2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Thyroid Nodule Segmentation and Classification Using Deep Convolutional Neural Network and Rule-based Classifiers

Atefeh Shahroudnejad[1,2], Roberto Vega[2], Amir Forouzandeh[2], Sharanya Balachandran[1,2], Jacob Jaremko[1,2], Michelle Noga[1], Abhilash Rakkunedeth Hareendranathan[1], Jeevesh Kapur[2,3], Kumaradeven Punithakumar[1]

Abstract— **Thyroid cancer has a high prevalence all over the world. Accurate thyroid nodule diagnosis can lead to effective treatment and decrease the mortality rate. Ultrasound imaging is a safe, portable, and inexpensive tool for thyroid nodule monitoring. However, the widespread use of ultrasound has also resulted in over-diagnosis and over-treatment of nodules. There is also large variability in the assessment and characterization of nodules. Thyroid nodule classification requires precise delineation of the nodule boundary which is tedious and time-consuming. Automatic segmentation of nodule boundaries is highly desirable, however, it is challenging due to the wide range of nodule appearances, shapes, and sizes. In this study, we propose an end-to-end pipeline for nodule segmentation and classification. A residual dilated UNet (resDUnet) model is proposed for nodule segmentation. The output of resDUnet is fed to two rule-based classifiers to categorize the composition and echogenicity of the segmented nodule. We evaluate our segmentation method on a large dataset of 352 ultrasound images reviewed by a certified radiologist. When compared with ground-truth, resDUnet gives a higher Dice score than the standard UNet (82% vs. 81%). Our method requires minimal user interaction and it is robust to reasonable variations in the user-specified region-of-interest. We expect the proposed method to reduce variability in thyroid nodule assessment which results in more efficient and cost-effective monitoring of thyroid cancer.**

I. Introduction

There has been a substantial increase in the number of thyroid cancer cases reported in the United States over the past decades [1], [2]. A total of 2170 deaths were attributed to thyroid cancer in 2019 [2]. Fortunately, almost 90% of nodules seen in the thyroid are benign and are unlikely to grow in size and become cancerous, even if they grow [3]. Nodules larger than 1 cm are considered suspicious based on their echogenic texture and they are recommended for biopsy via fine-needle aspiration (FNA). However, FNA is highly invasive, costly, and sometimes inconclusive. When combined with the low specificity of physical examinations, this results in over-diagnosis and over-treatment of thyroid nodules, which is a major financial burden on healthcare systems. Ultrasound imaging is the primary diagnostic tool for thyroid examination, but it is subject to high variability in interpretation. To standardize the reporting and characterization of thyroid nodules, the American College of Radiology introduced the Thyroid Imaging Reporting and Data System (TIRADS) [4], which is based on five characteristics of the thyroid nodule including echogenicity, composition, shape, margin and echogenic foci. There are various modifications of the TIRADS to reduce the variability in reporting nodules. However, the fundamental limitation is that individual characteristics of the nodule are determined manually making the assessment subjective.

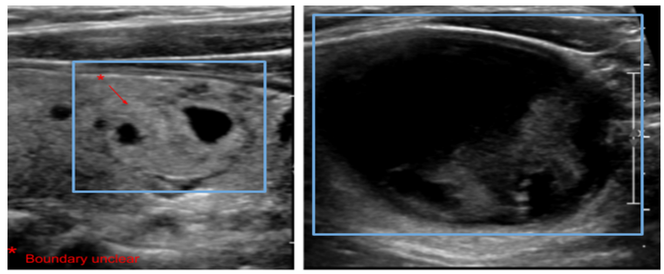


Fig. 1. Examples of thyroid nodules as seen in ultrasound. Note that the size, echogenicity and composition varies based on type of nodule.

A key aspect of thyroid nodule assessment is precise segmentation of the boundary of the nodule. As shown in Figure 1, the pixel intensity distribution inside the nodules varies considerably depending on the nodule composition. Moreover, nodules with unclear boundaries, and ultrasound images characteristics, especially heavy noise and artifacts, make the nodule segmentation task more challenging. Hence, manual segmentation of thyroid nodules is usually time-consuming and tedious. To address these issues, a computer-aided diagnosis (CAD) system is needed to automatically segment nodules, eliminate this variability, and provide fast and reliable risk stratification of the segmented nodules. There are some feature-based and conventional approaches proposed for thyroid and nodule segmentation. Koundal *et al.* uses spatial information with neutrosophic clustering and level-sets for segmentation of thyroid nodules [5]. A similar approach with neutrosophic speckle reduction is proposed in [6]. Illanes *et al.* uses wavelet decomposition and K-means clustering to identify thyroid tissue patches in ultrasound images [7]. In similar lines, patch-based and region-of-interest (ROI) based approaches have also been used to classify thyroid nodules [8][9].

