Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/media

Thyroid Region Prior Guided Attention for Ultrasound Segmentation of Thyroid Nodules

Haifan Gong[a,c], Jiaxin Chen[b], Guanqi Chen[a], Haofeng Li[c], Guanbin Li[a,*], Fei Chen[d,**]

[a] *School of Computer Science and Engineering, Research Institute of Sun Yat-sen University in Shenzhen, Sun Yat-sen University, Guangzhou, 510000, China*
[b] *School of Mathematics and Computer Science, Nanchang University, Nanchang, 330000, China*
[c] *Shenzhen Research Institute of Big Data, The Chinese University of Hong Kong (Shenzhen), Shenzhen, 518000, China*
[d] *Zhujiang Hospital, Southern Medical University, Guangzhou, 510000, China*

ARTICLE INFO

Article history:

Keywords: Thyroid nodule, Ultrasound image, Segmentation, Multi-task learning, Attention modeling, Benchmark

ABSTRACT

Ultrasound segmentation of thyroid nodules is a challenging task, which plays an vital role in the diagnosis of thyroid cancer. However, the following two factors limit the development of automatic thyroid nodule segmentation algorithms: (1) existing automatic nodule segmentation algorithms that directly apply semantic segmentation techniques can easily mistake non-thyroid areas as nodules, because of the lack of the thyroid gland region perception, the large number of similar areas in the ultrasonic images, and the inherently low contrast images; (2) the currently available dataset (i.e., DDTI) is small and collected from a single center, which violates the fact that thyroid ultrasound images are acquired from various devices in real-world situations. To overcome the lack of thyroid gland region prior knowledge, we design a thyroid region prior guided feature enhancement network (TRFE+) for accurate thyroid nodule segmentation. Specifically, (1) a novel multi-task learning framework that simultaneously learns the nodule size, gland position, and the nodule position is designed; (2) an adaptive gland region feature enhancement module is proposed to make full use of the thyroid gland prior knowledge; (3) a normalization approach with respect to the channel dimension is applied to alleviate the domain gap during the training process. To facilitate the development of thyroid nodule segmentation, we have contributed TN3K: an open-access dataset containing 3493 thyroid nodule images with high-quality nodule masks labeling from various devices and views. We perform a thorough evaluation based on the TN3K test set and DDTI to demonstrate the effectiveness of the proposed method. Code and data are available at https://github.com/haifangong/TRFE-Net-for-thyroid-nodule-segmentation.

1. Introduction

Thyroid nodule is one of the very common endocrine complaints which grows in the thyroid gland. Due to its potential of being malignant, there is an imperative need to examine the thyroid regularly Durante et al. (2018). With the unique advantages (e.g., relatively inexpensive, easy-to-use, portable) of ultrasound imaging, the prevailing examination approach to the thyroid is based on ultrasonic imaging equipment. Nevertheless, as the ultrasound examination progress is not standardized comparing to other image modalities (e.g., CT, MRI, X-ray), the diagnose of thyroid nodules is highly dependent on the experience and skill of the clinicians according to Yuan et al. (2022). Considering that inexperienced clinicians could easily cause misdiagnosis, many computer-aided diagnosis systems proposed by Wang et al. (2020); Yang et al. (2021); Manh et al.; Gong et al. (2022) have been developed for auxiliary

[*] Corresponding author *email: liguanbin@mail.sysu.edu.cn* (Guanbin Li)
[**] Corresponding author *email: gzchenfei@126.com* (Fei Chen)

diagnosis of thyroid diseases. The automatic segmentation of thyroid nodules is a fundamental component for developing an intelligent diagnosis system and is also valuable for ultrasound-guided thyroid puncture biopsy or resection of nodules Chen et al. (2020).

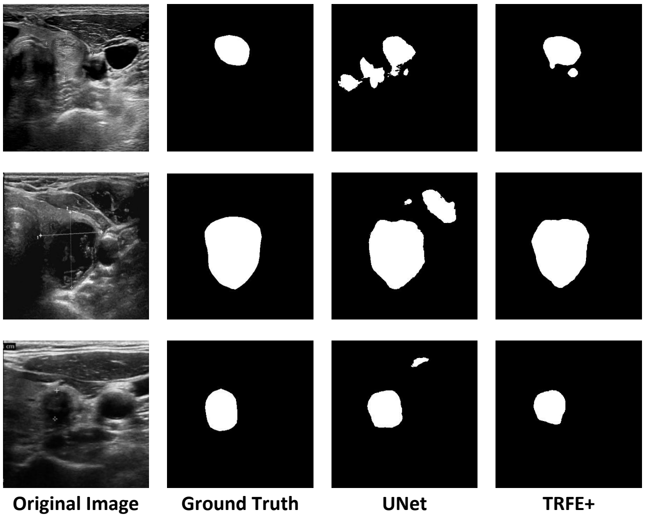

Fig. 1. Several visualization examples of our TRFE+ comparing to UNet Ronneberger et al. (2015). We can see that the UNet mis-regard the non-thyroid region as the thyroid nodule.

The previous thyroid nodule segmentation algorithms are mainly based on conventional segmentation models Ronneberger et al. (2015); Ma et al. (2017); Ying et al. (2018); Kumar et al. (2020) or attention mechanism Pan et al. (2021). However, these methods usually do not explicitly constrain the thyroid nodules to be located in the thyroid gland area, which leads to the algorithms generating incorrect location of thyroid nodules outside the thyroid gland region. Furthermore, these models can't use a large amount of *independently labeled dataset* (i.e., separate training data labeled with either thyroid gland region or thyroid nodule), which is a very common situation in the medical domain since the clinical annotations are expensive.

Another issue is that the current publicly available benchmark Pedraza et al. (2015) for thyroid nodules segmentation is limited and monolithic. The developed algorithms are usually based on ultrasonic data from a single center (i.e., the imaging captured from single equipment with fixed setting), which deviates from the realistic application scenario. Therefore, there is a need to construct a data set that contains ultrasound thyroid imaging from different devices of different settings. Furthermore, the currently public available dataset only contains one thyroid nodule per imaging, which is not consistent with the incidence of thyroid nodules in the actual scenario. Thus, we propose a new thyroid nodule segmentation dataset which contains various images of different nodule size, views, intensity, and from different ultrasonic devices as shown in Fig. 2.

In this work, we present **T**hyroid **R**egion prior guided **F**eature **E**nhancement network (**TRFE+**), a thyroid nodule segmentation framework that aims to improve the thyroid nodule segmentation performance by making use of the *independently labeled dataset*. However, there exist several intrinsic issues: (1) In our **T**hyroid **N**odule **3,493** (**TN3K**) data set, the size and number of thyroid nodules in different images are very different; while the size of the thyroid glands in the **T**hyroid **G**land **3,585** (**TG3K**) data set is relatively fixed, which makes it difficult for the encoder to effectively identify the nodules area. (2) Data from different sources will lead to unstable training. If we directly form a batch of data from different sources, it will be difficult to correct the deviations between samples during the training process. To address the above-mentioned issues, the TRFE+ contains a shared backbone encoder for feature representation learning and three separate decoders for thyroid gland region segmentation, thyroid nodule discovery, and nodule size prediction, respectively. To take full advantage of the features learned from the decoder of thyroid gland segmentation, we further design a thyroid region feature enhancement module to simultaneously capture important feature channels and filter out the part of the gland prior features that are not related to nodules. Finally, we apply the group normalization operation to alleviate the domain shift between the two *independently labeled dataset*. In summary, we make the following five contributions:

1. We introduce a novel multi-task learning-based network to simultaneously segments the thyroid gland regions and nodule regions, forcing the same backbone network to capture the location of nodules within the gland region. A nodule size prediction task is designed to constrain the encoder to be aware of the nodule size, thus avoids the over-fitting of the thyroid gland region prior information.
2. We redesign the thyroid region prior guided feature enhancement module by taking the relationship between both the channels and position into account, enabling better feature fusion between the gland prior feature map and the nodule feature map.
3. We demonstrate that the normalization of features from the channel dimension helps to alleviate the domain shift between the independently labeled training data and the test data from different sources.
4. We construct TN3K: an open-access dataset of thyroid nodule images with high-quality nodule masks labeling, aiming at facilitating the research of thyroid nodule segmentation.
5. We demonstrate the effectiveness of the proposed method on two test sets with a full evaluation, both qualitatively and quantitatively.

