AUTOMATED SEGMENTATION OF THE THYROID GLAND ON CT USING MULTI-ATLAS LABEL FUSION AND RANDOM FOREST

Jiamin Liu, Divya Narayanan, Kevin Chang, Lauren Kim, Evrim Turkbey, Le Lu, Jianhua Yao, Ronald M. Summers

Imaging Biomarkers and Computer-aided Diagnosis Laboratory
Radiology and Imaging Sciences, National Institutes of Health Clinical Center
Building 10 Room 1C224 MSC 1182, Bethesda, MD 20892-1182

ABSTRACT

The thyroid gland is an important endocrine organ. For a variety of clinical applications, a system for automated segmentation of the thyroid is desirable. Thyroid segmentation is challenging due to the inhomogeneous nature of the thyroid and the surrounding structures which have similar intensities. In this paper, we propose a fully automated method for thyroid detection and segmentation on CT scans. The thyroid gland is initially estimated by a multi-atlas segmentation with joint label fusion algorithm. The segmentation is then corrected by supervised statistical learning-based voxel labeling with a random forest algorithm. Multi-atlas label fusion transfers expert-labeled thyroids from atlases to a target image using deformable registration. Errors produced by label transfer are reduced by label fusion that combines the results produced by all atlases into a consensus solution. Then, random forest employs an ensemble of decision trees that are trained on labeled thyroids to recognize various features. The trained forest classifier is then applied to the estimated thyroid by voxel scanning to assign the class-conditional probability. Voxels from the expert-labeled thyroids in CT volumes are treated as positive classes and background non-thyroid voxels as negatives. We applied our method to 73 patients using 5 as atlases. The system achieved an overall 0.70 Dice Similarity Coefficient (DSC) if using the multi-atlas label fusion only and was improved to 0.75 DSC after the random forest correction.

Index Terms – thyroid gland segmentation, multi-atlas label fusion, random forest

1. INTRODUCTION

The thyroid is a small gland, shaped something like the letters H or U. It is located in the lower front part of the neck and surrounds the trachea. The thyroid is one of the important endocrine glands in the body. It produces thyroid hormones which have various impacts in the body, such as helping to regulate metabolism and energy usage. Thyroid volume estimation plays an important role in both diagnosis and treatment [1]. In radiation treatment planning, for patients with head and neck cancer, radiation therapy requires a precise delineation of the thyroid gland to be spared on the pre-treatment planning CT images to avoid thyroid dysfunction.

In the current clinical workflow, the thyroid gland is normally manually delineated slice by slice by radiologists or radiation oncologists. This is time consuming and error prone, due to various interpretations of slices of CT scans. Therefore, a system for automated segmentation of the thyroid is desirable. However, automated segmentation of the thyroid is challenging because the thyroid is inhomogeneous and surrounded by structures, such as veins and lymph nodes, which have similar CT intensities. Moreover, the shape and size of thyroid glands can vary considerably. An example CT section through the thyroid is shown in Figure 1.

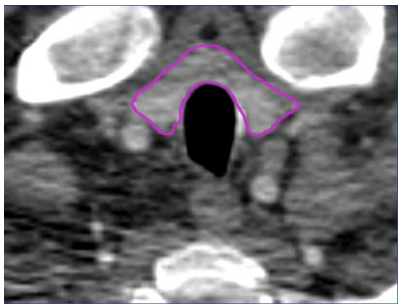

Figure 1. CT image of the thyroid (pink).

In this work, we propose an automated method for thyroid segmentation. The thyroid gland is initially segmented by a multi-atlas label fusion algorithm [2]. Multi-atlas segmentation uses more than one atlas to compensate for potential bias associated with a single atlas and applies label fusion to produce a consensus solution. Several studies have shown that multi-atlas segmentation outperforms single-atlas methods [3,4]. Then, the segmentation is corrected by supervised learning-based voxel labeling with a random forest algorithm [5]. The major contribution of this work is that supervised learning (random forest) is combined with

registration-based segmentation (multi-atlas label fusion) for more accurate thyroid segmentation.

2. METHODS

Our method consists of three major steps: (1) region of interest, (2) initial thyroid segmentation by multi-atlas label fusion, and (3) segmentation correction by random forest.

2.1 Region of interest

The region of interest (ROI) is determined automatically by using the segmentation of the trachea and spine to create a bounding box that includes the thyroid. The trachea is segmented using thresholding followed by morphological operations. The spine is segmented using a watershed algorithm followed by a directed graph search [6]. All segmentation methods do not require any user interaction. Figure 2 shows how the ROI is automatically determined. The upper bound of the ROI is indicated by the superior portion of the trachea and the height of the ROI is set as 6 cm for all patients.

