IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 4 December 2024, accepted 25 December 2024, date of publication 27 December 2024, date of current version 6 January 2025.

Digital Object Identifier 10.1109/ACCESS.2024.3523766

RESEARCH ARTICLE

RED-Net: A Neural Network for 3D Thyroid Segmentation in Chest CT Using Residual and Dilated Convolutions for Measuring Thyroid Volume

MIN-JI KIM[1], JIN-A KIM[1], NAAE KIM[2], YUL HWANGBO[1], HYUN JEONG JEON[3,4], DONG-HWA LEE[3,4], AND JI EUN OH[1]

[1]Heathcare AI Team, National Cancer Center, Goyang-si 10406, South Korea
[2]Research and Development Business Foundation, National Cancer Center, Goyang-si 10408, South Korea
[3]Department of Internal Medicine, Chungbuk National University Hospital, Cheongju-si 28644, South Korea
[4]Department of Internal Medicine, Chungbuk National University College of Medicine, Cheongju-si 28644, South Korea

Corresponding author: Ji Eun Oh (jieun12@ncc.re.kr)

This work was supported in part by the National IT Industry Promotion Agency (NIPA) funded by Korean Government [Ministry of Science and ICT (MSIT)] through the Development of AI Precision Medical Solution (Doctor Answer 2.0) under Grant S0252-21-1001, and in part by the National Cancer Center under Grant 2310840.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board (IRB)under Application Nos. NCC2021-0151 and 2022-06-001-001.

ABSTRACT Unlike the lungs or the heart, the thyroid gland is not a primary target in chest computed tomography (CT) scans and is relatively small; hence, it is difficult for radiologists to always clinically delineate it in chest CT to incidentally detect a goiter. We designed a residual and dilated convolution neural network (RED-Net), which automatically measures thyroid volume by segmenting the thyroid gland in contrast-enhanced chest CT scans. Its fundamental structure comprises a residual downsampling and upsampling pathway, complemented by a parallel dilated convolution module. This combination allows the model to extract features at multiple scales and capture contextual information to effectively segment even tiny thyroid glands in the complex anatomical structures observed in chest CT scans. Additionally, we constructed training and validation sets comprising CT scans of 1,150 adults (aged $\geq$19 years) who underwent chest CT scans at the National Cancer Center and included data of those without a history of thyroid nodules, C73 diagnosis, or thyroid surgery before scanning procedure. We evaluated the performance of our method on a test dataset (600 patients) comprising chest CT scans of individuals collected at Chungbuk National University Hospital using the same criteria. The results showed that it achieved state-of-the-art performance with a Dice similarity coefficient of 0.8901.

INDEX TERMS Chest CT scans, dilated convolution, goiter, RED-Net, residual blocks, thyroid segmentation, thyroid volume.

I. INTRODUCTION

Chest computed tomography (CT) scans are widely used to detect pulmonary and cardiac pathologies. The thyroid gland is typically visible in these scans, allowing incidental detection of thyroid diseases. However, because the primary purpose of chest CT is to identify structural abnormalities in the lungs, heart, and major blood vessels, thyroid abnormalities may be overlooked by the radiologist while interpreting the scans.

Thyroid diseases associated with structural abnormalities include thyroid cancer, thyroid nodule, and goiter. Goiters are

The associate editor coordinating the review of this manuscript and approving it for publication was Hengyong Yu.

often associated with conditions such as Graves' disease and Hashimoto's thyroiditis [1]. A goiter can be diagnosed by accurately measuring an increased thyroid volume in chest CT, thereby enabling the diagnosis of autoimmune thyroid diseases [1], [2]. However, failure to diagnose goiter in chest CT can potentially have adverse medical and socioeconomic consequences.

However, as comprehensive examinations of thyroid lesions in chest CT scans can considerably increase the workload of radiologists, an automated tool that can measure thyroid volume can facilitate goiter diagnosis and could offer exceptional convenience, thereby ensuring that goiter is not overlooked.

Previous studies have employed machine-learning methods such as progressive learning vector quantization [3] and multi-atlas label fusion [4], as well as the U-Net architecture, for thyroid segmentation [5], [6]. However, most of them used neck CT scans that focused on the thyroid region, and it is unclear whether these models would offer the same performance for chest CT scans.

Therefore, this study proposes a residual and dilated convolution neural network (RED-Net), designed to segment the thyroid gland and automatically measure the thyroid volume from contrast-enhanced chest CT scans, which are routinely performed for non-thyroid-related purposes. RED-Net combines the strengths of residual blocks with the ability of parallel dilated convolution to capture multiscale information. This combination allows the model to effectively learn representations of complex anatomical structures in 3D CT scans and delineate the thyroid gland. We evaluated RED-Net on a test set comprising 600 samples, wherein it achieved state-of-the-art performance with a DSC of 0.8901.

II. RELATED WORKS

Chang et al. [3] used a progressive learning vector quantization neural network to segment thyroid in CT images. They demonstrated that it can effectively segment the thyroid glands, with an average sensitivity of 88.43%. Narayanan et al. [4] used multi-atlas label fusion (MALF) and random forest (RF) algorithm to automatically segment the thyroid gland in CT images. They found that MALF with RF offered better segmentation performance, with a DSC of 0.76 ± 0.11, which was significantly better than the individual results of the MALF and RF methods.

The U-Net architecture introduced by Ronneberger et al. [7] has become the cornerstone of medical image segmentation. The original 2D U-Net employed an encoder-decoder structure with skip connections, enabling effective localization and context integration. Çiçek et al. [8] extended this architecture to 3D to develop 3D U-Net, which offered significantly better segmentation performance for volumetric data by leveraging the spatial context of 3D images.

The DeepLab series, a suite of semantic segmentation models, incorporates advanced segmentation techniques such as atrous convolutions and atrous spatial pyramid pooling for multiscale context capture from 2D images. The first version [9] contained a fully connected conditional random field (CRF) to enhance boundary localization, whereas later versions [10], [11] improved dense-feature extraction and eliminated the need for DenseCRF post-processing. Subsequently, DeepLabv3+ [12] introduced a decoder module and offered better segmentation accuracy at object boundaries and enhanced efficiency by using depthwise separable convolutions. SegResNet [13], developed by Myronenko et al., features a residual encoder–decoder architecture with a variational autoencoder branch for enhanced feature extraction.

Oktay et al. [14] introduced the attention U-Net, a modification of the U-Net architecture that incorporates attention mechanisms to enhance segmentation accuracy by focusing on relevant regions in the image. Additionally, transformer-based models, such as UNETR [15], have been introduced to overcome the limitations of convolutional neural networks (CNNs), using a vision-transformer encoder to effectively capture long-range dependencies and multiscale contexts. Moreover, the Swin transformer, with its shifted-window self-attention mechanism, enables efficient processing of large images. Swin-UNETR combines this transformer with a U-shaped network to effectively capture multiscale features and long-range dependencies for medical image segmentation.

Several studies have explored advanced segmentation techniques in medical imaging, demonstrating the growing impact of deep-learning. He et al. [5] used a deep CNN to segment the thyroid gland in non-contrast-enhanced head and neck CTs. Wen et al. [6] proposed a model that combines the HRNet architecture with a cSE attention mechanism based on U-Net to segment the thyroid gland in localized CT images. Specifically, the HRNet extracts multiscale features, whereas the cSE block enhances important channel features, thereby aiding in delineating the thyroid gland for radiotherapy.

