IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 12, 2020, accepted October 25, 2020, date of publication November 4, 2020, date of current version November 18, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3035772

Esophagus Segmentation in Computed Tomography Images Using a U-Net Neural Network With a Semiautomatic Labeling Method

XIAO LOU[1,2,3], YOUZHE ZHU[4], KUMARADEVAN PUNITHAKUMAR[3], (Senior Member, IEEE), LAWRENCE H. LE[3], (Member, IEEE), AND BAOSHENG LI[2,1]

[1]Laboratory of Image Science and Technology, School of Computer Science and Engineering, Southeast University, Nanjing 210096, China
[2]Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan 250117, China
[3]Department of Radiology and Diagnostic Imaging, University of Alberta, Edmonton, AB, Canada
[4]Department of Radiation Oncology, The First Hospital of Lanzhou University, Lanzhou 730000, China

Corresponding author: Baosheng Li (bsli@sdfmu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 81530060 and Grant 81874224, in part by the National Key Research and Development Program of China under Grant 2016YFC0105106, in part by the Provincial Key Research and Development Program of Shandong under Grant 2017CXZC1206, in part by the Academic Promotion Program of Shandong First Medical University and Shandong Academy of Medical Sciences under Grant 2019LJ004, and in part by the China Scholarship Council (CSC).

ABSTRACT Esophagus segmentation in computed tomography images is challenging due to the complex shape and low contrast of the esophagus. Fully automated segmentation is feasible with recent convolutional neural network approaches, such as U-Net, which reduce variability and increase reproducibility. However, these supervised deep learning methods require radiologists to laboriously interpret and label images, which is time-consuming, at the expense of patient care. We propose an esophagus segmentation method using a U-Net neural network combined with several variations of backbones. We also propose a semiautomatic labeling method with detection and execution components to solve the labeling problem. The detection component identifies the category to which each slice belongs using the bag-of-features method. The edges in each category are clustered using contour moments and their topological levels as features. In the execution component, the assumed esophageal contours are predicted by the clustered model. A convex hull approach and level set algorithm yield the final esophageal contours, which are employed to train the neural network. Several backbones are implemented as the encoder of the U-Net network to extract features. The predictions are then compared with those obtained via manual labeling by a radiologist and the segmentation results generated by the proposed semiautomatic method. The experimental evaluations demonstrate that the utilization of ResneXt50 and InceptionV3 as backbones with U-Net is more effective than that with other backbones. A three-dimensional rendering of the segmented model is performed to exhibit the prediction. The results demonstrate that the proposed method outperforms previously published methods.

INDEX TERMS Computed tomography, deep learning, esophagus, segmentation, U-Net.

I. INTRODUCTION

In the examination of thoracic tumors, the discovery of esophageal carcinoma has gradually increased in recent years [1], [2]. Determining the location and size of the esophagus in computed tomography (CT) images is important for radiologists in target volume delineation [3], [4]. However, this task is often difficult and time-consuming due to the complex structure of the esophagus, the randomness of the position in each slice, and the low contrast against the surrounding organs and tissues [5], [6]. Automatic segmentation techniques have been proposed for various application domains in medical imaging [7]–[12], and a similar approach for esophagus segmentation could help radiologists complete this task more accurately and efficiently. In addition, a precise segmentation technique could assist medical physicists in enclosing a lesion more easily so that the radiation dose to organs at risk can be reduced in radiotherapy treatment planning.

The associate editor coordinating the review of this manuscript and approving it for publication was Mehul S. Raval.

There have been few reported studies on esophagus segmentation, and the methods presented in most published articles require a large amount of prior knowledge. In addition, the utilized methods and models are sensitive to the data supplied by the user and the parameters involved. In a study by Rousson *et al.* in 2006 [13], a probabilistic spatial model was employed to extract the centerline of the esophagus, and the outer wall of the esophagus was approximated by elliptical shape on each slice using a region-based criterion. This method requires pre-provisioning of the locations of two pixels in the esophagus and pre-segmentation of the left atrium and aorta. In the same year, Huang *et al.* published a semiautomatic esophagus segmentation method [14]. Based on a segmented slice, the method propagated the contour to other slices that are registered using optical flow. In an article published by Fieselmann *et al.* [15] in 2008, the esophageal contours of each axial slice drawn in advance were transformed to the frequency domain for segmentation. In a study by Feulner *et al.* [16], a detection and connection method was applied to estimate the approximate shape of the esophagus, and a classifier was utilized to obtain fine-grained segments. The work of Kurugol *et al.* [17] used tissues surrounding the esophagus, that were segmented manually in a specific slice, to determine the location of the esophageal centerline, which served as the initialization for level set segmentation. Yang *et al.* [18] proposed an automatic method by selecting a subset of optimal atlases for multi-atlas segmentation. This method relies on cloud data and the existence of an esophageal atlas.

In recent years, deep convolutional neural network methods have been applied to esophagus segmentation. In 2017, Hao *et al.* [19] used a fully convolutional network (FCN) to develop an esophageal tumor classifier with expert-labeled tumor regions during training. In the same year, Trullo *et al.* [20] modified the FCN architecture to improve the accuracy of esophageal location and segmentation performance. In 2019, Chen *et al.* [21] compared the performance of esophagus segmentation using several types of evolutionary FCNs, which predicted the images in an esophageal CT scan slice by slice.

In this article, an esophagus segmentation method, which is based on the U-Net neural network [22] that utilizes training data generated with a semiautomatic labeling method, is proposed. Several studies have demonstrated that U-Net is highly effective for semantic medical image segmentation [23]–[26]. The proposed deep learning approach provides an option for accurate and automated esophagus segmentation. However, the ground truth of a dataset is obtained by manual labeling, which is time-consuming and a limiting factor in neural network training due to the amount of available data. A typical scan of the entire esophagus consists of more than 80 slices, and a radiologist must manually label a large number of slices in each scan to completely utilize the data set. Obtaining a high-performance neural network with big data that solely relies on manual labeling is impractical. To address this challenge, a semiautomatic segmentation method is proposed in this study to accelerate the generation of the ground truth.

The proposed semiautomatic segmentation method relies primarily on contour moments and the topological level of the contour to cluster the edges. In the proposed method, esophageal contours can be predicted by a clustered model. After a morphological operation and other postprocessing, the convex hull and level set algorithms are employed to generate the final confirmed esophageal contours. These contours are then employed to train the U-Net neural network. Further details on the semiautomatic method are provided in Section II. Several backbones, *e.g.,* ResNet, ResneXt, InceptionNet, and EfficientNet are utilized as the encoder of the U-Net network's downsampling step to convert the input into a certain feature representation. The entire test set is appended by manual labeling and is applied to estimate the accuracy of the prediction.

