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

Integrating User-Input into Deep Convolutional Neural Networks for Thyroid Nodule Segmentation

Rajshree Daulatabad[1], Roberto Vega[2], Jacob L. Jaremko[2,4], Jeevesh Kapur[2,3],
Abhilash R. Hareendranathan[4], Kumaradeven Punithakumar[4]

***Abstract*— Delineation of thyroid nodule boundaries is necessary for cancer risk assessment and accurate categorization of nodules. Clinicians often use manual or bounding-box approach for nodule assessment which leads to subjective results. Consequently, agreement in thyroid nodule categorization is poor even among experts. Computer-aided diagnosis systems could reduce this variability by minimizing the extent of user interaction and by providing precise nodule segmentations. In this study, we present a novel approach for effective thyroid nodule segmentation and tracking using a single user click on the region of interest. When a user clicks on an ultrasound sweep, our proposed model can predict nodule segmentation over the entire sequence of frames. Quantitative evaluations show that the proposed method out-performs the bounding box approach in terms of the dice score on a large dataset of 372 ultrasound images. The proposed approach saves expert time and reduces the potential variability in thyroid nodule assessment. The proposed one-click approach can save clinicians time required for annotating thyroid nodules within ultrasound images/sweeps. With minimal user interaction we would be able to identify the nodule boundary which can further be used for volumetric measurement and characterization of the nodule. This approach can also be extended for fast labeling of large thyroid imaging datasets suitable for training machine-learning based algorithms.**

***Index Terms*— Thyroid nodule detection, Thyroid nodule tracking, Deep learning, Ultrasound image segmentation, medical diagnosis**

I. INTRODUCTION

Thyroid cancer is very common in North American populations with a steady increase in the number of cases diagnosed every year [1], [2]. In 2019, a total of 52,070 cases of thyroid cancer were diagnosed in the United States alone [3] and it is currently the commonest cancer in women aged between 20 and 34 [3]. Computed Tomography (CT) [4], [5] and Magnetic Resonance Imagining (MRI) [6] have been used for thyroid nodule assessment, however, these modalities are expensive. Alternatively, the ultrasound is safe, easily portable and inexpensive modality compared to MRI and CT, hence, it can be used as a viable means for universal thyroid cancer screening. A major concern with mass imaging-based screening programs is the chance of a large number of false positives given that more than 90% of thyroid nodules are benign [7]. Over-diagnosis often results in needless and painful biopsies or in some cases over-treatment. Ultrasound examination, in particular, is highly user-dependent due to the variability in scan acquisition and human interpretation. The American College of Radiology (ACR) introduced the Thyroid Imaging Reporting and Data System (TIRADS) [8] which recommends reporting five characteristics of the nodule - echogenicity, composition, shape, margin and presence of calcification. However, these individual characteristics are determined manually making it user-dependent. Artificial Intelligence (AI) techniques that quantify the nodule with minimal user interaction can reduce this variability in reporting. AI techniques has been an excellent aid to radiologists [9] in general and it can be used for early detection of thyroid cancer and for quantitative measurement of nodule size in follow-up scans. With use of AI techniques we can segment the nodule with high accuracy and reliability leading to more accurate categorization of the five nodule characteristics. However, image segmentation from ultrasound is challenging due to specific characteristics such as speckle noise, blurry nodule boundaries, variations in probe position, low contrast, reflection and shadowing effects. Manual segmentation in thyroid ultrasound images is time-consuming, tedious and prone to user bias [4]. On the other hand, fully automated techniques often lack precision due to the wide variety of echogenic textures seen in nodules. Consequently, most Computer Assisted Diagnostic (CAD) systems for thyroid nodule segmentation are semi-automated and use a region-of-interest (ROI) to localize the nodule.

In this paper, we develop an approach to reduce the initial user interaction to a single mouse click. We also extend our approach to segmentation and tracking of a nodule in ultrasound cine sweeps which provides more complete and three dimensional visualization of the nodules.

