Surface-GCN:
Learning Interaction Experience for Organ Segmentation in 3D Medical Images

Fengrui Tian[a,b], Zhiqiang Tian[a], Zhang Chen[a], Dong Zhang[b,c], Shaoyi Du[b]

[a]School of Software Engineering, Xi'an Jiaotong University, Xi'an 710049, China
[b]Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710049, China
[c]School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an 710049, China

Version typeset February 4, 2023

Corresponding author: Zhiqiang Tian and Shaoyi Du. Email: zhiqiangtian@xjtu.edu.cn, dushaoyi@gmail.com

Abstract

Background: Accurate segmentation of organs has a great significance for clinical diagnosis, but it is still hard work due to the obscure imaging boundaries caused by tissue adhesion on medical images. Based on the image continuity in medical image volumes, segmentation on these slices could be inferred from adjacent slices with a clear organ boundary. Radiologists can delineate a clear organ boundary by observing adjacent slices.

Purpose: Inspired by the radiologists' delineating procedure, we design an organ segmentation model based on boundary information of adjacent slices and a human-machine interactive learning strategy to introduce clinical experience.

Methods: We propose an interactive organ segmentation method for medical image volume based on Graph Convolution Network (GCN) called Surface-GCN. First, we propose a Surface Feature Extraction Network (SFE-Net) to capture surface features of a target organ, and supervise it by a Mini-batch Adaptive Surface Matching (MBASM) module. Then, to predict organ boundaries precisely, we design an automatic segmentation module based on a Surface Convolution Unit (SCU), which propagates information on organ surfaces to refine the generated boundaries. In addition, an interactive segmentation module is proposed to learn radiologists' experience of interactive corrections on organ surfaces to reduce interaction clicks.

Results: We evaluate the proposed method on one prostate MR image dataset and two abdominal multi-organ CT datasets. The experimental results show that our method outperforms other state-of-the-art methods. For prostate segmentation, the proposed method conducts a DSC score of 94.49% on PROMISE12 test dataset. For abdominal multi-organ segmentation, the proposed method achieves DSC scores of 95%, 91%, 95%, and 88% for the left kidney, gallbladder, spleen, and esophagus respectively. For interactive segmentation, the proposed method reduces 5-10 interaction clicks to reach

i

the same accuracy.

Conclusions: To overcome the medical organ segmentation challenge, we propose a Graph Convolutional Network called Surface-GCN by imitating radiologist interactions and learning clinical experience. On single and multiple organ segmentation tasks, the proposed method could obtain more accurate segmentation boundaries compared with other state-of-the-art methods.

Contents
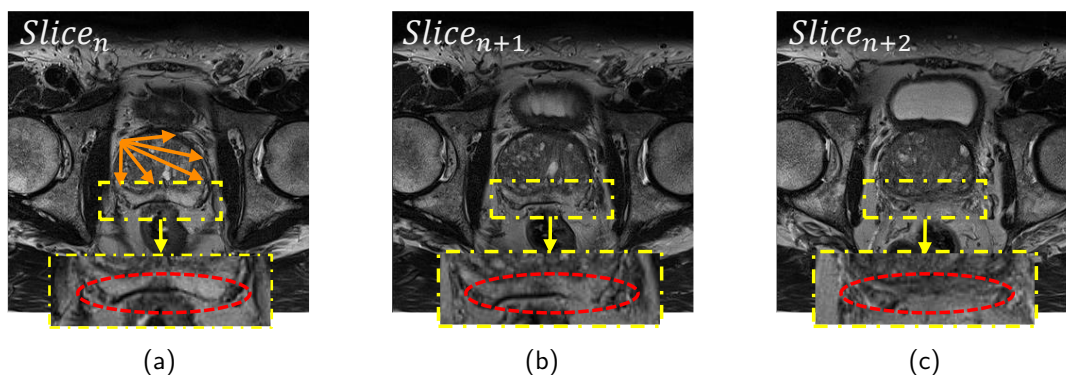
iii

I. Introduction

Figure 1: Intra-slice and inter-slice dependencies in prostate MR images. Intra-slice dependencies are pixel relationships in one MR image slice (like the orange arrow in (a)), and inter-slice dependencies are pixel relationships among continuous MR slices (like the red ellipses in (b) and (c)), both of which could contribute to prostate boundary prediction in some low-contrast regions or obscure organ boundaries.

