IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 1, 2021, accepted April 19, 2021, date of publication April 27, 2021, date of current version May 14, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3075294

Double U-Nets for Image Segmentation by Integrating the Region and Boundary Information

WEI GUO[1], HANXUN ZHOU[2], ZHAOXUAN GONG[1], AND GUODONG ZHANG[1]
[1]School of Computer, Shenyang Aerospace University, Shenyang 110136, China
[2]Department of Information Science, Liaoning University, Shenyang 110036, China

Corresponding author: Guodong Zhang (zhanggd@sau.edu.cn.com)

This work was supported in part by the Liaoning Education Department under Grant JYT19053, in part by the National Natural Science Foundation of Liaoning Province under Grant 2020-MS-239, in part by the Key Scientific Research Projects of Liaoning Provincial Department of Education under Grant LZD202002, in part by the Aviation Science Foundation under Grant 2019ZE054009, and in part by the Teaching Reform Project of Liaoning University under Grant 2020YBXM12.

ABSTRACT The existing CNN-based segmentation methods use the object regions alone as the labels to train their networks, and the potentially useful boundaries annotated by radiologists are not used directly during the training. Thus, we proposed a framework of double U-Nets to integrate object regions and boundaries for more accurate segmentation. The proposed network consisted of a down-sampling path followed by two symmetric up-sampling paths. The down-sampling path learned the low-level features of regions and boundaries, and two up-sampling paths learned the high-level features of regions and boundaries, respectively. The outputs from the down-sampling path were concatenated with the corresponding ones from two up-sampling paths by skip connections. The outputs of double U-Nets were the predicted probability images of object regions and boundaries, and they were integrated to calculate the dice loss with the corresponding labels. The proposed double U-Nets were evaluated on two datasets: 247 radiographs for the segmentation of lungs, hearts, and clavicles, and 284 radiographs for the segmentation of pelvises. Compared with the baseline U-Net, our double U-Nets improved the mean dices and reduced the 90% Hausdorff distances for the ''difficult'' objects (lower lungs, clavicles, and pelvises), and the integration of ''difficult'' object regions and boundaries can improve the segmentation results compared with the use of object regions alone. However, for the ''easy'' objects (entire lungs and hearts) or ''very difficult'' objects (pelvises in lateral and implanted images), the integration did not improve the segmentation performance.

INDEX TERMS Image segmentation, double U-Nets, integrate regions and boundaries.

I. INTRODUCTION

Computer-aided diagnosis (CAD) has become one of the major topics in medical imaging and diagnostic radiology. It can assist radiologists in image interpretation and decision making. The segmentation of organs and abnormalities in medical images is an important first step in a CAD system, which will also provide volume and shape parameters for clinical quantitative analysis.

Routinely, the delineation of objects is mostly based on manual procedures, which is time-consuming and tedious, especially for those in 3D images. Therefore, some automatic methods [1], [2] have been developed for segmentation. However, it is challenging for an automatic segmentation with high performance due to several factors such as the low contrast of images, variables object sizes of different patients, and similarity between the shapes of nearby objects.

Recently, with the availability of a large number of well-annotated data and advancements of graphics processing units (GPUs), convolutional neural networks (CNNs) with more layers have significantly improved the performance levels in computer vision and image processing [3]–[5]. Inspired by the remarkable successes of CNNs, many CNN architectures have been successfully utilized in the segmentation of various organs and abnormalities in medical images [6]–[8].

The high-performance CNNs-based methods usually require a large number of manual annotations for optimizing the massive amount of network parameters. For example, the residual network won the ImageNet Large Scale Visual Recognition Challenge in both the classification and single-object location tasks, and employed more than 10 million

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

annotations for successfully training [9]. However, due to expensive expert annotation and privacy issues, there are not sufficient training images and annotations in practice. To this end, some groups employed the affine transformation to generate samples and then used the generated samples to train their networks [10], [11]. Except for augmenting the dataset, we can also make full use of the potentially useful boundaries of objects for more effective training.

The reference standards of objects to be segmented are usually the boundaries delineated by radiologists. However, the existing CNN-based [6]–[23] methods usually employ the object regions, instead of the boundaries, as the labels for training their networks. If we fully take advantage of both regions and boundaries for object segmentation, the results will be better than those by using the region alone. Therefore, we proposed a framework of double U-Nets to integrate the information of object regions and boundaries and applied it to the segmentation of objects in radiographs.

II. RELATED WORKS

In the early time, CNNs were usually used to segment images in a patch-based way [7], [12]. The fixed-size patches were extracted for all pixels in the entire image and then used as the inputs to train and test a classification CNN. The classification result for each patch was assigned as the segmentation result of its center point in the entire image. Apparently, the high overlap between the adjacent patches will cause massive redundant operations.

A fully convolutional network (FCN) [13] was proposed by Long in 2015, and it employed de-convolutional layers to up-sample the feature maps for producing outputs with the same size as the input. Therefore, it can segment all the pixels in the entire image end-to-end and achieve a faster processing speed for image segmentation than the patch-based methods. Many architectures based on the FCN were proposed for a variety of applications in medical segmentation. He *et al.* [14] employed the FCN guided by the distinctive curve to segment the rectums, prostates, and bladders. Trullo *et al.* [15] combined the FCN with the conditional random fields to segment the organs at risk in thoracic CT scans. However, because the FCN ignores some useful information about the shallow location, the segmented details are usually not very accurate at the fine-scale.

