

Received 22 April 2024, accepted 22 May 2024, date of publication 31 May 2024, date of current version 10 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3407795

RESEARCH ARTICLE

BFG&MSF-Net: Boundary Feature Guidance and Multi-Scale Fusion Network for Thyroid Nodule Segmentation

JIANUO LIU¹, JUNCHENG MU¹, HAORAN SUN¹, CHENXU DAI¹,
ZHANLIN JI^{1,2}, (Member, IEEE), AND IVAN GANCHEV^{3,4,5}, (Senior Member, IEEE)

¹Hebei Key Laboratory of Industrial Intelligent Perception, North China University of Science and Technology, Tangshan 063210, China

²College of Mathematics and Computer Science, Zhejiang A&F University, Hangzhou 311300, China

³Telecommunications Research Centre (TRC), University of Limerick, Limerick, V94 T9PX Ireland

⁴Department of Computer Systems, University of Plovdiv "Paisii Hilendarski," 4000 Plovdiv, Bulgaria

⁵Institute of Mathematics and Informatics—Bulgarian Academy of Sciences, 1040 Sofia, Bulgaria

Corresponding authors: Chenxu Dai (daichenxu@ncst.edu.cn), Zhanlin Ji (zhanlin.ji@gmail.com), and Ivan Ganchev (ivan.ganchev@ul.ie)


This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFE0135700; in part by Bulgarian National Science Fund (BNSF) under Grant KP-06-IP-CHINA/1; and in part by the Telecommunications Research Centre (TRC), University of Limerick, Ireland.

ABSTRACT Accurately segmenting thyroid nodules in ultrasound images is crucial for computer-aided diagnosis. Despite the success of Convolutional Neural Networks (CNNs) and Transformers in natural images processing, they struggle with precise boundaries and small-object segmentation in ultrasound images. To address this, a novel BFG&MSF-Net model is proposed in this paper, utilizing four newly designed modules: (1) a Boundary Feature Guidance Module (BFGM) for improving the edge details capturing; (2) a Multi-Scale Perception Fusion Module (MSPFM) for enhancing the information capture by combining a novel Positional Blended Attention (PBA) with the Pyramid Squeeze Attention (PSA); (3) a Depthwise Separable Atrous Spatial Pyramid Pooling Module (DSASPPM), used in the bottleneck to improve the contextual information capturing; and (4) a Refinement Module (RM) optimizing the low-level features for better organ and boundary identification. Evaluated on the TN3K and DDTI open-access datasets, BFG&MSF-Net demonstrates effective reduction of boundary segmentation errors and superior segmentation performance compared to commonly used segmentation models and state-of-the-art models, which makes it a promising solution for accurate thyroid nodule segmentation in ultrasound images.

INDEX TERMS Ultrasound image, thyroid nodule, segmentation, deep learning, boundary feature guidance, multi-scale fusion.

I. INTRODUCTION

The thyroid, a vital endocrine organ, plays an indispensable role in maintaining normal physiological functions of the human body [1], [2]. Among various nodular lesions, thyroid nodules are highly prevalent endocrine disorders growing within the thyroid. Moreover, they represent one of the most commonly diagnosed abnormalities, necessitating regular thyroid examinations due to their potential malignancy [3],

The associate editor coordinating the review of this manuscript and approving it for publication was Carmelo Militello .

[4]. Utilizing ultrasound imaging, with its distinct advantages such as non-invasiveness, user-friendliness, and cost-effectiveness, has become the predominant method for thyroid examinations [5].

In medical practice, the interpretation of thyroid ultrasound images involves experienced clinicians who must assess multiple critical features, including shape, edges, composition, echogenicity, and abnormal lesions [6], [7]. However, this subjective diagnostic process introduces variability among observers and heavily relies on the clinician's years of expertise. Challenges such as low contrast and speckle noise

in ultrasound images further complicate the accurate diagnosis and decision-making process. Therefore, there is an urgent need for artificial intelligence (AI)-based computer-aided diagnostic systems, which to assist in thyroid disease diagnosis [8], [9], [10]. Automated segmentation of thyroid nodules is a fundamental component in developing intelligent diagnostic systems and holds significant value for ultrasound-guided thyroid fine-needle aspiration or nodule excision procedures [11].

Currently, image segmentation technology has sparked widespread research interest in the field of medical imaging [12], [13]. Unlike natural image segmentation, medical images often come with speckle noise, and the lesion areas typically have fuzzy boundaries, making it challenging for traditional models to accurately identify and segment the lesions based on low-level features [14]. Due to the lack of detailed information in the images, relying solely on semantic features usually struggles to obtain accurate boundary information [15]. Convolutional Neural Networks (CNNs) can fully analyze and extract complex texture features in medical image processing tasks [16], and can achieve satisfactory results in improving processing accuracy and generalization ability, which provides strong support for clinical diagnosis and treatment [17], [18].

Zhou et al. [19] cleverly combined Conditional Random Fields (CRFs) with graph-cut methods and refined segmentation by integrating spatial correlations. Subsequently, Fully Convolutional Networks (FCNs) replace the fully connected layer behind the traditional CNN with a convolutional layer, so that the output of the network is a thermal map rather than a category.

Chen et al. [22] proposed DeepLab, which uses an extended convolution and the Atlas Spatial Pyramid Pooling (ASPP) architecture to capture multi-scale contextual information [23].

In recent years, researchers have made many attempts and explorations in the use of CNNs for thyroid nodule segmentation of ultrasound images. Ying et al. [24] proposed a thyroid nodule segmentation model based on a cascaded CNN for more accurate localization of the thyroid nodules in images. Kumar et al. [25] proposed a multi-prong CNN to enlarge the receptive field and improve the accuracy of thyroid nodule segmentation by expanding the convolutional layer. Pan et al. [26] proposed a Semantic Guided U-Net (SGUNET) network to solve the problem of shallow features in the decoder being susceptible to noise interference from ultrasound images. Zhang et al. [27] proposed a method for thyroid nodule segmentation using a cascaded U-Net architecture. The network initially employs U-Net [28] for rough nodule localization and then performs fine segmentation through a second U-Net to obtain the final result. Considering that most researchers designed models for the location, size, and susceptibility to noise of thyroid nodules, but ignored the blurred boundaries of thyroid nodules in ultrasound images which made it difficult to distinguish the background from the target region, the existing thyroid nodules segmentation

networks could not describe the contour of nodules well.

To address the shortcomings of the current models for thyroid nodule segmentation, we propose a novel segmentation model, named BFG&MSF-Net, based on boundary-guided multi-scale fusion. By means of boundary guidance, the edge details are mainly captured, so that the network training pays attention to the edge division in the segmentation process, and the influence of background on the segmentation of thyroid nodules is reduced to a greater extent. The proposed model is built on U-Net [28] by adopting ConvNeXt [29], [30] as a backbone.

To address the limitations of traditional U-Net based models in feature extraction and information fusion, the proposed BFG&MSF-Net model introduces the following novel designs:

- 1) To cope with the issue of edge detail loss during feature extraction due to downsampling, a newly designed Boundary Feature Guidance Module (BFGM) is introduced for cleverly combining local edge information and global positional semantic information. By enhancing the edge feature extraction, BFGM effectively improves the model's ability to capture edge details.
- 2) A newly designed Multi-Scale Perception Fusion Module (MSPFM) is proposed as a replacement of traditional feature fusion modules. Its innovative design significantly improves the model's ability to extract thyroid feature information of different scales in the images by combining a novel Positional Blended Attention (PBA) with the Pyramid Squeeze Attention (PSA) [31], allowing the model to outperform state-of-the-art models in image-processing tasks.
- 3) In the intermediate bottleneck, a novel Depthwise Separable Atrous Spatial Pyramid Pooling Module (DSASPPM) is incorporated to expand the receptive field of the convolutional layer, fully capture contextual information of thyroid nodules of different scales, and integrate higher-level semantic information.
- 4) Finally, a newly designed Refinement Module (RM) is introduced as the last feature extractor in the model, aiming to optimize low-level features for better identification of thyroid glands and their boundaries.

Experimental results, obtained on two open-access datasets, demonstrate that the proposed BFG&MSF-Net model outperforms all similar models considered and allows to effectively reduce the boundary segmentation errors in thyroid nodule image segmentation tasks.

The rest of this paper is organized as follows. Section II introduces related work. Section III provides a detailed description of the overall architecture of the proposed model, along with the design of its novel modules. Section IV presents performance evaluation of the proposed model in comparison to other popular models, based on two open-access thyroid nodule datasets using standard evaluation metrics. Finally, Section V summarizes the main findings of this paper and suggests directions for future research work.

II. RELATED WORK

With the continuous development of image segmentation technology, numerous works have emerged in the field of medical image segmentation, contributing major breakthroughs. Methods for thyroid nodule segmentation in medical images can generally be classified into two categories: non-CNN and CNN based segmentation methods [32].