D'Aviero et al. [16] investigated the use of commercial deep learning-based auto-segmentation software for delineating organs at risk in head and neck radiotherapy. Their results showed that the software provided acceptable segmentations for most structures, significantly reducing inter-observer variability and time consumption, although manual adjustments were necessary for less accurate contours. Similarly, Xie et al. [17] proposed a comprehensive pipeline for lung nodule detection and robotic biopsy path planning from chest CT images, emphasizing accurate lung parenchyma segmentation and 3D reconstruction to improve clinical workflows.

Other applications of deep learning-driven segmentation include the use of MultiResUNet by Arsenescu et al. [18] for 3D ultrasound reconstruction of carotid arteries and thyroid glands, achieving high Dice similarity coefficients in detecting atherosclerosis. Lu et al. [19] proposed a two-stage method for lumbar spine segmentation in CT images, combining U-Net for localization and a novel 3D XUnet for precise segmentation, proving effective for spinal anomaly detection. For orbital segmentation in CT images,

Li et al. [20] employed a semi-supervised framework using a paired copy-paste strategy, achieving remarkable accuracy even with limited labeled data. Alkhadrawi et al. [21] developed a deep-learning model to segment orbital muscle and fat for diagnosing thyroid eye disease, highlighting its potential to improve clinical decision-making through rapid volumetric assessment. These advancements underscore the transformative role of deep-learning in enhancing segmentation accuracy, efficiency, and clinical utility across diverse medical imaging modalities.

III. METHODS

RED-Net is a specialized deep-learning model for thyroid-gland segmentation from chest CT scans. It employs residual blocks [13] and integrates dilated convolutions [11] to enhance its ability to capture multiscale contextual information, which is crucial for accurately segmenting complex anatomical structures in chest CT scans.

A. DATASET ACQUISITION

We collected chest CT scans of adults (aged ≥19 years) who underwent contrast-enhanced chest CT scanning in the National Cancer Center (NCC) and Chungbuk National University Hospital (CBNUH). The data of those with a prior diagnosis of C73 (malignant neoplasm of the thyroid gland) or thyroid nodules (classified as D34, D440, or E041) before the CT scans were excluded, as these conditions can significantly alter the thyroid appearance or characteristics in CT scans. Additionally, patients who underwent thyroid surgery before the CT scan were excluded to ensure that the dataset contained only scans of individuals with intact thyroid anatomy.

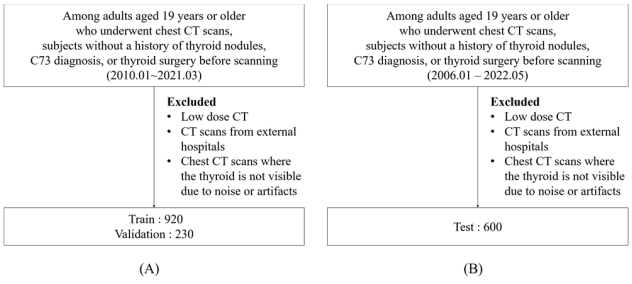

FIGURE 1. Flow diagrams for the train and test datasets. (A) Train dataset from NCC (B) Test dataset from CBNUH.

The dataset was further refined by eliminating scans that did not meet the quality and coverage requirements. Furthermore, low-dose CT scans were excluded because they typically have a lower resolution and could compromise the model accuracy. Additionally, scans from external hospitals were excluded to ensure consistency in imaging protocols and quality. Chest CT scans in which the thyroid gland was not visible due to noise or artifacts were also excluded as they lacked the necessary anatomical information for analysis.

Based on these criteria, CT scans of 1,150 patients conducted between January 2010 and March 2021 at the NCC and those of 600 patients conducted between January 2006 and May 2022 at CBNUH were obtained. Expert doctors specializing in thyroid diseases manually created the ground truth labels. They applied their extensive expertise, ensured consistency throughout the labeling process, and reviewed each other's work.

The NCC data were divided into an 80:20 ratio for training and validation (920 and 230 data points, respectively), whereas the CBNUH data were used for testing.

The study was approved by an IRB (IRB No. NCC2021-0151 and 2022-06-001-001) and exempt from obtaining patient consent for the following reasons. As a retrospective study, we utilized de-identified medical records from the Clinical Research Search Portal, involving adults who had undergone chest CT scans before the study period. It was not feasible to obtain consent from all eligible patients, as some were no longer attending the outpatient clinic, and others had passed away. Additionally, since the study relied on pre-existing medical records, it was impossible to ascertain or infer any refusal of participation. There was no harm to the patients, and the risk posed to them was minimal. Therefore, the study's results do not impact the patients involved.

B. DATA PREPROCESSING

To preprocess the CT images and labels, we employed a series of transformations to standardize the data and augment those included in the training set. Initially, we loaded the image and label data and confirmed that they were in the channel-first format. Subsequently, their intensity values were normalized by scaling them to a standardized range of 0–1, thereby ensuring consistent intensity levels across all images.

After normalization, we isolated the region of interest by cropping the foreground based on intensity values. This was achieved by automatically identifying all voxels with non-zero intensity values, which represent the anatomical structures within the image. A bounding box encompassing these voxels was computed to crop out the background, retaining the thyroid gland and surrounding tissues.

The images were oriented using the right-anterior-superior (RAS) coordinate system to maintain their spatial consistency. Next, the spacing was adjusted to standardize the voxel dimensions to (1.0, 1.0, 1.0) mm by using bilinear and nearest-neighbor interpolations for the images and labels, respectively. To augment the dataset and address class imbalance, we performed random cropping based on positive and negative label distributions. Specifically, the images were randomly flipped along all three axes (x, y, z) with a probability of 0.10 to introduce variability. Additionally, random 90° rotations were performed up to three times with the same probability.

A similar transformation pipeline was used for the validation and test data. We loaded the images and labels, ensured channel-first orientation, and scaled the image intensities between zero and one. The images were oriented using the RAS coordinate system, and the spacing was adjusted to (1.0, 1.0, 1.0) mm, ensuring that both the training and validation