The U-Net was considered as one of the most well-known CNN architectures for image segmentation [16], which stemmed from the FCN and achieved great success in biomedical image segmentation. It employs down-sampling and up-sampling paths to form a symmetric U-shaped fully convolutional network. The skip connections in the U-Net concatenate the outputs from the down-sampling path with those from the up-sampling path to well combine the shallow locations with deep semantics of images for more accurate segmentation. Since the appearance of U-Net, some similar network architectures [17]–[20] have been proposed for various segmentation of medical images.

III. MATERIALS

Our double U-Nets were evaluated on two datasets: 247 chest radiographs with lungs, hearts, and clavicles, and 284 radiographs with pelvises.

A. JSRT DATASET

The 247 chest radiographs were collected from a standard digital image database [21], which was created by the Japanese Society of Radiological Technology (JSRT) in cooperation with the Japanese Radiological Society. The chest radiographs comprise 2048×2048 pixels with a spatial resolution of 0.175 mm and 12-bit gray levels.

Two observers delineated 3 anatomical structures, i.e. lung, heart, and clavicle in each chest radiograph independently [22]. They were a medical student and a computer science student, and both of them were specialized in the medical image. They were also instructed by an experienced radiologist before their delineation. The segmentation results of the medical student were taken as the reference standards in this study.

B. PELVIS DATASET

The pelvis dataset consisted of 284 radiographs in the Dicom format, which were collected from a hospital in Shanghai, China. Seventy-five out of 284 were images with an implant, and the other 209 images were without an implant. Eighty-six out of 284 were lateral images, and the other 198 were posteroanterior images. The radiographs comprise 1031×1325 pixels with a spatial resolution of 0.125 mm. The distance from the source to the detector was 1770 mm, and the exposure time was 66 s. The radiologist delineated the pelvises in radiographs, and we used them as the reference standards to evaluate the segmentation methods.

IV. METHODS

A. FRAMEWORK OF DOUBLE U-NETS

The architecture of double U-Nets is shown in Fig. 1 for the segmentation of lungs in a chest radiograph, which integrated the lung's region and boundary information. It included a down-sampling path followed by two symmetric up-sampling paths, and the symmetric paths formed a double U-shaped fully convolutional network. The down-sampling path learned the low-level features of regions and boundaries, and two up-sampling paths learned the high-level features of regions and boundaries, respectively. The skip connections concatenated the outputs from the down-sampling path with the corresponding ones from two up-sampling paths. Two U-Nets were employed to predict the probability images of object region and boundary, respectively.

All the paths were divided into five stages with five feature resolutions. Each stage included two blocks, and each block included a 3×3 convolution layer followed by a batch normalization layer and a ReLU activation layer. In the down-sampling (up-sampling) path, the second block was followed by a max-pooling (up-convolution) layer with a kernel size

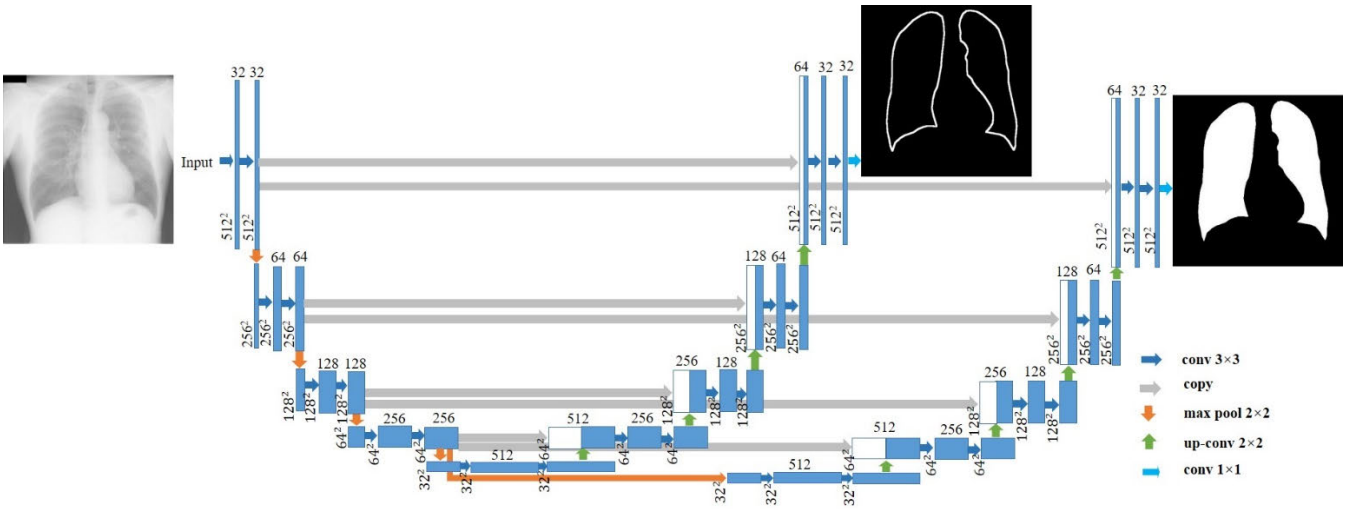

FIGURE 1. Double U-Nets architecture. The blue and white boxes represent the multi-channel feature maps and the copied feature maps, respectively. The numbers on the top and lower-left edge of each box denote the number of feature channels and the size of feature maps, respectively. The arrows of different shapes and colors indicate different operations. Each convolution throughout the network is followed by a batch normalization (BN) and a rectified linear unit (ReLu), which are omitted in this diagram.

of 2×2. Two last layers in two up-sampling paths employed a 1×1 convolution to transform the feature maps with 32 channels to those with 1 channel. Then, two sigmoid functions converted two final feature maps for the region and boundary into their corresponding segmentation probability images. For the testing, the predicted probability image of the region was converted to the binary one as the final segmentation result by thresholding at 0.5.