Organ segmentation is a vital method in medical image analysis for many applications. For example, prostate Magnetic Resonance (MR) image segmentation is an important step in the treatment of prostate disease, especially for prostate cancer, which is a general type of cancer in human[1,2]. Accurate Computed Tomography (CT) image segmentation of abdominal multi-organ is very significant for computer-assisted diagnosis systems that aim to assist radiologists for lesion diagnosis and treatment delivery[3]. In practice, manually delineating target organs on medical images is a tedious and time-consuming task[4]. Thus automatic methods for organ segmentation are studied over the past decades[5,6]. Recently, Convolutional Neural Networks (CNNs) based deep learning technologies have reached significant progress in many research areas[7], particularly in computer vision[8,9,10]. In medical image analysis, CNNs have shown an impressive performance on medical image segmentation tasks. U-Net[11], proposed by Ronneberger *et al.*, V-Net[12], proposed by Milletari *et al.* and their variations[13,14,15,16] are the most representative ones and have a wide application in many segmentation fields. One of the reasons for the success of these methods is the skip connections in their models that integrates the features maps learned from different resolution levels. Thus both low-level and high-level information of the target could be captured in concert.

Although segmentation performance has been improved by previous CNN-based methods, there are many difficulties in establishing an accurate medical organ segmentation

scheme. The boundary between different organs and their surrounding tissues is obscure and low contrast. Confusing background textures and the variations on the shape or size of target organs could make a challenge for achieving accurate segmentation. A concrete example is shown in Fig. 1c. The prostate boundary circled by the red dotted ellipse is obscure. In contrast, the boundary circled in Fig. 1a is relatively clear. It is believed that the boundary information from neighboring slices, such as the red dotted ellipses in Fig. 1a and Fig. 1b, contributes to predicting the obscure organ boundary.

Based on the above observation, we classify and differentiate two types of organ boundary dependencies in medical volume: intra-slice dependency and inter-slice dependency. The former is the correlation of pixels within the identical slice as the orange arrow shown in Fig. 1a. In contrast, inter-slice dependency is the correlation of pixels crossing neighboring slices as red dotted ellipses shown in Fig. 1. How to design a neural network to capture both dependencies effectively remains a challenging problem. In traditional CNN-based segmentation methods, increasing the size of the convolution kernel in each slice or upgrading the convolution kernel from 2 dimensions to 3 dimensions may indirectly improve the capability of capturing intra-slice dependency or inter-slice dependency. However, due to complicated organ variations on shape or size, regularly square or cube convolution kernel may not utilize those details for segmentation in a good manner.

To overcome the above-mentioned challenges, in this study, the organ segmentation is considered as a regression problem, and a Graph Convolutional Network (GCN)[17] based segmentation method is designed. Organs are mostly connected in space, *i.e.,* the organ boundary in medical volume could be built upon a closed surface. In each slice of medical volume, it is supposed that the organ boundary could be represented by a closed curve composed of a series of control vertices. Hence, segmenting organs is to predict the vertex positions of the closed curve in each slice.

Although above-mentioned methods could build a segmentation network that maximizes the use of organ boundary information, it is difficult to build a fully automatic neural network with the lack of training data. Therefore, many researchers focus on interactive image segmentation methods[18,19,20,21,22,23]. An obvious advantage of the interactive segmentation method compared to the fully automatic method is that it could achieve better segmentation performance with the guidance of human interaction. However, above-mentioned methods

I.   INTRODUCTION

use human annotation to fine-tune one image. We design an interactive segmentation method that refines organ segmentation surface in medical volume. In this way, the radiologists' experience could be further utilized to reach accurate segmentation result.

As mentioned above, we propose an interactive organ segmentation method in medical images called Surface-GCN. In order to simulate the information propagation of the two dependencies, a new graph convolution operation unit is designed to take advantage of the shape of the target organ called Surface Convolution Unit (SCU). We proposed the Surface Feature Extraction Network (SFE-Net) to capture the surface features of each organ. Then, the proposed SFE-Net is supervised by the proposed Mini-batch Adaptive Surface Matching (MBASM) module as an encoder to extract surface features of the organ. Our Surface-GCN architecture based on SCU is proposed to predict organ surface in the medical volume simultaneously using the extracted features. The predicted organ segmentations can be optimized through Surface-GCN by imitating radiologists' clicks on the control points. The entire network is trained in an end-to-end manner.

The rest of this paper is organized as follows. In Sec. II., our segmentation method is introduced followed by the details of each part. In Sec. III., Experiments on several datasets are discussed to demonstrate the superiority of our method. Sec. IV. concludes this paper.

II.   Methods
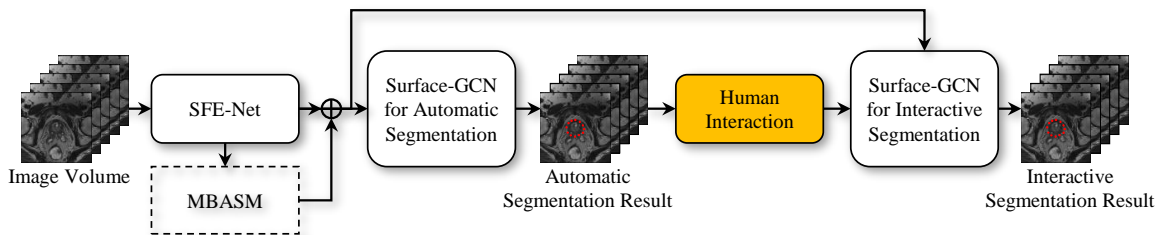
Figure 2: The overview of the proposed model. First, the proposed Surface Feature Extraction Network (SFE-Net) is introduced to extract the features of the target organ, which is supervised by the Mini-batch Adaptive Surface Matching (MBASM) module. Then the proposed Surface-GCN uses the features to automatically generate the closed surface represented as the target organ boundary. Finally, the model learns experience by radiologist interactions to refine the automatic segmentation results.