II. METHOD

The main architecture of the proposed method is presented in this section, and a flow chart of the method is provided in Fig. 1. First, a semiautomatic segmentation method is created to generate the ground truth of the esophagus in the CT images. Second, the size of the training and validation sets is augmented by horizontal and vertical shifts, shearing, rotation, scaling, and horizontal flipping with a normalization operation. Third, the expanded dataset is processed by various types of backbones, and the data produced by the different backbones are employed in the experiment. Last, U-Net is used to train the processed training and validation sets. The model after training is utilized to predict the esophageal region of the test set.

A. SEMIAUTOMATIC SEGMENTATION LABELING METHOD

Conventional supervised deep learning methods utilize manual labeling to generate the ground truth for training, which is relatively inefficient. For esophageal CT images, a complete scan usually consists of more than 80 tomographic images depending on the slice thickness and length of the esophagus. Producing the entire training and validation dataset by manual labeling would thus be very time consuming and laborious. Therefore, a semiautomatic segmentation method is proposed, and its results are applied as the ground truth for training.

The semiautomatic segmentation method consists of two stages: detection and execution. A flow diagram of the method is provided in Fig. 2. In the training stage, all thoracic CT images that are involved in the experiment are first classified by the bag-of-features method [27], [28]. Based on its clinical definition, the esophagus is divided into four main parts: the cervical part, upper thoracic part, middle thoracic part, and lower thoracic part. The esophagus categories after classification in this study are much more fine-grained than those in general clinical applications. The purpose of this classification step is to define a certain point where every slice is located vertically. As a result, the features

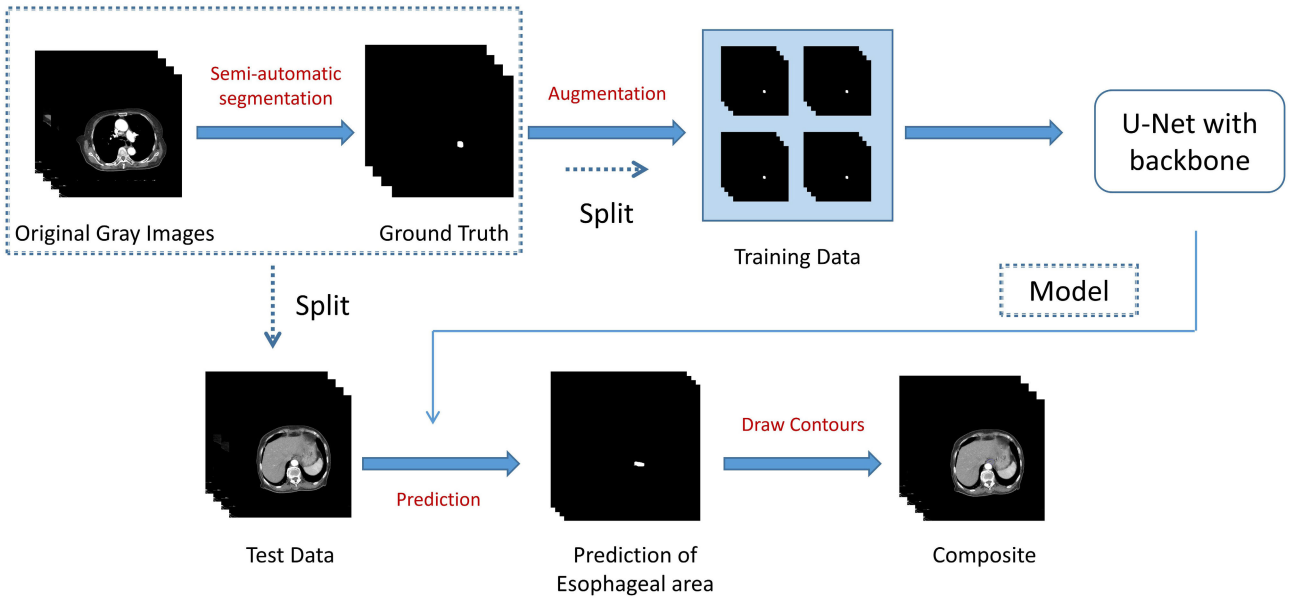

FIGURE 1. Flow chart of the proposed method for segmenting the esophagus in CT images.

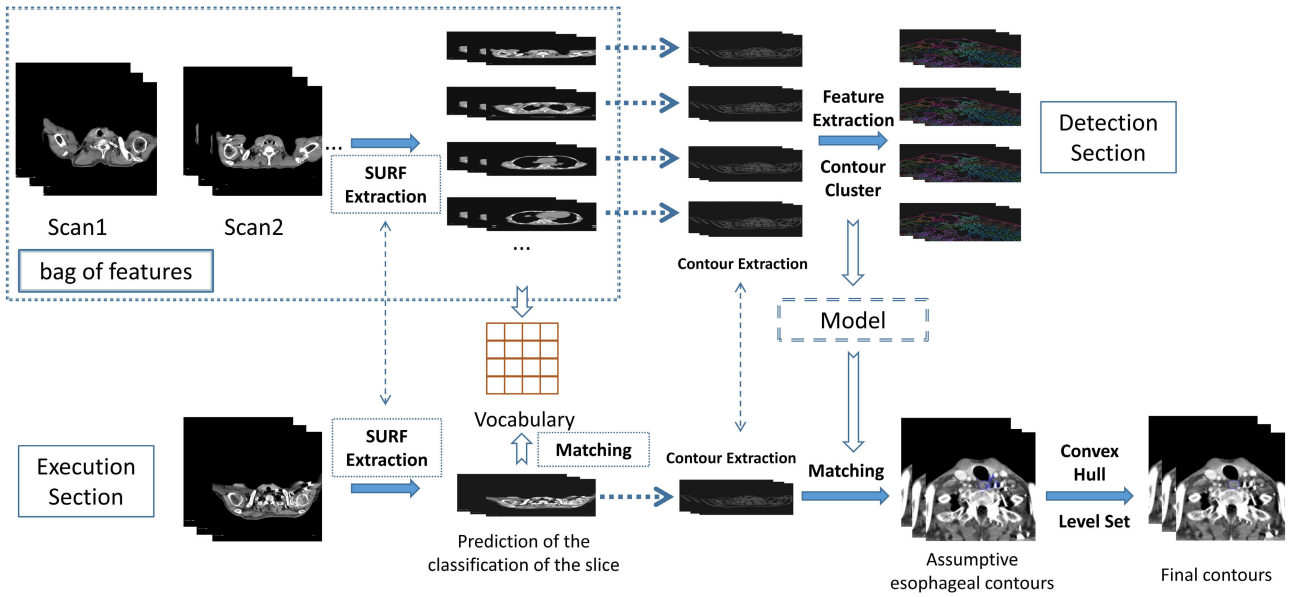

FIGURE 2. Main steps of the semiautomatic segmentation method. The detection and execution processes are shown at the top and bottom of the figure respectively.

of images in the same category of the esophagus are sufficiently similar, which benefits the subsequent clustering process for the contours. Second, the edges in all images are extracted by the Canny operator after a series of preprocessing operations (*e.g.,* contrast stretching). Interpolation and morphological operations are employed on the edges in every image to make them smoother and more consecutive. Third, the Canny operator is employed again while the contours of the processed edges are estimated and their moments (first- to third-order geometric moments and central moments) are calculated. Additionally, the topological levels of the contours are determined by the preorder traversal method in a topological tree [29], and the level of every contour is recorded. Fourth, the moments and topological levels of the contours are regarded as features of the edges. Last, the contours are clustered by the K-means algorithm [30], and the cluster of the esophageal contours can be selected visually from the image.

