MTN-Net: A Multi-Task Network for Detection and Segmentation of Thyroid Nodules in Ultrasound Images

Leyao Chen, Wei Zheng, and Wenxin Hu(✉)

School of Data Science and Engineering, East China Normal University,
Shanghai, China
wzheng@admin.ecnu.edu.cn, wxhu@cc.ecnu.edu.cn

Abstract. Automatic detection and segmentation of thyroid nodules is crucial for the identification of benign and malignant nodules in computer-aided diagnosis (CAD) systems. However, the diverse sizes of thyroid nodules in ultrasound images, nodules with complex internal textures, and multiple nodules pose many challenges for automatic detection and segmentation of thyroid nodules in ultrasound images. In this paper we propose a multi-task network based on Trident network, called MTN-Net, to accurately detect and segment the thyroid nodules in ultrasound images. The backbone of MTN-Net can generate scale-specific feature maps through trident blocks with different receptive fields to detect thyroid nodules with different sizes. In addition, a novel semantic segmentation branch is embedded into the detection network for the task of segmenting thyroid nodules, which is also valid for the complete segmentation of nodules with complex textures. Furthermore, we propose an improved NMS method, named TN-NMS, for combining thyroid detection results from multiple branches, which can effectively suppress falsely detected internal nodules. The experimental results show that MTN-Net outperforms the State-of-the-Arts methods in terms of detection and segmentation accuracy on both the public TN3K dataset and the public DDTI dataset, which indicates that our method can be applied to CAD systems with practical clinical significance.

Keywords: Ultrasound image · Thyroid nodule · Detection · Segmentation · Multi-task network

1 Introduction

Thyroid nodule is a common clinical problem [1] and its incidence rate has risen rapidly worldwide. Ultrasound imaging technology has the characteristics of non-invasive, non-radioactive, convenient and inexpensive [2]. It is the primary tool for the diagnosis of thyroid nodule diseases. The diagnosis of thyroid nodules in ultrasound images depends on experienced clinicians [3]. However, due to the low contrast and low signal-to-noise ratio of ultrasound images, it hinders

G. Memmi et al. (Eds.): KSEM 2022, LNAI 13370, pp. 219–232, 2022.
https://doi.org/10.1007/978-3-031-10989-8_18

clinicians from making effective diagnosis. In order to solve this problem, more and more computer-aided diagnosis(CAD) systems are developed to assist in the diagnosis of thyroid diseases. In traditional CAD systems, the Region of Interest (ROI) of nodules is first defined manually by the clinicians, which is very time consuming and highly dependent on the clinicians' experience, and then the nodules are segmented based on the ROI. Therefore, automatic detection and segmentation of thyroid nodules is essential for CAD systems. The detection of thyroid nodules is used to predict the bounding boxes of nodules, and then automatic segmentation of nodules is performed based on the bounding boxes, which can effectively reduce the workload of clinicians.

In recent years, many deep learning methods have been proposed and applied to the detection and segmentation of thyroid nodules in ultrasound images.

Thyroid Nodule Detection Methods. Thyroid nodule detection models in ultrasound images can be divided into two types: two-stage models and one-stage models. In order to obtain higher detection precision, the two-stage models are usually applied to the detection of thyroid nodules. Li et al. [4] proposed an improved Faster R-CNN [12] for thyroid papillary carcinoma detection. By using the strategy of layer concatenation, the detector can extract the features of surrounding region around the cancer regions, which improves the detection performance. Liu et al. [5] replaced the layer concatenation strategy with Feature Pyramid Network(FPN) [13] and added it to Faster R-CNN [12] to construct a multi-scale detection network, which can extract the features of nodules with different scales. Abdolali et al. [6] replaced the network backbone from Faster R-CNN [12] to Mask R-CNN [14] with higher performance, using a well-designed loss function and transfer learning strategy to achieve high accuracy on a small dataset. These two-stage detection models mentioned above can obtain high precision in thyroid nodule detection, but the detection speed is lower than the one-stage models. In order to detect thyroid nodules with different scales, Song et al. [7] utilized a multi-scale SSD [15] model with spatial pyramid module to achieve high detection accuracy. To fully extract multi-scale features from feature maps, shahroudnejad et al. [8] constructed a one-stage model with FPN for detecting and classifying pyramid nodules, which can extract global and local information from feature maps. The above detection methods fully extracted thyroid nodule features at different scales by adding modules that extract multi-scale features, such as the connection between low-level and high-level layers, and FPN, thereby improving the accuracy of detecting thyroid nodules.

Thyroid Nodule Segmentation Methods. Ying et al. [9] proposed a cascaded convolutional neural network that first segmented the Region of Interest(RoI) containing thyroid nodules, and then used a VGG network to accurately segment thyroid nodules on the basis of RoI. Wang et al. [10] constructed a cascade segmentation network based on DeepLabv3plus [16]. The rough location of nodules was first obtained, and then the nodules were segmented accurately based on the rough location, which eliminated the influence of the area around the nodules on the segmentation results, and thus obtained more accurate segmentation results.