The predicted probability images of object region and boundary were combined with their corresponding labels for calculating the dice loss, and the information of object boundary and region was integrated by optimizing the dice loss function. Compared with using the object region, the boundary information provided additional local information for object segmentation. Therefore, we integrated the information of both regions and boundaries to train a CNN for achieving better results of object segmentation.

B. LOSS FUNCTION FOR DOUBLE U-NETS

The *dice* has been used as an objective function in CNN-based segmentation methods, and the *dice* for bi-segmentation between the predicted binary image P and the ground truth binary image G can be written as [23]

$$Dice = \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i + \sum_i^N g_i}. \tag{1}$$

where p_i and g_i represent the pixel value in the predicted binary image P and the ground truth binary image G, respectively. N denotes the total number of pixels in the object.

In the case of multiclass segmentation tasks, the *dice* is extended to the *multiclass dice* (*M-Dice*) as

$$M - Dice = \prod_j^M Dice_j. \tag{2}$$

where M denotes the number of classes for segmentation. When the *dices* of all classes are equal to each other, the *M-Dice* is the largest. The corresponding *multiclass dice loss* (*M-loss*) is defined as

$$M - Loss = 1 - M - Dice. \tag{3}$$

In this study, the outputs of our double U-Nets were the predicted probability images, and they were employed to calculate the *M-Loss*, instead of the predicted binary images P, for more convenient calculation during the process of training. When the values of *dice* for both regions and boundaries were large and equal to each other, the *M-Dice* was largest, and the *M-Loss* is smallest. The predicted probability images of object region and boundary were integrated by optimizing the multiclass loss function.

C. EVALUATION METRICS FOR SEGMENTATION RESULTS

We employed two metrics to evaluate the segmentation results in our study. The first one is dice, which measures the overall consistency between the segmented result and its reference. The value of dice 0 gives no overlap between the two objects, and the value of 1 produces a perfect overlap.

The second is 90% Hausdorff distance (HD). The HD is the maximum distance between two objects and defined as

$$\begin{aligned} d_H(x, y) &= \max\{d_{XY}, d_{YX}\} \\ &= \max\{\max_{x\in X}\min_{y\in Y} d(x, y), \\ &\qquad \max_{y\in Y}\min_{x\in X} d(x, y)\} \end{aligned} \tag{4}$$

where X and Y denote the boundaries of the segmented result and its reference standard, respectively. The HD assesses the extreme dissimilarity between two boundaries and is too sensitive to very few outlier points with large distances. To end this, we employ the 90% HD, which is calculated based on

the 90th percentile of distances between the boundary points on the segmented result and its reference standard.

Further, we employed the mean values of dices and 90% HDs to evaluate the segmentation results for our two datasets.

D. IMPLEMENTATION OF OUR DOUBLE U-NETS

We trained a unique network for each object in chest radiographs. The networks' architecture and initial parameters were the same for all the networks. The experimental platform was equipped with an Intel E5 CPU with 128 GB RAM and an NVIDIA GTX-2080Ti GPU. The double U-Nets were implemented by Python based on a deep learning library of Keras. The weights of networks were initialized by He_normal distribution, and the networks were trained by minimizing the multiclass dice loss function with Adam optimization. We initialized the learning rate as 0.0003, and the batch size as 4.

The 5-fold cross-validation methods were used to evaluate the segmentation of objects in both datasets. We evenly divided all the cases into 5 subsets. For each fold, we used 4 subsets for training the networks, and the remaining one subset for testing. The processing of training and testing for a fold was repeated 5 times until all the images were tested.

The chest radiographs in the JSRT dataset were resized to 512 $\times$ 512 pixels for training and testing our networks, and pelvis radiographs were resized and padded to 420 $\times$ 420 pixels. The boundaries of lungs, hearts, and clavicle were dilated on both sides 4, 3, and inward 3 times, respectively. The boundaries of pelvises were dilated on both sides 3 pixels.

V. RESULT

A. SEGMENTATION RESULTS FOR LUNGS, HEARTS, AND CLAVICLES

1) SUBJECTIVE EVALUATION OF SEGMENTATION RESULTS

Figure 2 compares the segmentation results of lungs, hearts, and clavicles in chest radiographs by the baseline U-Net and our double U-Nets. The blue, red, and green curves indicate the delineated objects by radiologists, U-Net, and double U-Nets, respectively. The architecture and initial parameters of U-Net are the same as those used in our double U-Nets.

The U-Net did not segment the lower parts of the lungs well as shown in Fig. 2. Although the blurred boundaries near the diaphragm degraded the segmentation performance of U-Net, they did not worsen the boundaries segmented by our double U-Nets. Overall, our double U-Nets achieved better results than the U-Net with more accurate and smoother boundaries for the segmentation of lungs, hearts, and clavicles.

2) OBJECTIVE EVALUATION OF SEGMENTATION RESULTS

Due to the blurred boundaries of the lungs near the heart and diaphragm, the segmentation for the lower-left part of the lungs was more difficult than for other parts. Therefore, we divided the entire image into four parts with a size of 256 $\times$ 256 pixels and compared the segmentation result for each of the parts. Figure 3 shows the mean dices and 90% HDs for the segmentation of lungs, hearts, and clavicles by U-Net and our double networks.

a: LUNG

The segmentation result of U-Net for the lower-left part was not very good with a mean dice of 0.93. However, our double U-Nets improved the value to 0.96 by integrating the information of boundary and region. The double U-Nets improve the mean dice for the lower-right part a little from 0.97 to 0.98. For the two upper parts and entire images, the U-Net and double U-Nets achieved the same mean dice. Because the region of poor segmentation was not large in the lower part of images, it did not degrade the performance of U-Net for the entire image very much.

