Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/mbe

MBE, 21(3): 4351–4369.
DOI: 10.3934/mbe.2024192
Received: 19 November 2023
Revised: 15 February 2024
Accepted: 20 February 2024
Published: 26 February 2024

Research article

Biomedical image segmentation algorithm based on dense atrous convolution

Hong'an Li[1,3], Man Liu[1], Jiangwen Fan[1] and Qingfang Liu[2,*]

[1] College of Computer Science and Technology, Xi'an University of Science and Technology, Xi'an, 710054, China

[2] Information Center, Shijiazhuang Posts and Telecommunications Technical College, Shijiazhuang 050021, China

[3] State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing, 100191, China

* **Correspondence:** Email:liuqf@sjzpc.edu.cn.

Abstract: Biomedical images have complex tissue structures, and there are great differences between images of the same part of different individuals. Although deep learning methods have made some progress in automatic segmentation of biomedical images, the segmentation accuracy is relatively low for biomedical images with significant changes in segmentation targets, and there are also problems of missegmentation and missed segmentation. To address these challenges, we proposed a biomedical image segmentation method based on dense atrous convolution. First, we added a dense atrous convolution module (DAC) between the encoding and decoding paths of the U-Net network. This module was based on the inception structure and atrous convolution design, which can effectively capture multi-scale features of images. Second, we introduced a dense residual pooling module to detect multi-scale features in images by connecting residual pooling blocks of different sizes. Finally, in the decoding part of the network, we adopted an attention mechanism to suppress background interference by enhancing the weight of the target area. These modules work together to improve the accuracy and robustness of biomedical image segmentation. The experimental results showed that compared to mainstream segmentation networks, our segmentation model exhibited stronger segmentation ability when processing biomedical images with multiple-shaped targets. At the same time, this model can significantly reduce the phenomenon of missed segmentation and missegmentation, improve segmentation accuracy, and make the segmentation results closer to the real situation.

Keywords: biomedical image segmentation; deep learning; dense atrous convolution; dense residual pooling; multi-scale features; attention mechanism

1. Introduction

In recent years, the number of cancer patients has been continuously increasing and cancer has become a common concern. Early detection and treatment of cancer can greatly improve the survival rate of patients. In this context, using medical image segmentation technology to accurately identify and segment the lesion area of patients can help doctors accurately assess the patient's condition and formulate appropriate treatment plans. However, the tissue structure of biomedical images is relatively complex, with not only a large number of organs but also significant differences in images of the same part of different individuals, which greatly increases the difficulty of medical image segmentation [1]. At the same time, the problem of missegmentation and missed segmentation in existing medical image segmentation methods greatly affects the accuracy of image segmentation, thereby interfering with doctors making correct judgments and affecting the patient's treatment process. Therefore, it is particularly important for the treatment of diseases to efficiently utilize computer technology to improve the accuracy of medical image segmentation and reduce the phenomenon of missegmentation and missed segmentation.

The essence of image segmentation is to predict the value of each pixel in an image, but without precondition, the color value of a pixel is diverse and the result of color prediction for the same pixel is inaccurate. Traditional segmentation methods rely on detecting the grayscale changes of pixels in the target region and extracting the boundaries by using the abrupt changes of pixel grayscale and the discontinuity of the image to achieve image segmentation. Chen et al. used the single-scale Harris corner point method combined with a statistical algorithm, spatial domain algorithm, and dynamic queuing method to detect the edge contour features of ultrasound images and enhance and fuse the features of the detected edge information [2]. Aslam et al. used the Sobel edge detection algorithm combined with the correlation thresholding method to find different regions using closed contours; they finally segmented tumors from the images according to the different intensities of the information within the contours and achieved good segmentation results [3]. Eijnatten et al. divided the pixel points in the image into several classes by converting the grayscale image into a binary image, and segmented the image into different regions according to the grayscale difference between the target object to be segmented and the background [4].

The biomedical image segmentation algorithm based on deep learning can be traced back to 2015, when Long et al. proposed the famous fully convolutional neural networks (FCN) [5]. Based on this, Ronneberger et al. proposed a U-shaped convolutional neural network (U-Net) for biomedical image segmentation, which is a symmetric structure with encoding and decoding paths [6]. The feature fusion between the two paths is performed by jumping connections, which effectively alleviates the situation that the semantic information and the spatial domain information in the FCN model are incompatible. The highest accuracy was achieved in the segmentation of cell images and liver CT (Computed Tomography) images at that time. Oktay et al. improved based on U-Net and proposed AtU-Net (Attention U-Net), which improves segmentation efficiency by introducing the attention mechanism in the decoding stage [7]. Zhou et al. proposed U-Net++, in which the encoder and decoder subnetworks are connected by a series of dense hopping paths. The redesigned hopping paths can reduce the semantic gap between the feature mappings of the encoder and decoder subnetworks, which can process similar semantic images more easily [8]. Alom et al. formed R2U-Net (recurrent residual convolutional neural network based on U-Net) by combining residual networks, U-Net, and recurrent neural networks, which ensures better feature representation for segmentation tasks through feature accumulation in the convolutional layer, and these models have greatly advanced the development progress in the field of medical image segmentation [9]. Based on the idea of feature fusion, Huang et al. proposed DenseNet, which fuses any two layers of the network into a convolutional neural network with dense connectivity. The input feature of the current layer is the set of the output feature mapping of all previous network layers during the network propagation [10]. It can maximize the preservation of feature information in the network propagation process, alleviate the gradient disappearance problem, and make the network easier to train and have a certain regularity effect. Based on the research of V-net (a fully convolutional neural network for volumetric medical image segmentation), Zhu et al. proposed a three-dimensional (3D) end-to-end FCN called semantic V-net (SV-net), which consists of a downsampling path and an upsampling path, with the downsampling path used for feature extraction and the upsampling path used for recovering downsampling features. During the downsampling process, features are automatically extracted from the image through convolution operators. The ultimate goal is to automatically extract thyroid-related ophthalmopathy (EOM) and optic nerve (ON) from orbital CT through this network [11]. Li et al. proposed an improved FCN network, namely, AtFcn, which makes full use of a FCN and attention mechanism to achieve pixel-level accurate segmentation of images of arbitrary size [12]. Liu et al. proposed an OCTA (Optical Coherence Tomography Angiography) retinal vessel segmentation network, which mainly includes a dual-branch encoder based on adaptive gated axial transformer and residual module, a decoder, and a point repair module based on residual network. The encoder branch of the network exchanges a large amount of global and local information through feature interaction units, thereby preserving a large amount of detailed information. The point repair module of the network re-predicts the uncertain points in the low visibility region in the OCTA image. The various modules in this network work together to finally achieve the precise segmentation of the retinal vascular [13]. Mu et al. proposed an attention-based multi-scale supervised fully convolutional encoder-decoder network (ARU-Net), which combines depth-aware attention gates and multi-scale supervision strategy to achieve accurate segmentation of intracranial aneurysms and adjacent arteries in three-dimensional rotational angiography (3DRA) images [14].