To remove the mistake recognition of non-thyroid areas as nodules, Gong et al. [11] embedded a priori guided feature module of thyroid region into the nodule segmentation model for the first time, which improved the accuracy of nodule localization and enhanced the segmentation performance of thyroid nodules. The above-mentioned thyroid nodule segmentation methods first remove the influence of irrelevant regions, and then perform further segmentation on the Region of Interest(RoI), thus reducing the false recognition of non-nodular regions as nodules.

Although many deep learning methods have been applied to the detection and segmentation tasks of thyroid nodules, most of them only complete one of the two tasks. Only a few methods can detect and segment thyroid nodules simultaneously. Among them, thyroid nodule detection methods achieve high accuracy while maintaining high efficiency, but there are still many problems in detecting thyroid nodules with extreme sizes, nodules with complex internal texture, and multiple nodules. It leads to missed detection of small nodules, false detection of intermediate nodules, and false detection of tissue similar to nodules as nodules. In addition, thyroid nodule segmentation method achieves high accuracy while there are still many challenges to be solved in becoming a real-time system.

To address the above problems, we propose a multi-task thyroid nodule detection and segmentation model based on Trident network [17], called MTN-Net. It is embedded with a novel semantic segmentation branch for accurate segmentation of thyroid nodules, and it includes an improved NMS algorithm, called TN-NMS, for combining the thyroid nodule detection results from multiple branches. Therefore, MTN-Net achieves significant effects on the detection of thyroid nodules with different sizes and thyroid nodules with complex internal texture, and effectively suppresses the false detection of intermediate nodules in large nodules.

The main contributions of this paper can be summarized as follows:

- We propose a multi-task network based on Trident network [17] for the detection and segmentation of thyroid nodules in ultrasound images, which can generate specific scale feature maps through trident block [17] with different receptive fields. So it is effective in detecting thyroid nodules with different sizes.
- A novel semantic segmentation branch based on FCN [18] is embedded into the detection network to complete the segmentation task of thyroid nodules, which is valid for completely segmenting the thyroid nodules with complex texture.
- We propose an improved NMS algorithm called TN-NMS to fuse the detection results from multiple branches, which can successfully suppress the false detection results of internal nodules in large nodules.

The rest of this paper is as follows: we first describe the details of our proposed model and the feature generation in Sect. 2. We then introduce the experimental setup and results in Sect. 3. Finally, we conclude our work and indicate future directions in Sect. 4.

2 Method

2.1 Overall Architecture

The proposed MTN-Net is a multi-branch two-stage thyroid nodule detection and segmentation model based on Trident network [17]. Figure 1 illustrates the overall architecture of our proposed MTN-Net. The network is composed of backbone, extended Faster R-CNN head, and TN-NMS algorithm. We adopt ResNet-101 with trident blocks as the backbone, in which the conv4_x stage consists of trident blocks containing three branches. It can fully extract the multi-scale features of thyroid nodules in ultrasound images, and thus contributing to the detection of thyroid nodules with different sizes. Additionally, we add a novel semantic segmentation branch to the extended Faster R-CNN head to accomplish the thyroid nodule segmentation task. Finally, an improved NMS algorithm called TN-NMS is used to combine the detection results of thyroid nodules from multiple branches.

Ultrasound images of thyroid nodules are input to the backbone to generate feature maps with different receptive fields. They are then fed into the extended Faster R-CNN head to produce the corresponding detection and segmentation results, which are eventually combined by the TN-NMS algorithm to generate the output results.

