



Automatic segmentation of thyroid with the assistance of the devised boundary improvement based on multicomponent small dataset

Yifei Chen^{1,2} · Xin Zhang¹ · Dandan Li¹ · HyunWook Park² · Xinran Li³ · Peng Liu⁴ · Jing Jin¹ · Yi Shen¹

Accepted: 23 February 2023 / Published online: 15 March 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Deep learning has been widely considered in medical image segmentation. However, the difficulty of acquiring medical images and labels can affect the accuracy of the segmentation results for deep learning methods. In this paper, an automatic segmentation method is proposed by devising a multicomponent neighborhood extreme learning machine to improve the boundary attention region of the preliminary segmentation results. The neighborhood features are acquired by training U-Nets with the multicomponent small dataset, which consists of original thyroid ultrasound images, Sobel edge images and superpixel images. Afterward, the neighborhood features are selected by min-redundancy and max-relevance filter in the designed extreme learning machine, and the selected features are used to train the extreme learning machine to obtain supplementary segmentation results. Finally, the accuracy of the segmentation results is improved by adjusting the boundary attention region of the preliminary segmentation results with the supplementary segmentation results. This method combines the advantages of deep learning and traditional machine learning, boosting the accuracy of thyroid segmentation accuracy with a small dataset in a multigroup test.

Keywords Thyroid ultrasound images · Image segmentation · Multicomponent neighborhood ELM · Small dataset

1 Introduction

The thyroid, an essential endocrine gland, can affect the growth and development of the human body. More seriously, a diseased thyroid can even cause death, which has received more attention, due to its increasing incidence [1, 2]. Medical imaging technology is an important tool that is widely used in diagnosing diseases, treatment, and postoperative patient monitoring. Various technologies, including computed tomography, ultrasound, X-ray and magnetic resonance imaging, play irreplaceable roles in clinics [3–6]. To test whether a thyroid is healthy, imaging technologies have been used in preliminary clinical examinations due to their advantages of intuition and noninvasion [7, 8]. Among them, ultrasound (US) technology is more widely utilized to diagnose the thyroid due to its low-cost, nonradiation and real-time

characteristics [9–11]. These specialties make US an indispensable diagnostic means in clinical settings. Thus, US images are of great value in medical diagnosis, and much information can be provided as a useful reference for physicians.

Since the estimation of useful clinical information depends on the boundary and shape of the thyroid, thyroid segmentation based on US images is a necessary process for obtaining useful information [12]. Traditional segmentation labels are drawn manually by physicians, which is time-consuming and laborious. Traditional machine learning (ML) methods are first utilized to eliminate or reduce the burden of manual segmentation. For instance, Garg et al. proposed a method based on a feed-forward neural network for segmenting thyroid [13], and Selvathi et al. utilized ELM and support vector machine (SVM) to segment thyroid glands based on US images and compared the results [14]. Additionally, Chang et al. adopted a radial basis function network for segmenting the thyroid and evaluating the volume of the thyroid [15], and Gomathy et al. achieved thyroid segmentation by using principal component analysis [16]. In these papers, researchers have used a variety of traditional ML methods to segment the

✉ Xin Zhang
zhangxin7030@gmail.com

Extended author information available on the last page of the article.

thyroid gland. Traditional ML can be trained with relatively small datasets. Nevertheless, because of the high complexity of medical images and the widespread requirement for precise segmentation, it is not easy to accurately segment the thyroid using only traditional ML.

With the rapid development of deep learning, deep learning has also been used to solve various computer vision problems, including medical images, target region segmentation, recognition and disease diagnosis [17–21]. Naturally, the application of deep learning to thyroid segmentation is also an area of interest. Many deep learning networks have been used for segmenting the thyroid, such as U-Net, a fully convolutional network (FCN), and down/upsampling [21–24]. For example, Chen et al. utilized U-Net for segmenting nodules in thyroid ultrasound images [21], and Nandamuri et al. introduced an FCN to segment the thyroid based on US images [22]. Chu et al. proposed a mark-guided deep learning model based on U-Net to segment ultrasound thyroid nodules [23], and Zhu et al. designed V-net by using downsampling and upsampling to achieve the semantic segmentation of CT thyroid images [24]. In the above papers, deep learning methods have been used to address thyroid segmentation. However, when using deep learning networks to implement segmentation, the accuracy is often affected by the requirement of large number of labeled images. Thus, many papers have investigated target segmentation with no or limited labels in different fields. For instance, Lu et al. presented the CO-attention Siamese Network to achieve zero-shot video object segmentation, obtaining better unsupervised segmentation results [19, 20], and Wu et al. proposed a symmetric driven generative adversarial network to segment brain tumors without labels [25]. Meanwhile, when there is only a few annotated samples, Lu et al. devised an attentive graph neural network to acquire more accurate results based on few-shot segmentation [26], Abdel-Basset et al. proposed a method based on few-shot learning to accurately segment COVID-19 infections from a limited segmentation labels [27], and Guo et al. presented the multi-level semantic adaptation few-shot method to segment cardiac image sequences under limited labels [28]. Nevertheless, in addition to the fact that producing labels is laborious, acquiring medical images is also difficult. Therefore, it is difficult to obtain tens of thousands of images and labels, which are usually needed for deep learning, when the targets are medical images. This paper intends to improve the segmentation accuracy of the thyroid in the case of inadequate data.

In the proposed method, a multicomponent neighborhood extreme learning machine (ELM) is devised for obtaining supplementary segmentation results to improve the thyroid segmentation results. The boundary region of the preliminary segmentation results, acquired by utilizing the original

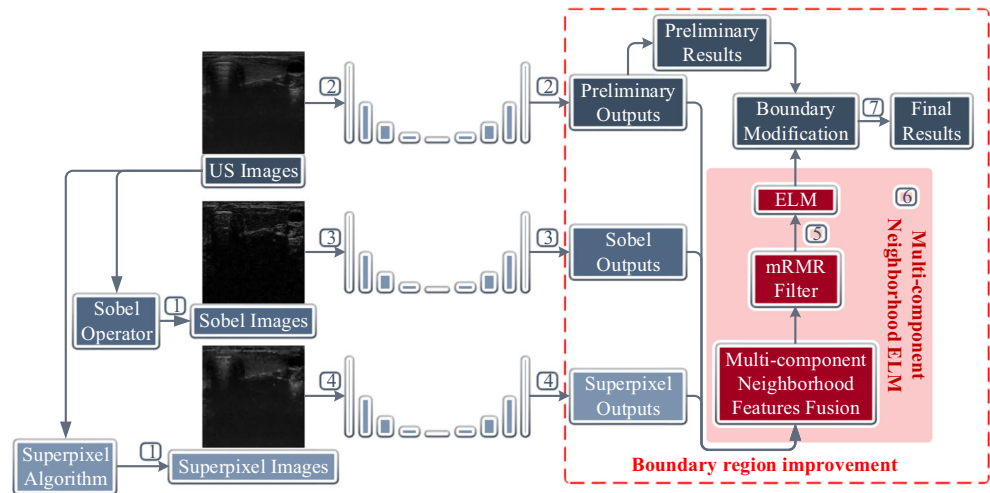
US thyroid images to train a U-Net, is complemented by the supplementary segmentation results. Additionally, to acquire multicomponent outputs, two types of images are extracted based on the original images and used to obtain segmentation results by using U-Net. The rest of this paper is organized as follows. The detailed methods involved in the proposed method are introduced in Section 2, and the experiment is introduced in Section 3. Subsequently, Section 4 compares and analyzes all the experimental results. Finally, the conclusion is given in Section 5.

2 Methods

In the proposed method, the thyroid segmentation results are boosted for small-dataset US images. The final segmentation results are acquired by improving the boundary region of the preliminary segmentation results with the supplementary segmentation results. Since the edge information in the image and the boundary of the segmentation results are closely related, the Sobel operator [29] is applied to process the original images. Additionally, the superpixel images containing the class information between the neighborhoods that facilitate segmentation are obtained by using the superpixel algorithm [30]. The segmentation outputs obtained by these two kinds of images pay more attention to the edge components and the neighborhood relationship in the images, which are beneficial for subsequent segmentation. Afterward, to obtain three kinds of segmentation outputs that are focused on different components of the thyroid US images, three types of images are used to train a U-Net [31] separately. Then, the devised multicomponent neighborhood ELM is utilized to obtain the supplementary segmentation results. The overall flow of the proposed method is shown in Fig. 1, and the detailed flow is as follows:

1. The image preprocessing is shown as Flow 1 in Fig. 1. The original thyroid images are utilized to generate Sobel edge images and superpixel images.
2. Three deep learning networks are shown as Flows 2, 3 and 4, respectively, in Fig. 1. The original images, Sobel edge images and superpixel images are used to train U-Nets. Subsequently, three sets of outputs can be obtained.
3. The supplementary segmentation is shown as Flow 6 in Fig. 1. The neighborhood features, extracted based on each set of outputs, are fused and selected by min-redundancy and max-relevance (mRMR) [32] filter. This process is shown as Flow 5 in the dotted box. Furthermore, the final supplementary segmentation results are obtained by reconstructing the classification results generated by ELM [33].

