Nodule Localization in Thyroid Ultrasound Images with a Joint-Training Convolutional Neural Network

Ruoyun Liu^{1,2} · Shichong Zhou^{3,4} · Yi Guo^{1,2} · Yuanyuan Wang^{1,2} · Cai Chang^{3,4}

Published online: 30 June 2020

© Society for Imaging Informatics in Medicine 2020

Abstract

The accurate localization of nodules in ultrasound images can convey crucial information to support a reliable diagnosis. However, this is usually challenging due to low contrast and image artifacts, especially in thyroid ultrasound images where nodules are relatively small in most cases. To address these problems, in this paper, we propose a joint-training convolutional neural network (CNN) for thyroid nodule localization in ultrasound images. Considering the advantage of the faster region-based CNN (Faster R-CNN) in detecting natural targets, we adopt it as the basic framework. To boost the representative power and noise suppression capability of the network, the attention mechanism module is embedded for adaptive feature refinement along the channel and spatial dimensions. Furthermore, in the training process, we annotate the training set in a novel way, called joint-training annotation, by exploiting the fake foreground (FFG) area around the nodule as a spatial prior constraint to improve the sensitivity to small nodules. Ablation experiments are conducted to verify the effectiveness of our proposed method. The experimental results show that our method outperforms others by a mean average precision (mAP) of 0.93 and achieves an intersection over union (IoU) of 0.9, indicating that the localization results agree well with the ground truth. Furthermore, extended experiments on breast nodule datasets are also conducted to verify the generalizability of the proposed approach. Above all, the proposed algorithm is of considerable significance for accurate thyroid nodule localization in ultrasound images and can be generalized to other types of nodules, thereby providing trustworthy assistance for clinical diagnosis.

Keywords Thyroid nodule localization · Ultrasound images · Attention mechanism module · Joint-training annotation · Fake foreground

Ruoyun Liu and Shichong Zhou contributed equally to this work.

✉ Yi Guo
guoyi@fudan.edu.cn

✉ Yuanyuan Wang
yywang@fudan.edu.cn

¹ Department of Electronic Engineering, Fudan University, No. 220, Handan Road, Yangpu District, Shanghai 200433, China

² Key Laboratory of Medical Imaging Computing and Computer Assisted Intervention (MICCAI) of Shanghai, Shanghai 200032, China

³ Department of Ultrasound, Fudan University Shanghai Cancer Center, Shanghai, China

⁴ Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China

Introduction

Thyroid cancers are nodules of the thyroid gland and originate from follicular or parafollicular cells. In most cases, a definitive etiology for thyroid cancer is not known, but radiation exposure (especially to children) and some hereditary conditions are associated with an increased risk for developing thyroid cancer. Early diagnosis of thyroid cancer contributes to the process of further disease treatment and improves survival expectancy [1–5]. Ultrasound has become one of the most commonly performed imaging modalities in clinical practice for detecting thyroid nodules and can provide crucial information to support reliable clinical diagnosis and disease evaluation [6]. To quantitatively assess the characteristics of nodules, lesions should first be separated from the background [7]. Hence, accurate localization of the thyroid nodule is generally a prerequisite for computer-aided disease segmentation, tracking, and diagnosis. In clinical studies, this process is usually performed manually and is time consuming, tedious, and

highly dependent on the experience of radiologists [8–11]. To improve diagnostic performance and reduce human intervention, the demand for computer-aided medical image processing is increasing [12]. Nonetheless, the precise localization of nodules in ultrasound images is still challenging due to low contrast and a substantial number of interferential image artifacts, such as shadowing and speckle noise. In addition, thyroid nodules are relatively small in many cases, which also increases the difficulty of accurate localization. As shown in Fig. 1, the left image presents the thyroid ultrasound image, and the right image is the expected nodule location result.