At the same time, we have also noticed that customizing segmentation methods for specific tasks can lead to fragmentation between various segmentation tasks, so many advanced studies have focused on improving the consistency of segmentation models. Qin et al. proposed a unified, universal, and open framework for local image segmentation, namely, FreeSeg. It is mainly divided into two stages: The first stage extracts universal mask proposals, and the second stage uses CLIP (Contrastive Language–Image Pretraining) to perform zero sample classification on the masks generated in the first stage. At the same time, an adaptive prompt learning method was proposed to encode arbitrary tasks and categories into compact textual abstraction, improving the robustness of the model in multiple tasks and different scenarios [15]. Kirillov et al. proposed a model that can segment everything, namely, SAM (segment everything model). SAM is a prompt-based model that exhibits strong zero sample generalization ability, greatly promoting the development of basic models. Its architecture mainly consists of three components: image encoder, prompt encoder, and fast mask decoder. The image encoder uses a pretrained visual converter (ViT) using MAE (masked autoencoders), the computational complexity of ViT is the square of the number of pixels, and its development in visual tasks is limited by its enormous computational cost [16]. The prompt encoder mainly includes two sets of prompts, namely, sparse prompts (points, boxes, text) and dense prompts (masks), which are represented in different ways. The mask decoder maps the image embedding, prompt embeddings, and an output token to a mask. The SAM network has achieved good results in multiple tasks [17]. However, the high computational cost of this model limits its widespread application in industry scenarios. To address this issue, Zhao et al. subsequently proposed a high-performance accelerated alternative method called FastSAM, which consists of two stages: all-instance segmentation and prompt-guided selection. FastSAM uses the YOLOv8-Seg (an instance segmentation model based on YOLO series object detection framework) method for the all-instance segmentation phase. After successfully segmenting all objects or regions in the image using the YOLOv8 method, the second stage task is to use various prompts to identify specific objects of interest. It mainly involves point prompts, box prompts, and text prompts. Compared with the SAM model, FastSAM has lower computational costs while ensuring segmentation performance [18].

Medical image segmentation has gone through three stages: manual segmentation, traditional segmentation, and methods based on deep learning; among which, manual segmentation has the highest accuracy, but it is often time-consuming and laborious, and manual segmentation varies according to doctors' experience and subjective judgment, and different doctors may have different segmentation results for the same biomedical image. Traditional segmentation algorithms have good generality, but the segmentation accuracy is low. In recent years, the rapid development of image segmentation based on deep learning can significantly improve the efficiency of image segmentation, but for the segmentation of biomedical images with large differences in feature size, missed segmentation and missegmentation can easily occur, resulting in low segmentation accuracy. To address the above problems, based on the structure design of U-Net, this paper proposes a biomedical image segmentation method based on dense atrous convolution. The method designs a dense atrous convolution module based on the initial structure and atrous convolution with the aim of extracting multi-scale information from images. Compared to traditional convolutions, atrous convolution can support exponential expansion of receptive domains to adapt to multi-scale contextual information without losing image resolution, which is often overlooked by most existing models. In addition, fully extracting multi-scale features of images is of great significance in biomedical image segmentation tasks, and effectively understanding and utilizing these features is also a highly important aspect in achieving medical image segmentation tasks. Therefore, this method fully utilizes the multi-scale information of the extracted image by introducing dense residual pooling. Finally, this method restores the size of the feature map through upsampling during the decoding stage, introduces an attention gate during the decoding stage, and enhances the target features by reducing the background weight. The specific content of this paper's innovation is as follows:

(1) A dense atrous convolution module has been added between the encoding and decoding paths of the U-Net network. This module is designed based on the Inception structure and atrous convolution and extracts multi-scale image features through multi-channel and multi-scale convolution kernels, enhancing the network's ability to extract image features. At the same time, it avoids the problem of gradient explosion and vanishing when extracting multi-scale image features.

(2) This paper designs a dense residual pooling module to widen the network using the Inception structure to make full use of the acquired image features. The combination of this module and the dense atrous convolution module not only improves the network's feature extraction ability, but also enhances the network's ability to process image features.

(3) Introducing the attention mechanism in the decoding stage highlights the target by increasing the

weight of the target area while suppressing the influence of irrelevant areas on segmentation, thereby improving segmentation accuracy and reducing the occurrence of missegmentation and missed segmentation.

2. Materials and methods

2.1. Related work

2.1.1. Inception structure

Convolutional neural networks usually extract local features during convolutional operations. Since the relevant features in an image may be far apart, smaller convolutional kernels often fail to learn the true features, and larger convolutional kernels may result in less detailed extracted features. The Inception structure can efficiently express the sparse structure of features, aiming to solve the computational redundancy caused by the stacking of convolutional layers. It was first proposed in 2015 by Szegedy et al. [19]. After that, from Incep-tion-V2 [20], Inception-V3 [21] (Ioffe and Szegedy; Szegedy et al.), to Inception-V4 [22] (Szegedy et al.), Inception networks have been continuously improved and innovated to achieve better performance.

The Inception structure can solve the computational redundancy caused by the accumulation of convolutional layers and deepen the number of layers and breadth of the network. By connecting convolutional kernels of different scales in parallel, multiple branches can be used to extract multi-scale image information, enhance the generalization ability of the network, and improve the learning ability of the convolutional neural network for features [23].

Multiple Inception modules in series will form the Inception structure. As shown in Figure 1, the Inception module consists of four channels, each of which uses convolutional kernels of different sizes to extract rich features [24]. To reduce computational complexity, each channel contains a 1 × 1 convolution kernel for dimensionality reduction. Each Inception module cascades the feature maps of the four channels through mapping to the next Inception module, forming the whole Inception structure and enhancing the feature extraction capability of the network.

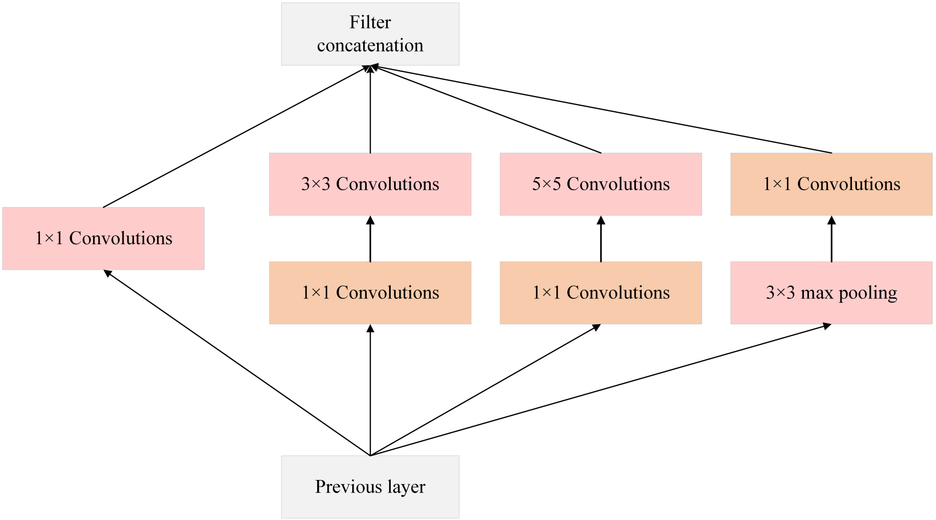
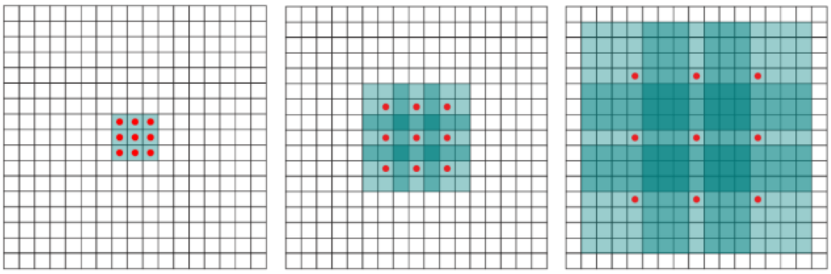


Figure 1. Original inception module diagram [24].