The position of the esophagus in different slices is predicted by the bag-of-features method employed in the training

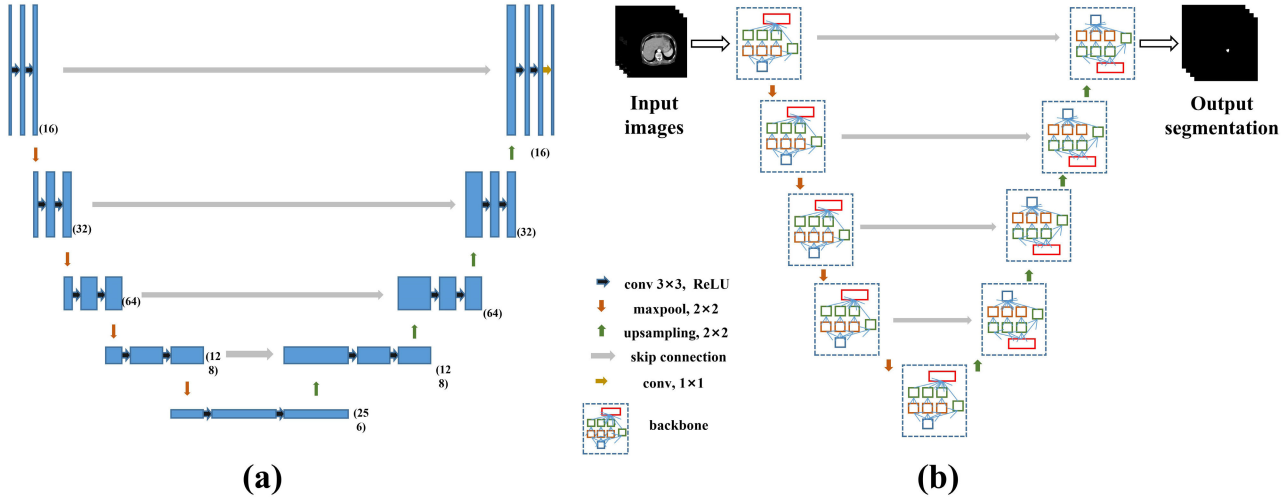

FIGURE 3. (a) Architecture of the U-Net neural network. (b) Schematic chart of the data flow via the network combined with the backbone unit. ResNet, SE-ResNet, ResNeXt, InceptionNet, etc. are employed as backbones in this work.

stage. Thus, in the execution stage, the initial steps are strictly identical to those in the detection stage. Preprocessing and feature extraction by means of the speeded up robust features (SURF) algorithm [31], [32] are performed . The SURF features in every slice are matched with the vocabularies created in the bag-of-features method. The categories of the slices are determined, and the edges can be extracted according to the categories. The moments and topological levels can also be calculated. With the model trained as shown in the last step of Fig. 2, the assumed contours that belong to the esophagus are predicted. Thereafter, postprocessing, such as morphological operations and node removal, is performed. The convex hull method is used to depict the external boundary of the contours, which eliminates some of the isolated islands and confusions of the internal edges that are enclosed by esophageal contours. The external boundaries serve as the initialization of the level set algorithm [33], [34].

B. ARCHITECTURE OF THE U-NET NEURAL NETWORK

Since 2015, the development of the FCNs has led to a wave of research in the field of image segmentation [35]. In the same year, the U-Net neural network was developed. This network is characterized by using a small amount of data as the training and validation sets, and having a concise network structure and a high training speed.

The architecture of the U-Net network is shown in Fig. 3(a). The U-Net network is composed of an encoder section and a decoder section. The left part is the encoder section, which corresponds to the downsampling phase of a traditional classification network. The right part is the decoder section, which corresponds to the upsampling phase. The gray arrows in the middle are skip connections, which stitch shallow features to deep features. The shallow layers can usually capture the simple features of an image, such as borders and colors, whereas deeper convolution operations capture more abstract features. This network uses both shallow features and deep features, which enables the decoder to learn relevant features that are lost in the encoder phase. The general schematic of the U-Net based neural networks with different backbones, such as ResNet, SE-ResNet, ResNeXt, and InceptionNet, is given in Fig. 3(b). Although these neural networks consist of encoder and decoder components and skip connections similar to the original U-Net, the structure and sequence of layers vary for each backbone at different levels of the encoder and decoder components.

III. EXPERIMENT

In this experiment, all CT images were provided by the Department of Radiation Oncology at Shandong Cancer Hospital. There were 17 scans in this experiment, with a total of 1,397 slices. Each scan involved the entire esophagus from the cervical part to the lower thoracic part. The original size of the image in each slice was 512 × 512 pixels, which corresponds to a spatial area of 404 mm × 404 mm. The thickness of a slice ranged from 3 mm to 5 mm.

The entire dataset was divided into a training set, validation set, and test set, in a ratio 13:2:2, and the specific number of images in each set were 1066, 170, and 161, respectively. Due to the random location of the esophagus in the thorax, differences in body size, and unintentional shifting of patients when lying inside the CT gantry, the dataset was adjusted in accordance with the variations in the esophageal location. Therefore, the training set was augmented with horizontal and vertical shifting, shearing, rotation, scaling, and horizontal flipping, which were performed automatically using the *ImageDataGenerator* function of the Keras module at each epoch.

A. COMPUTATIONAL FACILITY

The experiment was based on a supercomputer with four NVIDIA GeForce GTX 2080 Ti graphics processing units affiliated with the University of Alberta. The process for

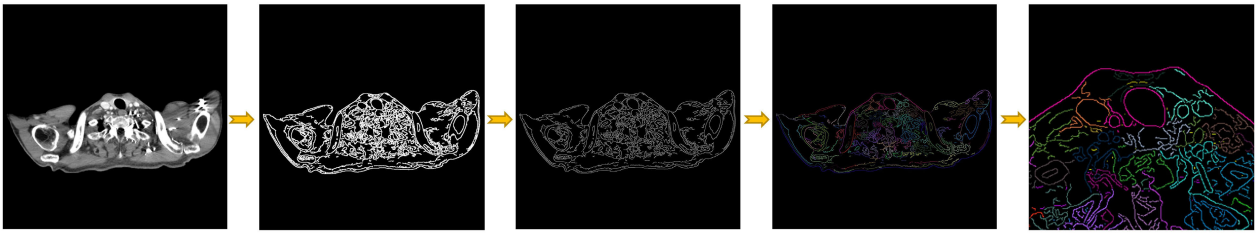

FIGURE 4. An example of the main training processes of the semiautomatic segmentation method using CT images: (a) an image after interpolation and smoothness; (b) the first Canny operator and morphology operation on (a); (c) the second Canny operator on (b); (d) the K-means clustering after feature extraction on (c); (e) enlargement of the esophageal area in (d).

training the neural network was configured in the Python 3.7 environment based on Keras and TensorFlow modules. The semiautomatic segmentation method was implemented in Visual Studio C++ 2019 with the OpenCV library.

