

Received May 10, 2019, accepted June 13, 2019, date of publication June 19, 2019, date of current version July 10, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2923760

U-Net Plus: Deep Semantic Segmentation for Esophagus and Esophageal Cancer in Computed Tomography Images

SHUCHAO CHEN¹, HAN YANG², JIAWEN FU¹, WEIJIAN MEI², SHUAI REN¹, YIFEI LIU², ZHIHUA ZHU², LIZHI LIU², HAOJIANG LI², AND HONGBO CHEN¹

¹School of Life and Environmental Sciences, Guilin University of Electronic Technology, Guilin 541004, China

²State Key Laboratory of Oncology in South China, Sun Yat-sen University Cancer Center, Collaborative Innovation Center for Cancer Medicine, Guangzhou 510060, China

Corresponding authors: Haojiang Li (lihaoj@sysucc.org.cn) and Hongbo Chen (hongbochen@163.com)

This work was supported in part by the National Science Foundation of China under Grant 81760322, Grant 81460273, and Grant 81572652, in part by the Science and Technology Planning Project of Guangdong Province under Grant 2017B020226004, and in part by the Innovation Program for College Student under Grant 201810595010.

ABSTRACT The effective segmentation and 3-D rendering of the esophagus and esophageal cancer from the computed tomography (CT) images can assist doctors in diagnosing esophageal cancer. Irregular and vague boundary causes great difficulty in the segmentation of esophagus and esophageal cancer. In this paper, U-Net Plus is proposed to segment esophagus and esophageal cancer from a 2-D CT slice. In the new network architecture, two blocks are employed to enhance the feature extraction performance of complex abstract information, which can effectively resolve irregular and vague boundaries. A block is a skip connection operation that is similar to convolution. The architecture is trained through a dataset of 1924 slices from 10 CT scans and tested through 295 slices from 6 CT scans. The training and test datasets are expanded tenfold to simulate the segmentation of the 3-D CT image. Using the new framework, we report a 0.79 ± 0.20 dice value and 5.87 ± 9.91 Hausdorff distance. A semi-automatic scheme is then designed for the 3-D segmentation of the esophagus or esophageal cancer. The 3-D rendering of the esophagus or esophageal cancer is implemented to assist in the diagnosis of esophageal cancer.

INDEX TERMS Esophageal cancer, image segmentation, deep learning, computed tomography (CT).

I. INTRODUCTION

Eastern Asia and Eastern and Southern Africa show a high rate of esophageal cancer [1]. Esophageal cancer ranks seventh in terms of incidence and sixth in overall mortality [2]. Computed tomography (CT) is a very important diagnostic technique for esophageal cancer because CT images of chest, abdomen, and pelvis can be used to evaluate tumor metastasis to adjacent tissues or distant organs, such as liver and lymph nodes; this information is important for the diagnosis and prognosis of esophageal cancer [3], [4]. Accurately extracting the region of esophagus and esophageal cancer from CT images shows high clinical application value.

The automatic segmentation of esophagus and esophageal cancer is very difficult. The wall of the esophagus from the

lumen outward consists of mucosa, connective tissue, layers of muscle fibers between layers of fibrous tissue, and an outer layer of connective tissue. As shown in Fig. 1, the boundary of the esophagus or esophageal cancer is irregular and vague, and air holes occasionally exist in the esophagus. Even experienced medical experts cannot accurately delineate the boundary of esophagus or esophageal cancer through single-slice CT images.

Automating the contour extraction for esophagus has been attempted. Rousson *et al.* [5] described a method that combines a spatial prior of the esophagus centerline with a histogram-based appearance model. The method is semi-automatic and requires two manually placed points on the centerline and a segmentation of the left atrium and the aorta as input. Markov chain model, followed by a “detect and connect” approach, has been used to obtain the maximum a posteriori estimate of the approximate esophagus shape [6].

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

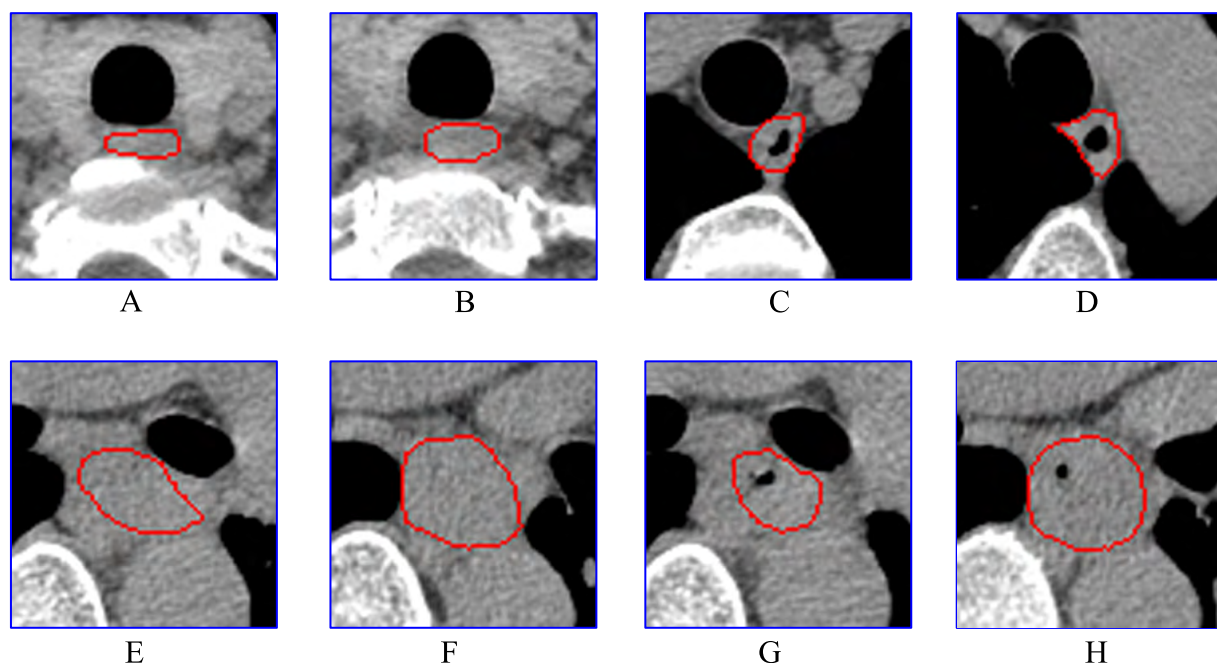


FIGURE 1. Esophagus and esophageal cancer in CT scans. The red curves are the border of the normal esophagus or the esophageal cancer. A & B the normal esophagus without air hole; C & D the normal esophagus with air holes; E & F the esophageal tumor without air hole; G & H the esophagus tumor with air holes.