Motivated by the success of deep learning in image segmentation tasks, recent approaches focus on deep models for

*Atefeh Shahroudnejad was supported by the MITACS Accelerate program supported by MEDO.ai.

[1]Atefeh Shahroudnejad, Sharanya Balachandran, Jacob Jaremko, Michelle Noga, Abhilash Rakkunedeth Hareendranathan, and Kumaradeven Punithakumar are with the Department of Radiology and Diagnostic Imaging, University of Alberta shahroud@ualberta.ca

[2]Atefeh Shahroudnejad, Roberto Vega, Amir Forouzandeh, Sharanya Balachandran, Jeevesh Kapur, and Jacob Jaremko are with MEDO.ai, 32 Carpenter Street, Singapore

[3]Jeevesh Kapur is also with the Department of Diagnostic Imaging, National University Hospital, Singapore

nodule and thyroid segmentation. Ma *et al.* proposes a deep multi-layer convolutional neural network (CNN) with 15 convolutional layers and 2 pooling layers for thyroid nodule segmentation. An 8 layer fully convolutional network (FCN) is proposed in [10]. Zhou *et al.* uses the UNet architecture [11] to interactively segment thyroid nodules based on a user-defined ROI [12]. Poudal *et al.* developed a 3D UNet model [13] for thyroid segmentation which outperformed level-sets, graph cut, and a pixel-based classifier in terms of Dice score and Hausdorff distance [14]. A major modification to the UNet model proposed recently is UNet++ [15] which is a deeply supervised encoder-decoder network where skip connections of the standard UNet are replaced with nested dense layers. This would reduce the semantic gap between features extracted from corresponding layers of the encoder and decoder paths of the network.

In this paper, we propose adding residual connections to convolution blocks of the UNet model and also embedding dilated convolution layers within the architecture. Residual structure improves learning ability and accelerates convergence [16]. The additional convolution layers generate richer multi-scale features to map the encoding and decoding paths of the UNet. We compare our network with the standard UNet and UNet++.

II. Methodology

An overview of the proposed CAD system is shown in Figure 2. The input to the CAD system is a 2D ultrasound image. The network generates a segmentation mask for the nodule based on a user-defined ROI (indicated by the red rectangular box in Figure 2). The network is able to delineate nodules in transverse (TRX) and sagittal (SAG) thyroid ultrasound scans. Using the segmentation mask and pixel information in the original image, two rule-based classifiers are developed to categorize the echogenicity and composition of the nodule to estimate the risk. TIRADS classification defines composition categories as cystic, spongiform, mixed solid cystic, and solid and, echogenicity categories as anechoic, hyperechoic, isoechoic, hypoechoic, Very hypoechoic. The composition classifier consists of five different image processing based stages including (1) morphological erosion to remove peripheral calcification, (2) hard thresholding to find solid percentage, (3) adaptive histogram equalization for image enhancement, (4) Isodata thresholding to find dark regions, and (5) morphological closing to find the number of liquid blobs. Results of these stages are compared to the predefined clinical rules and the final decision about the composition category is made. The echogenicity classifier is composed of the following stages: (1) using composition stages to find solid part of the nodule, (2) morphological dilation and erosion to find both the area surrounding the nodule and the area inside the nodule (exclude the nodule boundary), (3) adaptive thresholding to remove very dark and very bright pixels from both areas, (4) calculating the average of brightness for the remained pixels of each area and comparing them with each other. The decision about echogenicity is made based on the clinical rules and the comparison result.

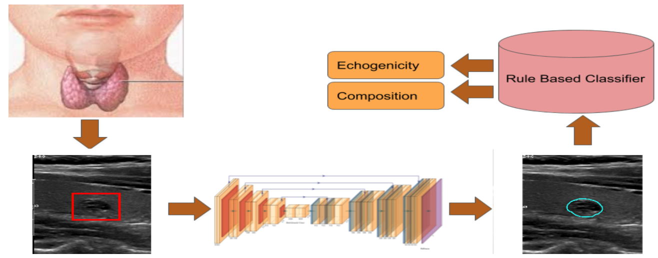


Fig. 2. Overview of the proposed CAD system with a deep learning based segmentation of the nodule boundary and rule-based classifiers.