2. Related works

The approaches for thyroid nodule segmentation have been reviewed in Chen et al. (2020), which can be roughly divided into the conventional counter, shape, region-based methods, and the prevailing deep learning-based methods. Since conventional approaches like ACWE Gui et al. (2017) or VRS Alrubaidi et al. (2016) requires to manually preset the potential nodule region, deep learning-based methods (e.g., FCN Long et al. (2015), UNet Ronneberger et al. (2015), SegNet Badrinarayanan et al. (2017), Deeplab Chen et al. (2018)) has shown its advance not only by making the inference automatically, but also significantly outperform the conventional methods. Thus, we initially review the deep learning-based semantic segmentation algorithms, then we introduce the approaches for thyroid nodule segmentation.

2.1. Deep Learning for Image Segmentation

Deep convolutional neural networks have recently greatly facilitated the progress of visual semantic segmentation and have shown remarkable progress. A large amount of advanced deep neural network structures have been proposed in the research area of semantic segmentation as well as medical image segmentation. With the key idea that fully convolutional networks could take input of the arbitrary size and produce spatially dense prediction efficiently, Long et al. (2015) proposed FCN, which is a new scheme for the semantic segmentation task. In the meanwhile, Ronneberger et al. (2015) proposed the UNet and its typical variants like UNet++ proposed by Zhou et al. (2019), have shown its great advance in the medical image segmentation domain with its encoder-decoder structure. The encoder progressively down-samples the image features and generates coarse contextual features that focus on contextual patterns, and the decoder progressively up-samples the contextual features and fuses them with fine-grained local visual features.

Other prevailing encoder-decoder based structure also includes Deeplab Chen et al. (2018), SegNet Badrinarayanan et al. (2017), etc. To resolve the performance degradation caused by insufficient context information extraction, Feng et al. (2020) proposed CPFNet which efficiently fuse the pyramid context information. To deal with diversity of size, color, and the textual of lesions, Fan et al. (2020) proposed the PraNet which aggregates the features in high-level layers using a parallel partial decoder to guide the learning of the lesions. Recently, Transformer Vaswani et al. (2017) based visual models have attracted great attention with their superiority in discovering the long-range relationship between the samples. Chen et al. (2021) proposed the TransUnet to segment the image. The transformer encodes markup image blocks from convolutional neural network (CNN) feature maps as input sequences to extract global context. Wu et al. (2022) used the transformer to segment the lesions. While the decoder up-samples the encoded features and then combines them with high-resolution CNN feature maps to achieve accurate positioning. By combining UNet with Transformers to recover local spatial information, it becomes a powerful framework for medical image segmentation tasks.

2.2. Thyroid Nodule Segmentation

For thyroid nodule segmentation, Abbasian Ardakani et al. (2019) use the hybrid filtering approach for thyroid nodule segmentation. Ma et al. (2017) first leveraged the convolutional neural network to segment the thyroid nodule. Ying et al. (2018) proposed a cascaded framework that first eliminates the influence of irrelevant regions with a UNet Ronneberger et al. (2015), and then applies the VGG Simonyan and Zisserman (2015) network to segment the thyroid nodules. Kumar et al. (2020) proposed a neural network to simultaneously segment the gland, the solid nodule, and the cystic nodule of the thyroid. Pan et al. (2021) proposed a thyroid nodule segmentation approach called SGUNet which is based on the guidance of the semantic feature. SGUNet abstracts a single channel pixel-wise semantic feature map from the high-dimensional features in each decoding layer, which is treated as high-level semantic guidance to low-level features in order to obtain a more accurate nodule position. Song et al. (2022) proposed a dual-branch pseudo-label-based method to localize thyroid nodules in ultrasound images. Sun et al. (2022) proposed TNSNet to achieve accurate segmentation of thyroid nodules by constraining the edges of thyroid nodules with soft labels. Chen et al. (2022) proposed that thyroid nodules are classified into different types according to their cystic and solid characteristics, and designed an encoder network that can sense the type of thyroid nodules to achieve more accurate segmentation of thyroid nodules. Yu Yu et al. (2022) proposed to segment the thyroid under weak supervision with self-attention mechanism. However, none of the existing thyroid nodule segmentation methods explicitly constrain the thyroid nodules to be located in the thyroid gland region, which we believe is an important reason for the relatively poor performance considering the inherently low contrast of ultrasound images.

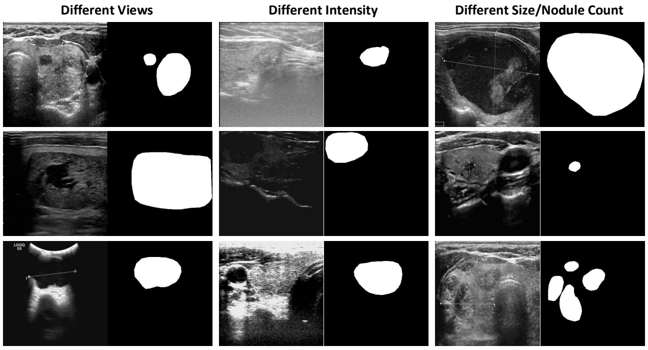

Fig. 2. Illustration of ultrasonic thyroid images from the proposed TN3K dataset and their pixel-wise annotations. From left to right: different views, intensities and sizes.

3. Dataset

To advance the research towards a computer-aided-diagnosis system for automatic thyroid nodule diagnosis, we contribute **TN3K**: a challenging thyroid nodule segmentation dataset including 3,493 ultrasound images with pixel-wise labels. The TN3K dataset was collected at Zhujiang Hospital, South Medical University, and has received the appropriate approvals from the institutional ethics committees.

3.1. Motivation

To the best of our knowledge, the only publicly available thyroid nodule dataset without any usage restriction is provided by Pedraza et al. (2015), which we call DDTI. However, this dataset only contains 637 images with pixel-wise lesion masks, which is limited for the deep learning-based model training and evaluation. Furthermore, in clinical applications, the composition of thyroid ultrasound image dataset is complex. As shown

Table 1. Summary of the ultrasonic thyroid image used in this study.

Dataset	Train	Test	Ultrasounic Imaging Device
DDTI Pedraza et al. (2015)	-	637	TOSHIBA Nemio 30, TOSHIBA Nemio MX
TG3K Wunderling et al. (2017)	3585	-	GE Logiq E9
TN3K	2879	614	GE Logiq E9, ARIETTA 850, RESONA 70B

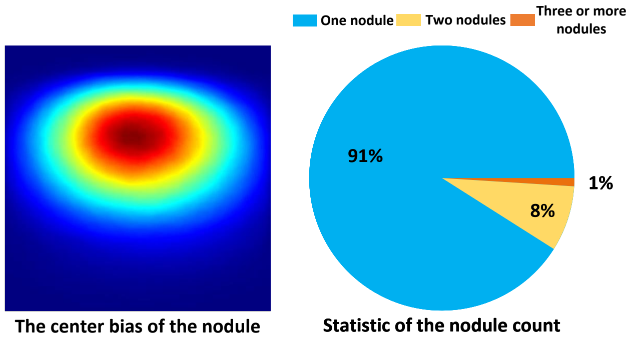

Fig. 3. Center bias and the thyroid nodule statistics information of TN3K dataset.

in Fig. 2, the imaging of thyroid could be captured from different views (e.g., front view, side view, vertical view) or different intensities (e.g., high brightness, low brightness). Besides, the multi-nodules situation is not considered in DDTI, which is a very common phenomenon during the ultrasonic thyroid examination. Thus, we build the TN3K dataset, which contains abundant ultrasound thyroid images from real-world scenarios.

3.2. Sample Collection and Annotation

The TN3K dataset is collected from various ultrasonic imaging systems including the GE Logiq E9, ARIETTA 850, and RESONA 70B. We select the samples with the following criteria: (1) At least one thyroid nodule is in the image; (2) The image doesn't contain the blood signal; (3) Only one representative image (i.e., the nodule closer to the center) is retained among the images from the same perspective or the same area of a patient. After selecting the samples with the above criteria, we obtain 3,493 images from 2,421 patients. Then we preprocess these images by ensuring that each image is converted to gray-scale and the non-ultrasound image area is cropped. We first ask volunteers to label the nodules under the guidance of an experienced radiologist. Then a radiologist is asked to check the annotation he confirmed and further invite a senior clinician to perform verification on the confusing thyroid nodule images. It is worth noting that if there are multi-nodules in the image, we label each nodule with no connecting mask. The statistics information about this dataset is shown in Fig. 3. The degree of the center bias is obtained by computing the average nodule mask over all samples in the TN3K dataset, which indicates the potential position of the nodule in image. Furthermore, there are about 9% ultrasonic images with two or above thyroid nodules.

