Multi-Task Refined Boundary-Supervision U-Net (MRBSU-Net) for Gastrointestinal Stromal Tumor Segmentation in Endoscopic Ultrasound (EUS) Images

XINYI LI1, YI GUO1, FEI JIANG2, LEIMING XU3, FENG SHEN3, ZHENDONG JIN2, AND YUANYUAN WANG1

1Department of Electronic Engineering, Fudan University, Shanghai 200433, China
2Department of Gastroenterology, Changhai Hospital, Shanghai 200433, China
3Department of Gastroenterology, Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai 200092, China

ABSTRACT The diagnosis of risk level of gastrointestinal stromal tumor (GIST) is of great clinical significance. The morphology of GIST in endoscopic ultrasound (EUS) images has been normally used by radiologists to diagnose the risk level of GISTS. Hence, accurate segmentation of GISTS in EUS images is a crucial factor to influence the diagnosis. U-net, an elegant network, has been commonly used in medical images. However, due to the plain architecture and complicated up-sampling path of U-net, classical U-net does not perform well in segmenting GISTS in EUS images with diverse size, heavy shadow and ambiguous boundary. Hence, this paper proposes a novel multi-task refined boundary-supervision U-net (MRBSU-net) for GIST segmentation in EUS images. In our network, multi-task refined U-net (RU-net) is set to deal with heavy shadow and diverse size. Boundary cross entropy in loss function of multi-task RU-net boosts the influence of small size tumors and the refinement avoid the noise information in EUS images propagating to the higher resolution layers. Then we design a refined boundary-supervision U-net (RBSU-net) to solve the ambiguous problem. The boundary supervision in RBSU-net leads the network focus on finding boundary in the down-sampling part and segmenting region on the up-sampling path. At last, we put multi-task RU-net in front of the RBSU-net to increase the stability of the network, what is called MRBSU-net. Extensive experiments have been designed to evaluate the performance of the proposed network. The comparison experiments include the results from traditional U-net, generative adversarial network (GAN) and Deep Attentional Features (DAF). The results of our proposed method perform best among all the comparison methods, which proves that the proposed network could be potentially used in clinic.

INDEX TERMS Gastrointestinal stromal tumor, endoscopic ultrasound, tumor segmentation, multi-task, boundary-supervision.

I. INTRODUCTION
Gastrointestinal stromal tumors (GISTs) are uncommon tumors of the GI tract. They arise in very early forms of special cells called interstitial cell of Cajal [1]. Several imaging modalities are applied into diagnosis and follow-up treatment of GISTs, such as the computerized tomography (CT), the magnetic resonance imaging (MRI) and the endoscopic ultrasound (EUS). Due to its real-time nature, non-invasive, inexpensive, and non-radiation, endoscopic ultrasound is used to assess patient specific gastrointestinal structure and function [2]. The risk level plays a significant role in the pre-operative treatment. Risk level determines a patient whether to undergo the targeted therapy or not. The lower risk level group (LRG) and the higher risk group (HRG) are two level groups of GISTs. In particular, HRG patients should take the targeted therapy before surgery, while LRG patients do not need. Recently, a computer-aided diagnosis model has

The associate editor coordinating the review of this manuscript and approving it for publication was Gina Tourassi.
Due to the nature of US images, in particular, acoustic shadows, speckle noise, simple structure, ambiguous boundary and poor contrast [3], [4], US images are difficult for automatic segmentation [5]. A limited number of automatic segmentation methods usually based on machine learning methods have been proposed up to now. As medical records being increasingly digitalized, the success of machine learning methods on computer vision tasks reaches an opportune time[6]. Deep learning methods have illustrated remarkable accuracy and efficiency in segmenting a scope of objects [7], [8]. It has become a useful research approach in many fields, and several segmentation methods depending on convolutional neural networks (CNNs) have been proposed on medical imaging [9]–[11]. End-to-point network, such as patch-based CNN pixel classification [12] is time consuming. To overcome these disadvantages, plenty of end-to-end networks come into being. One of the most popular networks is the fully convolutional neural networks (FCNs), the basis of end-to-end network in semantic segmentation [13]. O. Ronneberger proposed U-net in 2015 to segment medical images, achieving a profound impact [14]. Recently, U-net always achieves the state-of-the-art performance on medical image segmentation. To decrease the influence of information loss during the up-sampling, U-net propagated the context information to the higher resolution layer by connecting the formal information to the current feature maps. Although this elegant architecture has solved the problem of various locations, it leads to the shadow information transferring to the resolution layer. Moreover, in addition to the traditional US image problems, Gastro EUS (G-EUS) image exists its own problems: 1) GISTs are diverse size, as shown in Figure 1. 2) Shadows are always similar to the tumors, leading to the misjudgment (Figure 2.) which cause the difficulties of training

a traditional U-net. Traditional cross-entropy loss function of region does not work on small size tumors. Obviously, even if those tumors were segmented as background, it has tiny influence on the loss function of region.

