SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

September 2022, Vol. 65 192102:1–192102:12
https://doi.org/10.1007/s11432-020-3079-8

Difficulty-aware bi-network with spatial attention constrained graph for axillary lymph node segmentation

Qing XU[1], Xiaoming XI[2*], Xianjing MENG[3], Zheyun QIN[1], Xiushan NIE[2], Yongjian WU[1], Dongsheng ZHOU[4], Yi QU[5], Chenglong LI[2] & Yilong YIN[1*]

[1] *School of Software, Shandong University, Jinan* 250101*, China;*
[2] *School of Computer Science and Technology, Shandong Jianzhu University, Jinan* 250101*, China;*
[3] *School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan* 250014*, China;*
[4] *Shandong Provinvial Qianfoshan Hospital, The First Hospital Affiliated with Shandong First Medical University, Jinan* 250014*, China;*
[5] *Department of Geriatrics, Qilu Hospital of Shandong University, Jinan* 250012*, China*

Received 13 June 2020/Revised 21 August 2020/Accepted 29 September 2020/Published online 7 January 2022

Abstract Axillary lymph node (ALN) segmentation in ultrasound images is important for the diagnosis and treatment of breast cancer. Recently, deep learning methods for automatic medical image segmentation have improved significantly. However, two problems arise. (1) A unified model is often employed to segment all images without considering the difficulty diversity. (2) The relationship between elements in the learned class probability map is disregarded. To address these two issues, we propose a novel difficulty-aware bi-network with a spatial attention constrained graph. First, a difficulty grading module (DGM) is developed to learn the difficulty grade of input images. Based on the difficulty grade of images, a novel bi-network architecture is proposed to segment the image adaptively using different branches. In complex branches, a novel spatial attention module (SAM) and graph-based energy with spatial attention constraint are proposed. The learned spatial attention map can provide additional discriminative information. Moreover, the graph-based segmentation framework can capture the relationship between pixels, further improving the segmentation performance for complex images. We conducted an experiment on our ultrasound database using 216 cases. The overall dice similarity coefficient, Jaccard coefficient, volumetric overlap error, and false positive rate are 83.41%, 74.4%, 12.02%, and 13.36% for ALN segmentation, respectively. The comparison results demonstrated that the proposed method outperforms other deep learning methods.

Keywords ultrasound image, axillary lymph nodes segmentation, difficulty-aware segmentation, graph with spatial attention

Citation Xu Q, Xi X M, Meng X J, et al. Difficulty-aware bi-network with spatial attention constrained graph for axillary lymph node segmentation. Sci China Inf Sci, 2022, 65(7): 192102, https://doi.org/10.1007/s11432-020-3079-8

1 Introduction

Lymph nodes (LNs) are central to the immune system as they can filter fluids as well as capture virus and bacteria. When LNs appear abnormal, it often denotes pathological changes within their attributive areas [1]; this abnormality is useful for disease diagnosis [2]. In addition, the evaluation of axillary lymph node (ALN) status (i.e., metastatic or non-metastatic) in patients with breast cancer provides useful and reliable information for surgical and postsurgical management [3]. Therefore, the evaluation of ALNs is meaningful for the diagnosis and treatment of breast cancer.

Ultrasound is a widely used imaging modality for evaluating ALNs because it is non-invasive and typically available in hospitals [4, 5]. However, the interpretation of ultrasound images is dependent on radiologists' subjective expertise, resulting in inter-/intra-observer variations. Hence, an automatic computer-aided diagnosis (CAD) system must be developed.

[*] Corresponding author (email: fyzq10@126.com, ylyin@sdu.edu.cn)

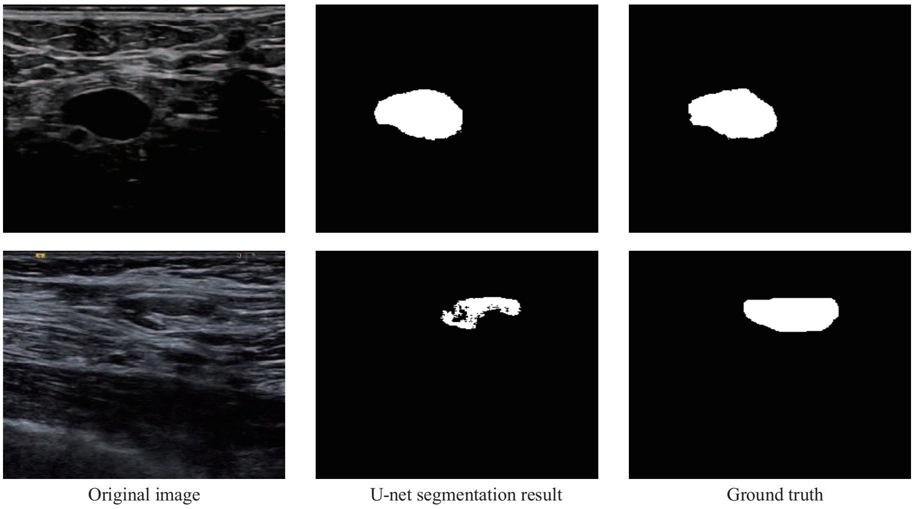


Figure 1 Two segmentation samples of U-net.

Automatic segmentation of ALNs is the foundation of CAD systems. It is the basis of many clinical applications such as tissue characterization and medical diagnosis [6]. This work aims to develop an effective segmentation method to automatically delineate ALN from surrounding tissues on Ultrasound images. By segmenting ALN accurately, doctors are able to acquire the properties of ALN, including the size, edge, and internal echo. These properties are useful for identification of ALN. Moreover, accurate identification of ALN involvement in patients with breast cancer is important for prognosis and therapy decisions [7].