3.3. Dataset Summary

Table. 1 summarizes the three ultrasound datasets used in our experiments. TG3K is the dataset with pixel-wise thyroid gland masks. TN3K and DDTI denote the data set with the thyroid nodule label. All the non-ultrasonic regions have been removed.

DDTI: This dataset is provided by the Pedraza et al. (2015), which contains 637 ultrasonic thyroid imaging with pixel-wise labels from a single device. As the dataset is limited, we treat this dataset as the external testset to evaluate the performance and the generalization ability of the algorithms.

TG3K: This dataset is acquired from 16 ultrasonic videos proposed by Wunderling et al. (2017). It is originally designed to segment the thyroid gland region from the videos. In this work, we first extract the frames from the videos. Based on the concern that the gland image is useless if it only contains a small part of the gland, we construct the thyroid gland imaging segmentation dataset with the rule that the proportion of the thyroid gland area to the image should be greater than 0.06. After that, we get 3,585 images for training.

TN3K: The proposed dataset is split into the training set and the test set with the criteria that the images from the same patient only appear in a certain subset. Thus, the training set contains 2,879 images while the test set contains 614 images.

4. Methodology

In this section, we introduce our proposed thyroid region prior guided feature enhance network named TRFE+, for thyroid nodule segmentation in ultrasonic images, which is shown in Fig. 4. The TRFE+ is mainly composed of three parts: a shared encoder, three separate decoders, and an adaptive thyroid region prior guidance module. The encoder is designed to extract the high-dimensional feature representation for the image. The decoders are designed for thyroid gland region segmentation, thyroid nodule discovery, and nodule size prediction, respectively. The adaptive region prior guidance module is designed to use the thyroid gland region prior information to improve the segmentation performance of nodules.

4.1. Multi-task Learning Framework

The pipeline of the TRFE+ is displayed in Fig. 4, which is consisted of a shared encoder backbone to learn the feature representation and three separate decoders for thyroid gland segmentation, nodule size prediction, and nodule segmentation, respectively. During the training process, we select one image from each of the TN3K and TG3K dataset to form a mini-batch, the nodule image and gland image are denoted in blue block and green block, respectively. The gland images are fed into

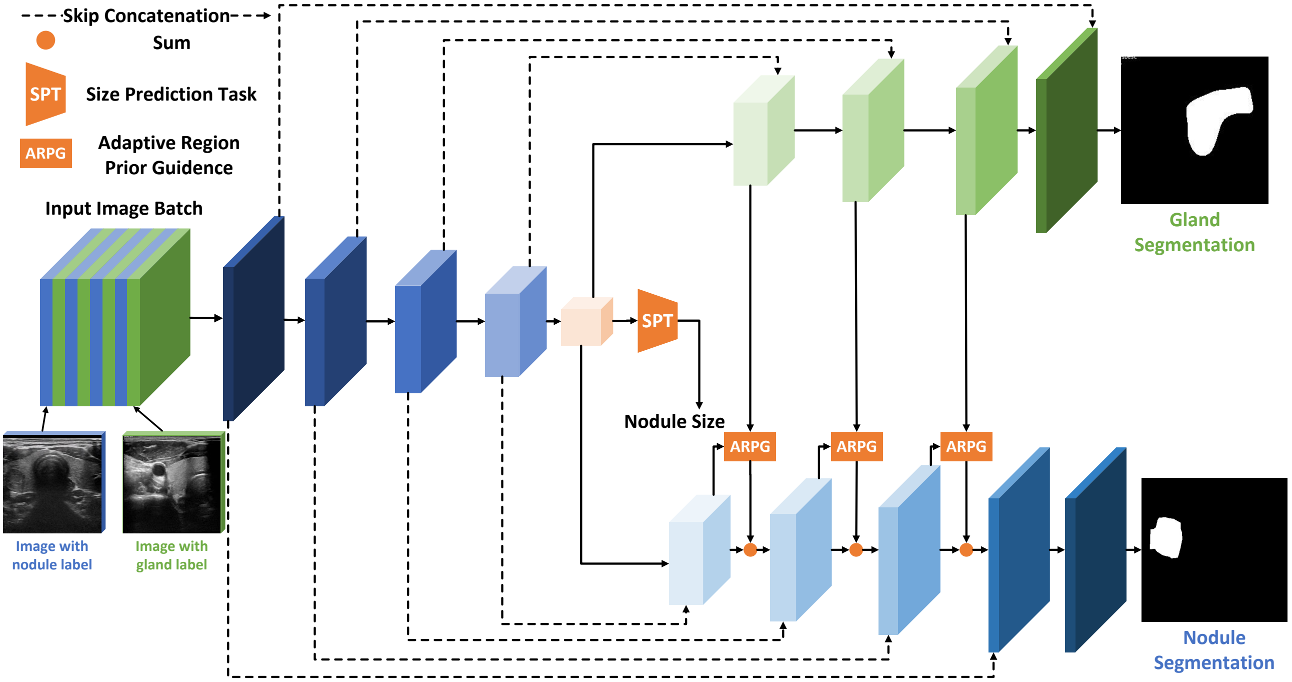

Fig. 4. The pipeline of the thyroid region prior guided feature enhancement network (TRFE+). The upper branch in green learns to segment the thyroid gland. The middle SPT block refers to the nodule size prediction task. The lower branch in blue aims to segment the thyroid nodule region. ARPG refers to the adaptive region prior guidance module.

the upper branch while the nodule images are sent to the inferior branch for loss back-propagation. It is worth noting that the nodule images are also fed to the upper branch for gland region discovery, and the obtained features are further fed to the region prior guidance module to improve the accuracy of the nodule region segmentation. We take the vanilla encoder and the decoder of UNet Ronneberger et al. (2015) in the thyroid gland segmentation branch of TRFE+. We add a refinement layer after the decoder of the nodule segmentation branch because our task focuses on the nodule segmentation. The details of the network architecture are shown in Table 1 of the supplemental materials.

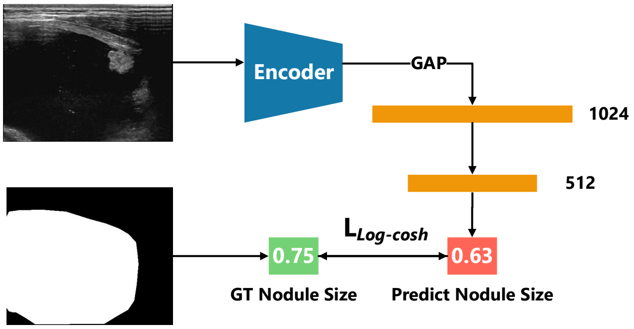

Fig. 5. The pipeline of the size prediction task. GAP refers to the global average pooling processing.

To better make use of the thyroid region prior, we dig deeper into the relationship between the gland region prior and the nodule segmentation task. We discover that not all nodules need the supervision of the thyroid gland. As we can see in Fig. 5, due to the inconsistent size between the thyroid gland and the thyroid nodule, the gland detection branch may regard the region in the nodule as the region out of the gland for large nodules. In this situation, the thyroid gland's prior information could be harmful. Thus, for the large nodules the thyroid gland prior information could be harmful. Thus, we proposed a **S**ize **P**rediction **T**ask (**SPT**) to force the encoder to be aware of the nodule size. Specifically, we add a size prediction module after the fifth layer of the encoder. Let *MLP* be the Multiple-Layers Perceptron network, the predicted size s^{pred} is obtained by:

$$s^{pred} = sigmoid(MLP(f_5)) \tag{1}$$

where *sigmoid* indicates the sigmoid function, f_5 denotes the feature map of the fifth layer. To avoid the over-fitting of the model, we adopts a smooth version of the MSE loss (i.e., Log-cosh loss) Natekin and Knoll (2013) to calculate the loss of the size prediction task L_{size}:

$$L_{size} = \frac{1}{N_{nodule}} \sum_{i=1}^{N_{nodule}} log(cosh(s_i^{pred} - s_i^{gt})) \tag{2}$$

where N_{nodule} denotes the number of the thyroid nodule images in a mini-batch, s^{gt} represents the ground-truth area ratio that the nodule occupied in the images. *log* and *cosh* indicates the log function and the hyperbolic cosine function, respectively.