Deep learning is a representation learning method that can automatically extract complex features from the raw data that are suited to a particular task [13]. In recent years, deep learning has become a dominant research method in clinical fields. Several localization and detection approaches based on the convolutional neural network (CNN) have been introduced to medical imaging [14, 15]. Hussain et al. [16] applied an effective deep CNN-based approach for tight kidney ROI localization in CT images, where an aggregated orthogonal decision CNN performing voxel-wise predictions was proposed to locate kidneys. In [17], a learning-based approach to locate the fetal abdominal standard plane in US videos by constructing a domain-transferred deep CNN was employed by Chen et al. More recently, some excellent object detection networks, such as faster region-based CNN (Faster R-CNN) [18] and you only look once (YOLO) [19], have been presented. Compared with CNN-based algorithms, Faster R-CNN proposes a region proposal network (RPN) and integrates feature extraction, candidate box extraction, boundary box regression, and classification into a network; this greatly improves the detection performance on natural targets [20–25]. Ribli et al.

[26] optimized both the object detection and classifier part of the Faster R-CNN model to locate lesions in mammograms. They also used back propagation and stochastic gradient descent with weight decay to fine-tune the training process. However, for thyroid ultrasound images, the low contrast and massive number of artifacts interfere with the feature presentation procedure of the network. Moreover, there is minimal feature information that can be learned and extracted for the localization of small targets, such as thyroid nodules. Relatedly, Chi et al. [27] propose a thyroid image classification method by fine-tuning the existing deep convolutional neural network GoogLeNet. Ma et al. [28] employ a hybrid CNN-based method to detect thyroid nodules from 2D ultrasound images, which is a cascade network based on a special splitting method and two CNN architectures with different layers. Although these researchers have yielded encouraging results, their methods place greater emphasis on segmentation and classification.

To address the aforementioned difficulties, in this paper, we propose a joint-training CNN for thyroid nodule localization in ultrasound images. Faster R-CNN is adopted as our basic framework. However, different from the traditional Faster R-CNN method, we introduce the attention mechanism module to the baseline network. Attention has an important influence on vision, and it is verified that incorporating an attention mechanism into basic CNN architectures can enhance the feature representation ability of the network, thereby improving the performance of different tasks [29]. Hence, we plug the attention mechanism module into the convolutional blocks and annotate the training dataset in a novel way to obtain more accurate localization results for thyroid nodules. Our main contributions are summarized as follows:

Fig. 1 Accurate thyroid nodule localization result. **a** and **b** present the thyroid ultrasound image and the expected localization result, respectively

1. To enhance the anti-interference ability, the attention mechanism module is introduced to the network, which can achieve adaptive feature refinement and boost the representation power of the network.
2. To improve the sensitivity to small nodules, we annotate the training dataset in a novel way called joint-training annotation, which can utilize the fake foreground (FFG) area around the nodule as the spatial prior constraint and capture more information to obtain more accurate thyroid nodule localization results for small targets.
3. After training the model on the thyroid nodule dataset, we also conduct an extended experiment on a breast nodule dataset to verify the adaptability of the proposed method.

The remainder of this paper is organized as follows. **Methods** section provides the principle of our joint-training CNN algorithm. The materials and experimental procedures are described in **Experiments** section. In **Results and Discussion** section, a discussion of the experimental results is presented. **Conclusion and Future Work** section ends with a summary of this paper and outlook for future research.

Methods

A schematic representation of the proposed method is shown in Fig. 2. Due to the low contrast of some thyroid ultrasound images, they are first processed by adaptive preprocessing operations, where adaptive histogram equalization [30] is adopted to increase the global contrast. Then, all the thyroid ultrasound images are annotated in a special way called joint-

training annotation before being sent to the proposed joint-training CNN. In the testing stage, the thyroid nodule localization results are directly generated by the trained model.

Adaptive Preprocessing

Histogram equalization methods have been widely used in the field of image processing for improving both contrast and structure visibility, therein utilizing the overall intensity distribution in the image as characterized by the normalized cumulative histogram [31]. Since some thyroid nodule ultrasound images have very low contrast, the nodules are poorly recognized. Hence, these images are preprocessed by adaptive histogram equalization to improve the image quality before training. We adopt the statistical mean pixel intensity of 50 as the threshold. If 80% of the pixel intensities in an ultrasound image are below this threshold, the operation is necessary. Visual examples are shown in Fig. 3, which also shows a comparison of the original image and the preprocessed image.