In this paper, we propose a novel multi-task refined boundary-supervision U-net (MRBSU-net) and a modified loss function to improve GISTs segmentation in EUS images with small size and ambiguous boundary.

Our main contributions are:

1) A refined U-net (RU-net) is proposed to decrease the influence of the heavy shadow. The encoder part in U-net is lower as well as the decoder part in the U-net has too many feature channels to propagate trashy information to higher resolution layers by the skip connections. For instance, the noise and shadows would be magnified and propagated in the decoder part [15]. RU-net reduce the concatenation between the previous layer and higher resolution layer, which could avoid the shadow information propagating to the higher resolution layers. Meanwhile, the convolutional layers in the up-sampling path are cut down.

2) To intensify the influence of loss function in small size tumor, we design a multi-task RU-net which consists of two RU-nets with the shared down-sampling part, a variant of Deep Contour-Aware Networks (DCAN) [16]. A boundary loss is added in the loss function to increase the impact on misjudgment of small size tumor. Summing the boundary loss and region loss up add a bonus of a multi-task RU-net, getting boundary and region information respectively without any post-processing.

3) We all know the advantages of fusing the tumor and boundary [17]. Nonetheless boundary details require deep supervision [18]. Due to boundary information losing in the down-sampling path, we propose a refined boundary-supervision U-net (RBSU-net) which focus on finding boundary in the down-sampling part and segmenting region on the up-sampling path. And we add a boundary-based loss in the final encoder part to reinforce the boundary restriction for those tumors with ambiguous boundary.

4) At last, we concatenate feature maps from two branches in multi-task RU-net before the classification and feed it to RBSU-net to increase the stability of the network.
This paper is organized as follows: In Section 2, the description of our MRBSU-net method are provided. The materials and experiments are introduced in Section 3. Section 4 contains an explanation for the experimental results. Discussion and conclusions are presented in Section 5 and Section 6, respectively.

II. METHOD
A. OVERVIEW
The proposed method was aimed at segment GISTs from background. Due to the limitation of the EUS images, the method should have the capability of differentiate GISTs from shadows and not be sensitive to noise. In this section, we describe the architecture of our proposed MRBSU-net method in details, as shown in Figure 3. We start with introducing the U-net for end-to-end training. Furthermore, we use the multi-task background features with auxiliary supervision to generate a good likelihood maps of GISTs. Then we concatenate two probability maps from the two branches and feed to a block of a refined boundary-supervision U-net (RBSU-net) followed by a softmax in the final classification layer.

B. MULTI-TASK RU-net
U-net achieved the state-of-the-art performance on image segmentation, especially medical image segmentation [14]. The network can be trained in an end-to-end way like U-net, which redeems image as input and outputs the probability map directly. The elegant architecture propagates context information to higher resolution layers, where the missing context is extrapolated the input image. This process solves the problem of various locations. The down-sampling path is designed to extracting the deep level invisible information, while the up-sampling path calculating the score pixel by pixel. However, the shallow depth of U-net and relatively small receptive field cause the poor performance. The shallow depth and complicated convolutional layers lead to the misjudgment of shadow and boundary to be matte.

Hence, to solve the limitations of U-net, we design a RU-net which includes an encoder part and a decoder part, shown in Figure 4. The decoder part contains 5 convolutional blocks with max pooling, which are widely used in the CNN for image classification [12]. Each block consists of 1~3 convolutional layers similar to that in VGG-16 network [19]. The decoder part extracts features. Different from the original U-net architecture, we simplify the last three up-sampling blocks replaced by the linear interpolation and cancel the convolutional layers in every up-sampling block to slick the boundary and avoid learning more redundant information. Moreover, sum operation instead of the concatenation operation combine the conciseness of the FCN-8s with the information in the mirroring position of the input image.

Afterwards, the classification scores from U-net relies on the intensity information from the given input. However, the network with single receptive input size cannot deal with the large differences in shape properly. Hence, we proposed a multi-task learning network instead of redeeming segmentation as an individual pixel-wise classification problem. In Figure 5, two of the above RU-nets share the encoder part and two different decoding branches are used for two tasks, tumor region classification and tumor boundary classification. At last, RU-net with multi-task features extract from input A can be trained by minimizing the loss $\text{Loss}_1$, a combination of loss $L_{E1}$ and $L_{R1}$, as shown following:

$$\text{Loss}_1 (A; W_1, W_2) = L_{E1} (A; W_1) + L_{R1} (A; W_2) \quad (1)$$

where $W_i$ represents the parameters of neural network, $L_E$ is the loss of tumor edge and $L_R$ is the loss of tumor region. All these losses are cross entropy losses:

$$L_x (A; W) = \sum_n \sum_i \log P_x (l_x (a_n,i); a_n,i, W) \quad (2)$$
$L_x$ represents the loss function of each task, $a_{n,i}$ is the $i$-th pixel in the $n$-th image and $P_x$ is the predicted probability of $a_{n,i}$ belonging to class $l_x$.