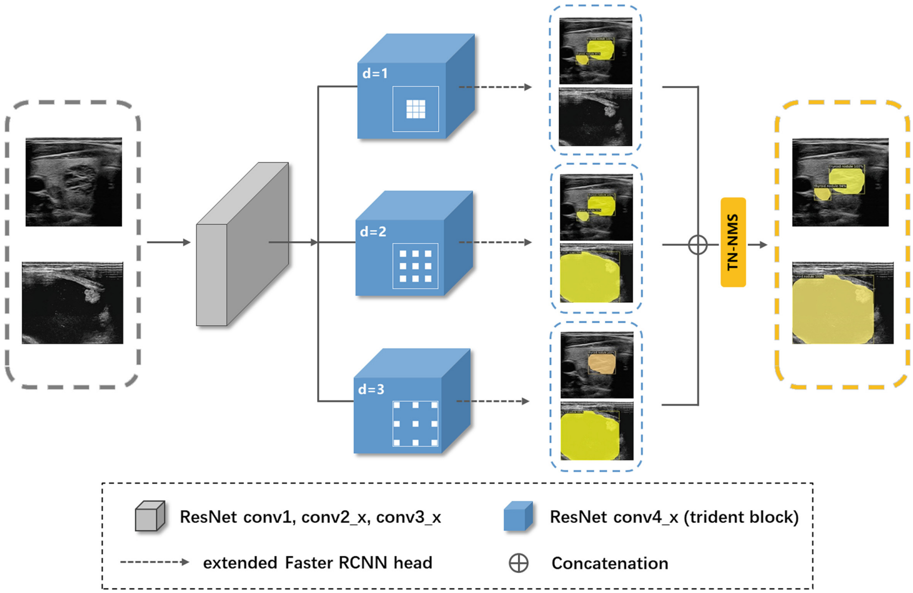

d=1
d=2
d=3
TN-NMS
ResNet conv1, conv2_x, conv3_x
ResNet conv4_x (trident block)
extended Faster RCNN head
Concatenation

Fig. 1. The architecture of proposed MTN-Net. MTN-Net is comprised of backbone (ResNet-101 with trident blocks), extended Faster R-CNN head, and TN-NMS algorithm.

2.2 Semantic Segmentation Branch

We use a novel semantic segmentation branch based on FCN [18] to segment thyroid nodules. This semantic segmentation branch is embedded into the Faster R-CNN detection head and parallel to the bounding-box classification and regression. In addition, we add an RoIAlign [14] layer in Faster R-CNN head to remove the rough space quantization of RoIPool [19], which can improve the accuracy of mask prediction at pixel level. The extended Faster R-CNN head is displayed in Fig. 2. Different from the existing extended Faster R-CNN heads mentioned in [14], our extended Faster R-CNN head has a novel semantic segmentation branch capable of segmenting thyroid nodules with complex textures more completely. We add four convolution layers before the deconvolution layer of the semantic segmentation branch to fully obtain the features in the Region-of-Interest(RoI), so as to completely segment the nodules with complex internal texture. Meanwhile, we add L_{mask} to the loss function. For some predicted boxes that do not contain thyroid nodules, the proposed semantic segmentation branch can suppress some incorrectly detected boxes through L_{mask}.

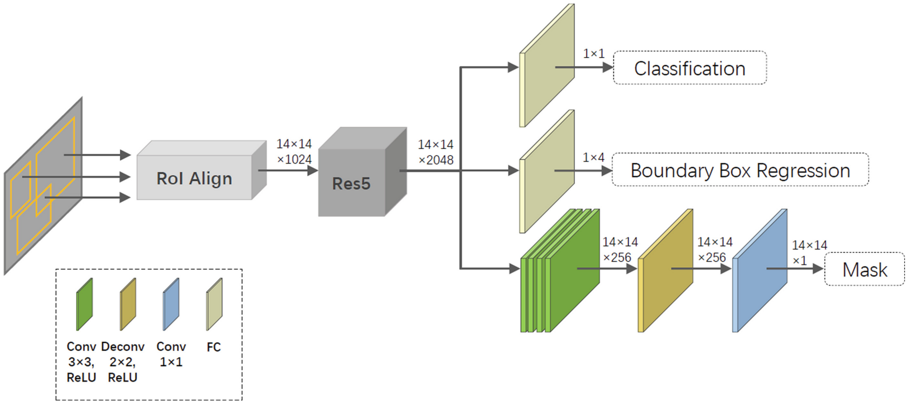
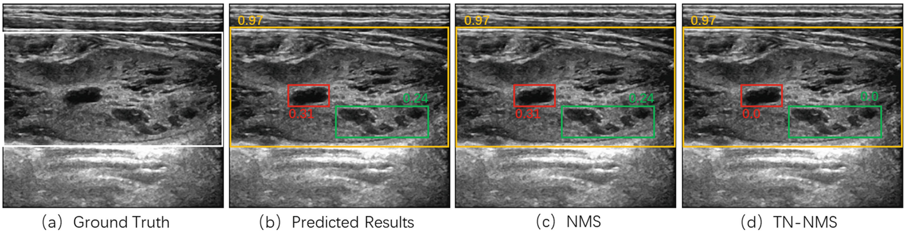


Fig. 2. The architecture of our extended Faster R-CNN head, in which a novel semantic segmentation branch is embedded to complete the segmentation task of thyroid nodules.