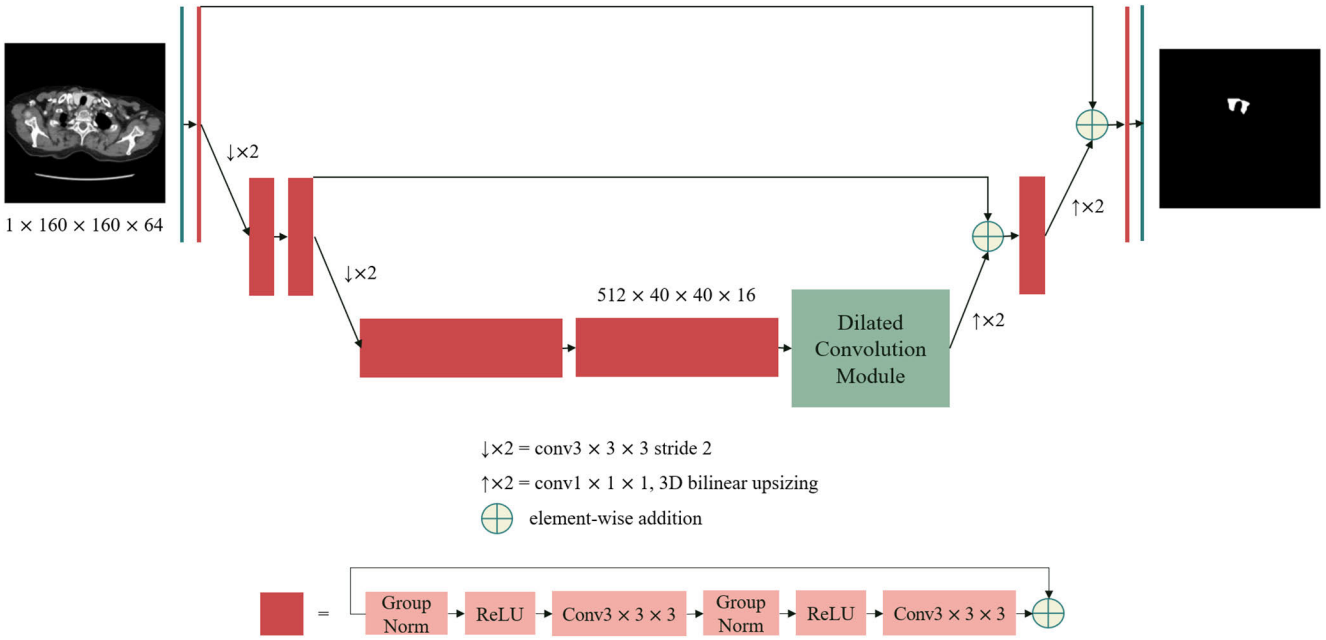

FIGURE 2. Architecture of the proposed RED-Net for thyroid-gland segmentation in chest CT scans. First, a 3D convolution layer is employed to extract the initial features, followed by downsampling path composed of residual blocks for group normalization and the rectified linear unit (ReLU) activation function, where the spatial resolution is halved using stride-2 convolutions. At the network bottleneck, the dilated convolution module (DCM) integrates multiscale information through parallel convolutions with different dilation rates. The upsampling path mirrors the downsampling one, and bilinear upsampling is employed to restore the original resolution.

datasets were subjected to identical preprocessing, whereas data-augmentation techniques such as cropping and random flipping were not employed.

C. DOWNSAMPLING PATH

The model begins with a convolution layer that processes the input image using a $3 \times 3 \times 3$ 3D convolution. This layer converts the input channels into 8, establishing a foundation for feature extraction in the subsequent layers. Each downsampling block consists of several ResNet [28] blocks, each comprising two convolutions with group normalization [29] and ReLU, followed by an additive identity skip connection. Subsequently, a stride-2 convolution is employed to decrease the spatial resolution and increase the feature size by 2. The output of the second downsampling is only four times smaller than that of the input image. We decided against further downsizing to preserve more spatial context and employed a parallel dilated convolution at the network bottleneck to capture multiscale information.

D. DILATED CONVOLUTION MODULE (DCM)

The DCM comprises parallel convolutions with different dilation rates (1, 6, 12, and 18) to gather contextual information from multiple scales and a global pooling layer to capture the global context. The convolution and pooling processes are followed by batch normalization and ReLU, respectively. The outputs of these branches are concatenated along the channel dimension and passed through final $1 \times 1 \times 1$ convolutional block to create the DCM feature map. This map integrates information from all scales, thereby enhancing the ability of the model to distinguish fine details in an image. We also used spatial dropout rate of 0.5 after the initial encoder convolution.

E. UPSAMPLING PATH

The upsampling path mirrors the downsampling path but operates in reverse to reconstruct the image resolution. It includes multiple layers that progressively restore the spatial dimensions of the feature maps while reducing their depth. Each upsampling block comprises an upsampling operation, followed by a residual block. The upsampling operation reduces the number of features by a factor of two (using $1 \times 1 \times 1$ convolutions) and doubles the spatial dimension (using 3D bilinear upsampling). Feature maps from the corresponding downsampling layers (stored during encoding) are added to the upsampled maps to recover the spatial details lost during downsampling. The final layer in the upsampling path has the same spatial dimensions as the original image. After upsampling, a final convolution layer is employed to convert the feature maps into the desired number of output channels (two in our case).

F. LOSS FUNCTION

We used the Dice loss (DL) function to accurately segment the thyroid gland in 3D chest CT images. This function, proposed by Milletari et al. [30], addresses class imbalance by

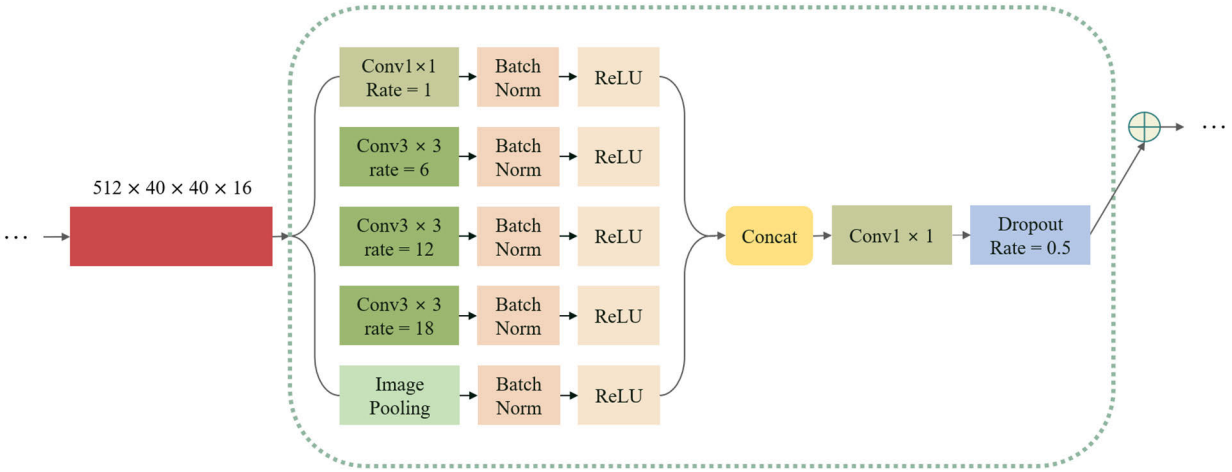

FIGURE 3. Architecture of the dilated convolution module (DCM) that uses parallel convolutions with different dilation rates (1, 6, 12, 18) to capture multiscale features and implement global image pooling. Batch normalization and ReLU are applied after each operation. The outputs are concatenated and passed through a 1 × 1 convolution, and a dropout rate of 0.5 is employed in the final feature map.

maximizing the overlap between the segmentation prediction and the ground truth, measured using the DSC. This approach ensures that the network can accurately detect and segment the thyroid gland without the need for sample reweighting strategies, which are often used in traditional methods to address foreground–background imbalances [30]. The DL function is computed as follows:

$$\mathrm{DL} = 1 - \frac{2\left|X \bigcap Y\right| + \varepsilon}{|X| + |Y| + \varepsilon} \tag{1}$$

where X represents the ground-truth label matrix of the thyroid gland, Y denotes the label matrix predicted by the model, and ε is a small constant added to avoid division by zero.

G. OPTIMIZATION

To train our RED-Net model, we employed the Adam optimizer [25] with an initial learning rate of $a_0 = 1 \times 10^{-4}$, which was progressively decreased according to the following schedule.

$$\alpha_e = \alpha_0(1 - \frac{e}{N_e}) \tag{2}$$

where e denotes the current epoch and N_e is the number of training epochs, which was set as 150 in our experiments. Additionally, a batch size of two was used for training, and the input images were randomly fed into the network to ensure learning variability and robustness.