Yang *et al.* [7] developed an online atlas selection approach to select a subset of optimal atlases for multi-atlas segmentation to automatically delineate the esophagus. The presence of air bubbles is an obstacle in the atlas selection method. The skeleton-shaped model has been used to guide 3-D esophagus segmentation. This method is sensitive to noise in the image, resulting in over-segmentation for the target object [8].

Deep learning has been recently used for the segmentation of normal esophagus or esophageal cancer. A convolutional neural network is trained to generate an esophagus probability map, and the active contour model and random walk are used to estimate the location of the esophagus [9]. Hao *et al.* [10] utilized fully convolutional network (FCN) to establish an esophagus–tumor classifier on the training dataset with expert-labeled tumor regions. Trullo *et al.* [11] changed the FCN architecture to improve the localization accuracy and performance for esophagus segmentation.

FCN is an end-to-end network that can effectively solve the problem of image semantics segmentation. The deep architecture achieves improved performance. However, the training error rate in a deep plain network is high because the gradient disappears easily in a deeper architecture [12]. LinkNet, SegNet, and U-Net have been developed on the basis of FCN. SegNet provides good performance with competitive inference time and provides the most efficient memory-wise inference [13]. SegNet is used in medical image segmentation, such as in the segmentation of lung field [14], cross-sectional brain MRI [15], and lumbar spinal stenosis [16]. LinkNet utilizes an information theoretical approach based on minimum description length to automatically adjust the number of regions and the temporal relationships among them [17]. U-Net has been developed on

the basis of FCN and considers skip connection between encoder and decoder. This process can effectively combine the features from shallow and deep layers through multipath confusion, which solves the spatial loss of feature map and improves the accuracy of semantic segmentation. U-Net has been applied in semantic medical image segmentation [18]–[23]. However, it has failed to produce satisfactory results because of the vague and irregular border of esophagus and esophageal cancer. To address the inefficiency of feature extraction of skip connection layer in U-Net network, we design U-Net Plus as the repeated process of U-Net feature extraction in the present study.

In summary, the main contributions of this study are as follows.

First, the network architecture of two encoder–decoder blocks with skip connection is adopted to improve the capability of complex abstract information processing. This characteristic helps improve the segmentation capability of irregular and vague boundary of esophageal or esophageal cancer.

Second, a semi-automatic scheme is designed to segment the esophagus or esophageal cancer from 3-D CT images. In this scheme, only few image parameters must be provided.

Third, 3-D rendering esophagus or esophageal cancer will help doctors diagnose esophageal cancer.

II. PROPOSED METHOD

In this section, the proposed method is described in detail. As shown in Fig. 2, the new method involves two stages. The first stage is training. The images in the training set are the input of the network after preprocessing. Data augmentation is implemented to enhance robustness of the network.

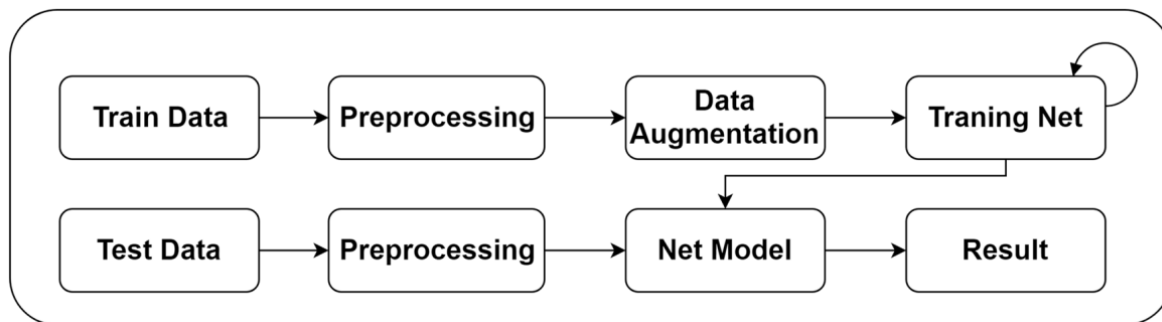


FIGURE 2. Flowchart of proposed method for the segmentation of esophagus and esophageal cancer.

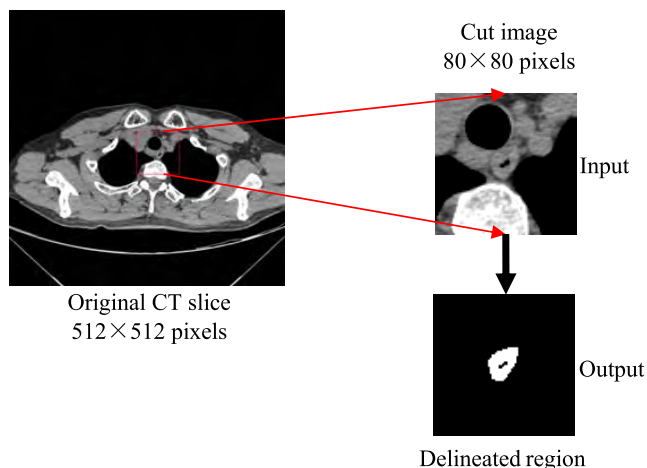


FIGURE 3. Description of input and output for the network.

The network is trained several times to obtain the best parameters. In the testing stage, the preprocessed test images are used to test the performance of the network.

A. IMAGE PREPROCESSING

The window center and window width in digital imaging and communications in medicine (DICOM) information are used to adjust the CT image with DICOM format to bitmap format. As shown in Fig. 3, a rectangular area of 80×80 pixels, which includes esophagus or esophageal cancer, is selected as input to the network for each CT slice. The delineated region of esophagus or esophageal cancer is considered as the desired output of the network.

B. ARCHITECTURE OF U-NET PLUS

The network architecture of U-Net Plus is illustrated in Fig. 4. The encoder–decoder architecture is the main structure of many semantic segmentation networks and performs well in pixel-level classification. U-Net Plus uses two blocks of encoder–decoder architecture to enhance the capability of feature extraction.

U-Net Plus adopts the FCN mode and is an end-to-end network architecture. The convolutional operation is applied to extract the features. All convolutions involve one step in

all directions. All convolutions in all layers, except for the final layer, comprise 64 channels, and the rectified linear unit is connected [24]. In the last layer, the convolutional size is 1×1 , whereas the channel involves two, which mainly maps the component features to the required number of classes. The 1×1 convolution can be used to improve the robustness of the network and is used in deep networks [25], [26].

Down-sampling is carried out by the max pool operation. The size of the pooling filter is 2×2 . The plane and channel involve two and one steps, respectively. After down-sampling, the receptive field of the convolution is increased, which contributes to obtaining the global features of the image and expressing abstract information.