Joint-Training CNN

According to the characteristics of ultrasound images and the nature of thyroid nodules, we propose a joint-training CNN for thyroid nodule localization in ultrasound images. Faster R-CNN is adopted as the basic framework, and the attention mechanism module is introduced to the baseline network ResNet50 [32] to enhance the noise suppression capability of the network. Furthermore, in the training process, we annotate the training set in a novel way called joint-training annotation to improve the sensitivity to small nodules and

Fig. 2 Overview of the proposed method

Fig. 3 Thyroid ultrasound images with different contrasts. **a** displays the image requiring adaptive histogram equalization. However, for **b**, this operation is unnecessary. **c** presents the adaptive preprocessed image

obtain more accurate thyroid nodule localization results. Our network structure is demonstrated in Fig. 4. As shown in the figure, the input images are initially processed by a 77 convolutional layer, followed by batch normalization (BN) and rectified linear unit (ReLU). The conv block and identity block are basic blocks in ResNet, as described in [32]. The number of feature maps sequentially increases from the lower to higher layer. Finally, the output is generated by a fully connected layer.

Attention Mechanism Module

Since convolution operations extract information features by mixing channel and spatial information together, the feature learning ability of the network is limited due to the strong noise and interference in the thyroid ultrasound image. Recently, [29] proposed a general module called the convolutional block attention module (CBAM) for adaptive feature refinement in the training procedure. Inspired by their

Fig. 4 The network structure of the proposed joint-training CNN (where k , n , and s denote the kernel size, number of feature maps, and strides, respectively)

Fig. 5 Attention mechanism module integrated with a ResBlock in ResNet

work, to increase the representation power and noise suppression capability of the proposed network, we introduce an attention mechanism module to enhance features meaningful for classification along the channel and spatial dimensions.

For an input thyroid ultrasound image, the convolution layer is first exploited to extract its feature maps. Then, this feature map is fed into the attention mechanism module to generate two attention maps in the channel and spatial dimensions in succession. The channel attention part of this module focuses on “what” is meaningful to a given input image, and the spatial attention one concentrates on “where” is an informative part [29]. These two parts complement each other. Extending the previous work described in [29], we add a comparison process to the module to strengthen the robustness of the selected features. It is known that in the binary classification problem, the features within one class are similar, while the correlation between two classes is relatively low. Thus, we compare the feature maps extracted from the predicted nodule area and the background in order to ensure the accuracy of the feature refinement operation. Specifically, the correlation coefficient of the two sets of feature maps is calculated, and those features with smaller correlations (i.e., those for which the correlation coefficient is less than 0.2, the threshold for relevance) are selected for weighting. Then, the weighted attention maps are multiplied consecutively to the input feature maps for the adaptive learning of features. The entire attention process can be described as follows:

$$F' = M_C(F) \times F \quad (1)$$

Fig. 6 An example of the joint-training annotation method

$$F'' = M_S(F') \times F' \quad (2)$$

where F' and F'' denote the input feature map and the enhanced output feature map, respectively; M_C is the channel attention map; and M_S is the spatial attention map.

To compute the attention maps more efficiently, both the average-pooling and max-pooling operations are used to aggregate feature information about the thyroid nodule and highlight the informative regions in the ultrasound image. During multiplication, the attention values are broadcast along the two dimensions of channel and space; then, following the comparison process, robust features that are meaningful for classification in these feature maps can be enhanced. Figure 5 depicts a diagram of the attention mechanism module integrated with a ResBlock in ResNet50 as an example.

Moreover, combining the attention mechanism module with the convolutional block in ResNet50 can be used to exploit the interchannel relationship of features, thus improving the feature representation power of the network model and further increasing the accuracy of localization.

Joint-Training Annotation

Nodules in thyroid ultrasound images are relatively small in many cases. Hence, to improve the sensitivity to small thyroid nodules and obtain more accurate localization results, we propose an innovative method called joint-training annotation to label the training dataset. This method can utilize the neighboring area around the nodule, which we called the FFG area, as the spatial prior constraint, and capture more information to assist the localization. Traditionally, a bounding box drawn tightly around the nodule is used as the input. However, the feature information in the bounding box is limited for small nodules. From a clinical point of view, the surrounding peripheral microenvironments are meaningful for the nodule. Hence, we add a supplementary part of the outer area as the FFG in the process of making the training set. The size of the FFG area is randomly selected, which ranges from 20 to 50% of the bounding box. More specifically, the coordinates of the bounding box drawn tightly around the nodule are $[x_1, x_2, y_1, y_2]$, and the coordinates of FFG are $[x_3, x_4, y_3, y_4]$, where $\frac{x_2 - x_1}{x_4 - x_3} \in [\frac{6}{5}, \frac{3}{2}]$ and $\frac{y_2 - y_1}{y_4 - y_3} \in [\frac{6}{5}, \frac{3}{2}]$. In this way, the original binary classification problem is converted into a detection problem of three categories, and these two labels are