II. METHODOLOGY AND EXPERIMENTAL SETUP

A. Overview

An overview of our approach is given in Figure 1. The nodule to be segmented in the ultrasound image is indicated by the user via a mouse click. The user clicks on the ultrasound image where the nodule of interest is present. This click point is converted into a 4×4 pixel dot image located centrally at the click as shown in Fig. 1

This dot-image is a binary image of the same size as the original ultrasound image and has a pixel-value of 1 on the

*Rajshree Daulatabad was supported by Mitacs Accelerate Program, York University, University of Alberta and MEDO.ai.

[1]Rajshree Daulatabad is with York University, Toronto, Canada (rajshree@16bit.ai).

[2] Roberto Vega, Jeevesh Kapur and Jacob L. Jaremko are with MEDO.ai, Edmonton, Canada.

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

[4] Jacob L. Jaremko, Abhilash R. Hareendranathan and Kumaradeven Punithakumar are with the Department of Radiology and Diagnostic Imaging, University of Alberta, Edmonton, Canada.

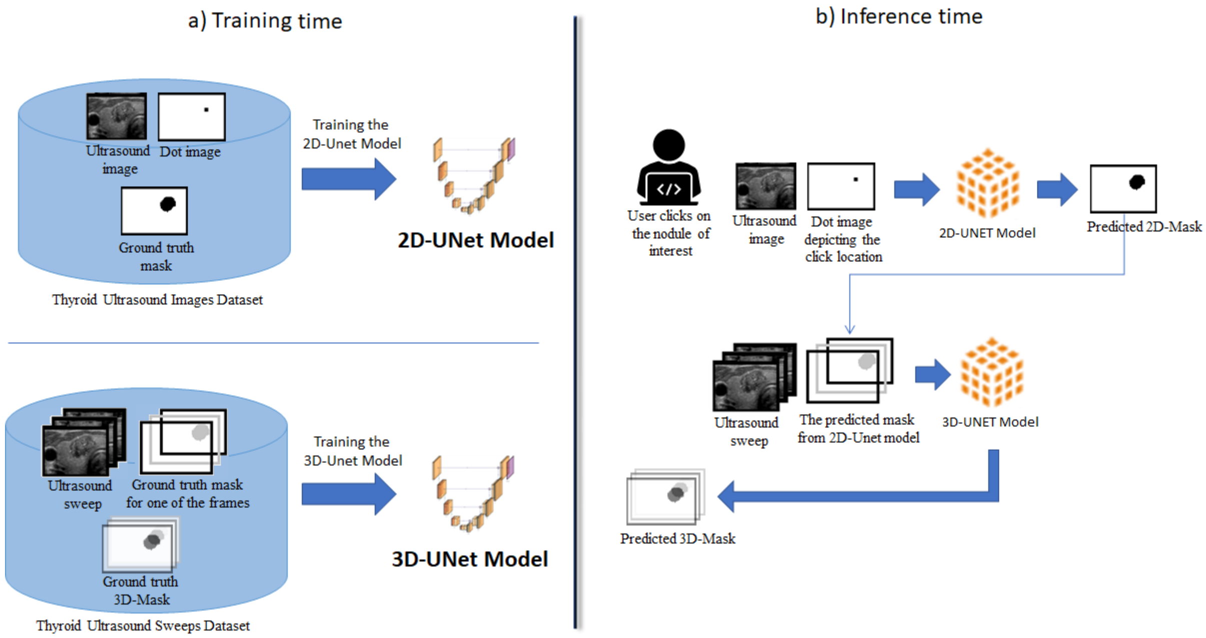

Fig. 1. Overview of the proposed approach for segmenting the nodule from a single 2D image (upper half of (a)) and a cine sweep (lower half of (a)). The proposed approach requires only a single mouse click to perform the thyroid nodule segmentation across the ultrasound sweep predicting the 3D mask for the nodule as shown above in the inference time (b).

click point and 0 elsewhere. The original ultrasound image and the dot image are both inputs to the DL model. We use a deeper, modified UNet model which is considered as the baseline model for medical image segmentation tasks and train with inputs in 2 channels, channel 1 shows the original ultrasound image and channel 2 shows the dot-image both resized (from 600×800 to 512×512) to fit the UNet input shape. The final network output is a mask indicating nodule extent in the ultrasound sweep images. This approach is further extended to ultrasound sweep as shown in the inference time Fig. 1 b). At the inference time, we use the 2D-UNet output along with the ultrasound sweep frames as input to the 3D-UNet model predicting the nodule boundaries across the frames. We also compare our model with the typical use case of bounding box approach where the user provides with a region of interest as a bounding box around the nodule (both tightly bound as well as loosely bound with 10% margin error is experimented). Refer section III for more details.

