Computing and Informatics, Vol. 41, 2022, 1383–1400, doi: 10.31577/cai_2022_5_1383

RGN-NET: A GLOBAL CONTEXTUAL AND MULTISCALE INFORMATION ASSOCIATION NETWORK FOR MEDICAL IMAGE SEGMENTATION

Zhixin Zhang, Shuhao Jiang, Xuhua Pan*

Information Engineering Department
Tianjin University of Commerce
Tianjin, 300134, China
e-mail: {Zhangzhixin, Jiangshuhao, Panxuhua}@tjcu.edu.cn

Abstract. Segmentation of medical images is a necessity for the development of healthcare systems, particularly for illness diagnosis and treatment planning. Recently, convolutional neural networks (CNNs) have gained amazing success in automatically segmenting medical images to identify organs or lesions. However, the majority of these approaches are incapable of segmenting objects of varying sizes and training on tiny, skewed datasets, both of which are typical in biomedical applications. Existing solutions use multi-scale fusion strategies to handle the difficulties posed by varying sizes, but they often employ complicated models more suited to broad semantic segmentation computer vision issues. In this research, we present an end-to-end dual-branch split architecture RGN-Net that takes the benefits of the two networks into greater account. Our technique may successfully create long-term functional relationships and collect global context data. Experiments on Lung, MoNuSeg, and DRIVE reveal that our technique reaches state-of-the-art benchmarks in order to evaluate the performance of RGN-Net.

Keywords: Image segmentation, medical image processing, attention mechanism, deep learning, global context extract

Mathematics Subject Classification 2010: 68T20

* Corresponding author

1 INTRODUCTION

In identifying and treating ocular fundus disorders, cardiovascular diseases, gastrointestinal diseases, etc., medical image analysis plays a crucial role. Concurrently, the identification of the focal lesion and the accompanying clinical evaluations are always performed by seasoned topic specialists. However, certain early indications are difficult to detect using the aforementioned conventional methods. Automatic Medical Image Segmentation (MIS), an interdisciplinary recognition scenario derived from widely used artificial intelligence methodologies [1, 2, 3, 4, 5], can optimize patient management and filter potential subjective factors in clinical decision-making. It can assist medical professionals in making more accurate and efficient diagnoses of the corresponding disorders.

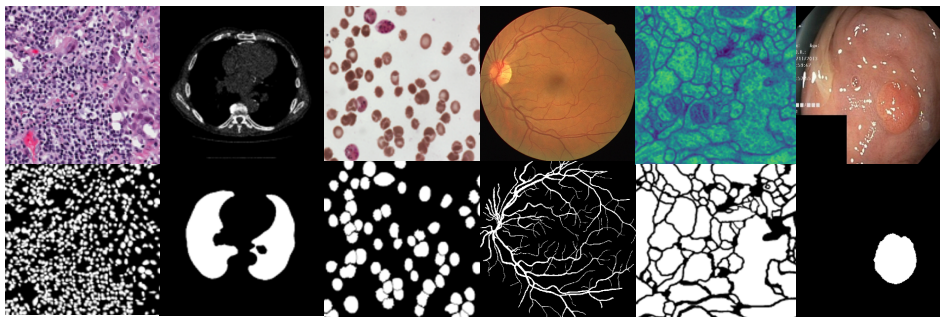

Figure 1. The results of segmenting medical images. The binary pictures in the second row correlate to the original medical image from left to right in the first row.

The objective of image segmentation is to distinguish the target region of an image from the background; it is a typically unstructured problem due to the differences in medical imaging principles and the properties of the tissues themselves, as illustrated by the diverse medical image segmentation in Figure 1. In addition, image creation, including CT and magnetic resonance imaging (MRI) [6], is influenced by a number of additional variables, including noise, tissue mobility, and individual variances.

In comparison to natural pictures, medical images have poor resolution, low contrast, and scattered targets, but they are very accurate and stable. Early techniques were mostly centered on edge detection and template matching due to these properties. Specifically, the circular or elliptical Hough transform was employed to segment the optic disc, whereas template matching was utilized to segment the spleen in MRI and the ventricles in brain CT.

In this decade, the rapid development of hardware computing power has enabled deep learning techniques [7, 8] to make significant advances in computer vision. For instance, convolutional neural networks (CNN) have demonstrated a high ability to learn discriminative visual representations with high capability and can be applied to a variety of computer vision tasks, such as target tracking. In this study,

an end-to-end network is selected as the design of the MIS encoder-whole decoder, and our approach can learn global contextual information and multi-scale feature information more precisely without significantly increasing the parameters, which can be essential for semantic segmentation tasks. The following is a summary of the key contributions made by this work:

1. This work presents a dual attention module to improve the capture of the long-distance dependency of spatial direction and precise position data.
2. Our technique incorporates the hole convolution with variable expansion factors for multi-scale information fusion into the encoder-decoder network in order to extract optimal multi-scale information in order to increase segmentation accuracy.
3. Our method incorporates an attention gating technique to increase the integration level of the feature map derived by upsampling and downsampling.
4. Extensive trials, including segmentation of retinal vessels, segmentation of pulmonary regions, segmentation of cell configurations, segmentation of cancerous skin, etc., have confirmed the suggested method. The findings demonstrate that the suggested strategy outperforms state-of-the-art techniques for certain tasks.

2 RELATED WORK

In this part, we outline the most significant advancements in medical picture segmentation, highlight advancements in feedback attention networks, and highlight recent contributions to image segmentation approaches based on Transformer architecture.