Fig. 2 is an overview of the proposed Surface-GCN. First, the SFE-Net is proposed to extract the surface features of the target organ. Second, the extracted features are incorpo-

Last edited $Date$:

rated into the MBASM module that aims to help the proposed Surface-GCN model see the organ surface. Third, Surface-GCN uses the features to generate the segmentation surface of the organ. Last, human interactions are introduced to Surface-GCN as the expert experience to fine-tune organ surface.

II.A. Surface Feature Extraction Network
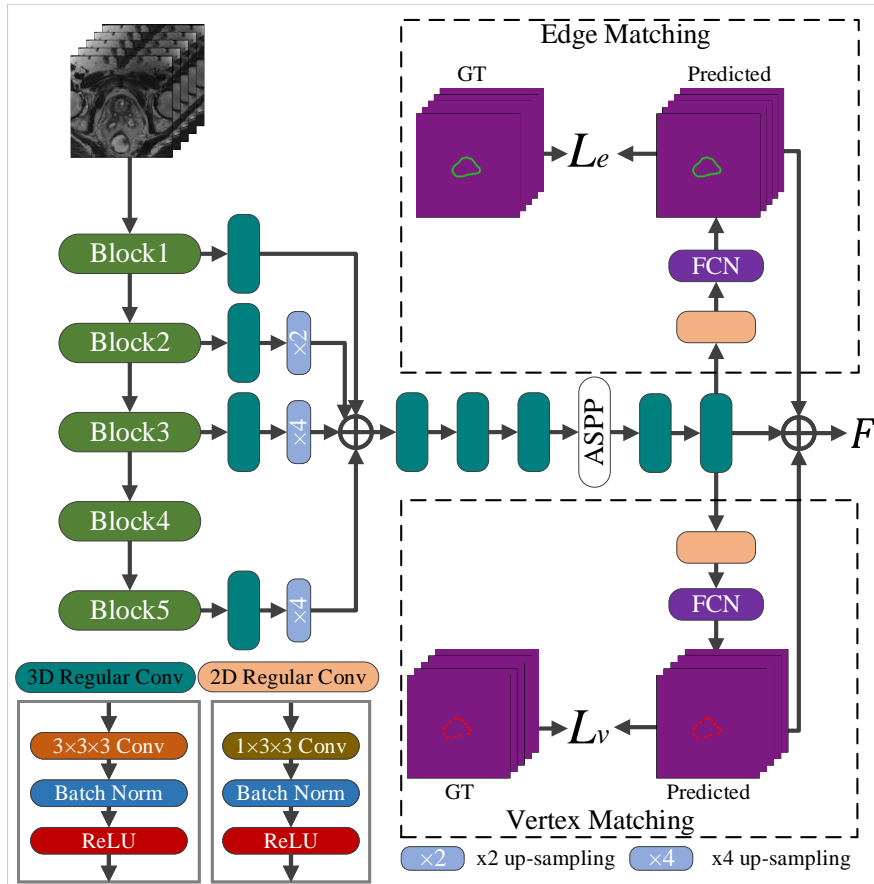
Figure 3: The proposed SFE-Net and MBASM module. Details of Blocks 1-5 are listed in Table 1. Three 3×3×3 convolution operations are used to encode organ surface features. Then the features are supervised by the MBASM module that is composed of one Edge Matching branch and one Vertex Matching branch.

To help the proposed model infer the obscure part of target organ boundary, the SFE-Net is presented to capture multi-scale feature of target organ and to enlarge the receptive field both in spatial and temporal dimension. As shown in Fig. 3, each slice cropped along the bounding box is resized into 224×224 and then fed into five convolution blocks. The details of five blocks are shown in Table 1. To increase the CNN reception field, We follow

II. METHODS   II.A. Surface Feature Extraction Network

Table 1: Backbone of SFE-Net. The proposed SFE-Nethas 5 different blocks. In the first and second block, the normal convolution is added to capture image details. Atrous convolutions are added at the Blocks 3, 4, and 5 to capture the global feature that contributes to segmentation.