A. NON-CNN BASED MEDICAL IMAGE SEGMENTATION

Non-CNN based segmentation includes methods rooted in computer vision and image processing technologies, designed to delineate structures or targets in medical images for subsequent analysis and diagnosis. These methods include region-based [33], threshold-based [34], and deformable model based decomposition [35], each offering unique advantages and applicability.

Region-based segmentation involves a gradual expansion or reduction of the segmented area. This type of methods starts operating on a specific image region and incrementally adds pixels that adhere to predefined rules to the target area or removes pixels that deviate from the rules, thus achieving the purpose of combining other pixels of similar quality. From a global point of view, this type of methods divides the whole image into different sub-regions.

The threshold-based methods are suitable for images with different grayscale of the background and target. The basic idea is to calculate one or more grayscale thresholds according to the grayscale features of the image. The pixel gray value obtained from the image is compared with the calculated threshold one by one, and the pixels are divided into appropriate categories according to the comparison results.

The deformable model based decomposition is based on the classification and geometric shape of the point cloud data. It is a method of comparing point clouds with known geometric shapes (cylinders, cones, spheres, etc.), classifying points with the same mathematical characteristics into one group, and dividing the known geometric shapes within the point clouds. In addition to being less affected by noise, the computation speed is also higher than splitting by edge information. This adjustment process considers various pixel characteristics such as brightness, texture, gradient, and so on.

Although non-CNN based segmentation methods have certain practicability in medical image processing, especially in scenarios requiring low computational complexity and fast processing, these methods still have certain limitations in dealing with scenarios involving complex computation and poor image quality. However, the advent of CNN technology has led to an increasing preference for deep learning methods in modern medical image segmentation. These methods excel in handling complex medical images and irregular structures by automatically extracting features and training on large datasets, consequently yielding more accurate segmentation results.

B. CNN BASED MEDICAL IMAGE SEGMENTATION

Non-CNN based segmentation methods typically rely on prior medical knowledge and necessitate manual intervention by experienced doctors, which may introduce subjective differences and consequently lead to errors. Therefore, addressing this issue has become increasingly urgent. With the continuous advancement of large-scale data collection and computer technology, deep learning CNNs have exhibited strong potential in the field of computer vision and have been extensively employed in practical projects [36].

In the medical domain, deep learning technology has found widespread application in tasks such as cancer- and tumor detection, classification, and segmentation. Deep learning methods yield highly accurate results and significantly contribute to assisting doctors in making precise diagnoses [37]. Research indicates that employing deep learning for automating tumor tissue segmentation not only enhances work efficiency but also aids doctors in making accurate judgments. Consequently, this technology holds promising application prospects in the medical field.

The U-Net model proposed by Ronneberger et al. [28] represents an image segmentation network grounded on deep learning principles, achieving precise image segmentation through an encoder-decoder architecture and skip connections. U-Net has demonstrated outstanding performance across various medical image segmentation tasks. Zhou et al. [20] introduced UNet++, comprising a series of U-Nets and decoders of varying depths. These decoders are densely interconnected at the same resolution through redesigned skip connections [21]. Despite its enhanced performance, the UNet++ model is characterized by complexity, necessitates additional learnable parameters, and includes redundant components for specific tasks.

Oktay et al. [38] introduced an additional focusing gate into a U-shaped structure for medical image segmentation, coupled with an attention gate (AG) mechanism to implicitly generate soft region proposals. The resultant model, called Att U-Net, accentuates salient features beneficial for specific tasks, allowing it to focus on crucial aspects during medical image segmentation by selectively fusing information during the feature fusion stage. Diakogiannis et al. [39] proposed deep residual U-Net (Res-Unet) model, still grounded on the U-Net architecture. A series of stacked residual units replace ordinary neural units as basic blocks to construct a deep Res-Unet, effectively increasing the number of network training layers. However, as the network depth increases, the training time extends significantly.

ResNet (Residual Neural Network) is a deep convolutional neural network, proposed by He et al. [40], which performs well in image recognition tasks. The core idea is to introduce residual blocks that pass input directly into later layers through skip connections. This design solves the problem of gradient disappearance and gradient explosion in deep neural networks, allowing the network to be trained more stably

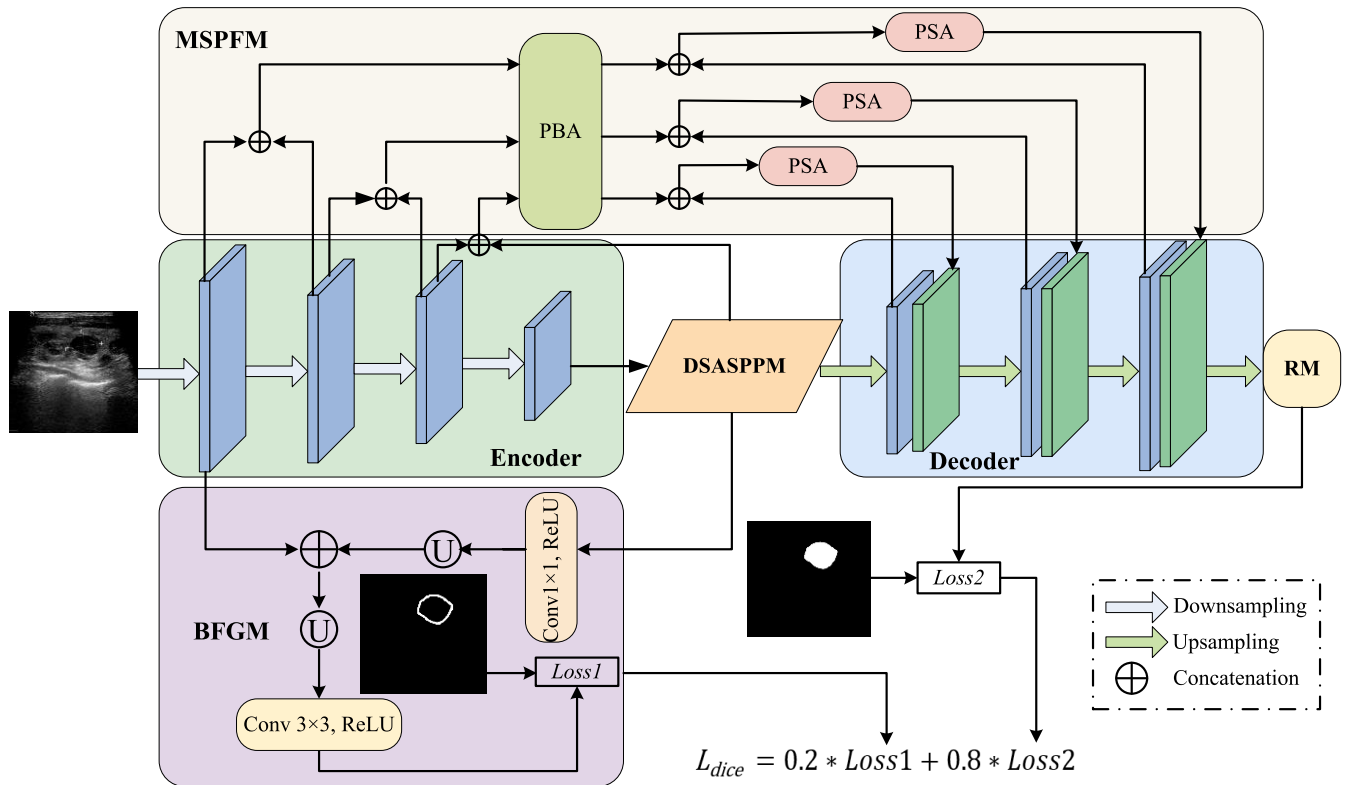


FIGURE 1. The proposed BFG&MSF-Net model.

and be stacked with more layers as to extract more complex features.

Researchers also explore integrating a self-attention mechanism into CNNs to enhance network performance. Rajakumar et al. [41] introduced the Seg-Net model with a lightweight and convenient network structure, which improves the training speed, but sacrifices accuracy to a considerable extent.

CNN-based medical image segmentation methods excel in learning image features, effectively integrating low-level and high-level semantic information, and autonomously and accurately obtaining segmentation results. This reduces manual intervention, saves doctors' time, and provides a foundation for further tumor analysis. Facilitating the formulation of subsequent surgical plans, the adoption of deep learning methods for medical image segmentation significantly advances computer-aided diagnosis and furnishes a robust framework for clinical application.

III. PROPOSED MODEL: BFG&MSF-NET

In the context of image segmentation tasks performed by deep learning models, edges are often considered to occupy a relatively small proportion due to their smaller area. However, this paper takes a different perspective, emphasizing the crucial role of edge features in the decoding process of deep learning models. In contrast to traditional views, we argue that by fully leveraging edge features and positional

information to constrain the decoding process, the accuracy of thyroid nodule image segmentation can be effectively improved.