2.1.2. Atrous convolution

Traditional image segmentation algorithms use larger convolution kernels to increase the perceptual field and, finally, upsampling to restore the image size. In this process, the feature map undergoes shrinking and enlarging, which degrades the image accuracy. Compared with traditional image segmentation methods, atrous convolution automatically enlarges the image field by extracting sparse features and inserting cavities into the convolution kernel to form atrous convolution kernels with different expansion rates to obtain different sizes of the image field [25]. The size of the feature map can be kept constant while expanding the field so that the image accuracy will not be degraded during the feature extraction.

The atrous convolution with different expansion rates is shown in Figure 2, where the expansion rates are 1, 2, and 4, respectively [26].



(a) Expansion rate = 1 (b) Expansion rate = 2 (c) Expansion rate = 4

Figure 2. Schematic diagram of atrous convolution [26].

The size of the convolution kernel after the atrous convolution and the size of the original convolution kernel are calculated as

$$k' = k + (k + 1)(d - 1) \tag{2.1}$$

where $k^{'}$ denotes the size of the actual convolution kernel after expansion, k denotes the size of the original convolution kernel, and d denotes the expansion rate. When the expansion rate of the atrous convolution is 1, the effect of the atrous convolution is the same as that of the normal convolution. When the atrous convolution is used for image segmentation, the image feature map resolution can be maintained and the feature information in the image can be located more accurately.

2.1.3. Attention mechanism

The attention mechanism is similar to the human perception process, using top information to guide the bottom-up pre-feedback process, and has been applied to deep and recurrent neural networks [27]. Since the attention module pools resources to process useful information, the training time is significantly reduced. Each attention mask in the attention module can be used not only as a feature selector in forward pushing, but also as a gradient updater in back propagation. In recent years, the attention module has been used in a lot of stacking structures to mine more image depth features by stacking attention modules [28].

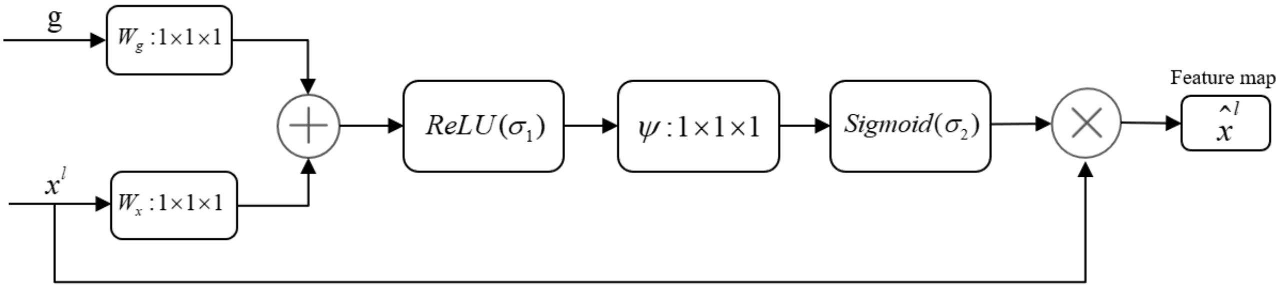


Figure 3. Attention module diagram.

The incremental nature of the stacked structure can refine the processing of complex images, and the features become clearer as the depth increases. In this paper, we use the attention mechanism in the decoding stage to extract the image depth features, suppress the background interference by increasing the target region weight, refine the processing of the image, make the features clearer, and reduce the information redundancy.

The internal structure of the attention module is shown in Figure 3. The inputs to this module are the feature map x^l and the upsampled feature map g. The feature map x^l is a coded feature of the same resolution passed through a jump connection, and the upsampled feature g can be regarded as a gating signal to enhance the learning ability of feature x^l. After the 1×1 convolution of the two inputs, the two feature maps are fused and activated by the ReLU (Rectified Linear Unit) function. The combined features are convolved again and the Sigmoid activation function is used to obtain the final attention coefficient AG, which is

$$AG = \sigma(C_{1\times1}(ReLU(C_{1\times1}(G) \times C_{1\times1}(X)))) \times X \tag{2.2}$$

where X, G correspond to the feature map x^l and feature map g in the above figure, $C_{1\times1}$ denotes 1×1 convolution, and σ denotes Sigmoid activation function.

2.2. Segmentation model of this paper

2.2.1. Network structure

The convolutional segmentation network based dense atrous is designed based on U-Net, and the network structure consists of encoding and decoding paths, as shown in Figure 4. The encoding stage of the network contains four downsampling modules, which can reduce the feature map size and learn more semantic information about the image. In the middle of the encoding and decoding paths of the network, the dense atrous convolution module is designed by combining Inception structure and atrous convolution to fully extract the multi-scale information in the image, and after the dense atrous convolution module, dense residual pooling is designed and introduced to fully utilize the extracted multi-scale information by concatenating multiple residual pooling modules.

In the decoding stage of the network, the feature map size is recovered by upsampling, and the attention gate is introduced in the decoding stage to enhance the target features by reducing the background weights. The convolutional kernel sizes for the encoding and decoding stages are shown in Table 1.

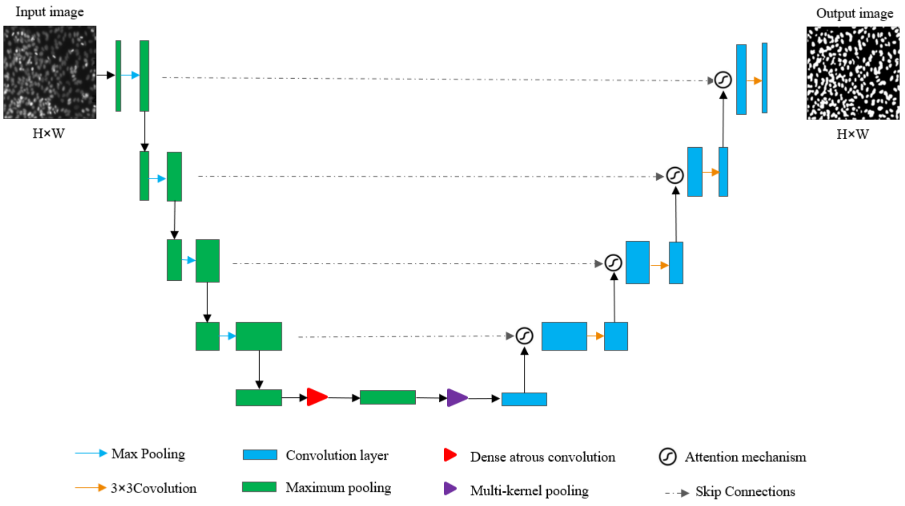


Figure 4. Network structure of this paper.

Table 1. Network structure parameters of this paper.