Block Name	Dilation Size	101-SFE-Net	152-SFE-Net
Block1	$1\times1\times1$	$3\times7\times7$, 64	
Block2	$1\times1\times1$	$\begin{bmatrix} 1\times1\times1,\ 64 \\ 3\times3\times3,\ 64 \\ 1\times1\times1,\ 256 \end{bmatrix} \times3$	$\begin{bmatrix} 1\times1\times1,\ 64 \\ 3\times3\times3,\ 64 \\ 1\times1\times1,\ 256 \end{bmatrix} \times3$
Block3	$1\times1\times1$	$\begin{bmatrix} 1\times1\times1,\ 128 \\ 3\times3\times3,\ 128 \\ 1\times1\times1,\ 512 \end{bmatrix} \times4$	$\begin{bmatrix} 1\times1\times1,\ 128 \\ 3\times3\times3,\ 128 \\ 1\times1\times1,\ 512 \end{bmatrix} \times8$
Block4	$1\times2\times2$	$\begin{bmatrix} 1\times1\times1,\ 256 \\ 3\times3\times3,\ 256 \\ 1\times1\times1,\ 1024 \end{bmatrix} \times23$	$\begin{bmatrix} 1\times1\times1,\ 256 \\ 3\times3\times3,\ 256 \\ 1\times1\times1,\ 1024 \end{bmatrix} \times36$
Block5	$1\times4\times4$	$\begin{bmatrix} 1\times1\times1,\ 512 \\ 3\times3\times3,\ 512 \\ 1\times1\times1,\ 2048 \end{bmatrix} \times3$	$\begin{bmatrix} 1\times1\times1,\ 512 \\ 3\times3\times3,\ 512 \\ 1\times1\times1,\ 2048 \end{bmatrix} \times3$

common practice[24] that adds Atrous Convolution at Blocks 3, 4, and 5. To obtain multi-scale image features, three $3\times3\times3$ convolution operations are used to encode features from 4 convolution blocks (Block 1, 2, 3, and 5 in Fig. 3). Trilinear upsampling operations are then used to concatenate features from different sizes. The concatenated features are fed into the ASPP[24] block to obtain the final output features F_c.

II.B. Mini-batch Adaptive Surface Matching Module

We denote F_c As the feature from the CNN encoder, which is more high-level information and could not focus on boundary details. To supervise the proposed SFE-Net to learn the features of the organ surface, as shown in Fig. 3, the proposed the MBASM module consists of one Edge Matching branch and one Vertex Matching branch. Each branch consists of one $1\times3\times3$ convolution layer and one fully connected layer. The Edge Matching branch and the Vertex Matching branch using F_c predict the possibility that whether the edges and vertices of the organ surface exist at a certain resolution, *e.g.*, $N_G = 28 \times 28$ grids in each slice, respectively. 28×28 resolution is a balance between more precise details and more computation memory in practice. We actually have tried larger resolution ($N_G = 112\times112$) and found larger resolution could hardly improve segmentation performance (from 94.49% DSC score on automatic prostate segmentation to 94.47% DSC score) but need more computational

resources. However, when we use a smaller resolution ($N_G = 7 \times 7$), the performance drops drastically (from 94.49%DSC score to 92.21% DSC score). So the 28×28 resolution is a balanced point between a higher score and fewer computational resources. The vertices in the Vertex Matching branch are sampled from the organ surface. The edges in the Edge Matching branch are generated by connecting the sampled vertices. For example, given an image volume, the organ boundary is obtained from corresponding ground truths at N_G resolution. The vertices in the Vertex Matching branch are up-sampled from the organ boundary. We use the algorithm in OpenCV to get the organ boundary from the ground truth. Then we up-sample the boundary to fitting the needed resolution. The edges in the Edge Matching branch are generated by connecting the sampled vertices. Cross-entropy (CE) is used to train two branches. However, as shown in Fig. 3, the edges (green areas in the Edge Matching branch) and the vertices (red areas in the Vertex Matching branch) always occupy a small part of the entire grid. There would be a severe area imbalance between the organ boundary area and other areas. To solve this problem, the α-balanced CE is generally used.

$$L_E(\mathcal{E}, \hat{\mathcal{E}}) = \sum_{s=0}^{S} \Big(\alpha_e \hat{\mathcal{E}}^s \log(\mathcal{E}^s) + (1 - \alpha_e)(1 - \hat{\mathcal{E}}^s) \log(1 - \mathcal{E}^s) \Big), \tag{1}$$

$$L_V(\mathcal{V}, \hat{\mathcal{V}}) = \sum_{s=0}^{S} \Big(\alpha_v \hat{\mathcal{V}}^s \log(\mathcal{V}^s) + (1 - \alpha_v)(1 - \hat{\mathcal{V}}^s) \log(1 - \mathcal{V}^s) \Big), \tag{2}$$

