 2021:3357-64.