4.2. Adaptive Region Prior Guidance Module

To take full use of the gland position prior knowledge, we have designed a region prior guidance (RPG) module in our

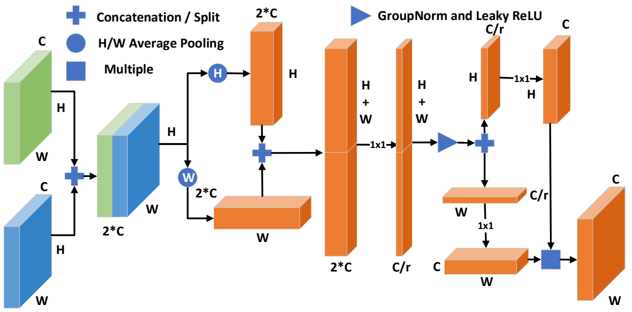

Fig. 6. The adaptive region prior guidance (ARPG) module. The gland segmentation feature map, the nodule segmentation feature map, and the fusion features are represented by blocks in green, blue, and orange, respectively. C, H, and W denotes channel dimension, height, and width of the feature map, respectively.

previous work Gong et al. (2021). However, the RPG module does not significantly improve the performance of the multi-task learning diagram due to the following facts: The explicit feature enhancement module was not able to select the useful semantic information, while the implicit feature enhancement module ignores the relationship between the different channels. Inspired by Hu et al. (2018); Woo et al. (2018); Hou et al. (2021) and the idea of matrix decomposition Geng et al. (2021), we re-design the RPG module by taking both the channel relationship and the position information into account. The new tailor-designed **A**daptive **R**egion **P**rior **G**uidance (**ARPG**) module is shown in Fig. 6.

To give a formulaic definition, we denote $g(\cdot)$ as the 1×1 convolution operation, $\sigma(\cdot)$ as the sigmoid function, $[\cdot, \cdot]$ as the concatenation operation, x_{tg} as the feature map of thyroid gland segmentation branch shaped $C \times H \times W$. Analogously, x_{tn} denotes the feature map shaped $C \times H \times W$ of thyroid nodule segmentation branch. We first concatenate them into one unit x_{concat}:

$$x_{concat} = [x_{tg}, x_{tn}] \tag{3}$$

After that, a direction-aware pooling strategy is applied to capture the long-range spatial information in one direction. Given the input x_{concat} feature map shaped $2C \times H \times W$, we generate the feature vector with the following operation:

$$v_h = \frac{1}{W} \sum_{0 \le i < W} x_{concat}(h, i) \tag{4}$$

Analogously, the vertical feature vector is obtained by:

$$v_w = \frac{1}{H} \sum_{0 \le j < H} x_{concat}(j, w) \tag{5}$$

After that, we first transpose the v_w shaped $C \times 1 \times W$ into v_w^T the shape $C \times W \times 1$, then we concatenate the feature vector v_h and v_w^T into one vector shaped $2C \times (H + W) \times 1$, and use the 1×1 convolution operator to discover the channel relationship by squeezing the channel dimension into C/r, where r is a squeeze factor set to 16 in our experiments. This progress is defined as:

$$v_{squeeze} = g([v_h, v_w^T]) \tag{6}$$

After that, we normalize and activate the feature vector with group normalization Wu and He (2018) and leaky ReLU Maas et al. (2013) function. We split the squeezed vector $v_{squeeze}$ into the vertical feature vector z_h and horizontal feature vector z_w. We transposed the z_w shaped $C/r \times W \times 1$ into the z_w^T shaped $C/r \times 1 \times W$ Then z_h and z_w are expanded with the $1 * 1$ convolution to the feature channels of C followed by the sigmoid function, as follows:

$$x_h = \sigma(g(z_h)) \tag{7}$$

$$x_w = \sigma(g(z_w^T)) \tag{8}$$

where x_h denotes the horizontal feature vector shaped $C \times H \times 1$ and x_w denotes the vertical feature vector shaped $C \times 1 \times W$. Finally, the fused feature map x_{fuse} is obtained by:

$$x_{fuse} = x_h \times x_w \tag{9}$$

The ARPG module is added to the first three layers of the decoder. The feature fusion process is formulated as:

$$y_i = x_{fuse\ i} + x_{tn\ i} \tag{10}$$

where i denotes the feature map index of the decoder.

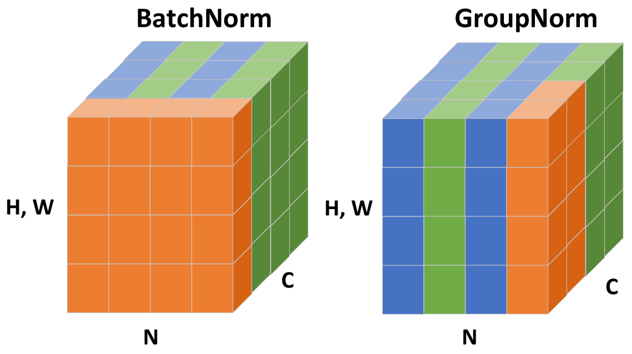

Fig. 7. BatchNorm and GroupNorm. The blue cubes denote the thyroid nodule image, and the green cubes indicate the thyroid gland image. The orange cubes are normalized by the same mean and variance by computing their values.

4.3. Group Normalization

One inherent issue caused by our designed multi-task learning framework is the distribution gap between the TN3K and the TG3K. Since the view and the scale of the images in mixed mini-batch could be different, conventional *BatchNorm* Ioffe and Szegedy (2015) could be helpless due to the domain shift. Thus, we propose a simple yet effective solution that replaces the *BatchNorm* operation with the *GroupNorm* Wu and He (2018) operation. The intuitive visualization result is shown in Fig. 7. The left cube shows how *BatchNorm* works while the right cube shows how *GroupNorm* works. As can be observed, the *BatchNorm* normalizes the features by the mean and variance computed within a mini-batch, i.e., the *N* dimension, which helps to reduce the difficulty of optimization and makes deeper neural networks easier to converge. However, when the samples of a batch come from a different domain or the number of samples in a batch are small, the neural networks could be under-performed. The *GroupNorm* normalizes the batch from the perspective of channel C, which avoids the influence of the different domain samples in the mini-batch. With the GroupNorm operation, our framework achieves lower loss and faster convergence.

4.4. Loss Function

We design a multi-task loss L_{total} to optimize our TRFE+ with the supervision of a thyroid nodule region mask, thyroid gland mask, and thyroid nodule size. The formula is defined as:

$$L_{total} = \sum_{i=1}^{N_{nodule}} L_{nodule\ i} + \lambda * \sum_{j=1}^{N_{gland}} L_{gland\ j} + L_{size} \tag{11}$$

where N_{gland} and N_{nodule} are set to half of the number of samples in a mini-batch, respectively. L_{nodule} and L_{gland} represent the Dice loss function L_{Dice} of the nodule segmentation and the gland segmentation, respectively. λ is a trade-off parameter between the nodule segmentation branch and gland segmentation branch, which is set to 0.5. L_{size} denotes the loss of the nodule size prediction task.

5. Experiments

5.1. Implementation Details

All the models are trained with the NVIDIA Tesla A100 GPU with 40 GB memory. The framework is implemented in PyTorch 1.8.1 with the CUDA 11.1. If the ImageNet pre-trained encoder is available, we use the ImageNet pre-trained weight to initialize the model. For other situations, the weight of the model is initialized by Kaiming-initialization He et al. (2015). Stochastic gradient descent (SGD) is used to optimize the model for 50 epochs. 'Poly' learning rate policy is applied, where $lr = baselr \times \left(1 - \frac{epoch}{epoch_{total}}\right)^{power}$. The $baselr$ is set to 0.01 while $power$ is set to 0.9. The batch size is set to 16. During the training phases, all the images are resized to 224 × 224 with a random horizontal flip at the probability of 0.5. All images are simply resized to 224 × 224 during the model inference. The segmentation performance is obtained by calculating the mean and variance of the five-fold cross-validation (i.e., 2,303 images for training and 576 images for validation) results. To train the TRFE Gong et al. (2021) and the proposed TRFE+, we combine the training set of TN3K with the 2,303 images from TG3K dataset to form a hybridized training dataset that contains 4606 images. The architecture of the proposed TRFE+ is shown in Table 2.