2.1 Medical Image Segmentation with Convolutional Neural Network

Most of the existing CNN semantic segmentation architecture is based on a fully convolutional network (FCN) [9], or encoder-decoder architecture, such as [10]. In the task of image classification, the fully connected layer at the end will compress the two-dimensional matrix information in the original image, which will result in the loss of spatial information of the image. It will have a great impact on the use of convolution for image segmentation.

The advent of FCN created a precedent for convolutional neural networks for image segmentation. The basic idea is to replace the fully connected layer in the traditional convolutional neural network model with a convolutional layer and then use the deconvolution operation to upsample the final output feature map and skip connections to improve the rough upsampling pixel positioning. For example, Ben-Cohen et al. [11] first explored the use of FCN to complete the segmentation task of liver and tumor in CT images. Drozdzal et al. [12] proposed very deep FCN by using short skip connections. The authors showed that a very deep FCN with long and short skip connections achieved a better result than the original one. Compared with CNN based on fixed-size input, FCN can accept input of any size and generate

sums through effective reasoning and learning. The corresponding size output of the original image. There are also many previous works that have proposed a series of SegNet [13], DeepLab [14], DANet [15], etc.

U-Net [16] is the most famous network architecture in the field of medical image segmentation. It is an FCN-based segmentation network proposed by Ronneberger and others in the ISBI challenge [17]. After that, a series of deformation structures based on U-Net were extended on this basis. For example, the dense connection operation (UNet++) [18] is added to the U-Net network, UNet++ adds more jump connection paths and up-sampling convolution blocks to compensate for the encoder and linguistic gap between decoders. Other work integrates the residual idea into the U-Net network. In 2019, Ibtehaz et al. [19] used the MultiResUNet network proposed by Resnet for reference, which used the residual idea to transform the convolutional block and jump connection in U-Net. Mehta et al. [20] proposed the Y-Net network structure in 2018, adding a classification task of breast cancer images to the task of segmentation of breast cancer biopsy images. Based on U-Net, Y-Net introduces the residual connection of the residual network to help improve the segmentation results, and at the same time, adds a second branch for the classification of breast cancer pictures on this basis, which is a multi-task learning algorithm.

The U-Net technique of recurrent neural networks is used in another portion of the study. Alom et al. [21] proposed the R2U-Net network design in 2018, which incorporates the U-Net framework. BCDU-Net was suggested in 2019 by Azad et al. [22]. This is an alternative method for integrating recurrent neural networks into the U-Net network. Adding LSTM to CNN mostly addresses the gradient disappearance and gradient explosion issues that arise during extended sequence training. For medical picture segmentation, Alom et al. [23] developed Recurrent Convolutional Neural Network (RCNN) and Recurrent Residual Convolutional Neural Network (R2CNN) based on U-Net models. CNN and LSTM are combined to create ConvLSTM [24]. BCDU-Net implements two-way ConvLSTM in the jump connection and integrates the appropriate feature maps of the encoding and decoding phases in a nonlinear fashion to give more accurate segmentation results.

2.2 Use Transformer for Medical Image Segmentation

The transformer was first developed for use in machine translation jobs and has now evolved to advanced levels in several NLP applications. In order to adapt the concept of the transformer to image tasks and computer vision problems, numerous recent research have made significant modifications to transformer design. Child et al. [25], for instance, suggested a sparse transformer (Sparse Transformer) that employs a scalable method to global self-attention. In response to the success of Transformer, the current Vision Transformer (ViT) [26] established the most sophisticated classification on ImageNet by immediately applying the transformer to full-size pictures with global self-attention. In contrast to the CNN-based technique, ViT must be pre-trained on its own massive dataset. Some networks, such as TransUNet [27] and Medical Transformer [28], are built on the enhance-

ment of ViT in an effort to lessen the difficulty of training ViT. As a vision backbone, [29] proposes a Swin Transformer, an efficient and effective hierarchical vision Transformer. Swin Transformer demonstrated state-of-the-art performance on a variety of vision tasks, including image classification, object identification, and semantic segmentation, using the shifted windows technique. In addition, some work, such as (UNETR: Transformers for 3D Medical Image Segmentation) [24], segments 3D medical image data using the transformer design and achieves high performance.

In this work, we considered the poor performance of the CNN network in learning long-term spatial dependence, which may seriously affect the segmentation performance of challenging tasks. The advantage of the transformer is the ability to model long-term dependencies. Therefore, we have combined the advantages and disadvantages of the two networks.

3 METHODOLOGY

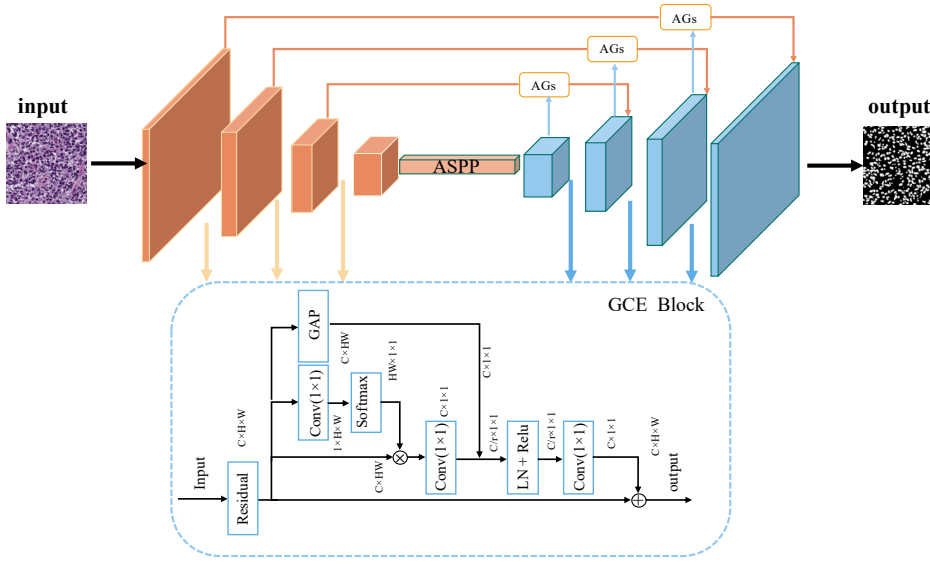

Figure 2. This schematic is about the framework of RGN-Net. It consists of an encoder-decoder network, multi-scale fusion, and a multi-hybrid attention mechanism. LayerNorm (LN) in the middle rounded rectangle denotes the normalization of the channel direction by calculating the global average value.