2.3 TN-NMS

NMS is utilized to merge the detection results from multiple branches in Trident network [17]. It is described as [20]:

$$S_i = \begin{cases} S_i, & iou\left(\mathcal{M}, b_i\right) < N_t \\ 0, & iou\left(\mathcal{M}, b_i\right) \geq N_t \end{cases} \tag{1}$$



(a) Ground Truth (b) Predicted Results (c) NMS (d) TN-NMS

Fig. 3. The results of a thyroid nodule with complex internal texture being correctly detected (yellow) and incorrectly detected (red and green) along with their confidence scores. Since the *iou* (0.03) of the red and yellow boxes, as well as the *iou* (0.11) of the green and yellow boxes, are much smaller than the threshold 0.5, the results of incorrect detections cannot be suppressed using the NMS algorithm (as shown in (c)). In contrast, the *niou* (1.0) of the red box and the yellow box, as well as the *niou* (1.0) of the green box and the yellow box exceed the threshold 0.9, so the TN-NMS algorithm can successfully suppress the bounding boxes of these false detections (as shown in (d)) (Color figure online)

Algorithm 1: TN-NMS

```
Input: $Boxes = \{b_1, ..., b_N\}$, $Scores = \{s_1, ..., s_N\}$,
    $N_{t_1}$, $N_{t_2}$;
    $Boxes$ is the list of detection boxes from three branches;
    $Scores$ contains corresponding detection scores from three branches;
    $N_{t_1}$ is the NMS threshold;
    $N_{t_2}$ is the nIoU threshold;
Output: $\mathcal{R}$ is the merged result from three branches;
    $Scores$ is the scores corresponding to the detection boxes in the merged
    result;
begin
    $\mathcal{R} \leftarrow \{\}$
    while $Boxes \neq empty$ do
        $m \leftarrow \text{argmax}\,\{Scores\}$
        $\mathcal{M} \leftarrow b_m$
        $\mathcal{R} \leftarrow \mathcal{R} \bigcup \mathcal{M}$; $Boxes \leftarrow Boxes - \mathcal{M}$
        for $b_i$ in $Boxes$ do
            if $iou\,(\mathcal{M}, b_i) \geq N_{t_1}$ or $niou\,(\mathcal{M}, b_i) \geq N_{t_2}$ then
                $Boxes \leftarrow Boxes - b_i$
                $Scores \leftarrow Scores - s_i$
            end
        end
    end
    return $\mathcal{R}, Scores$
end
```

The input data in Eq. 1 consists of an ordered list of detection boxes $Boxes$ with scores $Scores$ and a threshold N_t. S_i represents a re-scoring function, $\mathcal{M}$ is the box with the highest score in $Boxes$, b_i indicates the currently selected box in $Boxes$, iou denotes the intersection area divided by the union area of two boxes, N_t is a threshold indicating whether the currently selected box b_i should be removed. NMS starts by selecting the bounding box $\mathcal{M}$ with the highest score in $Boxes$, calculates the iou of the remaining bounding boxes b_i in $Boxes$ and $\mathcal{M}$, then deletes the bounding box b_i whose iou is greater than the threshold N_t, which is usually set to 0.5. However, the area of intermediate nodules detected by mistake is usually much smaller than that of large nodules, resulting in the iou of their corresponding bounding boxes less than 0.5, and thus the NMS algorithm is unable to suppress the results of these false detections, as shown in Fig. 3(c). Therefore, in order to suppress the bounding box of these intermediate nodules, we propose a new calculation method for thyroid nodule detection, named $niou$, which represents the intersection region of b_i and $\mathcal{M}$ divided by the region of b_i. It is described as:

$$niou\left(\mathcal{M}, b_i\right)=\frac{\mathcal{M} \cap b_i}{b_i} \tag{2}$$

The $niou$ calculated by the bounding box of incorrectly detected nodules and correctly detected nodules is usually equal to or close to 1.0, so that the results of incorrect detection above the threshold 0.9 are successfully suppressed, as shown in Fig. 3(d). Meanwhile, we add $niou$ to the NMS algorithm and propose an improved NMS algorithm, named TN-NMS, which is used to combine the detection results of three branches and is described as:

$$S_i=\begin{cases}S_i, & iou\left(\mathcal{M}, b_i\right)<N_{t_1} \text{ and } niou\left(\mathcal{M}, b_i\right)<N_{t_2} \\ 0, & iou\left(\mathcal{M}, b_i\right) \geq N_{t_1} \text{ or } niou\left(\mathcal{M}, b_i\right) \geq N_{t_2}\end{cases} \tag{3}$$

where N_{t_1} and N_{t_2} are thresholds that determine whether the currently selected bounding box b_i should be removed from $Boxes$. The detailed process of TN-NMS is shown in Algorithm 1. In each step of TN-NMS, the scores of all detection boxes that overlap with $\mathcal{M}$ are updated, then the detection boxes with a score of 0 are removed from $Boxes$, hence the computational complexity of each step of TN-NMS is $\mathcal{O}(N)$, where N is the number of detection boxes in Boxes. Therefore, for N detection boxes in $Boxes$, the computational complexity of the TN-NMS algorithm is $\mathcal{O}(N^2)$, which is the same as that of the NMS algorithm.