A. Network Architecture

The UNet architecture uses a series of convolution and max pooling operations in the encoding path to learning image features. The spatial and contextual information of these features is reconstructed in the decoding path through transposed convolutions and skip connections from the encoder. The newly proposed resDUnet improves segmentation results by adding residual shortcut connections [16] in building blocks and embedding dilated convolution layers in the bottleneck part of the network (Figure 3). Residual connections have the ability to eliminate the vanishing and exploding gradient problem, which leads to more consistent training of the neural network [16]. Dilated convolution layers apply 3×3 convolution with different dilation rates, which is defined as:

$$(F *_l k)(p) = \sum_{s+lt=p} F(s)k(t). \tag{1}$$

where $*_l$ is the dilated convolution operator. F and k are discrete function and discrete 3x3 filter, respectively. Considering that dilated convolutions increase receptive field while keeping the resolution, and different dilation rates apply different receptive fields, more robust features are extracted in different scales.

B. Network Training

We leveraged Keras with Tensorflow backend as a deep learning framework in Python programming language for the design and implementation of resDUnet. We trained the network with Dice coefficient based loss function and Adam optimizer [17] with an initial learning rate of 10^{-5}. Keras EarlyStopping strategy was used to avoid over-fitting. The maximum epoch number has been set to 100. Both input and output size for the model are $128 \times 128 \times 1$. The total number of trainable parameters of resDUnet is 1741833. Our model has been trained on GeForce GTX 1080 GPU with 12GB memory.

III. Experimental Results

A. Ultrasound Image Dataset

The ultrasound image data used for training and testing models were obtained retrospectively from our data collection center. This study was approved by the health research

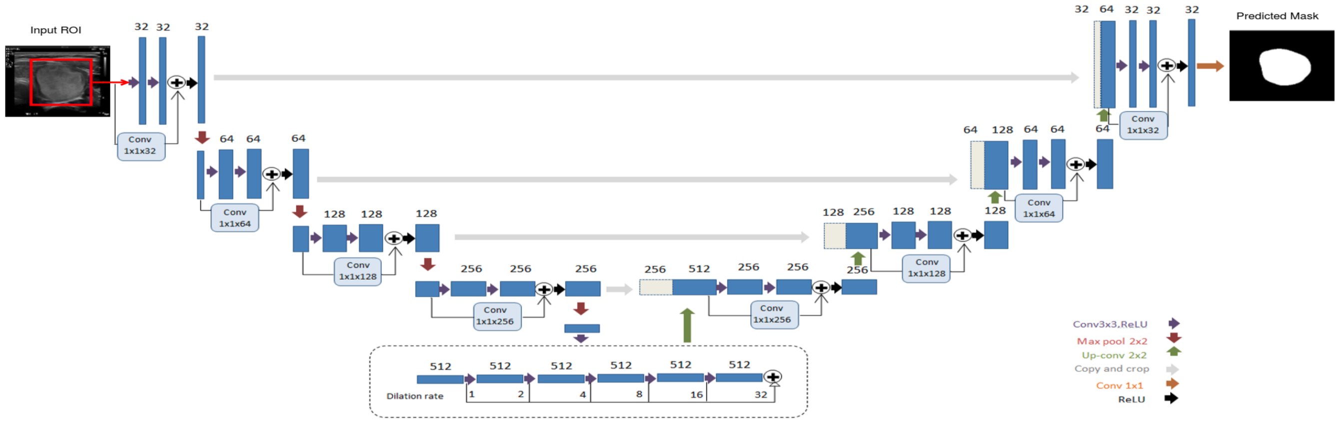
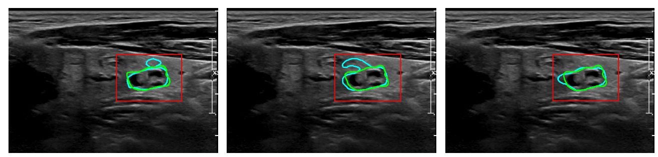


Fig. 3. Schematic architecture of the proposed resDUnet.

ethics board of the University of Alberta. The data comprises SAG and TRX cine-sweeps acquired using Philips i22 and GE Logiq ultrasound scanners.

The training set includes 4266 2D image slices containing thyroid nodules which have been acquired from 63 SAG and TRX sweeps of 41 patients. The test set includes 352 images from 141 patients. The boundary of each nodule was manually delineated by medical experts. Using the segmentation masks, we generated ROIs around nodules which were used as the input to the network. In order to account for variability in the manually selected ROI, we randomly changed the centroid and dimensions of the ROI which generated an augmented dataset of 12798 images. Moreover, each nodule was categorized based on echogenicity and composition by a medical expert.



(a) UNet (b) UNet++ (c) Proposed resDUNet

Fig. 4. Qualitative comparison of ground-truth (green) and segmentation results (blue) for different methods. ROI has shown by a red rectangle.