We propose a multi-scale fusion segmentation network model, named BFG&MSF-Net, based on boundary features. The importance of edge features is fully considered in the design of the model, whereby more accurate image segmentation is realized by combining local edge information with global positional semantic information. Experiments show that BFG&MSF-Net is superior to existing models in reducing boundary segmentation errors and improving the performance of segmentation tasks.

A. OVERALL ARCHITECTURE

The overall architecture of the proposed BFG&MSF-Net model, shown in Figure 1, consists of six main modules: an encoder, a decoder, a novel Boundary Feature Guidance Module (BFGM), a newly designed Multi-Scale Perception Fusion Module (MSPFM), a novel Depthwise Separable Atrous Spatial Pyramid Pooling Module (DSASPPM), and a newly designed Refinement Module (RM). Taking full advantage of ConvNeXt's powerful ability to extract spatial features, texture features, and global context information, we use the most advanced ConvNeXt as a backbone of BFG&MSF-Net for different levels of image feature extraction. BFGM fuses local edge information extracted by the encoder with global positional semantic information to

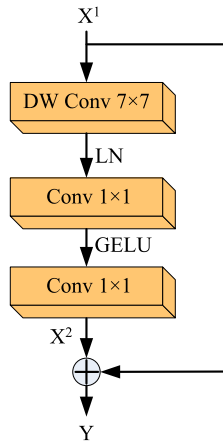


FIGURE 2. The ConvNeXt [29], utilized as a backbone network in the encoder of the proposed BFG&MSF-Net model.

acquire unit-specific edge features. MSPFM enhances the model’s ability to capture important information in images. DSASPPM expands the receptive field of the convolutional layer, fully captures contextual information of thyroid nodules of different scales, and integrates higher-level semantic information. RM augments internal learning to capture lower-level hidden features.

B. ENCODER

To enhance the model’s performance in thyroid nodule segmentation, we chose ConvNeXt as a backbone network for the encoder. This decision was based on the fact that ConvNeXt is pretrained on a large-scale image dataset, allowing it to learn generic image feature representations and better generalize to the relatively smaller thyroid nodule datasets. The use of ConvNeXt provides our model with robust feature extraction capabilities and establishes a solid foundation for the thyroid nodule segmentation task. Illustrated in Figure 2, ConvNeXt is an advanced CNN, inspired by the hierarchical architecture of the Swin Transformer. The ConvNeXt output Y is obtained as follows:

$$X^2 = Conv_{1 \times 1}(GELU(Conv_{1 \times 1}(LN(DWConv_{7 \times 7}(X^1)))))) \quad (1)$$

$$Y = X^1 + X^2 \quad (2)$$

where X^1 denotes the input feature map. First, a 7×7 depthwise (DW) convolution is used, aiming to preserve the global receptive field and emphasize the understanding of the entire image, rather than focusing solely on local feature information, [41]. Then, a LayerNorm (LN) is applied for normalization processing to solve the problem of excessive statistical deviation. After that, two 1×1 ordinary convolutions are used to refine the extracted features. A GELU activation function is utilized between these two convolutions to accelerate the model convergence. Finally, the obtained feature map X^2 is fused with the input feature map X^1 ,

by adding their elements one by one, to get the output feature map Y .

C. BOUNDARY FEATURE GUIDANCE MODULE (BFGM)

To address the issue of edge detail loss caused by down-sampling in image segmentation tasks, a newly designed Boundary Feature Guidance Module (BFGM) is introduced here. BFGM ingeniously combines local edge information with global positional semantic information to enhance the model’s ability to capture edge features. The design philosophy of BFGM is based on our belief that leveraging edge features and positional information to constrain the decoding process in deep learning models contributes to improving the accuracy of image segmentation. By guiding the model’s attention to the image boundaries, BFGM helps enhance the model’s segmentation performance in edge regions.

From the visualization studies of neural networks [19], it is evident that shallow features extracted by a network model largely preserve the edge information of images. However, these edge details are acquired solely through local information, lacking a comprehensive understanding of advanced semantics and positional information. In contrast, deep-layer features of the network, owing to their larger receptive field, encompass richer semantic and positional information. Therefore, BFGM integrates shallow features $M^{(1)}$ containing local edge information and advanced features $M^{(5)}$ containing semantic and positional information. On top of this integration, edge features are extracted using a convolutional layer with a kernel size of 3×3 . The specific computational process is described below.

First, a 1×1 convolution is applied to transform the channel dimensions of $M^{(5)}$. Subsequently, after performing Batch Normalization (BN) and Rectified Linear Unit (ReLU) activation, upsampling is applied. Following this, the feature fusion with $M^{(1)}$ results in $O^{(1)}$, as shown below:

$$O^{(1)} = M^{(1)} + Upsample(ReLU(BN(Conv_{1 \times 1}(M^{(5)})))) \quad (3)$$

where $Conv_{1 \times 1}$ denotes a 1×1 convolution, aimed at adjusting the channel dimension of $M^{(5)}$ to match that of $M^{(1)}$; BN denotes Batch Normalization, which accelerates training convergence and enhances generalization capability; ReLU represents the activation function, introducing a non-linear factor while effectively avoiding the issue of vanishing gradients; $Upsample$ employs bilinear interpolation for upsampling, intending to adjust the image dimensions of $M^{(5)}$ to match those of $M^{(1)}$.

Next, the fused result is upsampled, followed by the extraction of edge information using a 3×3 convolutional kernel, thereby obtaining the edge prediction result $O^{(Edge)}$, as follows:

$$O^{(Edge)} = ReLU(BN(Conv_{3 \times 3}(Upsample(O^{(1)})))) \quad (4)$$

Finally, the Binary Cross-Entropy with Logits Loss (BCE-WithLogitsLoss) function [44] is employed as a loss function to calculate the error loss ($Loss_1$) between the predicted result $O^{(Edge)}$ and the true edge information $Y^{(Edge)}$.

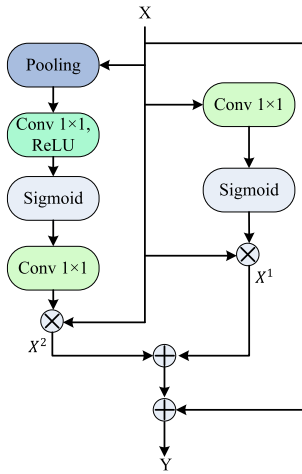


FIGURE 3. The newly designed Positional Blended Attention (PBA) submodule of MSPFM, utilized by the proposed model.

D. MULTI-SCALE PERCEPTION FUSION MODULE (MSPFM)

The innovative work and novel strategy of using a multi-scale perception fusion (MSPF) to replace the traditional feature fusion allows to achieve significant progress in enhancing the performance of image processing tasks by leveraging the synergistic action of two key components of the newly designed MSPF module (MSPFM), namely a novel Positional Blended Attention (PBA) and the Pyramid Squeeze Attention (PSA) [31].

Initially, features from adjacent layers in the encoder are fused to explore relationships between channel dimensions. Subsequently, PBA is employed to enhance long-range dependency relationships between deep semantic features. This allows the model to effectively focus on information from different positions in the images. The hybrid attention mechanism combines spatial and channel information to enable the model to focus on the importance of a specific location in the images, thereby enhancing the perception of key information. The output is fused with the corresponding output of the decoding layer for a second round of feature fusion, and then a new decoding layer is formed by the PSA convolutional blocks.

1) POSITIONAL BLENDED ATTENTION (PBA) SUBMODULE

The main function of the PBA submodule of MSPFM is to integrate channel attention and spatial positional information. PBA can selectively emphasize the importance of different positions in images by applying channel attention weight to the encoder feature map and introducing spatial positional information. This mixed attention mechanism makes the model more flexible and accurate in learning image semantic information and optimizes the perception of local buildings. The PBA architecture is shown in Figure 3.

First, a channel-wise transformation of the input X is performed through a 1×1 convolution. Subsequently, a Sigmoid activation function is applied to map the result into

the range $[0,1]$, obtaining channel attention weights. Finally, an element-wise multiplication is used to multiply the original input with the channel attention weights, yielding refined feature X^1 , emphasizing important information along the channel dimension, as specified in (5).

Simultaneously, the input X undergoes pooling through a MaxPool operation to capture some spatial information. After that, a 1×1 convolution is applied to the pooled result along the channel dimension. Next, non-linearity is introduced through a ReLU activation function, and finally, a Sigmoid activation function maps the result into the range $[0,1]$, obtaining spatial attention weights. Ultimately, an element-wise multiplication is employed to multiply the original input X with the spatial attention weights, producing refined feature X^2 , emphasizing important information along the spatial dimension, as specified in (6):

$$X^1 = X \odot \text{Sig}(\text{Conv}_{1 \times 1}(X)) \quad (5)$$

$$X^2 = X \odot \text{Conv}_{1 \times 1}(\text{Sig}(\text{ReLU}(\text{Conv}_{1 \times 1}(\text{MaxPool}(X)))))) \quad (6)$$

where Sig represents the Sigmoid function, \odot denotes an element-wise multiplication, MaxPool represents maximum pooling, and ReLU represents the activation function introducing a non-linear factor while effectively avoiding the problem of gradient vanishing. It is noteworthy that these two steps are performed simultaneously.