2.4 Loss Function

As shown in Fig. 1, the proposed network is a multi-task network, whose loss function combines the loss of classification, bounding box regression and segmentation. In order to improve performance, we add weighting factors to the loss function of each task. Therefore, the total loss function on each Region of Interest(RoI) is defined as follows:

$$L_{\text{total}}=\lambda_{cls} * L_{\text{cls}}+\lambda_{\text{box}} * L_{\text{box}}+\lambda_{\text{mask}} * L_{\text{mask}} \tag{4}$$

where L_{cls}, L_{box}, L_{mask} indicate classification loss, bounding box regression loss and mask segmentation loss respectively. λ_{cls}, λ_{box}, λ_{mask} are weighting factors of each component. We use the cross entropy loss function to calculate the classification loss of thyroid nodules, and utilize the smooth L1 loss for boundary box regression. The definitions of these two tasks are the same as those defined in [19]. Besides, we adopt the binary cross entropy loss to calculate the mask segmentation loss defined on the foreground proposals. Therefore, the loss of mask segmentation task is defined as follows:

$$L_{\text{mask}} = -\frac{1}{n^2} \sum_{0 \leq i,j \leq n} BCE\left(y_{ij}, y_{ij}^*\right) \tag{5}$$

where n is the length and width of each mask, y_{ij} is the predicted value and y_{ij}^* is the growth truth of each class. Furthermore, weighting factors can help optimize the performance of classification, detection and segmentation tasks.

3 Experiments

3.1 Dataset and Preprocessing

We evaluated the proposed architecture on the public thyroid nodule region segmentation dataset called TN3K provided in [11], which contains 3493 ultrasound images obtained from 2421 patients. In addition, we compare the performance of our proposed method with State-of-the-Arts methods on the public DDTI dataset [21]. It contains 347 thyroid ultrasound images from 299 patients with thyroid disease, annotated by radiologists for thyroid nodule segmentation results. All the cases in the DDTI dataset are from the IDIME Ultrasound Department, one of the largest imaging centers in Colombia.

In order to adopt these two datasets to thyroid nodule detection and segmentation, we add the bounding box annotation for object detection. Besides, we use the operation of adaptive histogram equalization for each image to transform the gray level of the image, so as to improve the contrast of the image. In addition, we perform data augmentation operations on the preprocessed images used for training, including random mirror flip, random left-right flip, random clipping, random sharpening, random increase or decrease of image contrast.

3.2 Implementation Details

The proposed network is implemented in PyTorch 1.8.1. The experimental codes are modified on the basis of Detectron2 [22], and many default configuration parameters are used for model training and inference. The model is trained on two NVIDIA Tesla P100 GPUs with a batch size of 16, and the backbone of the network is pre-trained on MS-COCO [23]. In our experiments, N_{t_1} and N_{t_2} in TN-NMS are set to 0.5 and 0.9 respectively, and λ_{cls}, λ_{box}, λ_{mask} of loss function are set to 2, 5 and 2 respectively. Moreover, the model is trained with the stochastic gradient descent optimizer and the learning rate of warmup and cosine annealing for 50 epochs, whose learning rate increases linearly to 0.05 in the first 1000 iterations, then decreases gradually in the form of cosine annealing. The total time of model training is 20 h, and the inference time of each image is 0.85 s.

3.3 Evaluation Metrics

For the evaluation, in order to accurately quantify the performance of our model, standard COCO metrics including AP (Average Precision), AP_{50} and metrics for evaluating the Average Precision of objects with different size, including AP_S (less than 32×32), AP_M (from 32×32 to 96×96), AP_L (greater than 96×96) are used as evaluation metrics. Since the smallest thyroid nodule contained in the DDTI dataset are larger than 32×32 pixels in size, AP_S cannot be used as an evaluation metric for the DDTI dataset. Therefore, we measure the thyroid nodule detection and segmentation performance of AP, AP_{50}, AP_M, AP_L on the DDTI dataset.

3.4 Ablation Study

In order to validate the performance of our proposed architecture, the evaluation metrics of detection and segmentation are used to quantify the comparison between our proposed model and baseline model. The baseline is Trident network with a mask prediction branch proposed in [14], which includes a 2×2 deconvolution layer with stride 2 and a 1×1 convolution layer for predicting mask. Baseline/ResNet-101 backbone refers to the baseline network with ResNet-101 as the backbone. Then we respectively add semantic segmentation branches and TN-NMS algorithm on baseline, which is denoted as bNet+S and bNet+T.

Table 1. Ablation studies on the detection of thyroid nodules.

