# XXX

Niu Wentao[a], XXX[a], XXX[b], Zhenyu Lei[a], Shangce Gao[a,*]

[a]*Faculty of Engineering, University of Toyama, Toyama-shi, 930-8555 Japan.*
[b]*Department of Computer Science and Technology, Beijing University, Beijing 100000, China.*

**Abstract**

Medical image segmentation serves as an important tool in the treatment of various medical diseases. However, achieving precise and efficient segmentation remains challenging due to the intricate structures and variations. Although neural network methods based on U-shaped structures have shown impressive results, they often lack effective representation of global-local features, leading to insufficient extraction of multi-scale and contextual information in medical image segmentation tasks. To tackle these challenges, we propose a novel approach: a dilated dendritic model with deep supervision, namely 3DL-Net. It integrates the flexible dilated convolution mechanism into the segmentation architecture, aiming to expand the model's receptive field and capture richer global features. Additionally, in contrast to other segmentation architectures, we innovatively introduce the processing of local feature of shallow structures through the dendritic neuron model in medical images into 3DL-Net. This is the first time dendritic learning has been employed at the channel level and represents a pioneering approach to the local feature process. During the training process, 3DL-Net incorporates a deep supervision mechanism that utilizes our designed loss function. Due to the intermediate supervision signals at various network stages, providing feedback at multiple levels, the model refines its predictions across various scales, contributing to further enhanced segmentation outcomes. To evaluate the effectiveness of our proposed method, we conducted extensive experiments on three medical image datasets to demonstrate significant improvements in segmentation accuracy compared to state-of-the-art models.

*Corresponding author
*Email address:* gaosc@eng.u-toyama.ac.jp (Shangce Gao)

Our mDice metrics on three datasets achieved 86.61%, 87.87%, and 85.06%, surpassing the second-best models by 3.85%, 1.54%, and 0.95%.
*Keywords:* Medical image segmentation, Dendritic learning, Dilated convolution, Global-local feature,

---

## 1. Introduction111

Medical image segmentation plays a crucial role in the accurate extraction and delineation of pathological regions [1]. This process enhances the clarity of anatomical or pathological structures within images, aiding in accurate clinical assessments, disease diagnosis, and treatment planning. It plays a significant role in healthcare decision-making and patient care [2]. Current medical image segmentation tasks span across various fields and imaging modalities, including computed tomography (CT) and ultrasound. These tasks encompass the segmentation of diverse organs such as the brain, lungs, liver, kidneys, and other vital structures [3, 4, 5].

In order to help clinicians make accurate diagnoses, segmentation of some key objects in medical images and extraction of features from the segmented region are necessary. Early image segmentation relied on conventional computer vision techniques, such as thresholding [6, 7], edge detection [8, 9], region growing [10], level set [11], and graph cut [12, 13]. These methods utilized specific image features, pixel-level processing, and empirical rules to segment images. Despite the notable contributions of traditional segmentation methods, image segmentation remains one of the most challenging topics in computer vision due to the difficulty of feature representation [14]. In particular, extracting discriminative features from medical images is more difficult than for normal RGB images, since the former are often faced with challenges such as complex background images, noise interference, and variations in target shape [15]. Moreover, it further involves managing large volumes of data and addressing issues related to inaccurate location labeling [16, 17]. Especially in the clinical process, these challenges become more critical. Clinical specialists seek better methods to address it instead of relying on time-consuming manual procedures for organ segmentation [18].

The emergence of deep learning technologies has effectively addressed challenges

such as complex backgrounds, noise interference, and variations in target shape [19, 20], leading to the development of new paradigms. The pivotal paradigm involves eliminating the need for manual crafting of features in medical image segmentation. Convolutional neural networks (CNNs) have emerged as a fundamental tool in medical image segmentation, capable of automatically learning hierarchical features from raw image data. These features encompass a wide range of visual characteristics, including edges, textures, shapes, and other discriminative patterns present in medical images. Due to their insensitivity to variations in image noise, blurring, and contrast, CNNs yield excellent segmentation results for medical images. This relieves healthcare professionals from the laborious task of manually generating features, enabling them to concentrate on directly diagnosing segmented lesion areas. This shift notably enhances the efficiency of clinical diagnosis.

With the continuous advancement of CNNs, image segmentation technology has achieved significant breakthroughs, particularly in semantic segmentation. Semantic segmentation involves dividing the input image into distinct regions with specific semantic meanings, enabling the identification of semantic categories for each region through pixel-wise annotation [21]. Subsequent developments in semantic segmentation methods based on CNNs have introduced notable models such as FCN [22], U-Net [23], and SegNet [24]. However, despite the notable achievements in semantic segmentation using CNN-based approaches, certain challenges persist. One of the primary obstacles involves addressing class imbalance, which refers to the uneven distribution of samples among different classes or regions of interest (ROI) [25]. The issue of class imbalance in medical image segmentation often leads to unsatisfactory segmentation results. This challenge becomes evident when minority classes are situated within the ROI, leading to potential neglect or insufficient model learning regarding critical details within these minority categories [26]. An appropriate loss function constitutes a pivotal approach to mitigating this issue. Conventional loss functions, primarily relying on accuracy, might not effectively address this problem. In response to these challenges, employing specialized loss functions such as Intersection over Union (IoU) [27], focal loss [28], dice loss [29], and others for assessing model performance is emphasized.

Another significant concern is the effective integration of multi-scale and contextual information. It significantly impacts the ability of CNN-based models in various medical image segmentation tasks to comprehend the local image structure and global feature process. Current CNN-based deep learning models often encounter challenges such as the omission of spatial information and insufficient representation of features [30]. The inherent bias of the convolutional layer often restricts the extracted information to the main regions of the segmented area, thus limiting its ability to accurately perceive intricate edge variations and the amalgamation of numerous small segmented regions [31]. The lack of multi-scale contextual information and the absence of semantic details during pooling operations exacerbate this insufficient feature representation. Notably, medical pathology images demonstrate more significant scale differences in lesion areas compared to typical natural images. Therefore, addressing these challenges necessitates the introduction of more effective multi-scale modules [32]. Specifically, CNN-based methods hierarchically extract features from shallow to deep through convolutional operations. However, these approaches frequently overlook the semantic distinctions between deep and shallow features, making it challenging to capture both global and local features in medical images [33]. Moreover, traditional CNNs architectures typically prioritize the local receptive field during feature extraction, neglecting the importance of global context information. Therefore, model optimization should involve extracting features across various levels of semantic representations to enhance segmentation tasks. To overcome this limitation, advanced approaches have been introduced, such as incorporating probabilistic graphical models into deep learning architectures [34] and utilizing pyramid scene parsing networks (PSPN) [35]. These methods incorporate graph-based and pyramid-based mechanisms to capture multi-scale and contextual information, thereby enhancing the representation of features. Additionally, AAUnet [36], RRCNet [37], LEA U-Net [38] utilize dilation convolution to expand the receptive field, combining various dilation rates to extract semantic features at different scales. These methods employ dilated convolution to obtain more global features, enabling the model to better capture long-range correlations and global contextual information in the image. Nonetheless, the extraction of coarse-grained features still results in the loss of fine-grained details pertaining

to small targets, thereby constraining the performance of their detection. It is essential to propose methods for integrating global coarse-grained and local fine-grained information.

The introduction of the aforementioned enables the enhancement of segmentation accuracy by simply deepening the network. However, in this process, the significance of shallow features near the output layer is often overlooked. The model may fail to capture some of the finer local structures and characteristics in the image, as they tend to provide relatively surface-level feature representations. Deep features in both natural and medical images exhibit some commonalities, as evidenced by the successful application of models trained on natural images to medical images through transfer learning. Particularly, generic features related to texture and shape are shared [39, 40]. Conversely, shallow features in natural and medical images exhibit substantial content and feature differences. Natural images encompass diverse scenes, emphasizing colors, larger textures, and shapes. In contrast, medical images, where color features are less pronounced, emphasize complex edge contours and focus on depicting the density features hidden within tissues and organ structures. They predominantly highlight characteristics related to grayscale, edges, and shapes. Nevertheless, traditional CNNs networks find it difficult to provide high-precision local feature analysis capabilities [41]. CNNs usually prioritize the extraction of high-level features while neglecting or inadequately representing low-level features, resulting in the loss of fine-grained details. Moreover, they cannot provide an interpretable basis for segmenting loca features, hindering doctors from giving diagnoses aligned with the underlying pathology. To address these challenges and refine the processing of local features, we propose a novel module for refining local features, introducing the dendritic neuron model (DNM). In contrast to the overly simplistic traditional McCulloch-Pitts model of neurons in CNNs [42], the DNM represents a biologically plausible model designed to emulate the information processing mechanisms of neurons in the human brain [43]. It is introduced to overcome the limitations of traditional artificial neurons, which are often considered too simplistic [44]. The DNM exhibits robust feature mapping capabilities. Through feature multiplexing and weighting, it can simulate dendritic processes to capture complex, nonlinear relationships, thereby addressing the challenges posed

by local features. This approach enables the extraction of subtle local features from shallow representations, thereby enhancing their representation and providing a more interpretable basis for the classification at the final layer.

To meet the challenges in medical image segmentation outlined previously, we propose a novel segmentation method, a dilated dendritic model with deep supervision, 3DL-Net. This method is trained with deep supervision and a loss function tailored for imbalanced class issues. By employing dilated convolutions, 3DL-Net captures broader contextual information, enhancing the representation and understanding of global features. Additionally, the integration of dendritic neurons, a distinct type of neuron from the commonly used McCulloch-Pitts neurons, offers greater biological interpretability and aims to refine the processing and representation of local features with superior precision. The contributions of this work can be summarized as follows:

1) We leverage dilated convolutions with diverse dilation rates to capture contextual information and multi-scale details from varying receptive fields, thereby obtaining richer representations of global features. This strategy enhances our understanding of the overall structure of surrounding tissues and improves segmentation accuracy for complex boundaries.
2) A novel artificial neuron, DNM, has been introduced and employed at the channel level for the first time to address challenges associated with local feature representation. This approach facilitates the refinement of boundaries and enables superior, precise processing of local features and fine-grained structures.
3) We introduce 3DL-Net, a novel medical image segmentation method designed to address the limitations of representing global-local features. To the best of our knowledge, this is the first attempt to tackle this challenge by leveraging DNM and dilated convolutions. Experimental evaluations demonstrate 3DL-Net remarkable accuracy and robustness, surpassing established state-of-the-art methods in the three public medical image segmentation datasets.

The remainder of this paper is structured as follows: Section 2 provides an overview of relevant literature and previous research in the medical image segmentation field. Section 3 details the proposed method in this study. The experimental design and

results are discussed in Section 4. Finally, Section 5 presents conclusions and outlines future work.

## 2. Related work

### *2.1. Medical image segmentation based on deep learning*

CNNs have made significant strides in advancing segmentation methods over the years. One of the pioneering models in this field is the FCN [22], which is the first to propose an end-to-end approach for segmentation. It enables pixel-wise classification through fully convolutional layers. This breakthrough brings segmentation tasks into the realm of deep learning and marks the beginning of a new era for image segmentation. Following FCN, the U-Net architecture, proposed by Ronneberger et al. [23], gained popularity for its success in biomedical image segmentation. U-Net is characterized by its U-shaped architecture, which combines contracting and expanding paths to effectively capture both low-level and high-level features. This design has proven particularly valuable in scenarios with limited training data and has been widely adopted in various medical imaging tasks. Another noteworthy model is SegNet [24], which employs an encoder-decoder architecture with skip connections to reconstruct segmented images while preserving spatial information. This architectural design enables SegNet to better capture the spatial positions of objects and accurately delineate their contours.

In recent years, inspired by the success of Transformer in computer vision [45, 46, 47, 48], attention-based visual Transformer methods have been introduced to leverage pixel correlations within medical images, aiming to enhance the effectiveness of medical image segmentation. Initially applied in conjunction with convolutional networks or to replace specific components of such networks, a groundbreaking approach emerged with Dosovitskiy's introduction of the Vision Transformer [49]. It departs from the reliance on convolutional networks by directly applying the Transformer to sequences of image patches. This departure from traditional convolutional approaches has demonstrated remarkable success in image classification tasks, as evidenced by its excellent performance on various benchmarks. TransUNet [46] stands out as the pioneer in utilizing a Transformer architecture for medical image segmentation challenges.