Finally, the refined features X^1 and X^2 are added to the original features X in the form of a residual connection to obtain the final output Y , as follows:

$$Y = X + X^1 + X^2 \quad (7)$$

2) PYRAMID SQUEEZE ATTENTION (PSA) SUBMODULE

In PSA [31], channel splitting is performed first as to extract multi-scale features for spatial information on each channel feature map. Then, a SEWeight module is utilized to extract channel attention for different scale feature maps, obtaining channel attention vectors for each scale. Next, a SoftMax function is applied to the multi-scale channel attention vectors for feature recalibration, yielding new weights for multi-scale channel interaction. Finally, the recalibrated weights are element-wise multiplied with the corresponding feature maps to output a feature map where multi-scale feature information is attentively weighted. As a result, the output feature map exhibits richer multi-scale information representation capability.

E. DEPTHWISE SEPARABLE A TROUS SPATIAL PYRAMID POOLING MODULE (DSASPPM)

Introducing a novel Depthwise Separable Atrous Spatial Pyramid Pooling Module (DSASPPM) into the bottleneck of the proposed model allows it to simultaneously fuse global and local information in specific dimensions. Shown in Figure 4, DSASPPM significantly enhances the perception of the model at different scales through dilated convolutions and pooling operations, which is crucial for handling the

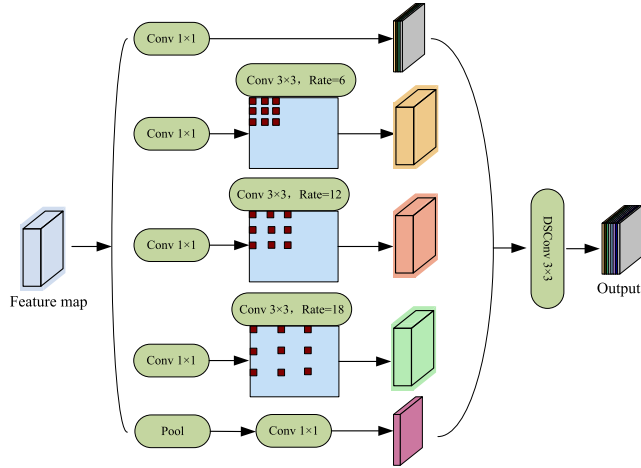


FIGURE 4. The novel Depthwise Separable Atrous Spatial Pyramid Pooling Module (DSASPPM), utilized by the proposed model.

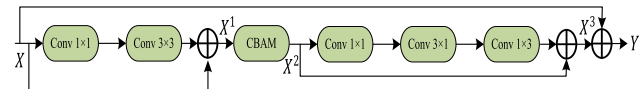


FIGURE 5. The novel Refinement Module (RM), utilized by the proposed model.

diversity and complexity of thyroid nodules. Additionally, since the bottleneck layer is typically a critical position in the network where feature dimensions are reduced, introducing DSASPPM helps maintain feature richness within a relatively small space. This strategy enables powerful integration of multi-scale features under more efficient computational conditions.

The multi-scale information fusion of DSASPPM contributes to more accurately capturing crucial information in images, thereby enhancing the model’s segmentation accuracy. In thyroid segmentation tasks, considering both local and global information is crucial for accurately extracting details such as nodule edges. This comprehensive multi-scale perception makes the model more adaptable to different nodule features, thereby improving the overall performance of thyroid nodule segmentation.

F. REFINEMENT MODULE

Image segmentation faces a series of challenges, including dealing with boundary blurriness and irregular shapes, and the inability to detect small organs. The newly designed Refinement Module (RM) is located after the decoder of the proposed model, with the main function to capture spatial and channel information so as to locate pixels more accurately. RM enhances internal learning to effectively capture low-level hidden features, thereby improving the accuracy of medical image segmentation. This module helps address complex image characteristics, providing more refined and accurate results for segmentation tasks. The RM architecture is shown in Figure 5.

The input X is first introduced with non-linearity through a 1×1 convolution, followed by capturing more spatial features through a 3×3 convolution. Afterwards, a residual fusion is performed to aid in smooth gradient propagation, avoiding gradient vanishing or exploding issues. Subsequently, channel attention weighting is applied through a Convolutional Block Attention Module (CBAM), making the network focus more on crucial channel information and enhancing the weight of effective features. This part of RM processing is detailed as follows:

$$X^1 = X + Conv_{3 \times 3}(\text{ReLU}(\text{BN}(Conv_{1 \times 1}(X)))) \quad (8)$$

$$X^2 = CBAM(X^1) \quad (9)$$

Afterwards, a series of asymmetric convolution operations are employed to capture directional features, enhancing the model’s ability to perceive information in different directions. Finally, the original features are added in the form of residual connections to produce X^3 and output Y , as follows:

$$X^3 = X^2 + Conv_{1 \times 3}(Conv_{3 \times 1}(Conv_{1 \times 1}(X^2))) \quad (10)$$

$$Y = X^3 + X \quad (11)$$

In summary, the entire module strengthens the connection between channel and spatial features, enhancing the weight of effective features, and is conducive to extracting effective features of organs in medical images. RM can capture hidden features between pixels, thus achieving finer segmentation and improving the model segmentation accuracy.

G. LOSS FUNCTION

The Combo loss [45], a combination of the Dice loss and the Binary Cross-Entropy (BCE) loss, was used as a loss function in the experiments, defined as follows:

$$L_{combo} = 0.5 * L_{bce} + L_{dice} \quad (12)$$

where L_{bce} denotes the BCE loss [46], calculated as follows:

$$L_{bce} = - \sum_{i=1}^N [S_{GT} \ln(S_{pred}) + (1 - S_{GT}) \ln(1 - S_{pred})] \quad (13)$$

and L_{dice} represents the result of the edge loss ($Loss1$) and the global loss ($Loss2$), based on the Dice loss function [47], calculated as follows:

$$L_{dice} = 0.2 * Loss1 + 0.8 * Loss2 \quad (14)$$

$$Loss1 = 1 - 2 \frac{\sum_{i=1}^N S_{GT_b} S_{pred_b}}{\sum_{i=1}^N S_{GT_b}^2 + \sum_{i=1}^N S_{pred_b}^2} \quad (15)$$

$$Loss2 = 1 - 2 \frac{\sum_{i=1}^N S_{GT} S_{pred}}{\sum_{i=1}^N S_{GT}^2 + \sum_{i=1}^N S_{pred}^2} \quad (16)$$

$Loss1$ represents the difference between the boundary key point prediction S_{pred_b} and the boundary point label S_{GT_b} , whereas $Loss2$ represents the difference between the segmentation prediction S_{pred} and the real segmentation map S_{GT} . The use of $Loss1$ and $Loss2$ is depicted in Figure 1.

TABLE 1. Summary of thyroid ultrasound image datasets used in the experiments.

Dataset	Training set's images	Validation set's images	Test set's images	Ultrasonic image devices utilized
TN3K	2303	576	614	GE Logiq E9, ARIETTA 850, RESONA 70B
DDTI	511	63	63	TOSHIBA Nemio 30, TOSHIBA Nemio MX

The Combo loss is designed to address both the pixel-level classification loss and region-level overlap loss simultaneously, providing a comprehensive evaluation of segmentation performance. This approach effectively handles the challenge of imbalanced positive and negative samples, reducing sensitivity to sample imbalances and enhancing the model's generalization ability. Furthermore, it minimizes differences not only between segmentation predictions S_{pred} and ground-truth segmentation maps S_{GT} but also those between boundary keypoint predictions S_{pred_b} and boundary point labels S_{GT_b} .

IV. EXPERIMENTS AND RESULTS

A. DATASETS

To evaluate the proposed model in comparison to other existing models used for the same purpose, two open-access datasets, containing thyroid ultrasound images, were utilized in the experiments – the Thyroid Nodule 3493 (TN3K) dataset and the DDTI dataset (Table 1).

The TN3K dataset was made public by Gong et al. [48] in order to drive advancements in the field of thyroid nodule segmentation. This dataset focuses on images of thyroid nodules and includes 3,493 ultrasound images of 2,421 patients. These images have been converted to grayscale form and are accompanied by high-resolution mask labels indicating the positions of thyroid nodules in the images. For the experiments, 2,303 images were selected for model training, 576 images for model validation, and 614 images for model testing. The size of all input images was set to 256×256 pixels.

The DDTI dataset, introduced by Pedraza et al. [49], consists of a moderate collection of 637 images, each with pixel-level lesion masks. For the experiments, 511 images were selected for model training, 63 images for model validation, and 63 images for model testing.