where $\mathcal{E}^s$ denotes the predicted edges in s^{th} slice, and $\mathcal{V}^s$ denotes the predicted vertices in s^{th} slice. $\hat{\mathcal{E}}^s$ and $\hat{\mathcal{V}}^s$ represent the edges and vertices from the ground truth in s^{th} slice, respectively. α_e and α_v are the weighting factors of two branches. S is the number of slices in medical volume. Since the losses are calculated by the sum of the loss in each slice, adding the S in the denominator is a numerical process that ensures the stability of the loss. $\mathcal{V} = \{\mathcal{V}^1, \mathcal{V}^2, \cdots, \mathcal{V}^S\}$ and $\mathcal{E} = \{\mathcal{E}^1, \mathcal{E}^2, \cdots, \mathcal{E}^S\}$ denote the set of edges and vertices predicted in medical volume, respectively. $\hat{\mathcal{E}}$ and $\hat{\mathcal{V}}$ are the set of edges and vertices from the corresponding ground truth, respectively. The edges and vertices from the organ surface are predicted simultaneously.

α_e and α_v are the key parameters to solve the area imbalance problem. Its value balances the weight of the boundary area and non-boundary area. However, the shape and size of

II.   METHODS   II.B.   Mini-batch Adaptive Surface Matching Module

different organs and different parts in target organ vary greatly. So the values of α_e and α_v ought to change drastically to fit different situations. In this study, a mini-batch adaptive CE is proposed to alleviate the area imbalance problem. Concretely, α_e and α_v are calculated by the following equations.

$$\alpha_e = 1 - \frac{\sum\limits_{b=0}^{B} |\mathcal{E}_b^s|}{B \cdot S \cdot N_G}, \tag{3}$$

$$\alpha_v = 1 - \frac{\sum\limits_{b=0}^{B} |\mathcal{V}_b^s|}{B \cdot S \cdot N_G}, \tag{4}$$

where B denotes the mini-batch size. $\sum\limits_{b=0}^{B} |\mathcal{E}_b^s|$ and $\sum\limits_{b=0}^{B} |\mathcal{V}_b^s|$ are the number of the pixels of the edges and vertices at the sth slice in the current mini-batch, respectively. Then the predicted edges and vertices denoted as F_e, F_v are concatenated with F_c.

$$F = concat\{F_c, F_e, F_v\}. \tag{5}$$

The aim of Fc, Fe, Fv is to provide strong features F for the GCN module. We believe that F must meet some basic rules i.e., F must be related to the organ boundary. To achieve this, we add two branches here to supervise F. Edge and vertex are two interpretable features related to different information on organ boundaries. On the one hand, we design edge and vertex branches that use ground truths to supervise F. On the other hand, the results of two branches can be interpretable features for segmentation. The final features F are introduced to the following GCN module to predict the target organ surface. The predicted edges and vertices could help the following modules capture more organ boundary information in medical volume.

II.C. Surface-GCN for Automatic Segmentation

In this study, the convolution kernel is built directly around the organ surface. It builds the information connection through the surface, instead of traditional CNN that builds the information propagation locally without the guidance from the organ shape. The organ boundary information is propagated through the vertices on the segmentation surface. In this way, the obscure organ boundary could be inferred by the boundary information on
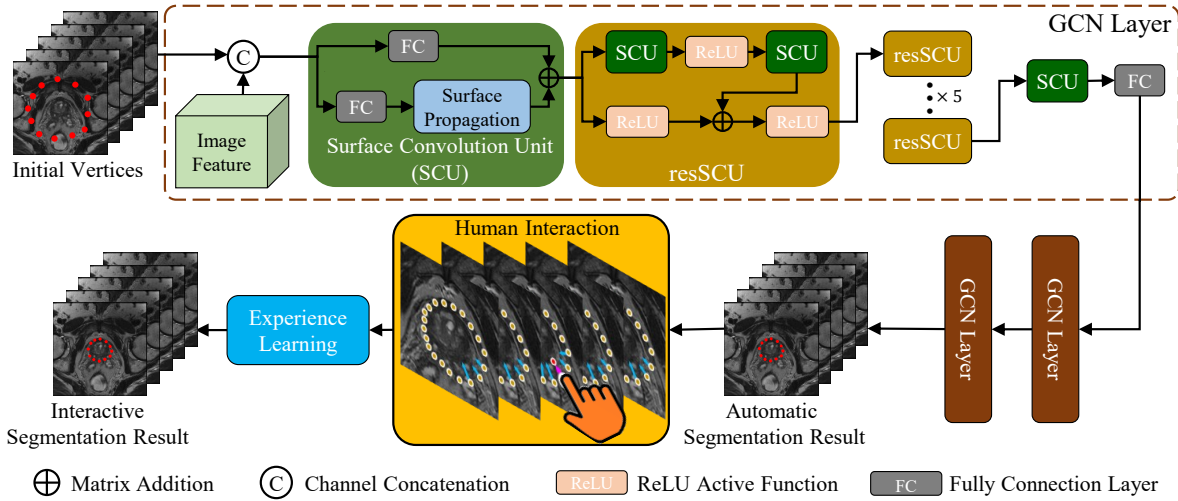
Figure 4: The overview of the proposed GCN module. We first initialize the GCN vertices along a circle at the image center. The initial vertices concatenated with the corresponding image features are fed into three GCN layers to obtain automatic segmentation result. Then the radiologist interaction experience is introduced to our module to progressively fine-tune segmentation output.