It ingeniously combines the local feature extraction capability of CNNs with the Transformer's unique ability to capture long-range relationships. Building upon this innovation, SwinUnet [50] further advances the TransUNet framework by introducing a significant modification, substituting the traditional Transformer with a swin Transformer backbone [51]. In this extension, SwinUnet adopts an asymmetric swin Transformer-based decoder with patch expansion layers for efficient upsampling, which aims to restore the spatial resolution of feature maps. In the most recent study, BRAUnet++ [52] combines the advantages of CNNs and Transformer to learn global semantic information and minimize local spatial information loss, and amplify the global dimension-interaction of multi-scale features. However, it's important to note that while Transformer models excel at capturing global relationships, there are challenges in handling certain image segmentation tasks, particularly those requiring the detailed analysis of local features. Additionally, in the domain of image processing, pre-training Transformer models may necessitate larger datasets to achieve optimal performance.

### *2.2. Segmentation methods through enhancing global feature representation*

Global features are essential for understanding the contextual information of medical images. While the receptive field of features in the last layer of the network theoretically encompasses a large portion of the input image, in practice, the empirical receptive field is often much smaller. This limitation renders it insufficient to adequately capture global features. Recognizing the limitation of insufficient global feature representation due to limited receptive fields, ParseNet [53] introduces global average pooling. This method aggregates the context features from the last layer or any target layer, thereby enhancing the network's ability to capture global context information. By integrating global features into the local feature map, it provides sufficient global context information. However, traditional average pooling only takes into account local information on a fixed scale. As a result, traditional networks are limited to learning a relatively narrow range of feature representations and cannot effectively capture multi-scale features of the image. In addition, pyramid pooling [35] allows the network to comprehensively capture the overall information of an image, including global structure and context at different scales, by extracting feature maps at different scales.

Pooling is a straightforward and effective method; however, it introduces a potential limitation. The downsampling process may lead to the exclusion of small-sized targets, thereby limiting the network's ability to accurately delineate fine-grained details and small objects within the image. The introduction of dilated convolution solves these problems well while expanding the receptive field. It preserves the image details better and provides a clearer representation of the features. Inspired by this, DeepLab [54] utilizes dilated convolutions, also known as hole convolutions, for semantic image segmentation. It achieves an expanded receptive field and captures contextual information at multiple scales by employing dilated convolutions with varying dilation rates, eliminating the need for pooling layers. This allows the model to preserve both spatial resolution and fine-grained details. They subsequently propose a method combining pyramid structure and dilated convolution to further improve the network's ability to perceive multi-scale global information [55]. Building on this foundation, RRCNet [37] similarly captures more global information by leveraging diverse receptive fields provided by dilated convolution. The difference is that RRCNet incorporates six deep supervised modules to guide the network learning to predict accurate segmentation masks at different scales. It enhances the model's capability to capture both fine-grained details and coarse-grained semantics from features at various scales.

Building upon the aforementioned study, we propose modules designed to enhance the representation of global features. By incorporating dilation convolution with a pyramidal structure and integrating deep supervision into the network, our method improves its representation of global features. However, while dilation convolution expands the receptive field to capture more global information, it may lead to the lost of detailed local features [56]. Therefore, it is crucial to ensure a more detailed and enriched process for local features while considering the enhancement of global features [57]. In this paper, we not only propose modules for enhancing global feature representation but also address the challenges associated with local feature representation through dendritic learning.

*2.3. Dendritic learning*

Traditional artificial neural networks, rooted in McCulloch-Pitts neurons [42], have been fundamental in shaping the field. Contemporary research is now focusing on architectures that provide increased biological interpretability. One intriguing avenue involves the integration of dendritic neurons, drawing inspiration from the intricate dendritic structures seen in biological neurons [43]. Dendrites, the branching extensions of neurons, play a pivotal role in processing incoming neural signals within the brain [58]. The conceptual utility of dendrites in neuroscience underscores the limitations of traditional neuron description, which neglects dendrites' computational properties [59, 60]. The complexity of neuronal behavior surpasses simple point neuron descriptions by including dendrites in their design for a more efficient capture of their true function and the crucial role dendrites play in information processing and overall neural complexity.

Emulating the adaptive behavior of dendrites, dendritic neurons aim to overcome limitations in traditional neural network architectures by effectively capturing local features. The concept of dendritic learning has emerged as a compelling research direction, aiming to incorporate the advantages of biological neurons into artificial intelligence. Consequently, the incorporation of dendritic learning into deep learning models has shown promise in various domains, including image classification, object recognition, and natural language processing [61]. By developing artificial dendrites, Li et al. [62] present a novel fully integrated neural network with synapses, dendrites, and soma, demonstrating significant improvements in capturing local features and overall performance, particularly in terms of reduced power consumption and improved accuracy in tasks like digit recognition. Egrioglu et al. [63] propose a new recurrent dendritic neuron model artificial neural network with a particle swarm optimization-based training algorithm, demonstrating superior forecasting performance in time series prediction. Gao et al. [64] introduce a complex-valued dendritic neuron model by extending DNM from a real-valued domain to a complex-valued domain. Moreover, numerous research endeavors have integrated the notable capability of dendritic networks in effectively addressing nonlinear problems, resulting in improved accuracy and robustness in the field of image classification [65, 66, 67].

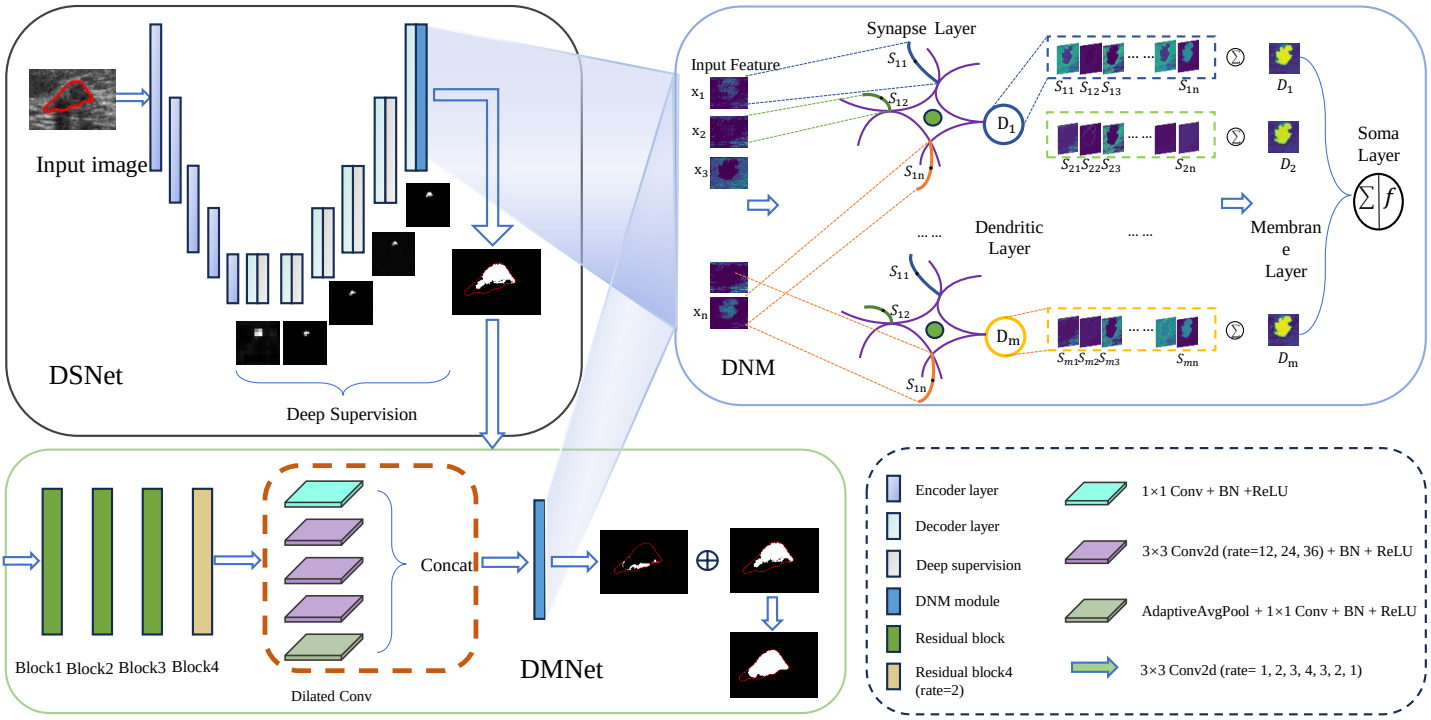


Figure 1: The basic framework of 3DL-Net, the instance in DSNet and DMNet and the feature maps in the DNM module are the result under random selection.

Inspired by the structure and function of retinal ganglion cells, dendritic networks are considered to be a more biologically plausible type of artificial neuron [65, 68, 69]. The design of such neural networks fits with the need for interpretability in medical imaging tasks [70]. However, although dendritic networks have demonstrated excellent performance in a wide range of fields, their application to image segmentation tasks is still relatively limited. We aim to further optimize the structure of dendritic networks to better suit the needs of such tasks. We propose modifying the synapse, dendritic, and soma layers of the dendritic network to incorporate channel-level features, thereby enhancing its applicability to image segmentation tasks. We aim to address specific challenges in medical image analysis, such as accurate delineation of structures and refining the processing of local features, by optimizing the structure of dendritic networks for image segmentation tasks.

## 3. Methodology

### *3.1. Architecture of 3DL-Net*

This section provides a detailed exposition of our proposed method. It combines three integral modules: the preliminary segmentation module with deep supervision based on variant SegNet (DSNet), the multi-scale contexual module for missed lesions area detection (DMNet), and the shallow feature process utilizing DNM. Each section provides a detailed explanation of these components, with the architecture of 3DL-Net shown in Figure 1.

The initial step of our segmentation process employs the DSNet module, which conducts the preliminary segmentation of the input medical images. Subsequently, the output of DSNet is fed into DMNet, which specializes in detecting and rectifying missed segmentation regions, further refining the segmentation results. By integrating multi-scale features using dilated neural networks, we can effectively capture fine-grained details and semantic information in images, enhancing adaptability to structures of various sizes and shapes in medical images. This integration also improves the representation and understanding of global features, providing a comprehensive view of the image context and relationships. Throughout the entire segmentation process, DNM plays a pivotal role by cascading at the end of DSNet and DMNet. The shallow features of the two modules are further refined to obtain a richer local feature representation.

### *3.2. Preliminary segmentation module with deep supervision based on variant SegNet*

As previously mentioned, the segmentation networks for medical images require the ability to process global-local features. The utilization of multi-scale information at different network levels to enhance global feature representation is an effective method. The shallow layers of the network mainly learn the edge and color information of the image, while the deep layers are responsible for capturing higher-level semantic details. Therefore, it is beneficial to leverage layers of varying scales and integrate highly semantic multiscale feature maps. In addition, extracting location information of anatomical structures or abnormalities from different imaging modalities in medical

imaging, including CT, ultrasound, and MRI, can significantly improve the accuracy and reliability of segmentation results [71]. This information provides contextual information about the entire image, which contributes to the extraction of global features.

In consideration of these issues, we employ SegNet [24] as the backbone. SegNet has a concise yet effective encoder-decoder structure that enables efficient extraction of features from images and captures highly semantic multiscale information. The decoder in the SegNet network recovers spatial information by performing accurate upsampling using a max-pooling index. Skipping connections enhances context capture by integrating features from different levels of encoders and decoders while preserving spatial information. This is also beneficial for processing local features, as it allows the network to better understand the relationship between the local features and the context information. These superiorities are an explicit response to the challenges faced by backbone networks.