B. EVALUATION METRICS

The proposed model was compared with other existing models using four popular metrics, namely the Intersection over Union (IoU), Dice Similarity Coefficient (DSC), recall, and precision.

IoU calculates the overlap rate between the candidate bound and the ground truth bound, whereby a ratio of 1 of their intersection to their union represents exact overlapping.

TABLE 2. Quantitative evaluation of the proposed BFG&MSF-Net model in comparison to other segmentation models.

Dataset	Model	IoU (%)	DSC (%)	Recall (%)	Precision (%)
TN3K	U-Net	66.70	75.56	82.63	78.67
	UNet++	68.80	78.85	88.81	76.14
	ResNet-18	71.66	80.54	88.73	79.07
	Att U-Net	74.09	82.64	86.21	84.57
	Seg-Net	68.14	77.27	81.20	81.72
	BFG&MSF-Net (proposed)	78.73	86.82	87.48	88.72
DDTI	U-Net	61.23	74.10	77.26	75.42
	UNet++	58.75	72.43	81.12	69.44
	ResNet-18	61.64	74.31	78.55	75.41
	Att U-Net	62.51	76.02	77.33	78.06
	Seg-Net	53.60	67.98	80.37	62.83
	BFG&MSF-Net (proposed)	69.35	81.02	81.19	83.64

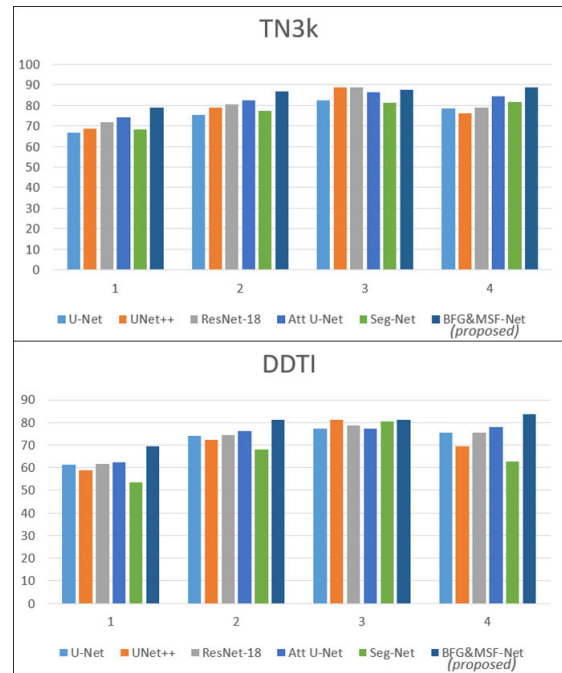


FIGURE 6. Visualization of the quantitative evaluation of the proposed BFG&MSF-Net model in comparison to other segmentation models.

IoU is commonly used for tasks such as object detection and semantic segmentation. It is calculated as follows:

$$IoU = \frac{TP}{FN + TP + FP} \tag{17}$$

where TP represents the number of cases correctly classified as positive, FP represents the number of cases incorrectly classified as positive, and FN represents the number of cases incorrectly classified as negative.

DSC describes the similarity between predicted and ground-truth values. It yields a result of 0 when there is no intersection between them, and 1 when both values are

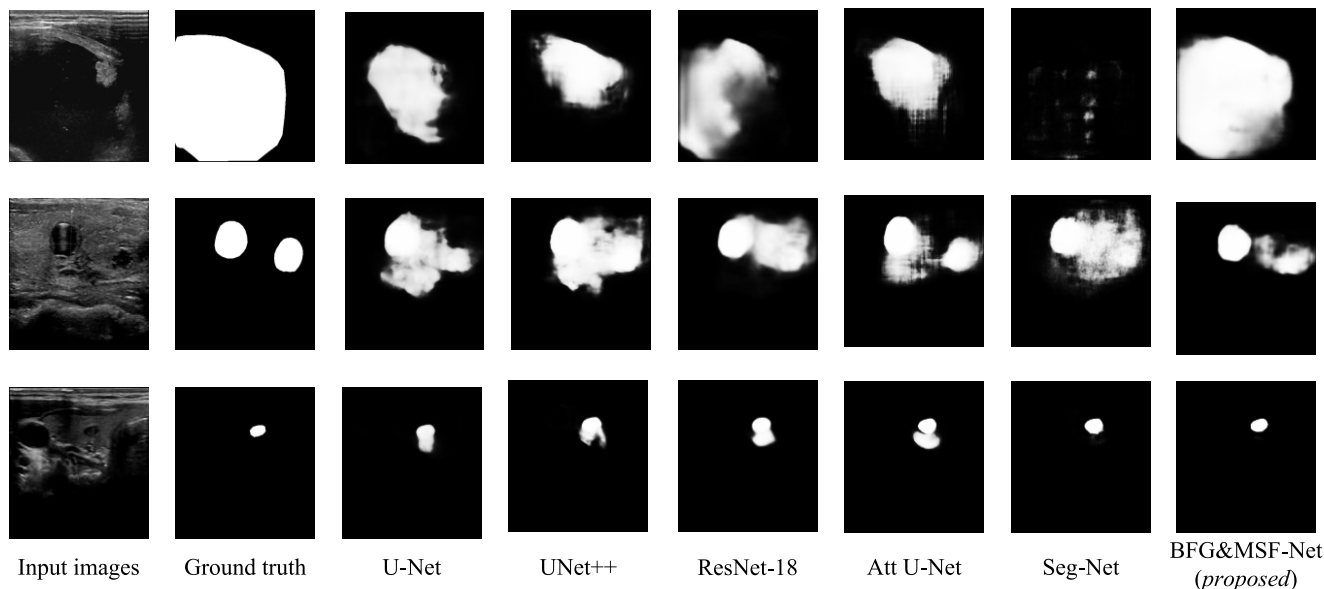


FIGURE 7. Qualitative evaluation of the proposed BFG&MSF-Net model in comparison to other segmentation models, based on TN3K dataset.

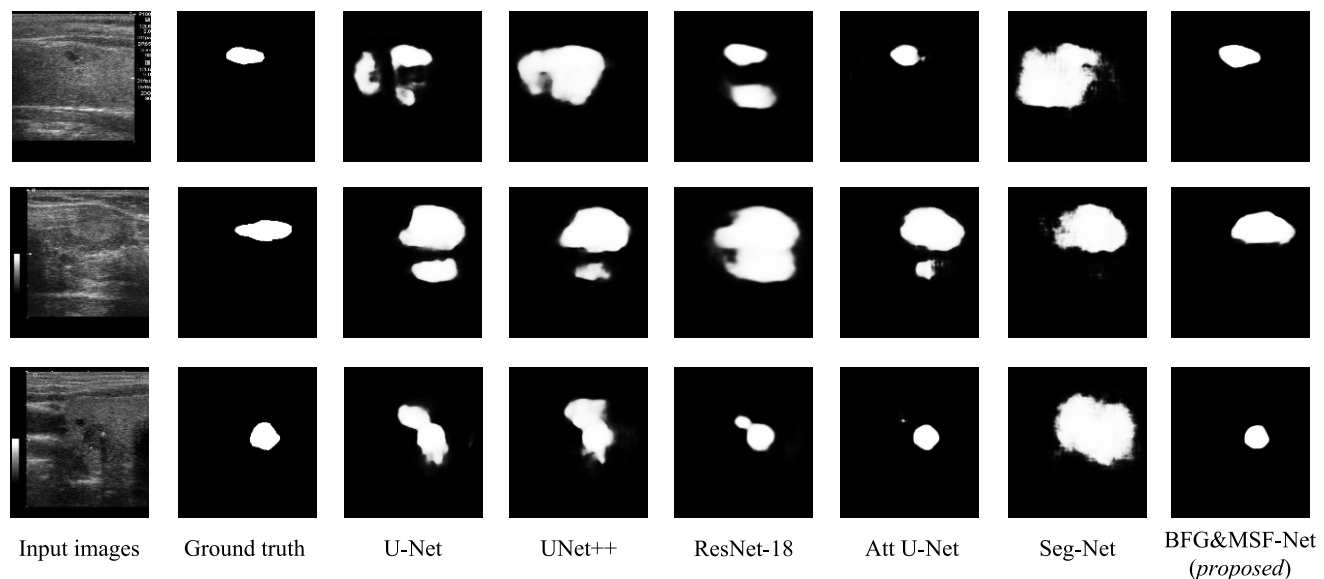


FIGURE 8. Qualitative evaluation of the proposed BFG&MSF-Net model in comparison to other segmentation models, based on DDTI dataset.

identical. It is calculated, as follows:

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{18}$$

Recall, also known as sensitivity or true positive rate, refers to the proportion of relevant samples that were correctly retrieved by a model among all samples that should have been retrieved. It is calculated, as follows:

$$Recall = \frac{TP}{TP+FN} \tag{19}$$

Precision, also known as positive predictive value, refers to the proportion of correctly predicted positive samples among

all samples predicted as positive by a model. It is calculated, as follows:

$$Precision = \frac{TP}{TP + FP} \tag{20}$$

C. EXPERIMENTAL ENVIRONMENT

The hardware configuration used in the experiments included an Intel Core i5-12490 processor with a clock speed of 3.0 GHz and a single NVIDIA RTX3060 graphics card with 12 GB of VRAM. Based on the server configuration, and to ensure normal model training, the hyperparameters for the neural network during the training process were set as

TABLE 3. Ablation study experimental results, obtained on TN3K dataset.