C. MULTI-TASK RBSU-net (MRBSU-net)

Although multi-task RU-net solves the various size in GISTs, it is still unstable and cannot avoid the boundary information losing during down-sampling. So it is difficult to segment GIST with ambiguous boundary and great noise as before. This is because in the encoder part causing spatial information loss. Thus, we propose a RBSU-net to solve cases with ambiguous boundary. Due to the information loss in such deep network, we add an auxiliary boundary supervision in the first up-sampling block of RU-net to supervise the network deeply, as shown in Figure 6. So the architecture is similar to the RU-net. The encoder part contains 5 convolutional blocks with max pooling and each block includes 1~3 convolutional layers. The auxiliary boundary supervision is added to reinforce the boundary in the early stage to give accurate contextual features for the higher resolution layer. The loss is:

$$\text{Loss}_2 (A; W_1, W_2) = \text{Loss}_{E_2} (A; W_1) + \text{Loss}_{R_2} (A; W_2)$$

The overview of the proposed network can be seen in Figure 3. To deal with the diverse size of GISTs, we add a multi-task RU-net in the front of the RBSU-net, which offers a good likelihood maps in both region and boundary of GISTs. Then the outputs from two branches are concatenated towards the boundary definitions whereas layers R1 and R2 are focused on the definition of the regions belonging to the tumor. Therefore, the total loss in our proposed network is:

$$L_{\text{total}} = \sum_{k=1}^{2} \text{Loss}_k (A; W)$$

$$\text{Loss}_k (A; W) = \text{Loss}_{E_k} (A; W) + \text{Loss}_{R_k} (A; W)$$

D. DATA PREPROCESSING

Normalization and data augmentation are the two parts of the data preprocessing. All of the data are normalized from 0 to 1. The datasheet has been augmented to ten times by rotation and folding. All the images are resized into 256$\times$256.

### Table 1. Summary of datasheet in segmentation.

|                | Number of patients | Number of images |
|----------------|--------------------|------------------|
| Training set   | 625                | 1252             |
| Validation set | 5                  | 8                |
| Testing set    | 275                | 545              |
| Total          | 905                | 1805             |

E. IMPLEMENTATION DETAILS

The MRBSU-net is trained by an optimizer with learning rate $= 10^{-4}$, batch size $= 8$ and epoch $= 200$. All the kernels are initialized randomly from 0 to 1.

III. MATERIALS AND EXPERIMENTS

A. MATERIALS

In this study, 905 patients went through biopsy and gastroscopy to detect the GIST from 2008 to 2018 in 19 different hospitals [2]. G-EUS images were collected from one of following ultrasound machines, Olympus EU-ME1, Olympus EU-ME2, Fuji SU9000, Fuji SU8000 and Olympus Alpha 10. All patients had provided written informed consent for imaging and clinical data to be donated for the research [2].

Basic information of all patients was recorded in the information center. The ages of patients [534 females (1062 images), 371 males (743 images)] are in scope of 21 to 87 years old (mean age $57.8 \pm 10.9$ years). The datasheet contains 905 patients (1805 images). And all patients have only one tumor in our datasheet. Images collected from April in 2010 to December 2016 are selected as the training set including 625 patients (1252 images), 5 patients (8 images) as the validation set and images collected in range of January in 2017 to November in 2018 as the individual testing set consisting of 275 patients (545 images). A summary of the patient characteristics is shown in TABLE 1.

B. EVALUATION INDEX

To evaluate the proposed approach, 545 test images are used. The Dice similarity coefficient (DSC), mean absolute deviation (MAD), and Hausdorff distance (HD) are applied to evaluate the results of the proposed method [15].

DSC is a statistic used to gauge the similarity of two samples, measuring the overlapped areas between segmentation result and the ground truth, as:

$$DSC (A, B) = \frac{2 \times n (\Omega_A \cap \Omega_B)}{n (\Omega_A) + n (\Omega_B)}$$

where, $n(.)$ is a total pixels in the area. A is the segmentation results, while B is ground truth. $\Omega_A$ and $\Omega_B$ are the segmented area of A and B.