Fig. 7 Thyroid nodule localization results with/without the adaptive preprocessing method. **a** presents the localization result without this operation where the nodule is unable to be detected. **b** presents the preprocessed one

interrelated, which can be utilized as the implicit spatial prior constraint information. As a result, even a relatively small nodule, which is prone to be ignored, can also be identified accurately in the localization procedure. Figure 6 depicts an example of a joint-training annotation method.

Experiments

In this section, a set of experiments is designed to verify the effectiveness of the proposed joint-training CNN method.

Dataset

A total of 500 thyroid ultrasound images were collected from the Department of Ultrasound, Fudan University Shanghai Cancer Center, China. These images were collected on different scanners, and the image size was $256 \times$

256 pixels. The data used in this article are all ultrasound images of malignant thyroid nodules. Among the data, 251 images were classified as small nodules (i.e., with a maximum diameter of less than 2 cm), 193 samples were classified as medium-sized nodules (i.e., with a maximum diameter between 2 and 4 cm), and the remaining 56 samples were classified as large nodules (i.e., with a maximum diameter of greater than 4 cm). Manual delineation by experienced doctors served as the ground truth, and we converted this annotation to bounding box coordinates. Due to limitations in the amount of available data, in this study, we conducted a tenfold cross-validation with the same dataset. Specifically, we randomly divided the original dataset into ten nonoverlapping groups to avoid intersection of data from the same patient across folds. Among these, 8 were utilized for training, one for validation, and one for testing, in which the validation set was used to tune the parameters of the CNNs, and the testing set was used to

Table 1 Comparison of the thyroid nodule localization results from different methods based on mAP value

Method	Joint-training CNN	Joint-training CNN without joint-training annotation	Faster R-CNN	YOLO
IoU threshold	mAP			
0.50	0.91 ± 0.05	0.82 ± 0.07	0.68 ± 0.10	0.63 ± 0.07
0.75	0.90 ± 0.08	0.77 ± 0.14	0.53 ± 0.15	0.30 ± 0.11
0.90	0.87 ± 0.11	0.68 ± 0.17	0.36 ± 0.19	0.11 ± 0.19

Fig. 8 Comparative results for two methods. **a** and **b** illustrate the thyroid nodule localization result of basic Faster R-CNN and YOLO, respectively. **c** serves as the ground truth

assess the performance of the model [28]. Here, 400 samples were chosen randomly as the training dataset, 50 images comprised the validation dataset, and the remaining 50 images were selected for testing.

Implementation

In the training process, an ImageNet-pretrained ResNet50 is employed as our baseline network. We train the model using an Adam optimizer [33] with weight decay = 10^{-5} , batch size = 10, and epoch = 100.

The preprocessing step and adaptive histogram equalization are implemented using MATLAB. Then, the entire training and testing process is implemented using the TensorFlow library [34] with an Nvidia GTX 1080Ti graphics card to increase the training speed.

Comparing Algorithms

In this subsection, we empirically show the effectiveness of our design network through comparison with a YOLO [19] network and ablation studies. For the ablation studies, we successively analyze the contribution of each component, including CBAM and joint-training annotation. The compared algorithms are as follows.

- 1) Faster R-CNN [18]
- 2) YOLO [19]
- 3) Joint-training CNN without joint-training annotation
- 4) Joint-training CNN

To ensure the fairness of the comparative experiments, the same optimizer, decay rate, batch size, and number of epochs are chosen for these four experimental methods, which are

mentioned before. The training dataset and testing dataset are also identical.

Extended Experiment

To verify the generalization of our trained model, we perform the training procedure on the thyroid nodule dataset while validating the testing experiments on both the thyroid nodule and breast nodule datasets. The breast nodule testing dataset includes 300 ultrasound images, also collected from the Department of Ultrasound, Fudan University Shanghai Cancer Center, China.