Model versions	Components							Evaluation Metrics			
	U-Net	Conv-NeXt	DSASPPM	PSA	PBA	RM	BFGM	IoU (%)	DSC (%)	Rec (%)	Pre (%)
v0 (baseline)	√							66.70	75.56	82.63	78.67
v1	√	√						75.58	83.27	86.12	86.39
v2	√	√	√					75.77	84.12	86.69	85.93
v3	√	√	√	√				76.78	85.08	85.00	88.90
v4	√	√	√	√	√			77.35	85.82	89.32	85.21
v5	√	√	√	√	√	√		78.03	86.39	89.81	85.65
v6 (proposed)	√	√	√	√	√	√	√	78.73	86.82	87.48	88.72

TABLE 4. Ablation study experimental results, obtained on DDTI dataset.

Model versions	Components							Evaluation Metrics			
	U-Net	ConvNeXt	DSASPPM	PSA	PBA	RM	BFGM	IoU (%)	DSC (%)	Rec (%)	Pre (%)
v0 (baseline)	√							61.23	74.10	77.26	75.42
v1	√	√						65.73	77.66	84.61	75.61
v2	√	√	√					67.26	78.98	77.56	83.99
v3	√	√	√	√				68.48	79.94	80.22	83.32
v4	√	√	√	√	√			68.64	80.33	78.24	86.21
v5	√	√	√	√	√	√		69.11	80.80	80.97	84.47
v6 (proposed)	√	√	√	√	√	√	√	69.35	81.02	81.19	83.64

follows: the batch size was set to 4, the number of epochs was set to 100, validation was performed on every epoch, and the Adam optimizer was used for training. The initial learning rate was set to 1×10^{-4} , with a decay factor of 1×10^{-4} to prevent overfitting, a momentum of 0.9, and a minimum learning rate of 1×10^{-5} . The network architecture was implemented using PyTorch.

D. PERFORMANCE COMPARISON WITH COMMONLY USED SEGMENTATION MODELS

In order to validate the superiority of the proposed model in performing thyroid nodule segmentation, it was compared with commonly used segmentation models such as U-Net [28], UNet++ [20], ResNet-18 [40], Att U-Net [38], and Seg-Net [41]. The obtained quantitative results of different models are shown in Table 2 and Figure 6, presented separately for each of the two datasets used – TN3K and DDTI. These results demonstrate that the proposed BFG&MSF-Net model outperforms all other segmentation models on both datasets according to all evaluation metrics, except for recall on TN3K.

In addition to the quantitative evaluation demonstrating the outstanding performance of the proposed BFG&MSF-Net model, Figures 7 and 8 contain some illustrative examples confirming that it works indeed better than other models in segmenting thyroid nodules, based on both TN3K and DDTI datasets. Specifically, Figure 7 illustrates three different cases of thyroid nodules of different sizes (i.e., large, double, and

small nodules), whereas Figure 8 shows the presence of thyroid nodules of different shapes and sizes.

In the case of large nodules (first row in Figure 7), the segmentation results of the other models deviate significantly from reality, while BFG&MSF-Net, due to its ability to capture global contextual information, can accurately segment the nodules' contours and maintain good regional continuity. In addition, BFG&MSF-Net achieves better segmentation results than other models in segmenting double nodules (second row in Figure 7) and small nodules (third row in Figure 7).

In Figure 8, one can clearly see that the proposed model can accurately distinguish thyroid nodules from the blurred background, along with distinguishing their edges from the surrounding background. The visual results of other models include edge partition errors and inaccurate target locations.

On both datasets, the fuzzy area of the renderings of the five compared models (U-Net, UNet++, ResNet-18, Att U-Net, and Seg-Net) is obviously larger than that of the proposed model.

E. ABLATION STUDY

To assess the performance of different components of the proposed model, we conducted ablation study experiments on both datasets, utilizing U-Net as a baseline.

Tables 3 and 4 provide a detailed classification of the various component compositions resulting in different model versions, as well as their segmentation performance results on the TN3K and DDTI datasets, respectively (the best results

TABLE 5. Segmentation performance comparison of the proposed BFG&MSF-Net model with state-of-the-art models.

Year	Model	DSC (%) on TN3K dataset	DSC (%) on DDTI dataset
2023	SAM [50]	81.26	-
2023	AMSeg [51]	84.21	74.81
2023	TRFE+ [48]	83.33	75.37
2023	BPAT-UNet [52]	83.64	-
2023	SAMUS [53]	84.45	-
2023	Tnseg [54]	85.71	74.93
2022	LCA-Net [55]	82.08	-
2021	TRFE [56]	81.19	69.04
2020	SGUNET [57]	80.80	70.60
2020	WU-Net [58]	81.27	-
2024	BFG&MSF-Net (proposed)	86.82	81.02

are shown in **bold**). The presented results indicate that the newly designed modules, presented in the previous section, all contribute positively to enhancing the segmentation performance of the proposed model. As shown in Table 3, the addition of these modules (one after the other) to the baseline resulted in a gradual increase in all metrics on the TN3K dataset, except for few cases of, recall (w.r.t. v5) and precision (w.r.t. v3). Similarly, as illustrated in Table 4, the incremental inclusion of these modules in the baseline resulted in a progressive enhancement in all metrics on the DDTI dataset, except for precision (w.r.t. v4). After integrating all designed modules, the sixth model version (v6), i.e., the proposed BFG&MSF-Net model, demonstrated an increase of 12.03 percentage points in IoU, 11.26 percentage points in DSC, 4.85 percentage points in recall, and 10.05 percentage points in precision, compared to the baseline (U-Net) on the TN3K dataset. On the DDTI dataset, the sixth model version demonstrated an increase of 8.12 percentage points in IoU, 6.92 percentage points in DSC, 3.93 percentage points in recall, and 8.22 percentage points in precision, compared to the baseline. These results demonstrate the effectiveness of each designed module in helping the proposed BFG&MSF-Net model to effectively perform the thyroid nodule segmentation task.

F. PERFORMANCE COMPARISON WITH STATE-OF-THE-ART MODELS

In addition, we conducted a comparison of the segmentation performance of the proposed BFG&MSF-Net model with that of state-of-the-art models, based on their reported in relevant literature. The comparative results on both datasets, presented in Table 5, demonstrate that BFG&MSF-Net exhibits superior performance compared to all other models. In this comparison, the decision to focus on DSC as a primary evaluation metric was intentional, given its widespread usage in the field of image segmentation and its recognition as a key metric there.

V. CONCLUSION AND FUTURE DIRECTIONS

This paper has proposed a Boundary Feature Guidance and Multi-Scale Fusion Network (BFG&MSF-Net) model

for thyroid nodule segmentation. The incorporation of the well-known backbone network, ConvNeXt, along with the newly designed Boundary Feature Guidance Module (BFGM), Multi-Scale Perception Fusion Module (MSPFM), Depthwise Separable Atrous Spatial Pyramid Pooling Module (DSASPPM), and Refinement Module (RM), allows to effectively enhance the model's ability to capture edge details and comprehensively perceive image information, leading to more accurate thyroid nodule segmentation. The obtained experimental results have demonstrated that the proposed model outperforms all existing models, used for similar purposes, thus providing an improved assistance for doctors in early disease diagnostics and treatment planning.

However, a limitation of the presented research lies in the experimental validation being conducted solely on ultrasound thyroid images. The generalizability of the proposed model to ultrasound images of other organs, such as breast and prostate, or different types of medical images, such as CT, Positron Emission Computed Tomography (PET), or Magnetic Resonance Imaging (MRI), has not been verified due to lack of access to such images. In the future, we aim to validate the performance of the proposed BFG&MSF-Net model across a broader spectrum of medical images and, in addition, will continually refine and iterate it.