Fig. 1 Overall flow of the proposed method



4. The boundary modification is shown as Flow 7 in Fig. 1. The boundary attention of the preliminary results is modified by the supplementary segmentation results, achieving the final segmentation results.

The proposed method includes the following innovations:

1. The proposed method obtains sufficient samples based on the depth features of the deep learning network under a small dataset. Additionally, the sufficient samples are utilized by the designed machine learning method to obtain more precise segmentation.
2. In addition to the original images, two component images, Sobel edge images and Superpixel images, are integrated to obtain more depth features for supplementary segmentation, improving the boundary attention region.
3. A multicomponent neighborhood ELM is designed to extract and utilize multicomponent neighborhood features. The features are optimized by mRMR in the devised ELM to obtain a subset that is more beneficial for pixel classification.

2.1 Image preprocessing methods

To obtain segmentation outputs that focus on different characteristics of the thyroid US images, the Sobel operator and the superpixel algorithm are used in the proposed method for preprocessing the original thyroid US images. The Sobel edge images contain the edge information of the original images, and the superpixel images contain the class information between the neighborhoods that benefits segmentation. In addition to the original image, the two preprocessed images are also used to train U-Net separately. Based on the preliminary and preprocessed outputs obtained from the images focusing on different components, multicharacteristics neighborhood features can be extracted.

- 1) Sobel operator: To acquire the images that contain edge information, the Sobel operator is utilized to process the original images. The Sobel operator was first formally published by Irwin Sobel in 1968 [29]. The Sobel operator, a discrete differentiation operator, is an important method in the computer vision field and is widely used in edge detection [34]. The principle of the Sobel operator is to calculate the approximation gradient of the image luminance with discrete factorial operators. When addressing thyroid US images, it is equivalent to calculating the approximation of the gradient of the images. The schematic is shown in Fig. 2, and the brief procedure is as follows:

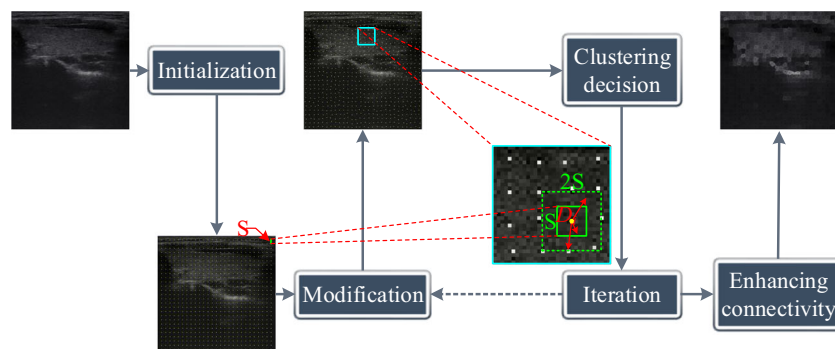
First, two 3×3 Sobel convolution kernels, the horizontal Sobel convolution kernel and the vertical convolution kernel are used to calculate Sobel edges. Then, the horizontal Sobel edge and vertical Sobel edge are obtained through the processing of original images. Finally, the Sobel edge image is acquired by integrating two directions of edge images. The whole formula is shown as follows:

$$I_{sobel} = \sqrt{(S_x * I)^2 + (S_y * I)^2}. \quad (1)$$

where I_{sobel} is the Sobel edge image, S_x is the horizontal convolution kernel, S_y is the vertical convolution kernel, and I is the original image. In the proposed method, Sobel edge images are used as inputs of the U-Net. The Sobel edge images are utilized to obtain segmentation outputs, which will be used to extract features attending to the edge information of images.

- 2) Superpixel algorithm: The images extracted by utilizing the superpixel algorithm contain neighborhood information that is beneficial for segmentation, which

Fig. 2 The process of the Simple Linear Iterative Clustering (SLIC)



can reflect the class information between the neighborhoods. The algorithm utilized for extracting superpixel images in the proposed method is simple linear iterative clustering (SLIC). This algorithm was first proposed by Achanta et al. in 2010 [30], and can be used for acquiring superpixel images with better edge adherence [30, 35]. Additionally, this algorithm can maintain image boundary information. The superpixel images obtained by SLIC have the advantages of high quality, compactness and nearly uniform.

First, the seed points of the superpixel images are initialized, and then these seed points are modified to make them not located on the edge. After initialization, the K seed points are evenly distributed in the image (a total of N pixels), the step size is $S = \sqrt{N/K}$, and the size of each superpixel is $S \times S$. In the modification, each initial seed point is moved to the pixel with the smallest gradient in the neighborhood of that seed point. Subsequently, the distances of all pixels in the search regions to the corresponding seed points are calculated. The distance metric of the j th pixel in the neighborhood of the i th seed point is calculated as follows:

$$D_{ij} = \sqrt{\left(\frac{\sqrt{(l_j - l_i)^2}}{C}\right)^2 + \left(\frac{\sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}}{S}\right)^2}. \quad (2)$$

where D_{ij} is the distance metric, the first item in the radical is the color distance in CIELAB space, the second item in the radical is the spatial distance in CIELAB space, l is the brightness of the image, C is the maximum color distance in an image, x and y are the horizontal and vertical dimensions of a pixel, respectively, and S is the maximum spatial distance. Finally, the superpixel images obtained after several iterations are processed to increase the connectivity. The discontinuous pixels are reassigned to neighboring superpixels. After the whole process, superpixel images are obtained for subsequent training; the SLIC process is

shown in Fig. 2. Compared to the original image, there are no pixels with extremely small grayscale changes in the superpixel images. Additionally, according to SLIC, the class information between neighborhoods is well stored in the superpixel images, which can reflect variations between neighborhoods.

2.2 Image segmentation network

To acquire the segmentation outputs for extracting features focusing on different characteristics of the images, the original image, Sobel edge image and superpixel image are utilized for training U-Nets. Ronneberger et al. first proposed U-Net in 2015, and it won the ISBI cell tracking challenge in 2015 for segmenting neuronal structures [31]. With the development of deep learning and the increase in computing power, various segmentation tasks are implemented by U-Net. Naturally, despite the complexity of medical images, U-Net can achieve better performance in the segmentation of medical images [31, 36].

U-Net is a U-shaped network that consists of a contracting path, an expansive path and skipconnections. The structure of this network is shown in Fig. 3. The contracting path is composed of convolution layers and max pooling layers, and the expansive path is composed of upconvolution layers and convolution layers. U-Net has the advantage of obtaining better results with fewer images, making it more suitable to be applied to medical image segmentation. Therefore, in this paper, U-Net was used three times for segmentation of the thyroid based on different kinds of images, including original US images, Sobel edge images and superpixel images.

2.3 Devised multicomponent neighborhood ELM

To utilize the effective features extracted from three kinds of segmentation outputs to obtain the supplementary segmentation results, a multicomponent neighborhood ELM is devised in the proposed method. In the whole process, the multicomponent segmentation outputs are first expanded so that the edge pixels of the output images also have

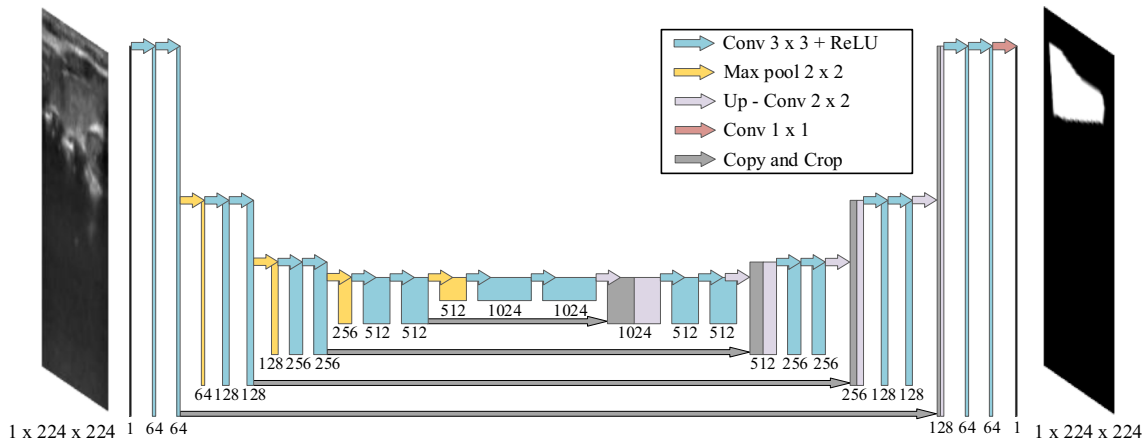


Fig. 3 The structure of U-Net

neighborhood features. The output images are appended with two pixels on every horizontal and vertical image boundary. Then, the neighborhood features with 5×5 pixels are extracted for each pixel. Additionally, the label corresponding to the central pixel of the neighborhood feature is adopted as the target for the final training, and all neighborhood features are fused. The neighborhood features with $5 \times 5 \times 3$ pixels are resized to 75×1 .