Results and Discussion

Performance of the Adaptive Preprocessing

In the proposed joint-training CNN method, some of the thyroid ultrasound images adopt adaptive preprocessing by adaptive histogram equalization to improve the image quality. We compare the nodule localization results obtained with/without the adaptive preprocessing in Fig. 7. Before adaptive preprocessing, processing of 132 thyroid ultrasound images in the training dataset and 19 images in the testing dataset failed to generate the localization results. After this operation, the number of thyroid nodules that were unable to be detected decreased to 82 and 14, respectively. The statistical improvement demonstrates that the proposed adaptive preprocessing operation is necessary for the chosen low-contrast ultrasound images and improves the overall localization performance.

Fig. 9 Comparison of the thyroid nodule localization results acquired with/without the attention mechanism module. **a** illustrates the results with the proposed attention mechanism module, **b** displays the results with CBAM, and **c** displays the results of baseline network. **d** serves as

the ground truth. The first and second rows present the general thyroid ultrasound image and the image with strong noise and shadows, respectively. The third row depicts the incomplete thyroid nodule localization result

Thyroid Nodule Localization

For detecting natural images, the mean average precision (mAP) values differ greatly over different intersection over union (IoU) thresholds, which are larger than 0.50. Thus, to better evaluate the localization capability and robustness of these algorithms, we use the average mAP

over different IoU thresholds from 0.50 to 0.90 for evaluation. Table 1 presents the comparison of mAP values for the four abovementioned methods at different thresholds. As shown in the figure, compared to other methods, our proposed method achieves better localization accuracy and recall rate for most images; thus, the mAP values show almost no change in the 0.50 to 0.90 interval, which

Table 2 Comparison of thyroid nodule localization results of different methods based on mAP value

Method	Joint-training CNN	Baseline with CBAM	Joint-training CNN without attention mechanism module
IoU threshold	mAP		
0.50	0.93 ± 0.05	0.85 ± 0.08	0.66 ± 0.11

Fig. 10 Comparison of the thyroid nodule localization results obtained with/without joint-training annotation method. **a** presents the results with the joint-training annotation method and **b** displays the results without it.

c serves as the ground truth. The first, second and third rows illustrate large, medium and small sizes of nodules in ultrasound images, respectively

decreases from 0.91 ± 0.05 to 0.90 ± 0.08 and to 0.87 ± 0.11 , indicating the superiority of our proposed method in terms of IoU and mAP. Without the joint-training annotation, the mAP values decrease from 0.82 ± 0.07 to 0.77 ± 0.14 and to 0.68 ± 0.17 when the IoU threshold increases from 0.50 to 0.75 and to 0.90. In addition, the localization results of other methods fluctuate greatly in terms of numerical values. For Faster R-CNN, the mAP values fall sharply from 0.68 ± 0.10 to 0.53 ± 0.15 and to 0.36 ± 0.19 . For YOLO, the values drop from 0.63 ± 0.07 to 0.30 ± 0.11 and to 0.11 ± 0.19 , illustrating that the localization accuracy is relatively low.

Comparison with YOLO v3

To determine a more suitable basic framework for the thyroid nodule localization task, we first compare the experimental results of the basic Faster R-CNN and YOLO, which are both state-of-the-art object detection methods. Visual examples are provided in Fig. 8.

As shown in Table 1 and Fig. 8 above, the conventional Faster R-CNN has higher localization accuracy than YOLO and acquires a more complete nodule localization result, which is more similar to the ground truth. This indicates that the number of proposed regions extracted

Fig. 11 Three different methods of annotation. **a**, **b**, and **c** display the fixed 20% annotation method, the fixed 50% annotation method, and the proposed randomly annotation method, respectively

by YOLO is much lower than Faster R-CNN. Moreover, the RPN design of the Faster R-CNN is equivalent to estimating each position of the feature map using a sliding window, which provides a facility for it to excel. Compared to its counterparts, Faster R-CNN optimizes the generation of region proposals, making it truly end-to-end training to ensure improved performance. In contrast, the YOLO network simplified the network and performs the target detection as a regression problem, and after inferring the input image, it obtains the position of all objects in the image as well as its category and the corresponding confidence probability. In contrast, Faster R-CNN takes a separate module to implement these steps. It divides the detection process into two parts: the object category and the object position. The model training process of the YOLO method relies more on the annotation data; thus, if the object to be detected has an unconventional shape or scale, the localization results of YOLO are not ideal. Because the YOLO network utilizes multiple downsampling layers, the feature extraction procedure of the network is blocked, leading to poor localization performances and thus reducing the localization accuracy. Therefore, the accuracy and mAP value of YOLO decrease considerably. Based on the above reasons, we ultimately adopt Faster R-CNN as our basic framework.