The MAD and HD are defined as the average and maximum of the surface distance errors (SDE) among the
TABLE 2. Summary of patient characteristics in different set in medical application step.

| Risk level | Training Patients(images) | Validation Patients(images) | Testing Patients(images) |
|------------|---------------------------|----------------------------|-------------------------|
| The LRG    | 315(672)                  | 17(17)                     | 81(167)                 |
| Low        | 247(483)                  | 13(13)                     | 64(142)                 |
| The HRG    | 74(152)                   | 18(18)                     | 22(49)                  |
| High       | 38(64)                    | 6(6)                       | 10(22)                  |

whole field, respectively, as:

\[
MAD(X, Y) = \frac{1}{2} \left\{ \frac{1}{N_x} \sum_{x \in X} d(x, Y) + \frac{1}{N_y} \sum_{y \in Y} d(y, X) \right\}
\]

\[
HD(X, Y) = \max \left\{ \frac{\max_{x \in X} d(x, Y)}{x \in X}, \frac{\max_{y \in Y} d(y, X)}{y \in Y} \right\}
\]

where, \(x\) and \(y\) represent point in the regions \(X\) and \(Y\); \(d(x, Y)\) is the minimum distance from point \(x\) to the region \(Y\); \(d(y, X)\) is the minimum distance from point \(y\) to the region \(X\) and \(N_x\) and \(N_y\) are the size of regions \(X\) and \(Y\), respectively.

C. SEGMENTATION EXPERIMENTS ON DIFFERENT IMPROVEMENTS AND METHODS

To illustrate the effects on each improvement and assess the segmentation quality, the following improvement methods and other deep learning based methods are compared:

1. U-net
2. Refined U-net (RU-net)
3. RU-net with multi-task contextual features (multi-task RU-net)
4. RBSU-net
5. Generative adversarial network (GAN) [20]
6. Deep attentional features (DAF) [21]

All experiments are implemented in MATLAB R2015b and Python 3.5 on a 3.06-GHz Intel(R) Xeon(R) CPU and an Nvidia Titan Xp GPU. The U-net, refined U-net, refined U-net with boundary supervision in the first up-sampling block, BSU-net and RBSU-net are trained and tested in TensorFlow.

D. CLINICAL GISTS CLASSIFICATION EXPERIMENTS

To verify the practicability of the proposed method, we classify the GISTS under the segmentation results of the proposed model. Differentiating HRG GISTS from LRG GISTS has great significance. This part illustrates the influence of the segmentation method on clinical diagnosis. 439 high-throughput GISTS features are designed to demonstrate characteristics of GISTS in EUS images [2]. The tumor boundaries are defined and then the features are extracted. Hence, the all these depends on the segmentation result. In this part, we extracted features from the outputs of the proposed method. A least absolute shrinkage and selection operator (LASSO) model is still the selection method to select features. After that, a random forest (RF) classifier is applied to differentiate the HRG cases from the LRG ones. Cross-validation was included in the training stage. Due to the unbalanced datasheet, we chose 30 LRG cases and 24 HRG cases randomly as validation set, while others were set as training set. And the experiment was bootstrapped for 10 times. More details can be referred to previous work [4].

In this step, from each hospital, about three-fourths patients of each of four risk levels are still selected as the training set [2]. The datasheet is divided into another 3 parts, as shown in TABLE 2: 1) Training set including 674 patients (1371 images), 2) Validation set including 54 patients (54 images) and 3) Testing set including 177 patients (380 images).

IV. RESULTS

A. OVERVIEW

Figure 7. illustrates the loss of training set and validation set against the training epochs.

The model starts to be stable with a low loss after 200 epochs in Figure 7.

TABLE 3. shows the evaluation indices of the proposed MRBSU-net and other compared methods. U-net, the state-of-the-art network, performs best among other traditional deep learning methods.

In TABLE 3. results from the proposed method achieves a mean DSC of 0.92±0.056, a mean HD of 11.72±10.73 pixels and a mean MAD of 3.14±2.37 pixels. All three metrics are the highest among other comparative method. Results from the multi-task RU-net achieves a mean DSC of 0.88±0.15, a mean HD of 15.89±18.83 pixels and a mean MAD of 5.12±11.24 pixels, while with respect to the RBSU-net.
TABLE 3. Comparative results of the proposed method and other methods on testing set.

| Method          | U-net[14]       | RU-net          | Multi-task RU-net | RBSU-net        | MRBSU-net       |
|-----------------|-----------------|-----------------|-------------------|----------------|-----------------|
| DSC             | 0.71±0.28       | 0.82±0.24       | 0.88±0.15         | 0.85±0.21      | 0.92±0.056      |
| HD (pixels)     | 37.41±34.33     | 21.84±26.21     | 15.89±18.83       | 14.73±24.69    | 11.72±10.73     |
| MAD (pixels)    | 13.80±20.81     | 7.81±14.73      | 5.12±11.24        | 6.56±14.14     | 3.14±2.37       |
| T-test on DSC   | 16.54           | 9.76            | 5.83              | 6.81           | ----            |
| P value on DSC  | 4.6*10^{-5}     | 1.19*10^{-21}   | 7.08*10^{-9}      | 1.55*10^{-11}  | ----            |
| Testing time consuming (s/image) | 0.060 | 0.042 | 0.044 | 0.042 | 0.090 |
| Miss detection  | 157             | 75              | 34                | 33             | 6               |

a mean DSC of 0.85±0.21, a mean HD of 14.73±24.69 pixels and a mean MAD of 6.56±14.14 pixels are achieved. Also, all of P values reach at a magnitude of 10^{-10}. These details in TABLE 3. show the significant improvement of multi-task enhances the DSC which focus on the overall segmentation performance and the improvement of boundary supervision devotes to find the region of interest. Time consuming in TABLE 3. shows that the proposed model cost only 0.09 second per image increasing in testing stage. Hence, we concatenate two branches from multi-task RU-net and feed them to the RBSU-net.