Building on this, incorporating deep supervision into the method further enhances the feature representation and extraction capabilities of the network. As indicated in Figure 1, DSNet inserts supervision signals at various levels of the decoder hierarchy; deep supervision guides the learning process of the backbone network. This guidance encourages the generation of multi-scale features, enabling the backbone network to effectively capture both local and global information from medical images [72]. Moreover, DSNet extends the original SegNet architecture by introducing an additional middle layer. It is utilized as the first output for deep supervision, providing an initial coarse result. This improvement allows the network to receive additional supervision signals at an early stage of feature extraction, facilitating the learning process and improving the network's ability to capture both local and global features effectively. As illustrated in Figure 2, DSNet produces multiple segmentation masks by introducing the deep supervision module: DS1 for the additional intermediate layer, DS2 for the intermediate layer, and DS3, DS4, and DS5 for each layer of the upsampling process. For each output mask, the feature map's size is initially upsampled through nearest-neighbor interpolation to restore the resolution, facilitating comparison with the original image. Subsequently, these masks undergo dimension reduction and normalization to obtain pixel-level probabilities. The final mask result, DS6, encompasses more comprehen-

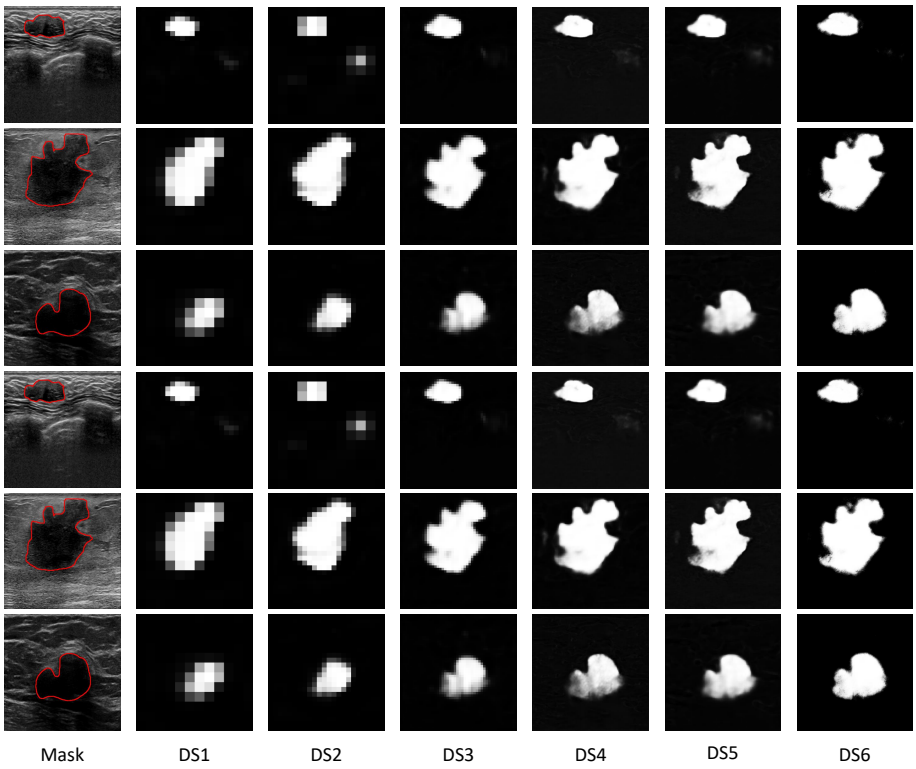


Figure 2: The visualization of deep supervision.

sive feature information with higher confidence. Consequently, DS6 is considered the final output of DSNet and acts as the input for the next stage.

### *3.3. Multi-scale contexual module for missed lesions area detection*

In medical image segmentation, one of the challenges is the lack of contextual information and global features at different scales. This deficiency may lead to the presence of missed detection areas in the medical segmentation task, as factors such as image noise, variability in tissue appearance, or artifacts could disrupt the process. In addition, the CNN-based method may fail to accurately capture these larger lesions due to its limited perception field and may overlook certain areas of interest. For instance, tumors in the middle and later stages tend to be larger in size compared to those in the early stages. Furthermore, when upsampling and downsampling, spatial information

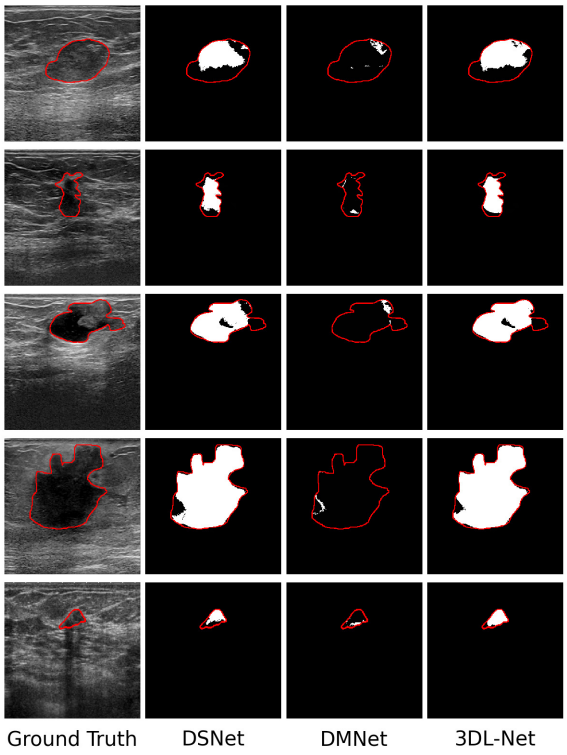


Figure 3: The visualization results of DSNet and DMNet.

may be lost due to the encoder-decoder structure.

In response to these challenges, we propose the DMNet. The size of the perceptual field essentially dictates the extent of contextual information that can be effectively utilized [14]. Larger receptive fields can bring about a richer representation of global features. Inspired by this, DMNet extends the DSNet architecture by incorporating a missed lesions area detection module based on dilated convolution. It enhances sensitivity to missing lesion areas or fine anatomical details through multi-scale capabilities. Besides, it is beneficial to address the limitation of missing spatial information and inadequate representation of global features, which enables more effective detection and aggregation of spatial information at different scales.

Figure 3 displays distinct components: the labeled region corresponding to the lesion area in the breast cancer image, the prediction outcomes generated by DSNet,

the prediction masks by DMNet, and the ultimate composite prediction masks. Owing to the initial predictions from DSNet, there are numerous undetected areas within the lesion region. We develop DMNet to integrate dilation convolution with varying dilated rates, with the goal of capturing the residuals of the lesion region. It results in segmentation with more complete details.

Inspired by the PSPN [35] and dilated convolutions [54], we leverage a pyramid structure and dilated convolutions to effectively capture global features and contextual information at multiple scales, aiming to address the challenges of spatial information loss and inadequate representation of global features in medical image segmentation tasks. Figure 1 The first three blocks of ResNet50 are used in their original structure, while the last block utilizes dilation convolution with a dilation rate set to 2. the architecture of DMNet, specifically, this module first consists of an initial feature extraction branch composed of the residual block. The first three blocks of ResNet50 are used in their original structure, while the last block utilizes dilation convolution with a dilation rate set to 2. Additionally, there are five branches of $3 \times 3$ dilated convolutions with different dilation rates, followed by a final branch of DNM. After receiving the feature outputs from DSNet, DMNet initiates feature extraction through the initial feature extraction branch, resulting in five feature maps. DMNet then combines the extracted features from different scales and levels of abstraction by concatenating these maps. Subsequently, the missed lesion areas identified by DMNet are obtained through the DNM module. Our method fuses these areas with the preliminary rough segmentation results from DSNet to refine the final prediction mask.

The pyramid structure of DMNet facilitates the extraction of both global and local features across multiple scales. This improves the model's ability to recognize lesion areas and structures of different sizes. Additionally, it effectively addresses the challenge of losing global feature information during the continuous downsampling process of the encoder through dilated convolutions. Each layer employs diverse dilation rates to attain varied receptive fields for capturing multi-scale information. This structure enables small convolutional kernels to capture fine features, while large convolutional kernels can extract large-scale features. This aids in effectively detecting the lesion region by comprehensively grasping structures and features in the image.

Notably, besides the enhancements in network structure, we employ dendritic learning instead of a 1×1 convolution function at the output of both DSNet and DMNet to achieve the final feature mapping. Further details on the specific implementation of DNM and the visualized feature map will be provided in the next section.

### *3.4. Dendritic neuron module*

To enhance the performance of the backbone and effectively map the feature information after rich local feature extraction, we introduce the DNM module. It compensates for the lost of local feature information caused by the larger receptive field introduced by the dilated convolution. As illustrated in Figure 1, both DSNet and DMNet incorporate our proposed DNM module.

In contrast to prior research [66, 65, 73], our method concentrates on optimizing the convolution and introduces an additional preprocessing step to enhance the utilization of the dendritic layer. The input image undergoes operations, such as convolutional and pooling layers, within the segmented backbone network. These operations produce a set of maps containing high-level features, which are subsequently utilized for dendritic learning. These features represent an abstract understanding of the complex shape and texture information pertaining to the lesion region in the image.

To incorporate dendritic learning into the model using a convolution-like approach, we reshape the utilization method of dendritic layers. Initially, we perform multiple dimensionality transformations on input features. Subsequently, we apply layer normalization to these features, ensuring optimal data scaling within the appropriate range. The input data is guided to the synapse layer, where input feature vectors are replicated along channel dimensions to correspond with the number of dendritic branches, promoting desired feature reuse. Following feature reuse, the feature maps undergo initial processing in the synapse layer, involving multiplexing and weighting. This process yields multiple sets of feature maps that more effectively capture the input image features. The processing in the synapse layer can be formulated as follows:

$$S_{ij} = \mathrm{ReLU}(k \cdot (w_{ij} \cdot \mathrm{Norm}(x_i) - q_{ij})) \tag{1}$$

Here, $S_{ij}$ represents the output of the $j$-th dendritic branch corresponding to the $i$-th

element in the input vector $x$. The computational process involves element-wise multiplication between the weights $w_{ij}$ and the input $x$, followed by subtraction of the threshold $q_{ij}$. Subsequent to this operation, the rectified linear unit (ReLU) activation function is applied, introducing non-linearity. The ReLU activation function serves to rectify negative values, thereby preserving salient features, attenuating noise, and disregarding extraneous information. The amplification of the resultant signal is modulated by the parameter $k$. Equation 1 encapsulates the crux of information integration and processing within the dendritic layer. The learnable parameters, including $k$, $w_{ij}$, and $q_{ij}$, are randomly initialized in the range $(0, 1)$.

After the individual feature maps $S_{ij}$ traverse the synapse layer, they are subsequently channeled into the dendritic layer. Within this layer, each feature map undergoes a concatenated addition computation, skillfully orchestrated by the dendritic layer. In this intricate process, the $j$-th dendritic branch adeptly consolidates the $N$ input signals $S_{ij}$, culminating in the following computation:

$$D_j = \sum_{i=1}^{N} \text{Norm}(S_{ij}) \tag{2}$$

In this expression, $D_j$ serves as the symbolic representation of the amalgamated output emanating from the $j$-th dendritic branch, encapsulating the synergized information gleaned from the $N$ input signals.

Following the intricate computations in the dendritic Layer, the subsequent stage involves the membrane layer which autonomously oversees the segmentation prediction process across multiple layers. In this phase, the membrane layer accumulates the signals from all dendritic branches. Through a summation operation, this layer combines the outputs of the $M$ dendritic branch representation, denoted as:

$$O = \sum_{j=1}^{M} D_j \tag{3}$$

Here, $O$ symbolizes the profound integration of inputs derived from the intricate processes within the dendritic layer, ultimately representing the collective output of this stage.

The soma Layer concludes the processing pipeline by generating the final prediction. Employing a sigmoid activation function, it processes the output of the membrane

layer. Yields the final output prediction $P$ according to the equation:

$$P = \frac{1}{1 + e^{-k_s(Y - q_s)}} \tag{4}$$

Here, $k_s$ and $q_s$ represent additional learnable parameters, initially set randomly within the range of $(0, 1)$. Throughout the training phase, the optimization of these learnable parameters is facilitated by the Adam optimizer.

The introduced DNM module empowers shallow feature processing, enabling the acquisition of finer local feature representations and more accurate segmentation masks in medical image segmentation tasks. This refinement of local features effectively addresses the issue of loss of fine-grained details in shallow features of medical image segmentation, providing a biologically plausible and interpretable basis. The challenge of losing local feature information due to the large receptive fields introduced by dilated convolution in DMNet is effectively mitigated. In Figure 4, input feature maps are preprocessed and passed through the synapse layer, resulting in $M$ groups of enriched shallow features. Each group then undergoes processing by the dendritic and membrane layers to produce $M$ middle feature maps. Finally, the soma layer integrates the middle feature maps to generate the final prediction results. The DNM module executes nonlinear feature mapping, enhancing the model's capacity to capture intricate details and fine-grained structures in medical images, thereby improving segmentation accuracy.