Encoding stage	Convolution kernel size	Decoding stage	Convolution kernel size
1	[3,3,3,64]	6	[3,3,1024,512]
2	[3,3,64,128]	7	[3,3,512,256]
3	[3,3,128,256]	8	[3,3,256,128]
4	[3,3,256,512]	9	[3,3,128,64]
5	[3,3,512,1024]	10	[3,3,64,1]

2.2.2. Dense atrous convolution

The structure of biomedical images is complex, and it is difficult to extract image boundary information and deep linguistic information for a small perceptual field model, so we use atrous convolution to increase the perceptual field of the model.

Based on the Inception structure and atrous convolution, we propose a dense atrous convolution module to encode deep semantic feature maps. As shown in Figure 5, the dense atrous convolution has four channels with an increasing number of convolutions in each channel, and the perceptual fields of the four channels are 3, 7, 9, and 19, respectively. Each channel is activated by 1×1 convolution. In this module, the features extracted from the four channels are fused with the original features through the jump connection to avoid the gradient disappearance and gradient explosion effectively.

With the support of multiple perceptual fields of the dense atrous convolution module, it is possible to extract the boundary information of biomedical images and the contextual relationship with other regions. Padding is used to keep the size of the feature map constant after convolution with cavities, and the size of the feature map after convolution with cavities is calculated as follows.

$$S = \frac{a + 2p - d \times (f - 1) - 1}{l} + 1 \tag{2.3}$$

where S denotes the output feature map size, a denotes the input feature map size, p denotes the number of layers filled with 0 by Padding operation, *d* denotes the expansion rate of the atrous convolution, *f* denotes the original convolution kernel size, and *l* denotes the convolution step size.

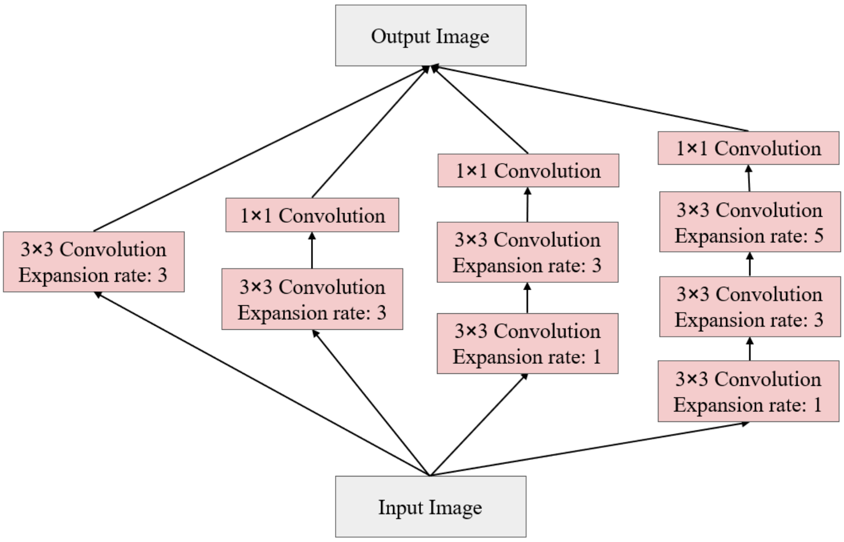


Figure 5. Dense atrous convolution module.

2.2.3. Dense residual pooling

One of the challenges in biomedical image segmentation is the great variation in target size, and biomedical images have more complex tissue structures. Not only is the number of organs large, but also the images of the same part of an individual are extremely different, and the same part of the same body can be very different at different times, so it is important to enhance the segmentation network for feature extraction of different sizes.

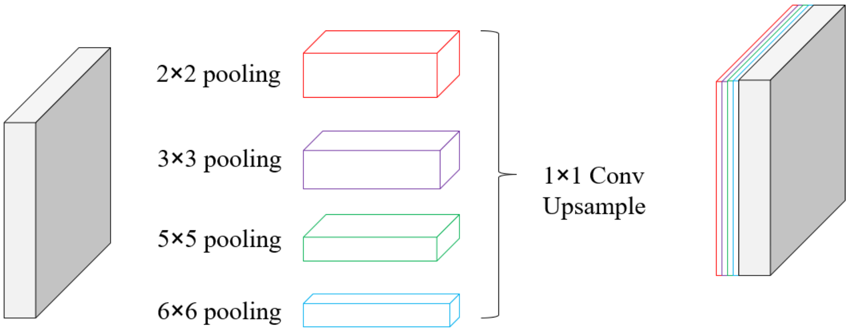


Figure 6. Dense residual pooling module.

The size of the perceptual field roughly determines how much contextual information we can use. Dense atrous convolution can enhance the extraction of features of different sizes, but ordinary pooling cannot effectively process multiple features after extraction, so this paper proposes a dense residual pooling to enhance the processing of features after extraction of dense atrous convolution, relying on multiple effective fields of view to detect targets of different sizes.

As shown in Figure 6, the dense residual pooling module uses four different sizes of perceptual fields to encode global contextual information, and the four pooling kernels are 2×2, 3×3, 5×5, and 6×6. The four branches output feature maps of various sizes. To reduce the feature dimensionality and computational cost, we use 1×1 convolution after each pooling branch. It reduces the size of the feature map to 1/N of the original dimension, where N denotes the number of channels in the original feature map. The low-dimensional feature map is upsampled to obtain the same dimensional features as the original feature map, and, finally, the original features are combined with the upsampled feature map.

3. Experiment and analysis

3.1. Experimental data and environment

This paper aims to improve the problem of missegmentation and missed segmentation in complex biomedical image segmentation. To verify the effectiveness and stability of this method, experiments were conducted on the Dsb2018Cell dataset (https://www.kaggle.com/c/data-science-bowl-2018) and the Luna dataset (https://www.kaggle.com/kmader/finding-lungs-in-ct-data), respectively.

Dsb2018Cell is a dataset for cell image segmentation, provided by the Data Science Bowl 2018 competition organized by Kaggle, to promote the application of computer vision and machine learning technologies in the field of biomedical image analysis. This dataset contains 576 original cell images of 256 × 256 pixels and their corresponding nuclei manually annotated with segmentation results. Luna is an important medical imaging dataset used for lung nodule detection and analysis, which includes manually segmented two-dimensional and 3D images of the lungs. This dataset is widely used in various medical image analysis competitions and research. In this paper, we only use 2-dimensional CT images, which contain 267 images and their corresponding labeled images, and we uniformly resize their original images to 256×256 pixels.

All experiments were performed on the same computer with a 64-bit Windows 10 operating system, Intel(R) Core(TM) i9-10900X CPU @ 3. 70 GHz 3. 70 GHz), NVIDIA GeForce RTX 2080Ti graphics card, and scientific computing libraries Python3. 7, PyTorch1. 7. 0, CUDA11.1. Github repository link: https://github.com/mmywcbysj/Biomedical-Image-Segmentation-Code.

3.2. Experimental procedure and evaluation index

Four common segmentation evaluation metrics, Iou (Intersection of Union), Dice (Dice Coefficient), Hd (Hausdorff distance) and Loss, were used to evaluate the segmentation results.

Iou is a common index of image segmentation, which indicates the similarity between the area of the segmented object and the original object. The value range is [0,1]; the larger the value means the segmentation result is closer to the real result and the better the segmentation effect. The formula is

$$Iou(A, B) = \frac{A \bigcap B}{A \bigcup B} \tag{3.1}$$

where *A* represents the area of the prediction result of the model and *B* represents the area of the manual labeling result.