Subsequently, the fused features are selected by the mRMR filter to obtain more useful features. The mRMR filter is a feature selection method proposed in 2005 by Peng et al. [32]. This method can select the features effectively and has been used in many fields [21, 37, 38]. The mRMR filter aims to minimize the relevance between the features while maximizing the relevance between the feature and the corresponding category. In this method, the relevance is calculated based on mutual information. The formula of mRMR is shown as follows:

$$\begin{cases} \max(R_{fc}), R_{fc} = \frac{1}{|F|} \sum_{f_i \in F} I(f_i; c_i) \\ \min(R_{ff}), R_{ff} = \frac{1}{|F|^2} \sum_{f_i, f_j \in F} I(f_i; f_j) \\ I(x; y) = \int \int p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy \end{cases} \quad (3)$$

where R_{fc} is the relevance between feature and category, R_{ff} is the relevance between features, $I(\cdot)$ is the mutual

information, F is the feature set, f_i and f_j are the features in the corresponding feature set, c_i is the corresponding category, and $p(\cdot)$ and $p(\cdot, \cdot)$ are the probability density functions. In this paper, 75 neighborhood features are selected by mRMR. After selection, the relevance between the feature and classification category is guaranteed, while some redundant features are eliminated.

Finally, the ELM is trained with the selected features and the corresponding targets, and the outputs of the ELM are reconstructed to the segmentation results. ELM is a kind of feedforward neuron network that was first proposed in 2006 by Huang et al. [33]. This ML method has been widely used in computer vision because of its ease of use, faster learning speed, strong generalization ability and so on. In ELM, the parameters of the hidden layer are set randomly and do not need to be updated in training. During the learning process, only the output weights of ELM need to be updated during learning [33, 39]. First, the parameters of the hidden nodes are generated randomly according to any continuous probability distribution. Then, the output matrix of the hidden layer is calculated. Finally, the output weights of the ELM are solved according to the loss function with the objective of minimizing the error. The loss function is defined as follows:

$$L_{ELM} = \|H\beta - T\|^2, \beta \in R_{L \times M}. \quad (4)$$

Fig. 4 The devised multicomponent neighborhood ELM

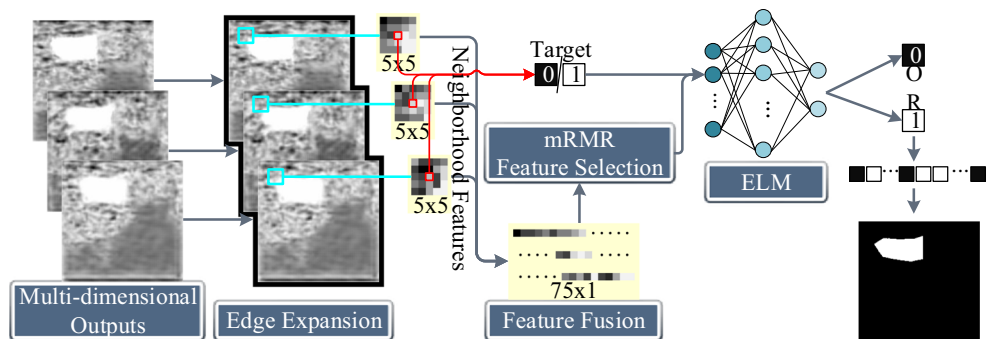
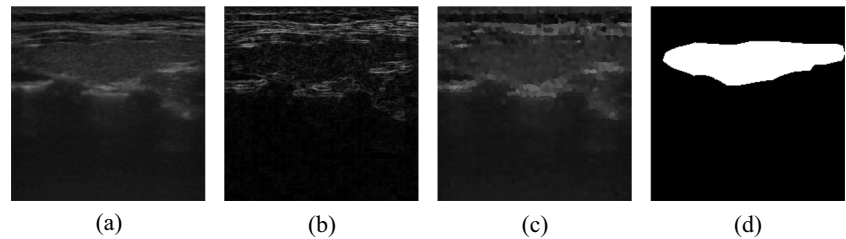


Fig. 5 The specimen of the U-Net dataset: (a) the original US thyroid image, (b) the Sobel edge image, (c) the superpixel image and (d) the segmentation label



where L_{ELM} is the loss function, H is the output matrix of the hidden layer, β is the output weights, T is the training target, L is the number of hidden nodes and M is the number of output nodes. In this paper, the hidden nodes are set to 300, and the output nodes are set to 2. The selected features are utilized from training the ELM to acquire the supplementary segmentation results. The proposed ELM is shown in Fig. 4.

3 Experiments

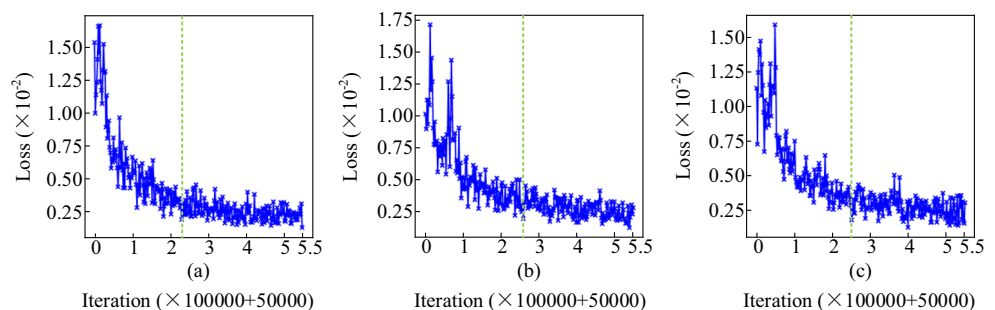
The experiments are carried out according to the entire method presented in Section 2. The original thyroid US images and the corresponding labels were provided by physicians from the Heilongjiang Provincial Key Laboratory of Trace Elements and Human Health and Endemic Disease Control Center at Harbin Medical University. All images were verified by the physicians, and the total number of images was 1,595. In this section, dataset construction, training U-Net and the devised ELM, and the boundary modifications are illustrated in detail.

Since the training of the U-Net and the devised multicomponent neighborhood ELM are involved in the proposed method, the datasets are constructed twice in this paper. Regardless of which dataset is constructed, it is important to ensure that the thyroid images from the same patient are classified into the same subset. First, before constructing the dataset for training the U-Net, all images are preprocessed by the Sobel operator and superpixel algorithm. Subsequently, three kinds of images are utilized to construct a dataset. Each group of samples is divided into a training set and a test set. The number of samples

used for training a U-Net is 1,251, and the number of test samples is 344. The specimen of the images and labels are shown in Fig. 5. Then, the dataset used for training the multicomponent neighborhood ELM consists of three segmentation outputs acquired from the U-Nets. The extracted features are three 5×5 squares obtained from three outputs, and the training target is the pixel in the segmentation label corresponding to the center of the square. In this dataset, all the U-Net test images were equally divided into 8 subsets for the 8 test groups. There are 43 images in each subset, and 50,176 5×5 squares can be extracted from each image. Thus, the number of samples in the training set is 2,157,568, and the number of squares in the test set is 15,102,976.