### *3.5. Loss function*

During the training, a combined loss function incorporating Binary Cross-Entropy Loss ($\mathcal{L}_{\text{BCE}}$) and Self-Adaptive Focal Loss ($\mathcal{L}_{\text{a-Focal}}$) is proposed. This choice arises from the suitability of BCE for pixel binary classification tasks and focal loss for segmentation tasks. It can focus more on hard or misclassified pixels and alleviate class imbalance. The proposed self-adaptive nature of our $\mathcal{L}_{\text{a-Focal}}$ is designed to address challenges specific to medical segmentation by dynamically adjusting weights based on the proportion of lesion areas in different data samples. $\mathcal{L}_{\text{BCE}}$ is given by:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^{N} (t_i \cdot \log(p_i) + (1 - t_i) \cdot \log(1 - p_i)) \tag{5}$$

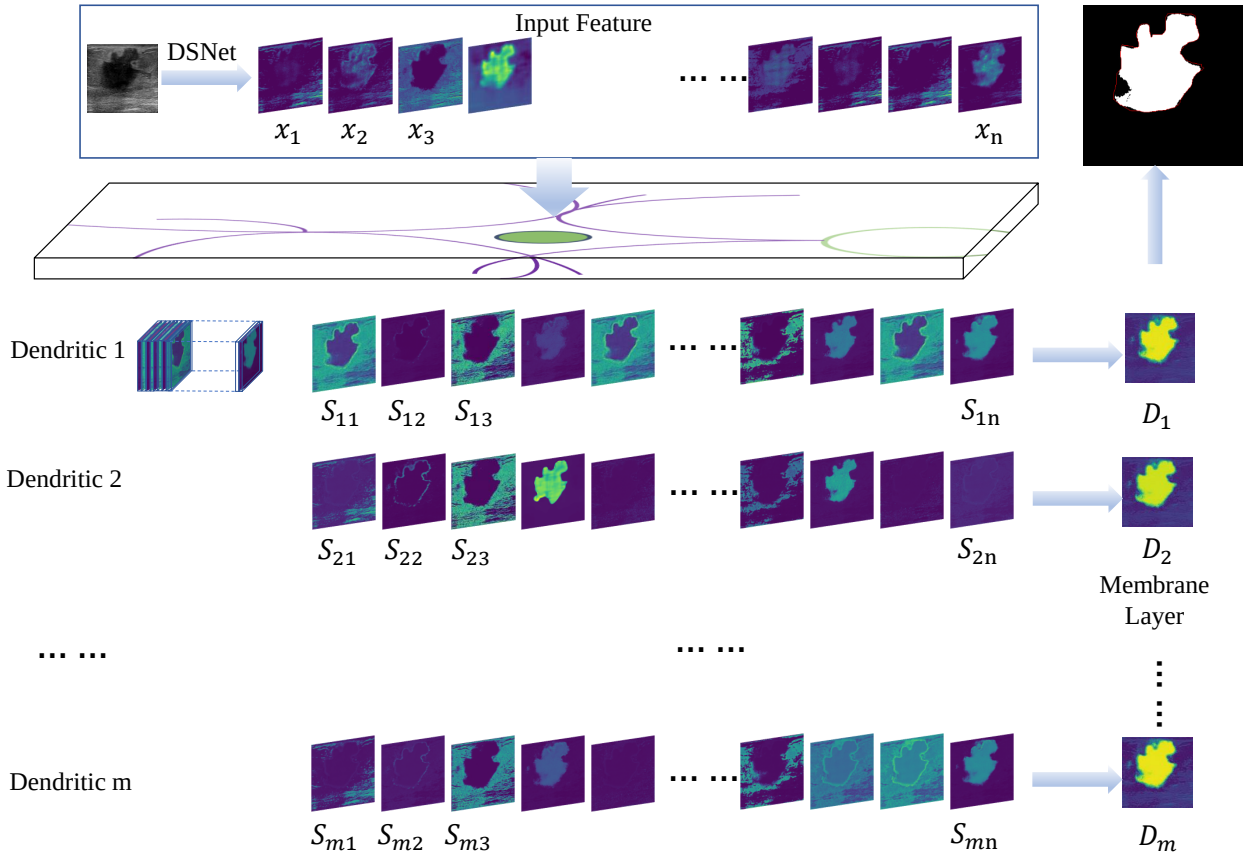


Figure 4: Visualization of the shallow feature processing of DNM.

where $N$ is the total number of pixels, $t_i$ represents the ground truth label for pixel $i$, and $p_i$ represents the predicted probability.

$\mathcal{L}_{\text{a-Focal}}$ is designed to alleviate the challenges caused by uneven distribution of classes and is formulated as follows:

$$\mathcal{L}_{\text{a-Focal}} = -\frac{1}{N}\sum_{i=1}^{N} \alpha_i \cdot (1 - p_i)^{\gamma_i} \cdot \log(p_i) \tag{6}$$

where $\alpha_i$ is dynamically adjusted based on the confidence level, and $\gamma_i$ serves as a focal factor adjusting the loss based on the predicted probability $p_{g_i}$.

$$\gamma_t = (1 - p_g) \cdot \mathbb{I}_{0.15 \le p_g \le 0.85} + 0.85 \cdot \mathbb{I}_{p_g = 0.15} + 0.15 \cdot \mathbb{I}_{p_g = 0.85} \tag{7}$$

This expression defines the focal factor $\gamma_t$ used in the $\mathcal{L}_{\text{a-Focal}}$. It adjusts the loss differently based on the predicted probability $p_g$. Specifically: When $p_g$ is within the range [0.15, 0.85], $\gamma_t$ is set to $1 - p_g$, emphasizing learning from examples that are more uncertain (closer to 0.5). When $p_g$ is within the range [0, 0.15], $\gamma_t$ is fixed at 0.85, giving more emphasis to examples predicted with high confidence as positive. When $p_g$ is within the range [0.85, 1], $\gamma_t$ is fixed at 0.15, providing more emphasis to examples predicted with high confidence as negative. This configuration is designed to mitigate overfitting issues associated with unbalanced samples, particularly when the lesion area varies significantly in size.

The final combined loss, denoted as $\mathcal{L}$, aggregates $\mathcal{L}_{\text{BCE}}$ and $\mathcal{L}_{\text{a-Focal}}$ for deep supervision, weighted by a factor $k$:

$$\mathcal{L} = \sum_{i=1}^{N} \mathcal{L}_{\text{BCE}} + k \cdot \sum_{i=1}^{N} \mathcal{L}_{\text{a-Focal}} \tag{8}$$

The weighting factor, denoted as $k$, enables nuanced adjustment in emphasizing the $\mathcal{L}_{\text{a-Focal}}$ component. This parameter plays a key role in striking a balance between the two final segmentation loss results and the intermediate auxiliary segmentation loss results embedded in our proposed loss function. This parameter serves as an adaptive adjustment between pixel-wise binary classification and addressing the imbalance in segmentation weights. Through the modulation of $k$, we exercise control over the relative impact of each loss component during the overall training process. A higher $k$ accentuates the influence of the $\mathcal{L}_{\text{a-Focal}}$, customized to address segmentation challenges by dynamically adapting to variations in lesion area proportions. Conversely, a lower $k$ prioritizes the ($\mathcal{L}_{\text{BCE}}$), enhanced for pixel-wise binary classification weighting. This nuanced control through the parameter $k$ provides researchers and practitioners with a flexible mechanism to refine the model's learning objectives, allowing them to formulate the weight tailored to the specific requirements of their medical image segmentation tasks. Experimentation with diverse $k$ values facilitates the optimization of the network's performance in alignment with the dataset nuances and segmentation goals.

## 4. Experiment

### *4.1. Datasets and implementation details*

To evaluate the effectiveness of 3DL-Net in medical image segmentation, a comprehensive set of experiments is conducted on three widely used public datasets. This section provides details on the experimental datasets, setups, comparative experiments, ablation experiments, parameter discussions, and visualization of the results.

The study utilizes three published datasets: the moderately-sized Breast Ultrasound (BUS) dataset [74], a significantly smaller subset of the BUS dataset (STU) [75], and a relatively larger COVID-19 dataset [76]. The inclusion of diverse medical imaging modalities, including CT and ultrasound, in these datasets highlights the generalizability and robustness of our proposed method. The first dataset is the BUS dataset, collected in 2012 from the UDIAT Diagnostic Centre of the Parc Taulí Corporation, Sabadell (Spain), utilizing a Siemens ACUSON Sequoia C512 system equipped with a 17L5 HD linear array transducer operating at 8.5 MHz. The dataset comprises 163 images from various women, with an average image size of $760 \times 570$ pixels. Each image contains one or more lesions. Among the 163 lesion images, 53 depict cancerous masses, while 110 showcase benign lesions. The STU dataset only comprises 42 BUS images, each with an average size of $128 \times 128$ pixels. These images are acquired by the Imaging Department of the First Affiliated Hospital of Shantou University, utilizing the GE Voluson E10 ultrasonic diagnostic system. The final dataset consists of COVID-19 lesion masks and their corresponding frames compiled from three public datasets. The dataset comprises 2729 pairs of images and corresponding ground truth masks. To maintain consistency across datasets, the masks of all various types of lesions have been uniformly mapped to the color white.

During the experiments, the following hardware and software configurations are employed: Intel(R) Xeon(R) Silver 4110 CPU @ 2.10 GHz, NVIDIA GeForce RTX A6000, and PyTorch 2.1.0 as the backend. To ensure equitable comparisons in training assessments, the COVID-19 dataset is partitioned using a consistent data partitioning strategy, the BUS dataset and the STU dataset are subjected to quadruple cross-validation studies, which are a frequently utilized method in medical image segmen-

tation. The input images are resized to 384 × 384 pixels and normalized to a mean value of [0.485, 0.456, 0.406] with a standard deviation of [0.229, 0.224, 0.225]. Additionally, a ResNet backbone pre-trained on ImageNet [77] is utilized as the DMNet encoder. In the training phase, a warm-up strategy is adopted, and the Adam optimization algorithm is employed to optimize the network. The initial learning rate is set to $5e-5$, epoch is set to 100, and the batch size is set to 16. The learning rate underwent a warm-up during the first epoch, gradually increasing from a very small value to the set initial learning rate and then decaying slowly.

To ensure a fair comparison among experiments, each folded cross-validation experiment randomly selected 25% of the validation data. The best-performing model on the validation set, determined after loss convergence, is utilized for all methods. During the testing process, images is resized to 384 × 384 without employing post-processing optimization strategies. To underscore the reliability and authority of our chosen evaluation metrics, we employed established criteria as recommended by prior research [78, 79, 80]. Specifically, evaluation of segmentation performance for BUS and pneumonia involved six established metrics: Precision, Recall, mDice, mIou, Specificity, and F1.

### *4.2. Evaluation metrics*

Drawing from recent work, we leverage comparable segmentation metrics to conduct a comprehensive evaluation of our model. Six evaluation indicators are utilized to assess the model's performance. In this section, we provide a meticulous exposition of each metric.

In the context of medical image segmentation, the definitions for true positive (TP), true negative (TN), false positive (FP) and false negative (FN) are as follows:

TP: Pixels are accurately identified as part of the diseased area when the predicted parameter value is above the defined threshold (indicating the presence of the disease).

FP: Pixels are mistakenly identified as part of the diseased area when, in reality, the predicted parameter value is below the defined threshold (indicating the absence of the disease).

TN: Pixels are correctly recognized as not part of the diseased area when the predicted parameter value is below the defined threshold (indicating the absence of the disease).

FN: Pixels representing the disease but incorrectly identified as not part of the diseased area because the predicted parameter value is above the defined threshold (indicating the presence of the disease).

Precision represents the ratio of TP redictions to all positive predictions, and Recall represents the proportion of TP predictions to all positive ground truth pixels. Specificity assesses the proportion of TN predictions to all negative ground truth pixels. The Jaccard index quantifies the similarity between predicted and ground truth segmentations. The mDice Coefficient, similar to mIou, quantifies the similarity between predicted and ground truth segmentations, particularly focusing on overlapping areas. In the realm of image segmentation, the F1 score serves as a comprehensive metric, amalgamating model accuracy and recall to deliver a holistic evaluation of the model's performance in the segmentation task. These six indicators serve as our benchmarks for evaluating the model. The specific formulas are as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{9}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{10}$$

$$\text{mDice} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \tag{11}$$

$$\text{mIou} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{12}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{13}$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{14}$$

### *4.3. Comparative experiments*

In this subsection, the 3DL-Net model's segmentation outcomes across three datasets are presented, and we conduct a comparative analysis against various medical image segmentation models. Our assessment comprehensively explores the model's learning capability and generalization, considering both quantitative and visualization analysis. For quantitative evaluation, performance metrics, including those in Eqs.(9) to (14) are employed, comparing the 3DL-Net model with other models. U-Net [23], U-Net++ [81], and SegNet [24] are the traditional methods for image segmentation tasks. AttU-Net [82] is the classical method that introduces the mechanism of attention. Transunet [46] and BRAUNet++ [52] are the classical and state-of-the-art methods based on transformers. AAUNet [36] and RRCNet [37] are state-of-the-art methods that use dilated convolution, and MBSNet [83] is one of the latest methods for medical image segmentation. Additionally, qualitative results are visually presented for each model, accompanied by an in-depth analysis of selected cases. We have highlighted the best results for each metric in bold and the second best in italics. Additionally, we have emphasized the most frequently bolded result, representing the best method, and indicated the second best method with underlining.