5.2. Evaluation Metrics

To quantitatively evaluate the segmentation performance of the proposed method, we select the following metrics:

- IoU (Intersection Over Union) = TP/(FP + FN);
- DICE (dice coefficient) = 2*TP/(FP + FN + 2 * TP);
- Specificity = TN/(FP+TN);
- PR (Precision) = TP/(TP+FP);
- SE (Sensitivity) = RE (Recall) = TP/(TP+FN);
- Accuracy = (TN+TP)/(TN+TP+FN+FP);
- F1-score = (2*PR*RE)/(PR+RE);
- AUC (Area Under the ROC Curve): this is obtained by the IoU score of all the five folds segmentation results;
- p-value: this is obtained by the t-test in terms of IoU difference between the proposed TRFE+ and other models;
- HD95: the qualified metric of segmentation boundaries by computing the top 95% maximum distance between the predicted boundaries and ground truth.

where TP, FP, TN, FN indicate true positive, false positive, true negative, and false negative, respectively. The F1-score is the harmonic mean of the precision and recall. Any p-value less than 0.05 demonstrates that the proposed method performs significantly better than the other compared methods.

Table 2. The architecture of the TRFE+. "E" and "D" are "Encoder" and "Decoder" for short, respectively.

	output size	layer				
E	224 × 224	3 × 3, 64 3 × 3, 64				
	2 × 2 max pool, stride 2					
	112 × 112	3 × 3, 128 3 × 3, 128				
	2 × 2 max pool, stride 2					
	56 × 56	3 × 3, 256 3 × 3, 256				
	2 × 2 max pool, stride 2					
	28 × 28	3 × 3, 512 3 × 3, 512				
	2 × 2 max pool, stride 2					
	14 × 14	3 × 3, 1024 3 × 3, 1024				
D	output size	layer	output size	layer	output size	layer
	adaptive avg pool		2 × 2 upsample, stride 2		2 × 2 upsample, stride 2	
	1024	Fc layer	28 × 28	3 × 3, 512 3 × 3, 512	28 × 28	3 × 3, 512 3 × 3, 512
	512	Fc layer	2 × 2 upsample, stride 2		ARPG module	
	predicted nodule size	Fc layer	56 × 56	3 × 3, 256 3 × 3, 256	2 × 2 upsample, stride 2	
			2 × 2 upsample, stride 2		56 × 56	3 × 3, 256 3 × 3, 256
			112 × 112	3 × 3, 128 3 × 3, 128	ARPG module	
			2 × 2 upsample, stride 2		2 × 2 upsample, stride 2	
			224 × 224	3 × 3, 64 3 × 3, 64	112 × 112	3 × 3, 128 3 × 3, 128
			224 × 224	3 × 3, 3	ARPG module	
					2 × 2 upsample, stride 2	
					224 × 224	3 × 3, 64 3 × 3, 64
					224 × 224	3 × 3, 32
					224 × 224	3 × 3, 32
					224 × 224	3 × 3, 3

5.3. Comparison with the State-of-the-art Methods

We thoroughly evaluate the proposed method on the test set of TN3K and DDTI which is shown in Table 3. By comparing it with the existing state-of-the-art segmentation methods, the superiority of our TRFE+ has been demonstrated. The FCN and the SegNet are trained with the ImageNet pre-trained VGG Simonyan and Zisserman (2015) backbone, while the Deeplabv3+, CPFNet, TransUNet are trained with the ImageNet pre-trained ResNet He et al. (2016) backbone. The

Table 3. Comparisons with the state-of-the-art semantic segmentation models on the TN3K testset and DDTI. The best result is shown in bold.

TN3K testset	AUC	F1-score	Accuracy	IoU	Dice	HD95	p-value
UNet Ronneberger et al. (2015)	95.01	76.43$_{\pm 0.67}$	96.46$_{\pm 0.11}$	65.99$_{\pm 0.66}$	79.51$_{\pm 1.31}$	18.44$_{\pm 0.75}$	<0.0001
SGUNet Pan et al. (2021)	92.88	76.54$_{\pm 0.43}$	96.54$_{\pm 0.09}$	66.05$_{\pm 0.43}$	79.55$_{\pm 0.86}$	18.16$_{\pm 0.71}$	<0.0001
TRFE Gong et al. (2021)	93.74	78.18$_{\pm 0.72}$	96.71$_{\pm 0.07}$	68.33$_{\pm 0.68}$	81.19$_{\pm 1.35}$	17.96$_{\pm 1.24}$	<0.0001
FCN Long et al. (2015)	95.37	78.39$_{\pm 0.25}$	96.92$_{\pm 0.04}$	68.18$_{\pm 0.25}$	81.08$_{\pm 0.50}$	16.93$_{\pm 0.77}$	<0.0001
SegNet Badrinarayanan et al. (2017)	96.06	77.02$_{\pm 0.85}$	96.72$_{\pm 0.12}$	66.54$_{\pm 0.85}$	79.91$_{\pm 1.69}$	17.13$_{\pm 0.89}$	<0.0001
Deeplabv3+ Chen et al. (2018)	95.80	80.52$_{\pm 0.40}$	**97.19**$_{\pm 0.05}$	70.60$_{\pm 0.49}$	82.77$_{\pm 0.98}$	13.92$_{\pm 0.89}$	0.053
CPFNet Feng et al. (2020)	95.85	80.46$_{\pm 0.37}$	97.17$_{\pm 0.06}$	70.50$_{\pm 0.39}$	82.70$_{\pm 0.78}$	13.56$_{\pm 0.82}$	<0.05
TransUNet Chen et al. (2021)	92.78	79.05$_{\pm 0.47}$	96.86$_{\pm 0.05}$	69.26$_{\pm 0.55}$	81.84$_{\pm 1.09}$	14.92$_{\pm 0.39}$	<0.0001
TRFE+	**97.44**	**81.21**$_{\pm 0.30}$	97.04$_{\pm 0.10}$	**71.38**$_{\pm 0.43}$	**83.30**$_{\pm 0.26}$	**13.23**$_{\pm 0.63}$	**-**
DDTI	AUC	F1-score	Accuracy	IoU	Dice	HD95	p-value
UNet Ronneberger et al. (2015)	84.81	53.49$_{\pm 4.81}$	90.94$_{\pm 0.53}$	42.59$_{\pm 4.16}$	59.74$_{\pm 7.99}$	40.43$_{\pm 6.29}$	<0.0001
SGUNet Pan et al. (2021)	82.98	57.09$_{\pm 2.38}$	91.30$_{\pm 0.35}$	45.90$_{\pm 2.12}$	62.92$_{\pm 4.15}$	34.62$_{\pm 2.12}$	<0.0001
TRFE Gong et al. (2021)	78.60	63.68$_{\pm 1.99}$	92.13$_{\pm 0.25}$	52.72$_{\pm 1.70}$	69.04$_{\pm 3.34}$	34.60$_{\pm 3.39}$	<0.0001
FCN Long et al. (2015)	85.99	64.99$_{\pm 2.19}$	92.67$_{\pm 0.22}$	53.80$_{\pm 2.03}$	69.96$_{\pm 3.98}$	31.03$_{\pm 2.54}$	<0.0001
SegNet Badrinarayanan et al. (2017)	86.03	59.05$_{\pm 2.73}$	91.84$_{\pm 0.33}$	48.36$_{\pm 2.35}$	65.19$_{\pm 4.59}$	37.32$_{\pm 3.40}$	<0.0001
Deeplabv3+ Chen et al. (2018)	89.68	69.86$_{\pm 1.28}$	**93.51**$_{\pm 0.24}$	59.23$_{\pm 1.21}$	74.40$_{\pm 2.39}$	25.45$_{\pm 0.78}$	<0.001
CPFNet Feng et al. (2020)	90.74	70.65$_{\pm 1.33}$	93.25$_{\pm 0.27}$	59.70$_{\pm 1.10}$	74.77$_{\pm 2.18}$	24.79$_{\pm 1.17}$	<0.05
TransUNet Chen et al. (2021)	85.18	70.65$_{\pm 1.62}$	93.11$_{\pm 0.33}$	59.28$_{\pm 1.78}$	74.43$_{\pm 3.50}$	**24.37**$_{\pm 1.06}$	<0.05
TRFE+	**92.56**	**72.20**$_{\pm 1.10}$	93.18$_{\pm 0.26}$	**60.47**$_{\pm 1.08}$	**75.37**$_{\pm 2.14}$	24.60$_{\pm 1.23}$	**-**

proposed TRFE+ considerably exceeds the previous TRFE by 3.19% Jaccard score on TN3K testset and significantly outperforms the TRFE by 7.75% Jaccard index on the DDTI dataset. By taking the advantage of the tailor-designed multi-task learning diagram, the proposed TRFE+ even outperformed other state-of-the-art methods equipped with the powerful ImageNet pre-train backbone.