Recently, deep convolutional networks such as fully convolutional networks (FCNs) [8] and U-net [9] have been widely used for image segmentation. An encoder is used to learn the discriminative features, whereas a decoder is used to obtain accurate pixel-level classification results.

However, two problems arise when using these deep learning-based segmentation methods for ALN segmentation. (1) All images are treated equally without considering the difficulty diversity; the network also utilizes a unified network to segment all images. Generally, for some complex images, irregular distribution may be caused by the different abnormal degrees of lesion or imaging factors, such as different devices or imaging protocols [10]. In these images, complicated intensity distributions often occur, resulting in a large intraclass variance. Figure 1 shows two segmentation examples obtained by U-net. As shown in Figure 1, U-net can achieve an accurate segmentation result for the first image because it has a uniform intensity distribution. By contrast, U-net fails to segment the second image owing to the complicated intensity distribution. (2) The relationship between elements in the class probability map is disregarded, resulting in limited performance improvement. Traditional deep learning methods learn a class probability map and label each pixel using the class having the maximum probability. However, they disregard relationship between the pixels, thereby resulting in increased sensitivity to noise and increased error rate.

To overcome the two aforementioned issues, in this study, we propose a novel difficulty-aware bi-network with spatial attention constrained graph for ALN segmentation. The proposed method primarily comprises two components: (1) Difficulty grading module (DGM). The image difficulty recognition network is developed to learn the difficulty grade of the images, i.e., simple or complex. In this study, complex images have complicated characteristics such as complicated intensity distribution and texture. (2) Segmentation network. It comprises two functionality-specialized branches: an simple branch and a complex branch. In the simple branch, U-net is used as the basic backbone of the network. To learn more effective features, the novel squeeze-and-excitation (SE)-U-net is developed by incorporating an SE block [11]. In the complex branch, the SE-U-net is first used for segmentation. Subsequently, a spatial attention module (SAM), which contains a local smoother and global spatial map learner, is proposed. It can generate a spatial attention map. The final segmentation result is obtained by minimizing a novel graph-based energy by incorporating the spatial attention map. A novel spatial constrain term is proposed to guarantee that the segmentation result is consistent with the spatial attention map. The spatial attention map provides discriminative information that is robust to intensity variation. Moreover, the graph-based segmentation framework can capture the relationship between pixels, further improving the segmentation performance of complex images. Experimental results on our database demonstrate that the proposed method outperform state-of-the-art methods. The contribution of this study is as follows:

(1) A novel difficulty-aware bi-network with spatial attention constrained graph is proposed to segment ALNs in ultrasound images. It primarily comprises a DGM and a segmentation network.

(2) A DGM is developed to guarantee that an image can be segmented adaptively using different models, thereby improving the efficiency of the segmentation model.

(3) In the proposed segmentation network, a complex branch is developed to focus on complex images by introducing a novel spatial attention constrained graph framework. An SAM is proposed to generate a spatial attention map that is robust to intensity variation. In addition, the graph framework is utilized to capture the relationship between pixels, further improving the segmentation performance of complex images.

2 Related work

Considering that graphs can capture effective relationships between various parameters, graph-based methods are widely used for LN segmentation. Detabs et al. [12] combined region growing and graph cut to segment pelvic LNs. Zhy et al. [13] proposed a method based on texture feature and graph cut for LN segmentation in ultrasound images. To exploit structure information, Kuo et al. [14] proposed a novel nested-graph-cut method to segment LNs in three-dimensional (3D) ultrasound images. Zhang et al. [1] proposed a modified graph cut with the incorporation of an elliptical shape constraint to segment cervical LNs on sonograms. To solve the problem of spatially varying distributions of LN parenchyma and fat, Kuo et al. [15] developed a novel graph framework by introducing locally adaptive energy. In addition, other traditional segmentation methods such as snake model and region growing have been used for LN segmentation. To exploit effective edge information, Zhang et al. [16] proposed an improved gradient vector flow snake model combined with the edge flow method. Furthermore, region growing was used for LN segmentation [17]. Zhang et al. [18] proposed a multiscale fuzzy c-means method integrated with particle swarm optimization for ultrasound image segmentation. Considering the elliptical prior of ALNs, Meinel et al. [19] developed an elliptical model to segment ALNs in ultrasound image data.

Compared with traditional methods, deep learning methods can be used to learn more effective features, resulting in significant performance improvement. Long et al. [8] proposed an FCN exhibiting a typical encoder-decoder network architecture. The encoder was used to extract discriminative features, while the decoder was used to obtain an output that is of the same size as the input. In addition, a skip connection was developed to fuse low-level and high-level features to further improve the segmentation performance. Using the FCN architecture, Zhang et al. [2] proposed coarse-to-fine stacked fully convolutional nets for LN segmentation in ultrasound images. To improve segmentation efficiency, Segnet was proposed, which adopted an encoder-decoder network structure that introduced pooling indices [20]. To further improve the effectiveness of feature learning, Deeplab series were proposed. Deeplabv1 introduced atrous convolution and conditional random field to learn more effective contextual information and multiscale information [21]. To learn more effective multiscale information, the variance of atrous spatial pyramid pooling (ASPP) were developed in Deeplabv2 and Deeplabv3 [22, 23]. The latest Deeplabv3+, with has an additional decode module compared with its previous version, achieved the best performance [24]. As the typical encoder-decoder network for medical image segmentation, U-net [9] and its variant [25–28] were proposed. Unlike the FCN, in U-net, the concatenation operation of the feature map was introduced in the skip connection.