H. EVALUATION INDICES

1) DSC

The DSC [26] metric is used to quantify the similarity between two sets. It is commonly employed in image-segmentation tasks to determine the overlap between the predicted and ground-truth segmentations. It is expressed as

$$\mathrm{DSC} = \frac{2\left|X \bigcap Y\right|}{|X| + |Y|} \tag{3}$$

where X and Y denote the sets of predicted and ground-truth segmentation pixels (or voxels), respectively.

The DSC evaluates the average overlap between the predicted and actual segmentations, where a value of 1 indicates perfect agreement and 0 indicates no overlap, and is sensitive to the degree of overlap between the two sets.

2) JSC

The JSC [27], or intersection over union (IoU), is widely used in segmentation tasks to evaluate the overlap between two sets, normalized by their total coverage. It is defined as:

$$\mathrm{JSC} = \frac{\left|X \bigcap Y\right|}{\left|X \cup Y\right|} \tag{4}$$

where X and Y are the predicted and ground-truth segmentation sets, respectively. The JSC quantifies the overall accuracy of the predicted segmentation by calculating the ratio of overlapping regions between the prediction and the ground truth. It ranges from zero (no overlap) to one (perfect overlap), with higher values indicating better segmentation performance.

3) HAUSDORFF DISTANCE (HD) (95TH PERCENTILE)

The HD [28] is the maximum distance between the boundaries of the two sets. A lower HD indicates a better boundary alignment between the predicted and ground-truth segmentations. The HD between two sets, X and Y, is defined as:

$$\mathrm{H}(X, Y) = \max\left(\sup_{x \in X} \inf_{y \in Y} \mathrm{d}(x, y), \sup_{y \in Y} \inf_{x \in X} \mathrm{d}(x, y)\right) \tag{5}$$

where d is the Euclidean distance between points $x \in X$ and $y \in Y$, sup is the supremum (maximum value), and inf is the infimum (minimum value). The 95th percentile HD (HD95) is calculated by identifying the 95th percentile distance from each point in X to the nearest point in Y to mitigate the influence of outliers.

4) SENSITIVITY

Sensitivity (SE) [29], or true-positive (TP) rate, is an essential metric for evaluating the capability of a model to correctly identify positive cases. In this study, TP indicates the correct thyroid-gland-region identifications, whereas false negative (FN) indicates the number of voxels in which it is incorrectly classified as background. SE is calculated as follows:

$$\mathrm{SE} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{6}$$

This equation computes the proportion of actual positives correctly identified by the model. A high SE indicates the effectiveness of the model in detecting TP cases, which is especially crucial for medical diagnoses, wherein missing a positive case can have serious consequences. Therefore, SE is a critical measure of a model's ability to accurately classify the intended region.

5) POSITIVE PREDICTIVE VALUE (PPV)

PPV [30], also known as precision, measures the proportion of correctly predicted positive regions among all those predicted as positive. Thus, it indicates the number of correctly predicted positive regions, reflecting the model's ability to avoid false positives (FPs). It is computed as

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{7}$$

where TP and FP are the numbers of correct and incorrect thyroid-gland region predictions, respectively. Thus, precision quantifies the accuracy of the positive predictions, and a high precision indicates a low FP rate, indicating that the model predictions are likely to be correct.

IV. RESULTS

A. QUANTITATIVE ANALYSIS OF RESULTS

Table 1 lists the results for the NCC validation set on the DSC, HD95, JSC, SE, and PPV metrics. Evidently, RED-Net outperformed the other segmentation models across most evaluation metrics, demonstrating its superior capability in accurately segmenting the thyroid-gland region in chest CT scans.

Specifically, it achieved the highest DSC (0.9094) and JSC (0.8361), indicating excellent overlap and segmentation accuracy. Furthermore, it exhibited superb accuracy for boundary segmentation, with an HD95 of 37.931. This is notably lower than those of attention U-Net (223.441), Swin-UNETRv2 (163.850), and SegResNet (67.296), indicating that RED-Net can more accurately capture the segmentation boundaries of the thyroid-gland region. However, DeepLabv3 achieved the lowest HD95 (10.122), slightly outperforming RED-Net.

Although RED-Net and DeepLabv3 both use residual connections and multi-scale context modules employing dilated (atrous) convolutions—subtle architectural differences may contribute to the observed performance disparity in the HD95 metric. A key difference lies in the normalization layers used: DeepLabv3 employs batch normalization, while RED-Net uses group normalization. Additionally, the ASPP module in DeepLabv3 concatenates multi-scale features followed by extra convolutional layers. Furthermore, differences in the ResNet backbone configurations, compared to RED-Net's encoder, may influence feature representations and the models' abilities to capture boundary information.

Nevertheless, RED-Net exhibited a superior SE (0.9192) and maintained a high PPV (0.9036), indicating that Table 1 lists the results for the NCC validation set on the DSC, HD95, JSC, SE, and PPV metrics. Evidently, RED-Net outperformed the other segmentation models across most evaluation metrics, demonstrating its superior capability in accurately segmenting the thyroid-gland region in chest CT scans.

TABLE 1. Results of RED-Net and other segmentation models for the NCC validation set.

	Attention U-Net	Swin-UNETR v2	DeepLabv3	SegRes Net	RED-Net
DSC	0.7589	0.8751	0.8883	0.8880	*0.9094*
JSC	0.6189	0.7874	0.8009	0.8019	*0.8361*
HD95	223.441	163.850	*10.122*	67.296	37.931
SE	0.9148	0.8892	0.8725	0.9112	*0.9192*
PPV	0.6599	0.8757	*0.9080*	0.8714	0.9036

The results presented in Table 2 confirm the superior segmentation capability of RED-Net than the other models. It achieved the highest DSC of 0.8901 and a JSC of 0.8087, indicating excellent segmentation accuracy, and a relatively low HD95 of 42.032, outperforming most models, except for DeepLabv3; however, DeepLabv3 did not perform as well on other metrics. Moreover, RED-Net maintained a high SE (0.9162) and PPV (0.8744), demonstrating a balanced performance in identifying and predicting TPs. These results indicate that RED-Net consistently outperformed the other models on key segmentation-performance metrics.

TABLE 2. Results of RED-Net and other segmentation methods for the CBNUH test set.

	Attention U-Net	Swin-UNETR v2	DeepLabv3	SegRes Net	RED-Net
DSC	0.7367	0.8442	0.8637	0.8651	*0.8901*
JSC	0.5965	0.7481	0.7676	0.7700	*0.8087*
HD95	208.758	179.106	*31.085*	66.414	42.032
SE	*0.9183*	0.8782	0.8666	0.9034	0.9162
PPV	0.6350	0.08398	0.8714	0.8399	*0.8744*

Table 3 lists the MAEs of all segmentation models for the measured thyroid volume compared with the labeled thyroid size. RED-Net achieved the lowest MAE of 1.10 cm^3 on the NCC validation set and 1.74 cm^3 on the CBNUH test set,

FIGURE 4. (a) (e) visualize the segmentation results of RED-Net, SegResNet, DeepLabv3, Swin-UNETRv2, and Attention U-Net for five consecutive axial slides. The numbers below each slide are DSCs for that slide. The enlarged images inside the slides are enlarged thyroid segmentation areas. The arrows indicate nodules inside the thyroid, and the models other than RED-Net fail to recognize these nodules as thyroid glands.

indicating superior accuracy in volume estimation compared to other models. Accurate volume estimation is critical for clinical assessments, and RED-Net's performance in this metric further establishes its practical utility. Accurate volume estimation is critical for clinical assessments, and RED-Net's performance in this metric further establishes its practical utility.