Table 1: The comparison experiment results (mean ± std) in the BUS dataset.

| Methods | U-Net | U-Net++ | AttU-Net | SegNet | TransUNet | AAUNet | MBSNet | BRAUNet++ | *RRCNet* | **3DL-Net** |
|---|---|---|---|---|---|---|---|---|---|---|
| Precision(%) | 88.31 ± 2.58 | 86.59 ± 3.12 | 85.27 ± 4.81 | 79.70 ± 3.63 | 80.36 ± 6.84 | 78.83 ± 2.40 | 86.93 ± 0.86 | 84.81 ± 3.32 | 81.77 ± 4.12 | **88.36 ± 2.03** |
| Recall(%) | 77.06 ± 5.74 | 80.44 ± 2.45 | 75.52 ± 7.31 | 66.81 ± 6.24 | 73.53 ± 7.92 | 82.22 ± 3.84 | 73.05 ± 2.36 | 72.31 ± 1.41 | 82.82 ± 1.70 | **85.67 ± 2.00** |
| mDice(%) | 81.46 ± 5.44 | 82.76 ± 2.79 | 78.27 ± 4.20 | 72.25 ± 3.33 | 75.84 ± 5.66 | 78.14 ± 2.41 | 78.17 ± 1.92 | 76.99 ± 2.17 | 80.40 ± 2.31 | **86.61 ± 0.70** |
| mIou(%) | 69.55 ± 7.15 | 71.38 ± 3.68 | 66.04 ± 5.03 | 56.85 ± 4.19 | 61.86 ± 7.35 | 69.10 ± 2.98 | 65.53 ± 1.92 | 63.18 ± 3.01 | 71.81 ± 2.74 | **76.85 ± 1.04** |
| Specificity(%) | 99.45 ± 0.20 | 99.30 ± 0.24 | 99.36 ± 0.26 | 99.18 ± 0.15 | 98.98 ± 0.50 | 98.82 ± 0.35 | 99.46 ± 0.20 | 99.39 ± 0.13 | 99.01 ± 0.39 | **99.52 ± 0.10** |
| F1(%) | 84.99 ± 4.06 | 84.63 ± 2.84 | 81.48 ± 2.49 | 75.72 ± 2.11 | 77.95 ± 5.40 | 80.49 ± 2.95 | 82.30 ± 1.05 | 80.70 ± 2.60 | 82.29 ± 2.41 | **87.47 ± 1.18** |

#### *4.3.1. Quantitative analysis*

Table 1 illustrates that 3DL-Net achieves a dominant advantage over second place on the BUS dataset. In particular, we improve 5.04% on mIoU and outperform U-Net++ by about 3.85% on mDice. Compared to RRCNet, our method achieves a Recall

Table 2: The comparison experiment results (mean ± std) in the STU dataset.

| Methods | U-Net | U-Net++ | AttU-Net | SegNet | TransUNet | AAUNet | MBSNet | BRAUNet++ | *RRCNet* | **3DL-Net** |
|---|---|---|---|---|---|---|---|---|---|---|
| Precision (%) | 88.28 ± 4.32 | 87.37 ± 7.53 | **90.74 ± 2.65** | 81.54 ± 4.08 | 90.35 ± 1.62 | 89.30 ± 2.82 | 87.41 ± 4.81 | 90.03 ± 2.97 | 88.35 ± 5.40 | 90.22 ± 0.36 |
| Recall (%) | 84.84 ± 3.44 | 83.53 ± 1.27 | 79.33 ± 6.85 | 85.01 ± 4.81 | 82.01 ± 2.01 | 81.66 ± 5.44 | 85.00 ± 5.69 | 80.51 ± 5.65 | 85.32 ± 3.30 | **85.71 ± 3.84** |
| mDice (%) | 86.14 ± 2.35 | 85.08 ± 3.66 | 84.30 ± 3.95 | 83.59 ± 3.39 | 85.97 ± 1.30 | 85.18 ± 2.71 | 86.03 ± 3.46 | 84.84 ± 2.01 | 86.54 ± 2.99 | **87.87 ± 2.05** |
| mIou (%) | 75.89 ± 3.68 | 74.49 ± 5.57 | 73.18 ± 5.99 | 71.97 ± 5.02 | 75.40 ± 1.98 | 74.26 ± 4.05 | 75.74 ± 5.21 | 73.71 ± 3.06 | 76.46 ± 4.68 | **78.41 ± 3.23** |
| Specificity %) | 98.42 ± 0.66 | 97.93 ± 1.90 | 98.82 ± 0.12 | 97.35 ± 0.32 | 98.70 ± 0.25 | 98.57 ± 0.88 | 98.09 ± 0.90 | 98.76 ± 0.70 | 98.40 ± 0.93 | **98.83 ± 0.12** |
| F1 (%) | 87.17 ± 3.10 | 86.17 ± 5.54 | 87.36 ± 2.58 | 82.53 ± 3.52 | 88.10 ± 1.27 | 87.16 ± 1.94 | 86.68 ± 3.63 | 87.31 ± 1.02 | 87.41 ± 4.00 | **89.02 ± 1.08** |

Table 3: The comparison experiment results (mean ± std) in the COVID-19 dataset.

| Methods | U-Net | U-Net++ | AttU-Net | SegNet | TransUNet | AAUNet | MBSNet | BRAUNet++ | *RRCNet* | **3DL-Net** |
|---|---|---|---|---|---|---|---|---|---|---|
| Precision (%) | 91.92 ± 1.81 | **92.85 ± 0.81** | 90.84 ± 1.95 | 89.27 ± 0.73 | 91.57 ± 4.17 | 92.74 ± 0.95 | 91.04 ± 0.65 | 91.38 ± 0.57 | 80.43 ± 1.04 | 82.09 ± 0.17 |
| Recall (%) | 74.77 ± 1.52 | 74.16 ± 0.85 | 75.28 ± 1.63 | 75.40 ± 0.59 | 62.37 ± 2.61 | 72.87 ± 1.12 | 75.83 ± 0.46 | 70.37 ± 0.28 | 88.38 ± 1.14 | **88.39 ± 0.18** |
| mDice (%) | 82.27 ± 0.15 | 82.27 ± 0.19 | 82.15 ± 0.24 | 81.48 ± 0.22 | 73.71 ± 0.37 | 81.43 ± 0.34 | 82.53 ± 0.20 | 79.27 ± 0.07 | 84.11 ± 0.07 | **85.06 ± 0.03** |
| mIou (%) | 70.09 ± 0.22 | 70.13 ± 0.29 | 69.94 ± 0.35 | 69.14 ± 0.31 | 58.85 ± 0.47 | 68.93 ± 0.47 | 70.51 ± 0.25 | 66.01 ± 0.10 | 72.82 ± 0.12 | **74.16 ± 0.04** |
| Specificity (%) | 99.86 ± 0.04 | 99.88 ± 0.02 | 99.85 ± 0.03 | 99.84 ± 0.01 | 99.85 ± 0.09 | **99.89 ± 0.02** | 99.85 ± 0.02 | 99.86 ± 0.01 | 99.74 ± 0.02 | 99.76 ± 0.00 |
| F1 (%) | 82.44 ± 0.18 | 82.45 ± 0.21 | 82.30 ± 0.23 | 81.75 ± 0.20 | 74.09 ± 0.42 | 81.60 ± 0.33 | 82.74 ± 0.16 | 79.51 ± 0.07 | 84.25 ± 0.08 | **85.12 ± 0.03** |

improvement of about 2.85%, emphasizing lesion detection capability and reducing the risk of missed detection. Our Precision outperformed the closest competitor by approximately 0.05%. The Specificity of our method outperforms U-Net by approximately 0.07%. Finally, F1 Score shows a significant improvement, outperforming RRCNet by about 5.18%. These results demonstrate that our method indeed significantly improves segmentation accuracy on the BUS dataset.

Table 2 demonstrates the performance on the STU dataset. 3DL-Net performs slightly worse than AttU-Net in Precision, but this metric alone may not provide a comprehensive assessment of segmentation performance. Thus it primarily evaluates the classifier's accuracy in identifying lesion regions. Notably, AttU-Net exhibits significantly lower scores in Recall, mDice, and mIoU. This discrepancy suggests that it may not meet the essential requirements for medical image segmentation. Conversely, our model excels in other metrics, with improvements of 3.32%, 1.54%, 4.46%, 0.01%, and 2.12% compared to the second-best results in Recall, mDice, mIoU, Specificity,

and F1 Score.

Table 3 demonstrates the results of 3DL-Net and contrasting methods on the COVID-19 dataset. Although we have not achieved optimal results on Precision, it's worth noting that mDice may hold more significance in medical image segmentation tasks. This metric emphasizes the discovery and accurate segmentation of positive samples, such as lesions, which is crucial in medical diagnosis [84]. Moreover, mDice inherently combines aspects of both precision and recall metrics. For example, a higher mDice score indicates better overlap between predicted and ground truth segmentation masks, implying both high precision and recall. This is exemplified by the performance of U-Net++, which, despite achieving the best precision, exhibits mediocre performance on several other metrics, suggesting the importance of considering overall segmentation quality rather than focusing solely on precision. In addition, while our Specificity may not rank highest among the metrics, it solely evaluates the model's ability to correctly identify healthy regions (true negatives). Given that healthy regions constitute a significant portion of medical image segmentation tasks, many methods tend to achieve high scores in Specificity. However, relying solely on this metric may not provide a comprehensive assessment of segmentation effectiveness. Therefore, it's essential to consider a diverse range of metrics to accurately evaluate the model's performance in delineating both pathological and healthy regions. 3DL-Net achieved the best results in other indicators, highlighting our superior capability in lesion detection. It achieves 88.39% ± 0.18, 85.06% ± 0.03, 74.16% ± 0.04, and 85.12% ± 0.03 in Recall, mDice, mIoU, and F1 Score, with improvements of 0.01%, 0.95%, 1.34%, and 0.87% compared to the second-best results.

The significant gap between 3DL-Net and existing methods can be explained by the fact that the representation of global-local features has significantly improved. The proposed DMNet is able to obtain multi-scale semantic features at each level. In contrast to the U-Net, SegNet, and U-Net++ that use the same scale feature map in cascade, DMNet helps to preserve finer-grained details. In addition, the proposed DNM module provides more refined processing of local features. The limitation of insufficient local feature representation brought about by the introduction of dilated convolution is also addressed compared to RRCNet and AAUNet. In addition, the standard devia-

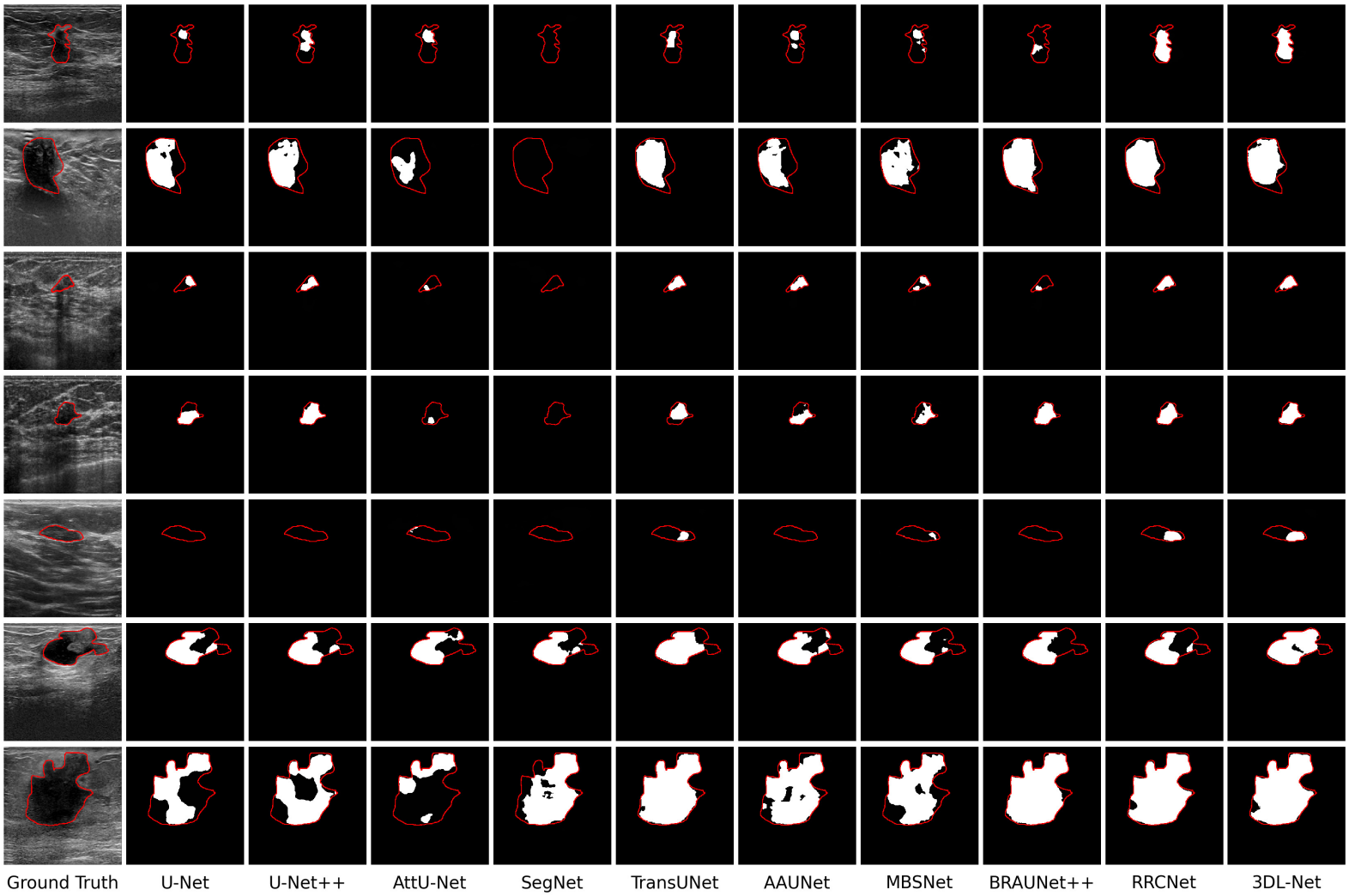


Figure 5: Visual Comparison with State-of-the-Art Methods on BUS dataset. White pixels represent the predicted values, and red curves represent the ground truth values.

tion values in the table indicate that our method exhibits significantly better robustness compared to other methods. This stability is particularly crucial in medical imaging segmentation tasks. It is essential for accurate lesion area detection.