Deep learning methods have been applied to ALN status prediction [7] and thyroid nodules classification [29]. To predict ALN status, Zheng et al. [7] first proposed a parallel pretrained ResNet model to learn features that can be combined with clinical parameters. Subsequently, they classified the features using an support vector machine (SVM) model. Wang et al. [29] proposed pretraining a VGG16 model via fine-tuning for thyroid nodule classification.

3 Method

Figure 2 shows the proposed framework. It primarily comprises two components: (1) DGM. It is used to learn the difficulty grade of the input images, i.e., simple or complex. (2) Segmentation network. It

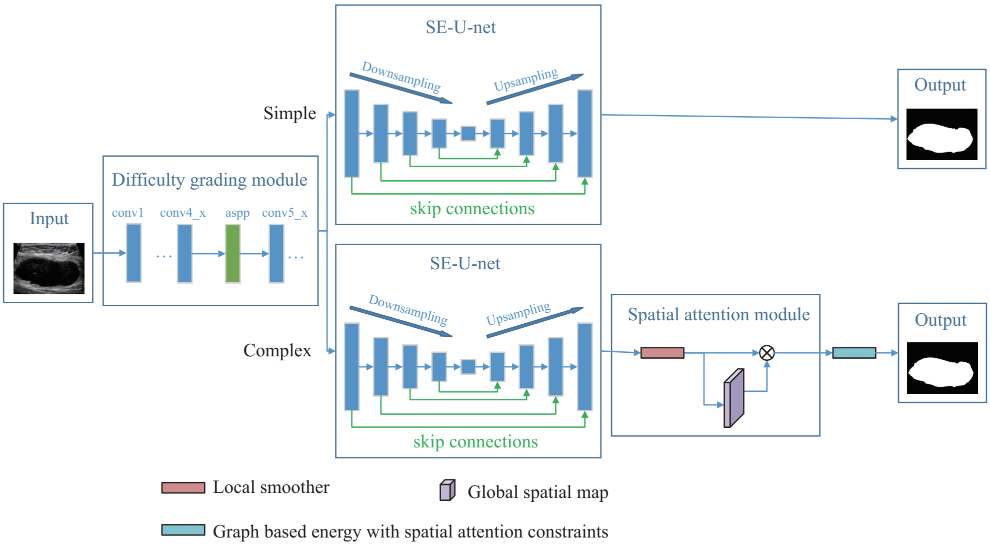
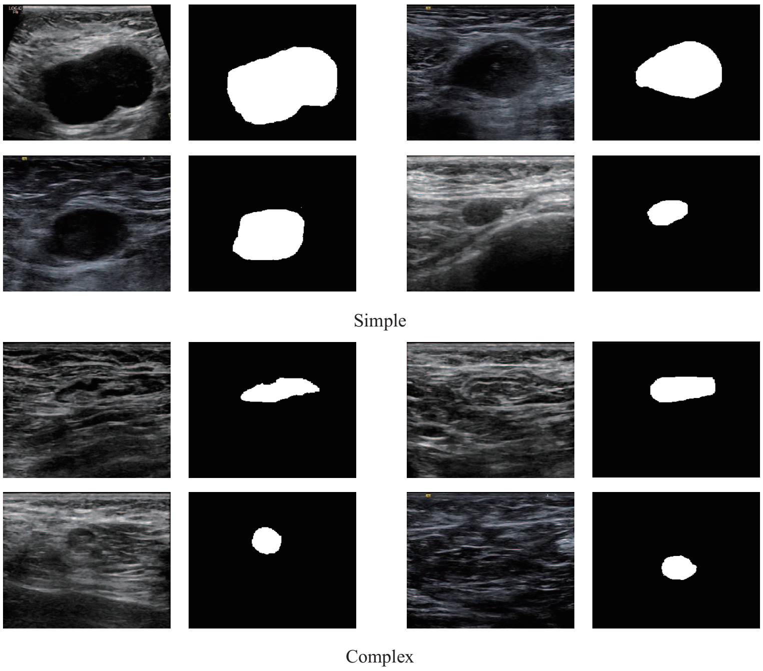


Figure 2 (Color online) The framework of the proposed method.


Simple

Complex

Figure 3 The examples of hard image and easy image.

comprises two functionality-specialized branches: an simple branch and a complex branch. In the simple branch, SE-U-net is developed for simple image segmentation. In the complex branch, SE-U-net is first used for segmentation. Subsequently, an SAM that contains a local smoother and a global spatial map learner is proposed to generate a spatial attention map. The final segmentation result is obtained by minimizing a novel graph-based energy by incorporating the spatial attention map.

3.1 DGM

A DGM is used to generate the difficulty grade of images, i.e., complex or simple. In this study, the DGM generates results based on the image complexity. We assume that complicated characteristics occur in complex images. As shown in Figure 3, simple and complex images exhibit uniform and complicated intensity distributions, respectively.

To exploit the prior, we modeled the problem of difficulty grading as a classification problem and

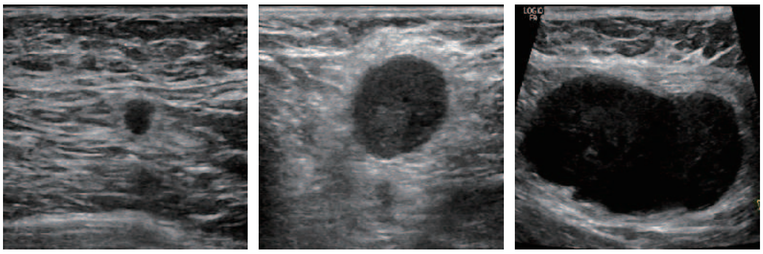


Figure 4 ALN with different sizes.