TABLE I
COMPARISONS AGAINST UNET AND UNET++.

Model	Large >50k pixels 51 images	Medium 10k-50k pixels 150 images	Small <10k pixels 151 images	Total 352 images
	Dice	Dice	Dice	Dice
UNet [11]	90%	87%	73%	81%
UNet++[15]	91%	87%	71%	80%
resDUnet	91%	87%	**74%**	**82%**

B. *Qualitative and Quantitative Analysis*

To evaluate the performance of the proposed resDUnet, we compare our model with standard UNet [11] and one of its variants, UNet++ [15]. As we can see in Table. I, the average Dice score of resDUNet is 82% on the entire test set, which is higher than UNet and UNet++. To better realize the performance of resDUnet on different sizes of nodules, we categorize nodules into 3 subgroups (greater than 50k pixels, as large, 10k–50k pixels, as a medium, and less than 10k pixels, as small nodules) and compare the average Dice score in each subgroup.

Figure 4 shows the qualitative comparison of resDUnet with UNet and UNet++. As we can see, resDUnet segmentation result is closer to the ground-truth than the two other networks. Moreover, it is able to delineate the nodule even with a relatively large ROI. Boxplot diagram of the Dice scores has been presented in Figure 5. We also use two rule-based classifiers to categorize the composition and echogenicity of segmented nodules. The agreement of these categorizations with ground-truth assessments of radiologists, in terms of Kappa score, is 0.68 for echogenicity and 0.51 for composition, which are relatively high scores.

IV. DISCUSSION

Earlier feature-based thyroid and nodule segmentation approaches generally make implicit assumptions about the data. In contrast, our proposed method is data-driven and does not make any prior assumptions about the data.

Regarding nodule classification, instead of directly classifying nodules based on their ROIs, we develop a more explainable workflow by preliminary segmenting nodule boundaries. We compared the segmentation results obtained from standard UNet, UNet++, and resDUNet models on our dataset. Our resDUnet segmentation results were closest to the ground-truth, which indicates that extracted features in resDUnet are complementary to each other. The improvement in the Dice score was most significant in small nodules (as shown in Figure 5), which potentially increases the sensitivity in screening. We also designed a rule-based engine that predicts the composition and echogenicity of the segmented nodule based on the pixel intensities within the segmentation mask and surrounding area. The predicted values of composition and echogenicity showed high agreements with expert assessment (Kappa score is 0.68 for echogenicity and 0.51 for composition) considering the agreement between experts, which ranges between $(0.45 - 0.57)$ for echogenicity and

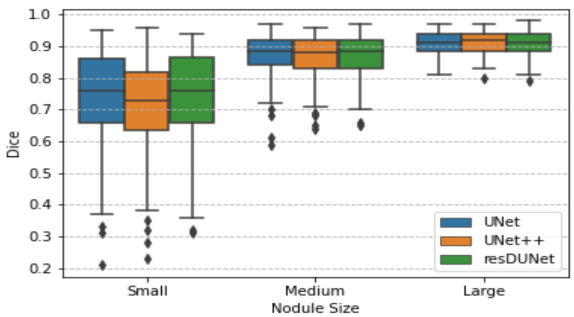


Fig. 5. Boxplot of Dice scores of UNet, UNet++ and resDUnet segmentation algorithms on three categories of nodules (small, medium and large), and for the entire test set.

$(0.59-0.64)$ for composition [18], [19]. Agreement between non-expert readers is less for both echogenicity $(0.34-0.46)$ and composition $(0.18-0.36)$ [18], [19]. The agreement of our CAD system with an expert radiologist is higher than the range of agreement between experts for echogenicity and close to the experts agreement in composition.

The limitation of the proposed method is that it requires initial user interaction for defining prior ROI. Hence, for future work, we will develop a fully automated CAD system to detect thyroid nodules first and then segment these nodules without the need for user-defined ROIs. There is also scope for improvement in the segmentation of large nodules which have heterogeneous textures. As an extension of the current work, we also plan to collect more data in each nodule size category and develop a multi-head ensemble network to integrate different nodule sizes and refine the final segmentation mask.

V. Conclusions

In this paper, we developed a CAD system for analyzing and categorizing thyroid nodules from cine-sweeps acquired in TRX and SAG orientations. The proposed network created more accurate segmentation masks, in terms of Dice score than UNet and UNet++. We also proposed two rule-based classifiers to categorize nodules based on their segmentation masks. The results demonstrated high agreements with categorizations made by radiologists. The proposed CAD system is expected to reduce inter-observer variability, may result in an overall improvement in thyroid cancer management.