Each training epoch (920 cases) required 33 min to complete on a single GPU. Therefore, model training over 150 epochs entailed six days. Additionally, the inference time for each model was 2.41 s.

B. QUALITATIVE ANALYSIS OF RESULTS

As illustrated in Figure 4, the performances of five deep-learning models for thyroid-region segmentation in chest CT scans were compared: RED-Net, SegResNet, DeepLabv3, Swin-UNETRv2, and Attention U-Net.

TABLE 3. MAE of measured thyroid volume compared with labeled thyroid size. (Unit:cm^3).

	Attention U-Net	Swin-UNETR v2	DeepLab v3	SegRes Net	RED-Net
NCC	6.22	2.51	1.23	1.52	*1.10*
CBNUH	8.53	3.72	*1.58*	2.07	1.74

Each column presents the results of one model, along with the corresponding ground-truth and original images for reference.

Figure 4 shows five consecutive axial slices, with the DSC of each slice displayed below it. The areas pointed out by arrows indicate nodules within the thyroid gland. RED-Net successfully included these nodules as part of the thyroid region, closely aligning with the ground truth, while the other models failed to do so, resulting in segmentation

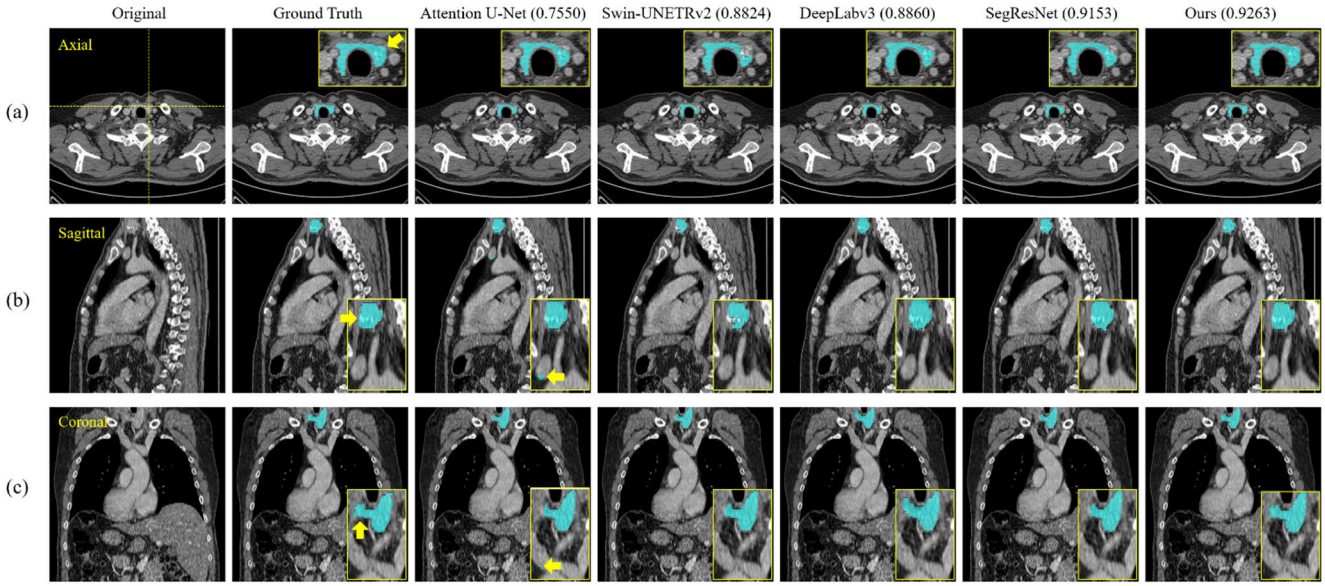

FIGURE 5. (a) is the same as Figure 4 (a). (b) shows the sagittal plane, (c) shows the coronal plane, and the vertical line of the original image of (a) corresponds to the sagittal plane, and the horizontal line corresponds to the coronal plane. The numbers next to the model names are the volumetric DSCs of the Chest CT scan in Figures 4 and 5.

distortions. Figure 5 illustrates the corresponding sagittal and coronal slices of the same example. Swin-UNETRv2, as seen in Figure 5(b), failed to identify the nodule, leading to its exclusion from the thyroid region. Similarly, Attention U-Net incorrectly segmented non-thyroid areas as part of the thyroid gland, as highlighted by the arrows. In contrast, as shown in Figure 5(c), RED-Net demonstrated the most accurate segmentation, aligning closely with the ground truth and effectively capturing the thyroid boundaries with minimal errors.

Figures 6 and 7 further validate these observations. Attention U-Net and Swin-UNETRv2 showed under-segmentation along the thyroid boundaries. While SegResNet and DeepLabv3 provided reasonably accurate segmentations that closely aligned with the ground truth, they exhibited slight inaccuracies, particularly in capturing fine details of the thyroid gland. Figure 7(b) shows that non-thyroid regions were misclassified as thyroid tissue in Attention U-Net. In contrast, RED-Net consistently delivered superior results, accurately delineating the thyroid gland and minimizing segmentation errors.

These superior results of RED-Net can be attributed to its architectural design, which combines residual connections and dilated convolutions to effectively capture both local details and global contextual information. The residual connections facilitate the training of deeper networks by mitigating the vanishing gradient problem, allowing the model to learn complex features essential for accurate segmentation. Dilated convolutions expand the receptive field without increasing the number of parameters, enabling the model to integrate multi-scale features crucial for identifying the thyroid gland amidst surrounding structures.

C. ABLATION STUDY

The results of the ablation study are presented in Tables 4 and 5, demonstrating the performance of the proposed RED-Net model compared to its ablated variants on the NCC validation set and the CBNUH test set. Each model was evaluated using several metrics, including DSC, JSC, HD95, SE, PPV, and MAE, to assess segmentation quality and overall performance.

RED-Net exhibited superior performance across most evaluation metrics, showcasing its effectiveness in segmenting thyroid-gland regions in chest CT scans. It achieved the highest DSC (0.9094) and JSC (0.8361), indicating strong overlap and segmentation accuracy. Additionally, the HD95 value of 37.931 underscores RED-Net's exceptional boundary segmentation accuracy, outperforming the w/o Res. (42.173) and w/o DCM (63.131) variants. Moreover, RED-Net maintained a high SE (0.9192) and PPV (0.9036), reflecting its robust ability to identify true positives and predict thyroid-gland regions with high precision. The MAE of RED-Net was the lowest (1.10 cm^3), indicating its capability to accurately measure thyroid volume, further validating the significance of incorporating both residual blocks and the Dilated Convolution Module (DCM).