B. Dataset

The ultrasound images used were retrieved retrospectively from our institution's Picture Archiving and Communication System (PACS). The study was approved by the human research ethics board of the University of Alberta. We used a total of 865 ultrasound scans for training which included images with various categories of nodules such as cystic, solid, mixed-cystic solid and spongiform. A separate set of 372 ultrasound images extracted from patients who were not part of the training set was used for testing. All images were labeled by medical experts using a web-based open-source image segmentation software (www.dataturks.com) and reviewed by a board certified radiologist with more than 15 years of experience.

To increase robustness to user input, we augmented the dataset $(Size = 71110)$ by sampling multiple click points in the neighbourhood of the centroid of the ground-truth mask. Also to avoid possible size bias we balanced the number of small (less than 100 pixels), medium (100 to 300 pixels) and large (more than 300 pixels) sized nodules in the dataset. As a result, the model does not favour a particular nodule size and instead predicts the nodule boundaries based on the user click.

C. U-Net Architecture

The proposed deep neural network is based on the popular U-Net architecture introduced in [10]. We have modified the original architecture to adapt it to the specific task of thyroid nodule segmentation.

Our modified U-Net accepts 512×512 size of input image with two channels. The input is fed to the U-Net encoder segment (also called contracting path). The U-Net encoder consists of sequential and repeated application of set of two 2D convolutional layers with no padding and ReLU activation followed by a max pooling layer. Max pooling reduces the image dimensions by half in each layer. Each pair of 2D convolutional layer doubles the number of filters starting from 32 until 2048 in our case. In total, there are 6 such blocks in encoder followed by two 2D convolutional layers. The output of the encoder which is of size $8\times8\times2048$ is fed to U-Net decoder segment (also called expansion path of U-Net). The decoder segment consists of sequential and repeated application of 2D convolutional transpose (also called deconvolution) and its concatenation

Dataset	Train accuracy	Test accuracy	Test accuracy segregated (S, M, L)
Augmented	91.12%	84.00%	91.19% (S), 82.62% (M), 74.4% (L)
Only Large	95.43%	80.01%	- (S), - (M), 80.50% (L)
Small and Medium	91.66%	82.04%	82.87% (S), 82.89% (M), -% (L)

TABLE I

DICE SCORE RESULTS OF THYROID NODULE SEGMENTATION USING PROPOSED TWO-CHANNEL 2D-UNET APPROACH. S=SMALL, M=MEDIUM AND L=LARGE NODULE SIZE RESPECTIVELY.

User Interaction	Train accuracy	Test accuracy
Zero-Click (Baseline 2D-UNet)	98.33%	60.83%
Bounding Box approach with 0% margin error	97.02%	90.34%
Bounding Box approach with 10% margin error	97.50%	83.57%
One-Click (Proposed method)	91.12%	84.00%

TABLE II

COMPARISON OF DICE SCORE RESULTS FOR THE PROPOSED ONE-CLICK METHOD WITH ZERO-CLICK AND BOUNDING BOX APPROACHES. THE PROPOSED ONE-CLICK APPROACH YIELDED SIGNIFICANTLY MORE ACCURATE RESULTS THAN ZERO-CLICK APPROACH WHILE OFFERING SLIGHTLY HIGHER ACCURATE RESULTS WITH REDUCED MANUAL INPUT THAN THE BOUND BOX APPROACH WITH 10% MARGIN ERROR.

with the corresponding layer in contracting path followed by two 2D convolutional layers. The 2D convolutional transpose doubles the image size and halves the filter size. And hence such 6 repeated layers play the role of expansion of the input dimensions. The output is $512 \times 512 \times 1$ size consisting of the probability of each pixel being mask. The 3D-UNet is a modification of the 2D-UNet such that all the parameters are same as before except for the additional third dimension in every layer for handling 3D data. Due to increased resources requirements for 3D UNet model(computational expense and memory) , we resized the input sweep from $512 \times 512 \times 512$ to $128 \times 128 \times 128$.