Up-sampling is carried out by deconvolution. The size of deconvolution is 3×3 , and the plane and channel involve two and one steps, respectively. The deconvolution aims to match the size of the feature map after down-sampling with the size of the current feature map. Up-sampling conveys information from a large field to a small field of view, allowing information in different fields to communicate with one another.

The skip connection links the corresponding down-sampling and up-sampling feature maps. This process can solve the problem of spatial information loss caused by down-sampling and improves focus on convolution feature extraction.

A block in the network architecture includes skip connection and other operations, such as convolution, pool operation, and deconvolution. Although numerous feature extraction operations are performed in block, feature extraction after skip connection must be based on previous input information obtained after skip connection. Think of skip connection and operations between it as skip connection operation. Skip connection operation extracts the features of the input information, and the generated output information must be connected to the input information before it can be used as the input for the next operation.

The skip connection combines the feature information of small field of view and large field of view, which can substantially use the global features. However, the feature maps of the small field of view are from very early features. Even after many convolutions, the features of the large field of view still have strong correspondence with the features of the

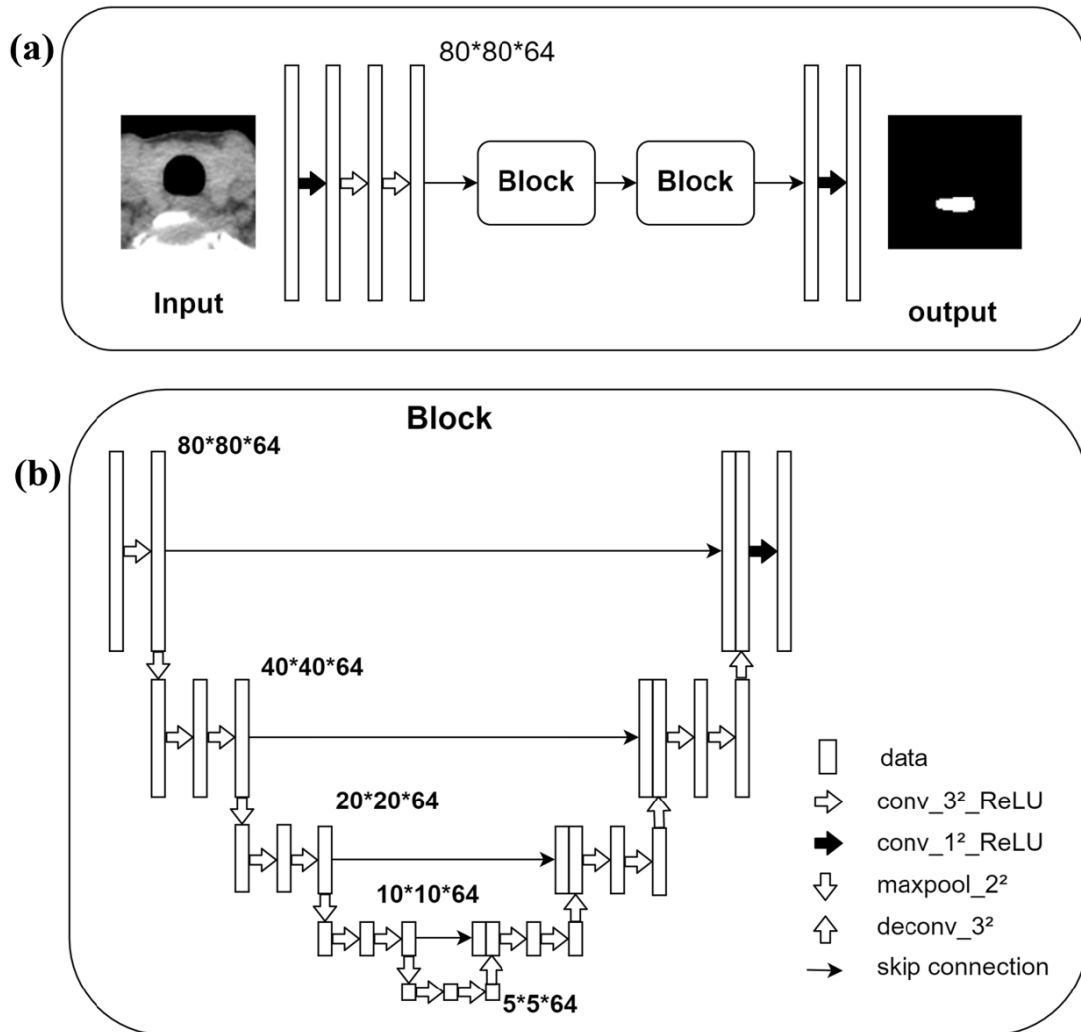


FIGURE 4. Architecture of U-Net Plus. (a) Main architecture of U-Net Plus. (b) Detailed architecture of the block. conv: convolution. ReLU: rectified linear unit.

small field of view. This process is similar to a convolution operation. The convolution operation only extracts features from its input information. A block is a skip connection operation which is similar to convolution, but its feature extraction capability is much stronger than convolution. Two blocks are used to enhance the feature extraction performance of complex abstract information.

C. TRAINING AND TESTING

The preprocessed esophageal cancer image is used as the direct input information of the network, and the corresponding label image is used as the supervision of the network. The label image is a binary image. The targets are the esophageal and esophageal cancer area, which are the foreground in the label image and the background in other parts.

The energy function SoftMax is used to obtain the probability distribution map; and the cross-entropy loss

function [27] and Adam optimizer [28] are used to train the network.

The size of the input image and the probability map are the same, and their positions on the pixels correspond one by one because the network fills the boundary when convoluting.

The model fits rapidly on some data but slowly on others. If all data are used for each iteration in training, the model may show over-fitting on some features. To reduce over-fitting, we raise the threshold of the loss value and perform training only when the loss value of a batch of data is greater than the threshold value. The initial threshold is set to 0.04. The weight parameters of the trained model are saved for testing.