B. EVALUATION CRITERIA

Although the semiautomatic segmentation method was created to serve as the ground truth, manual labeling was also employed in the test set to evaluate the performance of the proposed method. The mean surface distance (MSD), Hausdorff distance (HD), and Dice coefficient (DC) were calculated in pairs of contours to compare the differences in the ground truth generated by the semiautomatic segmentation method, prediction of the test set, and manual labeling.

Let us assume that there are two contours in an image that can be depicted as discrete point sets $A = (a_1, a_2, \ldots, a_p)$ and $B = (b_1, b_2, \ldots, b_q)$. The MSD can be defined as

$$M(A, B) = max[m(A, B), m(B, A)], \tag{1}$$

where

$$m(A, B) = (1/p) \sum_{i=1}^{p} \min_{b \in B} \|a_i - b\|, \tag{2}$$

$$m(B, A) = (1/q) \sum_{i=1}^{q} \min_{a \in A} \|b_i - a\|, \tag{3}$$

$min_{b \in B} \|a_i - b\|$ and $min_{a \in A} \|b_i - a\|$ represent the minimum distances between point a_i and A and between b_i and B respectively. Equations (2) and (3) represent the average distances from all points in A to B and from all points in B to A respectively. The HD is defined as

$$H(A, B) = max[h(A, B), h(B, A)], \tag{4}$$

where $h(x, y)$ is the maximum distance between x and y

$$h(x, y) = \max_{a \in x} \min_{b \in y} \|a - b\|, \tag{5}$$

$x(y)$ is $A(B)$ or $B(A)$. The DC is used to estimate the degree of overlap between two areas. Assuming that the two contours are padded, the DC can be defined as

$$DC = (2|X \cap Y|)/(|X| + |Y|), \tag{6}$$

where X and Y represent the padded areas of two targeted areas.

IV. RESULTS AND ANALYSIS

The images in Fig. 4 illustrate the primary steps in the detection stage of the semiautomatic segmentation labeling method. The final step visually presents the assumed esophageal contours, and the cluster of the esophagus can then be selected. Fig. 5(a) presents the assumed esophageal contours after prediction by the model trained with the K-means algorithm in the execution stage. Fig. 5(b) presents the results after the refinement of these contours and the convex hull operation. The polygon produced by the convex hull method is regarded as the initialization of the level set method. By evolving this partial segmentation method for approximately 9 seconds, the final periphery of the esophagus appears, as illustrated in Fig. 5(c).

To train the neural network, the Adam optimizer was employed [36], and the loss was evaluated by a combination of the binary focal loss and Dice loss. The learning rate was modified by a factor of 0.2 when the evaluation metric stopped improving, and 5 was chosen as the number of epochs to wait with no improvement before changing the learning rate. An epoch of 100 was selected, and the best model from the loss curve was employed for prediction.

Although the ground truth was not utilized in the training, it was valuable in the evaluation of the results. The ground truth coincided with the manual labeling to some extent. However, the segmentation is not an entirely automatic technique. The purpose of training the ground truth by the deep learning method is that once the model is trained, the esophagus can be predicted automatically. According to previous studies, the precision of the prediction is as expected.

We utilized a total of 26 different backbones in the neural network, including ResNet [37] (ResNet34, ResNet50, ResNet101, ResNet152), SE-ResNet [38] (SE-ResNet18, SE-ResNet34, SE-ResNet50, SE-ResNet101), ResNeXt [39](ResNeXt50, ResNeXt101), SE-ResNeXt [39] (SE-ResNeXt50, SE-ResNeXt101), DenseNet [40] (DenseNet121, DenseNet169, DenseNet201), InceptionV3 [41], InceptionResNetV2 [42], MobileNet [43], MobileNetV2 [44], and EfficientNet [45] (EfficientNetB0, EfficientNetB1,

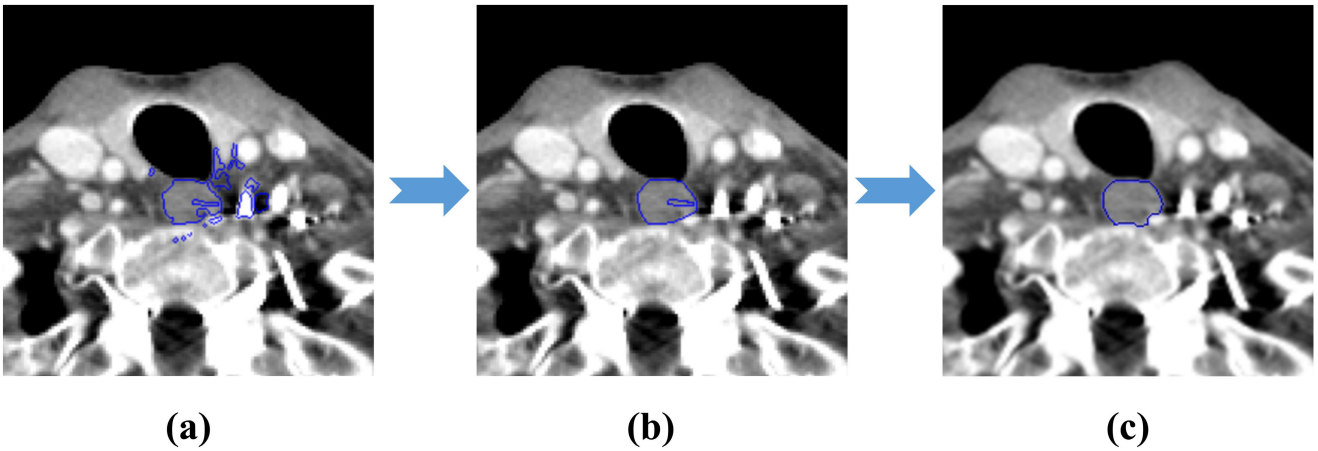

FIGURE 5. Example of the main execution processes of the semiautomatic segmentation method: (a) the assumed esophageal contours predicted by the *K*-means algorithm; (b) result after refinement of (a) and the convex hull operation; and (c) the final result after level-set operations.