The Dice measures the similarity index of two sets and the value range is [0,1]. The higher value means the better segmentation result, and the calculation formula is

$$Dice = \frac{2|A \bigcap B|}{|A| + |B|} \tag{3.2}$$

where A represents the set of samples segmented by the model in this paper and B represents the set of manually labeled samples.

Hd is a measure of the similarity between two sets of points, which represents the maximum value of the shortest distance between the segmentation result and the labeled result. The smaller the value, the smaller the image segmentation error and the better the quality.

$$H(A, B) = max(h(A, B), h(B, A)) \tag{3.3}$$

where $h(A, B) = \max_{a \in A} \left\{ \min_{b \in B} ||a - b|| \right\}$, $h(A, B)$, and $h(B, A)$ are the one way Hds from set A to set B and set B to set A, respectively. $h(A, B)$ first ranks the distance between each point a_i in set A to the nearest point b_j in set B, and finally takes the maximum value of this distance as $h(A, B)$.

The cross entropy loss function BCELoss (Binary CrossEntropy Loss) can be used not only for binary classification but also for multiclassification, and it has good results in multiclassification image segmentation problems. The smaller the loss value is, the closer the model segmentation result is to the real labeling result. The calculation formula is

$$BCELoss = -\frac{1}{N} \sum_{i=1}^{N} [y_i log(p_i) + (1 - y_i) log(1 - p_i)] \tag{3.4}$$

where N is the total number of biomedical image samples, y_i is the category to which the i sample belongs, and p_i is the predicted value of the i sample.

3.3. Analysis of experimental results

3.3.1. Ablation experiments

In order to verify the effectiveness of several modules in this paper for biomedical image segmentation, we conducted ablation experiments based on U-Net network with dense atrous convolution, multi-scale pooling module, and attention mechanism. The results are shown in Table 2.

Table 2. Results of ablation experiment.

Method	Iou	Dice	Hd	Loss
U-Net	0.8182	0.8863	4.3902	2.9444
U-Net+Attn	0.8144	0.8845	4.3800	4.2741
U-Net+ Attn+ DAC	0.8335	0.9024	4.3271	1.0289
U-Net+ Attn +MP	0.8287	0.9001	4.3657	1.1962
U-Net+ DAC+ MP+ Attn	0.9389	0.9629	4.5293	0.4640

Note: Bolded font is the optimal value of each column, Attn stands for attention mechanism, DAC stands for dense atrous convolution, and MP stands for dense residual pooling.

From the comparison results, we can see that when the U-Net network incorporates both dense atrous convolution modules and dense residual pooling modules, the Iou and Dice metrics have significantly improved, with values of 0.9389 and 0.9629, respectively. Compared with the U-Net network and the U-Net fusion attention mechanism network, the Iou metric has increased by 0.1207 and 0.1245, and the Dice metric has increased by 0.0766 and 0.0784, respectively. Compared with the U-Net network that incorporates an attention mechanism and dense atrous convolutional module, the Iou metric has improved by 0.1054 and the Dice metric has improved by 0.0766; and compared with the U-Net network that integrates the attention mechanism and dense residual pooling module, the IoU index has increased by 0.1102 and the Dice index has increased by 0.0628. This is because the dense atrous convolution module extracts multi-scale image features through multi-channel and multi-scale convolution kernels, enhancing the network's ability to extract image features. The dense residual pooling effectively utilizes the extracted image features through multiple residual pooling kernels. After introducing dense atrous convolution and dense residual pooling, attention mechanisms are added to increase target weights and reduce background weights to enhance target features, thereby improving the segmentation performance of the network. At the same time, the method proposed in this paper also performs the best on the Loss index, with no significant difference in the Hd index.

3.3.2. Comparison with other algorithms

In order to be more objective about the segmentation effect of different algorithms, the proposed network is compared with U-Net [6], R2U-Net [7], AtU-Net [29], and AtFcn [10]. All experiments were conducted in the same experimental environment, and the average of each index of 256 images in the Dsb2018Cell dataset was selected. The average segmentation data of each model training is shown in Table 3, and the results of quantitative analysis of the comparison experiments are shown in Figure 7.

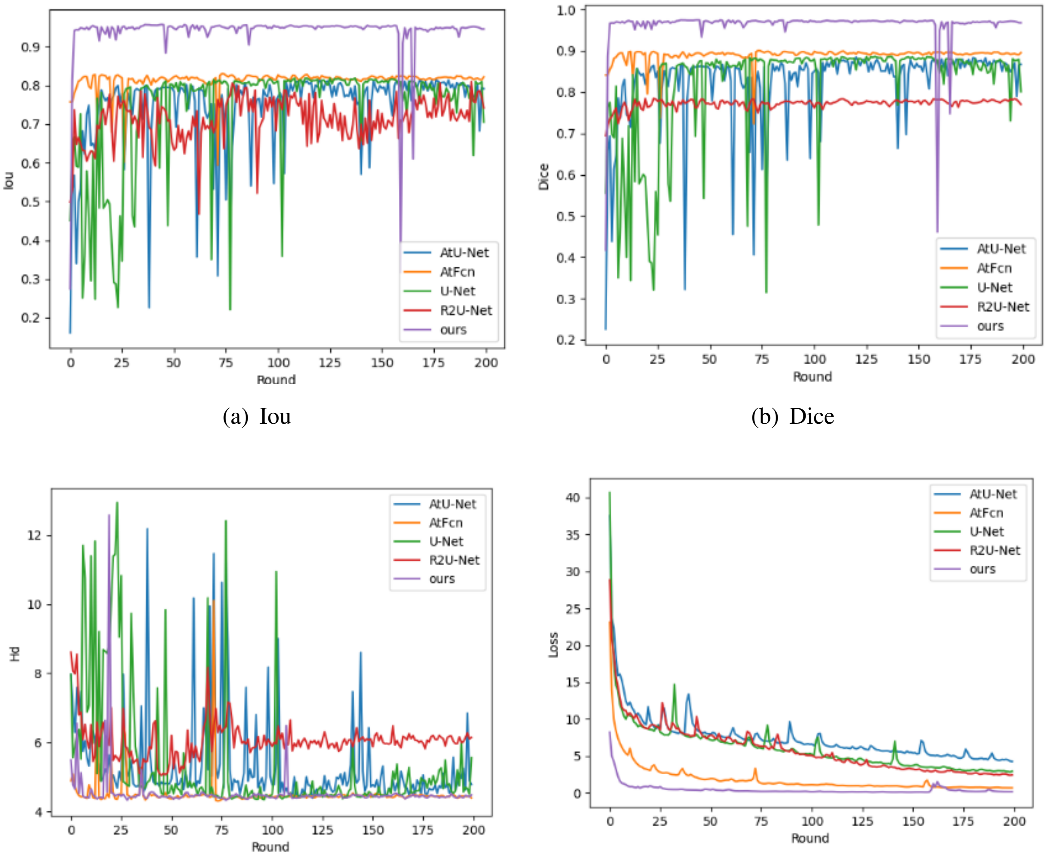
Table 3. Comparative experimental results under Dsb2018Cell dataset.

Method	Iou	Dice	Hd	Loss
U-Net	0.8182	0.8863	4.3902	2.9444
R2U-Net	0.7076	0.7999	5.2819	2.4472
AtU-Net	0.8144	0.8845	4.3800	4.2741
AtFcn	0.8293	0.8998	4.3155	0.7269
ours	0.9389	0.9629	4.5293	0.4640

Note: Bolded font is the best value of each column, all indicators are kept in four valid digits.