The trends observed in the NCC validation set were consistent in the CBNUH test set. RED-Net achieved the highest DSC (0.8901) and JSC (0.8087), confirming its superior segmentation accuracy. Its HD95 value (42.032) further highlights its accurate boundary delineation compared to w/o Res. (48.545) and w/o DCM (59.446). These results indicate that the integration of residual connections and dilated convolutions enhances multiscale feature extraction and boundary

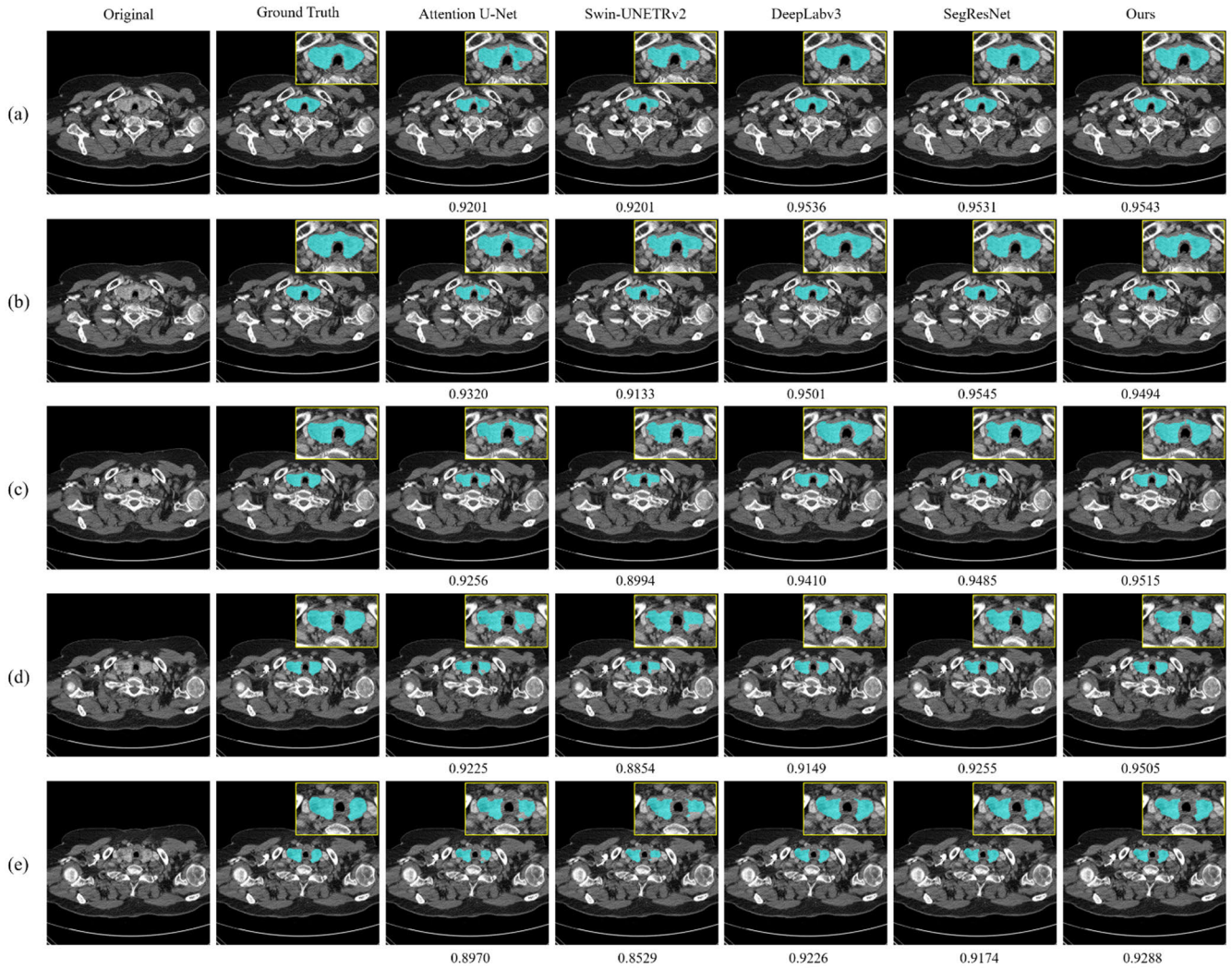

FIGURE 6. (a) (e) visualize the segmentation results of RED-Net, SegResNet, DeepLabv3, Swin-UNETRv2, and Attention U-Net for five consecutive axial slides. The numbers below each slide are DSCs for that slide. The enlarged images inside the slides are enlarged thyroid segmentation areas.

TABLE 4. Results of Ablation study for the NCC validation set.

	w/o Res. & w/o DCM	w/o DCM	w/o Res.	RED-Net
DSC	0.7583	0.8810	0.9024	*0.9094*
JSC	0.6304	0.7950	0.8306	*0.8361*
HD95	157.26	63.131	42.173	*37.931*
SE	0.9278	0.8878	*0.9391*	0.9192
PPV	0.6667	0.8866	0.8872	*0.9036*
MAE [a]	8.17	1.71	1.25	*1.10*

"w/o" means without and "Res." means residual convolutions.
[a]Unit:cm^3

detection. RED-Net also demonstrated a balanced performance on SE (0.9162) and PPV (0.8744), maintaining its robustness in identifying and predicting true positives. The MAE of RED-Net (1.74 cm^3) was comparable to the w/o Res. variant, indicating consistent accuracy in volumetric measurements.

The ablation study confirms that the inclusion of both residual connections and the Dilated Convolution Module in RED-Net plays a pivotal role in enhancing segmentation performance. The superior results of RED-Net across multiple datasets and evaluation metrics underscore its potential as a reliable and accurate tool for thyroid gland segmentation in chest CT imaging. By effectively combining residual learning and multiscale context aggregation, RED-Net outperforms its ablated variants, demonstrating improved overlap measures (DSC and JSC), boundary accuracy (HD95), sensitivity (SE), precision (PPV), and volumetric estimation (MAE). This highlights the synergistic effect of its architectural components in capturing the intricate structures of the thyroid gland and enhancing overall segmentation quality.

V. DISCUSSION

Most previous studies on thyroid segmentation have employed neck CT images [3], [5], [6], [16] which provide the clearest view of the thyroid gland. However, chest CT scans are more frequently conducted in clinical practice. During interpretation of chest CT scans, thyroid