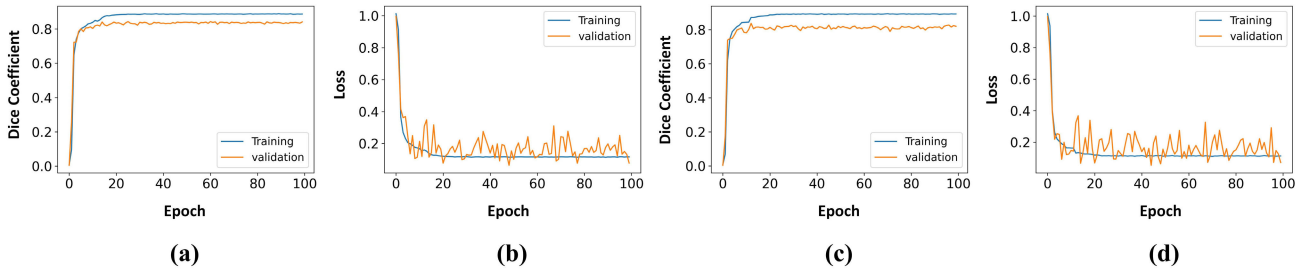

FIGURE 6. Accuracy and loss values at each epoch during the training process using the U-Net neural network: (a) and (b) using the backbone of ResneXt50; (c) and (d) using the backbone of InceptionV3. The segmentation accuracy in (a) and (c) is evaluated in terms of the DC. The loss values in (b) and (d) are evaluated in terms of a combination of the focal and Dice loss values.

TABLE 1. Esophagus segmentation results for the methods using different types of neural net backbones with the U-Net architecture.

Backbone	Neural net prediction vs semiautomatic segmentation			Neural net prediction vs manual labeling		
	MSD (mm)	HD (mm)	DC (%)	MSD (mm)	HD (mm)	DC (%)
ResNet101	1.23 ± 1.94	$\mathbf{3.30 \pm 1.75}$	85.73 ± 10.27	$\mathbf{1.07 \pm 1.02}$	$\mathbf{1.55 \pm 1.46}$	85.71 ± 9.10
ResneXt50	1.14 ± 1.25	3.63 ± 2.45	85.41 ± 9.53	1.14 ± 1.13	1.74 ± 1.71	$\mathbf{85.84 \pm 12.91}$
Se-ResNet50	0.96 ± 0.76	3.62 ± 2.01	86.60 ± 7.75	1.39 ± 2.43	1.91 ± 2.72	83.84 ± 16.47
InceptionV3	$\mathbf{0.86 \pm 0.65}$	$\mathbf{3.27 \pm 2.06}$	$\mathbf{87.74 \pm 7.16}$	$\mathbf{1.01 \pm 0.79}$	$\mathbf{1.50 \pm 1.37}$	$\mathbf{86.80 \pm 6.43}$
InceptionresnetV2	$\mathbf{0.89 \pm 0.75}$	3.35 ± 2.31	$\mathbf{87.04 \pm 8.98}$	1.15 ± 1.71	1.73 ± 1.90	85.79 ± 12.40
Efficientnetb3	1.57 ± 1.84	5.50 ± 3.99	82.23 ± 12.17	1.47 ± 3.13	1.73 ± 3.06	83.94 ± 15.82
Original U-Net	1.34 ± 1.55	2.10 ± 2.13	79.25 ± 21.80	1.07 ± 0.96	1.97 ± 1.87	76.07 ± 28.95

MSD, HD, and DC refer to mean surface distance, Hausdorff distance, and Dice coefficient, respectively. For each metric, the best two accuracy values are given in bold font.

EfficientNetB2, EfficientNetB3, EfficientNetB4, EfficientNetB5, EfficientNetB6). All the networks were trained independently. The best five results of the networks are presented in Table 1. The last item in Table 1 indicates the U-Net network trained without any backbone. The experimental evaluations demonstrate the advantage of utilizing different backbones with U-Net over the original U-Net approach for obtaining more accurate fully automated segmentation results. The best two accuracy values for each metric are presented in bold font, and the network with ResNeXt50 and InceptionV3 had a closer performance to manual labeling than the other methods. The accuracy and loss during the training of these two types of backbones are presented in Fig. 6. The accuracy was estimated by the DC, while the loss was estimated by a combination of the focal loss and Dice loss. The trained neural network model with the best validation accuracy was saved during the entire training process by utilizing *ModelCheckpoint* and *Callback* functions in the Keras module. The loss values of ResneXt50 and InceptionV3 in the best model were 0.063 and 0.053, respectively. Although the neural network training did not utilize manual labeling as the ground truth, the accuracy of the final results demonstrates that the semiautomatic segmentation method may be effective for labeling the CT images. The proposed