In the comparison experiment of Table 3, Our method performed the best on all other indicators except for Hd, which was slightly inferior. Compared with AtFcn, Iou and Dice increased by 0.1096 and 0.0631, respectively, and Loss decreased by 0.2629. Compared with AtU-Net, Iou and Dice increased by 0.1245 and 0.0784, respectively, and Loss decreased by 3.8101. Compared with R2U-Net, Iou and Dice increased by 0.2313 and 0.163, respectively, and Loss decreased by 1.9832. Compared with U-Net network, Iou and Dice have increased by 0.1207 and 0.0766, respectively, and Loss has decreased by 2.4804.

Both the Iou and Dice metrics compare the segmented images and labels in terms of segmentation



(a) Iou

(b) Dice

(c) Hd

(d) Loss

Figure 7. Graph of quantitative analysis results of different models.

effect, and their values can indicate the percentage of correctly segmented regions in the whole image area. The first and second columns of Table 3 show that the results of this paper are significantly better than other algorithmic models, indicating that the dense atrous convolution module and dense residual pooling module in the method proposed in this paper can effectively capture multi-scale features in images, thereby improving the accuracy of image segmentation. The Hd coefficient indicates the similarity relationship between point sets, and the lower value indicates the higher similarity. As shown in the third column of Table 3, there is no big difference between this paper and the AtFcn model with the best Hd coefficient. As shown in the fourth column of Table 3 and Figure 7(d), all models converge with the training iterations, but the method in this paper can achieve fast convergence and is optimal. This indicates that the attention mechanism improves the training efficiency by highlighting the target features.

The issue in nucleus segmentation is that the nuclei are too small and the spacing between nuclei is too small for the segmentation to cause adhesions. As shown in Figure 8, only this method and the AtFcn model can avoid cell adhesion during the segmentation process.

The U-Net and AtFcn models avoid the phenomenon of segmentation adhesion, improve the segmentation accuracy, and reduce the probability of false segmentation. Although both AtU-Net and R2U-Net have jump connections, they are not suitable for segmenting small targets such as cell nuclei and more pixels are mis-segmented. Compared with other segmentation models, the segmentation results of this model do not show cell adhesion, which is closer to the label of expert segmentation. In the second and fourth rows of Figure 8, this model can segment more correct nucleus regions without mis-segmentation at the cell and boundary. It is shown that dense atrous convolution and dense residual pooling can effectively extract image feature information and accurately identify boundary features when segmenting, thus achieving more accurate segmentation.

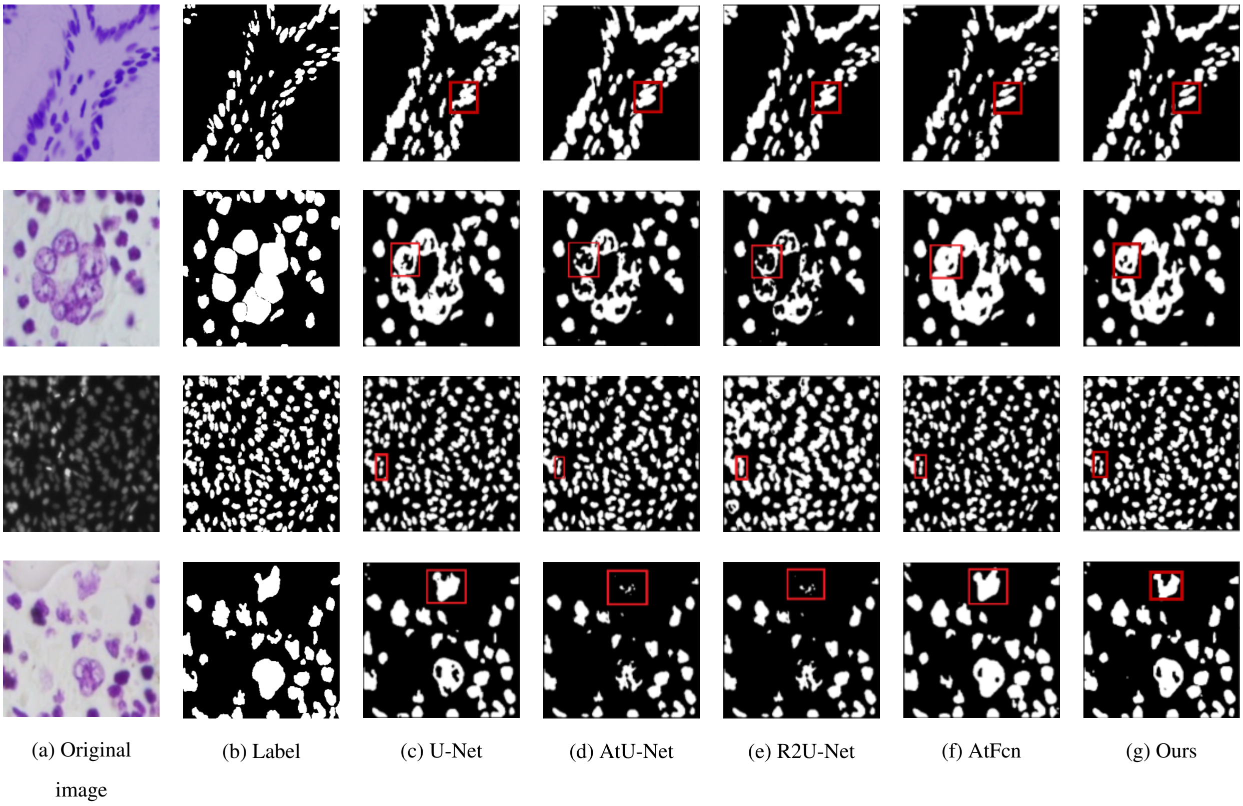


Figure 8. Comparison effect of cell nucleus image.

In order to verify the effectiveness of the algorithm in this paper, we continue the validation on the Luna lung dataset, and the comparison test is shown in Figure 9. Compared with the Dsb2018Cell nucleus dataset, the image complexity of the Luna lung dataset is lower, and the different models can achieve the segmentation purpose for the same images compared with the segmented labeled images. The first row of Figure 9 has significantly lower image complexity and clearer image boundaries, so the overall segmentation results are good. The third column of the second row of Figure 9 shows the segmentation results of the U-Net model, where the left lung is well segmented, but there is large missegmentation in the upper part of the right lung. It is worth noting that this method can extract the information of different scales in the image, so that the segmented image boundary at the boundary is clearer and missegmentation is avoided. In the fourth and fifth columns of Figure 9, the left and right lungs are well segmented, but there is missegmentation in the lower part of the right lung, indicating that the background features and the target features are similar, so we add an attention mechanism in the decoding stage to reduce the weight of the background features to highlight the target features, ultimately achieving more accurate segmentation.