Compared with the U-Net, double U-Nets decreased the mean 90% HDs for the segmentation of the entire lungs from 6.05 mm to 3.20 mm. Further, double U-Nets also achieved a smaller mean 90% HDs than the U-Net for each part of the images.

b: HEART AND CLAVICLE

For the U-Net, the mean dices were 0.93 and 0.92 for the segmentation of hearts and clavicles, respectively. And the double U-Nets achieved the values of 0.93 and 0.94, respectively. The U-Net achieved a very large mean 90% HD of 124.31 mm for hearts, and double U-Nets greatly reduced the value to 17.85 mm. Although the mean 90% HD of the U-Net for clavicle was not very large (14.90 mm), our double U-Nets greatly reduced the value to 4.80 mm.

B. SEGMENTATION RESULTS FOR PELVIS

1) SUBJECTIVE EVALUATION OF THE SEGMENTATION RESULTS

Figure 4 shows the segmentation results of pelvises with and without an implant in a posteroanterior and lateral radiograph by the baseline U-Net and our double U-Nets. The first row is the original images, and the second row is the segmentation results. The blue, red, and yellow curves indicate the delineated objects by radiologists, U-Net, and double U-Nets, respectively. The architecture and initial parameters of U-Net are the same as those used in our double U-Nets.

For a pelvis in a non-implanted posteroanterior image as shown in Fig. 6, the U-Net misidentified some bone and tissue, but our double U-Nets precisely segmented the pelvis with smooth boundary. Due to the few lateral and implanted images, overlap of bones, and high contrast of implants, both U-Net and double U-Nets did not segment the pelvises in the lateral and implanted images well.

2) OBJECTIVE EVALUATION OF SEGMENTATION RESULTS

Figure 5 shows the mean dices and 90% HDs for the segmentation of all the pelvises by U-Net and our double networks. Compared with the U-Net, the double U-Nets improved the mean dices from 0.85 to 0.87 and reduced the mean