B. EFFECTS OF THE CONCATENATION AND BOUNDARY-SUPERVISION

TABLE 3. demonstrates the metrics of the proposed model and other improved methods. The results from RU-net show that the DSC increase by 10% as well as HD and MAD decrease by 15.57 and 5.99, respectively. Compared with RU-net, results from multi-task RU-net and RBSU-net have different degrees of improvements in three metrics. DSC from multi-task RU-net rises by 6%, while the one from RBSU-net increases by 3%. HDs from two methods decrease in different degrees. It reduces by 7.11 in RBSU-net and decreases slightly worse in multi-task RU-net, about 5.95. While the improvement of MAD in multi-task RU-net boost more than that in RBSU-net. MAD from multi-task RU-net drops by 2.69 and that from RBSU-net reduces by 1.25. MRBSU-net who combines multi-task RU-net with RSBU-net performs best with a mean DSC of 0.92±0.056, a mean HD of 11.72±10.73 pixels and a mean MAD of 3.14±2.37 pixels.

Figure 8. and Figure 9. shows the five examples on small and normal size respectively. Figure 8. and Figure 9. illustrate the segmentation results of the proposed method and other compared improved methods, which shows the proposed method performs best among all the methods. Obviously, shadows of outputs from RU-net are cut down, but misjudgment still exists in RU-net (Figure 9.). Figure 8. shows that those methods are still unstable in small size tumors.

To illustrate the effects of the concatenation and boundary-supervision, we list the outputs of each up-sampling block in the RU-net, RU-net advanced, RBSU-net and MRBSU-net (Figure 10.). The case that cannot be accurately deal with RU-net (Figure 10.) shows multi-task RU-net has eliminated the shadow in the second up-sampling block, while RBSU-net could eliminate the shadow in the first up-sampling block. The combination of multi-task RU-net and RBSU-net could increase the stability of the network. What’s more, we also show the outputs of layer E1, layer R1 and layer R2 in the proposed MRBSU-net (Figure 11.) to demonstrate the effects of concatenation. We could see the heat of misjudgment part climbs up quickly to a background level, proving that multi-task RU-net also works on deleting shadows. All the outputs are presented in heat map to make the effects clearer.

C. SEGMENTATION RESULTS ON DIFFERENT METHODS

TABLE 3. demonstrates the metrics of the proposed model and other deep learning based methods. These results show the proposed method improves in all aspects including DSC, HD and MAD. The GAN method has a DSC of 0.72±0.32, an HD of 39.28±38.94 pixels, and an MAD of 15.96±25.74 pixels as well as the DAF method has a DSC of 0.68±0.31, an HD of 43.29±39.86 pixels, and an MAD of 16.94±25.45 pixels which demonstrates a statistically significant improvement over the GAN, DAF in all three metrics.

D. CLINICAL GISTs CLASSIFICATION RESULTS

We distinct 32 HRG cases from 145 LRG cases. This part compares the results segmented by the proposed method and those segmented manually. The verification experiment on the proposed method uses the same features and classifier as manual segmentation. 10 features are remained after feature selection. In the RF classifier, accuracy, the area under the receiver operating characteristic (ROC) curve (AUC), sensitivity, and specificity reach at 0.791, 0.808, 0.750, 0.800, respectively. The classification results are shown in TABLE 5. And the ROC of the two segmentation was shown in Figure 12. Although the metrics are still lower than those using manual segmentation, these results could indicate this method could be useful as a starting point for manual segmentation for reducing effort from radiologists. The classification results show the proposed method is of diagnostic values.

V. DISCUSSION

We propose a novel MRBSU-net method to segment GISTs automatically in EUS images, connecting a multi-task RU-net to a RBSU-net. Concatenated input and boundary supervision play an important role in performance improvement, especially in images with heavy shadows. Moreover, the clinical application also performs well.
A. OVERVIEW

Figure 7 shows the 195th epoch is the start of the stable epoch. We stop to train the model after 5 more epochs to make sure the stability. TABLE 4 shows the performance of three deep learning-based methods on GISTs segmentation. Although DSC of U-net is a little lower than that of GAN [20], the other two metrics of U-net, HD and MAD, perform much better than those of GAN. And all metrics in U-net and GAN are better than those in DAF [21]. These results prove U-net does perform well on medical images, providing theoretical basis for improvement of U-net.