REFERENCES

- [1] J. Kim, J. E. Gosnell, and S. A. Roman, "Geographic influences in the global rise of thyroid cancer," *Nature Rev. Endocrinol.*, vol. 16, no. 1, pp. 17–29, Jan. 2020.
- [2] S. A. Paschou, A. Vryonidou, and D. G. Goulis, "Thyroid nodules: A guide to assessment, treatment and follow-up," *Maturitas*, vol. 96, pp. 1–9, Feb. 2017.
- [3] E. K. Alexander and E. S. Cibas, "Diagnosis of thyroid nodules," *Lancet Diabetes Endocrinol.*, vol. 10, no. 7, pp. 533–539, 2022.
- [4] C. Durante, G. Grani, L. Lamartina, S. Filetti, S. J. Mandel, and D. S. Cooper, "The diagnosis and management of thyroid nodules: A review," *JAMA*, vol. 319, no. 9, pp. 914–924, 2018.
- [5] M. A. Savelonas, D. K. Iakovidis, I. Legakis, and D. Maroulis, "Active contours guided by echogenicity and texture for delineation of thyroid nodules in ultrasound images," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 4, pp. 519–527, Jul. 2008.
- [6] F. N. Tessler, W. D. Middleton, E. G. Grant, J. K. Hoang, L. L. Berland, S. A. Teefey, J. J. Cronan, M. D. Beland, T. S. Desser, M. C. Frates, L. W. Hammers, U. M. Hamper, J. E. Langer, C. C. Reading, L. M. Scoutt, and A. T. Stavros, "ACR thyroid imaging, reporting and data system (TI-RADS): White paper of the ACR TI-RADS committee," *J. Amer. College Radiol.*, vol. 14, no. 5, pp. 587–595, May 2017.
- [7] A. Tuysuzoglu, J. Tan, K. Eissa, A. P. Kiraly, M. Diallo, and A. Kamen, "Deep adversarial context-aware landmark detection for ultrasound imaging," in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018*. Cham, Switzerland: Springer, Sep. 2018, pp. 151–158.
- [8] V. T. Manh, J. Zhou, X. Jia, Z. Lin, W. Xu, Z. Mei, Y. Dong, X. Yang, R. Huang, and D. Ni, "Multi-attribute attention network for interpretable diagnosis of thyroid nodules in ultrasound images," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 69, no. 9, pp. 2611–2620, Sep. 2022.
- [9] I. L. Xi, Y. Zhao, R. Wang, M. Chang, S. Purkayastha, K. Chang, R. Y. Huang, A. C. Silva, M. Vallières, P. Habibollahi, Y. Fan, B. Zou, T. P. Gade, P. J. Zhang, M. C. Soulen, Z. Zhang, H. X. Bai, and S. W. Stavropoulos, "Deep learning to distinguish benign from malignant renal lesions based on routine MR imaging," *Clin. Cancer Res.*, vol. 26, no. 8, pp. 1944–1952, Apr. 2020, doi: 10.1158/1078-0432.ccr-19-0374.

- [10] W. Yang, Y. Dong, Q. Du, Y. Qiang, K. Wu, J. Zhao, X. Yang, and M. B. Zia, "Integrate domain knowledge in training multi-task cascade deep learning model for benign-malignant thyroid nodule classification on ultrasound images," *Eng. Appl. Artif. Intell.*, vol. 98, May 2021, Art. no. 104064.
- [11] J. Chen, H. You, and K. Li, "A review of thyroid gland segmentation and thyroid nodule segmentation methods for medical ultrasound images," *Comput. Methods Programs Biomed.*, vol. 185, Mar. 2020, Art. no. 105329, doi: [10.1016/j.cmpb.2020.105329](https://doi.org/10.1016/j.cmpb.2020.105329).
- [12] I. Qureshi, J. Yan, Q. Abbas, K. Shaheed, A. B. Riaz, A. Wahid, M. W. J. Khan, and P. Szczuko, "Medical image segmentation using deep semantic-based methods: A review of techniques, applications and emerging trends," *Inf. Fusion*, vol. 90, pp. 316–352, Feb. 2023.
- [13] K. Sharifani and M. Amini, "Machine learning and deep learning: A review of methods and applications," *World Inf. Technol. Eng. J.*, vol. 10, no. 7, pp. 3897–3904, 2023.
- [14] A. Kaur and G. Dong, "A complete review on image denoising techniques for medical images," *Neural Process. Lett.*, vol. 55, no. 6, pp. 7807–7850, Dec. 2023, doi: [10.1007/s11063-023-11286-1](https://doi.org/10.1007/s11063-023-11286-1).
- [15] J. Sun, C. Li, Z. Lu, M. He, T. Zhao, X. Li, L. Gao, K. Xie, T. Lin, J. Sui, Q. Xi, F. Zhang, and X. Ni, "TNSNet: Thyroid nodule segmentation in ultrasound imaging using soft shape supervision," *Comput. Methods Programs Biomed.*, vol. 215, Mar. 2022, Art. no. 106600, doi: [10.1016/j.cmpb.2021.106600](https://doi.org/10.1016/j.cmpb.2021.106600).
- [16] Q. Zhang, J. Xiao, C. Tian, J. Chun-Wei Lin, and S. Zhang, "A robust deformed convolutional neural network (CNN) for image denoising," *CAAI Trans. Intell. Technol.*, vol. 8, no. 2, pp. 331–342, Jun. 2023.
- [17] J. Li, J. Chen, B. Sheng, P. Li, P. Yang, D. D. Feng, and J. Qi, "Automatic detection and classification system of domestic waste via multimodel cascaded convolutional neural network," *IEEE Trans. Ind. Informat.*, vol. 18, no. 1, pp. 163–173, Jan. 2022.
- [18] X. Li, Y. Jiang, M. Li, and S. Yin, "Lightweight attention convolutional neural network for retinal vessel image segmentation," *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 1958–1967, Mar. 2021.
- [19] H. Zhou, J. Zhang, J. Lei, S. Li, and D. Tu, "Image semantic segmentation based on FCN-CRF model," in *Proc. Int. Conf. Image, Vis. Comput. (ICIVC)*, Aug. 2016, pp. 9–14.
- [20] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2018, pp. 3–11.
- [21] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense, volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016*, S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells, Eds. Cham, Switzerland: Springer, 2016, pp. 424–432.
- [22] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018, doi: [10.1109/TPAMI.2017.2699184](https://doi.org/10.1109/TPAMI.2017.2699184).
- [23] Y. Mo, Y. Wu, X. Yang, F. Liu, and Y. Liao, "Review the state-of-the-art technologies of semantic segmentation based on deep learning," *Neurocomputing*, vol. 493, pp. 626–646, Jul. 2022.
- [24] X. Ying, Z. Yu, R. Yu, X. Li, M. Yu, M. Zhao, and K. Liu, "Thyroid nodule segmentation in ultrasound images based on cascaded convolutional neural network," in *Neural Information Processing*, L. Cheng, A. C. S. Leung, and S. Ozawa, Eds. Cham, Switzerland: Springer, 2018, pp. 373–384.
- [25] V. Kumar, J. Webb, A. Gregory, D. D. Meixner, J. M. Knudsen, M. Callstrom, M. Fatemi, and A. Alizad, "Automated segmentation of thyroid nodule, gland, and cystic components from ultrasound images using deep learning," *IEEE Access*, vol. 8, pp. 63482–63496, 2020, doi: [10.1109/ACCESS.2020.2982390](https://doi.org/10.1109/ACCESS.2020.2982390).
- [26] H. Pan, Q. Zhou, and L. J. Latecki, "SGUNET: Semantic guided UNET for thyroid nodule segmentation," in *Proc. IEEE 18th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2021, pp. 630–634, doi: [10.1109/ISBI48211.2021.9434051](https://doi.org/10.1109/ISBI48211.2021.9434051).
- [27] Y. Zhang, H. Lai, and W. Yang, "Cascade UNet and CH-UNet for thyroid nodule segmentation and benign and malignant classification," in *Segmentation, Classification, and Registration of Multi-Modality Medical Imaging Data*, N. Shusharina, M. P. Heinrich, and R. Huang, Eds. Springer, 2021, pp. 129–134.
- [28] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham, Switzerland: Springer, 2015, pp. 234–241.
- [29] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11966–11976, doi: [10.1109/CVPR52688.2022.01167](https://doi.org/10.1109/CVPR52688.2022.01167).
- [30] S. Woo, S. Debnath, R. Hu, X. Chen, Z. Liu, I. S. Kweon, and S. Xie, "ConvNeXt v2: Co-designing and scaling ConvNets with masked autoencoders," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2023, pp. 16133–16142.
- [31] H. Zhang, K. Zu, J. Lu, Y. Zou, and D. Meng, "EPSANet: An efficient pyramid squeeze attention block on convolutional neural network," in *Proc. Asian Conf. Comput. Vis.*, 2022, pp. 1161–1177.
- [32] D. D. Patil and S. G. Deore, "Medical image segmentation: A review," *Int. J. Comput. Sci. Mobile Comput.*, vol. 2, no. 1, pp. 22–27, 2013.
- [33] Y. Zhong, J. Yang, P. Zhang, C. Li, N. Codella, L. H. Li, L. Zhou, X. Dai, L. Yuan, Y. Li, and J. Gao, "RegionCLIP: Region-based language-image pretraining," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 16772–16782.
- [34] K. Bhargavi and S. Jyothi, "A survey on threshold based segmentation technique in image processing," *Int. J. Innov. Res. Develop.*, vol. 3, no. 12, pp. 234–239, 2014.
- [35] T. McInerney and D. Terzopoulos, "Deformable models in medical image analysis: A survey," *Med. Image Anal.*, vol. 1, no. 2, pp. 91–108, Jun. 1996.
- [36] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu, Z. Yang, Y. Zhang, and D. Tao, "A survey on vision transformer," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 87–110, Jan. 2023.
- [37] Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang, "A review of deep learning based methods for medical image multi-organ segmentation," *Phys. Medica*, vol. 85, pp. 107–122, May 2021.
- [38] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.
- [39] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "ResUNet—A: A deep learning framework for semantic segmentation of remotely sensed data," *ISPRS J. Photogramm. Remote Sens.*, vol. 162, pp. 94–114, Apr. 2020, doi: [10.1016/j.isprsjprs.2020.01.013](https://doi.org/10.1016/j.isprsjprs.2020.01.013).
- [40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [41] G. Rajakumar, R. S. Jeba Leela, P. E. Darney, K. L. Narayanan, R. S. Krishnan, and Y. H. Robinson, "Seg-Net: Automatic lung infection segmentation of COVID-19 from CT images," in *Proc. 5th Int. Conf. Trends Electron. Informat. (ICOEI)*, Jun. 2021, pp. 739–744.
- [42] Y. Dai, F. Giesecke, S. Oehmcke, Y. Wu, and K. Barnard, "Attentional feature fusion," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 3559–3568.
- [43] C. Li, R. Du, Q. Luo, R. Wang, and X. Ding, "A novel model of thyroid nodule segmentation for ultrasound images," *Ultrasound Med. Biol.*, vol. 49, no. 2, pp. 489–496, Feb. 2023, doi: [10.1016/j.ultrasmedbio.2022.09.017](https://doi.org/10.1016/j.ultrasmedbio.2022.09.017).
- [44] BCEWithLogitsLoss. *PyTorch Documentation*. Accessed: May 31, 2024. [Online]. Available: <https://pytorch.org/docs/stable/generated/torch.nn.BCEWithLogitsLoss.html>
- [45] S. A. Taghanaki, Y. Zheng, S. K. Zhou, B. Georgescu, P. Sharma, D. Xu, D. Comaniciu, and G. Hamarneh, "Combo loss: Handling input and output imbalance in multi-organ segmentation," *Computerized Med. Imag. Graph.*, vol. 75, pp. 24–33, Jul. 2019, doi: [10.1016/j.compmedimag.2019.04.005](https://doi.org/10.1016/j.compmedimag.2019.04.005).
- [46] W.-C. Tu, M.-Y. Liu, V. Jampani, D. Sun, S.-Y. Chien, M.-H. Yang, and J. Kautz, "Learning superpixels with segmentation-aware affinity loss," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 568–576, doi: [10.1109/CVPR.2018.00066](https://doi.org/10.1109/CVPR.2018.00066).
- [47] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571, doi: [10.1109/3DV.2016.79](https://doi.org/10.1109/3DV.2016.79).