References

[1] Dawnelle Topstad and James A Dickinson, "Thyroid cancer incidence in canada: a national cancer registry analysis," *CMAJ open*, vol. 5, no. 3, pp. E612, 2017.

[2] Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal, "Cancer statistics, 2019," *CA: a cancer journal for clinicians*, vol. 69, no. 1, pp. 7–34, 2019.

[3] Cosimo Durante, Giuseppe Costante, Giuseppe Lucisano, Rocco Bruno, Domenico Meringolo, Alessandra Paciaroni, Efisio Puxeddu, Massimo Torlontano, Salvatore Tumino, Marco Attard, et al., "The natural history of benign thyroid nodules," *Jama*, vol. 313, no. 9, pp. 926–935, 2015.

[4] Edward G Grant, Franklin N Tessler, Jenny K Hoang, Jill E Langer, Michael D Beland, Lincoln L Berland, John J Cronan, Terry S Desser, Mary C Frates, Ulrike M Hamper, et al., "Thyroid ultrasound reporting lexicon: white paper of the acr thyroid imaging, reporting and data system (tirads) committee," *Journal of the American college of radiology*, vol. 12, no. 12, pp. 1272–1279, 2015.

[5] Deepika Koundal, Savita Gupta, and Sukhwinder Singh, "Automated delineation of thyroid nodules in ultrasound images using spatial neutrosophic clustering and level set," *Applied Soft Computing*, vol. 40, pp. 86–97, 2016.

[6] Deepika Koundal, Savita Gupta, and Sukhwinder Singh, "Computer aided thyroid nodule detection system using medical ultrasound images," *Biomedical Signal Processing and Control*, vol. 40, pp. 117–130, 2018.

[7] Alfredo Illanes, Nazila Esmaeili, Prabal Poudel, Sathish Balakrishnan, and Michael Friebe, "Parametrical modelling for texture characterization—a novel approach applied to ultrasound thyroid segmentation," *PloS one*, vol. 14, no. 1, pp. e0211215, 2019.

[8] Antonin Prochazka, Sumeet Gulati, Stepan Holinka, and Daniel Smutek, "Patch-based classification of thyroid nodules in ultrasound images using direction independent features extracted by two-threshold binary decomposition," *Computerized Medical Imaging and Graphics*, vol. 71, pp. 9–18, 2019.

[9] Qin Yu, Tao Jiang, Aiyun Zhou, Lili Zhang, Cheng Zhang, and Pan Xu, "Computer-aided diagnosis of malignant or benign thyroid nodes based on ultrasound images," *European Archives of Oto-rhino-laryngology*, vol. 274, no. 7, pp. 2891–2897, 2017.

[10] Xuewei Li, Shuaijie Wang, Xi Wei, Jialin Zhu, Ruiguo Yu, Mankun Zhao, Mei Yu, Zhiqiang Liu, and Shupei Liu, "Fully convolutional networks for ultrasound image segmentation of thyroid nodules," in *2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*. IEEE, 2018, pp. 886–890.

[11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[12] Shujun Zhou, Hong Wu, Jie Gong, Ting Le, Hao Wu, Qin Chen, and Zenglin Xu, "Mark-guided segmentation of ultrasonic thyroid nodules using deep learning," in *Proceedings of the 2nd International Symposium on Image Computing and Digital Medicine*. ACM, 2018, pp. 21–26.

[13] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2016, pp. 424–432.

[14] Prabal Poudel, Alfredo Illanes, Debdoot Sheet, and Michael Friebe, "Evaluation of commonly used algorithms for thyroid ultrasound images segmentation and improvement using machine learning approaches," *Journal of healthcare engineering*, vol. 2018, 2018.

[15] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 3–11. Springer, 2018.

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[17] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[18] Hyun Gi Kim, Jin Young Kwak, Eun-Kyung Kim, Seon Hyeong Choi, and Hee Jung Moon, "Man to man training: can it help improve the diagnostic performances and interobserver variabilities of thyroid ultrasonography in residents?," *European journal of radiology*, vol. 81, no. 3, pp. e352–e356, 2012.

[19] Sung Hun Kim, Chang Suk Park, So Lyung Jung, Bong Joo Kang, Jee Young Kim, Jae Jung Choi, Ye Il Kim, Jin Kyung Oh, Jung Suk Oh, Hanna Kim, et al., "Observer variability and the performance between faculties and residents: Us criteria for benign and malignant thyroid nodules," *Korean journal of radiology*, vol. 11, no. 2, pp. 149–155, 2010.