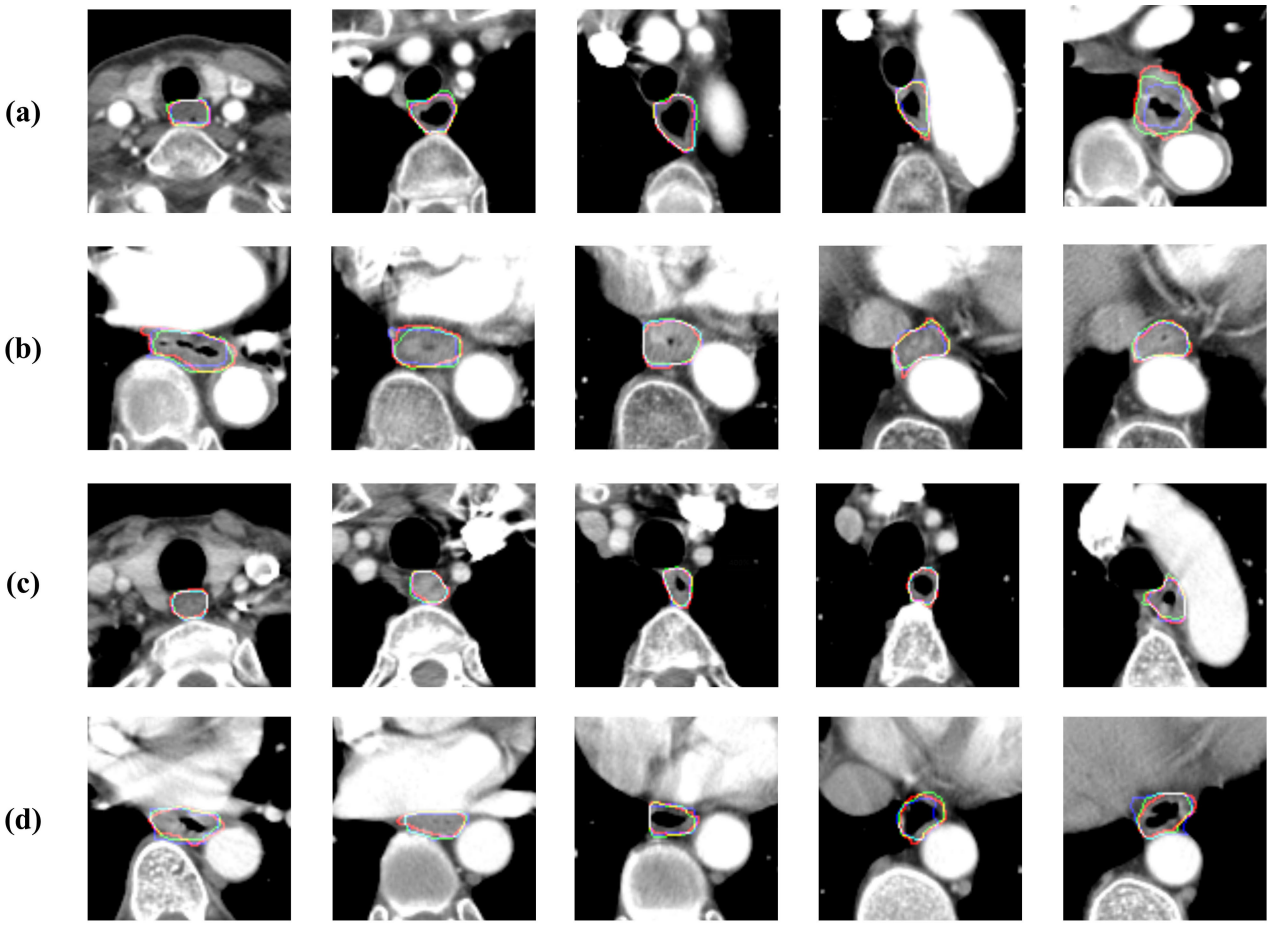

FIGURE 7. Examples of the CT images in the test set after the prediction of the model trained by U-Net with the InceptionV3 backbone. (a) and (b) are images from the same patient, and (c) and (d) are images from another patient. The blue contours represent the ground truth obtained by the semiautomatic segmentation; the green contours represent the prediction of the test set; the red contours represent the manual delineation. Overlapping pixels are replaced by composite colour.

method can thus help improve the efficiency of deep learning research. During the training phase, the trainable neural network parameters are computed only using the training set, and the validation set is employed merely to compute the neural network's prediction accuracy leading to fluctuations in the validation loss curve, as shown in Fig 6(b) and (d).

Fig. 7 presents examples of the results obtained by employing U-Net with the InceptionV3 backbone. The specimens in Fig. 7(a) and Fig. 7(b) represent a scan in the test set, whereas the specimens in Figs. 7(c) and (d) represent a different patient scan. Overlapping pixels are replaced by a composite color. The prediction coincides closely with the manual labeling, which is objectively reflected in Table 1 with an MSD of 1.01 ± 0.79 mm, HD of 1.50 ± 1.37 mm, and DC of 86.80 ± 6.43 (%).

To illustrate the area of the esophagus more distinctly, the regions of interest (ROIs) in Fig. 7 are enlarged. The process of prediction works globally in every slice. In contrast to the work of Chen *et al.* [21], prediction by the proposed method can be carried out concurrently in different slices, and its computational efficiency depends on the multithreading degree of the program and the computer hardware. Furthermore, it is not necessary for the proposed method to confirm the ROI for the prediction slice by slice. A comparison of the proposed method with those in the literature tested with different datasets is reported in Table 2, where the values indicate that the proposed method yielded more accurate results in terms of MSD, HD, and DC.

Fig. 8 presents an example of a three-dimensional (3D) rendering of the two-dimensional (2D) prediction results in Fig. 7. The slices are overlaid in the original sequence that corresponds to the natural esophageal morphology. The vertical scale is different from that in the slice plane before volume rendering. As illustrated in Fig. 8, each pixel represents an area of 0.79 mm $\times$ 0.79 mm, and the thickness of a slice is 5 mm. Therefore, the quantity of slices was augmented by interpolation to ensure the uniformity of the scales in different orientations. To observe the internal structure, the opacity index of the esophageal surface was adjusted to an adequate value. The 3D rendering was programmed

TABLE 2. Comparison of the proposed method with the other methods published in literature.

Authors	MSD (mm)	HD (mm)	DC (%)	Dimension	DataSet (Number of scans)
Rousson *et al.* [13]	Unknown	Unknown	80.25	2D	Unknown (20 images)
Fieselmann *et al.* [15]	Unknown	Unknown	80	2D	8
Feulner *et al.* [16]	2.3 ± 1.6	14.5	Unknown	2D	117
Kurugol *et al.* [17]	2.1 ± 1.9	Unknown	Unknown	3D	8
Yang *et al.* [18]	2.0 ± 0.4	14.5 ± 4.1	73 ± 6	2D	30
Hao *et al.* [19]	18	Unknown	76	2D	87
Trullo *et al.* [20]	Unknown	Unknown	72 ± 7	2D	30
Chen *et al.* [21]	Unknown	5.87 ± 9.91	79 ± 20	2D	6
Proposed method (InceptionV3)	1.01 ± 0.79	1.50 ± 1.37	86.80 ± 6.43	2D	17 (1397 images)

Some publications did not report certain evaluation metrics which are denoted as unknown in the table. Every method listed in this table was evaluated by different data set.

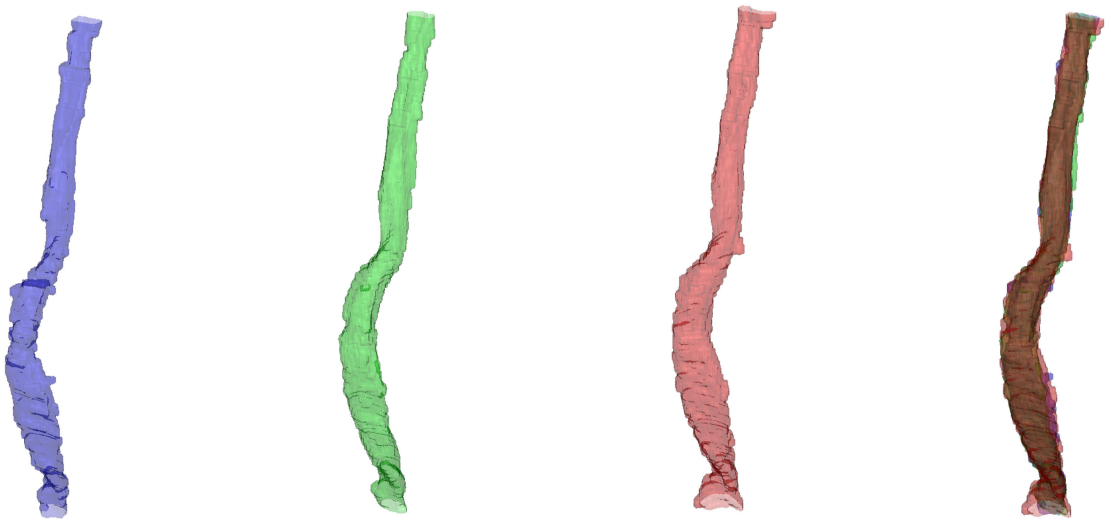

FIGURE 8. 3D rendering of the 2D slices of an integral esophagus in the test set. (a) represents the model generated by the semiautomatic segmentation method, (b) represents the prediction by the model, trained by the U-Net neural network with the InceptionV3 backbone, (c) represents manual labeling, and (d) represents a composite of (a), (b) and (c).

with the Visualization Toolkit library (Kitware Inc., Clifton Park, New York, USA) in the C++ language. This rendering visually reflects the performance of the proposed method, and has the potential to assist radiologists in observing the esophageal structures of the patients and making a relatively accurate diagnosis.