This section illustrates the MIS technique recommended in this article. This analogues of our algorithm include the encoder-decoder network, multiscale fusion, and multiple hybrid attention mechanisms. Figure 2 displays the specifics of framework.

3.1 Encoding and Decoding Module

Our RGN-Net is shown in Figure 2, in which we employ a ResNet network with pre-training and integrate a global context extraction network with an attention gate mechanism network for joint extraction of feature vectors with higher semantic representations. In Figure 2, red and blue tensor blocks, respectively, indicate the encoding and decoding routes. This has the benefit of plainly illustrating the forward propagation flow of our algorithm. Our encoder and decoder techniques are as follows:

1. We quantify an input medical image as $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_i, \ldots \mathbf{x}_N\}$, where N denotes the pixel number, and $\mathbf{x}_i$ has m-dimensional features. In Figure 2, input feature map of RGN-Net can be represent as $\mathbf{F}$, where $\mathbf{F} \in \mathbb{R}^{C \times H \times W}$; $\otimes$ is the operator of the tensor product; C, H, and W represent the number of channels, height, and width of the tensor, respectively.

2. In the encoder, there are four downsampling layers, with each layer consisting of two 3×3 convolutions, a Rectified Linear Unit (ReLU) function, and a 2×2 max-pooling operation. Each encoder step can yield doubled feature maps. After four iterations of downsampling, a high-dimensional feature extraction map may be obtained and fed into a multi-scale fusion module.

3. Regarding the decoder, our objective is to recover the high-level semantic utilizing the extracted features from the feature encoder and the multi-scale fusion module. A hops-connected architecture is utilized to collect comprehensive information from the encoder in order to compensate for information loss caused by continuous pooling and strung convolution operations. Two procedures comprise the decoder procedure: upscaling and deconvolution. The former expands the size of the picture using linear interpolation, whilst the latter fills the leftover positions with 0 before performing a convolution process.

3.2 Multi-Scale Feature Fusion

Using diverse convolution kernels, we may extract and combine information from various scales to provide a superior representation. We also believe that a broader receptive field can improve the detection and segmentation of tiny objects. Therefore, hole convolution is performed to generate more dense data, whilst Atrus Spatial Pyramid Polling (ASPP) is utilized for multi-scale information fusion [14]. Additionally, the varied image sizes provide an obstacle. Consequently, we execute multi-branch pooling in four sizes. The resultant fused feature map is given into the subsequent network layer.

3.3 Attention Mechanism of Segmentation Network

Coordinate attention, in contrast to channel attention, which turns the feature tensor into a vector via 2-dimensional global pooling, decomposes the channel attention into two 1-dimensional feature encoding processes that aggregate information along two spatial directions. Consequently, distant dependencies may be collected in one spatial direction, while precise position information can be maintained in another spatial direction. The produced feature maps are then encoded into a pair of orientation-sensitive and position-sensitive attention maps that may be deployed in conjunction with the input feature maps to improve the representation of intriguing items.

3.3.1 Self-Attention Mechanism for Context Extraction

To describe distant dependencies using the Global Context (GC) of the Non-Local Network (NLNet) structure, cutting-edge methodologies have implemented a self-attention mechanism. Nonetheless, NLNet is plagued by a high computational load. Significantly, the GC of the NLNet is almost same across positions, indicating that it is unnecessary to learn the GC by taking position dependence into account.

Informed on the work of Cao et al. [30], our technique captures all of the image's pixels. The structure of a Global Context Extract (GCE) block is shown in Figure 2, which is expressed as (1):

$$Z_i = x_i + \mathbf{W}_{v2}\mathrm{ReLU}\left(LN\left(\mathbf{W}_{v1}\sum_{j=1}^{n}\frac{e^{\mathbf{W}_{v3}x_j}}{\sum_{m=1}^{n}e^{\mathbf{W}_{v3}x_m}}x_j\right)\right) \tag{1}$$

where x_i represents each element in the feature map of one instance of input data $\mathbf{x}$, $1 \leq i \leq n$, $n = W \times H$, $\mathbf{x} \in \mathbb{R}^n$; $\frac{e^{\mathbf{W}_{v3}x_j}}{\sum_{m=1}^{n}e^{\mathbf{W}_{v3}x_m}}$ represents the weight for global attention pooling; $\mathbf{W}_v1$, $\mathbf{W}_{v2}$ and $\mathbf{W}_{v3}$ indicate the linear transformation matrices.

3.3.2 Attention Gates in RGN-Net Model

To acquire a large segmentation domain, a downsampling method that improves the feature graph might be considered. First, the coarse-grained position of the target item is determined, while its global connection is represented. In general, we are able to generate a global eigenvector to give AGs with information to delete irrelevant content [31].