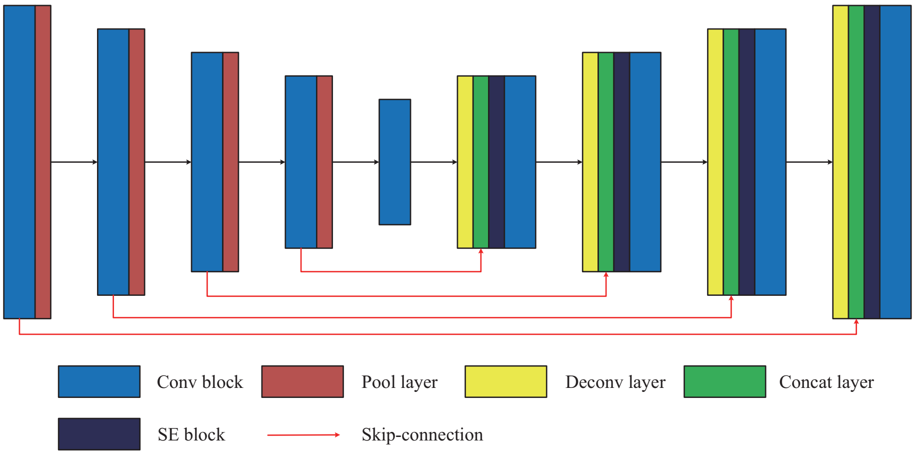


Figure 5 (Color online) The architecture of SE-U-net.

developed a difficulty recognition network. As a typical object recognition network, Resnet [30] was used as the basic backbone of the DGM. To capture more context information of ALNs with different sizes (shown in Figure 4), an ASPP-Resnet was developed by introducing ASPP [23].

The architecture of the ASPP-Resnet is based on Resnet50. ASPP is appended, followed by conv4_x. ASPP comprises multiple parallel filters with different rates that can fuse multiscale context information more effectively. In the proposed ASPP-Resnet, the ASPP module comprises a 1×1 convolution and three 3×3 convolutions with rates 6, 12, and 18.

Training data were collected to train the DGM. The label represents the difficulty grade of the training data. To generate a difficult grade groundtruth for each training image, U-net was first used to segment the images. Subsequently, the Jaccard coefficient (JACCARD) of the segmented image was calculated and compared with a specific threshold. If the JACCARD is larger than the threshold, then the groundtruth of the image is simple; otherwise, it is complex.

3.2 SE-U-net

For simple images, the segmentation network was constructed using the traditional encoder-decoder network architecture. As a typical medical image segmentation network, U-net was used as the basic backbone of the network.

Considering that the SE block [11] can improve the representational power of the learned feature, SE-U-net was developed by introducing an SE block in U-net (shown in Figure 5). The SE block was appended, followed by the concatenation layer.

The SE blocks mainly contains the SE operation (shown in Figure 6). In the squeeze operation, a $1{\times}1{\times}C$ global feature is generated using global average pooling for c feature maps. In the excitation operation, the $1{\times}1{\times}C$ global feature is transformed into a channel weight matrix using the gating mechanism. This weight matrix can capture channel-wise dependencies, thereby increasing the sensitivity to informative features, which is beneficial for improving the segmentation performance.

For simple images, SE-U-net can obtain accurate segmentation results. However, it is difficult to achieve satisfactory performances for complex images owing to their complicated characteristics. To improve the segmentation performance of complex images, in the complex branch, the generated class probability map is used as the prior to mine more effective information.

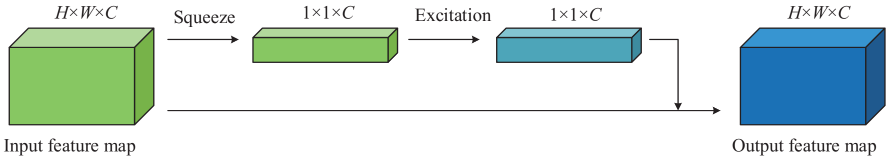


Figure 6 (Color online) Sequeeze-and-excitation block.

3.3 SAM

A novel SAM was developed in this study to generate a spatial attention map. In the proposed module, a local smooth filter was developed to ensure the smoothness of the prior probability map. Subsequently, a global spatial map was generated based on the developed center importance-based probability function. Finally, a spatial attention map was generated by combining the prior probability and global spatial maps.

3.3.1 *Local smooth filter*

To reduce noise in the class probability map generated via SE-U-net, a local smooth filter was developed. In fact, images have strong local correlations among spatially adjacent pixels. In other words, neighboring pixels exhibit similar characteristics.

Hence, the local smooth filter was constructed by combining the probability of neighboring pixels in a local region. The output for an arbitrary pixel (i, j) is calculated as follows:

$$g(i,j) = \frac{\sum_{(x,y)\in N_{ij}} (f(x,y) \times C(r-i+x, r-j+y))}{\sum_{(x,y)\in N_{ij}} C(r-i+x, r-j+y)}, \tag{1}$$

$$N_{ij} = \{(x,y) \mid i-r < x < i+r,\ j-r < y < j+r\}, \tag{2}$$

where f denotes the prior probability map generated via SE-U-net, C the weight matrix of the filter, and N_{ij} the local region whose center is pixel (i, j). r is the radius of region N.

3.3.2 *Global spatial map learner*

Generally, for pixels in a segmented ALN, the edge pixel has larger uncertain values than the central pixels. In other words, the probability that the central pixel belongs to the ALN should be higher.

Hence, a center importance-based probability function was developed to capture the global spatial distribution of the object. The global spatial probability of an arbitrary pixel (i, j) is calculated as follows:

$$p_{ij} = 1 - \left(\frac{2 \times s_{ij}^2}{L^2} + \frac{2 \times l_{ij}^2}{S^2}\right). \tag{3}$$

In the aforementioned equation, L and S represent the long and short axes of the segmented object obtained via SE-U-net, respectively. l_{ij} and s_{ij} represent the distance between pixel (i, j) and the long and short axes, respectively. As shown in the equation, we can infer that the pixels near the central pixels have a higher probability of belonging to an ALN. The global spatial map can be generated by obtaining the spatial probability of each pixel.