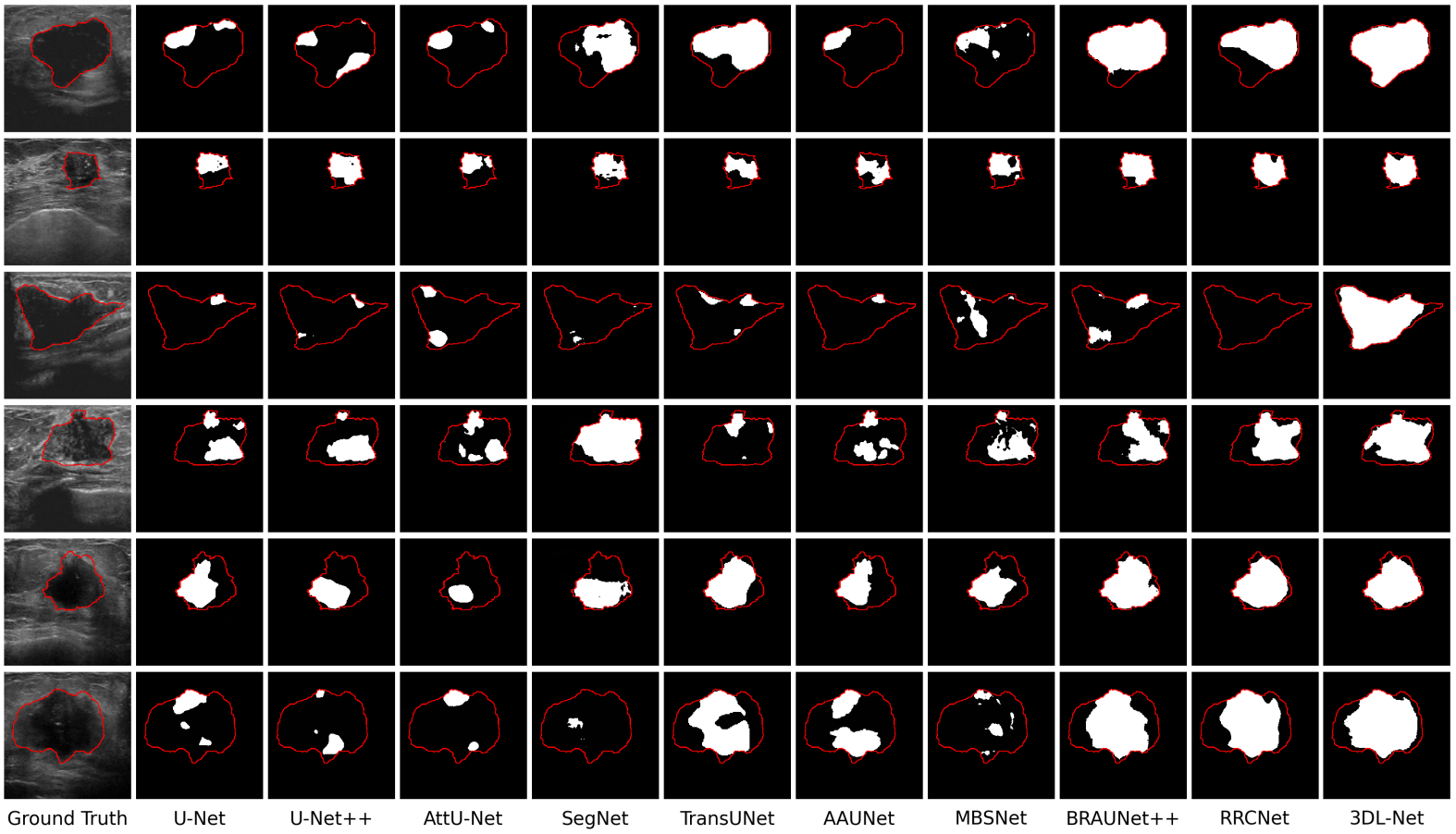


Figure 6: Visual Comparison with State-of-the-Art Methods on STU dataset. White pixels represent the predicted values, and red curves represent the ground truth values.

#### *4.3.2. Visualization analysis*

The visualization results on the BUS dataset are depicted in Figure 5. From the visualization images, it is evident that the segmented results exhibit a high degree of similarity to the original ground truth, confirming the accuracy of 3DL-Net. These visualization results highlight the superior segmentation performance of 3DL-Net compared to other methods. Similarly to the quantitative results, the visualization outcomes of each method demonstrate similarities. However, our proposed method demonstrates clear advantages, particularly when built upon the baseline of SegNet. We observe

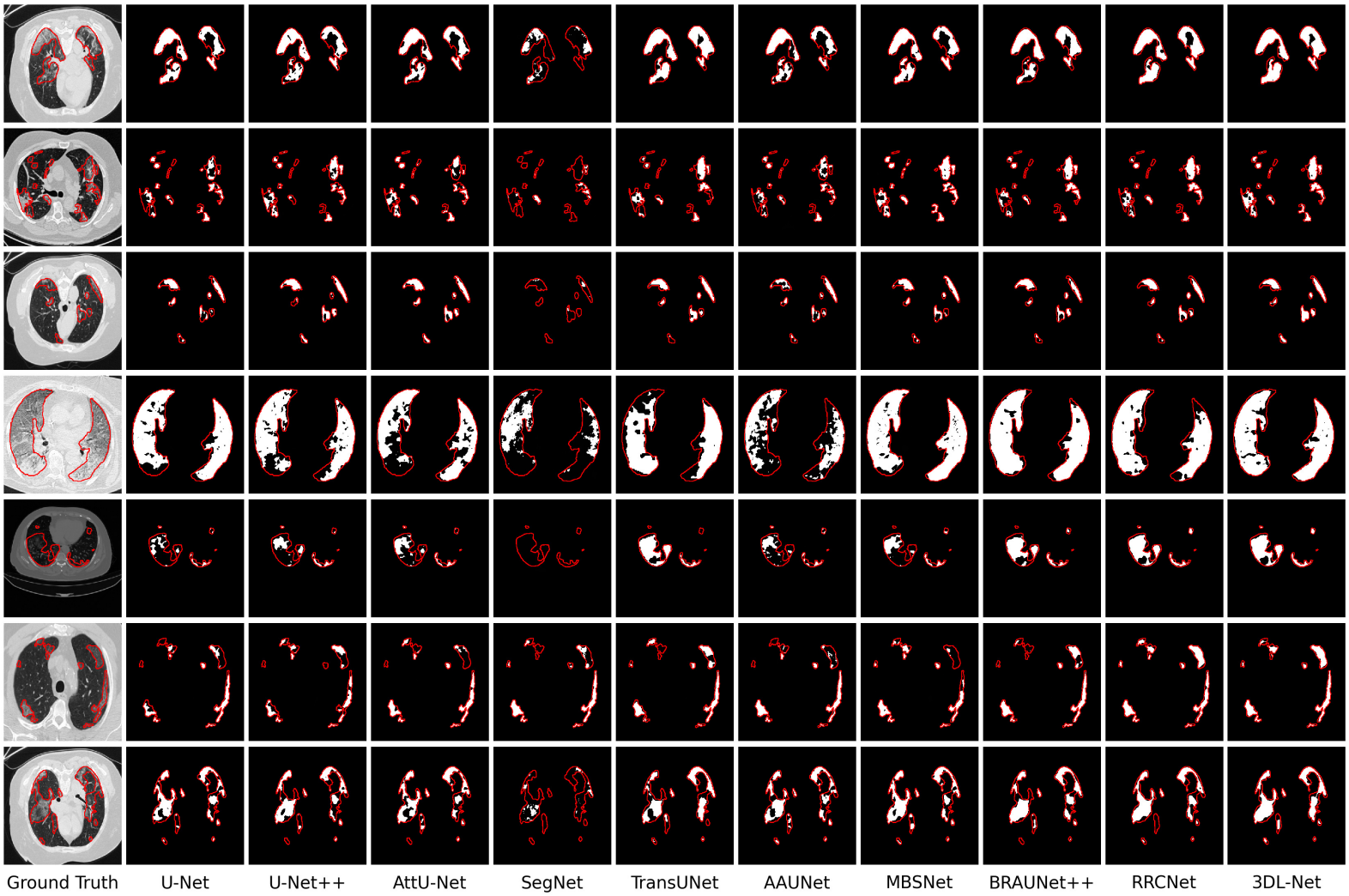


Figure 7: Visual Comparison with State-of-the-Art Methods on COVID-19 dataset. White pixels represent the predicted values, and red curves represent the ground truth values.

substantial improvements in accurately identifying numerous lesion sites, addressing the limitations of the original SegNet in refining boundaries, and facilitating enhanced delineation of local structures. Compared to U-Net and its variants, 3DL-Net, with its DMNet-enhancing global features, successfully detects many regions that would have been missed by segmentation. Additionally, compared to several state-of-the-art net-

works, 3DL-Net demonstrates significant improvements in refining boundaries. This improvement can be attributed to the fact that while these networks primarily focus on global feature extraction, they often overlook the issue of a lack of adequate representation of local features due to the inherent characteristics of dilated convolution. Our proposed DNM aims to optimize local feature processing, resulting in enhanced delineation of segmentation boundaries.

The visualization results on the STU dataset are depicted in Figure 5. Notably, STU is a small dataset, which poses a challenge due to its limited amount of tumor data, making it difficult for many networks to capture a comprehensive representation of lesion features. The proposed DMNet plays a crucial role in addressing this issue by capturing more global features through an expanded receptive field. This becomes particularly apparent when compared with traditional networks that do not effectively incorporate global features. As a result, BRAUNet++ and RRCNet achieve excellent segmentation results. BRAUNet++ utilizes a hierarchically constructed encoder-decoder architecture along with attention mechanisms to learn global semantic features. RRCNet, on the other hand, utilizes inflated convolution, which expands the receptive field of the network, and deep supervision to guide the learning process at multiple network depths. However, despite their strengths, they lack the capability to refine local features effectively. Thus, 3DL-Net achieves more accurate results in delineating the boundaries of the lesion areas, as it addresses the challenge of refining local features while capturing more global features, leading to superior segmentation boundary details.

The comprehensive and precise segmentation visualization results of lung images provide additional evidence supporting these observations. The visualization results on the COVID-19 dataset are illustrated in Figure 7. Compared to other methods, the various approaches demonstrate similar overall structures and layouts. They exhibit comparable efficacy in capturing global features. Due to the intricate nature of the focal area in COVID-19 and the presence of numerous minor lesions, the network faces challenges in accurately identifying numerous small-sized targets. However, 3DL-Net excels in delineating local anatomical structures and capturing fine details. This underscores the effectiveness of 3DL-Net in capturing and leveraging local features, which

can be attributed to the meticulous refinement facilitated by our proposed DNM module.