3.4 Loss Function

In order to complete the end-to-end deep learning framework RGN-Net, which is represented in Figure 2, we must train the model pixel-by-pixel to identify whether a pixel belongs to the foreground or background in the classification job. Cross-entropy, the most common loss function, is unsuitable for MIS applications because

the objects in medical images, such as retinal arteries and erythrocytes, usually occupy a restricted amount of space. In this work, the Dice coefficient loss function is substituted for the conventional cross entropy loss for quantifying segmentation performance when ground truth is available. Suppose k is the class label, where $k = \{1, 2, \ldots, K\}$, $K \in \mathbb{N}^+$. The ground truth labels vector and the predicted probabilities vector can be represented as $\mathbf{Y} = \{y_1(k), y_2(k), \ldots, y_i(k), \ldots, y_N(k)\}$, $\hat{y}_i(k) \in [0, 1]$ and $\hat{\mathbf{Y}} = \{\hat{y}_1(k), \hat{y}_2(k), \ldots, \hat{y}_i(k), \ldots, \hat{y}_N(k)\}$, $y_i(k) \in \{0, 1\}$, respectively. The formula of Dice loss function is shown as Equation (2).

$$L_{Dice} = 1 - \sum_{k=1}^{K} \frac{2\omega_k \sum_i^N \hat{y}_i(k) y_i(k)}{\sum_i^N \hat{y}_i^2(k) + \sum_i^N y_i^2(k)} \tag{2}$$

where N is the number of pixels. K and ω_k represent, respectively, the class number and class weight. According to the MIS binary classification constraint, $K = 2$. Due to the fact that the regularization factor successfully reduces model overfitting, we can construct the final loss function as Equation (3).

$$L_{final} = L_{dice} + \frac{\lambda}{2} \parallel \omega \parallel^2 \tag{3}$$

where $\frac{\lambda}{2} \parallel \omega \parallel^2$ is a regularization term to avoid overfitting; λ is the hyperparameter of regularization term.

4 PERFORMANCE EVALUATION

4.1 Datasets and Evaluation Criterion

To evaluate the effectiveness of our algorithm RGN-Net, we conducted trials on four medical picture segmentation datasets, including lung segmentation, DSB2018 cell segmentation, retinal vascular identification, and red blood cell segmentation. The DRIVE dataset is utilized for segmenting retinal vascular pictures, the lung dataset is used for lung cancer diagnosis and screening, and the ISBI dataset is used for segmenting cells.

To comprehensively evaluate the experimental results, we used seven evaluation metrics commonly used for medical image segmentation tasks, namely accuracy, specificity, sensitivity, precision, F1-Score, IoU, and Dice, which are shown as a cri-

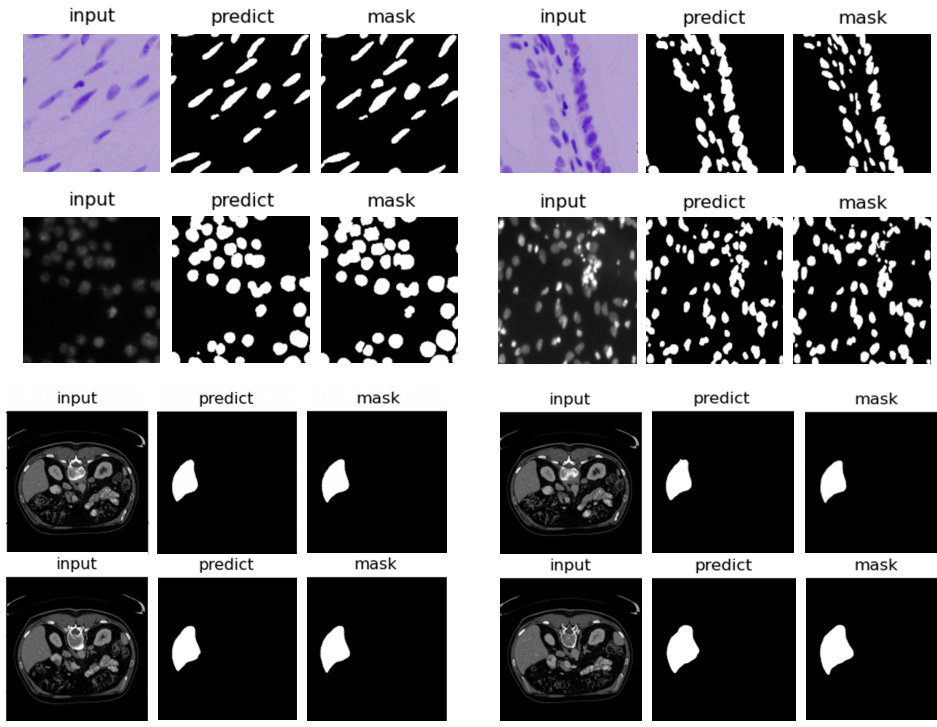

Figure 3. We extracted some prediction visualization results plots on several public benchmark datasets

terion group in Equation (4).

$$\begin{cases} Accuracy = \dfrac{TP + TN}{TP + FP + FN + TN}, \\ Sensitivity = \dfrac{TP}{TP + FN}, \\ F1 = \dfrac{2 \times (PR \times SN)}{PR + SN}, \\ IoU = \dfrac{TP}{TP + FN + FP}, \\ Dice = \dfrac{2 \times TP}{(TP + FN) + (TP + FP)}. \end{cases} \tag{4}$$

In this experiment, TP reflects the number of positive samples that were accurately anticipated. Similarly, TN indicates the number of accurately negative predicted samples, FP represents the number of positive predicted samples antici-

pated to be negative, and FN represents the number of negative predicted samples predicted to be positive. During the construction of the index, these four computed values account for every conceivable circumstance.