The spatial attention map is calculated as follows:

$$h(i,j) = \begin{cases} p_{ij} \times g(i,j), & (i,j) \in A, \\ g(i,j), & (i,j) \notin A. \end{cases} \tag{4}$$

In the aforementioned equation, A is the ellipse area around the center of the ALN via SE-U-net.

The spatial attention map was generated based on the spatial prior information, which is independent of the intensity distribution. The pixels in the ALN can be assigned with high attention, which is helpful for avoiding missing ALN structures in the segmentation result. Therefore, the spatial attention map can provide additional discriminative information, which is beneficial for performance improvement.

3.3.3 *Graph with spatial constraint-based segmentation*

Considering that graphs can effectively capture the relationship between nodes, a graph framework was introduced in the complex branch. Let $G(V, E)$ be a graph, where V and E represent the sets of nodes and edges, respectively. In the graph, elements in the prior probability map are regarded as the nodes of the graph, whereas the similarity between them are regarded as the edges of the graph.

To embed the graph into the segmentation process, we transformed the image segmentation problem into the minimum cut problem of the graph [31, 32]. In the general framework of graph cut, additional terminal nodes are introduced to be connected to all the common nodes via t-links. In addition, the edges between the common nodes are regarded as n-links.

The energy function for the graph cut is written as follows:

$$E(L) = \lambda R(L) + B(L). \tag{5}$$

In this function, $R(L)$ represents the region term, $B(L)$ the boundary term, and λ the tradeoff parameter. The region term energy corresponds to the sum of weights of t-links contained in the cut L; the boundary term energy corresponds to the sum of weights of n-links contained in the cut L. The region term is used such that the predicted label is consistent with the groundtruth, and the boundary term is used to smooth the boundary.

The energy $R(L)$ can be rewritten as follows:

$$R(L) = \sum_{p \in P} G(p), \tag{6}$$

$$G(p) = \begin{cases} -\ln f(p), & L_p = \text{“obj”}, \\ -\ln(1 - f(p)), & L_p = \text{“bac”}. \end{cases} \tag{7}$$

In Eq. (7), $f(p)$ is the class probability of the pixel p, and L_p is the label of pixel p.

The boundary term $B(L)$ is defined as in [31].

$$B(L) = \sum_{\{p,q\} \in N} S\{p, q\}, \tag{8}$$

$$S\{p, q\} = \mathrm{e}^{-\frac{(f(p) - f(q))^2}{2\sigma^2}} \times \frac{1}{\mathrm{dist}(p, q)}. \tag{9}$$

Variable N represents the set of edges in the graph.

To generate more accurate segmentation results with the incorporation of the spatial attention map, a novel spatial attention constraint term was formulated into the energy. Subsequently, Eq. (5) can be rewritten as follows:

$$E(L) = \lambda R(L) + B(L) + \gamma C(L). \tag{10}$$

The energy $C(L)$ can be rewritten as follows:

$$C(L) = \sum_{p \in P} H(p), \tag{11}$$

$$H(p) = \begin{cases} -\ln h(p), & L_p = \text{“obj”}, \\ -\ln(1 - h(p)), & L_p = \text{“bac”}. \end{cases} \tag{12}$$

In Eq. (12), $h(p)$ is the attention value of pixel p in the spatial attention map.

The proposed spatial attention constraint term aims to guarantee that the segmentation result is consistent with the spatial attention map. In the class probability map, some ALN pixels may be assigned a low probability of belonging to an ALN because large intraclass variance appears in the ALN region. However, they may be assigned a high attention value according to the spatial prior. Therefore, the spatial attention map can provide additional discriminative information for ALN segmentation. Moreover, the graph-based segmentation framework can capture the relationship between pixels, further improving the segmentation performance.

Table 1 Performance of different methods

Method	Jaccard (%)	DSC (%)	VOE (%)	FP (%)
SE-U-net	70.44	79.79	15.91	17.75
SAG-bi-net	72.36	82.01	16.35	18.52
DSAG-bi-net	**74.40**	**83.41**	**12.02**	**13.36**

4 Experiment

4.1 Experiment setting

A database was constructed by collecting ALN ultrasound images from Qianfushan Hospital of Shandong Province. The images in our database were collected from 216 patients (each patient contributed one image). ALNs were manually delineated by radiologists.

Images from 150 patients were selected as training images, whereas the remaining 66 images were used as test images. To improve the performance of the trained model, data augmentation was applied on the training images through rotation and translation. Subsequently, 1200 additional images were generated for training. In our experiments, the values of the parameters in Eq. (10) were set as follows: $\lambda = 5$, $\gamma = 0.5$, and $\sigma = 5$. The Pytorch [33] framework was implemented in our experiment. The initial learning rate, batch size, and training epoch of U-net were set as 0.001, 4, and 200, respectively. Cross-entropy loss was used as the segmentation loss. The stochastic gradient descent (SGD) optimizer was used for optimization, and the learning rate decay strategy was poly. The equipment used for the experiment was a 10-core 2.40 GHz Intel Xeon processor, with two Nvidia Titan XP GPUs and a 256 GB RAM.

In the experiment, JACCARD, volumetric overlap error (VOE), false positive rate (FP) and Dice's similarity coefficient (DSC) were used as the performance metric measures.