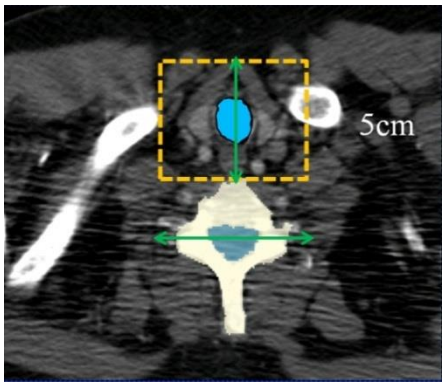

Figure 2: ROI (dashed yellow box) of thyroid gland and segmentation of spinal column (ivory), spinal canal (dark blue), and trachea (light blue).

2.2 Initial thyroid segmentation

The rationale of atlas-based segmentation is the segmentation of a target image correlates with the atlas image. A target image can be segmented by deforming the atlas images. The segmentations of the atlas are transferred to the target image by deformable registration. As an extension, multi-atlas segmentation makes use of more than one atlas to compensate for potential error associated with using a single atlas. The final segmentation is generated by label fusion which combines the results produced by all atlases to reduce the errors.

In this work, multi-atlas with joint label fusion [2] was used for initial thyroid segmentation. An atlas A is defined as a pair of images (I, L) where I is the gray scale image and L is the manually labeled thyroid gland. Let $U = (I_T, L_T)$ be a target image to be segmented (L_T denotes the target thyroid) and $A_1 = (I_1, L_1)$, ... , $A_n = (I_n, L_n)$ be n atlases. In order to be able to propagate the thyroid L_i from an atlas A_i to a target image U, a registration is needed between them. In this work, each atlas A_i is registered to the target image U by using diffeomorphic registration developed in Advanced Normalization Tools (ANTS) [7]. The registration resulted in n transformations u_i , describing transformations from A_i to U. These transformations are used to warp the atlas thyroid L_i to the target U. For each voxel in the target U, each atlas provides an opinion about the label. Finally, the probability of thyroid at voxel x in the target U is determined by

$$\hat{L}_T(x) = \sum_{i=1}^{n} \omega_i(x) u_i[L_i(x)] \quad (1)$$

$u_i[L_i(x)]$ is the warped i_{th} atlas thyroid and $\omega_i(x)$ is an optimal weight assigned to the i_{th} atlas with $\sum_{i=1}^{n} \omega_i(x) = 1$. $\omega_i(x)$ could be estimated by similarity-weighted voting strategies [8], where the atlas image that is more similar to the target image has higher weights. One limitation of this approach is that the voting weight for each atlas is estimated independently from other atlases, ignoring potential correlations among the atlases. To address this problem, we use the joint label fusion algorithm [2] estimates voting weights by simultaneously considering pairwise atlas correlations. As shown in [2], joint label fusion performed better than label fusion with independent weight estimation.

$$\omega_i(x) = \frac{M_x^{-1} 1_n}{1_n^t M_x^{-1} 1_n} \quad (2)$$

where M_x is a matrix measuring the correlation between i_{th} and j_{th} atlases and 1_n = [1; 1; …; 1] is a vector of size n. More detail about how to capture the probability that different atlases produce the same label error at location x by a dependency matrix M_x can be found in [2].

Five atlases (n=5) were used in our experiment. Previous published work [8] also shows that n=5 is an appropriate choice for the multi-atlas segmentation of heart and aorta in chest CT scans; more atlases will not significantly improve the segmentation performance. Figure 3(b) shows the probability of thyroid after multi-atlas label fusion.

2.3 Segmentation correction

The initial thyroid segmentation is corrected by supervised learning based voxel labeling with a random forest algorithm. A random forest classifier consists of a

number of trees that generates many classifiers and aggregates their results in order to make decisions. The advantages of the random forest include: (i) both training and classification are computationally efficient, (ii) classification output is probabilistic, (iii) can handle a large variety of features and features are not required to be normalized, and (iv) inherent feature selection.