Model	TN3K					DDTI			
	AP	AP$_{50}$	AP$_S$	AP$_M$	AP$_L$	AP	AP$_{50}$	AP$_M$	AP$_L$
Baseline/ResNet-101 backbone	54.7	85.7	32.8	47.3	61.1	49.0	85.1	42.0	57.2
bNet+S	54.9	86.6	30.6	47.2	62.5	49.8	87.2	43.7	57.5
bNet+T	54.7	86.6	**36.7**	48.0	61.6	49.7	85.5	43.0	57.4
Ours	**55.2**	**87.1**	34.4	**48.2**	**62.8**	**51.3**	**88.7**	**45.4**	**57.8**

As shown in Table 2, bNet+S improves 1.2% and 1.0% on AP_L for nodule segmentation on TN3K and DDTI, respectively, which indicates that semantic segmentation branche has high performance in segmenting large nodules. From Table 1, we can see that bNet+T has a 3.9% and 0.4% improvement on AP_S and AP_L for TN3K and 0.2% improvement on AP_L for DDTI, respectively, which demonstrates that the TN-NMS algorithm improves the detection performance of large and small nodules by suppressing the internal nodules in large nodules.

Table 2. Ablation studies on the segmentation of thyroid nodules.

Model	TN3K					DDTI			
	AP	AP$_{50}$	AP$_S$	AP$_M$	AP$_L$	AP	AP$_{50}$	AP$_M$	AP$_L$
Baseline/ResNet-101 backbone	56.2	84.6	32.0	50.1	61.7	46.7	84.3	41.1	53.2
bNet+S	56.5	85.6	31.4	50.0	**62.9**	47.7	85.5	43.4	54.2
bNet+T	56.3	85.4	**36.0**	50.6	62.5	47.0	84.4	41.3	53.6
Ours	**56.8**	**86.9**	35.5	**50.8**	**62.9**	**49.0**	**86.6**	**44.7**	**54.4**

When both are added into baseline, MTN-Net greatly enhances in all evaluation metrics compared to baseline. However, the AP_S of MTN-Net is lower than that of bNet+T. We consider that the semantic segmentation branch focuses too much on large nodules, and thus has lower performance on the detecting and segmenting small nodules, there by leading to the lower performance of MTN-Net than that of bNet+T.

3.5 Comparisons Against State-of-the-Arts Methods

We compared our framework MTN-Net with several state-of-the-art approaches, including Mask R-CNN [14], Cascade Mask R-CNN [24], Mask Scoring R-CNN [25], PointRend [26]. Mask R-CNN is a commonly used two-stage detection and segmentation model. And Cascade Mask R-CNN is a multi-head model based on Cascade R-CNN, which has higher detection accuracy than Mask R-CNN. Besides, Mask Scoring R-CNN adds a branch for scoring masks on the basis of Mask R-CNN, which enhances the accuracy of segmentation. Furthermore, PointRend is optimized for image segmentation at the edges of objects, resulting in better performance at the hard-to-segment edges of objects.

Table 3. Performance comparison of thyroid nodule detection on TN3K and DDTI.

Model	TN3K					DDTI			
	AP	AP$_{50}$	AP$_S$	AP$_M$	AP$_L$	AP	AP$_{50}$	AP$_M$	AP$_L$
Mask R-CNN	52.1	84.5	28.3	45.2	59.5	45.1	82.8	38.2	53.5
Cascade Mask R-CNN	53.9	84.8	31.0	47.1	61.3	47.5	84.3	43.0	54.0
Mask Scoring R-CNN	53.1	84.4	**37.3**	45.6	62.0	47.6	83.4	39.4	57.1
PointRend	54.5	85.0	37.1	47.3	61.3	47.4	82.0	40.0	56.1
Ours	**55.2**	**87.1**	34.4	**48.2**	**62.8**	**51.3**	**88.7**	**45.4**	**57.8**

Quantitative Analysis on TN3K. Tables 3 and 4 demonstrate the quantitative comparison results between our MTN-Net and other SOTA models on the public TN3K dataset. MTN-Net greatly improves AP, AP_{50}, AP_M, and AP_L against other SOTA models. However, the performance in detecting and segmenting

Table 4. Performance comparison of thyroid nodule segmentation on TN3K and DDTI.

Model	TN3K					DDTI			
	AP	AP_{50}	AP_S	AP_M	AP_L	AP	AP_{50}	AP_M	AP_L
Mask R-CNN	54.4	84.5	30.0	48.3	60.7	42.6	80.2	36.4	49.7
Cascade Mask R-CNN	55.5	84.7	31.6	49.8	62.0	45.6	84.4	41.0	51.0
Mask Scoring R-CNN	55.1	84.6	**37.5**	49.5	61.1	46.1	82.9	41.5	51.9
PointRend	56.2	85.8	36.5	50.3	62.3	46.4	81.0	39.5	53.2
Ours	**56.8**	**86.9**	35.5	**50.8**	**62.9**	**49.0**	**86.6**	**44.7**	**54.4**