To illustrate the performance of the proposed method in a more intuitive manner, we further provide the ROC curves for both the TN3K testset and the DDTI data set in Fig. 8. Comparing to the previous state-of-the-art methods like CPFNet Feng et al. (2020) or the powerful transformer-based TransUNet Chen et al. (2021), the proposed method performed well with the following reasons. Firstly, the proposed method efficiently takes the advantage of independently labeled thyroid gland data with a tailor-designed multi-task learning paradigm, which effectively avoids the neural network's misunderstanding of nodules in non-thyroid areas. Secondly, the proposed method includes a region feature enhance module and a size prediction task, which achieves better performance fully exploiting the thyroid region prior and the relationship between the nodule size and the prior intensity. In the end, we take the group norm as the normalization approach to avoid the distribution gap between the two different domains of the source data. Thus, the proposed method achieves much superior performance on the TN3K test set and DDTI dataset.

5.4. Ablation Study

We conduct a comprehensive ablation study in Table 4 to verify the effectiveness of different variants of our the proposed method. As we have demonstrated in our previous work that simply use the thyroid gland data to pre-train the UNet does not bring obvious performance improvement Gong et al. (2021), we mainly focus on the effectiveness of the re-designed method in this paper. In Table 4, the MTNet denotes the vanilla multi-task learning framework with two branches. The TRFE refers to the MTNet with the RPG attention module in Gong et al. (2021). The TRFE-M0 adds the ARPG module to the MTNet. The TRFE-M1 replaces the batch norm in TRFE-M0 with the group norm. TRFE-M2 indicates the TRFE-M1 network with the size prediction task. TRFE+ denotes the TRFE-M2 with the additional refinement layer.

By leveraging the group normalization that normalizes the feature from the channel perspective, our network could avoid suffering from the domain shift between the two *independently labeled dataset*. Furthermore, a tailor designed feature fusion module that taking both the channel relationship and the position information into account. The size prediction task can regularize the encoder to force the model to achieve better performance on the extremely large or small nodule segmentation. By adding the refinement layer after the vanilla decoder of the thyroid nodule segmentation branch, the proposed TRFE+ focuses more on the nodule segmentation task thus achieves better performance. It is worth noting that *GroupNorm* could improve the performance of the task when batch size is extreme small. However, the batch size in our paper is set to 16 which is not small. According to Wu and He (2018), *BatchNorm* outperformed *GroupNorm* when batch size is set to 16, but in this work *GroupNorm* significantly outperformed *BatchNorm*, which proves the effectiveness of our method in another way. Furthermore, we enforce the framework to learn the nodule size with an additional size prediction task, which alleviates the mismatch between the nodule size and the gland size. The sensitivity analysis of the trade-off parameter is shown in Tab. 6. As λ

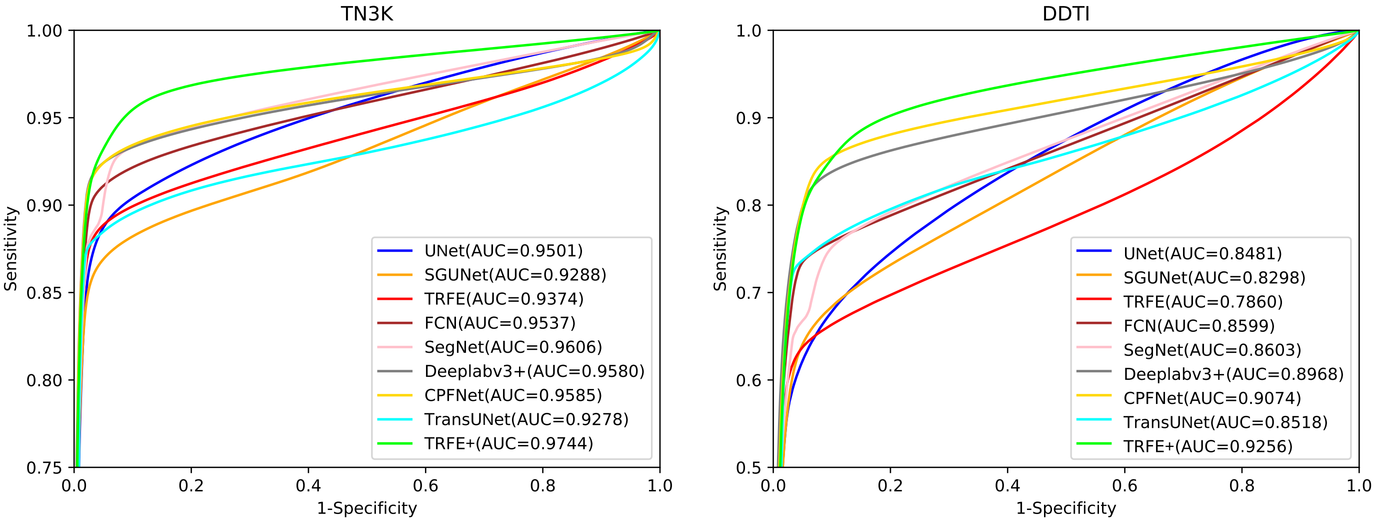

Fig. 8. ROC curves of different algorithms on the TN3K testset and the DDTI dataset.

Table 4. Ablation study on the TN3K test set. N.S. means not significant.

Models	Methods				IoU	p-Value
	ARPG	GN	SPT	R		
MTNet					$68.03_{\pm 0.41}$	<0.001
TRFE					$68.19_{\pm 0.43}$	<0.001
TRFE-M0	✓				$68.61_{\pm 0.45}$	<0.001
TRFE-M1	✓	✓			$70.65_{\pm 0.51}$	<0.05
TRFE-M2	✓	✓	✓		$71.11_{\pm 0.34}$	N.S.
TRFE+	✓	✓	✓	✓	$71.38_{\pm 0.43}$	-

Table 5. Sensitivity analysis of the λ value in equation 11

λ	0.25	0.5	0.75	1
Mean IoU	$71.23_{\pm 0.46}$	$71.38_{\pm 0.43}$	$71.28_{\pm 0.26}$	$70.94_{\pm 0.36}$

changes, the performance of the model does not change significantly, and the model performs best with λ value of 0.5. We also compare with the region guided attention machination named MGA Li et al. (2019) by replacing our ARPG module with their MGA module. The detailed comparison result is shown in Table 6. The analysis of size prediction task is shown in Tab. 8. We regard the nodule with an occupation ratio of more than 0.1 as a large nodule, and the experimental results are shown below. We can see that the size prediction task significantly boost the performance of our method on the large nodules, while also improve the performance on small nodules.

Table 6. Compairson with other location prior based attention method.

Method	IoU	Improvement
TRFE	$68.19_{\pm 0.43}$	-
TRFE+MGA Li et al. (2019)	$68.44_{\pm 0.38}$	+0.25
TRFE+ARPG	$68.61_{\pm 0.45}$	+0.42

Table 7. Analysis of size prediction task.

	Large	Small	Overall
w/o SPT	$77.4_{\pm 0.36}$	$66.1_{\pm 0.63}$	$70.65_{\pm 0.51}$
w SPT	$78.7_{\pm 0.26}$	$66.3_{\pm 0.39}$	$71.11_{\pm 0.34}$

6. Discussion

6.1. Qualitative analysis

The qualitative comparison result is shown in Fig. 9. We can observe that the vanilla UNet achieves a relatively inferior performance, due to the lack of the thyroid gland region prior knowledge. Deeplabv3+, CPFNet, and TransUNet achieve better performance to UNet. However, they still make some mistakes including regarding the non-nodule/non-gland region as the thyroid nodule (second raw and fifth raw in Fig.9) and unable to segment all the nodules in the image (first raw in Fig.9). In contrast, the proposed TRFE+ yields more accurate results. TRFE+ avoids the mistake of identifying non-thyroid gland areas as thyroid nodules as much as possible. Furthermore, with the help of group normalization, the TRFE+ shows much better performance on the DDTI dataset (last two rows of Fig.9).