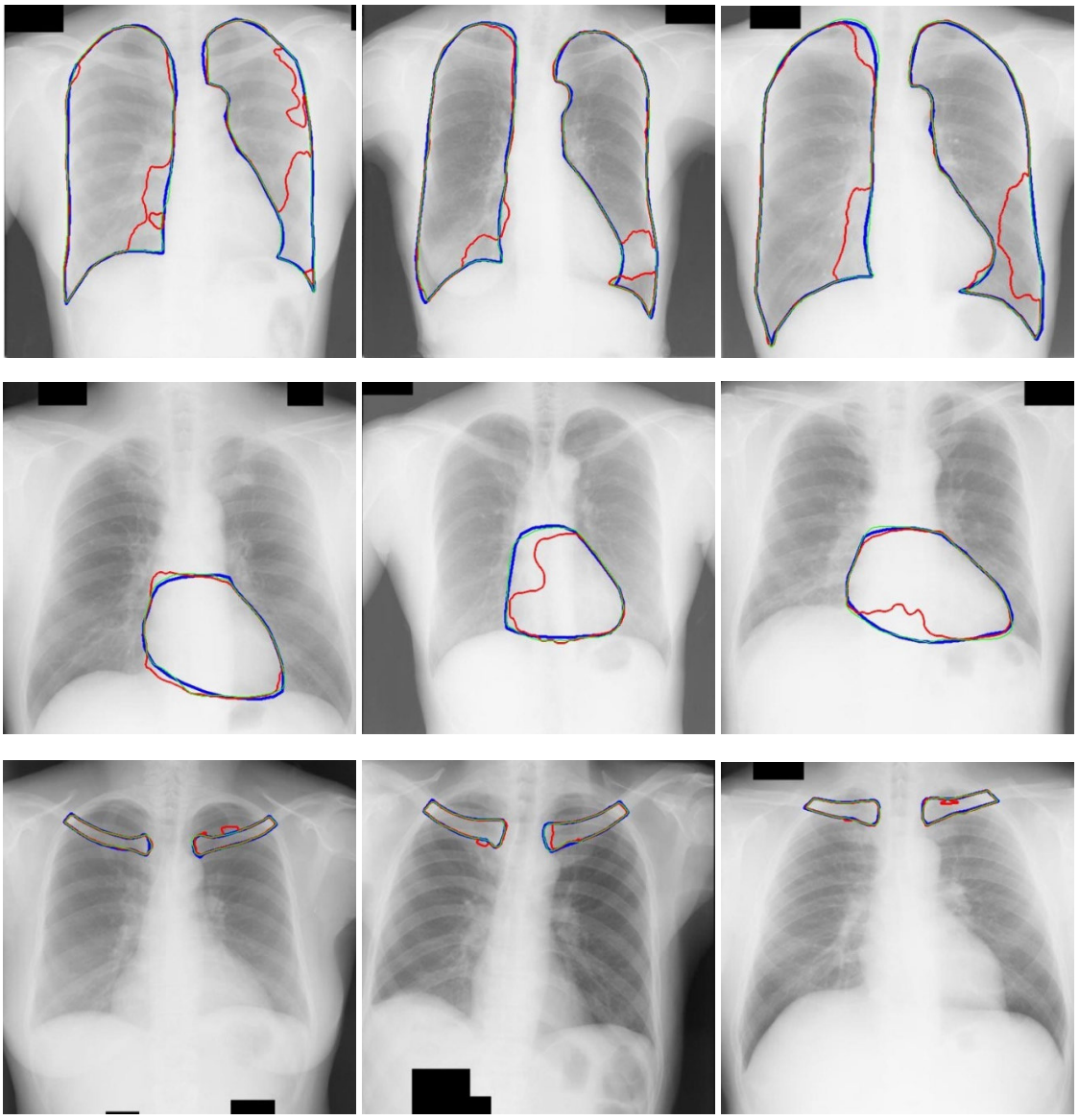

FIGURE 2. Segmentation results for lungs, hearts, and clavicles. The blue, red, and green curves indicate the delineated objects by radiologists, U-Net, and double U-Nets, respectively.

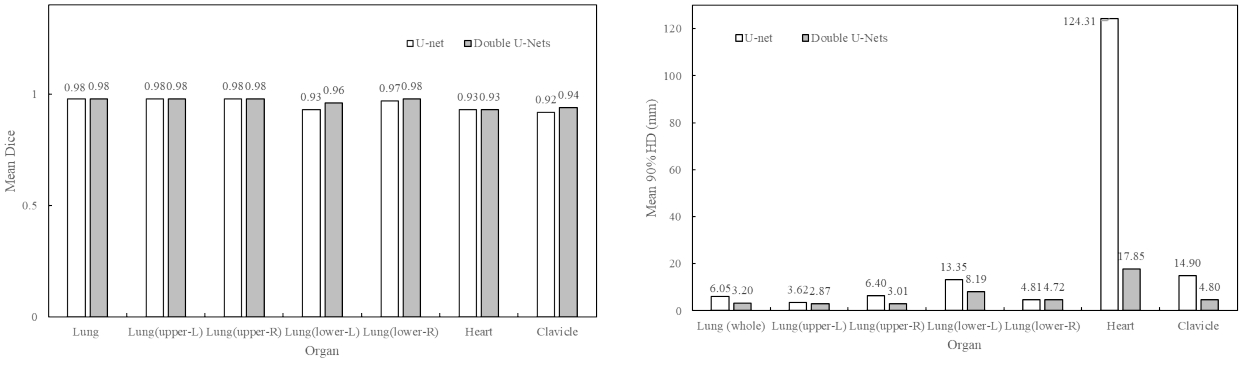

FIGURE 3. Mean Dices and 90% HDs for the segmentation of lungs, hearts, and clavicles by U-Net and double U-Nets.

90% HDs from 20.78 mm to 12.74 mm. In this study, the implanted and lateral images were difficult to be segmented accurately, and they degraded the segmentation performance of networks. Therefore, we also reported the segmentation results for the implanted images and lateral images, respectively.

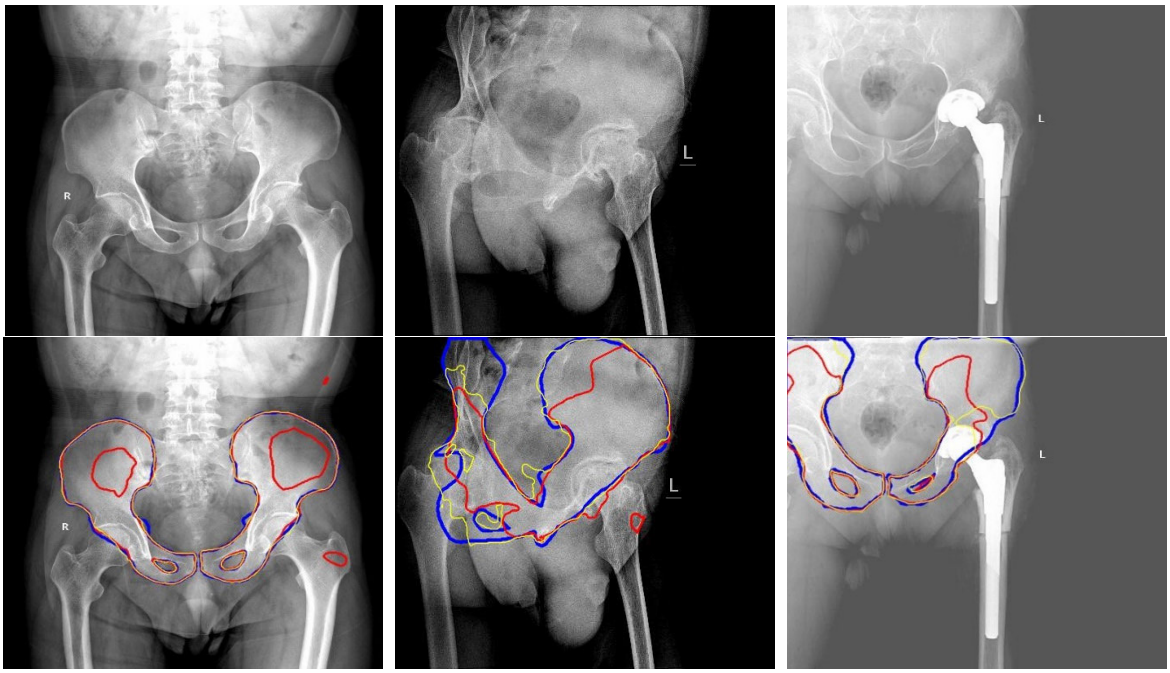

FIGURE 4. Segmentation results for the pelvises. The blue, red, and green curves indicate the delineated objects by radiologists, U-Net, and double U-Nets, respectively.