4.2 Experiment Configuration

Our method RGN-Net is built on the PyTorch deep learning library. We use the ResNet network, which has been pre-trained on ImageNet, as the backbone in the encoder stage, because it was experimentally found that ResNet outperforms other backbones, and we perform four downsampling operations to obtain a sufficiently comprehensive set of semantic features. We trained and tested the platform relying mainly on Ubuntu 20.04 system and two NVIDIA® RTX3090 graphics cards with 24 GB video memory. In the training process, we used a batch stochastic gradient descent (SGD) strategy, where the batch size was set differently for different datasets, with the main parameters being 8, 16, and 32. In addition, the network learning rate was set to 0.0001. In terms of optimizer selection, although we tested both Adam and SGD for comparative experiments, it was found that SGD usually achieves better performance, while Adam can converge in a shorter time. Our method adds more dropout layers to the network in order to prevent the overfitting phenomenon of the network training process, and the cut threshold is set to 0.7. Experimentally, this proves to be effective in reducing the overfitting phenomenon of the network.

4.3 Experimental Results

As shown in Figure 3, we visualize the visual segmentation result graph and ground truth comparison result graph of our method on several datasets. It can be seen that after several iterations of the model, the gap between the result graphs predicted by our algorithm and the ground truth result graphs is significantly reduced. This precisely proves the effectiveness of our method and the accuracy of our experimental design.

4.3.1 Cell Segmentation Datasets

The dataset utilized for the task of cell segmentation consisted of electron microscope photographs of cells annotated with cell outlines. Our mission is to precisely pinpoint the location of the cell outline in the image. Multiple cell segmentation datasets, including gland segmentation, red blood cell segmentation, and DSB2018 datasets, were used to test our approach. The glandular dataset comprises the anatomy of intestinal glands; the DSB2018 dataset includes a significant number of segmented nuclear pictures; and the erythrocyte dataset has 1 328 images. In addition, there are 1 328 basic facts that are related.

In Table 1, we present the common assessment metrics for the segmented cell dataset and erythrocyte dataset from the DSB 2018 release. Four assessment scores for the U-Net technique on the DSB2018 dataset are 0.8516, 0.8812, 0.8725, and

0.8833. UNet++ received scores of 0.9043, 0.9217, 0.9255, and 0.8974 across the four assessment measures. For precision, accuracy, IoU, and Dice on the DSB2018 dataset, the experimental results of our technique for the four separate evaluation metrics are 0.9340, 0.9339, 0.9422, and 0.9420, respectively, which are superior than the current state-of-the-art methods.

Methods	**Precision**	**Accuracy**	**IoU**	**Dice**
U-Net [16]	0.8516	0.8812	0.8725	0.8833
Attention U-Net [31]	0.8881	0.9093	0.9119	0.9235
R2U-Net [23]	0.8531	0.8732	0.9015	0.9033
UNet++ [18]	0.9043	0.9217	0.9255	0.8974
CE-Net [32]	0.9232	0.9243	0.9349	0.9156
DoubleU-Net [33]	0.9296	0.9307	0.9401	0.9233
ResUNet++ [34]	0.9330	0.9045	0.9231	0.9328
RGN-Net (Ours)	**0.9340**	**0.9339**	**0.9422**	**0.9420**

Table 1. Erythrocyte dataset performance comparison between the proposed network with state-of-the-art methods

4.3.2 Pulmonary Segmentation Dataset

The lung pictures in the Pulmonary Segmentation dataset consist of CT scans in two dimensions. The collection contains 267 photos, each with a dimension of 512 by 512 pixels. The 267 photos are divided into three groups: a training set, a test set, and a validation set. In the experimental data allocation, we utilize 80 % of the photos for training and the remaining images for testing and cross-validation. The assessment criteria consist of precision, sensitivity, IoU, and Dice. Table 2 displays the segmentation dataset findings for the lungs. Four distinct measures were employed to assess our experiments. U-Net achieved 0.9685, 0.9696, 0.9872, and 0.9784 for the four assessment measures, but UNet++ achieved 0.9812, 0.9734, 0.9815, and 0.9836. Our technique yielded the highest F1, accuracy, and AUC values, respectively 0.9912, 0.9980, and 0.9957.

Methods	**F1-Score**	**Sensitivity**	**Accuracy**	**AUC**
U-Net [16]	0.9658	0.9696	0.9872	0.9784
Attention U-Net [31]	0.9783	0.9784	0.9892	0.9834
RU-Net [35]	0.9638	0.9734	0.9836	0.9800
R2U-Net [23]	0.9832	**0.9944**	0.9918	0.9940
BCDU-Net [22]	0.9904	0.9910	0.9972	0.9946
UNet++ [18]	0.9812	0.9734	0.9815	0.9836
CE-Net [32]	0.9823	0.9831	0.9920	0.9812
RGN-Net (Ours)	**0.9912**	0.9862	**0.9980**	**0.9957**

Table 2. Performance comparison of the proposed network and current methods on a pulmonary dataset