6.2. Clinical Analysis

In this work, we proposed a model that using the independently labeled thyroid nodule images and thyroid gland images to improve the segmentation accuracy of the thyroid nodule for the first time. This is of great significance because despite the difficulty in annotating medical images, the TRFE+ can make full use of a large amount of respectively labeled data with the shared backbone and three decoders. Automatic thyroid nodule segmentation is not only valuable for the convenience of the clinicians, but also plays an important role for the diagnose of thyroid nodules as the size, shape, even the number of the nodules are statistically significant for thyroid nodule grading.

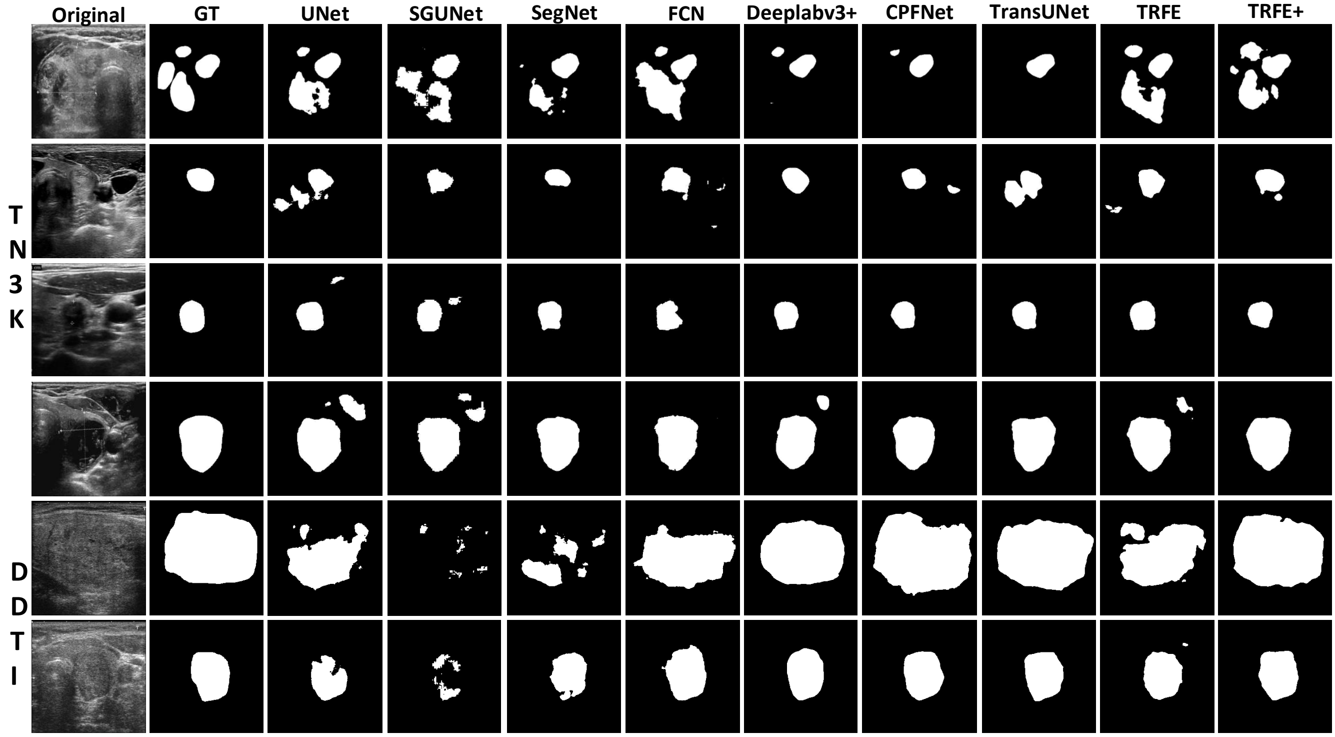

Fig. 9. Qualitative comparison of the algorithms on the TN3K test set and the DDTI dataset.

Based on the segmentation result, we can also locate the periphery region and use the periphery information Mohammadi et al. (2022) around the thyroid nodule to provide better diagnosis accuracy. In the TN3K dataset, we carefully label each nodule independently. Furthermore, our proposed dataset are acquired from different views with different equipment, which better adapts to the real world application of the nodules. However, as we can see from Table. 3, the models trained on the TN3K training set is unable to accurately segment the nodules in the DDTI dataset. Our proposed method significantly outperforms the previous methods by a large margin on the DDTI dataset which is from different domain. A potential explanation is that the batch norm will lead the information reveal of the dataset from a recent work Wu and Johnson (2021). The information reveal does do harm to the transfer ability of the model. Still, the domain shift is a valuable research topic for the further study of thyroid nodules segmentation. We also provide the morphemetric analysis according to Abbasian Ardakani et al. (2019) which is shown in Tab. 9. The proposed method significantly exceeds the UNet with a higher PCC score.

6.3. Limitation

One of the limitations of this work is that this model works not that well on the DDTI dataset, which is obtained from other devices. In the future, we think it will be a valuable topic to further utilize domain adaptation technologies to alleviate the domain shift between two domains. Besides, in this work, we neglect the thyroid nodule segmentation based on B-mode images, and we think it will be valuable to further apply our method to these images.

7. Conclusion

In this paper, we present a novel thyroid nodule segmentation framework named TRFE+ that takes the thyroid region prior as guidance for nodule segmentation. This framework is designed as a multi-task learning paradigm that simultaneously learns the nodule region segmentation, gland region segmentation, and the nodule size prediction. Besides, the ARPG module is incorporated to enhance the thyroid nodule segmentation by utilizing the feature of the thyroid gland segmentation branch. Moreover, we discover the fact that normalizing the data from channel dimension could not only alleviate the domain gap between the TN3K and TG3K during the training phase, but also boost the performance of heterogeneous testset. To the best of our knowledge, the proposed TRFE+ is the first to successfully utilize the independently labeled thyroid nodule data and thyroid gland data, which is valuable as there exists a large amount of independently labeled clinical data. Experiment results on the TN3K testset and the DDTI dataset have demonstrated the effectiveness of the proposed method. We have selected various metrics and conducted paired t-tests on the segmentation IoU of our method and others. The statistical analysis in Section V demonstrates competitive performance of the proposed method.

To promote research in the field of automatic thyroid diagnosis, we contribute TN3K: a challenging, heterogeneous, and precisely labeled benchmark for thyroid nodule segmentation.

Table 8. Morphemetric analysis according to Abbasian Ardakani et al. (2019) based on TN3K testset. "Avg." denotes the average result. "PCC" denotes the Pearson correlation coefficient. "GT" denotes the manual annotated result obtained by radiologist.

	X-coor	Y-coor	Eccentricity	Orientation	Diameter	Area	Perimeter	Convex	Solidity	Extent
GT Avg.	270.83	153.06	0.6695	3.9996	156.77	30390.05	618.71	33615.77	0.9416	0.7288
UNet Avg.	270.01	150.97	0.6928	2.0185	158.75	30051.89	657.73	33075.04	0.9060	0.6763
UNet PCC	0.6670	0.7614	0.2179	0.1276	0.7954	0.7866	0.7368	0.8932	0.1450	0.1233
TRFE+ Avg.	271.63	152.07	0.6915	3.2135	160.46	30763.45	641.96	33562.21	0.9298	0.7010
TRFE+ PCC	0.8635	0.9303	0.4039	0.3395	0.8117	0.7828	0.7414	0.8945	0.2159	0.1936

To guarantee the patient privacy not being revealed, the personal identities of all images have been removed and cannot be reconstructed. We plan to keep on constructing the dataset with more challenging situations such as instance-level nodule masks, labeling the grading clues (e.g., strong echo) from the treatment guidelines within the nodule. We believe these efforts will facilitate the future development of computer-aided thyroid nodule diagnosis.

Acknowledgments

This work was supported in part by the Guangdong Basic and Applied Basic Research Foundation (No.2020B1515020048), in part by the National Natural Science Foundation of China (No.61976250 and No.U1811463) and in part by the Guangzhou Science and technology project (No.202102020633). Dr. Fei Chen is supported by Guangdong Science and Technology Plan Project, 2022.

References

Abbasian Ardakani, A., Bitarafan-Rajabi, A., Mohammadzadeh, A., Mohammadi, A., Riazi, R., Abolghasemi, J., Homayoun Jafari, A., Bagher Shiran, M., 2019. A hybrid multilayer filtering approach for thyroid nodule segmentation on ultrasound images. Journal of Ultrasound in Medicine 38, 629–640.