The RU-net illustrates an obvious improvement from U-net in Figure 8 and Figure 9. It eliminates the interferences of shadows. Compared with U-net, the performance of RU-net shows that decrease of the convolutional layer could refine boundary, making it smoother. Attributing to its plain network architecture, U-net does not perform well in the scope of the EUS image segmentation. Compared with RU-net, the down-sampling path in U-net is not deep enough to extract features and the up-sampling path in U-net has too many channels to propagate the noise information in EUS images. However, RU-net is not stable in small size tumor (Figure 8).

FIGURE 8. Five small size examples and segmentation results of the proposed MRBSU-net and other methods.
For those small size tumors, RU-net cannot differentiate tumors from background. For the image with heavy shadow, RU-net cannot distinguish the tumor from heavy noise. Obviously, multi-task RU-net and RBSU-net have a certain effect on small size tumors due to the boundary supervision in loss function, as shown in Figure 8. Adding boundary cross entropy in loss function of multi-task RU-net boosts the influence of small size tumors, if they are redeemed as background. Compared with the weight in region cross entropy, that in boundary cross region is much higher, which could differentiate the small size tumors from background. Moreover, it could eliminate the small false positives, as shown in Figure 9. The boundary supervision in RBSU-net leads the network focus on finding boundary in the down-sampling part and segmenting region on the up-sampling path. Hence, it does work on large false positives, as shown in Figure 9. However, they are still not stable on small size tumors. The fifth example in Figure 8 illustrates a traditional case of over-supervision by RBSU-net, while RU-net redeem the noise nearby contour as the tumor.
The proposed method concatenates two branches from multi-task RU-net and feeds them to the RBSU-net, on the basis of empirical evidence in Figure 8., to increase the stability of the segmentation on small size tumors and to solve this oversupervision. Column 5 in Figure 9. shows RBSU-net does not work on tumors with noise inside, especially those noise similar to boundary. These explain the complementary effects on the combination of multi-task RU-net and RBSU-net from another respects.

Figure 8. and Figure 9. show the segmentation results of five examples of small and normal size GISTs. The segmentation performance of tumor with clear boundary is almost similar among five methods. However, in the other cases, results are relatively diverse. The MRBSU-net is more sensitive to the intensity diversity adjoining the contour, ascribing to localization information learned by MRBSU-net. Hence, the proposed model has the best segmentation accuracy and robustness among all the methods.

**B. EFFECTS OF THE CONCATENATION AND BOUNDARY-SUPERVISION**

To illustrate the effects of the concatenation, we calculate the DSC of the RBSU-net and MRBSU-net, listing in TABLE 3. These results indicate that the outputs with the
TABLE 4. Comparative results of the proposed method and other familiar methods.

| Method      | GAN[20]  | DAF[21] | U-net[14] | MRBSU-net |
|-------------|----------|---------|-----------|-----------|
| DSC         | 0.72±0.32| 0.68±0.31| 0.73±0.28 | 0.92±0.056|
| HD (pixels) | 39.28±38.94| 42.29±39.86| 37.41±34.33 | 11.72±10.73|
| MAD (pixels)| 15.96±25.74| 16.94±25.45| 13.80±20.81 | 3.16±2.37 |
| T-test on DSC| 14.15 | 17.87 | 16.54 | --- |
| P-value on DSC | 6.73*10^{-4} | 2.41*10^{-4} | 4.61*10^{-4} | --- |
| Time consuming (s/image) | 0.66 | 0.15 | 0.06 | 0.09 |
| Miss detection | 155 | 192 | 157 | 6 |

TABLE 5. Classification results using different methods.

| Method      | AUC     | Accuracy | Sensitivity | Specificity |
|-------------|---------|----------|-------------|-------------|
| Training set | 0.854±0.01 | 0.801±0.01 | 0.823±0.04 | 0.803±0.015 |
| Manual segmentation | 6 | 4 | 6 | |
| RBSU-net | 0.819±0.01 | 0.775±0.01 | 0.742±0.03 | 0.741±0.018 |
| Testing set | 0.879 | 0.822 | 0.853 | 0.814 |
| Manual segmentation | 4 | 4 | 5 | |
| MRBSU-net | 0.808 | 0.791 | 0.750 | 0.800 |

assistance of the multichannel is much closer to the ground truth. Also, number of miss detection, T-test and P-value in TABLE 3. shows the significant improvements of the proposed model. The contrast experiment of testing per image among all methods in TABLE 3. and TABLE 4. illustrates the proposed method achieves a better performance in a low time cost. And this lays the foundation of being applied into practice. The improvement of multichannel can be seen clearly in Figure 10. The outputs of the second up-sampling block in MRBSU-net has already found the location of the tumor, while RBSU-net has not. These improvements are also proved in the classification experiment, as shown in TABLE 5. In the classification experiment, AUC, accuracy, sensitivity and specificity of the MRBSU-net method achieves a higher quantitative value, showing that the proposed method could be used in clinic. These all proves the effects of the concatenation. Although the results from automatic segmentation method are slightly worse than those from manual segmentation, it still achieves clinical application requirements. Compared to the tedious and time-consuming manual segmentation by radiologists, the automatic segmentation method does have the significance influence in practice.