V. CONCLUSION

This article presents an esophagus segmentation approach that primarily relies on a semiautomatic labeling technique to produce annotations for training fully automated deep learning-based techniques. This approach enables the highly efficient generation of pixel-wise annotations and reduces the time required from an expert radiologist in producing annotated data. A standard evaluation using the MSD, HD, and DC indicates that the goodness of fit between the prediction and manual labeling in this study is superior to that in former studies. The proposed method thus has the potential to reduce the ionizing radiation-related hazard to organs at risk during the delineation of the gross target volume and provide guidance during image-guided radiation therapy.

For further research, we intend to increase the number of cases that will contribute to the precision of deep learning. However, we also have a schedule for exploiting a fully automatic labeling method to increase labeling efficiency.

REFERENCES

[1] A. K. Rustgi and H. B. El-Serag, "Esophageal carcinoma," *New England J. Med.*, vol. 371, no. 26, pp. 2499–2509, Dec. 2014.
[2] K. J. Napier, M. Scheerer, and S. Misra, "Esophageal cancer: A review of epidemiology, pathogenesis, staging workup and treatment modalities," *World J. Gastrointestinal Oncol.*, vol. 6, no. 5, pp. 112–120, May 2014.
[3] G. Holstege, G. Graveland, C. Bijker-Biemond, and I. Schuddeboom, "Location of motoneurons innervating soft palate, pharynx and upper Esophagus. Anatomical evidence for a possible swallowing center in the pontine reticular formation," *Brain, Behav. Evol.*, vol. 23, nos. 1–2, pp. 47–62, 1983.

[4] G. F. Hatch, III, L. Wertheimer-Hatch, K. F. Hatch, G. B. Davis, D. K. Blanchard, R. S. Foster, Jr., and J. E. Skandalakis, "Tumors of the esophagus," *World J. Surg.*, vol. 24, no. 4, pp. 401–411, Apr. 2000.

[5] E. E. Daniel, "Lower esophagus: Structure and function," in *Sphincters: Normal Function-Changes in Disease*. Boca Raton, FL, USA: CRC Press, 1992, pp. 49–66.

[6] M. Hashizume, S. Kitano, K. Sugimachi, and K. Sueishi, "Three-dimensional view of the vascular structure of the lower esophagus in clinical portal hypertension," *Hepatology*, vol. 8, no. 6, pp. 1482–1487, Nov. 1988.

[7] A. Makropoulos, S. J. Counsell, and D. Rueckert, "A review on automatic fetal and neonatal brain MRI segmentation," *NeuroImage*, vol. 170, pp. 231–248, Apr. 2018.

[8] A. Işın, C. Direkoğlu, and M. Şah, "Review of MRI-based brain tumor image segmentation using deep learning methods," *Procedia Comput. Sci.*, vol. 102, pp. 317–324, Jan. 2016.

[9] D. Poornima and A. G. Karegowda, "A review of image segmentation techniques applied to medical images," *Int. J. Data Mining Emerg. Technol.*, vol. 8, no. 1, pp. 78–94, Jul. 2018.

[10] J. E. Iglesias and M. R. Sabuncu, "Multi-atlas segmentation of biomedical images: A survey," *Med. Image Anal.*, vol. 24, no. 1, pp. 205–219, Aug. 2015.

[11] N. M. Zaitoun and M. J. Aqel, "Survey on image segmentation techniques," *Procedia Comput. Sci.*, vol. 65, pp. 797–806, Jan. 2015.

[12] M. A. M. Salem, A. Atef, A. Salah, and M. Shams, "Recent survey on medical image segmentation," in *Handbook of Research on Machine Learning Innovations and Trends*. Hershey, PA, USA: IGI Global, 2017, pp. 424–464.

[13] M. Rousson, Y. Bai, C. Y. Xu, and F. Sauer, "Probabilistic minimal path for automated esophagus segmentation," *Proc. SPIE*, vol. 6144, Mar. 2006, Art. no. 614449.

[14] T.-C. Huang, G. Zhang, T. Guerrero, G. Starkschall, K.-P. Lin, and K. Forster, "Semi-automated CT segmentation using optic flow and Fourier interpolation techniques," *Comput. Methods Programs Biomed.*, vol. 84, nos. 2–3, pp. 124–134, Dec. 2006.

[15] A. Fieselmann, S. Lautenschlager, F. Deinzer, M. John, and B. Poppe, "Esophagus segmentation by spatially-constrained shape interpolation," in *Bildverarbeitung für die Medizin*, Berlin, Germany: Springer, 2008, pp. 247–251.

[16] J. Feulner, S. K. Zhou, A. Cavallaro, S. Seifert, J. Hornegger, and D. Comaniciu, "Fast automatic segmentation of the esophagus from 3d ct data using a probabilistic model," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Berlin, Germany: Springer, 2009, pp. 255–262.

[17] S. Kurugol, E. Bas, D. Erdogmus, J. G. Dy, G. C. Sharp, and D. H. Brooks, "Centerline extraction with principal curve tracing to improve 3D level set esophagus segmentation in CT images," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2011, pp. 3403–3406.

[18] J. Yang, B. Haas, R. Fang, B. M. Beadle, A. S. Garden, Z. Liao, L. Zhang, P. Balter, and L. Court, "Atlas ranking and selection for automatic segmentation of the esophagus from CT scans," *Phys. Med. Biol.*, vol. 62, no. 23, pp. 9140–9158, Nov. 2017.

[19] Z. Hao, J. Liu, and J. Liu, "Esophagus tumor segmentation using fully convolutional neural network and graph cut," in *Proc. Chin. Intell. Syst. Conf.* Singapore: Springer, 2017, pp. 413–420.

[20] R. Trullo, C. Petitjean, D. Nie, D. Shen, and S. Ruan, "Fully automated esophagus segmentation with a hierarchical deep learning approach," *Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA)*, Sep. 2017, pp. 503–506.

[21] S. Chen, H. Yang, J. Fu, W. Mei, S. Ren, Y. Liu, Z. Zhu, L. Liu, H. Li, and H. Chen, "U-net plus: Deep semantic segmentation for esophagus and esophageal cancer in computed tomography images," *IEEE Access*, vol. 7, pp. 82867–82877, 2019.