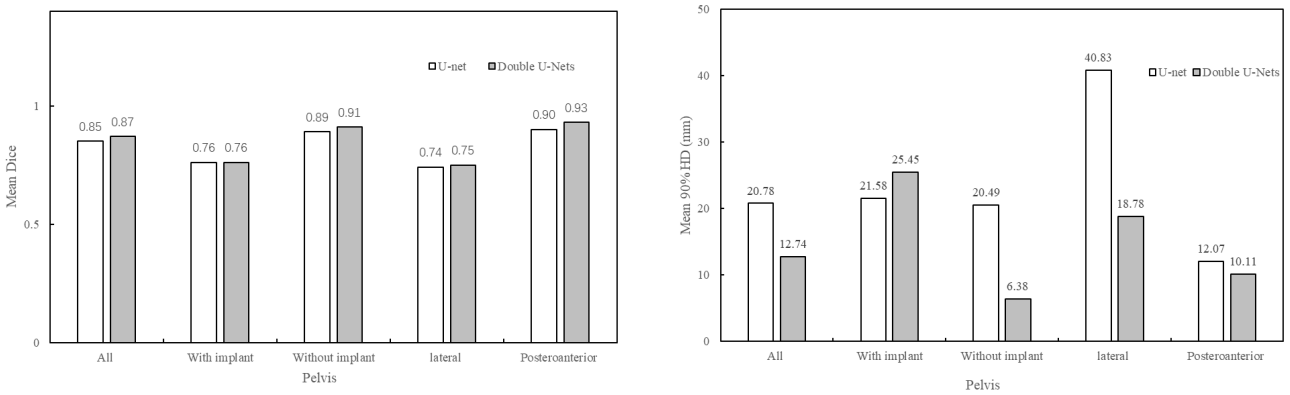

FIGURE 5. Mean Dices and 90% HDs for the segmentation of pelvises by U-Net and double U-Nets.

a: NON-IMPLANTED AND IMPLANTED IMAGES

For the non-implanted images, the double U-Nets achieved higher performance than the U-Net with a higher mean dice and a lower mean 90% HD. However, for the implanted images, both networks achieved poor segmentation performance with low mean dices and high mean 90% HDs due to the high contrast of implants and few training images.

b: POSTEROANTERIOR AND LATERAL IMAGES

For the posteroanterior images, the double U-Nets also improve the mean dice and reduced the mean 90% HD. For the lateral images, both networks did not segment the pelvises well due to the overlap of bones and few training data.

C. SEGMENTATION RESULTS FOR OTHER DATASETS

To further verify that the integration of object regions and boundaries can improve the segmentation results, we conducted an additional experiment on the datasets of ISBI 2019 SegTHOR and DRISHTI-GS1, and compared our double U-Nets with other state-of-the-art approaches.

1) DATASET

(1) The SegTHOR dataset included 60 scans with the manual delineations of the esophagus, heart, trachea, and aorta, and the whole dataset was randomly split into a training set of 40 scans and a testing set of 20 scans. The CT scans have a size of 512×512 pixels on XY-plane with spatial resolutions between 0.90 mm and 1.37 mm. The numbers of slices in a scan range from 150 to 284 with a Z-resolution between 2 mm and 2.5 mm.

(2) The DRISHTI-GS1 dataset consisted of 101 retinal images, which were collected from Aravind eye hospital, Indian. The images are in PNG uncompressed format and have an approximate size of 2047×1760 pixels. Four experts delineated the boundaries of optical discs (ODs) and optical

TABLE 1. Dice metric for SegTHOR 2019.

Method	Dice			
	Esophagus	Heart	Trachea	Aorta
(1) 3D V-Net/ ResNet	0.8651	0.9536	0.9276	0.9464
(2) 3D V-Net/ ResNe	0.8597	0.9459	0.9217	0.9433
(3) 2.5 D U-Net/ DenseNet /ResNet	0.8594	0.9500	0.9201	0.9484
(4) 3D V-Net/ ResNet	0.8166	0.9329	0.8910	0.9232
(5) 2D U-Net/ attention mechanism/ pixel shuffle	0.8303	0.9381	0.9088	0.9353
(6) 3D V-like++	0.7734	0.9414	0.8927	0.9233
(7) 2.5D U++	0.7518	0.9328	0.8885	0.8919
(8) 2D U-Net	**0.7462**	0.9433	0.9163	**0.8755**
(9) **Double 2D U-Net**	**0.7843**	0.9528	0.9287	**0.9321**

cups (OCs), and the reference standards of boundaries were determined by the majority of experts.

2) RESULT

a: RESULTS FOR ISBI 2019 SegTHOR CHALLENGE

We performed the comparison between the U-Net (8) and double U-Nets (9). As shown in Table 1, the U-Net achieved quite high-performance levels with the mean dices of 0.9433 and 0.9163 for the heart and trachea, respectively, and the double U-Nets also achieved similar results with the mean dices of 0.9528 and 0.9287. For the esophagus and aorta, the results of the U-Net were not very good with the mean dices of 0.7462 and 0.8755, respectively, however, double U-Nets improved the values to 0.7843 and 0.9321 by integrating the information of regions and boundaries.

Moreover, we downloaded the top-ranking teams' reports and compared with their approaches and results. Other teams used the training and testing sets to train and test their networks, respectively. Because we can only download the training set from the website, we used 40 scans of the training set to evaluate our networks by the 5-fold cross-validation method.