III. EXPERIMENTS

A. DATASET

Experimental CT scans are collected from SUN YAT-SEN University Cancer Center. A total of 16 thoracic CT scans

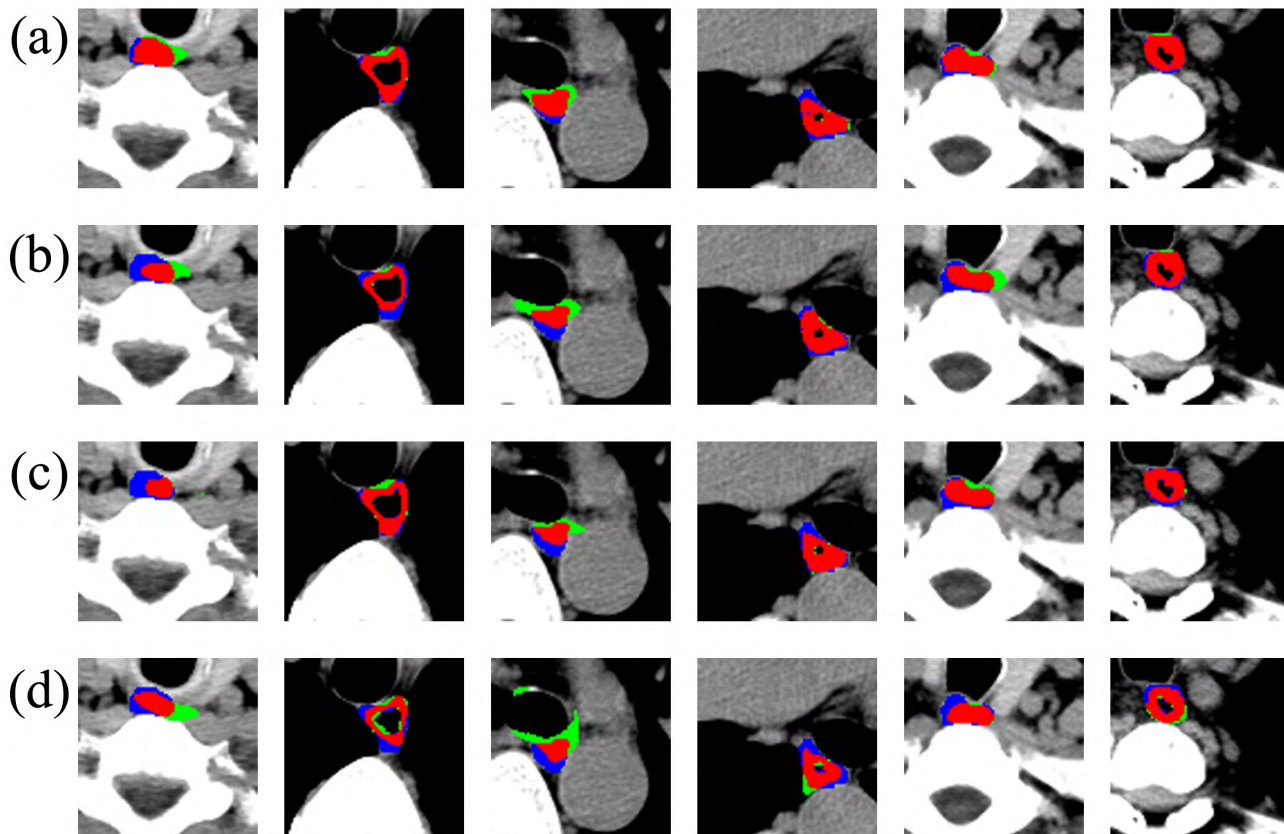


FIGURE 5. Visualization of segmentation results for normal esophagus. The results of (a) U-Net Plus, (b) U-Net, (c) LinkNet, and (d) SegNet. Red: overlapped region between the delineated region and segmented region (true positive). Blue: region in the delineated region that is not in the segmented region (false negative). Green: region in the segmented region that is not in delineated region (false positive).

from patients diagnosed with esophageal carcinoma are used for experimental esophagus segmentation. The imaging parameters are as follows: field of view = 376 mm × 376 mm and matrix = 512 × 512, slice thickness = 1 mm, and bits stored = 12. Two doctors with extensive experience in imaging diagnosis of esophageal cancer delineate the location of esophageal cancer and esophagus in all CT images as the region of interest (ROI). A total of 2,219 slices are labeled manually.

Among the 16 CT scans, 10 are selected as training set, and 6 are selected as test set. The esophagus or esophageal cancer areas in the CT images in training set CT scans are delineated. A total of 295 slices in test set are also delineated.

The data are augmented to improve the robustness of the segmentation algorithm. The region of esophagus or esophageal cancer does not always occur at the center of the selected rectangle but may be located anywhere in the rectangle. Therefore, 10 samples are used for one CT slice to improve the adaptability of the network. For one CT slice, all ROIs of 10 samples are located inside of the rectangle, and the centers of the ROI do not overlap with the center of the rectangle but randomly locate at different positions in the rectangle. Data augmentation is implemented on the training and testing data. Thus, the training and the test samples are expanded tenfold, that is, 19,240 samples in the training set and 2,950 samples in the test set.

B. EVALUATION METRIC

The following two pixel-level measurements are employed to compare the segmentation performance of the proposed approach with those of other methods: dice value (DV) and Hausdorff distance (HD) [29].

DV is defined as:

$$DV(A_S, A_T) = \frac{2A_{TS}}{A_S + A_T}$$

where A_T represents the target area, which is delineated manually by the doctors. A_S represents the segmented automatically area of esophagus or esophageal cancer, which is the result of the model testing. A_{TS} represents the overlap between A_T and A_S . DV measures the overlapping degree between target area and segmented area, and it always ranges between [0, 1]. The larger the DV, the higher is the consistency between manual and automated segmentation.

Another evaluation metric is HD, which is expressed as:

$$HD(A_T, A_S) = \max(\max_{P \in S} d(P, T), \max_{P' \in T} d(P', S))$$

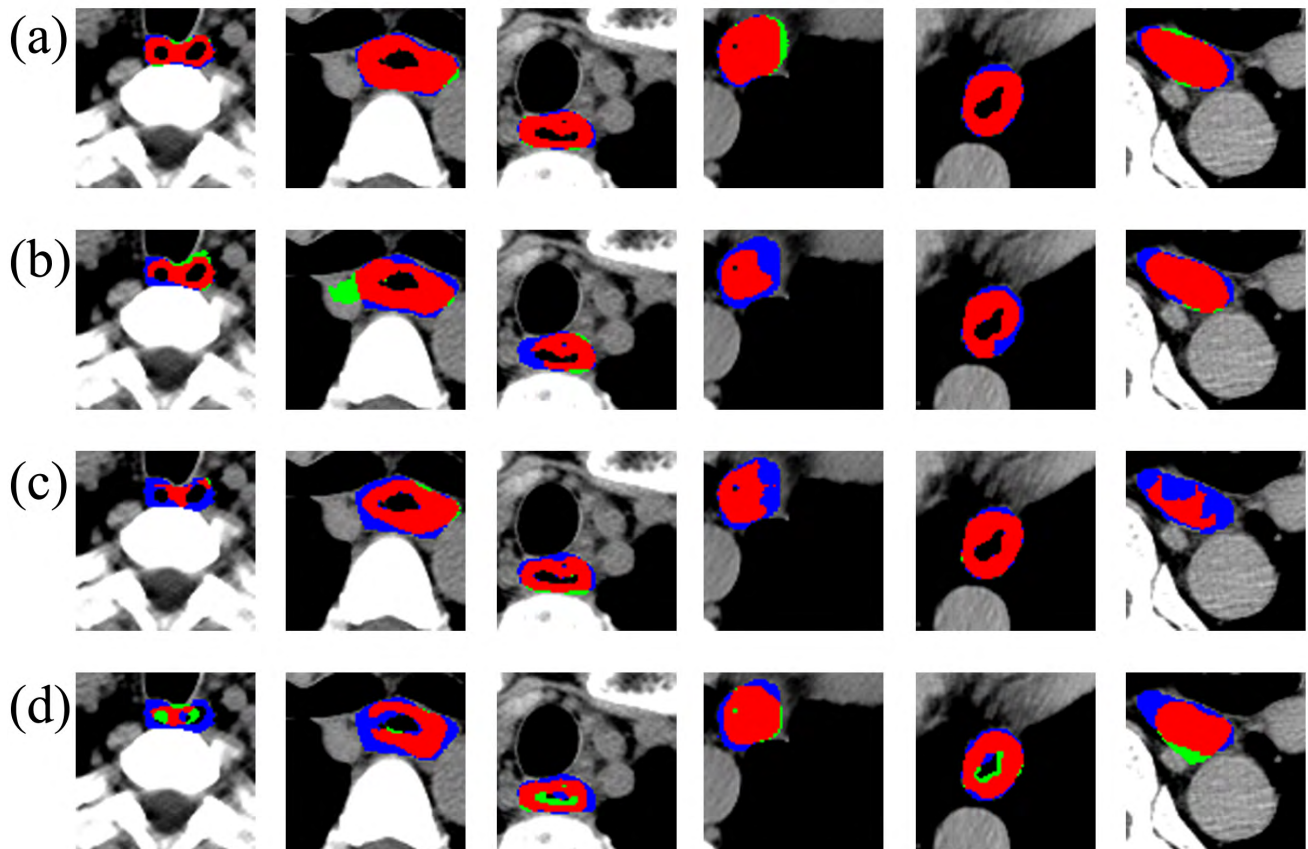
where

$$d(P, T) = \min_{P1 \in T} ||P - P1||$$

which is defined as the minimum distance from a point (P) on the segmented contour (S) to the nearest point ($P1$) of

TABLE 1. Comparison between the u-net plus and several popular networks.

	U-Net Plus	U-Net	LinkNet	SegNet
DV	0.79±0.20	0.72±0.19	0.66±0.29	0.61±0.24
HD(mm)	5.87± 9.91	7.04± 7.45	11.12± 19.74	14.45±27.00

**FIGURE 6.** Visualization of segmentation result for esophagus cancer. (a) The result of U-Net Plus; (b) the result of U-Net; (c) the result of LinkNet; (d) the result of SegNet. Red: overlapped region between the delineated region and segmented region (true positive); Blue: region in delineated region not in segmented region (false negative); Green: region in segmented region not in delineated region (false positive).

the target contour (T). HD measures how far two regions of a metric space are from each other. HD is the maximum of the two values and is always at $[0, \infty]$. As HD increases, the performance degrades.

C. EXPERIMENTAL PROCESS

Numerous experiments are carried out to demonstrate the effectiveness and feasibility of U-Net Plus from three aspects.

First, the performance of U-Net Plus is evaluated on our augmented dataset of 2,950 CT 2-D images for esophagus and esophageal cancer. The result of U-Net Plus is compared with that of U-Net, LinkNet, and SegNet.

Second, the performance of the proposed method is evaluated by comparing the results of segmentation for esophagus or esophageal cancer from existing literature.

Finally, a 3-D rendering experiment for esophagus and esophageal cancer is performed, and a semi-automatic segmentation scheme is designed.

IV. RESULTS AND ANALYSIS

A. SEGMENTATION RESULTS

Figs. 5 and 6 illustrate the segmentation results of six representative CT images of esophagus and six representative CT images of esophageal cancer for different methods, respectively. The delineated region and the segmented region are presented together to intuitively observe the segmentation effect. The red area represents the overlapped region between the delineated and segmented regions. The blue area represents the delineated region that is not located in the segmented region. The green area represents the region in

TABLE 2. Performance comparison with existing methods.

	Description of method	DV	HD (mm)
Feulner et al. [6]	Markov chain model	0.74± 0.14	7.19±0.78
Yang et al. [7]	Atlas selection	0.73 ± 0.06	14.5 ± 4.1
Grosgeorge et al. [8]	Skeleton-shaped model	0.78 ± 0.03	Unknown
Fechter et al. [9]	FCN and random walk	0.76± 0.09	13.63± 7.51
Hao et al. [10]	FCN	0.76	Unknown
Trullo et al. [11]	Modified FCN	0.72 ± 0.07	Unknown
Proposed method	U-Net Plus	0.79±0.20	5.87+9.91

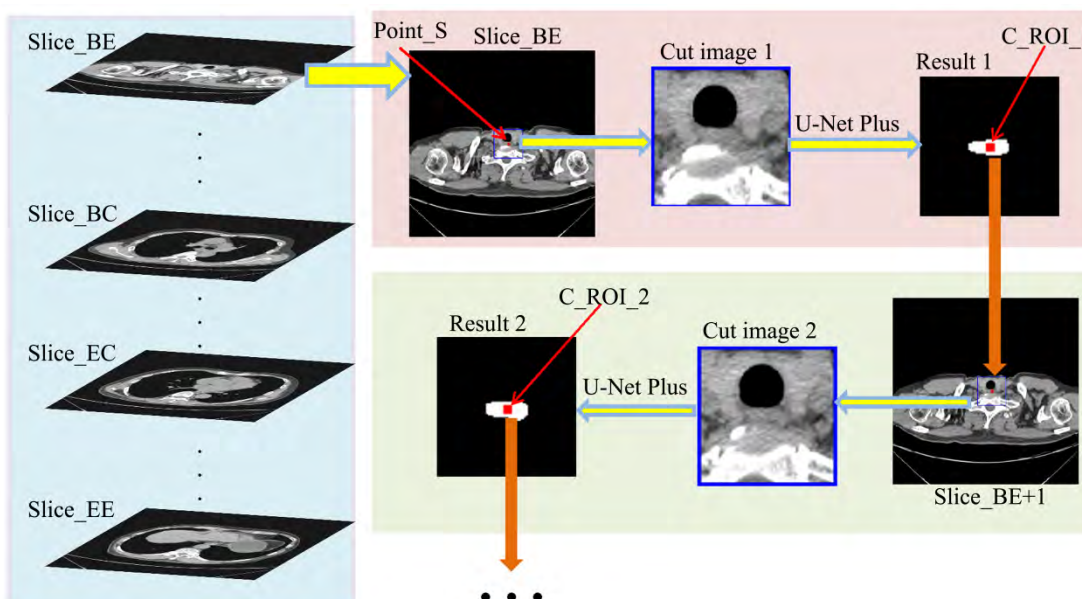


FIGURE 7. Scheme of semi-automatic segmentation for 3-D esophagus and esophageal cancer. Slice_BE and Slice_EE: begin and end slice of esophagus, Slice_BC and Slice_EC: begin and slice of esophageal cancer, Point_S: delineated start point for segmentation, and C_ROI: center of segmented ROI.

the segmented region that is not located in the delineated region.