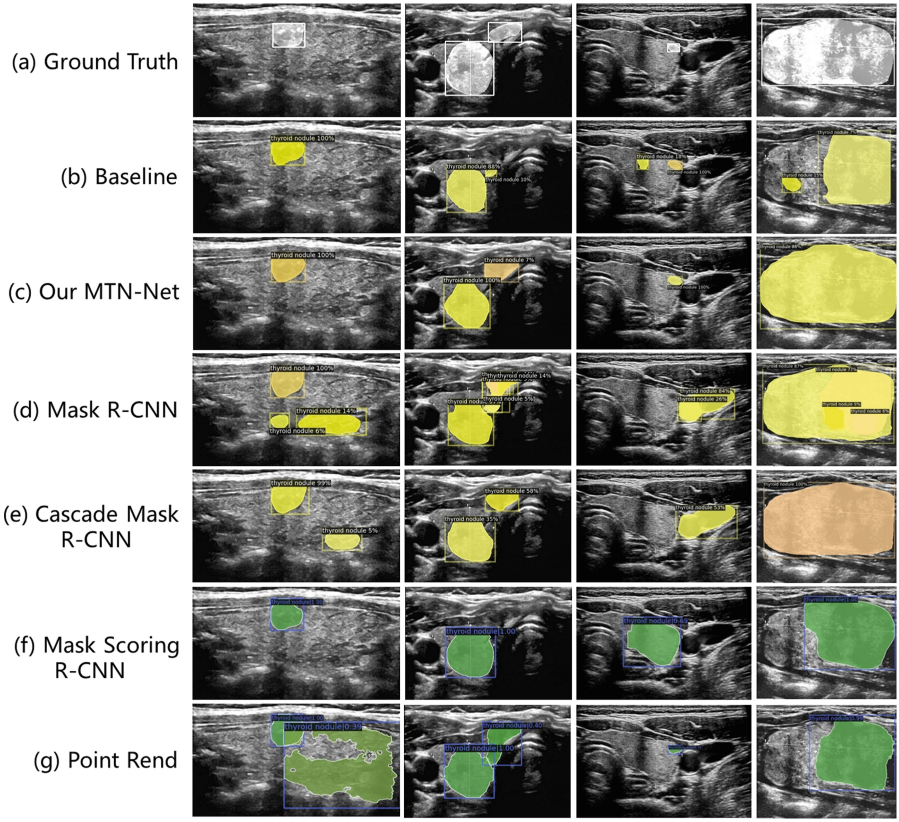


Fig. 4. Qualitative comparison of our MTN-Net and SOTA models. Among them, Baseline, Our MTN-Net, Mask R-CNN, Cascade Mask R-CNN (yellow) are implemented based on Detectron2, and Mask Scoring R-CNN and Point Rend (green) are implemented based on MMDetection [27] (Color figure online)

small nodules (less than 32 × 32 pixels) is inferior to Mask Scoring R-CNN and Point Rend. Since the appearance and texture of some small nodules are extremely similar to the surrounding tissues, MTN-Net is prone to mis-detect other tissues and organs as small nodules. Nevertheless, MTN-Net has high accuracy on both AP_M, and AP_L, which indicates its remarkable competitiveness in detecting and segmenting medium and large nodules.

Quantitative Analysis on DDTI. As shown in Tables 3 and 4, MTN-Net exceeds other SOTA models in the above metrics on the DDTI dataset. For thyroid detection, it increases 3.8%, 4.4%, 2.4%, and 0.7% for AP, AP_{50}, AP_M, and AP_L, respectively. For thyroid segmentation, the increases are 2.4%, 2.2%, 3.2%, and 1.2% for AP, AP_{50}, AP_M, and AP_L, respectively. This demonstrates that MTN-Net has an excellent performance in both nodule detection and segmentation when the nodule size is larger than 32 × 32 pixels.

Qualitative Analysis. Figure 4 illustrates the qualitative comparison results between our MTN-Net and other SOTA models. The first column of Fig. 4 shows that MTN-Net can successfully exclude false-positive detection results. And the second column of Fig. 4 illustrates that MTN-Net is able to accurately detect and segment multiple thyroid nodules. In addition, the third column of Fig. 4 displays that MTN-Net is significantly competitive in the detection of small nodules. Furthermore, the fourth column of Fig. 4 indicates that MTN-Net can not only completely segment large nodules with complex texture, but also effectively suppress internal nodules.

4 Conclusion

In this paper, we proposed a two-stage network for thyroid nodule detection and segmentation in ultrasound images. Our network is built on Trident network, which is capable of precisely detecting thyroid nodules with diverse sizes. The semantic segmentation branch added to the network is effective for fully segmenting large nodules with complex textures. In addition, we proposed an improved NMS algorithm to fuse the detection results from multiple branches, and it is useful to suppress the false detection of internal nodules. Consequently, our network achieves a remarkable competitiveness in detecting thyroid nodules with diverse sizes, segmenting completely nodules with internal texture, and suppressing incorrectly detected internal nodules. Experimental results demonstrate the effectiveness of the proposed method against other state-of-the-art methods. In the future, we will utilize self-supervision methods to further reduce the false positive rate of our model for thyroid nodule detection and segmentation in ultrasound images.