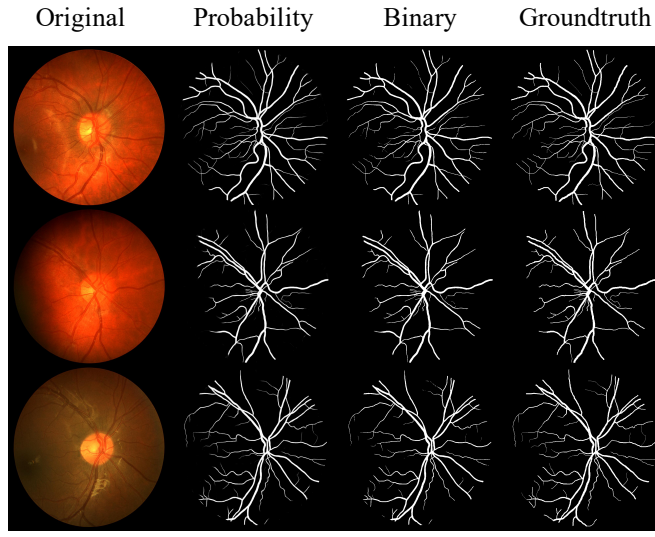

Figure 4. The segmentation results of RGN-Net on DRIVE dataset

Methods	**Accuracy**	**Specificity**	**Sensitivity**	**AUC**
U-Net	0.9531	0.9820	0.7537	0.9680
Attention U-Net	0.9629	0.9725	0.7884	0.9740
Deep Model	0.9495	0.9768	0.7763	0.9720
RU-Net	0.9553	0.9820	0.7726	0.9779
R2U-Net	0.9652	0.8303	0.7792	0.9245
BCDU-Net	0.9560	0.9786	0.8007	0.9789
UNet++	0.9656	0.9867	0.8234	0.9628
CE-Net	0.9545	0.9851	0.8309	0.9779
Backbone	0.9477	–	0.7781	0.9705
Fusion Mechanism	0.8247	0.9847	0.8140	0.9782
RGN-Net (Ours)	**0.9684**	**0.9937**	**0.8443**	**0.9868**

'–' represents the data is not available

Table 3. Performance comparison of the proposed network and the state-of-the-art methods on DRIVE dataset

4.3.3 Vessel Segmentation Dataset

DRIVE is a data collection designed to separate blood vessels from retinal pictures. It comprises of forty color retinal pictures, of which twenty are utilized for training and the remaining twenty for assessment. These photos were originally 565×584 pixels in size. Having such a sample dataset under normal conditions is insufficient to train a deep neural network. Given that deep learning networks must converge, the model must be backed with a significant quantity of high-quality

data so that the network may acquire more knowledge. Consequently, we use data augmentation methods and the following ways to overcome the aforementioned issue: First, random blocks were generated from the input photos. Twenty training photos produced a total of around 190 000 patches, of which 171 000 were utilized for training and the remaining images were used to DRIVE 19 000 patches of segmentation data for validation. The batch size employed as network input data was 64×64.

We show the segmentation results of RGN-Net on the DRIVE dataset in Figure 4. The four columns of data are the original color image, the predicted probability image, the predicted binary image, and the ground truth. In addition, we list in Table 3 other state-of-the-art works and quantitative results obtained by the proposed network RGN-Net on the DRIVE dataset. We used four different evaluation metrics to evaluate our experiments, where U-Net method obtained 0.9531, 0.9820, 0.7537, 0.9680 on DSB2018 dataset from four evaluation metrics. UNet++ obtained 0.9656, 0.9867, 0.8234, 0.9628 on four evaluation metrics. We can see that RGN-Net obtained 0.9656, 0.9867, 0.8234, 0.9628 on accuracy, specificity, sensitivity and AUC metrics achieved excellent results with values of 0.9684, 0.9937, 0.8443 and 0.9868, respectively.

5 CONCLUSION

In this study, we introduce the innovative combinatorial network RGN-Net, an end-to-end system for medical picture segmentation using deep learning. During downsampling and stepwise feature map extraction, we use an attention strategy based on the sum of two spatial orientations, which captures long-distance dependencies along one space while keeping precise position information along the other space. The resultant feature maps are then encoded as two direction-aware and position-sensitive attention maps that may be applied to the input feature maps to enhance the representation of significant objects. In addition, by cascading a multi-scale information extraction module in the network, our method takes into consideration the vast variety of medical image scales. To lower the weight of uninteresting areas while stitching downsampled and upsampled feature maps, we calculate the difference between the target region and the region of no interest. Our experimental findings demonstrate that our suggested technique may enhance the segmentation of medical pictures for a variety of applications, such as the segmentation of diverse cellular datasets and the identification of retinal vascular structures in the lungs. RGN-Net earns the greatest F1 score on the lung dataset, which corresponds to 99.12 % of the values discovered. This method should be applicable to additional 2D MIS initiatives. Future study will focus further on reducing the number of model parameters to reduce model complexity and enhance forecast accuracy.

Acknowledgment

The authors are thankful to reviewers for their insightful remarks and recommendations. Additionally, we would like to thank our coworkers and students for their assistance in our laboratory.

REFERENCES

[1] Liang, Z.—Powell, A.—Ersoy, I.—Poostchi, M.—Silamut, K.—Palaniappan, K. et al.: CNN-Based Image Analysis for Malaria Diagnosis. 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2016, pp. 493–496, doi: 10.1109/BIBM.2016.7822567.

[2] Karbalayghareh, A.—Qian, X.—Dougherty, E. R.: Optimal Bayesian Transfer Learning for Count Data. IEEE/ACM Transactions on Computational Biology and Bioinformatics, Vol. 18, 2021, No. 2, pp. 644–655, doi: 10.1109/TCBB.2019.2920981.

[3] Chiu, Y. C.—Hsiao, T. H.—Wang, L. J.—Chen, Y.—Chuang, E. Y.: Analyzing Differential Regulatory Networks Modulated by Continuous-State Genomic Features in Glioblastoma Multiforme. IEEE/ACM Transactions on Computational Biology and Bioinformatics, Vol. 15, 2018, No. 6, pp. 1754–1764, doi: 10.1109/TCBB.2016.2635646.