III. RESULTS

Dice score [11] was used as metric to quantify the performance of our segmentation models based on 2D-UNet architecture. Dice score is defined as the ratio of the size of the intersecting regions between predicted and ground truth regions and total size of the two regions together. We evaluated the accuracy separately for small, medium and large sized nodules as summarized in Table I. The segmentation accuracy in each sub-group improved when the model was trained exclusively on images of that group. Merging all the datasets together and training our model gave an overall Dice score accuracy of 84.00%. on the test set. Some example of the segmentation obtained from the proposed one-click approach are shown in Figure 2. In each figure, the ground truth is represented by the green contour and the predicted segmentation is represented by the red contour. The user click-point is shown as red-dot inside the nodule. Figure 3 represents few of results from the 3D-UNet model tracking the thyroid nodule in a sweep.

In Table II, we compare the Dice score results of the proposed approach with the bounding box approach (Two-click) and with the fully-automatic approach (Zero-click). We compare our results with the segmentation obtained from the 1) perfect tight bounding box (0% error) obtained from ground-truth segmentation, 2) a loose bounding box allowing 10% error from the ground-truth, and 3) the image as is with no bounding box. As summarized in Table II our approach gave a much higher Dice score than the zero-click approach (60.83% vs 84.00%) and slightly higher than the loose bounding box (83.57% vs 84.00%)

IV. DISCUSSION

We developed a new technique to segment the boundary of the thyroid nodule using a single mouse-click and compared it to existing bounding box approaches. The predicted segmentation from the model was close to ground-truth as shown in Figure 2. The output error was of independent of echotexture of the nodule. The fully automatic (zero click) approach performed poorly on the test set while a perfectly tight ROI around the nodule gave the best segmentation accuracy. However, practically defining such a tight bounding box is difficult to achieve as there are inevitably human errors and inter-observer variation in manually detecting the bounding box. Detecting a precise bounding box is particularly challenging for nodules with unclear or blurry boundaries. Therefore, we also calculated the test accuracy for a loose bounding box with 10% randomized error. Our method yielded better accuracy than the segmentation with loose bounding box and effectively reduced the user interaction.

As future work, we plan to develop a fully automatic technique for nodule segmentation based on the proposed one-click approach. Such an extension could be implemented by training a network to automatically identify candidate seed points in the image. We also plan to improve segmentation accuracy of our approach in nodules by training multiple networks for sub-groups of nodules using an initial non-

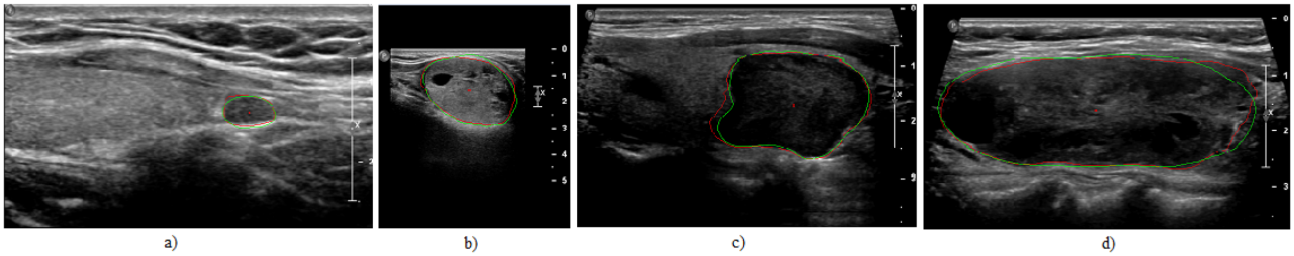

Fig. 2. Sample results from 2D-UNet. The diagram shows different nodule sizes in increasing order from a) to d). Green and red color contours represent the ground truth and prediction by the proposed 2D-UNet approach, respectively. The results demonstrate that agreement of predicted contour with ground-truth does not depend on the composition of the nodule.