After constructing the datasets, three U-Nets are trained using the U-Net dataset first to obtain multicomponent features. Since medical images differ significantly from other images, the images from different components are utilized to train a U-Net directly, and the hyperparameters are as follows: optimization - stochastic gradient descent, momentum - 0.99, weight decay - 0.0005, learning rate - from $1e-06$ to $1e-08$. The loss curves for three trainings are shown in Fig. 6. In this figure, the models are optimized near the green dotted line. Meanwhile, because the loss after the initialization of the network is much larger than the loss after stabilization, the starting point of each curve is modified to the appropriate position to make the curve clearer. Subsequently, the devised ELM is also trained with the training dataset. Before training, the neighborhood features, which consist of three 5×5 squares, are first fused before training the devised ELM. All squares are resized to 25×1 pixels, and then the resized features are combined in the order of original images, Sobel edge

Fig. 6 The curves of training loss: (a) the loss of the U-Net trained with original images, (b) the loss of the U-Net trained with Sobel edge images and (c) the loss of the U-Net trained with superpixel images



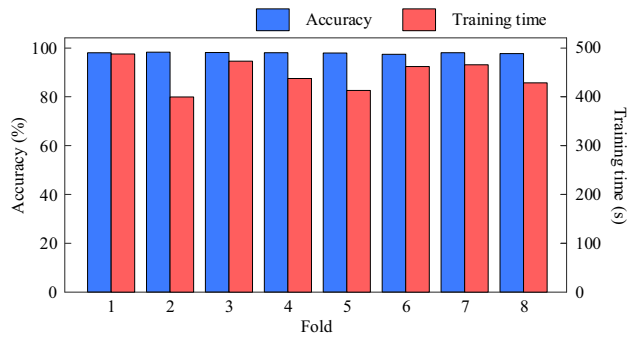


Fig. 7 The training accuracy and training time of the devised ELM

images, and superpixel images. Afterward, the combined features consisting of 75 pixels are selected by mRMR. After feature selection, 20% of the features are removed, and the selected features consist of 60 pixels. Finally, the selected features are used for training an ELM. In the experiment, the ELM was trained eight times, and 2,157,568 features in one test group were utilized in each training process. The training accuracy that can be correctly classified and the training time of 8 trainings are shown in Fig. 7. After each multicomponent neighborhood ELM training process, the 15,102,976 features from other groups are used for testing.

In this experiment, the improvement of the boundary attention region is implemented after constructing the datasets, the U-Net training and the process of multicomponent neighborhood ELM. The boundary attention regions of the preliminary segmentation results are adjusted by the supplementary segmentation results from the devised ELM to improve the final segmentation results. The inside regions of the preliminary results are preserved, and the boundary regions are added to the supplementary results. The schematic of the improvement of the boundary attention region is shown in Fig. 8.

To verify the proposed method and each technique, some ablation studies are added. The supplementary segmentation results and final segmentation results, obtained by using unselected neighborhood features with ELM, are compared (denoted as Supplementary (-mRMR) and Final (-mRMR), respectively). Then, the supplementary segmentation results

obtained by using original deep features and any two component depth features are compared (denoted as Supplementary (Original), Supplementary (Original+Sobel), Supplementary (Original+Superpixel) and Supplementary (Sobel+Superpixel), respectively), and the corresponding final segmentation results are also compared (denoted as Final (Original), Final (Original+Sobel), Final (Original+Superpixel) and Final (Sobel+Superpixel), respectively). Finally, more comparative experiments were conducted to demonstrate the improvement of the proposed method. Some commonly used segmentation methods and the methods proposed for thyroid segmentation, ultrasound image segmentation are compared, including FCN-8s, SegNet, MGU-NET, SV-net and VEU-Net [24, 29, 40–42].

4 Results and discussion

After the overall experiment, the preliminary segmentation results are obtained from the preliminary outputs, the supplementary segmentation results are acquired by the devised ELM, and then the final segmentation results can be acquired for each test group. In this section, these results are evaluated and analyzed. Furthermore, the segmentation results obtained from the other outputs and the unselected neighborhood features are also compared with the proposed method.

4.1 The preliminary segmentation results

Preliminary segmentation of thyroid US images is achieved by training a U-Net with original images and segmentation labels. The segmentation results are the binarization results of the corresponding outputs. Some examples of the preliminary segmentation outputs, preliminary segmentation results and the corresponding labels are shown in Fig. 9 (a), (b) and (d), respectively. Although the preliminary segmentation results are similar in shape to the segmentation labels, it is obvious that these segmentation results are not sufficiently precise. To intuitively analyze the accuracy of the preliminary segmentation results, the contours (blue lines) of the segmentation label are extracted

Fig. 8 The schematic of the improvement of boundary attention region

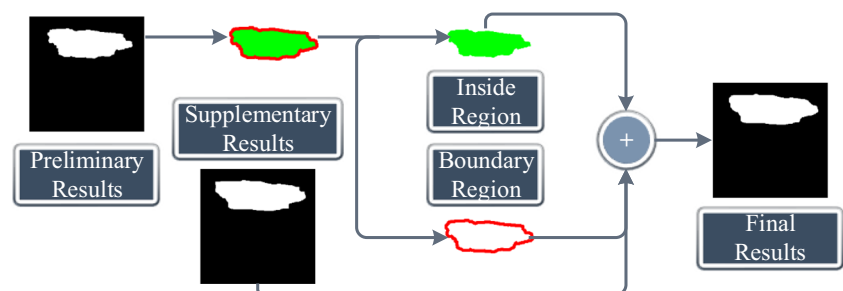
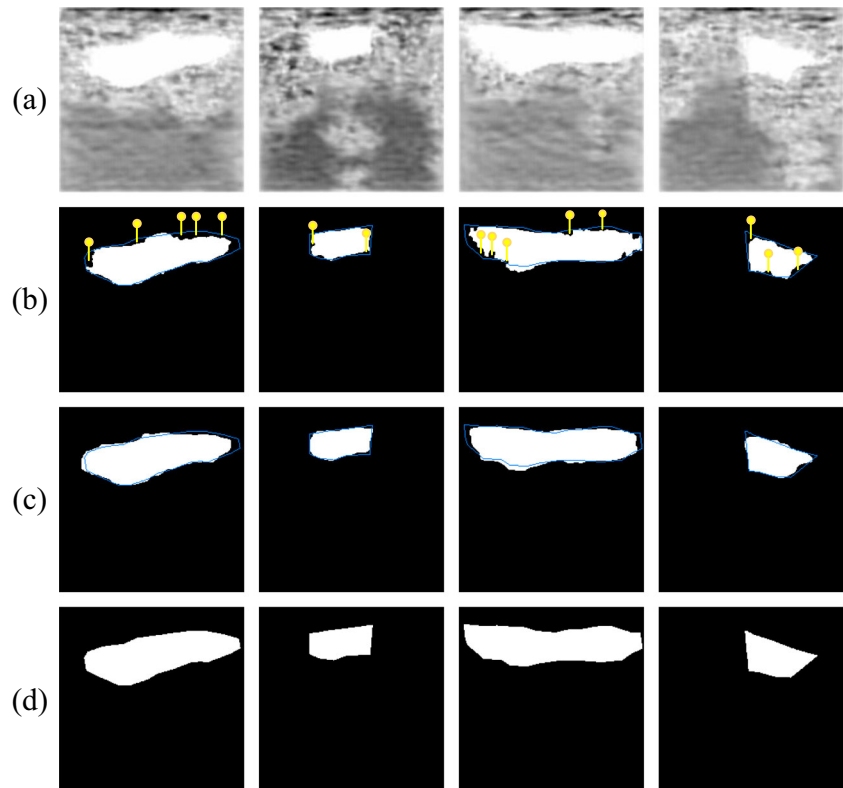


Fig. 9 Example images of the preliminary segmentation and supplementary segmentation: (a) the preliminary segmentation outputs, (b) the overlaid contours (blue lines) of the label on the preliminary segmentation results, (c) the overlaid contours (blue lines) of the label on supplementary segmentation results and (d) the segmentation labels



and overlaid on the preliminary segmentation results, as shown in Fig. 9 (b). The preliminary segmentation results cover most of the region of the thyroid, while only some of the boundaries are inaccurate. In particular, some areas near the contour are not correctly segmented.

To further analyze the accuracy of the segmentation results, the test set with 8 test groups is analyzed with IoU, MCC, F1 score and Hausdorff distance 95 (HD95) [26, 43–45]. The formulas of these indices are shown as follows:

$$IoU = \frac{TP}{FP + TP + FN}. \quad (5)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}. \quad (6)$$

$$F1 = \frac{2TP}{N + TP - TN}. \quad (7)$$

$$HD95 = \max \left\{ \max_{x \in X} \min_{y \in Y} hd \{x, y\}, \max_{y \in Y} \min_{x \in X} hd \{x, y\} \right\} \quad (8)$$

where N is the number of pixels in an image, TP is the number of pixels that are segmented correctly among the pixels in the thyroid region of the labels, FN is the number of pixels that are segmented incorrectly among the pixels in the thyroid region of the labels, TN is the number of

pixels that are segmented correctly among the pixels in the nonthyroid region of the labels, FP is the number of pixels that are segmented incorrectly among the pixels in the nonthyroid region of the labels, x is the pixel in the boundary line of segmentation results, y is the pixel in the boundary line of labels, X is the set of all x , Y is the set of all y , and hd is the Euclidean distance. The mean values of these indices are 0.7995 (IoU), 0.8782 (MCC), 0.8867 (F1), and 2.4469 (HD95), and the calculation results are shown in Fig. 10. Nevertheless, neither the average values nor the maximum values are sufficiently high. In this paper, just over 1,000 images are used for training, contrary to the requirement of tens of thousands of samples used to train deep learning networks. Simultaneously, the presence of a large number of complex nonthyroidal regions will aggravate the poor segmentation when the training set is not