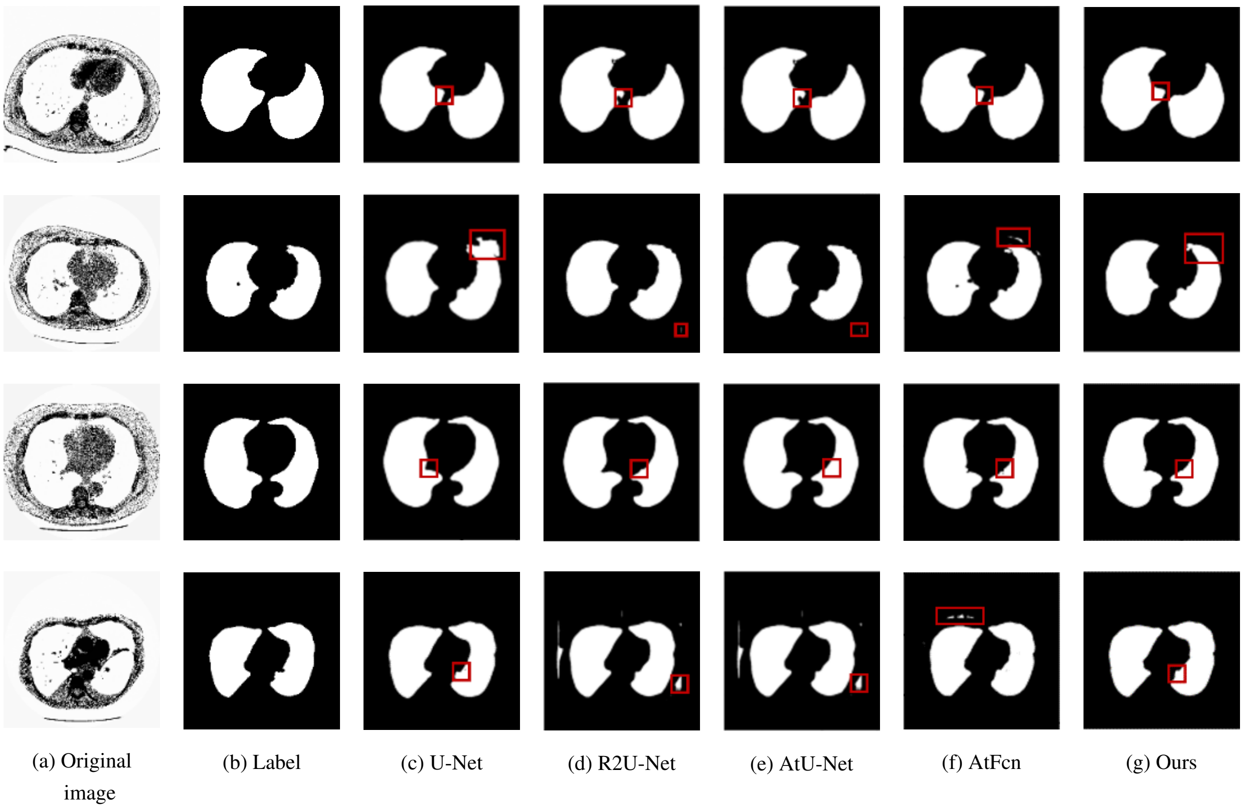


Figure 9. Comparison results of lung images.

4. Dicussion

In recent years, many advanced methods have emerged in the field of medical image segmentation. ARP-Net is a novel OCTA retinal vessel segmentation method based on the Adaptive gated axial transformer (AGAT), Residual and Point repair modules [13]. It has achieved good performance in various indicators, such as Dice of 0.9513, BACC (Balance Accuracy) of 0.9781, and JAC (Jaccard Index) of 0.9126 on the OCTA-3M dataset. ARU-Net is an attention-based multi-scale supervised fully convolutional encoder-decoder network that achieves segmentation tasks for intracranial aneurysms and adjacent arteries in 3DRA images [14]. The network has also achieved good performance on multiple indicators, such as SE (Sensitivity) of 0.8533, SP (Specificity) of 0.9978, and DSC (Dice Similarity Coefficient) of 0.8681. ARP-Net and ARU-Net both adopt encoding and decoding structures, and both focus on extracting multi-scale information from images. Unlike the method proposed in this paper, the encoder of ARP-Net exchanges a large amount of global and local information through feature interaction units, thereby preserving a large amount of detailed information. ARU-Net follows the classic U-Net framework and effectively emphasizes important targets on 3DRA images through depth-aware attention gates, thereby improving the accurate segmentation of small but critical vessels. At the same time, the ability of the network to integrate multilevel spatial and semantic information is improved through multi-scale supervision strategies to enhance the sensitivity of smaller aneurysms and vessels.

Our method is based on the Inception structure and atrous convolution to design a DAC module, which is added between the encoding and decoding paths of the U-Net. The DAC module extracts multi-scale image features through multi-channel and multi-scale convolution kernels, enhancing the network's ability to extract image features. Atrous convolution automatically enlarges the image field by extracting sparse features and inserting cavities into the convolution kernel to form atrous convolution kernels with different expansion rates to obtain different sizes of the image field. The size of the feature map can be kept constant while expanding the field so that the image accuracy will not be degraded during the feature extraction. At the same time, fully utilizing the extracted image features plays a crucial role in segmentation work. Therefore, this paper designs a dense residual pooling module, which, combined with dense atrous convolution, can achieve the goal of fully extracting and utilizing multi-scale image features. Our method also introduces attention mechanism in the decoding stage, highlighting the target by increasing the weight of the target area while suppressing the influence of irrelevant areas on segmentation, thereby improving segmentation accuracy and reducing the occurrence of missegmentation and missed segmentation. Although the above methods are slightly different from the algorithm proposed in this paper in the field of medical applications, the network structure design of different methods has brought great inspiration to our future research.

Customizing segmentation methods for specific tasks can lead to fragmentation between various segmentation tasks, so there have been many recent algorithms dedicated to improving the consistency of image segmentation networks. The FreeSeg model [15] is a unified, universal, and open framework for local image segmentation, which has good robustness in multiple tasks and different scenes. For example, it achieves 20.6% mAP (mean Average Precision) of unseen classes on COCO (a dataset that can be used for image recognition) and achieves 16.3%/15.4% mAP of seen/unseen classes on ADE20k (large scale datasets for scene analysis). SAM [16] is a prompt-based model with strong zero sample generalization ability, which breaks through segmentation boundaries and greatly promotes the development of basic models. Although the SAM model has strong zero sample generalization ability, its application in industrial production is limited due to its high computational cost. Subsequently, in order to address the issues of SAM, the FastSAM model was proposed [16]. Compared to the SAM model, FastSAM achieves performance comparable to SAM while running at a speed (32 × 32) 50 times faster than SAM (64 × 64) 170 times faster. Good operating speed makes FastSAM the choice for industrial applications.

In the future, we will conduct detailed comparative analysis with various advanced segmentation models, aiming to design more lightweight models and improve the universality of our models in other image segmentation tasks.

5. Conclusion

In order to efficiently process biomedical images for more accurate computer-aided diagnosis, this paper proposes a biomedical image segmentation network based on dense atrous convolution, which consists of an encoding path and a decoding path. In the middle of the encoding and decoding paths, the dense atrous convolution is proposed by combining Inception structure and atrous convolution, which can expand the size and number of perceptual fields and extract the multi-scale information in the images. By introducing dense residual pooling, the extracted effective information can be more fully utilized. The feature map size is recovered by upsampling in the decoding stage, and attention gates are introduced in the decoding stage to enhance the target features by reducing the background weights. Experiments on the nucleus and lung datasets show that this method is closer to the real segmentation results, reduces the occurrence of missegmentation, and improves the segmentation accuracy. It can provide more accurate biomedical image data for computer-aided diagnosis and treatment. However, although our model has achieved good segmentation results and reduced the occurrence of missegmentation and missed segmentation, the results show that our method still needs further improvement when segmenting finer structures. In addition, our method faces certain challenges in balancing image segmentation accuracy and model complexity, making it difficult to achieve the lightweight design of the model while maintaining high segmentation accuracy. In the future, our research will mainly focus on reducing model complexity and improving model robustness. Meanwhile, we will consider improving the generalization ability of our method and applying it to images in other fields, such as remote sensing image segmentation.