### *4.4. Parameter analysis*

$k$ is an important parameter in the loss function. Table 4 presents the outcomes of our parameter $k$ modulation using the BUS dataset, highlighting the optimal segmentation performance achieved with $k$ set to 0.1. While a value of 0.05 for $k$ yields excellent results in Precision, Recall is not satisfactory. Notably, the F1 score, which comprehensively balances between Recall and Precision, underscores the superiority of setting $k$ to 0.1. This specific value of $k$ strikes the best balance between the ($\mathcal{L}_{\text{BCE}}$) and $\mathcal{L}_{\text{a-Focal}}$, yielding the most favorable segmentation outcomes on BUS.

Table 4: Parameter analysis results (mean ± std) for parameter $k$ in the BUS Dataset.

| $k$ | 0.01 | 0.05 | 0.1 | 0.3 | 0.5 | 0.7 |
|---|---|---|---|---|---|---|
| Precision (%) | 87.28 ± 4.60 | **90.06 ± 2.44** | 88.36 ± 2.03 | 87.77 ± 3.08 | 85.29 ± 3.33 | 85.29 ± 3.33 |
| Recall (%) | 80.42 ± 3.87 | 78.70 ± 4.06 | **85.67 ± 2.00** | 78.81 ± 2.95 | 80.74 ± 7.37 | 80.74 ± 7.37 |
| mDice (%) | 82.53 ± 2.77 | 83.18 ± 2.65 | **87.87 ± 2.05** | 81.83 ± 1.19 | 81.62 ± 5.93 | 81.62 ± 5.93 |
| mIou (%) | 71.56 ± 3.34 | 71.82 ± 3.54 | **76.85 ± 1.04** | 70.40 ± 1.67 | 69.80 ± 8.11 | 69.80 ± 8.11 |
| Specificity (%) | 99.54 ± 0.05 | **99.60 ± 0.14** | 98.83 ± 0.12 | 99.50 ± 0.10 | 99.39 ± 0.21 | 99.39 ± 0.21 |
| Accuracy (%) | 98.50 ± 0.52 | 98.47 ± 0.24 | **98.75 ± 0.43** | 98.45 ± 0.44 | 98.47 ± 0.50 | 98.47 ± 0.50 |
| F1 (%) | 84.80 ± 3.08 | 86.47 ± 2.13 | **87.47 ± 1.18** | 84.67 ± 1.46 | 83.37 ± 4.47 | 83.37 ± 4.47 |

In addition, we conducted a detailed analysis of the parameter $M$, representing the number of dendrites in dendritic learning. As shown in Table 5 our findings suggest that the choice of $M$ influences the model's performance in medical image segmentation tasks. Specifically, the model exhibits optimal segmentation performance whlie $M$ is set to 10, indicating that configuring the network with 10 dendrites is most suitable for effectively capturing microstructures in the feature maps received by the synapses layer in the DNM. However, we also observed that excessively large or small values of $M$ lead to a decline in performance, highlighting the need to adjust the number of dendrites within a certain range to balance the model's complexity and performance.

Table 5: Parameter analysis results (mean ± std) for parameter $M$ in the BUS Dataset.

| $M$ | 1 | 5 | 10 | 15 | 20 |
|---|---|---|---|---|---|
| Precision(%) | 86.17 ± 3.10 | 88.19 ± 5.40 | **88.36 ± 2.03** | 87.05 ± 4.76 | 85.99 ± 4.70 |
| Recall(%) | 70.94 ± 7.02 | 83.44 ± 3.41 | **85.67 ± 2.00** | 79.39 ± 5.69 | 78.53 ± 4.81 |
| mDice(%) | 76.10 ± 4.56 | 84.83 ± 1.39 | **87.87 ± 2.05** | 82.16 ± 3.89 | 81.44 ± 4.50 |
| mIou(%) | 62.41 ± 5.58 | 74.32 ± 2.08 | **76.85 ± 1.04** | 70.97 ± 5.31 | 69.70 ± 6.02 |
| Specificity(%) | 99.49 ± 0.15 | **99.57 ± 0.15** | 98.83 ± 0.12 | 99.48 ± 0.29 | 99.39 ± 0.26 |
| Accuracy(%) | 98.12 ± 0.47 | 98.70 ± 0.40 | **98.75 ± 0.43** | 98.47 ± 0.53 | 98.39 ± 0.37 |
| F1(%) | 80.69 ± 1.70 | 85.59 ± 1.38 | **87.47 ± 1.18** | 84.48 ± 3.56 | 83.64 ± 4.37 |

Table 6: The ablation experiment results (mean ± std) in the BUS dataset.

| Methods | Precision (%) | Recall (%) | mDice (%) | mIou (%) | Specificity (%) | F1 (%) |
|---|---|---|---|---|---|---|
| SegNet | 79.70 ± 3.63 | 66.81 ± 6.24 | 72.25 ± 3.33 | 56.85 ± 4.19 | 99.18 ± 0.15 | 75.72 ± 2.11 |
| SegNet + DC | 87.16 ± 1.73 | 83.24 ± 5.85 | 84.48 ± 3.01 | 73.95 ± 4.07 | 99.44 ± 0.15 | 85.76 ± 1.20 |
| SegNet + DP | 87.72 ± 2.40 | 83.67 ± 4.66 | 85.10 ± 3.56 | 74.69 ± 5.03 | 99.47 ± 0.26 | 86.38 ± 2.86 |
| SegNet + DC + DNM | 84.28 ± 3.17 | 68.28 ± 4.84 | 73.65 ± 3.19 | 59.72 ± 4.36 | 99.51 ± 0.16 | 78.59 ± 2.99 |
| SegNet + DP + DNM | 88.05 ± 3.55 | 80.64 ± 7.05 | 83.12 ± 2.87 | 71.94 ± 3.51 | 99.55 ± 0.16 | 85.42 ± 0.35 |
| SegNet + DNM + DC | 85.23 ± 9.01 | 69.11 ± 4.53 | 75.27 ± 6.02 | 61.94 ± 6.63 | **99.56 ± 0.18** | 79.89 ± 7.11 |
| SegNet + DNM + DP | 85.15 ± 3.07 | 77.67 ± 9.65 | 79.73 ± 7.46 | 68.04 ± 9.47 | 99.38 ± 0.23 | 82.27 ± 5.10 |
| SegNet + DNM + DC + DNM | 87.39 ± 4.26 | 83.30 ± 4.75 | 84.75 ± 3.32 | 74.37 ± 5.01 | 99.43 ± 0.16 | 86.02 ± 3.26 |
| SegNet + DNM + DM + DNM (3DL-Net) | **88.36 ± 2.03** | **85.67 ± 2.00** | **86.61 ± 0.70** | **76.85 ± 1.04** | 99.52 ± 0.10 | **87.47 ± 1.18** |

### *4.5. Ablation experiments*

To further validate the advancements of 3DL-Net, we conducted ablation studies on the BUS dataset. The experimental setup remained consistent with previous experiments, except for variations in the modules. Our objective is to validate the effectiveness of different modules within 3DL-Net, considering diverse network structures, with or without DNM, and the ablation results of the global feature and local feature modules. In addition, we discuss the comparative results of DMNet using cascade and pyramid structures. As depicted in Table 6, DM represents the method using a pyramid structure, while DC represents the method using a cascade approach. In the cascade

approach, we directly employed a cascade of multilayer dilation convolution, adjusting the parameters based on the study of [85, 37]. We cascade the outputs of multiple dilated convolutional layers, where each output directly connects to the input of the next layer, forming a continuous chain structure. This cascade aids in enhancing global feature representation by capturing hierarchical features and facilitating the progressive transfer and processing of information. Table 6 demonstrates that the introduction of DMNet in the baseline SegNet model resulted in significant improvements. On the foundation of DSNet, both cascade and pyramid structures exhibited performance improvement, with the best results achieved using DMNet with a pyramid structure. It is due to the ability of the proposed DMNet to capture features at different scales to better understand global information and local details. This enables the network to understand the image content more comprehensively. Notably, the performance improvement is limited when adding the DNM module alone. The optimal results are achieved when both DNM and DMNet are integrated. This is attributed to DMNet enhancing global feature representation, while DNM addresses the challenge of insufficient consideration of local features introduced by dilated convolution. A series of ablation experiments provide compelling evidence that 3DL-Net not only outperforms other methods but also benefits from the auxiliary branch, which includes DMNet, dilated neural networks, and dendritic modules, applied to 3DL-Net for segmenting breast tumors and pneumonia images. The results of thorough experiments and comparisons in this section conclusively establish 3DL-Net as the superior model.

## 5. Conclusion

We propose a medical image segmentation network that enhances global-local feature representation and refinement, leading to comprehensive integration of multi-scale contextual information. Among them, dendritic learning plays a pivotal role. 3DL-Net leverages a new generation of interpretable neurons to enhance the processing of local features and improve segmentation effectiveness. Unlike previous studies, we pioneer the application of dendritic learning at the channel level. In addition, it enhances the perceived range of the features by using a dilated convolution with a pyramid structure

to obtain global and local feature information at different levels. We evaluated two ultrasound datasets and one CT dataset. Due to our stability and precision results, we expect to demonstrate promising outcomes in additional medical segmentation tasks. 3DL-Net could potentially serve as a dependable computer-aided screening system for breast tumors and COVID-19 pneumonia, aiding physicians in lesion localization and early diagnosis.

**Acknowledgment**

This research was partially supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant JP22H03643, and Japan Science and Technology Agency (JST) Support for Pioneering Research Initiated by the Next Generation (SPRING) under Grant JPMJSP2145.

**References**

[1] J. E. Iglesias, M. R. Sabuncu, Multi-atlas segmentation of biomedical images: A survey, Medical Image Analysis 24 (1) (2015) 205–219.

[2] J. Duncan, N. Ayache, Medical image analysis: Progress over two decades and the challenges ahead, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (1) (2000) 85–106.

[3] D. Shen, G. Wu, H.-I. Suk, Deep learning in medical image analysis, Annual Review of Biomedical Engineering 19 (1) (2017) 221–248.

[4] N. Salpea, P. Tzouveli, D. Kollias, Medical image segmentation: A review of modern architectures, in: European Conference on Computer Vision, 2022, pp. 691–708.

[5] M. Antonelli, A. Reinke, S. Bakas, K. Farahani, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze, O. Ronneberger, R. M. Summers, et al., The medical segmentation decathlon, Nature Communications 13 (1) (2022) 4128.

[6] Y. Zhang, H. Qu, Y. Wang, Adaptive image segmentation based on fast thresholding and image merging, in: 16th International Conference on Artificial Reality and Telexistence–Workshops (ICAT'06), 2006, pp. 308–311.

[7] N. Senthilkumaran, S. Vaithegi, Image segmentation by using thresholding techniques for medical images, Computer Science & Engineering: An International Journal 6 (1) (2016) 1–13.

[8] V. Chalana, Y. Kim, A methodology for evaluation of boundary detection algorithms on medical images, IEEE Transactions on Medical Imaging 16 (5) (1997) 642–652.

[9] J. Mehena, Medical images edge detection based on mathematical morphology, Journal of Computer and Communication Technology 4 (1) (2013) 2.

[10] R. K. Justice, E. M. Stokely, J. S. Strobel, R. E. Ideker, W. M. Smith, Medical image segmentation using 3D seeded region growing, in: Medical Imaging 1997: Image Processing, Vol. 3034, 1997, pp. 900–910.

[11] B. N. Li, C. K. Chui, S. Chang, S. H. Ong, A new unified level set method for semi-automatic liver tumor segmentation on contrast-enhanced ct images, Expert Systems with Applications 39 (10) (2012) 9661–9668.

[12] Y. Boykov, G. Funka-Lea, Graph cuts and efficient N-D image segmentation, International Journal of Computer Vision 70 (2) (2006) 109–131.

[13] D. S. Manoharan, A. Sathesh, Improved version of graph-cut algorithm for CT images of lung cancer With clinical property condition, Journal of Artificial Intelligence and Capsule Networks 2 (4) (2020) 201–206.

[14] R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, A. K. Nandi, Medical image segmentation using deep learning: A survey, IET Image Processing 16 (5) (2022) 1243–1267.

[15] D. L. Pham, C. Xu, J. L. Prince, Current methods in medical image segmentation, Annual Review of Biomedical Engineering 2 (1) (2000) 315–337.

[16] M. H. Hesamian, W. Jia, X. He, P. Kennedy, Deep learning techniques for medical image segmentation: Achievements and challenges, Journal of Digital Imaging 32 (2019) 582–596.

[17] I. Scholl, T. Aach, T. M. Deserno, T. Kuhlen, Challenges of medical image processing, Computer Science-Research and Development 26 (2011) 5–13.

[18] B. McCrindle, K. Zukotynski, T. E. Doyle, M. D. Noseworthy, A Radiology-focused review of predictive uncertainty for AI interpretability in computer-assisted segmentation, Radiology: Artificial Intelligence 3 (6) (2021) e210031.