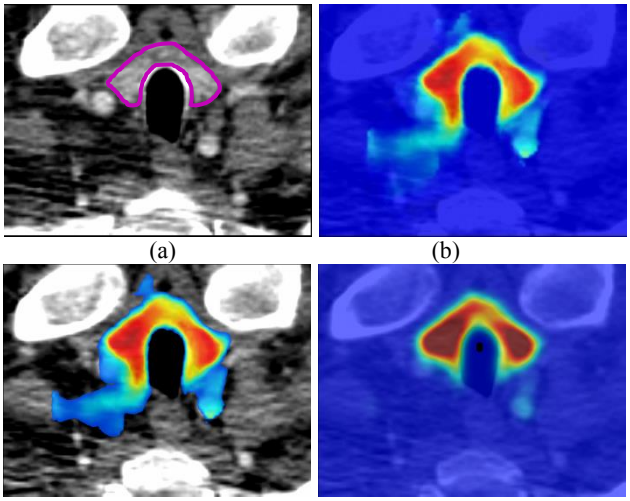

Figure 3. (a) Example of thyroid CT scan. (b) Thyroid probability map from the multi-atlas label fusion. (c) Voxels with thyroid probability ≥ 0.15. (d) Thyroid probability map after the random forest correction.

At each node, the classification tree randomly takes a subset of training samples and predicts the probability of the sample being in the thyroid class. Based on the features, the random forest continually splits the training samples at every node ("tree growth"), and assigns the partitions to the left/right nodes. This splitting is done with a random dimension. Tree growth continues up to a certain tree depth. In the testing phase, target voxels are pushed through each tree according to the learned split functions.

Here, random forest training starts with the subsampling of manually labeled data from the 5 atlases. Positive voxels were sampled on the labeled thyroids and negative voxels were sampled adaptively. First, all negative voxels within 2 voxels distance from the boundary of the labeled thyroid by distance transform were not used for training due to their low confidence of being negative or positive. Second, negatives beyond 2 voxels but within 7 voxels away were randomly sampled.

The two image features were CT attenuation and histogram of oriented gradient within a patch (7×7×7 voxels). Features were selected by maximizing the information gain function. In this work, the number of trees used in the random forests was 500. Each node of a tree assessed possible sample splits on 3 randomly selected features. A tree branch was terminated if a split node results in 2 or fewer samples in the branch.

Once trained, the forest was then applied to the initially estimated thyroid (all voxels with thyroid probability of greater than 0.15 (Figure 3c)). Each voxel was pushed through the tree starting at the root, and the path taken depended on the voxel's feature response and the feature threshold at the node learned at training. All samples ended up at a tree leaf node where its probability function representing the likelihood of the voxel being thyroid was given. The testing point reached to different leaf nodes in each tree of the forest, and the posterior probabilities collected at each leaf were averaged to determine the final distribution, to be then used to classify the voxel (Figure 3d).

3. RESULTS

Our patient population consisted of 73 patients with intravenous contrast-enhanced thoracic CT scans that included the thyroid. The reference standard consisted of manual tracings of the 73 thyroids by a medical student. The CT slice thickness was 1 mm, and the voxel spacings within an axial slice were in the range 0.7 - 0.9 mm. 5 scans served as the atlases for multi-atlas label fusion. For each atlas, the thyroid was manually segmented by using 3D Slicer. These 5 scans were also used for random forest training. The remaining 68 patients were used for thyroid segmentation and evaluation.

Figure 4 shows one example of thyroid segmentation before and after the random forest correction. More segmentation examples are shown in Figure 5.

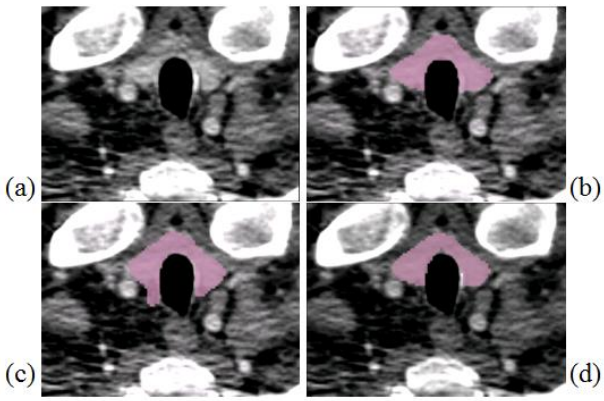

Figure 4. (a) Example of thyroid CT scan. (b) Manually-labeled ground truth. (c) Segmentation from the multi-atlas label fusion. (d) Segmentation after the random forest correction.

The segmentation performance was evaluated using the Dice Similarity Coefficient (DSC), a coefficient used to represent the ratio of volumetric overlap between two segmentations based on their average volume. Figure 6 shows the comparison of the DSC results across the various stages of our method. Before the label fusion, the DSC (average of the 5 DSCs) was 0.49 ± 0.12. This value was increased to 0.70 ± 0.12 after label fusion, and

increased to 0.75 ± 0.08 with the random forest for segmentation correction.

The surface error between the segmented thyroid and the ground truth was 1.14 ±1.25 mm after the multi-atlas label fusion and was reduced to 0.60 ± 0.37 mm after the random forest for correction. Figure 7 shows one example of thyroid surface comparison between the automated method and the ground truth.