4.2 Analysis of effectiveness of DGM

An experiment was performed to demonstrate the effectiveness of the DGM. We compared the proposed method with those of SE-U-net and SAG-net (network with spatial attention constrained graph). We used SE-U-net and SAG-net separately to segment all images. The proposed DSAG-bi-net can be regarded as a combination of SE-U-net and SAG-net with adaptive segmentation. Table 1 shows the results of the different methods. As shown, the proposed method and SAG-bi-net outperformed SE-U-net. This is because it is difficult for SE-U-net to manage the complicated characteristics of complex images, thereby resulting in performance degradation. By contrast, spatial attention and the graph framework can improve the segmentation performance of complex images, resulting in improved segmentation performance. Furthermore, it is clear that the proposed DSAG-bi-net performed better than SAG-net. This is because DSAG-bi-net outperformed SAG-net in simple image segmentation. For simple image segmentation, SE-U-net may outperform than SAG-net. However, SAG-net uses spatial attention additionally, which may introduce noise in simple images. By contrast, for DSAG-bi-net, the DGM can select SE-U-net to segment simple images adaptively, thereby achieving better performances.

4.3 Analysis of effectiveness of complex branch

An experiment was performed to evaluate the effectiveness of the complex branch. We tested the performances of SE-U-net and SAG-net on complex images. Figure 7 shows the performances of the two methods; as shown, SAG-net outperformed SE-U-net. In ALN segmentation, the ALN structure is often missing in the segmentation result. However, the introduced SAM can learn the spatial attention map, which can provide additional discriminative information for ALN segmentation. Moreover, the graph-based segmentation framework can capture the relationship between pixels, further improving the segmentation performance of complex images.

4.4 Analysis of effectiveness of SE-U-net

An experiment was performed to evaluate the effectiveness of SE-U-net. The results are shown in Table 2 As can be seen, SE-U-net achieved better performances than the others on all metrics. This is because the SE blocks assigned different weights to channels by modeling the relationships between them. Moreover, they can increase the sensitivity to informative features, which is beneficial for improving the segmentation performance.

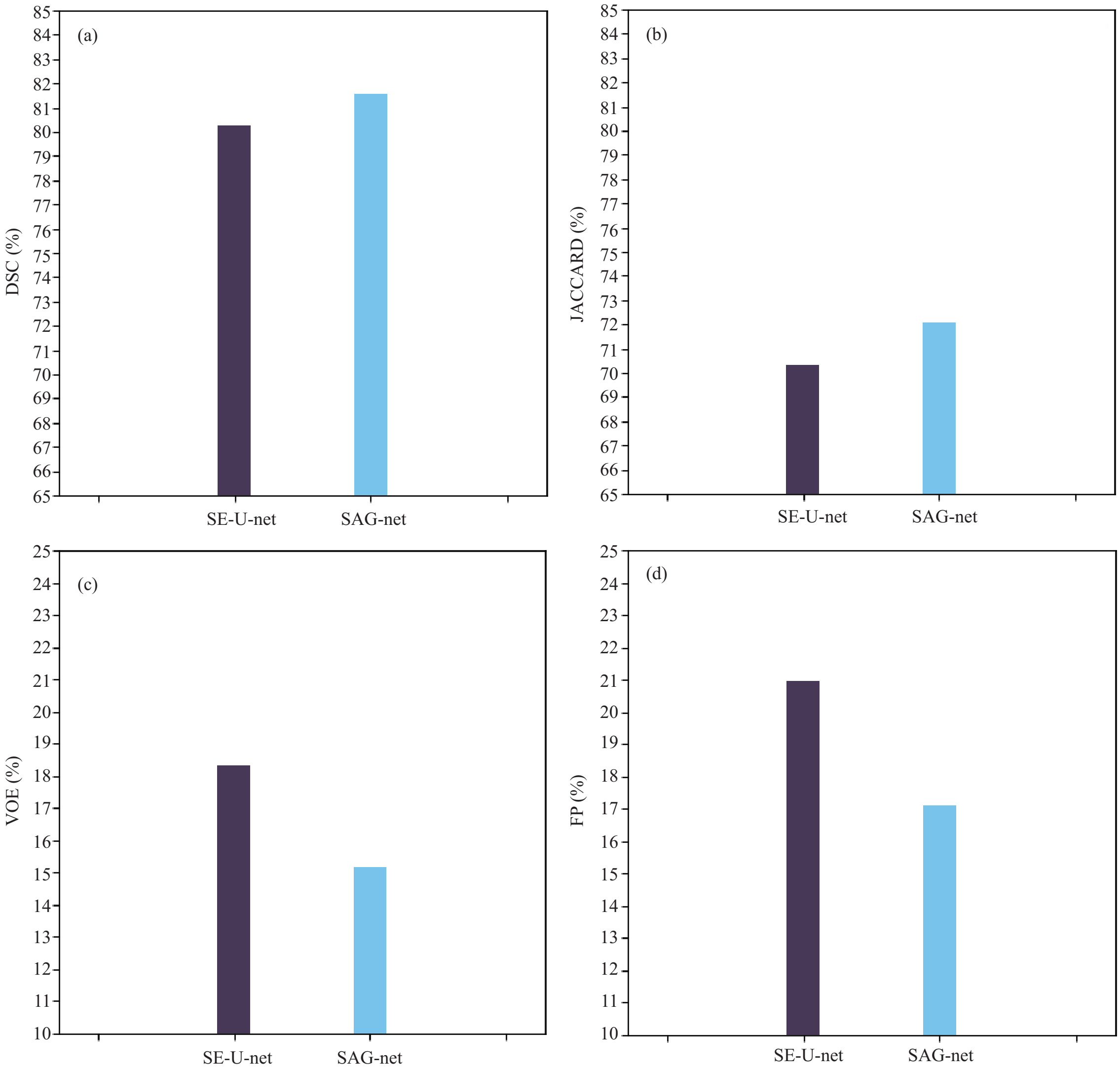


Figure 7 (Color online) Performance of SE-U-net and SAG-net on complex images. (a) DSC; (b) JACCARD; (c) VOE; (d) FP.