[19] Y. Xie, J. Zhang, C. Shen, Y. Xia, CoTr: Efficiently bridging CNN and Transformer for 3D medical image segmentation, in: Medical Image Computing and Computer Assisted Intervention – MICCAI 2021, 2021, pp. 171–180.

[20] S. Bao, A. C. Chung, Multi-scale structured CNN with label consistency for brain MR image segmentation, Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization 6 (1) (2018) 113–117.

[21] F. Lateef, Y. Ruichek, Survey on semantic segmentation using deep learning techniques, Neurocomputing 338 (2019) 321–348.

[22] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3431–3440.

[23] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, 2015, pp. 234–241.

[24] V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (12) (2017) 2481–2495.

[25] N. Japkowicz, S. Stephen, The class imbalance problem: A systematic study, Intelligent Data Analysis 6 (5) (2002) 429–449.

[26] M. Yeung, E. Sala, C.-B. Schönlieb, L. Rundo, Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation, Computerized Medical Imaging and Graphics 95 (2022) 102026.

[27] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, S. Savarese, Generalized intersection over union: A metric and a loss for bounding box regression, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 658–666.

[28] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, IEEE Transactions on Pattern Analysis and Machine Intelligence 42 (2) (2020) 318–327.

[29] R. Zhao, B. Qian, X. Zhang, Y. Li, R. Wei, Y. Liu, Y. Pan, Rethinking dice loss for medical image segmentation, in: 2020 IEEE International Conference on Data Mining (ICDM), 2020, pp. 851–860.

[30] F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolutions, in: International Conference on Learning Representations (ICLR), 2015.

[31] J. Lv, Y. Hu, Q. Fu, Y. Hu, L. Lv, G. Yang, J. Li, Y. Zhao, Local feature matters: Cascade multi-scale MLP for edge segmentation of medical images, IEEE Transactions on NanoBioscience 22 (4) (2023) 828–835.

[32] A. Srivastava, D. Jha, S. Chanda, U. Pal, H. D. Johansen, D. Johansen, M. A. Riegler, S. Ali, P. Halvorsen, MSRF-Net: A multi-scale residual fusion network for biomedical image segmentation, IEEE Journal of Biomedical and Health Informatics 26 (5) (2022) 2252–2263.

[33] Y. Li, Y. Zhang, J.-Y. Liu, K. Wang, K. Zhang, G.-S. Zhang, X.-F. Liao, G. Yang, Global Transformer and dual local attention network via deep- hierarchical feature fusion for retinal vessel segmentation, IEEE Transactions on Cybernetics 53 (9) (2023) 5826–5839.

[34] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, P. H. Torr, Conditional random fields as recurrent neural networks, in: Proceed-

ings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2015, pp. 1529–1537.

[35] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2881–2890.

[36] G. Chen, L. Li, Y. Dai, J. Zhang, M. H. Yap, AAU-Net: An adaptive attention U-Net for breast lesions segmentation in ultrasound images, IEEE Transactions on Medical Imaging 42 (5) (2023) 1289–1300.

[37] G. Chen, Y. Dai, J. Zhang, RRCNet: Refinement residual convolutional network for breast ultrasound images segmentation, Engineering Applications of Artificial Intelligence 117 (2023) 105601.

[38] J. Ouyang, S. Liu, H. Peng, H. Garg, D. N. Thanh, LEA U-Net: A U-Net-based deep learning framework with local feature enhancement and attention for retinal vessel segmentation, Complex & Intelligent Systems 9 (6) (2023) 6753–6766.

[39] V. Cheplygina, Cats or cat scans: Transfer learning from natural or medical image source data sets?, Current Opinion in Biomedical Engineering 9 (2019) 21–27.

[40] D. Karimi, S. K. Warfield, A. Gholipour, Transfer learning in medical image segmentation: New insights from analysis of the dynamics of model parameters and learned representations, Artificial Intelligence in Medicine 116 (2021) 102078.

[41] Z. Li, Y. Zheng, D. Shan, S. Yang, Q. Li, B. Wang, Y. Zhang, Q. Hong, D. Shen, ScribFormer: Transformer makes CNN work better for scribble-based medical image segmentation, IEEE Transactions on Medical Imaging (2024) 1–1.

[42] W. S. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity, The Bulletin of Mathematical Biophysics 5 (1943) 115–133.

[43] S. Gao, M. Zhou, Y. Wang, J. Cheng, H. Yachi, J. Wang, Dendritic neuron model with effective learning algorithms for classification, approximation, and prediction, IEEE Transactions on Neural Networks and Learning Systems 30 (2) (2019) 601–614.

[44] T. Zhou, S. Gao, J. Wang, C. Chu, Y. Todo, Z. Tang, Financial time series prediction using a dendritic neuron model, Knowledge-Based Systems 105 (2016) 214–224.

[45] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in Neural Information Processing Systems, Vol. 30, 2017.

[46] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, Y. Zhou, TransUNet: Transformers make strong encoders for medical image segmentation, arXiv preprint arXiv:2102.04306 (2021).

[47] R. Gu, G. Wang, T. Song, R. Huang, M. Aertsen, J. Deprest, S. Ourselin, T. Vercauteren, S. Zhang, CA-Net: Comprehensive attention convolutional neural networks for explainable medical image segmentation, IEEE Transactions on Medical Imaging 40 (2) (2020) 699–711.

[48] W. Wang, C. Chen, M. Ding, H. Yu, S. Zha, J. Li, Transbts: Multimodal brain tumor segmentation using transformer, in: Medical Image Computing and Computer Assisted Intervention – MICCAI 2021, 2021, pp. 109–119.

[49] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., An image is worth 16x16 words: Transformers for image recognition at scale, in: International Conference on Learning Representations (ICLR), 2021.

[50] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, M. Wang, Swin-Unet: Unet-like pure Transformer for medical image segmentation, in: European Conference on Computer Vision, 2022, pp. 205–218.

[51] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin Transformer: Hierarchical vision Transformer using shifted windows, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 10012–10022.

[52] L. Lan, P. Cai, L. Jiang, X. Liu, Y. Li, Y. Zhang, BRAU-Net++: U-Shaped hybrid CNN-Transformer network for medical image segmentation, arXiv preprint arXiv:2401.00722 (2024).

[53] W. Liu, A. Rabinovich, A. C. Berg, ParseNet: Looking wider to see b etter, in: International Conference on Learning Representations (ICLR), 2016.

[54] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille, DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs, IEEE Transactions on Pattern Analysis and Machine Intelligence 40 (4) (2017) 834–848.

[55] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 801–818.

[56] R. Hamaguchi, A. Fujita, K. Nemoto, T. Imaizumi, S. Hikosaka, Effective use of dilated convolutions for segmenting small object instances in remote sensing imagery, in: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), 2018, pp. 1442–1450.

[57] E. Liu, S. Li, S. Liu, Color enhancement using global parameters and local features learning, in: Proceedings of the Asian Conference on Computer Vision (ACCV), 2020.

[58] B. A. Richards, T. P. Lillicrap, P. Beaudoin, Y. Bengio, R. Bogacz, A. Christensen, C. Clopath, R. P. Costa, A. de Berker, S. Ganguli, et al., A deep learning framework for neuroscience, Nature Neuroscience 22 (11) (2019) 1761–1770.

[59] M. E. Larkum, Are dendrites conceptually useful?, Neuroscience 489 (2022) 4–14, Dendritic contributions to biological and artificial computations.

[60] Y. Tang, Z. Song, Y. Zhu, M. Hou, C. Tang, J. Ji, Adopting a dendritic neural model for predicting stock price index movement, Expert Systems with Applications 205 (2022) 117637.

[61] J. Ji, C. Tang, J. Zhao, Z. Tang, Y. Todo, A survey on dendritic neuron model: Mechanisms, algorithms and practical applications, Neurocomputing 489 (2022) 390–406.

[62] X. Li, J. Tang, Q. Zhang, B. Gao, J. J. Yang, S. Song, W. Wu, W. Zhang, P. Yao, N. Deng, et al., Power-efficient neural network with artificial dendrites, Nature Nanotechnology 15 (9) (2020) 776–782.

[63] E. Egrioglu, E. Baş, M.-Y. Chen, Recurrent dendritic neuron model artificial neural network for time series forecasting, Information Sciences 607 (2022) 572–584.

[64] S. Gao, M. Zhou, Z. Wang, D. Sugiyama, J. Cheng, J. Wang, Y. Todo, Fully complex-valued dendritic neuron model, IEEE Transactions on Neural Networks and Learning Systems 34 (4) (2023) 2105–2118.

[65] Z. Zhang, Z. Lei, M. Omura, H. Hasegawa, S. Gao, Dendritic learning-incorporated vision Transformer for image recognition, IEEE/CAA Journal of Automatica Sinica 11 (2) (2024) 539–541.

[66] J. Li, Z. Liu, R.-L. Wang, S. Gao, Dendritic deep residual learning for covid-19 prediction, IEEJ Transactions on Electrical and Electronic Engineering 18 (2) (2023) 297–299.

[67] G. Liu, J. Wang, Dendrite net: A white-box module for classification, regression, and system identification, IEEE Transactions on Cybernetics 52 (12) (2021) 13774–13787.

[68] E. Bas, E. Egrioglu, T. Cansu, Robust training of median dendritic artificial neural networks for time series forecasting, Expert Systems with Applications 238 (2024) 122080.

[69] C. Koch, T. Poggio, V. Torre, Retinal ganglion cells: A functional interpretation of dendritic morphology, Philosophical Transactions of the Royal Society of London. B, Biological Sciences 298 (1090) (1982) 227–263.

[70] Y. Zhang, P. Cai, Y. Sun, Z. Zhang, Z. Lei, S. Gao, A Lightweight multi-dendritic pyramidal neuron model with neural plasticity on image recognition, IEEE Transactions on Artificial Intelligence (2024) 1–13doi:10.1109/TAI.2024.3379968.

[71] M. Aljabri, M. AlGhamdi, A review on the use of deep learning for medical images segmentation, Neurocomputing 506 (2022) 311–335.

[72] Z. Fu, J. Li, Z. Hua, L. Fan, Deep supervision feature refinement attention network for medical image segmentation, Engineering Applications of Artificial Intelligence 125 (2023) 106666.

[73] X. Ning, W. Tian, Z. Yu, W. Li, X. Bai, Y. Wang, HCFNN: High-order coverage function neural network for image classification, Pattern Recognition 131 (2022) 108873.

[74] M. H. Yap, G. Pons, J. Marti, S. Ganau, M. Sentis, R. Zwiggelaar, A. K. Davison, R. Marti, Automated breast ultrasound lesions detection using convolutional neural networks, IEEE Journal of Biomedical and Health Informatics 22 (4) (2017) 1218–1226.

[75] Z. Zhuang, N. Li, A. N. Joseph Raj, V. G. Mahesh, S. Qiu, An RDAU-NET model for lesion segmentation in breast ultrasound images, PloS one 14 (8) (2019) e0221535.

[76] M. Maftouni, A. C. C. Law, B. Shen, Z. J. K. Grado, Y. Zhou, N. A. Yazdi, A robust ensemble-deep learning model for COVID-19 diagnosis based on an integrated CT scan images database, in: IIE Annual Conference. Proceedings, 2021, pp. 632–637.

[77] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.

[78] F. Hoorali, H. Khosravi, B. Moradi, Irunet for medical image segmentation, Expert Systems with Applications 191 (2022) 116399.

[79] G. Chen, Y. Liu, J. Qian, J. Zhang, X. Yin, L. Cui, Y. Dai, DSEU-net: A novel deep supervision SEU-net for medical ultrasound image segmentation, Expert Systems with Applications 223 (2023) 119939.

[80] A. Iqbal, M. Sharif, Bts-st: Swin transformer network for segmentation and classification of multimodality breast cancer images, Knowledge-Based Systems 267 (2023) 110393.

[81] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, 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, 2018, pp. 3–11.

[82] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, D. Rueckert, Attention U-Net: Learning where to look for the pancreas, in: Medical Imaging with Deep Learning, 2018.

[83] S. Jin, S. Yu, J. Peng, H. Wang, Y. Zhao, A novel medical image segmentation approach by using multi-branch segmentation network based on local and global information synchronous learning, Scientific Reports 13 (1) (2023) 6762.

[84] F. Milletari, N. Navab, S.-A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: 2016 fourth international conference on 3D vision (3DV), 2016, pp. 565–571.

[85] P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, G. Cottrell, Understanding Convolution for Semantic Segmentation, in: 2018 IEEE winter conference on applications of computer vision (WACV), 2018, pp. 1451–1460.