Table 3 Comparison of the different annotation methods based on mAP value

Method	20% FFG	50% FFG	[20%,50%] randomly selected
mAP	0.86 ± 0.03	0.90 ± 0.02	0.93 ± 0.05

Performance of the Attention Mechanism Module in the Joint-Training CNN

As mentioned previously, the feature learning ability of the network is limited due to the strong noise and interference in the thyroid ultrasound image. In our proposed joint-training CNN, the proposed attention mechanism module is plugged into the network to weight meaningful features for classification and enhance the representation power of the network. We compare the experimental results acquired with the proposed attention mechanism module, the localization results of baseline network with CBAM module, and the results without any attention mechanism module (baseline network) in Fig. 9. Furthermore, Table 2 presents the thyroid nodule localization performances of different annotation methods in terms of the mAP value, where the threshold is 0.50.

As shown in Fig. 9, for the general thyroid ultrasound image, these three methods achieve accurate nodule localization compared to the ground truth. For ultrasound images with strong noise and shadows, the added attention mechanism module can improve the localization confidence of the network through adaptive feature refinement and reduce the number of nodules unable to be detected. As shown in the third row of Fig. 9, for the incomplete localization result, this module improves the integrity of thyroid nodule localization. The proposed attention mechanism module outperforms CBAM because of its acquisition of more robust features. Furthermore, as shown in Table 2, the localization results for these three methods are 0.660.11, 0.850.08, and 0.930.05, respectively. The results indicate that the proposed method exhibits better thyroid nodule localization performance both visually and statistically.

Above all, the proposed attention mechanism module introduced to the baseline network can obtain adaptive feature

Table 4 Comparison of the thyroid nodule segmentation results for the different methods based on DICE value

Method	Joint-training CNN	Joint-training CNN without joint-training annotation	Faster R-CNN	YOLO
DICE	0.88 ± 0.07	0.76 ± 0.05	0.66 ± 0.11	0.41 ± 0.21

refinement, boost the representation power of the network, and enhance the anti-interference ability, thus improving the overall thyroid nodule localization performance.

Performance of the Joint-Training Annotation

As mentioned above, the joint-training annotation is adopted to improve the sensitivity to small thyroid nodules in ultrasound images. This method can utilize the FFG area around the nodule as a spatial prior constraint and capture more information to obtain more accurate thyroid nodule localization results for small targets. Representative examples are shown in Fig. 10. For the same image, the proposed method obtains more accurate localization results for small targets both visually and quantitatively.

As shown in Fig. 10, for large- or medium-sized nodules, the thyroid nodule localization results are similar to the ground truth. However, for small thyroid nodules, utilization of the proposed joint-training annotation improves sensitivity and yields accurate localization results.

Moreover, to determine the most appropriate method for FFG area selection, we compare the experimental results in three ways. In our proposed method, the size of the FFG area is randomly selected, ranging from 20 to 50% of the bounding box. Additionally, fixed 20% and 50% of the range are selected as our comparison methods, respectively. Figure 11 presents examples of three annotation methods, and Table 2 displays the thyroid nodule localization performances of the different annotation methods in terms of the mAP value, where the threshold

Fig. 12 Segmentation results with the proposed joint-training CNN. The red area serves as the ground truth, the green area is the segmentation result, and the white area in the middle is the overlap

Fig. 13 Localization results of breast ultrasound images

is 0.5. As shown in Table 3, the localization results by fixed 20% FFG, fixed 50% FFG, and randomly selected FFG are 0.86 ± 0.03 , 0.90 ± 0.02 , and 0.93 ± 0.05 , respectively. The results indicate that the proposed randomly selected method has superior thyroid nodule localization performance, and this selection is the most appropriate selection from both a clinical perspective and comparative experimental results.