Acknowledgments

This work was partly supported by the PhD pre-research Foundation of Shijiazhuang Posts and Telecommunications Technical College No.G99993073, and the Natural Science Basis Research Plan in Shaanxi Province of China under Grant 2023-JC-YB-517, and the Open Project Program of State Key Laboratory of Virtual Reality Technology and Systems Beihang University under Grant VR-LAB2023B08.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. L. Kong, Q. Wang, Y. Bao, H. Li, A survey on medical image segmentation based on deep leaning, *Radio Commun. Technol.*, **47** (2021), 121–130. https://doi.org/10.3969/j.issn.1003-3114.2021.02.001
2. J. Chen, L. Li, Automatic segmentation of fuzzy areas in ultrasonic images based on edge detection, *Autom. Instrument.*, **11** (2021), 19–22. https://doi.org/10.14016/j.cnki.1001-9227.2021.11.019
3. A. Aslam, E. Khan, M. M. S. Beg, Improved edge detection algorithm for brain tumor segmentation, *Proced. Computer Sci.*, **58** (2015), 430–437. https://doi.org/10.1016/j.procs.2015.08.057
4. M. Van Eijnatten, R. van Dijk, J. Dobbe, G. Streekstra, J. Koivisto, J. Wolff, CT image segmentation methods for bone used in medical additive manufacturing, *Med. Eng. Phys.*, **51** (2018), 6–16. https://doi.org/10.1016/j.medengphy.2017.10.008
5. J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2015), 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965

6. O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *International Conference on Medical image computing and computer assisted intervention*, (2015), 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

7. O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, et al., Attention u-net: Learning where to look for the pancreas, *arXiv preprint*, (2018), arXiv:1804.03999. https://doi.org/10.48550/arXiv.1804.03999

8. Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, J. Liang, Unet++: A nested u-net architecture for medical image segmentation, *Deep learning in medical image analysis and multimodal learning for clinical decision support*, (2018) 3–11. https://doi.org/10.1007/978-3-030-00889-5_1

9. M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, V. K. Asari, Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation, *arXiv preprint*, (2018), arXiv:1802.06955. https://doi.org/10.48550/arXiv.1802.06955

10. G. Huang, S. Liu, L. Van der Maaten, K. Q. Weinberger, Condensenet: An efficient densenet using learned group convolutions, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2018), 2752–2761. https://doi.org/10.1109/CVPR.2018.00291

11. F. Zhu, Z. Gao, C. Zhao, Z. Zhu, J. Tang, Y. Liu, et al., Semantic segmentation using deep learning to extract total extraocular muscles and optic nerve from orbital computed tomography images, *Optik*, **244** (2021), 167551. https://doi.org/10.1016/j.ijleo.2021.167551

12. H. Li, J. Fan, Q. Hua, X. Li, Z. Wen, M. Yang, Biomedical sensor image segmentation algorithm based on improved fully convolutional network, *Measurement*, **197** (2022), 111307. https://doi.org/10.1016/j.measurement.2022.111307

13. X. Liu, D. Zhang, J. Yao, J. Tang, Transformer and convolutional based dual branch network for retinal vessel segmentation in OCTA images, *Biomed. Signal Process. Control*, **83** (2023). https://doi.org/10.1016/j.bspc.2023.104604

14. N. Mu, Z. Lyu, M. Rezaeitaleshmahalleh, J. Tang, J. Jiang, An attention residual u-net with differential preprocessing and geometric postprocessing: Learning how to segment vasculature including intracranial aneurysms, *Med. Image Anal.*, **84** (2023), 102697. https://doi.org/10.1016/j.media.2022.102697

15. J. Qin, J. Wu, P. Yan, M. Li, R. Yuxi, X. Xiao, et al., FreeSeg: unified, universal and open-vocabulary image segmentation, in*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2023), 19446–19455. https://doi.org/10.1109/CVPR52729.2023.01863

16. T. Ma*, H. Zhao, X. Qin, A dehazing method for flight view images based on transformer and physical priori, *Math. Biosci. Eng.*, **20** (2023), 20727–20747. http://dx.doi.org/10.3934/mbe.2023917

17. A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, et al., Segment anything, *arXiv preprint*, (2023), arXiv:2304.02643. https://doi.org/10.48550/arXiv.2304.02643

18. X. Zhao, W. Ding, Y. An, Y. Du, T. Yu, M. Li, et al., Fast Segment Anything, *arXiv preprint*, (2023), arXiv:2306.12156. https://doi.org/10.48550/arXiv.2306.12156

19. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al., Going deeper with convolutions, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2015), 1–9. https://doi.org/10.1109/CVPR.2015.7298594
20. S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, in *International conference on machine learning*, (2015), 448–456. https://doi.org/10.48550/arXiv.1502.03167
21. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2016), 2818–2826. https://doi.org/10.1109/CVPR.2016.308
22. C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in *Proceedings of the AAAI conference on artificial intelligence*, **31** (2017). https://doi.org/10.48550/arXiv.1602.07261
23. T. Ma, C. Fu, J. Yang, J. Zhang, C. Shang, RF-Net: unsupervised low-light image enhancement based on retinex and exposure fusion, *Comput. Mater. Continua*, **77** (2023), 1103–1122. https://doi.org/10.32604/cmc.2023.042416
24. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al., Going deeper with convolutions, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2015), 1–9. https://doi.org/10.1109/CVPR.2015.7298594
25. Y. Wei, H. Xiao, H. Shi, Z. Jie, J. Feng, T. S. Huang, Revisiting dilated convolution: A simple approach for weakly and semi supervised semantic segmentation, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2018), 7268–7277. https://doi.org/10.1109/CVPR.2018.00759
26. D. M. Vo, S. W. Lee, Semantic image segmentation using fully convolutional neural networks with multi-scale images and multi-scale dilated convolutions, *Multimedia Tools Appl.*, **77** (2018), 18689–18707. https://doi.org/10.1007/s11042-018-5653-x
27. H. Li, Q. Zheng, W. Yan, R. Tao, X. Qi, Z. Wen, Image super-resolution reconstruction for secure data transmission in Internet of Things environment, *Math. Biosci. Eng.*, **18** (2021), 6652–6672. https://doi.org/10.3934/mbe.2021330
28. H. Cheng, J. Lu, M. Luo, W. Liu, K. Zhang, PTANet: Triple attention network for point cloud semantic segmentation, *Eng. Appl. Artif. Intell.*, **102** (2021), 104239. https://doi.org/10.1016/j.engappai.2021.104239
29. Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, et al., Ce-net: Context encoder network for 2d medical image segmentation, *IEEE Transact. Med. Imag.*, **38** (2019), 2281–2292. https://doi.org/10.1109/TMI.2019.2903562