other positions. The proposed GCN module is shown in Fig. 4. In each GCN layer, the graph is defined as $G = \{V, E\}$ where V, E denote the set of vertices and edges in graph, respectively. In each slice, $N = 40$ vertices form a closed curve as the target organ boundary.

Each vertex takes one pixel in each slice and represents the image features at that pixel. Therefore, $V = \{V^1, V^2, \cdots, V^S\}$ where $V^s = \{v_1^s, v_2^s, \cdots, v_N^s\}$ denotes the set of vertices in s^{th} slice of medical volume. $(X^s, Y^s) = \{(x_1^s, y_1^s), (x_2^s, y_2^s), \cdots, (x_N^s, y_N^s)\}$ denotes the positions of vertices in V^s. Hence the positions of vertices in V are defined as $(X, Y) = \{(X^1, Y^1), (X^2, Y^2), \cdots, (X^S, Y^S)\}$. Sampled image features $F_{sampled}$, *i.e.*, the image features F sampled from the corresponding position of V, concatenated with two advanced channels (X, Y) are fed into each GCN layer as shown in Eq. (6).

$$H^{(0)} = concat\{F_{sampled}, (X, Y)\}. \tag{6}$$

$H^{(0)}$ is the input feature of each GCN layer. $H^{(0)}$ is fed into the first SCU. $H^{(l+1)}$ is the output feature of the lth SCU. The SCU consists of two fully connected layers and one surface propagation layer. Inspired by [17], the lth SCU propagation rule can be formulated as Eq. (7).

$$H^{(l+1)} = \delta(\widetilde{A}^{(l)} H^{(l)} W_0^{(l)} + H^{(l)} W_1^{(l)}), \tag{7}$$

where $W_0^{(l)}, W_1^{(l)}$ are two trainable weight matrics. δ denotes the active function. $\widetilde{A}^{(l)} =$

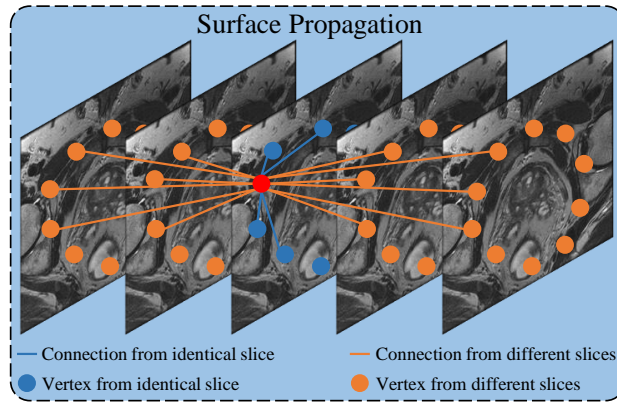II. METHODS II.C. Surface-GCN for Automatic Segmentation

Figure 5: Surface Propagation. In the proposed model, GCN vertices are connected crossing slices like a surface. Take the red vertex for example, the blue vertices are the adjacent vertices in the same slice, and the orange vertices are the adjacent vertices in different slices. The blue and orange dotted lines are intra-slice and inter-slice dependency connections, respectively. Further discussion about the proposed Surface Propagation is presented in supplemental material.

$A^{(l)} + I_N$ denotes the adjacent matrix of graph G with added self-connections. I_N is the identity matrix. $\widetilde{A}^{(l)}$ controls the vertex adjacency in the entire GCN model. Information propagations of surface features are learned by the $W_0^{(l)}$, $W_1^{(l)}$ in each layer. For $a_{ij} \in \widetilde{A}^{(l)}$, a_{ij} denotes the connection from vertex v_i to vertex v_j. $\widetilde{A}^{(l)}$ can also be expressed as the following equation.