As shown in Figs. 5 and 6, the best and worst segmentation effect is achieved by the proposed method and SegNet, respectively. U-Net Plus can achieve satisfactory segmentation results in the presence or absence of holes. The region of esophagus cancer with two holes can also be segmented by the proposed method but not by the other methods. The U-Net Plus results show the largest intersection of segmented and manual region and the highest shape consistency. The visualization results clearly illustrate that the proposed method can extract the region of esophagus or esophageal cancer from CT images.

Table 1 shows the average segmentation results for 2,950 test samples. U-Net Plus achieves the best segmentation performance with the highest DV (0.79) and lowest

HD (5.87 mm), followed by U-NET and LinkNet. SegNet obtains the lowest DV (0.61) and highest HD (14.45 mm). The segmentation result of U-Net Plus shows the highest consistency on the delineated region and the nearest boundary of result region to that of the delineated region.

The effectiveness and robustness of the proposed network are discussed from two aspects. First, the rectangular areas of esophagus or esophageal cancer in CT image are cut out as input features of the network, greatly reducing the effect of other tissues and organs of the human chest on target region segmentation. Second, the benefit from the optimized network architecture is discussed. In the new network architecture, we implement two U-Net feature extraction processes that improve the efficiency and capability of feature extraction. Consequently, the proposed network realizes accurate

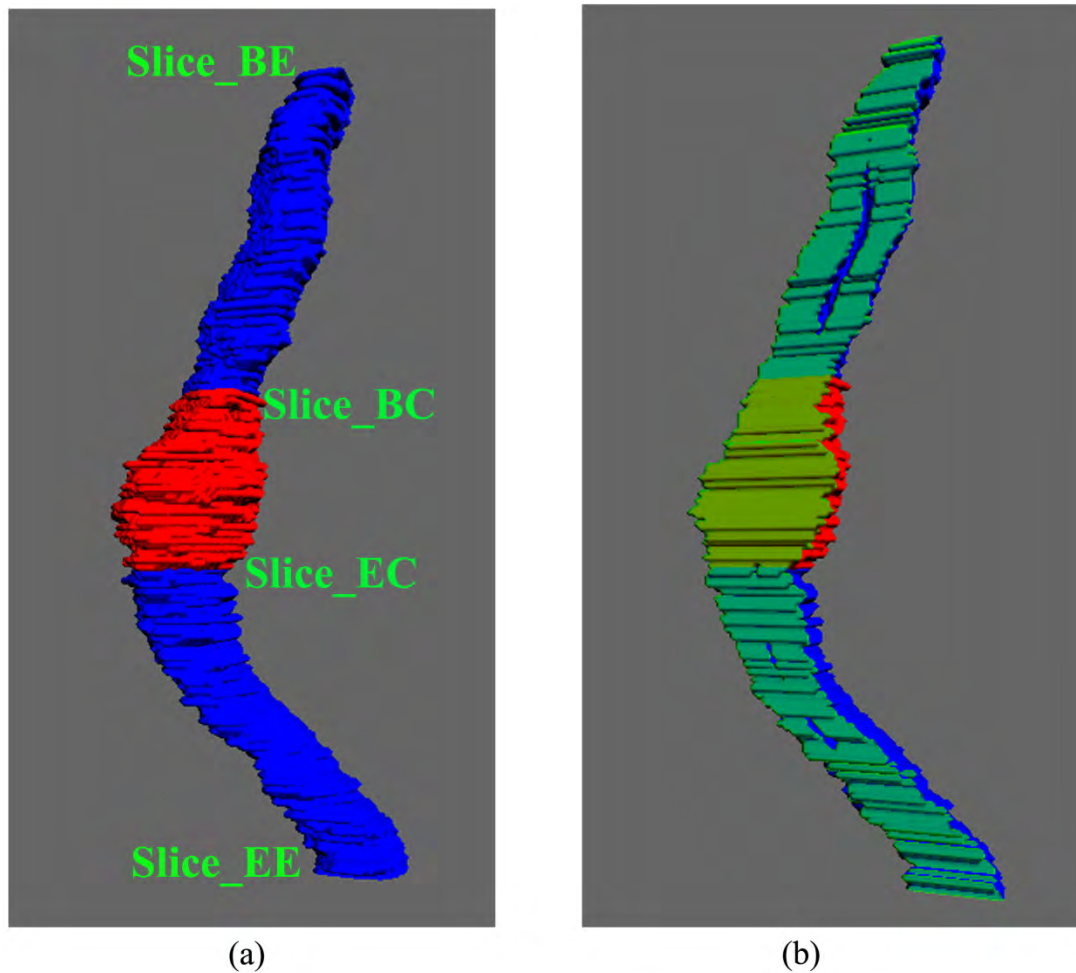


FIGURE 8. 3-D rendering for esophagus and esophageal cancer. (a) 3-D rendering for outer esophagus and esophageal cancer, (b) 3-D rendering for inner esophagus and esophageal cancer. Slice_BE and Slice_EE: begin and end slice of esophagus, Slice_BC and Slice_EC: begin and slice of esophageal cancer.

and robust segmentation for esophagus or esophageal cancer in thoracic CT images.

B. PERFORMANCE COMPARISON WITH EXISTING METHODS

The methods and effects of segmentation for esophagus or esophageal cancer in CT images have been previously reported. The performances of previously reported segmentation methods are summarized in Table 2. The DV metric is employed to evaluate the performance of segmentation in all of these methods. Three groups also evaluated segmentation performance using HD. The proposed method achieves the best performance, highest DV, and lowest HD.