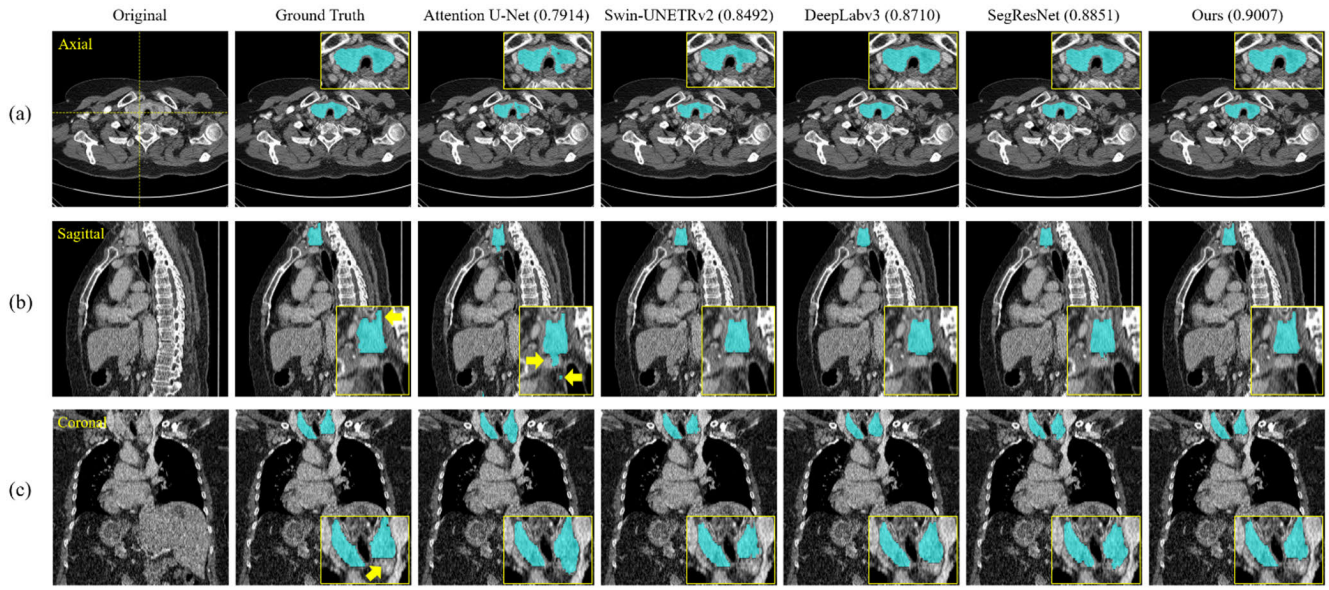

FIGURE 7. (a) is the same as Figure 6 (a). (b) shows the sagittal plane, (c) shows the coronal plane, and the vertical line of the original image of (a) corresponds to the sagittal plane, and the horizontal line corresponds to the coronal plane. The numbers next to the model names are the volumetric DSCs of the Chest CT scan in Figures 6 and 7.

TABLE 5. Results of Ablation study for the CBNUH test set.

	w/o Res.&DCM	w/o DCM	w/o Res.	RED-Net
DSC	0.7322	0.8744	0.8829	*0.8901*
JSC	0.6056	0.7861	0.7960	*0.8087*
HD95	152.34	59.446	48.545	*42.032*
SE	0.9149	0.8807	*0.9165*	0.9162
PPV	0.6484	*0.8807*	0.8600	0.8744
MAE [a]	9.58	*1.51*	1.71	1.74

"w/o" means without and "Res." means residual convolutions.
[a]Unit:cm^3

abnormalities are sometimes overlooked, and goiter cases may not be diagnosed. Therefore, this study aimed to segment the thyroid region in chest CT and measure its volume using various deep-learning models and ultimately developed an optimal model. Compared to the DSC of 0.76±0.11 (evaluated from 66 patients) of the previous study using multi-atlas label fusion and random forest classification [4] in automated segmentation of the thyroid gland on thoracic CT scans, our model has an overwhelmingly high DSC of 0.89 (evaluated from 600 patients).

In our experiments, we employed various network architectures and explored several alternative approaches. For example, we attempted to increase the batch size to eight, but owing to GPU memory limitations, this required cropping images to a smaller size, which led to performance decline. Conversely, increasing the network width (i.e., the number of features/filters) consistently enhanced the results. Notably, we found that optimal performance was achieved by reducing the downsampled feature map size to 4× the input image size, compared with downsizing to 8× or 16×. However, this study had some limitations. First, it used data from 1,150 patients from the NCC, which limited the dataset size. Second, the high computational demands of RED-Net led to prolonged inference times, indicating the need for further optimization.

VI. CONCLUSION

This study presents a significant advancement in thyroid segmentation and volume measurement from chest CT scans through the development of RED-Net. The proposed model effectively addresses the challenge of detecting thyroid abnormalities in routine chest CT scans, primarily focused on pulmonary and cardiac pathologies, by automating the process of thyroid segmentation. This contribution is crucial, as it ensures that thyroid diseases, such as goiter, are not overlooked, thereby improving diagnostic accuracy without increasing the radiologists' workload.

This paper proposed the RED-Net model, which employs residual blocks and DCM to delineate the thyroid gland in contrast-enhanced chest CT scans automatically. Thus, it can assist radiologists in identifying unsuspected goiters. The high performance of RED-Net, as demonstrated by its state-of-the-art DSC of 0.89, surpasses previous methods used for thoracic CT segmentation, which were primarily based on neck CT images or less accurate machine learning models. This study holds clinical significance by providing an automated and precise neural network model for thyroid segmentation in chest CT scans, offering the potential for early diagnosis and reducing the risk of undiagnosed thyroid conditions. Additionally, it opens the door for more efficient integration of deep-learning models in routine radiological practice, contributing to enhanced patient care and optimized resource allocation.

REFERENCES

[1] I. Oueslati, S. Salhi, M. Yazidi, F. Chaker, and M. Chihaoui, "A case of Hashimoto's thyroiditis following Graves' disease," *Clin. Case Rep.*, vol. 10, no. 10, p. 6466, Oct. 2022.

[2] I. C. Nam, K. H. Lee, J. Ryu, O. Kim, S. H. Kim, H. J. Baek, Y. Lee, T. N. Kim, M.-K. Kim, S.-J. Kim, and S. M. Kim, "Quantitative analysis of 3-dimensional volumetry and histogram of thyroid gland on neck computed tomography for patients with Hashimoto's thyroiditis," *J. Korean Soc. Radiol.*, vol. 73, no. 6, p. 367, 2015.

[3] C.-Y. Chang, P.-C. Chung, Y.-C. Hong, and C.-H. Tseng, "A neural network for thyroid segmentation and volume estimation in CT images," *IEEE Comput. Intell. Mag.*, vol. 6, no. 4, pp. 43–55, Nov. 2011.

[4] D. Narayanan, J. Liu, L. Kim, K. W. Chang, L. Lu, J. Yao, E. B. Turkbey, and R. M. Summers, "Automated segmentation of the thyroid gland on thoracic CT scans by multiatlas label fusion and random forest classification," *J. Med. Imag.*, vol. 2, no. 4, Dec. 2015, Art. no. 044006.

[5] X. He, B. J. Guo, Y. Lei, S. Tian, T. Wang, W. J. Curran, L. J. Zhang, T. Liu, and X. Yang, "Thyroid gland delineation in noncontrast-enhanced CTs using deep convolutional neural networks," *Phys. Med. Biol.*, vol. 66, no. 5, Mar. 2021, Art. no. 055007.

[6] X. Wen, B. Zhao, M. Yuan, J. Li, M. Sun, L. Ma, C. Sun, and Y. Yang, "Application of multi-scale fusion attention U-Net to segment the thyroid gland on localized computed tomography images for radiotherapy," *Frontiers Oncol.*, vol. 12, May 2022, Art. no. 844052.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," 2015, *arXiv:1505.04597*.

[8] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense, volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*. Cham, Switzerland: Springer, 2016, pp. 424–432.

[9] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," 2014, *arXiv:1412.7062*.

[10] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[11] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.

[12] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Computer Vision—ECCV*. Cham, Switzerland: Springer, 2018, pp. 833–851.

[13] A. Myronenko, "3D MRI brain tumor segmentation using autoencoder regularization," 2018, *arXiv:1810.11654*.

[14] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.

[15] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, and D. Xu, "Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (Lecture Notes in Computer Science). Cham, Switzerland: Springer, 2022, pp. 272–284.