[48] H. Gong, J. Chen, G. Chen, H. Li, G. Li, and F. Chen, "Thyroid region prior guided attention for ultrasound segmentation of thyroid nodules," *Comput. Biol. Med.*, vol. 155, Jul. 2023, Art. no. 106389.

[49] L. Pedraza, C. Vargas, F. Narváez, O. Durán, E. Muñoz, and E. Romero, "An open access thyroid ultrasound image database," *Proc. SPIE*, vol. 9287, pp. 188–193, Jan. 2015.

[50] D. Cheng, Z. Qin, Z. Jiang, S. Zhang, Q. Lao, and K. Li, "SAM on medical images: A comprehensive study on three prompt modes," 2023, *arXiv:2305.00035*.

[51] X. Ma, B. Sun, W. Liu, D. Sui, J. Chen, and Z. Tian, "AMSeg: A novel adversarial architecture based multi-scale fusion framework for thyroid nodule segmentation," *IEEE Access*, vol. 11, pp. 72911–72924, 2023, doi: [10.1109/ACCESS.2023.3289952](https://doi.org/10.1109/ACCESS.2023.3289952).

[52] H. Bi, C. Cai, J. Sun, Y. Jiang, G. Lu, H. Shu, and X. Ni, "BPAT-UNet: Boundary preserving assembled transformer UNet for ultrasound thyroid nodule segmentation," *Comput. Methods Programs Biomed.*, vol. 238, Jul. 2023, Art. no. 107614.

[53] X. Lin, Y. Xiang, L. Zhang, X. Yang, Z. Yan, and L. Yu, "SAMUS: Adapting segment anything model for clinically-friendly and generalizable ultrasound image segmentation," 2023, *arXiv:2309.06824*.

[54] X. Ma, B. Sun, W. Liu, D. Sui, S. Shan, J. Chen, and Z. Tian, "TNSeg: Adversarial networks with multi-scale joint loss for thyroid nodule segmentation," *J. Supercomput.*, vol. 80, no. 5, pp. 6093–6118, Mar. 2024, doi: [10.1007/s11227-023-05689-z](https://doi.org/10.1007/s11227-023-05689-z).

[55] Z. Tao, H. Dang, Y. Shi, W. Wang, X. Wang, and S. Ren, "Local and context-attention adaptive LCA-Net for thyroid nodule segmentation in ultrasound images," *Sensors*, vol. 22, no. 16, p. 5984, Aug. 2022, doi: [10.3390/s22165984](https://doi.org/10.3390/s22165984).

[56] H. Gong, G. Chen, R. Wang, X. Xie, M. Mao, Y. Yu, F. Chen, and G. Li, "Multi-task learning for thyroid nodule segmentation with thyroid region prior," in *Proc. IEEE 18th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2021, pp. 257–261, doi: [10.1109/ISBI48211.2021.9434087](https://doi.org/10.1109/ISBI48211.2021.9434087).

[57] M. Byra, P. Jarosik, A. Szubert, M. Galperin, H. Ojeda-Fournier, L. Olson, M. O'Boyle, C. Comstock, and M. Andre, "Breast mass segmentation in ultrasound with selective kernel U-Net convolutional neural network," *Biomed. Signal Process. Control*, vol. 61, Aug. 2020, Art. no. 102027.

[58] Y. Wu, X. Shen, F. Bu, and J. Tian, "Ultrasound image segmentation method for thyroid nodules using ASPP fusion features," *IEEE Access*, vol. 8, pp. 172457–172466, 2020.



HAORAN SUN was born in 2000. He received the B.S. degree from North China University of Science and Technology, in 2022, where he is currently pursuing the master's degree. His research interests include machine vision and graphic image processing.



CHENXU DAI received the B.S. and M.S. degrees from the School of Computer Science and Technology, North China University of Science and Technology. She is currently a Research Associate with North China University of Science and Technology. Her current research interests include neural networks, intelligent optimization algorithms, and deep learning.



ZHANLIN JI (Member, IEEE) received the M.Eng. degree from Dublin City University, Ireland, in 2006, and the Ph.D. degree from the University of Limerick, Ireland, in 2010. He is currently a Professor with North China University of Science and Technology and Zhejiang A&F University, China. He has authored/coauthored more than 100 research papers in refereed journals and conferences. His research interests include ubiquitous consumer wireless world (UCWW), the

Internet of Things (IoT), cloud computing, big data management, and data mining.



JIANUO LIU was born in 2000. She received the B.S. degree from North China University of Science and Technology, in 2022, where she is currently pursuing the master's degree. Her research interests include machine vision and graphic image processing.



JUNCHENG MU was born in 2000. He received the B.S. degree from North China University of Science and Technology, in 2022, where he is currently pursuing the master's degree. His research interests include machine vision and graphic image processing.



IVAN GANCHEV (Senior Member, IEEE) received the Engineering and Ph.D. degrees (summa cum laude) from Saint Petersburg University of Telecommunications, in 1989 and 1995, respectively. He is an International Telecommunications Union (ITU-T) Invited Expert and an Institution of Engineering and Technology (IET) Invited Lecturer, currently affiliated with the University of Limerick, Ireland, the University of Plovdiv "Paisii Hilendarski," Bulgaria, and the

Institute of Mathematics and Informatics—Bulgarian Academy of Sciences, Bulgaria. He was involved in more than 40 international and national research projects. He has authored/coauthored one monographic book, three textbooks, four edited books, and more than 300 research papers in refereed international journals, books, and conference proceedings. He has served on the TPC of more than 400 prestigious international conferences/symposia/workshops. He is on the editorial board of and has served as a guest editor for multiple prestigious international journals.

...