V. SEMI-AUTOMATIC 3-D SEGMENTATION

In the previous section, we proposed a method to segment esophageal or esophageal cancer in 2-D CT images. However, the structure of esophagus or esophageal cancer is 3-D. The diagnosis of esophageal cancer must be supported

by segmenting the esophagus or esophageal cancer from 3-D CT images. As shown in Fig. 7, a semi-automatic segmentation scheme is designed for 3-D esophagus or esophageal cancer in CT images. In the semi-automatic segmentation method, the following five parameters must be manually provided before image segmentation: begin and end slices of esophagus, begin and end slices of esophageal cancer, and start point for image segmentation. The parameter of start point helps in image segmentation, and the other parameters assist in the 3-D rendering of esophagus and esophageal cancer.

This method assumes that in thoracic CT acquisition, slice thickness is sufficiently small such that the center of the upper esophagus is in the next esophageal region. This condition is met in most cases. Therefore, the feature image of each slice for segmentation can be obtained through the center of esophagus in the upper slice. As shown in Fig. 7, C_ROI_1 is the center of segmented ROI in a slice, and the rectangular region can be defined by C_ROI_1. The main steps to segment the 3-D esophagus and esophageal cancer are as follows:

Step 1: A rectangle region of 80×80 pixels centered on the previously given start point is extracted.

Step 2: The cut CT image in the extracted rectangle region is used as the feature input of the trained U-Net Plus network. The output of U-Net Plus is the segmented region of esophagus or esophageal cancer in the given slice.

Step 3: If the slice is the end slice of esophagus, then the process of image segmentation is finished. Otherwise, the center of the segmented region is extracted and mapped to the next slice.

Step 4: In the new slice, another rectangular region located on the center of the segmented region of the previous slice is redefined. Step 2 is then performed.

For a CT slice, the region and center of segmented esophagus or esophageal cancer are saved for 3-D rendering.

VI. 3-D RENDERING

2-D segmented results are stacked before performing rendering operation. All slices between begin slice and end one are used for stacking. An empty space is first created for the 3-D model. Each pixel on 2-D segmented result image is then transferred to the empty space. The pixel's X and Y coordinates are adjusted to the center of segmented esophagus or esophageal cancer. The slice number with respect to the distance between each slice is considered as the Z coordinate. The stacking process is repeated until all slices are completed. All voxels in the 3-D space are connected as in the 2-D slices.

Fig. 8 shows the 3-D rendering for esophagus or esophageal cancer. Begin and end slices of esophagus and esophageal cancer are defined by the doctor. As shown in Fig. 8A, normal esophagus is shown in blue and esophageal cancer in red. The 3-D images can be rotated arbitrarily. In addition, the esophagus or esophageal cancer can be cut and displayed according to the center information of the segmentation region (Fig. 8B). Doctors can clearly observe the internal condition of esophageal or esophageal cancer and receive further assistance in esophageal cancer diagnosis.

VII. CONCLUSION

The effective segmentation and 3-D rendering of esophagus and esophageal cancer can assist doctors in diagnosing esophageal cancer. The irregular and vague boundary causes difficulty in the effective segmentation of esophagus and esophageal cancer. In this study, a U-Net Plus network, which is characterized by high feature extraction capability, is proposed for the segmentation of esophagus or esophageal cancer. The result of the proposed method is better than that of other method for semantic segmentation based on deep learning and that of the existing literature by the evaluation of DV and HD. A semi-automatic scheme is designed for the 3-D segmentation of esophagus or esophageal cancer. The 3-D rendering of esophagus or esophageal cancer is implemented to assist in the diagnosis of esophageal cancer.

Future works must focus on three aspects. First, the architecture of U-Net Plus must be modified to improve the capability of feature extraction for objects with irregular and

vague boundaries. Second, the training set and testing set must be expanded to further improve the performance of the network for esophagus or esophageal cancer segmentation. Finally, a method for automatic recognition of esophageal and esophageal cancer to reduce manual intervention and improve the automation of assistant diagnosis system for esophageal cancer must be developed.

ACKNOWLEDGMENT

The authors would like to thank EnPapers for polishing and modifying the language of this paper. (*Shuchao Chen and Han Yang contributed to this work equally.*)

REFERENCES

- [1] B. Gupta and N. Kumar, "Worldwide incidence, mortality and time trends for cancer of the oesophagus," *Eur. J. Cancer Prevention*, vol. 26, no. 2, pp. 107–118, 2017.
- [2] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *Cancer J. Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.
- [3] N. Kim, H. Cho, M. Yun, K. R. Park, and C. G. Lee, "Prognostic values of mid-radiotherapy ^{18}F -FDG PET/CT in patients with esophageal cancer," *Radiat. Oncol.*, vol. 14, no. 1, pp. 14–27, 2019.
- [4] R. T. H. M. Larue, R. Klaassen, A. Jochems, R. T. H. Leijenaar, M. C. C. M. Hulshof, M. I. Van Berge Henegouwen, W. M. J. Schreurs, M. N. Sosef, W. Van Elmpt, H. W. M. Van Laarhoven, and P. Lambin, "Pre-treatment CT radiomics to predict 3-year overall survival following chemoradiotherapy of esophageal cancer," *Acta Oncol.*, vol. 57, no. 11, pp. 1475–1481, 2018.
- [5] M. Rousson, Y. Bai, C. Xu, and F. Sauer, "Probabilistic minimal path for automated esophagus segmentation," *Proc. SPIE*, vol. 6144, Mar. 2006, Art. no. 614449.
- [6] J. Feulner, S. K. Zhou, M. Hammon, S. Seifert, M. Huber, D. Comaniciu, J. Hornegger, and A. Cavallaro, "A probabilistic model for automatic segmentation of the esophagus in 3-D CT scans," *IEEE Trans. Med. Imag.*, vol. 30, no. 6, pp. 1252–1264, Jun. 2011.
- [7] 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, 2017.
- [8] D. Grosgeorge, C. Petitjean, B. Dubray, and S. Ruan, "Esophagus segmentation from 3D CT data using skeleton prior-based graph cut," *Comput. Methods Med.*, vol. 2013, Jul. 2013, Art. no. 547897.
- [9] T. Fechter, S. Adebahr, D. Baltas, I. B. Ayed, C. Desrosiers, and J. Dolz, "A 3D fully convolutional neural network and a random walker to segment the esophagus in CT," Apr. 2017, *arXiv:1704.06544*. [Online]. Available: <https://arxiv.org/abs/1704.06544>
- [10] 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.
- [11] R. Trullo, C. Petitjean, D. Nie, D. Shen, and S. Ruan, "Fully automated esophagus segmentation with a hierarchical deep learning approach," in *Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA)*, Sep. 2017, pp. 503–506.
- [12] Q. Zhang, Z. Cui, X. Niu, S. Geng, and Y. Qiao, "Image segmentation with pyramid dilated convolution based on ResNet and U-Net," in *Proc. Int. Conf. Neural Inf. Process.*, Springer, 2017, pp. 364–372.
- [13] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [14] A. Mittal, R. Hooda, and S. Sofat, "LF-SegNet: A fully convolutional encoder-decoder network for segmenting lung fields from chest radiographs," *Wireless Pers. Commun.*, vol. 101, no. 1, pp. 511–529, 2018.
- [15] B. Khagi and G. R. Kwon, "Pixel-label-based segmentation of cross-sectional brain MRI using simplified SegNet architecture-based CNN," *J. Healthcare Eng.*, vol. 2018, Oct. 2018, Art. no. 3640705.