In TABLE 3., we could see the results of RBSU-net are all higher than those of RU-net which shows the effects of the boundary-supervision. Figure 11. shows outputs of three important nodes, two outputs from the branches of the multi-task RU-net and final outputs in heat map. These figures show how the MRBSU-net method works and effects of every improvement.

C. CLINICAL GISTS CLASSIFICATION APPLICATIONS

We are glad to see that the classification based on MRBSU-net segmentation performs well on such unbalanced datasheet. Notwithstanding the metrics are still lower than those using manual segmentation [2] because of several miss-detected cases, as shown in TABLE 3. and TABLE 4. Experiments demonstrate that the proposed method produces good diagnostic performances in the classification of GIST risk levels. It has been proved that the method could be used in practice which means doctors could abandon the tedious manual segmentation part to a valuable reference by the proposed method. And solving these extreme cases will become our next challenge.

VI. CONCLUSION

In this study, we propose an automatic segmentation method in EUS images, combining a multi-task RU-net with a RBSU-net, named as MRBSU-net. Compared with U-net and other improved U-nets, the MRBSU-net has the best performance in all quantitative indices, DSC, MAD and HD. Multi-task RU-net solves the various sizes in GISTs by intensifying the influence of loss function in small size tumor. Concatenation provides an initial boundary and initial tumor region, which is more efficient than single channel input. By minimizing the boundary-supervision energy function, the network focuses on finding the interest region. Finally, the network performs better and more stable in those cases with ambiguous boundary, shadows and variation in size by making use of multi-task RU-net and RBSU-net. Moreover, the segmentation results get a good hand in tumor classification experiment.

However, in some extreme cases, such as tumor with complicated internal texture structure, the proposed method cannot segment well. Our future work will focus on improving the structure of the deep learning network to handle all GIST cases in EUS.

ACKNOWLEDGMENT

The data in this work was supported by Beijing Friendship Hospital, Capital Medical University, Tongji Hospital, Tongji Medical College of Hust, Tongji Medical College Huazhong University of Science and Technology, Union Hospital, Tongji Medical College Huazhong University of Science & Technology, Renmin Hospital of Wuhan University, Qilu Hospital of Shandong University, West China Hospital, Sichuan University, The Third Xiangya Hospital of Central South University, Renji Hospital, Shanghai Jiaotong...
University of Medicine, Shanghai General Hospital, Xinhua Hospital Affiliated to Shanghai JiaoTong University of Medicine, The First Hospital of Zhejiang Province, Affiliated Hangzhou First People’s Hospital, Zhejiang University School of Medicine, The First Affiliated Hospital, Sun Yat-sen University, Nanfang Hospital, Zhongda Hospital of Southeast University, Changhai Hospital, PLA Army General Hospital, Nanjing Drum Towel Hospital, The Affiliated Hospital of Nanjing Medical School and Jiangsu Province Hospital.