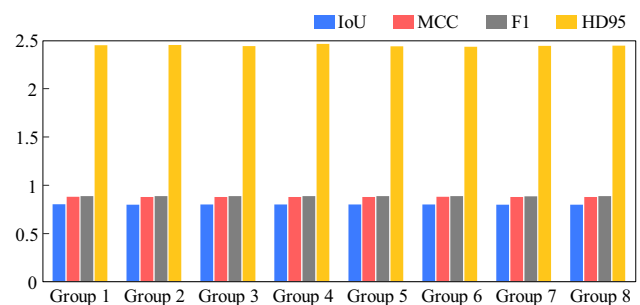


Fig. 10 The calculation results of preliminary segmentation results

Table 1 The accuracy indices and improvement of the supplementary segmentation results

Test set	Accuracy indices				Improvement			
	IoU	MCC	F1	HD95	IoU	MCC	F1	HD95
Group 1	0.8127	0.8863	0.8951	2.3391	1.68%	0.88%	0.92%	5.46%
Group 2	0.8095	0.8842	0.8931	2.3338	1.59%	0.84%	0.86%	5.75%
Group 3	0.8127	0.8862	0.8952	2.3244	1.94%	1.04%	1.08%	5.75%
Group 4	0.8110	0.8851	0.8940	2.3530	1.58%	0.83%	0.85%	5.45%
Group 5	0.8157	0.8881	0.8966	2.3361	2.12%	1.13%	1.15%	5.25%
Group 6	0.8184	0.8900	0.8985	2.3070	2.48%	1.32%	1.36%	6.25%
Group 7	0.8131	0.8865	0.8949	2.3246	2.13%	1.14%	1.15%	6.70%
Group 8	0.8164	0.8888	0.8974	2.3228	2.51%	1.32%	1.38%	6.17%
Mean	0.8173	0.8869	0.8956	2.3301	2.01%	1.06%	1.09%	5.85%

sufficient. Therefore, the preliminary segmentation results are imprecise and need to be further improved.

4.2 The supplementary segmentation results

The supplementary segmentation results are acquired from a trained multicomponent neighborhood ELM. Some supplementary segmentation results are shown in Fig. 9 (c). Compared to the segmentation results in Fig. 9 (b), the supplementary segmentation results are more similar to the labels. In Fig. 9 (c), the supplementary segmentation results are closer to the expected contours, and the gap between the contour and the segmentation result is significantly reduced. The reduced gap is marked with yellow flags in Fig. 9 (b). Although the supplementary results slightly exceeded contours at some parts, the reduction in the gap has a greater impact on the segmentation results. Therefore, this analysis shows that the accuracy of the supplementary segmentation results is improved.

Then, to clearly demonstrate the improvement of the supplementary results, accuracy indices are utilized to compare the supplementary segmentation results with the preliminary segmentation results. The average improvement rate (compared to the preliminary segmentation results) for each index on each test group is also calculated, and the results are shown in Table 1. The accuracy of all test groups is boosted, the mean value of the average

improvement rate is 2.01% in IoU, 1.06% in MCC, 1.09% in F1 score and 6.70% in HD95, and the mean value of IoU, MCC, F1 score and HD95 achieves 0.8137, 0.8869, 0.8956 and 2.3301, respectively. From these results, the segmentation results of the devised ELM have better performance in terms of accuracy. In all the test sets, the supplementary segmentation results of Group 6 achieve the best performance (marked as bold), and those of Groups 6, 7, and 8 achieve the maximum improvement on different indices (marked as blue).

Afterward, because the boundary attention region is used to modify the preliminary results, the boundary region of the supplementary segmentation results is further analyzed. To visualize the improvement of the supplementary segmentation results, the boundary attention region obtained from the multicomponent neighborhood ELM is shown in Fig. 11. The area that is covered by the preliminary segmentation results is marked as white, the improved area from the supplementary segmentation is marked as green, and the area that is not contained in the segmentation labels is marked as red. Furthermore, the blue area is not included in the supplementary segmentation results. The red and blue regions are the error parts and the eliminated error parts of the supplementary segmentation, respectively, whereas the green region is the error part of the preliminary segmentation. Although the error area exists, the largest area of the boundary is improved.

Fig. 11 Examples of the boundary attention regions of supplementary segmentation result

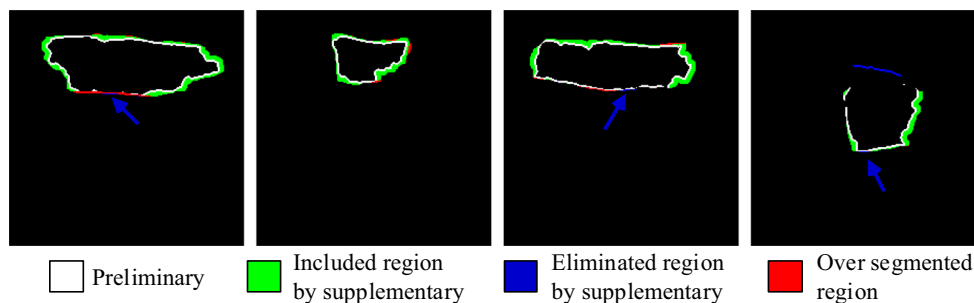
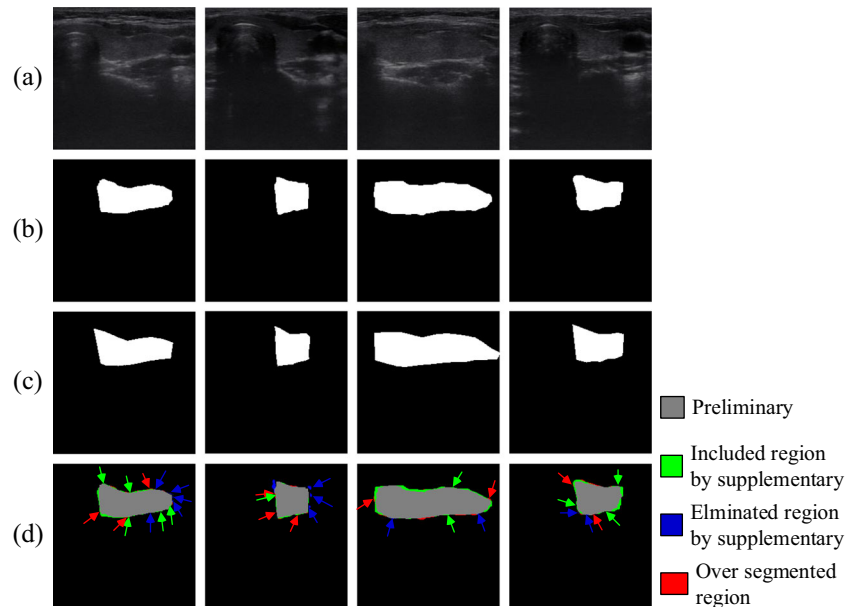


Fig. 12 Examples of the final segmentation results: (a) original thyroid ultrasound images, (b) the boosting final segmentation results, (c) the segmentation labels and (d) final segmentation with annotations



Meanwhile, the mean values of the average improvement are calculated for IoU, MCC and F1 score, which are 33.04%, 20.82%, and 20.96%, respectively. Therefore, the analysis and comparison of the supplementary segmentation results from several perspectives prove that the use of filtered multicomponent neighborhood features and devised ELM can introduce more information that is effective for segmentation.

4.3 The final segmentation results

The final segmentation results are obtained by improving the boundary attention region of the preliminary segmentation results. The preliminary segmentation results are boosted with the assistance of the supplementary segmentation results, which are acquired from the multicomponent neighborhood ELM. The images in eight test groups are preliminarily segmented and then improved by utilizing the devised ELM. Some examples of the final boosting results are shown in Fig. 12.

As shown in Fig. 12 (b) and (c), the shape of the final segmentation results is extremely close to that of the segmentation labels, although there is a slight difference

between them. Through the analysis of these results, it is clear that more accurate thyroid segmentation results are obtained under a small dataset by using the proposed method. To further demonstrate the improvement of the proposed method, the final segmentation results with annotations are shown in Fig. 12 (d). The gray area represents the area that has been correctly segmented by the preliminary segmentation, the green area represents the correct region added by the proposed method, the blue area represents the error area in preliminary segmentation that has been eliminated by the proposed method, the red area represents the final segmentation beyond target labels, and the arrows of corresponding colors point out some tiny areas. From these examples, it can be found that the inside region can be mostly covered by the preliminary segmentation results. The sum of the improvement area and the eliminated error is much larger than the error area. Therefore, the final segmentation results are boosted by using the multicomponent neighborhood ELM to improve the boundary attention region of the preliminary segmentation results.