- [16] A. S. Al-Kafri, S. Sudirman, A. Hussain, D. Al-Jumeily, F. Natalia, H. Meidia, N. Afriliana, W. Al-Rashdan, M. Bashtawi, M. Al-Jumaily, "Boundary delineation of MRI images for lumbar spinal stenosis detection through semantic segmentation using deep neural networks," *IEEE Access*, vol. 7, pp. 43487–43501, 2019.
- [17] A. Chaurasia and E. Culurciello, "LinkNet: Exploiting encoder representations for efficient semantic segmentation," Jun. 2017, *arXiv:1707.03718*. [Online]. Available: <https://arxiv.org/abs/1707.03718>
- [18] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," May 2015, *arXiv:1505.04597*. [Online]. Available: <https://arxiv.org/abs/1505.04597>
- [19] M. Zhang, X. Li, M. Xu, and Q. Li, "Image segmentation and classification for sickle cell disease using deformable U-Net," Oct. 2017, *arXiv:1710.08149*. [Online]. Available: <https://arxiv.org/abs/1710.08149>
- [20] 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: <https://arxiv.org/abs/1710.01248>
- [21] M. Salem, S. Valverde, M. Cabezas, D. Pareto, A. Oliver, J. Salvi, Á. Rovira, X. Lladó, "Multiple sclerosis lesion synthesis in MRI using an encoder-decoder U-NET," *IEEE Access*, vol. 7, pp. 25171–25184, 2019.
- [22] 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.
- [23] L. Wang, C. Xie, and N. Zeng, "RP-Net: A 3D convolutional neural network for brain segmentation from magnetic resonance imaging," *IEEE Access*, vol. 7, pp. 39670–39679, 2019.
- [24] A. F. M. Agarap, "Deep learning using rectified linear units (ReLU)," Mar. 2018, *arXiv:1803.08375*. [Online]. Available: <https://arxiv.org/abs/1803.08375>
- [25] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," Sep. 2014, *arXiv:1409.4842*. [Online]. Available: <https://arxiv.org/abs/1409.4842>
- [26] M. Lin, Q. Chen, and S. Yan, "Network in network," Dec. 2013, *arXiv:1312.4400*. [Online]. Available: <https://arxiv.org/abs/1312.4400>
- [27] K. Hu, Z. Zhang, X. Niu, Y. Zhang, C. Cao, F. Xiao, and X. Gao, "Retinal vessel segmentation of color fundus images using multiscale convolutional neural network with an improved cross-entropy loss function," *Neurocomputing*, vol. 309, pp. 179–191, Oct. 2018.
- [28] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Dec. 2014, *arXiv:1412.6980*. [Online]. Available: <https://arxiv.org/abs/1412.6980>
- [29] X. Du, W. Zhang, H. Zhang, J. Chen, Y. Zhang, J. C. Warrington, G. Brahm, and S. Li, "Deep regression segmentation for cardiac Bi-ventricle MR images," *IEEE Access*, vol. 6, pp. 3828–3838, 2018.



JIAWEN FU is currently pursuing the B.Sc. degree in biomedical engineering with the School of Life and Environmental Sciences, Guilin University of Electronic Technology. His research interests include medical image analysis and medical pattern recognition.



WEIJIAN MEI is currently pursuing the M.D. degree with the Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China. His research interests include thoracic surgery and surgical oncology.



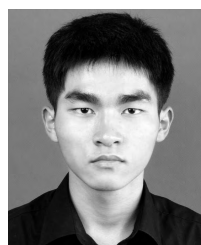
SHUAI REN is currently pursuing the M.Sc. degree in biomedical engineering with the School of Life and Environmental Sciences, Guilin University of Electronic Technology. His research interests include medical image analysis and machine learning.



YIFEI LIU received the M.D. degree in clinical medicine from Sun Yat-sen University, Guangzhou, China, in 2017. She is currently a Radiologist with the Sun Yat-sen University Cancer Center. Her research interests include medical imaging, medical image analysis, and machine learning.



ZHIHUA ZHU received the M.D. degree from Sun Yat-sen University. He trained as a Postdoctoral Fellow at Sun Yat-sen University. He is currently an Associate Professor with the State Key Laboratory of Oncology in South China, Cancer Center, Sun Yat-sen University. He has actively been engaged in the pre-clinical and clinical studies on chest tumors, including lung cancer, esophageal carcinoma, and breast cancer, with special interests in the palliative treatment of tumors and the individualized treatment of lung cancers.



SHUCHAO CHEN is currently pursuing the B.Sc. degree in biomedical engineering with the School of Life and Environmental Sciences, Guilin University of Electronic Technology. His research interests include computer vision and medical images analysis.



HAN YANG received the Ph.D. and M.D. degrees in clinical medicine from the Peking Union Medical College, China, in 2010. He is currently an Attender with Sun Yat-sen University Cancer Centre. His research interests include tumor growth, metastasis, and bioinformatics.



LIZHI LIU received the B.Sc. degree from North Sichuan Medical University, in 1996, and the M.Sc. and D.Sc. degrees from Sun Yat-sen University, in 2008 and 2016, respectively. He was a Visiting Associate Professor with the Department of Radiology and Imaging Sciences of Emory University. He is currently a Professor and a Radiologist with the Sun Yat-sen University Cancer Center. His research interests include oncologic imaging and medical image analysis.



HONGBO CHEN received the B.Sc. degree in physics from Xiangtan Normal University, in 1996, the M.Sc. degree in computer software and theory from Guangxi Normal University, in 2004, and the D.Sc. degree in biomedical engineering from Central South University, in 2009. He was a Visiting Professor with the Center of System Imaging, Emory University. He is currently a Professor with the School of Life and Environmental Sciences, Guilin University of Electronic Technology. His research interests include medical image analysis and medical pattern recognition.

• • •



HAOJIANG LI is currently a Radiologist with the State Key Laboratory of Oncology in South China, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Collaborative Innovation Center for Cancer Medicine, Department of Radiology, Sun Yat-sen University Cancer Center, Guangzhou, Guangdong, China. His areas of research interests include data mining in medical imaging and statistics.