REFERENCES

[1] V. T. DeVita, T. S. Lawrence, and S. A. Rosenberg, DeVita, Hellman, and Rosenberg’s Cancer: Principles & Practice of Oncology, 10th ed. Alphen aan den Rijn, The Netherlands: Wolters Kluwer, 2008.
[2] X. Li, F. Jiang, Y. Guo, Z. Jin, and Y. Wang, “Computer-aided diagnosis of gastrointestinal stromal tumors: A radiomics method on endoscopic ultrasound image,” Int. J. Comput. Assist. Radiol. Surg., vol. 14, no. 10, pp. 1635–1645, May 2019.
[3] P. N. T. Wells and M. Halliwell, “Speckle in ultrasonic imaging,” Ultrasonics, vol. 19, no. 5, pp. 225–229, Sep. 1981.
[4] R. M. Kirbyber, “Imaging artifacts in diagnostic ultrasound—A review,” Veterinary Radiol. Ultrason., vol. 36, no. 4, pp. 297–306, Jul. 1995.
[5] J. Ker, L. Wang, J. Rao, and T. Lim, “Deep learning applications in medical image analysis,” IEEE Access, vol. 6, pp. 9375–9389, 2018.
[6] H. Greenspan, B. Van Ginneken, and R. M. Summers, “Guest editorial on deep learning in medical imaging: Overview and future promise of an exciting new technique,” IEEE Trans. Med. Imaging, vol. 35, no. 5, pp. 1153–1159, May 2016.
[7] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, May 2015.
[8] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Noguez, J. Yao, D. Mollura, and R. M. Summers, “Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning,” IEEE Trans. Med. Imag., vol. 35, no. 5, pp. 1285–1298, May 2016.
[9] J. Ker, T. Kooi, B. E. Bejnordi, A. A. Setio, F. Cionnpi, M. Ghaffoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sanchez, “A survey on deep learning in medical image analysis,” Med. Image Anal., vol. 42, pp. 60–88, Dec. 2017.
[10] J. Minnema, M. Van Eijnatten, W. Krouw, F. Diblen, A. Mendrik, and J. Wolff, “CT image segmentation of bone for medical additive manufacturing using a convolutional neural network,” Comput. Biol. Med., vol. 103, pp. 130–139, Dec. 2018.
[11] D. C. Cirean, A. Giusti, L. M. Gambardella, and J. Schmidhuber, “Deep neural networks segment neuronal membranes in electron microscopy images,” in Proc. 25th Int. Conf. Neural Inf. Process. Syst. (NIPS), vol. 25, Jan. 2012, pp. 2852–2860.
[12] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in Proc. IEEE Trans. Pattern Anal., Boston, MA, USA, Jun. 2015, pp. 640–651.
[13] O. Ronneberger, F. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in Proc. Int. Conf. Med. Imag. Comput. Assist. Intervent., 2015, pp. 234–241.
[14] Y. Hu, Y. Guo, Y. Wang, J. Yu, J. Li, S. Zhou, and C. Chang, “Automatic tumor segmentation in breast ultrasound images using a dilated fully convolutional network combined with an active contour model,” Med. Phys., vol. 46, no. 1, pp. 215–228, Jan. 2019.
[15] H. Chen, X. Qi, L. Yu, and P.-A. Heng, “DCAN: Deep contour-aware networks for accurate gland segmentation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2016, pp. 2487–2496.
[16] L.-C. Chen, J. T. Barron, G. Papandreou, K. Murphy, and A. L. Yuille, “Semantic image segmentation with task-specific edge detection using CNNs and a discriminatively trained domain transform,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2016, pp. 4545–4554.
[17] I. Kokkinos, “Pushing the boundaries of boundary detection using deep learning,” in Proc. Int. Conf. Learn. Represent., 2016. [Online]. Available: https://arxiv.org/pdf/1511.07386.pdf

[18] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in Proc. Int. Conf. Learn. Represent., 2014.
[19] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, “Semantic segmentation using adversarial networks,” in Proc. NIPS, Barcelona, Spain, 2016.
[20] Y. Wang, Z. Deng, X. Hu, L. Zhu, X. Yang, X. Xu, F. Heng, and D. Ni, “Deep attentional features for prostate segmentation in ultrasound,” in Proc. Conf. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI), 2018, pp. 523–530.

XINYI LI is currently pursuing the master’s degree with the Department of Electronic Engineering, Fudan University. Her research is focus on the medical image processing. She works together with doctors and satisfy their clinical demands.

YI GUO received the Ph.D. degree in biomedical engineering from Fudan University, Shanghai, China, in 2013. She is currently an Associate Professor with the Department of Electronic Engineering, Fudan University. Her current research interests include medical signal and image processing.

FEI JIANG received the M.D. and Ph.D. degrees from Second Military Medical University, Shanghai, China, in 2011. She is currently an Associate Chief Physician with Shanghai Changhai Hospital. She is also focusing on the research of endoscopic ultrasonography.

LEIMING XU received the Ph.D. degree in medical from Second Military Medical University, Shanghai, China, in 2012. He is currently a Professor with the Department of Gastroenterology, Shanghai Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine. His research interests include endoscopy and digestive diseases.

FENG SHEN received the Ph.D. degree in medical from the Shanghai Jiaotong University School of Medicine, Shanghai, China, in 2014. He is currently an Associate Professor with the Department of Gastroenterology, Shanghai Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine. His research interests include endoscopy and digestive diseases.
ZHENDONG JIN received the master’s degree from First Military Medical University, in 1991. He is currently a Professor and the Chief Physician of the Shanghai Changhai Hospital. He is also the Ph.D. Supervisor of Second Military Medical University. He is also focusing on the research of application of endoscopic ultrasonography in digestive diseases.

YUANYUAN WANG received the B.Sc., M.Sc., and Ph.D. degrees in electronic engineering from Fudan University, Shanghai, China, in 1990, 1992, and 1994, respectively. From 1994 to 1996, he was a Postdoctoral Research Fellow with the School of Electronic Engineering and Computer Science, University of Wales, Bangor, U.K. In May 1996, he went back to the Department of Electronic Engineering, Fudan University, as an Associate Professor. He was then promoted to a Full Professor, in May 1998. He is currently the Director of the Biomedical Engineering Center, Fudan University. He is also the author or coauthor of six books and 500 research articles. His research interests include medical ultrasound techniques and medical image processing.

* * *