[4] Sheng, W.—Chen, S.—Sheng, M.—Xiao, G.—Mao, J.—Zheng, Y.: Adaptive Multisubpopulation Competition and Multiniche Crowding-Based Memetic Algorithm for Automatic Data Clustering. IEEE Transactions on Evolutionary Computation, Vol. 20, 2016, No. 6, pp. 838–858, doi: 10.1109/TEVC.2016.2524555.

[5] Shen, C.—Ding, Y.—Tang, J.—Guo, F.: Multivariate Information Fusion with Fast Kernel Learning to Kernel Ridge Regression in Predicting LncRNA-Protein Interactions. Frontiers in Genetics, Vol. 9, 2019, Art. No. 716, doi: 10.3389/fgene.2018.00716.

[6] Tsai, A.—Yezzi, A.—Wells, W.—Tempany, C.—Tucker, D.—Fan, A.—Grimson, W. E.—Willsky, A.: A Shape-Based Approach to the Segmentation of Medical Imagery Using Level Sets. IEEE Transactions on Medical Imaging, Vol. 22, 2003, No. 2, pp. 137–154, doi: 10.1109/TMI.2002.808355.

[7] Yu, Y.—Li, M.—Liu, L.—Wu, F. X.—Wang, J.: Tentative Diagnosis Prediction via Deep Understanding of Patient Narratives. 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2019, pp. 1000–1003, doi: 10.1109/BIBM47256.2019.8983129.

[8] Cheng, S.—Wu, Y.—Li, Y.—Yao, F.—Min, F.: TWD-SFNN: Three-Way Decisions with a Single Hidden Layer Feedforward Neural Network. Information Sciences, Vol. 579, 2021, pp. 15–32, doi: 10.1016/j.ins.2021.07.091.

[9] Long, J.—Shelhamer, E.—Darrell, T.: Fully Convolutional Networks for Semantic Segmentation. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3431–3440, doi: 10.1109/CVPR.2015.7298965.

[10] Jia, C.—Shi, F.—Zhao, M.—Zhang, Y.—Cheng, X.—Wang, M.—Chen, S.: Semantic Segmentation with Light Field Imaging and Convolutional Neural Networks.

IEEE Transactions on Instrumentation and Measurement, Vol. 70, 2021, pp. 1–14, doi: 10.1109/TIM.2021.3115204.

[11] Ben-Cohen, A.—Klang, E.—Kerpel, A.—Konen, E.—Amitai, M. M.—Greenspan, H.: Fully Convolutional Network and Sparsity-Based Dictionary Learning for Liver Lesion Detection in CT Examinations. Neurocomputing, Vol. 275, 2018, pp. 1585–1594, doi: 10.1016/j.neucom.2017.10.001.

[12] Drozdzal, M.—Vorontsov, E.—Chartrand, G.—Kadoury, S.—Pal, C.: The Importance of Skip Connections in Biomedical Image Segmentation. In: Carneiro, G., Mateus, D., Peter, L. et al. (Eds.): Deep Learning and Data Labeling for Medical Applications (DLMIA 2016, LABELS 2016). Springer, Cham, Lecture Notes in Computer Science, Vol. 10008, 2016, pp. 179–187, doi: 10.1007/978-3-319-46976-8_19.

[13] Badrinarayanan, V.—Kendall, A.—Cipolla, R.: SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 39, 2017, No. 12, pp. 2481–2495, doi: 10.1109/TPAMI.2016.2644615.

[14] Chen, L. C.—Papandreou, G.—Kokkinos, I.—Murphy, K.—Yuille, A. L.: DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 40, 2018, No. 4, pp. 834–848, doi: 10.1109/TPAMI.2017.2699184.

[15] Fu, J.—Liu, J.—Tian, H.—Li, Y.—Bao, Y.—Fang, Z.—Lu, H.: Dual Attention Network for Scene Segmentation. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 3141–3149, doi: 10.1109/CVPR.2019.00326.

[16] Ronneberger, O.—Fischer, P.—Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W. M., Frangi, A. F. (Eds.): Medical Image Computing and Computer-Assisted Intervention –MICCAI 2015. Springer, Cham, Lecture Notes in Computer Science, Vol. 9351, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[17] Codella, N. C. F.—Gutman, D.—Celebi, M. E.—Helba, B.—Marchetti, M. A.—Dusza, S. W.—Kalloo, A.—Liopyris, K.—Mishra, N.—Kittler, H.—Halpern, A.: Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC). 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), 2018, pp. 168–172, doi: 10.1109/ISBI.2018.8363547.

[18] Zhou, Z.—Siddiquee, M. M. R.—Tajbakhsh, N.—Liang, J.: UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In: Stoyanov, D., Taylor, Z., Carneiro, G. et al. (Eds.): Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (DLMIA 2018, ML-CDS 2018). Springer, Cham, Lecture Notes in Computer Science, Vol. 11045, 2018, pp. 3–11, doi: 10.1007/978-3-030-00889-5_1.

[19] Ibtehaz, N.—Rahman, M. S.: MultiResUNet: Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation. Neural Networks, Vol. 121, 2020,

pp. 74–87, doi: 10.1016/j.neunet.2019.08.025.

[20] Mehta, S.—Mercan, E.—Bartlett, J.—Weaver, D.—Elmore, J. G.—Shapiro, L.: Y-Net: Joint Segmentation and Classification for Diagnosis of Breast Biopsy Images. In: Frangi, A. F., Schnabel, J. A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (Eds.): Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. Springer, Cham, Lecture Notes in Computer Science, Vol. 11071, 2018, pp. 893–901, doi: 10.1007/978-3-030-00934-2_99.