Consequently, three accuracy indices are obtained to analyze the segmentation performance of the final

Table 2 The accuracy indices of the final segmentation results

Test set	IoU	MCC	F1	HD95	Test set	IoU	MCC	F1	HD95
Group 1	0.8157	0.8883	0.8971	2.3232	Group 5	0.8207	0.8915	0.9000	2.3247
Group 2	0.8137	0.8869	0.8958	2.3208	Group 6	0.8214	0.8920	0.9005	2.2991
Group 3	0.8148	0.8875	0.8965	2.2997	Group 7	0.8186	0.8902	0.8986	2.2812
Group 4	0.8150	0.8878	0.8967	2.3390	Group 8	0.8183	0.8901	0.8986	2.2875
–	–	–	–	–	Mean	0.8173	0.8893	0.8980	2.3094

segmentation results, as shown in Table 2. In this table, there are two groups with an IoU over 0.82, four groups with an MCC over 0.89, and two groups with an F1 score over 0.90. In the eight groups, the mean value of IoU is 0.8173, the mean value of MCC is 0.8893, the mean value of F1 score is 0.8980, and the mean value of HD95 is 2.3094. The best scores of three indices are obtained from Group 6 (marked as bold), where the values of IoU, MCC and F1 score are 0.8214, 0.8920 and 0.9005, respectively. The best score of HD95, 2.2812, is obtained from Group 7 (marked as bold). By comparing these results with the preliminary segmentation results, the precision of the final segmentation results is significantly improved. The IoU is increased by 0.0143-0.213, the MCC is increased by 0.0090-0.0132, the F1 score is increased by 0.0092-0.0134 and the HD95 is increased by 0.1140-0.1626. The accuracy of the final segmentation results is also proven to be strengthened by comparison with the results in Table 1.

To clearly validate the effect of multicomponent images and better explain the improvement of the final segmentation results, the saliency maps [46] are utilized to reflect the pixels that play important roles (30% pixels of target segmentation regions) in training U-Nets. In order to analyze some of the details in the saliency map in detail, rectangles were added in this figure. The comparison of segmentation results with segmentation error (including over segmented region and unsegmented region) and saliency maps is shown as Fig. 13. Comparing the samples in Fig. 13 (a) and (b), the segmentation error is significantly reduced. As Fig. 13 (c), (d) and (e), the important pixels of saliency maps are marked as Indian Red in the corresponding input image.

It can be seen that the multicomponent images pay more attention to different regions of the target when segmenting thyroids. This can be verified by the rectangles in the first row. In Fig. 13 (f), important pixels in different components are overlapped and the overlap can cover and be similar to the area of the thyroids, demonstrating the effectiveness of using multicomponent images. Then, as the rectangles in the second row, although the multicomponent images focus on the rectangular region, the details of attention are different. Therefore, depth features that can respond to different details are fused, achieving a higher improvement in the final segmentation. As the rectangles in the third and fourth row, the attention details are brought by Sobel edge images and superpixel images, respectively, improving the precision of segmentation.

Subsequently, the average improvement rates of the final segmentation to some comparison experiments are calculated. In primary comparisons, the final segmentation is compared with preliminary segmentation (noted as Original), supplementary segmentation (noted as Supplementary), segmentation results obtained by only using Sobel edge images (noted as Sobel), and the segmentation results obtained by only using superpixel images (noted as Superpixel). To further validate the proposed method and each technique, the ablation studies mentioned in Experiments are compared, and the calculation results of the average improvement are shown in Table 3. In this table, the lower the value of the average improvement is, the higher the value of the compared method on the corresponding index, proving that the compared method has better segmentation results. As for the calculation results, all four indices of the

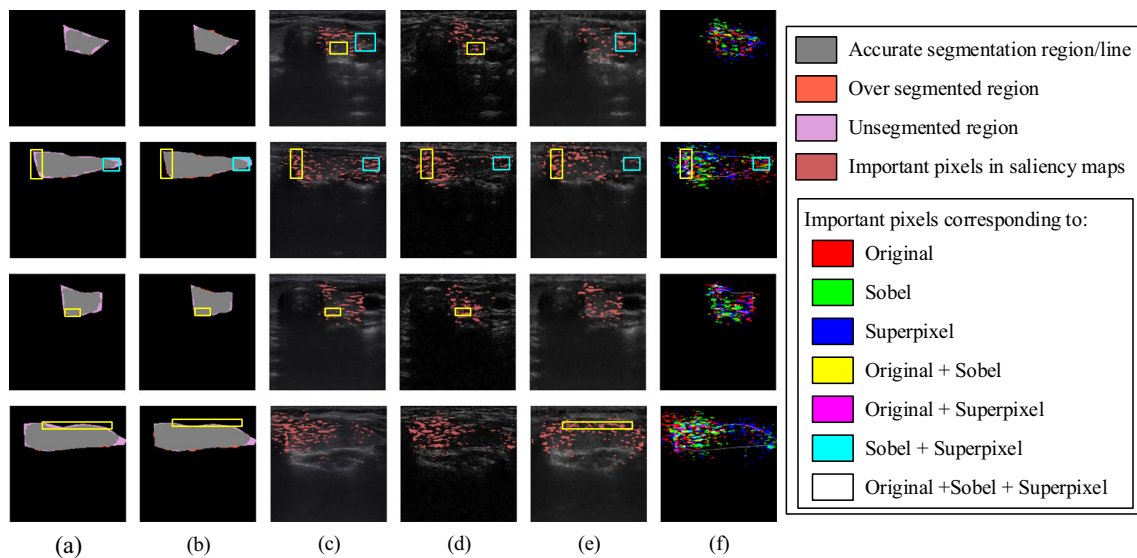


Fig. 13 Comparison of segmentation results with segmentation error and saliency maps: (a) preliminary segmentation results, (b) final segmentation results, (c) original images with important pixels in corresponding saliency maps, (d) Sobel edge images with important pixels

in corresponding saliency maps, (e) Superpixel images with important pixels in corresponding saliency maps and (f) Overlay of important pixels in three saliency maps

Table 3 The average improvement of the proposed method over the comparison experiment in the four accuracy indices

Comparison	Num	Average improvement			
		IoU	MCC	F1	HD95
Preliminary	1	2.43% ± 0.41%	1.33% ± 0.25%	1.35% ± 0.23%	6.85% ± 1.22%
Sobel	2	10.73% ± 0.57%	5.31% ± 0.23%	6.02% ± 0.31%	4.30% ± 0.94%
Supersixel	3	3.68% ± 0.60%	2.03% ± 0.32%	2.03% ± 0.34%	2.49% ± 0.98%
Supplementary	4	0.85% ± 0.53%	0.37% ± 0.21%	0.38% ± 0.23%	1.23% ± 0.62%
Supplementary(-mRMR)	5	0.94% ± 0.26%	0.53% ± 0.15%	0.53% ± 0.15%	2.68% ± 1.08%
Supplementary(Original)	6	0.90% ± 0.27%	0.53% ± 0.16%	0.51% ± 0.16%	2.71% ± 1.40%
Supplementary(Original+Sobel)	7	1.17% ± 0.15%	0.69% ± 0.09%	0.65% ± 0.08%	3.23% ± 0.57%
Supplementary(Original+Supersixel)	8	0.84% ± 0.34%	0.48% ± 0.19%	0.48% ± 0.20%	2.87% ± 1.98%
Supplementary(Sobel+Supersixel)	9	2.19% ± 0.32%	1.28% ± 0.18%	1.21% ± 0.18%	2.81% ± 1.44%
Final(-mRMR)	10	0.19% ± 0.13%	0.12% ± 0.08%	0.10% ± 0.07%	1.49% ± 1.08%
Final(Original)	11	0.83% ± 0.31%	0.48% ± 0.17%	0.46% ± 0.17%	2.49% ± 1.47%
Final(Original+Sobel)	12	0.54% ± 0.08%	0.33% ± 0.05%	0.30% ± 0.05%	3.18% ± 0.57%
Final(Original+Supersixel)	13	0.27% ± 0.12%	0.16% ± 0.07%	0.15% ± 0.06%	2.33% ± 1.78%
Final(Sobel+Supersixel)	14	0.66% ± 0.13%	0.42% ± 0.10%	0.37% ± 0.09%	2.91% ± 1.78%

proposed method showed significant improvements compared to the preliminary comparison (Comparisons 1, 2, 3, and 4). The precision of the final segmentation results is better than that of the preliminary results and supplementary results. Compared with Preliminary, the final segmentation results are improved by approximately 2.43% on IoU, 1.35% on MCC, 1.35% on F1 score and 6.85% on HD95 on average, and compared with Supplementary, the final segmentation results are improved by approximately 0.85% on IoU, 0.37% on MCC, 0.38% on F1 score and 1.23% on HD95.