Most teams (1, 2, 3, 4, 6, and 7) employed a coarse- and fine-resolution 3D/2.5D network for locating and segmenting an object, respectively, and architectures of networks were mainly based on the V-Net/U-Net with ResNet or DenseNet modules. Team (3) segmented four classes/organs simultaneously, and Team (1) fused the segmentation results for the single-class and multi-class. Team (2) used an additional abdomen dataset with 13 classes of organs and 50 scans to pre-train their model. Team (5) added the pixel shuffle and the attention model in their 2D U-Net.

Compared with other teams, we used a very simple network to locate and segment objects and obtained comparable results for the heart, trachea, and aorta. If we also use a multi-resolution strategy for locating and segmenting respectively, our final segmentation performance may be improved. Moreover, if we use additional residual or attention modules in our networks, our segmentation results may be also better. On the other hand, other teams used only the object regions as labels to train their networks. If they also integrate the object regions and boundaries, they may achieve more accurate segmentation results.

b: RESULTS FOR DRISHTI-GS1

Table 2 shows the segmentation results of U-Net, double U-Nets, and other methods. For ODs and OCs, compared with the U-Net, the double U-Nets networks improved the mean values of dices from 0.95 and 0.84 to 0.97 and 0.87, respectively. It indicates that the integration of regions and boundaries enables the double U-nets to produce better segmentation results than that used the region alone.

TABLE 2. Dice and accuracy metrics for DRISHTI-GS1.

Method	Dice		Accuracy	
	ODs	OCs	ODs	OCs
(1) Modified U-Net[24]	0.9644	0.8739	------	-----
(2) DeepLab V3+MobileNet [25]	0.9173	------	------	-----
(3) Recurrent fully convolution network [26]	------	------	0.976 4	0.977 8
(4) U-Net	0.9502	0.8396	------	-----
(5) **double 2D U-Net**	**0.9656**	**0.8721**	------	-----

We also compared the segmentation results of our double U-Nets with three state-of-the-art methods. Yu *et al.* [24] used two networks for locating and segmenting ODs and OCs. The segmentation network was the modified U-Net, and the pre-trained ResNet-34 was used as the encoding layers. Sreng *et al.* [25] adopted the combination of DeepLab V3 and MobileNet to segment ODs. Our double U-Net achieved comparable results as shown in Table 2.

Gao *et al.* [26] also used a CNN-based method to segment ODs and OCs, which is mainly composed of a multi-scale input layer, recurrent fully convolutional network, multiple output layer, and polar transformation. Because they used different metrics to evaluate their method, we cannot directly compare the segmentation results. But it can be believed that our segmentation performance is not much worse than that of theirs. Moreover, if they used the integration of object regions and boundaries in their networks, they will achieve better segmentation results compared with the use of regions alone.

VI. DISCUSSION

The reference standards of objects for segmentation are usually the boundaries delineated by radiologists. However, the U-Net discards the expensively delineated boundaries and employs the regions alone to train the network. For some "easy" objects (lungs and hearts), the region provides sufficient information and the U-Net achieves a good result. Double U-Nets does not improve the segmentation performance. But for some "difficult" objects (lower lungs, clav-

icles, and pelvises), the U-Net based on the region does not achieve a good segmentation result, and the boundary will provide additional useful information. The double U-Nets integrate the region and boundary information in the feature extraction and loss calculation, and such integration enables the network to learn more useful information for better segmentation. Furthermore, for some ''very difficult'' objects (pelvises in lateral and implanted images) with few training data, the integration of boundary and region also does not learn very useful features, and the integration does not improve the performance of segmentation.

In our study, we evaluated the effect of different boundary dilation schemes on the segmentation result. For the JSRT chest radiographs, we dilated the boundaries of lungs on both sides from 2 to 8 pixels, and the boundaries of hearts and clavicles on both sides and inward from 1 to 3 pixels, respectively. For the radiographs of pelvises, we dilated the boundaries of pelvises on both sides from 3 to 5 pixels. The segmentation results indicated that the mean dices for all objects with different dilation schemes remain similar. Although the mean 90% HDs of our double U-Nets were a little different for the different dilation schemes of all objects, most of them were lesser than those of the U-Net.

In our study, the architecture of U-Net was the same as one network used in the double U-Nets. The number of parameters in the U-Net was 8,641,697, and that in the double U-Nets was increased to 16,107,874. Although the number of parameters in the double U-Nets was almost twice of that in the U-Net, the calculation time was not doubled and only slightly increased due to the GPU parallel computing.

VII. CONCLUSION

In this study, we proposed a framework of double U-Nets to integrate the region and boundary information in the feature extraction and loss calculation, and applied it to the segmentation in two datasets. The experiments show that the integration of ''difficult'' object boundary and region can improve the segmentation results compared with the use of object region alone.

REFERENCES

[1] M. Han, J. Ma, Y. Li, M. Li, Y. Song, and Q. Li, ''Segmentation of organs at risk in CT volumes of head, thorax, abdomen, and pelvis,'' *Proc. SPIE Medical Imag.*, vol. 9413, Mar. 2015, Art. no. 94133.

[2] E. Schreibmann, D. M. Marcus, and T. Fox, ''Multiatlas segmentation of thoracic and abdominal anatomy with level set-based local search,'' *J. Appl. Clin. Med. Phys.*, vol. 15, no. 4, pp. 22–38, Jul. 2014.