Acknowledgements. This research is funded by East China Normal University-Qiniu Intelligent Learning Joint Laboratory. The computation is supported by ECNU Multifunctional Platform for Innovation (001).

References

1. Haugen, B.R., et al.: 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)
2. Savelonas, M.A., Iakovidis, D.K., Legakis, I., Maroulis, D.: Active contours guided by echogenicity and texture for delineation of thyroid nodules in ultrasound images. IEEE Trans. Inf. Technol. Biomed. **13**(4), 519–527 (2008)
3. Chen, J., You, H., Li, K.: A review of thyroid gland segmentation and thyroid nodule segmentation methods for medical ultrasound images. Comput. Methods Progr. Biomed. **185**, 105329 (2020)
4. Li, H., et al.: An improved deep learning approach for detection of thyroid papillary cancer in ultrasound images. Sci. Rep. **8**(1), 1–12 (2018)
5. Liu, R., Zhou, S., Guo, Y., Wang, Y., Chang, C.: Nodule localization in thyroid ultrasound images with a joint-training convolutional neural network. J. Digital Imaging **33**(5), 1266–1279 (2020)
6. Abdolali, F., Kapur, J., Jaremko, J.L., Noga, M., Hareendranathan, A.R., Punithakumar, K.: Automated thyroid nodule detection from ultrasound imaging using deep convolutional neural networks. Comput. Biol. Med. **122**, 103871 (2020)
7. Song, W., et al.: Multitask cascade convolution neural networks for automatic thyroid nodule detection and recognition. IEEE J. Biomed. Health Inform. **23**(3), 1215–1224 (2018)
8. Shahroudnejad, A., et al.: TUN-Det: a novel network for thyroid ultrasound nodule detection. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12901, pp. 656–667. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87193-2_62
9. Ying, X., et al.: Thyroid nodule segmentation in ultrasound images based on cascaded convolutional neural network. In: Cheng, L., Leung, A.C.S., Ozawa, S. (eds.) ICONIP 2018. LNCS, vol. 11306, pp. 373–384. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-04224-0_32
10. Wang, M., et al.: Automatic segmentation and classification of thyroid nodules in ultrasound images with convolutional neural networks. In: Shusharina, N., Heinrich, M.P., Huang, R. (eds.) MICCAI 2020. LNCS, vol. 12587, pp. 109–115. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-71827-5_14
11. Gong, H., et al.: Multi-task learning for thyroid nodule segmentation with thyroid region prior. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), pp. 257–261. IEEE (2021) https://doi.org/10.1109/ISBI48211.2021.9434087
12. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. Adv. Neural Inf. Process. Syst. **28**, 1–9 (2015)
13. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2117–2125 (2017). https://doi.org/10.48550/arXiv.1612.03144
14. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2961–2969 (2017). https://doi.org/10.48550/arXiv.1703.06870

15. Liu, W., et al.: SSD: single shot multibox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9905, pp. 21–37. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_2
16. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 801–818 (2018). https://doi.org/10.48550/arXiv.1802.02611
17. Li, Y., Chen, Y., Wang, N., Zhang, Z.: Scale-aware trident networks for object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6054–6063 (2019). https://doi.org/10.1109/ICCV.2019.00615
18. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015). https://doi.org/10.1109/CVPR.2015.7298965
19. Girshick, R.: Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448 (2015). https://doi.org/10.1109/ICCV.2015.169
20. Bodla, N., Singh, B., Chellappa, R., Davis, L.S.: Soft-NMS-improving object detection with one line of code. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 5561–5569 (2017). https://doi.org/10.1109/ICCV.2017.593
21. Pedraza, L., Vargas, C., Narváez, F., Durán, O., Muñoz, E., Romero, E.: An open access thyroid ultrasound image database. In: 10th International Symposium on Medical Information Processing and Analysis. vol. 9287, p. 92870W. International Society for Optics and Photonics (2015). https://doi.org/10.1117/12.2073532
22. Wu, Y., Kirillov, A., Massa, F., Lo, W.Y., Girshick, R.: Detectron2 (2019). https://github.com/facebookresearch/detectron2
23. Lin, T.-Y., et al.: Microsoft COCO: common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8693, pp. 740–755. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10602-1_48
24. Cai, Z., Vasconcelos, N.: Cascade R-CNN: high quality object detection and instance segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **43**(5), 1483–1498 (2019)
25. Huang, Z., Huang, L., Gong, Y., Huang, C., Wang, X.: Mask scoring R-CNN. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6409–6418 (2019). https://doi.org/10.1109/CVPR.2019.00657
26. Kirillov, A., Wu, Y., He, K., Girshick, R.: Pointrend: image segmentation as rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9799–9808 (2020). https://doi.org/10.1109/CVPR42600.2020.00982
27. Chen, K., et al.: MMDetection: Open MMLab detection toolbox and benchmark. arXiv preprint arXiv:1906.07155 (2019)