[22] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[23] M. Zhang, X. Li, M. Xu, and Q. Li, "Image segmentation and classification for sickle cell disease using deformable U-Net," 2017, *arXiv:1710.08149*. [Online]. Available: http://arxiv.org/abs/1710.08149

[24] B. S. Lin, K. Michael, S. Kalra, and H. R. Tizhoosh, "Skin lesion segmentation: U-Nets versus clustering," Sep. 2017, *arXiv:1710.01248*. [Online]. Available: http://arxiv.org/abs/1710.01248

[25] M. Salem, S. Valverde, M. Cabezas, D. Pareto, A. Oliver, J. Salvi, A. Rovira, and X. Llado, "Multiple sclerosis lesion synthesis in MRI using an encoder-decoder U-NET," *IEEE Access*, vol. 7, pp. 25171–25184, 2019.

[26] Y. Weng, T. Zhou, Y. Li, and X. Qiu, "NAS-Unet: Neural architecture search for medical image segmentation," *IEEE Access*, vol. 7, pp. 44247–44257, 2019.

[27] E. Nowak, F. Jurie, and B. Triggs, "Sampling strategies for bag-of-features image classification," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2006, pp. 490–503.

[28] H. Jégou, M. Douze, and C. Schmid, "Improving Bag-of-Features for large scale image search," *Int. J. Comput. Vis.*, vol. 87, no. 3, pp. 316–336, May 2010.

[29] S. Suzuki and K. Abe, "Topological structural analysis of digitized binary images by border following," *Comput. Vis., Graph., Image Process.*, vol. 30, no. 1, pp. 16–32, 1985.

[30] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, Jun. 2010.

[31] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2006, pp. 404–417.

[32] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.

[33] V. Caselles, F. Catté, T. Coll, and F. Dibos, "A geometric model for active contours in image processing," *Numerische Math.*, vol. 66, no. 1, pp. 1–31, 1993.

[34] R. Malladi, J. A. Sethian, and B. C. Vemuri, "Shape modeling with front propagation: A level set approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 2, pp. 158–175, Feb. 1995.

[35] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, Jun. 2015, pp. 3431–3440.

[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[37] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and F. F. Li, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, Jun. 2009, pp. 248–255.

[38] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[39] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1492–1500.

[40] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[41] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[42] C. Szegedy, S. Ioffe, and V. Vanhoucke, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-first AAAI Conf. Artif. Intell.*, 2017, pp. 1–12.

[43] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*. [Online]. Available: http://arxiv.org/abs/1704.04861

[44] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.

[45] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," 2019, *arXiv:1905.11946*. [Online]. Available: http://arxiv.org/abs/1905.11946

XIAO LOU received the B.Eng. degree in software engineering from Xidian University, Xi'an, China, in 2012, and the M.S. degree in physics from the Xi'an University of Technology, Xi'an, in 2015. He is currently pursuing the Ph.D. degree in computer science and technology with Southeast University, Nanjing, China.

From 2017 to 2019, he had an internship at the Shandong Cancer Hospital and Institute, Jinan, China. He was selected by the China Scholarship Council (CSC), in 2019, to be awarded a scholarship under the State Scholarship Fund to pursue studies at the University of Alberta, Edmonton, AB, Canada, as a Visiting Ph.D. Student for one year. He studies and conducts research at the Department of Radiology and Diagnostic Imaging, University of Alberta. His research interests include medical imaging, image processing, computer vision, and deep learning. He was a recipient of the National Scholarship from the Ministry of Education of China, in 2014. He received the title of Outstanding Graduate from the Xi'an University of Technology, in 2015.

YOUZHE ZHU received the M.B. degree in clinical medicine from Jining Medical University, Jining, China, in 2016, and the M.M. degree in oncology from the University of Jinan, Jinan, China, in 2019. She had an internship at the Affiliated Hospital of Jining Medical University, from 2015 to 2016, and the Shandong Cancer Hospital and Institute, from 2016 to 2019. She is currently a Radiologist with the Department of Radiation Oncology, The First Hospital of Lanzhou University, Lanzhou, China. Her research interests include esophageal carcinoma, breast carcinoma, and estimation of the efficacy of chemo-radiotherapy. She was awarded the Academic Scholarship by the University of Jinan, in 2017 and 2018.

KUMARADEVAN PUNITHAKUMAR (Senior Member, IEEE) received the B.Sc.Eng. degree (Hons.) in electronic and telecommunication engineering from the University of Moratuwa and the M.A.Sc. and Ph.D. degrees in electrical and computer engineering from McMaster University. From 2001 to 2002, he was an Instructor with the Department of Electronic and Telecommunication Engineering, University of Moratuwa. From 2008 to 2012, he was an Imaging Research Scientist with GE Healthcare, Canada. He is currently an Associate Professor (Research) with the Department of Radiology and Diagnostic Imaging, University of Alberta, and an Operational and Computational Director with the Servier Virtual Cardiac Centre, Mazankowski Alberta Heart Institute. His research interests include medical image analysis and visualization, machine learning, information fusion, object tracking, and nonlinear filtering. He was a recipient of the Industrial Research and Development Fellowship from the National Sciences and Engineering Research Council of Canada, in 2008, the GE Innovation Award, in 2009, and the Outstanding Associate Editor Award from IEEE Access, in 2018.

LAWRENCE H. LE (Member, IEEE) received the M.B.A. degree in finance and technology commercialization from the University of Alberta and the Ph.D. degree in earth physics from the University of Alberta, Edmonton, AB, Canada. He held a Natural Sciences and Engineering Research Council of Canada (NSERC) Postdoctoral Fellowship at the Schlumberger-Doll Research Laboratory, Ridgefield, CT, USA. He began his medical physics residency training at the Department of Radiology and Diagnostic Imaging (DRDI), University of Alberta, in 1994. He joined DRDI, University of Alberta, as a Member of Academic Staff, and Capital Health as a Clinical Medical Physicist, in 2000. He is currently a Clinical Professor, who leads the graduate program, with DRDI and the Edmonton Authorized Radiation Protection Agency within the Alberta Health Services. He is also a Senior Visiting Scholar with the State Key Laboratory of ASIC and System, Fudan University. He directs the Ultrasonic Bone Tissue Characterization and Imaging Group. He guides his students to use interdisciplinary methods to solve biomedical problems to improve patient care. His research interests include imaging, signal and image processing, simulation, inversion, and machine learning. He is a member of the American Association of Physicists in Medicine (AAPM) and the Canadian Organization of Medical Physicists (COMP).

BAOSHENG LI received the Ph.D. degree in biomedical engineering from Southeast University, in 2004. He worked as a Visiting Scholar at the Medical School, University of Maryland, from 1999 to 2000. He is currently a Distinguished Professor with the Shandong Cancer Hospital and Institute, Shandong First Medical University. His research interests include precision radiotherapy, radioimmunotherapy, and the application of organ motion analysis and functional imaging in radiotherapy for cancer.