Alrubaidi, W.M., Peng, B., Yang, Y., Chen, Q., 2016. An interactive segmentation algorithm for thyroid nodules in ultrasound images, in: International Conference on Intelligent Computing, Springer. pp. 107–115.

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. PAMI 39, 2481–2495.

Chen, F., Ye, H., Zhang, D., Liao, H., 2022. Typeseg: A type-aware encoder-decoder network for multi-type ultrasound images co-segmentation. Computer Methods and Programs in Biomedicine 214, 106580.

Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y., 2021. Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 .

Chen, J., You, H., Li, K., 2020. A review of thyroid gland segmentation and thyroid nodule segmentation methods for medical ultrasound images. Computer methods and programs in biomedicine 185, 105329.

Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation, in: ECCV, pp. 801–818.

Durante, C., Grani, G., Lamartina, L., Filetti, S., Mandel, S.J., Cooper, D.S., 2018. The diagnosis and management of thyroid nodules: a review. Jama 319, 914–924.

Fan, D.P., Ji, G.P., Zhou, T., Chen, G., Fu, H., Shen, J., Shao, L., 2020. Pranet: Parallel reverse attention network for polyp segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 263–273.

Feng, S., Zhao, H., Shi, F., Cheng, X., Wang, M., Ma, Y., Xiang, D., Zhu, W., Chen, X., 2020. Cpfnet: Context pyramid fusion network for medical image segmentation. IEEE transactions on medical imaging 39, 3008–3018.

Geng, Z., Guo, M., Chen, H., Li, X., Wei, K., Lin, Z., 2021. Is attention better than matrix decomposition?, in: 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7.

Gong, H., Chen, G., Wang, R., Xie, X., Mao, M., Yu, Y., Chen, F., Li, G., 2021. Multi-task learning for thyroid nodule segmentation with thyroid region prior, in: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), pp. 257–261. doi:10.1109/ISBI48211.2021.9434087.

Gong, H., Cheng, H., Xie, Y., Tan, S., Chen, G., Chen, F., Li, G., 2022. Less is more: Adaptive curriculum learning for thyroid nodule diagnosis, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 248–257.

Gui, L., Li, C., Yang, X., 2017. Medical image segmentation based on level set and isoperimetric constraint. Physica Medica 42, 162–173.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, in: Proceedings of the IEEE international conference on computer vision, pp. 1026–1034.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

Hou, Q., Zhou, D., Feng, J., 2021. Coordinate attention for efficient mobile network design, in: Proceedings of the IEEE conference on computer vision and pattern recognition.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7132–7141.

Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: International conference on machine learning, PMLR. pp. 448–456.

Kumar, V., Webb, J.M., Gregory, A., Meixner, D.D., Knudsen, J., Callstrom, M., Fatemi, M., Alizad, A., 2020. Automated segmentation of thyroid nodule, gland, and cystic components from ultrasound images using deep learning. IEEE Access 8, 63482–63496.

Li, H., Chen, G., Li, G., Yu, Y., 2019. Motion guided attention for video salient object detection, in: Proceedings of the IEEE/CVF international conference on computer vision, pp. 7274–7283.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation, in: CVPR, pp. 3431–3440.

Ma, J., Wu, F., Jiang, T., Zhao, Q., Kong, D., 2017. Ultrasound image-based thyroid nodule automatic segmentation using convolutional neural networks. International Journal of Computer Assisted Radiology and Surgery 12, 1895–1910.

Maas, A.L., Hannun, A.Y., Ng., A.Y., 2013. Rectifier nonlinearities improve neural network acoustic models, in: In Proc. icml, pp. 770–778.

Manh, V.T., Zhou, J., Jia, X., Lin, Z., Xu, W., Mei, Z., Dong, Y., Yang, X., Huang, R., Ni, D., . Multi-attribute attention network for interpretable diagnosis of thyroid nodules in ultrasound images. IEEE transactions on ultrasonics, ferroelectrics, and frequency control .

Mohammadi, A., Mirza-Aghazadeh-Attari, M., Faeghi, F., Homayoun, H., Abolghasemi, J., Vogl, T.J., Bureau, N.J., Bakhshandeh, M., Acharya, R.U., Ardakani, A.A., 2022. Tumor microenvironment, radiology, and artificial intelligence: Should we consider tumor periphery? Journal of Ultrasound in Medicine 41.

Natekin, A., Knoll, A., 2013. Gradient boosting machines, a tutorial. Frontiers in neurorobotics 7, 21.

Pan, H., Zhou, Q., Latecki, L.J., 2021. Sgunet: Semantic guided unet for thyroid nodule segmentation, in: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), pp. 630–634. doi:10.1109/ISBI48211.2021.9434051.

Pedraza, L., Vargas, C., Narváez, F., Durán, O., Muñoz, E., Romero, E., 2015. An open access thyroid ultrasound image database, in: 10th International Symposium on Medical Information Processing and Analysis, pp. 188–193. doi:10.1117/12.2073532.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: MICCAI, pp. 234–241.

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition, in: 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings.

Song, R., Zhu, C., Zhang, L., Zhang, T., Luo, Y., Liu, J., Yang, J., 2022. Dual-branch network via pseudo-label training for thyroid nodule detection in ultrasound image. Applied Intelligence , 1–17.

Sun, J., Li, C., Lu, Z., He, M., Zhao, T., Li, X., Gao, L., Xie, K., Lin, T., Sui, J., et al., 2022. Tnsnet: Thyroid nodule segmentation in ultrasound imaging using soft shape supervision. Computer Methods and Programs in Biomedicine 215, 106600.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need, in: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA, pp. 5998–6008.

Wang, L., Zhang, L., Zhu, M., Qi, X., Yi, Z., 2020. Automatic diagnosis for thyroid nodules in ultrasound images by deep neural networks. Medical Image Anal. 61, 101665. doi:10.1016/j.media.2020.101665.

Woo, S., Park, J., Lee, J.Y., Kweon, I.S., 2018. Cbam: Convolutional block attention module, in: Proceedings of the European conference on computer vision (ECCV), pp. 3–19.

Wu, J., Fang, H., Shang, F., Yang, D., Wang, Z., Gao, J., Yang, Y., Xu, Y., 2022. Seatrans: Learning segmentation-assisted diagnosis model via transformer, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 677–687.

Wu, Y., He, K., 2018. Group normalization, in: Proceedings of the European conference on computer vision (ECCV), pp. 3–19.

Wu, Y., Johnson, J., 2021. Rethinking "batch" in batchnorm. ArXiv abs/2105.07576.

Wunderling, T., Golla, B., Poudel, P., Arens, C., Friebe, M., Hansen, C., 2017. Comparison of thyroid segmentation techniques for 3d ultrasound, in: Medical Imaging, p. 1013317.

Yang, W., Dong, Y., Du, Q., Qiang, Y., Wu, K., Zhao, J., Yang, X., Zia, M.B., 2021. Integrate domain knowledge in training multi-task cascade deep learning model for benign-malignant thyroid nodule classification on ultrasound images. Eng. Appl. Artif. Intell. 98, 104064. doi:10.1016/j.engappai.2020.104064.

Ying, X., Yu, Z., Yu, R., Li, X., Yu, M., Zhao, M., Liu, K., 2018. Thyroid nodule segmentation in ultrasound images based on cascaded convolutional neural network, in: ICONIP, pp. 373–384.

Yu, M., Han, M., Li, X., Wei, X., Jiang, H., Chen, H., Yu, R., 2022. Adaptive soft erasure with edge self-attention for weakly supervised semantic segmentation: Thyroid ultrasound image case study. Computers in Biology and Medicine 144, 105347. URL: https://www.sciencedirect.com/science/article/pii/S0010482522001391, doi:https://doi.org/10.1016/j.compbiomed.2022.105347.

Yuan, Y., Li, C., Xu, L., Zhu, S., Hua, Y., Zhang, J., 2022. Csmnet: Automatic joint segmentation of intima-media complex and lumen in carotid artery ultrasound images. Computers in Biology and Medicine , 106119URL: https://www.sciencedirect.com/science/article/pii/S0010482522008277, doi:https://doi.org/10.1016/j.compbiomed.2022.106119.

Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J., 2019. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE transactions on medical imaging 39, 1856–1867.