[21] Alom, M. Z.—Yakopcic, C.—Taha, T. M.—Asari, V. K.: Nuclei Segmentation with Recurrent Residual Convolutional Neural Networks Based U-Net (R2U-Net). NAECON 2018 – IEEE National Aerospace and Electronics Conference, 2018, pp. 228–233, doi: 10.1109/NAECON.2018.8556686.

[22] Azad, R.—Asadi-Aghbolaghi, M.—Fathy, M.—Escalera, S.: Bi-Directional ConvLSTM U-Net with Densley Connected Convolutions. 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2019, pp. 406–415, doi: 10.1109/ICCVW.2019.00052.

[23] Alom, M. Z.—Hasan, M.—Yakopcic, C.—Taha, T. M.—Asari, V. K.: Recurrent Residual Convolutional Neural Network Based on U-Net (R2U-Net) for Medical Image Segmentation. 2018, doi: 10.48550/arXiv.1802.06955.

[24] Shi, X.—Chen, Z.—Wang, H.—Yeung, D. Y.—Wong, W. K.—Woo, W. C.: Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In: Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., Garnett, R. (Eds.): Advances in Neural Information Processing Systems 28 (NIPS 2015). Curran Associates, Inc., 2015, pp. 802–810.

[25] Child, R.—Gray, S.—Radford, A.—Sutskever, I.: Generating Long Sequences with Sparse Transformers. 2019, doi: 10.48550/arXiv.1904.10509.

[26] Chen, J.—He, Y.—Frey, E. C.—Li, Y.—Du, Y.: ViT-V-Net: Vision Transformer for Unsupervised Volumetric Medical Image Registration. 2021, doi: 10.48550/arXiv.2104.06468.

[27] Chen, J.—Lu, Y.—Yu, Q.—Luo, X.—Adeli, E.—Wang, Y.—Lu, L.—Yuille, A. L.—Zhou, Y.: TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. 2021, doi: 10.48550/arXiv.2102.04306.

[28] Valanarasu, J. M. J.—Oza, P.—Hacihaliloglu, I.—Patel, V. M.: Medical Transformer: Gated Axial-Attention for Medical Image Segmentation. In: de Bruijne, M., Cattin, P. C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., Essert, C. (Eds.): Medical Image Computing and Computer-Assisted Intervention – MICCAI 2021. Springer, Cham, Lecture Notes in Computer Science, Vol. 12901, 2021, pp. 36–46, doi: 10.1007/978-3-030-87193-2_4.

[29] Liu, Z.—Lin, Y.—Cao, Y.—Hu, H.—Wei, Y.—Zhang, Z.—Lin, S.—Guo, B.: Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. 2021, doi: 10.48550/arXiv.2103.14030.

[30] Cao, Y.—Xu, J.—Lin, S.—Wei, F.—Hu, H.: GCNet: Non-Local Networks Meet Squeeze-Excitation Networks and Beyond. 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2019, pp. 1971–1980, doi: 10.1109/ICCVW.2019.00246.

[31] Oktay, O.—Schlemper, J.—Le Folgoc, L.—Lee, M.—Heinrich, M.—Misawa, K. et al.: Attention U-Net: Learning Where to Look for the Pancreas. 2018, doi: 10.48550/arXiv.1804.03999.

[32] Gu, Z.—Cheng, J.—Fu, H.—Zhou, K.—Hao, H.—Zhao, Y.—Zhang, T.—Gao, S.—Liu, J.: CE-Net: Context Encoder Network for 2D Medical Image Segmentation. IEEE Transactions on Medical Imaging, Vol. 38, 2019, No. 10, pp. 2281–2292, doi: 10.1109/TMI.2019.2903562.

[33] Jha, D.—Riegler, M. A.—Johansen, D.—Halvorsen, P.—Johansen, H. D.: DoubleU-Net: A Deep Convolutional Neural Network for Medical Image Segmentation. 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), 2020, pp. 558–564, doi: 10.1109/CBMS49503.2020.00111.

[34] Jha, D.—Smedsrud, P. H.—Riegler, M. A.—Johansen, D.—De Lange, T.—Halvorsen, P.—Johansen, H. D.: ResUNet++: An Advanced Architecture for Medical Image Segmentation. 2019 IEEE International Symposium on Multimedia (ISM), 2019, pp. 2250–2255, doi: 10.1109/ISM46123.2019.00049.

[35] Jaeger, P. F.—Kohl, S. A. A.—Bickelhaupt, S.—Isensee, F.—Kuder, T. A.—Schlemmer, H. P.—Maier-Hein, K. H.: Retina U-Net: Embarrassingly Simple Exploitation of Segmentation Supervision for Medical Object Detection. In: Dalca, A. V., McDermott, M. B., Alsentzer, E., Finlayson, S. G., Oberst, M., Falck, F., Beaulieu-Jones, B. (Eds.): Proceedings of the Machine Learning for Health NeurIPS Workshop. Proceedings of Machine Learning Research (PMLR), Vol. 116, 2020, pp. 171–183.

Zhixin Zhang received his Master degree in computer science and technology from the Tianjin University of Technology. He is currently Lecturer in the College of Information Engineering, Tianjin University of Commerce. His research interests include image recognition and intelligence computing.

Shuhao Jiang received his Master degree in engineering from the Tianjin Normal University and his Ph.D. from the Tianjin University. He is currently Professor in the College of Information Engineering, Tianjin University of Commerce. His research interests include intelligence computing and natural language processing.

Xuhua Pan received his Master degree in computer science and technology from the Jilin University. He is currently Professor in the College of Information Engineering, Tianjin University of Commerce. His research interests include intelligence computing and data handling.