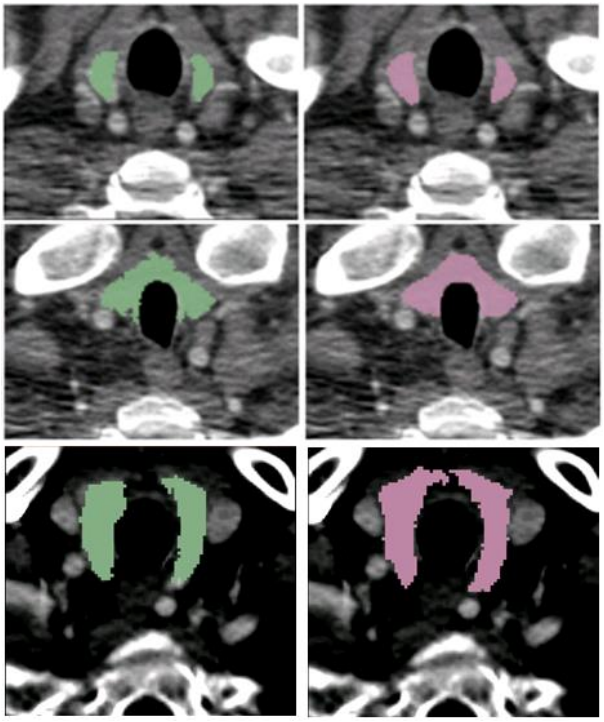

Figure 5: Example of automated thyroid segmentation (1st column) and ground truth segmentations (2nd column).

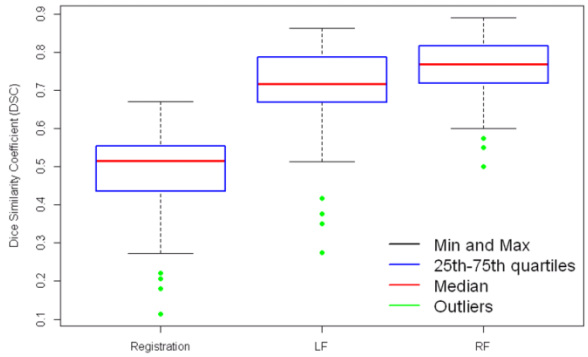

Figure 6: Dice Similarity Coefficient comparison after the registration, the label fusion (LF) and the random forest (RF) for correction.

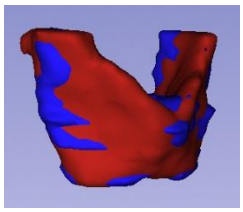

Figure 7. Thyroid surface comparison between the automated method (blue) and the ground truth (red).

4. CONCLUSION

We presented a method for automatic segmentation of the thyroid gland on CT scans. This is a challenging problem because the thyroid is inhomogeneous and surrounded by structures with similar CT attenuations. The thyroid is initially segmented by multi-atlas segmentation method and further corrected by the supervised learning method random forest. This thyroid gland segmentation could be used for organ specific radiation dose estimation and radiotherapy treatment planning.

REFERENCES

1. Shabana, W., Peeters, E., & De Maeseneer, M.,"Measuring thyroid gland volume: should we change the correction factor?" American Journal of Roentgenology, 2006, 186(1), pp. 234-236.
2. Wang, H., et al., "Multi-Atlas Segmentation with Joint Label Fusion," IEEE Trans Pattern Anal Mach Intell, 2012.
3. Rohlfing T, Brandt R, Menzel R, Maurer CR Jr., "Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains, " Neuroimage. 2004, 21(4), pp. 1428-42.
4. Heckemann, Joseph V. Hajnal , Paul Aljabar, Daniel Rueckert, Alexander Hammers, "Automatic anatomical brain MRI segmentation combining label propagation and decision fusion, " Neuroimage, 2006, 33(1), pp. 115–126.
5. Breiman, L., "Random forests," Machine Learning, 2001, **45**(1), pp. 5-32.
6. Yao J, OConnor SD, Summers RM, "Automated spinal column extraction and partitioning," Biomedical Imaging: Nano to Macro, IEEE International Symposium on, 2006, pp. 390 – 393.
7. Avants BB, Epstein CL, Grossman M, Gee JC, "Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain," Med Image Anal. 2008, 12(1), pp. 26-41.
8. van Rikxoort, E.M., et al., "Adaptive local multi-atlas segmentation: application to the heart and the caudate nucleus," Med Image Anal, 2010. **14**(1): pp. 39-49.