To verify the effectiveness of using three different component images, Comparisons 10, 11, 12, 13, and 14 are analyzed. It can be seen that the best segmentation accuracy can be obtained based on the multicomponent neighborhood features in four indices. The addition of supersixel images and Sobel edge images can improve thyroid segmentation. When compared to Comparisons 11, 12, and 13, the addition of another component to the original image can lead to an enhancement of the thyroid segmentation. In this comparison group, approximately 0.29% improvement in IoU, 0.15% improvement in MCC, and 0.16% improvement in F1 score was brought by Sobel edge images, and approximately 0.46% improvement in IoU, 0.32% improvement in MCC, 0.31% improvement in F1 Score and 0.16% improvement in HD95 was achieved by supersixel images. Furthermore, as shown in Comparisons 6, 9, 11, and 14, if the depth features of the original image are not used, the obtained supplementary segmentation results are poor, proving that the other two images are not sufficient to replace the original images. Sobel edge images retain the high-frequency component of the images,

which makes the segmentation outputs more focused on the regions with sharp changes in the image gradient. Supersixel images ignore pixels with extremely small grayscale variations, making the segmentation output more focused on the grayscale variations between neighborhoods. By adding two kinds of segmentation outputs, different components are introduced into the method to enrich the features, improving the final segmentation results.

Afterward, the optimization brought by improving the boundary attention region with supplementary segmentation can be validated by the proposed method compared to Comparisons 1 and 4. In addition, this optimization can also be proved by Comparisons 1, 6, and 11, Comparisons 1, 7, and 12, Comparisons 1, 8, and 13 and Comparisons 1, 9, and 14. Additionally, the optimization brought by the devised ELM can also be verified by Comparisons 1 and 6. Approximately 1.53% improvement in IoU, 0.80% improvement in MCC, 0.84% improvement in F1 score and 4.14% improvement in HD95 were achieved by the devised ELM. Then, when Comparisons 6 and 7 are analyzed, the addition of the deep features from Sobel edge images has a negative influence on thyroid segmentation. This is because the Sobel edge images only retain the gradient information of images, and the information is insufficient to obtain an accurate segmentation result. However, when all deep features are selected for supplementary segmentation, the segmentation results are improved. Thus, there are parts of Sobel depth features that do not contribute to the classification of pixel deep features. This can also justify the requirement for feature selection. The improvement of the proposed method caused by mRMR is shown in Comparison 10. Approximately 0.19% improvement in IoU,

Table 4 The comparison results of mean value on four indices

Comparison	Mean			
	IoU	MCC	F1	HD95
Proposed	0.8173	0.8893	0.8980	2.3094
FCN-8s	0.7849	0.8698	0.8777	2.3385
SegNet	0.8084	0.8837	0.8925	2.3291
MGU-NET	0.7657	0.8588	0.8567	2.3452
SV-net	0.8072	0.8838	0.8924	2.3202
VEU-Net	0.7901	0.8722	0.8815	2.3679

0.12% improvement in MCC, 0.10% improvement in F1 score and 1.49% improvement in HD95 is achieved, and this improvement can be verified by Comparisons 4 and 5.

Finally, to judge the performance of the entire thyroid segmentation method, the proposed method is compared with some segmentation method mentioned in Experiments, including FCN-8s, SegNet, MGU-NET, SV-net and VEU-Net. The mean values of the four indices are shown in Table 4. According to these calculation results, the proposed method ranked first on the four indices (marked in bold). The comparison methods, VEU-Net and SegNet, ranked second on two indices (marked as blue). Compared to the comparison methods, the proposed method is 0.0516-0.0089 higher on IoU, 0.0305-0.0055 higher on MCC, 0.0413-0.0055 higher on F1 score and 0.0585-0.0108 better on HD95. In general, combining all the comparison experiments (Tables 3 and 4), the segmentation results of the proposed method have a better overall performance.

4.4 Limitation

Because the ultrasound images utilized in this paper are cutouts provided by physicians, the proposed method has difficulty dealing with raw ultrasound images that are large in size and may have interfering markers. If the raw images need to be processed, target region identification can be added.

5 Conclusion

The purpose of the proposed method is to improve thyroid segmentation in ultrasound images under a small dataset. The proposed method integrates the advantages of deep learning and traditional machine learning. In this paper, a multicomponent dataset, which consists of original images, Sobel edge images and superpixel images, is utilized to improve the final segmentation results. Three kinds of images are used to train three U-Nets to obtain preliminary segmentation outputs, Sobel outputs

and superpixel outputs. The multicomponent features are extracted from the three trained U-Nets and are applied to train the multicomponent neighborhood ELM to acquire the supplementary segmentation results. Meanwhile, the mRMR feature selection algorithm is utilized in the devised ELM to further optimize the subset of neighborhood features. Finally, the precise final segmentation results are obtained by improving the boundary attention region of the preliminary segmentation results. In the proposed method, the mean values of IoU, MCC, F1 score and HD95 are 0.8173, 0.8893, 0.8980 and 2.3094, respectively, which are much better than those of the compared methods. Furthermore, on the three indices, the eight test groups are not only stable but also have better performance than the comparison experiments. Overall, it was demonstrated that by using the proposed method, the segmentation precision of the thyroid can be improved.

Acknowledgements The authors would like to thank the three funds that supported this work: National Natural Science Foundation of China No. 62173116, National Key R&D Program of China No. 2019YFC0117400 and China Scholarship Council No. 202106120077. The authors also would like to express our gratitude to the Detection and Control Laboratory of Harbin Institute of Technology.

Declarations

Conflict of Interests We declare that all authors have no conflict of interest. The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

- Kratzsch J, Pulzer F (2008) Thyroid gland development and defects. *Best Pract Res Clin Endocrinol Metab* 22(1):77–93
- Lopes NMD, Leins H, Armani A et al (2020) Thyroid cancer and thyroid autoimmune disease: a review of molecular aspects and clinical outcomes. *Pathol Res Pract* 216(9)
- Tenda ED, Yulianti M, Asaf MM et al (2020) The importance of chest CT scan in COVID-19. *Acta Med Indones* 52(1):68–73
- Sharifi Y, Sargolzaei M, Bakhshali MA et al (2020) Deep learning on ultrasound images of thyroid nodules. *Biocybernetics and Biomed Eng* 41(2)
- Lee J, Kim B, Park HW (2021) MC2-Net: motion correction network for multi-contrast brain MRI. *Magn Reson Med* 86(2)
- Lu L, Sun M, Lu Q, Wu T, Huang B (2021) High energy X-ray radiation sensitive scintillating materials for medical imaging, cancer diagnosis and therapy. *Nano Energy* 79
- Chandy A (2019) A review on IOT based medical imaging technology for healthcare applications. *J Innov Image Process* 1(01)
- Alexander A, McGill M, Tarasova A, Ferreira C, Zurkiya D (2019) Scanning the future of medical imaging. *J Am Coll Radiol*
- Calle S, Choi J, Ahmed S, Bell D, Learned KO (2021) Imaging of the thyroid: practical approach. *Neuroimaging Clin N Am* 31(3)
- Smith T, Kaufman CS (2021) Ultrasound guided thyroid biopsy. *Tech Vasc Interv Radiol* 2021 24(3):100768