[16] A. D'Aviero, A. Re, F. Catucci, D. Piccari, C. Votta, D. Piro, A. Piras, C. Di Dio, M. Iezzi, F. Preziosi, and S. Menna, "Clinical validation of a deep-learning segmentation software in head and neck: An early analysis in a developing radiation oncology center," *Int. J. Environ. Res. Public Health*, vol. 19, no. 15, p. 9057, Jul. 2022.

[17] R.-L. Xie, Y. Wang, Y.-N. Zhao, J. Zhang, G.-B. Chen, J. Fei, and Z. Fu, "Lung nodule pre-diagnosis and insertion path planning for chest CT images," *BMC Med. Imag.*, vol. 23, no. 1, p. 22, Feb. 2023.

[18] T. Arsenescu, R. Chifor, T. Marita, A. Santoma, A. Lebovici, D. Duma, V. Vacaras, and A. F. Badea, "3D ultrasound reconstructions of the carotid artery and thyroid gland using artificial-intelligence-based automatic segmentation—Qualitative and quantitative evaluation of the segmentation results via comparison with CT angiography," *Sensors*, vol. 23, no. 5, p. 2806, Mar. 2023.

[19] H. Lu, M. Li, K. Yu, Y. Zhang, and L. Yu, "Lumbar spine segmentation method based on deep learning," *J. Appl. Clin. Med. Phys.*, vol. 24, no. 6, p. 13996, Jun. 2023.

[20] W. Li, H. Song, D. Ai, J. Shi, Y. Wang, W. Wu, and J. Yang, "Semi-supervised segmentation of orbit in CT images with paired copy-paste strategy," *Comput. Biol. Med.*, vol. 171, Mar. 2024, Art. no. 108176.

[21] A. M. Alkhadrawi, L. Y. Lin, S. A. Langarica, K. Kim, S. K. Ha, N. G. Lee, and S. Do, "Deep-learning based automated segmentation and quantitative volumetric analysis of orbital muscle and fat for diagnosis of thyroid eye disease," *Investigative Ophthalmol. Vis. Sci.*, vol. 65, no. 5, p. 6, May 2024.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Computer Vision—ECCV*. Cham, Switzerland: Springer, 2016, pp. 630–645.

[23] Y. Wu and K. He, "Group normalization," *Int. J. Comput. Vis.*, vol. 128, no. 3, pp. 742–755, Mar. 2020.

[24] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

[25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[26] M. Hermsen, T. de Bel, M. den Boer, E. J. Steenbergen, J. Kers, S. Florquin, J. J. Roelofs, M. D. Stegall, M. P. Alexander, B. H. Smith, B. Smeets, L. B. Hilbrands, and J. A. van der Laak, "Deep learning–based histopathologic assessment of kidney tissue," *J. Amer. Soc. Nephrol.*, vol. 30, no. 10, pp. 1968–1979, 2019.

[27] T. Eelbode, J. Bertels, M. Berman, D. Vandermeulen, F. Maes, R. Bisschops, and M. B. Blaschko, "Optimization for medical image segmentation: Theory and practice when evaluating with dice score or Jaccard index," *IEEE Trans. Med. Imag.*, vol. 39, no. 11, pp. 3679–3690, Nov. 2020.

[28] R. Parikh, A. Mathai, S. Parikh, G. C. Sekhar, and R. Thomas, "Understanding and using sensitivity, specificity and predictive values," *Indian J. Ophthalmol.*, vol. 56, no. 1, pp. 45–50, 2008.

[29] Y. A. E.-W. Skaik, "Understanding and using sensitivity, specificity and predictive values," *Indian J. Ophthalmol.*, vol. 56, no. 4, p. 341, 2008.

[30] S. K. Talluri, "Positive predictive value," *BMJ*, vol. 339, p. 3835, Sep. 2009.

MIN-JI KIM received the Bachelor of Business Administration degree, in 2021, and the master's degree in artificial intelligence from the Department of Information Convergence Engineering, in 2023.

She is currently a Researcher with the Healthcare AI Team, National Cancer Center. Her research interests include computer vision and deep learning.

Ms. Kim received the Best Paper Award at the 19th International Conference on Multimedia Information Technology and Applications (MITA 2023).

JIN-A KIM received the bachelor's degree in radiology in South Korea, in 2015, and the degree in radiographer's license.

She has over seven years of experience as a Radiographer and is currently a Researcher with the Healthcare AI Team, National Cancer Center. Her research interests include CT medical imaging analysis and the application of artificial intelligence technology in radiology.

NAAE KIM received the bachelor's degree in materials science and engineering from Dongseo University, South Korea, in 2012.

She is currently a Researcher with the Research and Development Business Foundation, National Cancer Center. Her research interests include segmentation for accurately identifying and extracting lesions from medical images.

YUL HWANGBO was born in South Korea, in 1981. He received the degree in medical license, in 2006, the bachelor's, master's, and Ph.D. degrees from Seoul National University College of Medicine, in 2020, and the M.D. degree.

He became a Board-Certified Internal Medicine Specialist, in 2011, and a Specialist in endocrinology and metabolism, in 2016. Since 2015, he has been a Clinical Physician with the National Cancer Center, South Korea, and in 2019, he became a Senior Researcher. From 2018 to 2020, he was the Head of the IT Department. Since 2019, he has been with the Healthcare AI Team, National Cancer Center.

Dr. Hwangbo founded Weknew Inc., in 2021. He has been the Director of Content with Korean Society of Medical Informatics, since 2019, and has been the Director of Education with Korean Society of Artificial Intelligence in Medicine, since 2023.

HYUN JEONG JEON received the M.D. degree from Chungbuk National University College of Medicine, Cheongju-si, South Korea.

She completed the Internal Medicine Specialist Course with Chungbuk National University Hospital. Her fields of practice include the endocrine systems, diabetes, osteoporosis, and thyroid disorders. She became the Chief Clinical Officer of the Endocrinology Department, Cheongju St. Mary's Hospital. She is currently a Professor with Chungbuk National University College of Medicine.

Prof. Jeon served as a fellow for the Endocrinology Department, Seoul National University Hospital.

DONG-HWA LEE received the M.D., M.S., and Ph.D. degrees from Chungbuk National University College of Medicine, Cheongju-si, South Korea.

She completed her internship and Internal Medicine Specialist Training with Chungbuk National University Hospital. She was a Clinical Instructor with the Endocrinology Department, Seoul National University Hospital, and Bundang Seoul National University Hospital, later becoming a Clinical Professor with Bundang Seoul National University Hospital. She is currently an Associate Professor with the Endocrinology Department, Chungbuk National University College of Medicine.

Prof. Lee's specialties include diabetes, thyroid disorders, osteoporosis, and obesity. She is an active member of several professional societies, including Korea Geriatrics Society, Korean Society for the Study of Obesity, Korean Society of Lipid and Atherosclerosis, Korean Thyroid Association, Korean Diabetes Association, and Korean Endocrine Society.

JI EUN OH received the M.S. and Ph.D. degrees in radiological science from Yonsei University, South Korea, in 2008 and 2013, respectively.

In 2013, she joined the National Cancer Center Research Institute, South Korea, where she is currently an Assistant Professor with the Department of Cancer Biomedical Science and a Researcher with the Healthcare AI Team. Her research interests include medical image analysis, computer-aided diagnosis, and deep learning techniques for precision medicine and its applications.