Table 2 Performance of different methods

Method	Jaccard (%)	DSC (%)	VOE (%)	FP (%)
U-net	67.85	78.27	18.41	22.83
SE-U-net	**70.44**	**79.79**	**15.91**	**17.75**

4.5 Compared with other segmentation methods

We compared our method with typical deep networks, such as FCN-8s [8], U-Net [9], deeplabv3+ [24], SegNet [20], and Frrn [34]. The initial learning rate, batch size, and training epoch were set as 0.001, 4, and 200, respectively.

As shown in the Table 3, the proposed method achieved the best performance. Figure 8 shows segmentation examples of these methods. As can be seen, these methods achieve better performances on simple images. However, for complex images, traditional deep learning methods could not easily achieve satisfactory performances because the difficulty diversity of the images was disregarded. It is difficult to learn the effective knowledge for both simple and complex images using a unified model because large variances exist in these two image levels.

However, the proposed method can manage simple and complex images adaptively by introducing the DGM. To manage complex images, a complex branch was constructed by introducing spatial attention and a graph framework. Spatial attention can assign high spatial attention to pixels that may be segmented

Table 3 Performance of different methods

Method	Jaccard (%)	DSC (%)	VOE (%)	FP (%)
U-net	67.85	78.27	18.41	22.83
FCN8s	54.76	67.57	31.89	43.73
Deeplabv3+	65.95	77.70	16.14	19.96
SegNet	62.83	73.47	18.94	22.68
Frrn	64.77	76.23	34.54	54.29
Our method	**74.40**	**83.41**	**12.02**	**13.36**

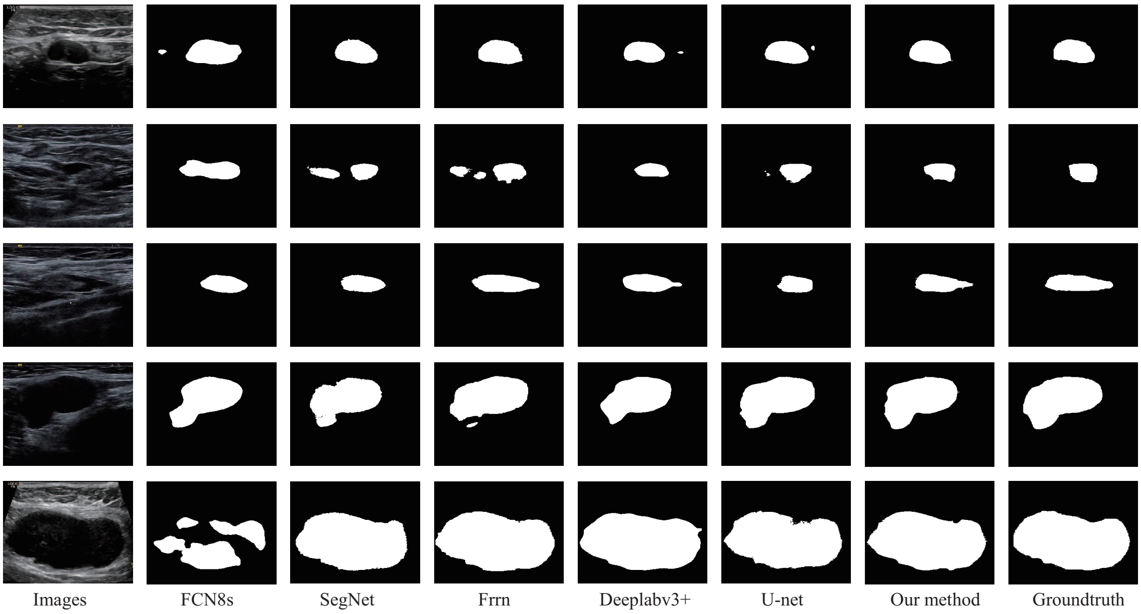


Figure 8 Examples of segmentation results of different methods.

incorrectly by SE-U-net. It can introduce additional discriminative information for these pixels, which is beneficial for the performance improvement. Moreover, the graph-based segmentation framework can capture the relationship between pixels, thereby further improving the segmentation performance of these complex images.

5 Conclusion

A novel difficulty-aware bi-network with spatial attention constrained graph for ALN segmentation was proposed herein. Traditional deep learning methods use a unified network to segment all images without considering the difficulty diversity. However, two network architectures were developed in the proposed method to adaptively segment images with different difficulty grades. The proposed network comprised a DGM and a segmentation network. The DGM was used to generate the difficulty grade of images. The developed segmentation network mainly comprised two functionality-specialized branches: an simple branch and a complex branch. The simple branch was used to manage simple image segmentation, and SE-U-net was developed by introducing an SE block. The complex branch focused on complex images. In the complex branch, SE-U-net was first used for segmentation. Subsequently, a SAM was proposed to generate a spatial attention map that can provide additional discriminative features. To investigate the spatial attention and relationship between pixels, a novel graph-based energy incorporating the spatial attention map was proposed. The experimental results on our database demonstrated that the proposed method outperformed state-of-the-art methods.

The proposed network is proposed to segment 2D image such as ultrasound image. However, it cannot be directly used for segmentation task of 3D image such as CT, MRI. Due to data difference between 2D data and 3D data, some challenges arise in 3D data segmentation task. For example, the proposed method ignores useful relationship information between slices in 3D data, result in the limitation of performance improvement. For future work, we will extend the proposed framework to 3D application.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61701280, 61801263, 61703235, 61701281), National Key R&D Program of China (Grant Nos. 2018YFC0830100, 2018YFC0830102), Natural Science Foundation of Shandong Province (Grant No. ZR2018BF012), and Foundation of Distinguished Associate Professor in Shandong Jianzhu University. The authors would like to greatly thank the anonymous reviewers for their valuable comments and suggestions.