$$\widetilde{A}^{(l)} = \begin{bmatrix} \widetilde{A}_{11}^{(l)} & \widetilde{A}_{12}^{(l)} & \cdots & \widetilde{A}_{1S}^{(l)} \\ \vdots & \vdots & \ddots & \vdots \\ \widetilde{A}_{S1}^{(l)} & A_{S2}^{(l)} & \cdots & \widetilde{A}_{SS}^{(l)} \end{bmatrix}, \tag{8}$$

where $\widetilde{A}_{mn}^{(l)}$ denotes the adjacent matrix from vertices in m^{th} slice V^m to vertices in n^{th} slice V^n. If $m = n$, $\widetilde{A}_{mn}^{(l)}$ represents the vertex connections in the identical slice. Otherwise, $m \neq n$ denotes that $\widetilde{A}_{mn}^{(l)}$ records the vertex connections in the different slices. As mentioned above, there are N vertices in one slice and hence $\widetilde{A}_{mn}^{(l)}$ is an $N \times N$ matrix. About the situation $a_{ij}^{intra} \in \widetilde{A}_{mn}^{(l)}$, $m = n$, the a_{ij}^{intra} simulates the information connection of intra-slice dependency. As shown in Fig. 5, take the red vertex v_{red}^s for instance, the vertex has intra-slice dependency connections with its neighboring blue vertices $\{v_{red-k_h}^s, \cdots, v_{red+k_h}^s\}$ in s^{th} slice. Therefore,

$$a_{ij}^{intra} = \begin{cases} 1 & i - j \leq k_h, \\ 0 & otherwise, \end{cases} \tag{9}$$

where $k_h = 2$ controls the number of intra-slice dependency connections.

In contrast, for $a_{ij}^{inter} \in \widetilde{A}_{mn}^{(l)}$, $m \neq n$, suppose $i > j$, $m > n$. a_{ij}^{inter} represents the infor-

---

Last edited $Date$:   II.C.   Surface-GCN for Automatic Segmentation

mation connection of inter-slice dependency. In Fig. 5, orange vertices denote the neighboring vertices of v_{red}^s in the neighboring slices. The inter-slice dependency connection with these vertices (the orange connections) is also established by the following equation.

$$a_{ij}^{inter} = \begin{cases} 1 & i - j < k_v, m - n \le k_v, \\ 0 & otherwise, \end{cases} \tag{10}$$

where $k_v = 2$ controls the number of inter-slice dependency connections. In total, $a_{ij} \in \widetilde{A}_{mn}^{(l)}$ can be formulated as Eq. (11).

$$a_{ij} = \begin{cases} 1 & m = n, i - j \le k_h, \\ 1 & m \neq n, i - j < k_v, m - n \le k_v, \\ 0 & otherwise. \end{cases} \tag{11}$$

Other types of vertex connections are further discussed in supplemental material. Following[25,26], a residual Surface Convolution Unit (resSCU) is designed to propagate vertex information by using residuals. The resSCU contains two SCUs and two ReLU active function layers. Inspired by[17,27], a multi-layer GCN architecture is incorporated to train the target organ in a coarse-to-fine manner. There are three GCN layers in the proposed module. The output from the prior layer is introduced into the next layer for finetuning. Due to the limitation of the capacity of one GCN layer, we use multiple GCN layers to refine the organ boundary step by step The output of each GCN layer is the vertex position in each slice. We use multi-layer GCN to regress the positions of vertices until they reach the boundary. The centripetal Catmull-Rom Spline (CRS)[28] is used to connect output vertices to formulate target organ contours.

II.D. Surface-GCN for Interactive Segmentation

For interactive segmentation, the proposed method aims to learn the experience that radiologists finetune the organ segmentation. In order to imitate radiologist interactions, an annotator simulation method is proposed, *i.e.,* moving the predicted inaccurate vertex to its correct location on the ground truth. Specifically, as shown in Fig. 4, the experience learning module is trained by shifting the vertex that has the largest Euclidean distance from the ground truth. The module learns the residual corrections from radiologists. The shifting of the vertex (x_i, y_i) defined as $(\Delta x_i, \Delta y_i)$ as two additional channels is fed into the module as

following Eq. (12).

$$H^{(0)}_{interact} = concat\{F_{sampled}, (X, Y), (\Delta x_i, \Delta y_i)\}. \tag{12}$$

Then the concatenated feature $H^{(0)}_{interact}$ is used to refine the organ surface location. Similar to the automatic segmentation module, the experience learning module is modeled using GCN. The vertices are connected in the image volume to simulate boundary dependency information propagation. However, in the experience learning module, radiologists' residual corrections are incorporated to help the module infer organ surfaces. Hence radiologists could click one vertex in an image slice, and the annotation information is used to refine the predicted organ surface both in the current slice and in its neighboring slices. It reduces the radiologist interaction clicks to reach the same segmentation accuracy.