[3] J. Donahue, ''DeCAF: A deep convolutional activation feature for generic visual recognition,'' in *Proc. 31st Int. Conf. Mach. Learn.*, 2014, vol. 32, no. 1, pp. 647–655.

[4] Y. LeCun, Y. Bengio, and G. Hinton, ''Deep learning,'' *Nature*, vol. 521, pp. 436–444, May 2015.

[5] D. Shen, G. Wu, and H. Suk, ''Deep learning in medical image analysis,'' *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, Jun. 2017.

[6] K. Men, J. Dai, and Y. Li, ''Automatic segmentation of the clinical target volume and organs at risk in the planning CT for rectal cancer using deep dilated convolutional neural networks,'' *Med. Phys.*, vol. 44, no. 12, pp. 6377–6389, Dec. 2017.

[7] N. Dhungel, ''Deep learning and structured prediction for the segmentation of mass in mammograms,'' in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCIA)*, 2015, pp. 605–612.

[8] J. Kleesiek, G. Urban, A. Hubert, D. Schwarz, K. Maier-Hein, M. Bendszus, and A. Biller, ''Deep MRI brain extraction: A 3D convolutional neural network for skull stripping,'' *NeuroImage*, vol. 129, pp. 460–469, Apr. 2016.

[9] K. He, ''Deep residual learning for image recognition,'' in Proc. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, May 2015, pp. 770–778.

[10] A. A. A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S. J. van Riel, M. M. W. Wille, M. Naqibullah, C. I. Sanchez, and B. van Ginneken, ''Pulmonary nodule detection in CT images: False positive reduction using multi-view convolutional networks,'' *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1160–1169, May 2016.

[11] H. R. Roth, L. Lu, J. Liu, J. Yao, A. Seff, K. Cherry, L. Kim, and R. M. Summers, ''Improving computer-aided detection using convolutional neural networks and random view aggregation,'' *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1170–1181, May 2016.

[12] F. Xing, Y. Xie, and L. Yang, ''An automatic learning-based framework for robust nucleus segmentation,'' *IEEE Trans. Med. Imag.*, vol. 35, no. 2, pp. 550–566, Feb. 2016.

[13] J. Long, E. Shelhamer, and T. Darrell, ''Fully convolutional networks for semantic segmentation,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[14] K. He, X. Cao, Y. Shi, D. Nie, Y. Gao, and D. Shen, ''Pelvic organ segmentation using distinctive curve guided fully convolutional networks,'' *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 585–595, Feb. 2019.

[15] R. Trullo, C. Petitjean, S. Ruan, B. Dubray, D. Nie, and D. Shen, ''Segmentation of organs at risk in thoracic CT images using a SharpMask architecture and conditional random fields,'' in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 1003–1006.

[16] O. Ronneberger, P. Fischer, and T. Brox, ''U-net: Convolutional networks for biomedical image segmentation,'' in *Proc. Med. Image Comput. Comput.-Assisted Intervent (MICCAI)*, 2015, pp. 234–241.

[17] 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.

[18] O. Çiçek, ''3D U-Net: Learning dense, volumetric segmentation from sparse annotation,'' in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2016, pp. 424–432.

[19] 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.

[20] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, ''Unet++: A nested u-net architecture for medical image segmentation,'' in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, vol. 11045. Cham, Switzerland: Springer, 2018, pp. 3–11.

[21] J. Shiraishi *et al.*, ''Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules,'' *AJR*, vol. 174, pp. 71–74, 2000.

[22] B. van Ginneken *et al.*, ''Segmentation of anatomical structures in chest radiographs using supervised methods: A comparative study on a public data,'' *Med. Image Anal.*, vol. 10, no. 1, pp. 19–40, 2006.

[23] X. Gu *et al.*, ''Segmentation and suppression of pulmonary vessels in low-dose chest CT scans,'' *Med. Phys.*, vol. 46, no. 8, p. 13648, 2019.

[24] S. Yu *et al.*, ''Robust optic disc and cup segmentation with deep learning for glaucoma detection,'' *Comput. Med. Imag. Graph.*, vol. 74, pp. 61–71, 2019.

[25] S. Sreng *et al.*, ''Deep learning for optic disc segmentation and glaucoma diagnosis on retinal images,'' *Appl. Sci.*, vol. 10, no. 14, p. 4916, 2020.

[26] J. Gao *et al.*, ''Joint disc and cup segmentation based on recurrent fully convolutional network,'' *PLoS ONE*, vol. 15, no. 9, pp. 1–23, 2020.

[27] J. Gao, ''Joint disc and cup segmentation based on recurrent fully convolutional network,'' *PLoS ONE*, vol. 15, no. 9, pp. 1–23, 2020.

WEI GUO was born in 1983. She received the Ph.D. degree from Northeastern University, China, in 2011. Since 2011, she has been working with Shenyang Aerospace University, China, where she is currently an Associate Professor. Her research interests include artificial intelligence, machine learning, and image processing.

HANXUN ZHOU was born in 1981. He received the Ph.D. degree from Northeastern University, China, in 2009. He is currently an Associate Professor with Liaoning University, China. His research interest includes machine learning.

ZHAOXUAN GONG was born in 1983. He received the Ph.D. degree from Northeastern University, China, in 2017. Since 2017, he has been working with Shenyang Aerospace University, China. His research interests include medical image processing and artificial intelligence.

GUODONG ZHANG was born in 1972. He received the M.S. degree from the South China University of Technology, China, in 2003. He is currently a Professor with Shenyang Aerospace University, China. His research interests include medical image analysis and machine learning.