References

1 Zhang J H, Wang Y Y, Shi X L. An improved graph cut segmentation method for cervical lymph nodes on sonograms and its relationship with node's shape assessment. Comput Med Imag Graph, 2009, 33: 602–607

2 Zhang Y Z, Ying M, Lin Y, et al. Coarse-to-fine stacked fully convolutional nets for lymph node segmentation in ultrasound images. In: Proceedings of IEEE International Conference on Bioinformatics and Biomedicine, 2016. 443–448

3 Chmielewski A, Dufort P, Scaranelo A M. A computerized system to assess axillary lymph node malignancy from sonographic images. Ultrasound Med Biol, 2015, 41: 2690–2699

4 Diepstraten S C E, Sever A R, Buckens C F M, et al. Value of preoperative ultrasound-guided axillary lymph node biopsy for preventing completion axillary lymph node dissection in breast cancer: a systematic review and meta-analysis. Ann Surg Oncol, 2014, 21: 51–59

5 Guo Q, Dong Z W, Zhang L, et al. Ultrasound features of breast cancer for predicting axillary lymph node metastasis. J Ultrasound Med, 2018, 37: 1354–1353

6 Cheng H D, Shan J, Ju W, et al. Automated breast cancer detection and classification using ultrasound images: a survey. Pattern Recogn, 2010, 43: 299–317

7 Zheng X Y, Yao Z, Huang Y N, et al. Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. Nat Commun, 2020, 11: 1236

8 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2015. 3431–3440

9 Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015. 234–241

10 Li C M, Huang R, Ding Z H, et al. A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI. IEEE Trans Image Process, 2011, 20: 2007–2016

11 Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks. IEEE Trans Pattern Anal Mach Intell, 2020, 42: 2011–2023

12 Debats O A, Litjens G J S, Barentsz J O, et al. Automated 3-dimensional segmentation of pelvic lymph nodes in magnetic resonance images. Med Phys, 2011, 38: 6178–6187

13 Zhy C M, Gu G C, Liu H B, et al. Segmentation of ultrasound image based on texture feature and graph cut. In: Proceedings of International Conference on Computer Science and Software Engineering, 2008. 795–798

14 Kuo J, Mamou J, Wang Y, et al. A novel nested graph cuts method for segmenting human lymph nodes in 3D high frequency ultrasound images. In: Proceedings of International Symposium on Biomedical Imaging, 2015. 372–375

15 Kuo J W, Mamou J, Wang Y, et al. Segmentation of 3-D high-frequency ultrasound images of human lymph nodes using graph cut with energy functional adapted to local intensity distribution. IEEE Trans Ultrason Ferroelect Freq Control, 2017, 64: 1514–1525

16 Zhang J H, Wang Y Y, Dong Y, et al. Sonographic feature extraction of cervical lymph nodes and its relationship with segmentation methods. J Ultrasound Med, 2006, 25: 995–1008

17 Bnouni N, Mechi O, Rekik I, et al. Semi-automatic lymph node segmentation and classification using cervical cancer MR imaging. In: Proceedings of International Conference on Advanced Technologies for Signal And Image Processing, 2018

18 Zhang Q, Huang C C, Li C L, et al. Ultrasound image segmentation based on multi-scale fuzzy c-means and particle swarm optimization. In: Proceedings of International Conference on Information Science and Control Engineering, 2012

19 Meinel L A, Bergtholdt M, Abe H, et al. Multi-modality computer-aided diagnosis system for axillary lymph node (ALN) staging: segmentation of ALN on ultrasound images. In: Proceedings of International Society for Optical Engineering, 2009

20 Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell, 2017, 39: 2481–2495

21 Chen L C, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs. 2014. ArXiv:1412.7062

22 Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans Pattern Anal Mach Intell, 2018, 40: 834–848

23 Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation. 2017. ArXiv:1706.05587

24 Chen L C, Zhu Y K, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of European Conference on Computer Vision, 2018. 833–851

25 Zhou Z W, Siddiquee M M R, Tajbakhsh N, et al. Unet++: a nested u-net architecture for medical image segmentation. 2018. ArXiv:1807.10165

26 Oktay O, Schlemper J, Folgoc L L, et al. Attention U-Net: learning where to look for the pancreas. 2018. ArXiv:1804.03999

27 Alom M Z, Hasan M, Yakopcic C, et al. Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation. 2018. ArXiv:1802.06955

28 Gu Z W, Cheng J, Fu H Z, et al. CE-Net: context encoder network for 2D medical image segmentation. IEEE Trans Med Imag, 2019, 38: 2281–2292

29 Wang Y F, Yue W W, Li X L, et al. Comparison study of radiomics and deep learning-based methods for thyroid nodules classification using ultrasound images. IEEE Access, 2020, 8: 52010–52017

30 He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2016. 770–778

31 Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts. IEEE Trans Pattern Anal Mach Intell, 2001, 23: 1222–1239

32 Boykov Y, Kolmogorov V. An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision. IEEE Trans Pattern Anal Mach Intell, 2004, 26: 1124–1137

33 Steiner B, DeVito Z, Chintala S, et al. PyTorch: an imperative style, high-performance deep learning library. In: Proceedings of Neural Information Processing Systems, 2019. 8026–8037

34 Pohlen T, Hermans A, Mathias M, et al. Full-resolution residual networks for semantic segmentation in street scenes. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2017. 3309–3318