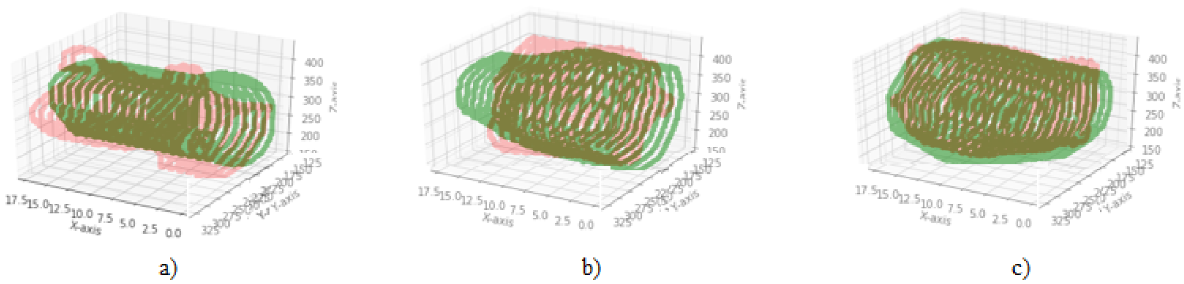

Fig. 3. Sample results from 3D-UNet showing the tracked thyroid nodule in given sweep vis-à-vis the ground truth. Green and red colors represent the ground truth and 3D-UNet prediction, respectively.

supervised clustering technique to identify optimal image clusters.

V. Conclusion

We showed that a single click from user can be used to segment a nodule from a thyroid ultrasound image. Further, we showed this one slice segmentation could be extended to the whole sweep using 3D-Unet, segmenting the particular nodule across the sweep. This segmentation can be used for quantifying the size of the nodule and for categorizing its risk of malignancy. We expect the proposed approach to save expert time and reduce the variability in manual segmentation of the nodule resulting in better thyroid cancer care.

VI. Compliance with Ethical Standards

The Institution's Ethical Review Board - Health Research Ethics Board of the University of Alberta, approved all experimental procedures involving human subjects.

VII. Acknowledgments

Rajshree Daulatabad was supported by the MITACS Accelerate program, MEDO.ai and her supervisor, Prof. Vassilios Tzerpos.

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] Joannie Lortet-Tieulent, Silvia Franceschi, Luigino Dal Maso, and Salvatore Vaccarella, "Thyroid cancer "epidemic" also occurs in low- and middle-income countries," *International journal of cancer*, vol. 144, no. 9, pp. 2082–2087, 2019.

[3] 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.

[4] Tomas Sakinis, Fausto Milletari, Holger Roth, Panagiotis Korfiatis, Petro Kostandy, Kenneth Philbrick, Zeynettin Akkus, Ziyue Xu, Daguang Xu, and Bradley J Erickson, "Interactive segmentation of medical images through fully convolutional neural networks," *arXiv preprint arXiv:1903.08205*, 2019.

[5] Jinlian Ma, Fa Wu, Tian'an Jiang, Jiang Zhu, and Dexing Kong, "Cascade convolutional neural networks for automatic detection of thyroid nodules in ultrasound images," *Medical physics*, vol. 44, no. 5, pp. 1678–1691, 2017.

[6] Ahmet Alkan, Seda Arslan Tuncer, and Mucahid Gunay, "Comparative mr image analysis for thyroid nodule detection and quantification," *Measurement*, vol. 47, pp. 861–868, 2014.

[7] E Chester Ridgway, "Medical treatment of benign thyroid nodules: have we defined a benefit?," *Annals of internal medicine*, vol. 128, no. 5, pp. 403–405, 1998.

[8] Franklin N Tessler, William D Middleton, Edward G Grant, Jenny K Hoang, Lincoln L Berland, Sharlene A Teefey, John J Cronan, Michael D Beland, Terry S Desser, Mary C Frates, et al., "Acr thyroid imaging, reporting and data system (ti-rads): white paper of the acr ti-rads committee," *Journal of the American college of radiology*, vol. 14, no. 5, pp. 587–595, 2017.

[9] Kunio Doi, "Computer-aided diagnosis in medical imaging: historical review, current status and future potential," *Computerized medical imaging and graphics*, vol. 31, no. 4-5, pp. 198–211, 2007.

[10] 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.

[11] Carole H Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M Jorge Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pp. 240–248. Springer, 2017.