11. Liu S, Wang Y, Yang X et al (2019) Deep learning in medical ultrasound analysis: a review. *Engineering* 5(2)
12. Gulame MB, Dixit VV, Suresh M (2021) Thyroid nodules segmentation methods in clinical ultrasound images: a review. *Mater Today Proc*
13. Garg H, Jindal A (2013) Segmentation of thyroid gland in ultrasound image using neural network. In: 2013 4th International conference on computing, communications and networking technologies
14. Selvathi D, Sharnitha VS (2011) Thyroid classification and segmentation in ultrasound images using machine learning algorithms. In: 2011 - International conference on signal processing, communication, computing and networking technologies
15. Chang CY, Lei YF, Tseng CH, Shih SR (2010) Thyroid segmentation and volume estimation in ultrasound images. *IEEE Trans Biomed Eng* 57(6):1348–1357
16. Gomathy V, Snekhaltha U (2015) Automated segmentation using PCA and area estimation of thyroid gland using ultrasound images. In: 2015 IEEE International conference on innovations in information embedded and communication systems
17. Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E (2018) Deep learning for computer vision: a brief review. *Comput Intell Neurosci* 2018
18. Liu YP, Li Z, Xu C, Li J, Liang R (2019) Referable diabetic retinopathy identification from eye fundus images with weighted path for convolutional neural network. *Artif Intell Med*
19. Lu X, Wang W, Ma C, Shen J, Shao L, Porikli F (2019) See more know more: unsupervised video object segmentation with co-attention siamese networks. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition
20. Lu X, Wang W, Shen J, Crandall D, Luo J (2022) Zero-shot video object segmentation with co-attention siamese networks. *IEEE Trans Pattern Anal Mach Intell*
21. Chen Y, Li D, Zhang X, Jin J, Shen Y (2021) Computer aided diagnosis of thyroid nodules based on the devised small-datasets multi-view ensemble learning. *Med Image Anal* 67(1):101819
22. Nandamuri S, China D, Mitra P, Sheet D (2019) SUMNet: fully convolutional model for fast segmentation of anatomical structures in ultrasound volumes. In: Proceedings - international symposium on biomedical imaging
23. Chu C, Zheng J, Zhou Y (2021) Ultrasonic thyroid nodule detection method based on U-Net network. *Comput Methods Programs Biomed* 199(1):105906
24. Zhu F, Gao Z, Zhao C et al (2021) Semantic segmentation using deep learning to extract total extraocular muscles and optic nerve from orbital computed tomography images. *Optik*
25. Wu X, Bi L, Fulham M et al (2021) Unsupervised brain tumor segmentation using a symmetric-driven adversarial network. *Neurocomputing* 455:242–254
26. Lu X, Wang W, Shen J, Crandall DJ, Van Gool L (2022) Segmenting objects from relational visual data. *IEEE Trans Pattern Anal Mach Intell*
27. Abdel-Basset M, Chang V, Hawash H, Chakraborty RK, Ryan M (2021) FSS-2019-nCov: a deep learning architecture for semi-supervised few-shot segmentation of COVID-19 infection. *Knowl-Based Syst*
28. Guo S, Xu L, Feng C, Xiong H, Gao Z, Zhang H (2021) Multi-level semantic adaptation for few-shot segmentation on cardiac image sequences. *Medical Image Anal*
29. Furnari A, Farinella GM, Bruna AR, Battiato S (2017) Distortion adaptive Sobel filters for the gradient estimation of wide angle images. *J Vis Commun Image Represent* 46(jul.):165–175
30. Angulakshmi M, Lakshmi Priya GG (2019) Walsh hadamard transform for simple linear iterative clustering (SLIC) superpixel based spectral clustering of multimodal MRI brain tumor segmentation. *IRBM*
31. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. *Springer Int Publ*
32. Xiao L, Wang C, Dong Y, Wang J (2019) A novel sub-models selection algorithm based on max-relevance and min-redundancy neighborhood mutual information. *Inf Sci* 486:310–339
33. Salaken SM, Khosravi A, Nguyen T, Nahavandi S (2017) Extreme learning machine based transfer learning algorithms: a survey. *Neurocomputing*
34. Sharifrazi D, Alizadehsani R, Roshanzamir M et al (1988) Fusion of convolution neural network, support vector machine and Sobel filter for accurate detection of COVID-19 patients using X-ray images. *Biomed Signal Process Control*
35. Boemer F, Ratner E, Lendasse A (2018) Parameter-free image segmentation with SLIC. *Neurocomputing*
36. Siddique N, Paheding S, Elkin CP, Devabhaktuni V (2021) U-net and its variants for medical image segmentation: a review of theory and applications. *IEEE Access*
37. Eroglu Y, Yildirim M, Cinar A (2021) Convolutional neural networks based classification of breast ultrasonography images by hybrid method with respect to benign, malignant, and normal using mRMR. *Comput Biol Med* 133(10016):104407
38. Togacar M, Ergen B, Comert Z (2019) Detection of lung cancer on chest CT images using minimum redundancy maximum relevance feature selection method with convolutional neural networks. *Biocybernetics Biomed Eng* 40(1)
39. Law A, Ghosh A (2019) Multi-label classification using a cascade of stacked autoencoder and extreme learning machines. *Neurocomputing* 358(SEP.17):222–234
40. Shelhamer E, Long J, Darrell T (2017) Fully convolutional networks for semantic segmentation. *Fully Convolutional Netw Semant Segmentation*
41. Badrinarayanan V, Kendall A, Cipolla R (2017) SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell*
42. Ilesanmi AE, Kummerle LB, Chaumrattanakul U, Makhanov SS (2021) A method for segmentation of tumors in breast ultrasound images using the variant enhanced deep learning. *Biocybernetics Biomed Eng* 41(2)
43. Moon WK, Lee YW, Ke HH et al (2020) Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. *Comput Methods Prog Biomed* 190:105361
44. Shen T, Gou C, Wang FY, He Z, Chen W (2019) Learning from adversarial medical images for X-ray breast mass segmentation. *Comput Methods Programs Biomed*
45. Dutande P, Baid U, Talbar S (2022) Deep residual separable convolutional neural network for lung tumor segmentation. *Comput Biol Med*
46. Ayhan MS, Kummerle LB, Kuhlewein L, Inhoffen W, Aliyeva G, Ziemssen F, Berens P (2022) Clinical validation of saliency maps for understanding deep neural networks in ophthalmology. *Medical Image Anal*

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



Yifei Chen was born in 1995. He received the B.S. degree and began the Joint Study of Bachelor and Doctoral Degree in 2017 from Harbin Institute of Technology, Harbin, China. His research interests include ultrasound image analysis, computer aided diagnosis, image segmentation and classification, machine learning and deep learning.



HyunWook Park has been a professor of the school of electrical engineering at Korea Advanced Institute of Science and Technology (KAIST), Korea since 1993. He received the BS degree in electrical engineering from Seoul National University, Seoul, Korea in 1981 and the MS and PhD degrees in electrical engineering from KAIST, Seoul, Korea in 1983 and 1988, respectively. His current research interests include image computing system, image compression, and medical imaging.



Xin Zhang was born in 1987. He received the B.S. degree from the Harbin Engineering University, Harbin, China, in 2009, and the M.S. and Ph.D. degrees from Harbin Institute of Technology, Harbin, in 2011 and 2015, respectively. He is now an Associate Professor with the Department of Control Science and Engineering at Harbin Institute of Technology. His research interests include AE detection in rail, signal processing, and pattern recognition.



Xinran Li was born in 2002. She is currently pursuing the B.S. degree with Mathematics and Applied Mathematics, Harbin Institute of Technology, Harbin, China. Her research interests include feature extracting, image processing and intelligent diagnosis of ultrasonic thyroid standard planes based on machine learning.



Dandan Li was born in 1978. She received the B.S., M.S., and Ph.D. degrees from Harbin Institute of Technology, Harbin, China, in 1997, 1999, and 2005. She is the technical leader of HIT-Nvidia AI and Control Joint Lab. She has profound project experience in medical image processing, intelligent diagnosis and intelligent curling robot, and her current research focuses on the application and algorithm research of artificial intelligence.



Peng Liu was born in 1978. She received the B.S., M.S., and M.D. degrees from Harbin Medical University, Harbin, China, in 2001, 2006, and 2009. She is a professor and director of the Institute for the Prevention and Control of Iodine Deficiency Diseases, Center for Endemic Disease Control, Chinese Center for Disease Prevention and Control. Her research interests include thyroid pathophysiology, iodine deficiency disease prevention and control.




Jing Jin was born in 1980. She received the B.S and Ph.D. degrees from Harbin Institute of Technology, Harbin, China, in 2003 and 2008. She is a Professor with the Department of Control Science and Engineering, Harbin Institute of Technology, and the Deputy Dean of the Harbin Institute of Technology for Artificial Intelligence. Her research interests include artificial intelligence, medical image diagnosis, intelligent curling robot, brain-computer interface and navigation technology.



Yi Shen (Member, IEEE) was born in 1965. He received the B.S., M.S., and Ph.D. degrees from Harbin Institute of Technology, Harbin, China, in 1985, 1988, and 1995, respectively. He is a Professor with the Department of Control Science and Engineering, Harbin Institute of Technology. His research interests include ultrasound signal processing, medical image analysis, fault diagnosis for control systems and flight vehicle control.

Affiliations

Yifei Chen^{1,2} · Xin Zhang¹  · Dandan Li¹ · HyunWook Park² · Xinran Li³ · Peng Liu⁴ · Jing Jin¹ · Yi Shen¹

Yifei Chen
cyf19950405@126.com

Dandan Li
beauty@hit.edu.cn

HyunWook Park
hwpark@kaist.ac.kr

Xinran Li
lixinran0708@gmail.com

Peng Liu
liup7878@163.com

Jing Jin
jinjinghit@hit.edu.cn

Yi Shen
yishen_hit@126.com

- ¹ Control Science and Engineering, Harbin Institute of Technology, Harbin, 150001, China
- ² Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon, 34141, Korea
- ³ Mathematics, Harbin Institute of Technology, Harbin, 150001, China
- ⁴ Heilongjiang Provincial Key Laboratory of Trace Elements and Human Health, Harbin Medical University, Harbin, 150081, China