Auxiliary to Segmentation Experiment

To illustrate the effectiveness of the proposed method, we utilized the thyroid nodule localization results of the aforementioned four methods to facilitate the segmentation experiment. The U-net framework is adopted for segmentation, and the model was trained with our training dataset. We regard the localization results as constraints and utilize the area tightly around the bounding box to conduct the testing experiments and obtain the segmentation results. The DICE coefficient is adopted as the evaluation index, and the performance is described in Table 4. Our proposed method has a DICE coefficient of 0.88 ± 0.07 , which is the highest among these four methods. The segmentation results confirm that our proposed method has a meaningful auxiliary contribution to the subsequent segmentation experiment and outperforms

other methods in terms of the evaluation index. Visual examples of the segmentation results with our proposed method are presented in Fig. 12.

Extended Study

To further illustrate the adaptability of our proposed method, we applied the trained model to 300 ultrasound images of breast nodules and performed extended testing experiments. The final result of the mAP value is approximately 0.84 ± 0.07 , which proves that our method has the ability to roughly locate breast nodules in ultrasound images, and so may be more extensively generalizable for the detection of lesions in ultrasound images of other anatomic structures. Figure 13 depicts some nodule localization results of breast ultrasound images, indicating that the proposed method has the potential and flexibility to be applied to other ultrasound image datasets.

Conclusion and Future Work

In summary, a novel joint-training CNN algorithm with an effective feature utilization method for accurate thyroid nodule localization in ultrasound images is proposed. First, after adaptive preprocessing, an attention mechanism module is

introduced to the network to improve the feature representation power and the anti-interference performance of the network. Second, the training dataset is annotated in a special way, called joint-training annotation, which utilizes the features of the neighboring FFG area to yield more accurate localization results of the thyroid nodule in ultrasound images. Finally, we verify the effectiveness and robustness of the proposed method through extensive experiments, and the results illustrate its superiority in terms of both mAP and IoU. Moreover, the testing experiment on a breast nodule dataset is also conducted to further illustrate the adaptability and flexibility of our trained model, demonstrating that the proposed approach can easily be extended to other, similar nodule ultrasound images.

In future work, other types of nodule ultrasound images or videos may be applied to verify the adaptability of the proposed algorithm. Extended applications, such as segmentation and tracking, will also be explored in future work.

Funding Information This work was funded by the National Natural Science Foundation of China (61871135, 81830058, 81627804) and the Science and Technology Commission of Shanghai Municipality (18511102904, 17411953400).

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Ethical Approval This article does not contain any studies with human participants performed by any of the authors.

References

- Haugen BR, Alexander EK, Bible KC: 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association Guidelines Task Force on thyroid nodules and differentiated thyroid cancer. *Thyroid* 26(1):1–133, 2016
- Gharib H, Papini E, Paschke R: Medical guidelines for clinical practice for the diagnosis and management of thyroid nodules. *Endocr Pract* 12(1):63–102, 2006
- Kloos RT, Eng C, Evans DB: Medullary thyroid cancer: management guidelines of the American Thyroid Association. *Thyroid* 19(6):565–612, 2009
- Haugen BR, Alexander EK, Bible KC: The American Thyroid Association (ATA) guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid* 26(1):1–133, 2016
- Gautherie, Michel: Thermobiological assessment of benign and malignant breast diseases. *Am J Obstet Gynecol* 147(8):861–869, 1983
- Gharib H, Papini E, Paschke R: American Association of Clinical Endocrinologists, Associazione Medici Endocrinologi, and European Thyroid Association medical guidelines for clinical practice for the diagnosis and management of thyroid nodules: executive summary of recommendations. *J Endocrinol Investig* 33:287, 2010
- Yap MH, Edirisinghe E, Bez H: Processed images in human perception: a case study in ultrasound breast imaging. *Eur J Radiol* 73(3): 682–687, 2010
- Calas MJG, Almeida RMVR, Gutfilen B, Pereira WCA: Intraobserver interpretation of breast ultrasonography following the bi-rads classification. *Eur J Radiol* 74(4):525–528, 2010
- Yap, Moi Hoon, Edirisinghe E, Bez H: Processed images in human perception: a case study in ultrasound breast imaging. *Eur J Radiol* 73(11):682–687, 2010
- Chang RF, Wu WJ, Moon WK, Chen DR: Improvement in breast nodule discrimination by support vector machines and speckle-emphasis texture analysis. *Ultrasound Med Biol* 29(5):679–686, 2003
- Noble JA, Boukerroui D: Ultrasound image segmentation: a survey. *IEEE Trans Med Imaging* 25(8): 987–1010, 2006
- Lee H, Chen YPP: Image based computer aided diagnosis system for cancer detection. *Expert Syst Appl* 42(2):5356–5365, 2015
- LeCun Y, Bengio Y, Hinton G: Deep Learning. *Nature* 521(5): 436–444, 2015
- Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I: Deep convolutional neural networks for computer-aided detection: cnn architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging* 35(5):1285–1298, 2016
- Litjens G, Kooi T, Bejnordi BE, Setio A, Ciompi F, Ghafoorian M: A survey on deep learning in medical image analysis. *Med Image Anal* 42(12):60–88, 2017
- Hussain MA, Amir-Khalili A, Hamarneh G, Abugharbieh R: Segmentation-free kidney localization and volume estimation using aggregated orthogonal decision cnns. In: *International Conference on Medical Image Computing & Computer-assisted Intervention*. Springer, Cham, 2017, pp 612–620
- Chen H, Ni D, Qin J, Li SL, Yang X, Wang TF: Standard plane localization in fetal ultrasound via domain transferred deep neural networks. *IEEE J Biomed Health Inform* 19(4):1627–1636, 2015
- Ren S, He K, Girshick R, Sun J: Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell* 39(6): 1137–1149, 2017
- Redmon J, Divvala S, Girshick R, Farhadi A: You Only Look Once: Unified, Real-Time Object Detection. In: *2016 IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp 779–788
- Fang J, Cheng J: Target detection and recognition based on improved Faster R-CNN. *J Image Signal Process* 8(1):43–50, 2019
- Szegedy C, Ioffe S, Vanhoucke V, Alemi A: Inception-v4, inception-resnet and the impact of residual connections on learning. *arXiv:1602.07261*, 2016
- Wang X, Girshick R, Mulam H, He K: Non-Local Neural Networks. *arXiv:1711.07971*, 2017
- Wei SE, Ramakrishna V, Kanade T, Sheikh Y: Convolutional Pose Machines. *arXiv:1602.00134*, 2016
- Neubeck A, Van Gool L: Efficient non-maximum suppression. In: *18th International Conference on Pattern Recognition (ICPR)*. 2006, pp 850–855
- Girshic R, Donahue J, Darrell T: Rich feature hierarchies for accurate object detection and semantic segmentation. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2014, pp 580–587
- Ribli D, Horváth A, Unger Z, Pollner P, Péter, Csabai I: Detecting and classifying lesions in mammograms with deep learning. *Sci Rep* 8:4165, 2018
- Chi J, Walia E, Babyn P, Wang J, Eramian M: Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network. *J Digit Imaging* 30(3):477–486, 2017

28. Ma J, Wu F, Jiang T, Zhu J, Kong D: Cascade convolutional neural networks for automatic detection of thyroid nodules in ultrasound images. *Med Phys* 44(5):1678-1691, 2017
29. Woo S, Park J, Lee JY: CBAM: Convolutional Block Attention Module. *arXiv:1807.06521*, 2018
30. Pizer SM, Amburn EP, Austin JD, Cromartie R, Zuiderveld K: Adaptive histogram equalization and its variations. *Comput Vis Graphics Image Process* 39(9):355-368, 1987
31. Ploquin M, Basarab A, Kouamé D: Resolution enhancement in medical ultrasound imaging. *J Med Imaging* 2(1):017001, 2015
32. He K, Zhang X, Ren S, Sun J: Deep residual learning for image recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp 770-778
33. Kingma D, Ba J: Adam: A method for stochastic optimization. *arXiv:1412.6980*, 2015
34. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J: TensorFlow: a system for large-scale machine learning. In: *Conference on Operating Systems Design and Implementation*. 2016, pp 265-283

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