II.E. Surface Matching Loss

Fan *et al.*[29] and Rubner *et al.*[30] proposed Chamfer Loss and Earth Mover's Loss respectively to evaluate the distance of two point sets. However, in this paper, so as to match the contours between the ground truth and the predicted surface precisely, it is assumed that vertices in each slice are in a well-defined order. In this way, it could solve the self-intersection problem[31,32]. However, the assumption is contradictory to the above two loss functions concerning the unordered point sets. To address this problem, we suppose these vertices are in a defined order. We connect these vertices in order to form a curve representing the predicted 2D boundary of the target organ in the medical image. Then we resample T=1000 points from the predicted boundary uniformly. At the same time, we also sample T=1000 points from the corresponding ground truth boundary. The points in the two sets are in the defined order. We match each point between two sets in an order and use Mean Absolute Error to calculate the distance of pair points. Since the points have been ordered, there are 1000 different matching strategies. We select the strategy with the minimum distance as the final loss. in s^{th} slice, the contours from the predicted organ and the corresponding ground truth are sampled to T discrete points, and the contour matching loss L_C is defined as Eq. (13).

$$L_C(gt^s, pc^s) = \min_{i \in [0, T-1]} \sum_{j=0}^{T-1} ||gt^s_j - pc^s_{(i+j)\%T}||, \tag{13}$$

where $gt^s = \{gt_0^s, gt_1^s, \cdots, gt_{T-1}^s\}$ and $pc^s = \{pc_0^s, pc_1^s, \cdots, pc_{T-1}^s\}$ are the sampled vertices from ground truth and predicted contour in s^{th} slice, respectively. $||gt_j^s - pc_{(i+j)\%T}^s||$ denotes that the Mean Absolute Error (L1 loss) is used to calculate the distance between two point sets. The $\mathrm{L}_{Contour}$ is calculated in all image slices simultaneously. Hyperparameter T controls the number of sampled discrete points that is used to approximate the contour of the ground truth. The contour details of the organ cannot be well captured if T is small, while it may cause a huge computational cost if the value of T is too large. It is set as $T = 1000$ in this study. The surface matching loss L_S is defined as Eq. (14).

$$L_S(gt, pc) = \sum_{s=0}^{S-1} L_C(gt^s, pc^s), \tag{14}$$

wherein $gt = \{gt^0, gt^1, \cdots, gt^{S-1}\}$ and $pc = \{pc^0, pc^1, \cdots, pc^{S-1}\}$ are the sampled vertices from the ground truths and the predicted contours in all slices. In general, the entire model is optimized by the total loss L_{total}.

$$L_{total} = L_S + L_V + L_E. \tag{15}$$

The model is trained in an end-to-end fashion. In some slices of medical volume the abdominal organ is split into disconnected components by surrounding tissues. These slices (about 2%—5% of total slices for the four organs) had to be removed in the training step because a closed curve is supposed to represent the organ boundary in each slice. One straightforward way for segmentation on these slices is to run the method on each part of disconnected components of target organs.

III.   Experiments

The proposed method was tested through two separate tasks: prostate MR image segmentation task and abdominal multi-organ CT image segmentation task. The prostate MR image segmentation task evaluates the segmentation accuracy of the proposed method on a single organ. In contrast, the abdominal multi-organ CT image segmentation task tests its robustness in segmenting multiple organs.

III.   EXPERIMENTS

III.A.   Dataset

III.A.1.   Prostate MR Image Segmentation

Two public prostate MR datasets and our in-house prostate MR dataset were used for training the proposed Surface-GCN. Three datasets consist of 142 subjects of prostate MR images. 50 subjects were from the MICCAI 2012 Prostate MR Image Segmentation (PROMISE12) Challenge dataset[33], 49 from International Symposium on Biomedical Imaging 2013 (ISBI2013), and 43 from in-house datasets. Our in-house dataset contains T2-weighted MR images that were scanned at 1.5T and 3.0T. The voxel size of these images is $0.625mm^3$, $0.875mm^3$ or $1mm^3$. The matrix sizes vary from 320×320 to 256×256. In each subject, the slices without ground truth masks were dropped out. The training validation ratio was 9 to 1. Thirty test subjects from PROMISE12 test dataset labeled by radiologists were used for evaluation. Each MR slice was labeled manually by 2 radiologists with 15 years and 3 years of experience, separately. Each radiologist labeled the slices three times. Subsequent manual segmentations were conducted one week apart with respect to each other. The labels from two radiologists were finally fused in a majority voting manner.

III.A.2.   Abdominal Multi-organ CT Image Segmentation

Following[16], 90 subjects of abdominal CT images[34] were used to test Surface-GCN performance of multi-organ segmentation. The CT images and some organ segmentations were drawn from 2 public datasets: The Cancer Imaging Archive (TCIA) Pancreas-CT dataset[35,36,37] and the Beyond the Cranial Vault (BTCV) segmentation challenge[38] dataset. The TCIA dataset contains 43 abdominal CT subjects[37]. In our experiment, 4 organs were used: left kidney, gallbladder, spleen, and esophagus. The spleen and left kidney are in large shape, and in contrast, the gallbladder and esophagus are relatively small and hard to segment. All the slices without any target organ were also removed. The images were cropped along the bounding box of different organs.

III.B.   Implementation Details

The proposed method was implemented in Python 3.6.10 and PyTorch 1.0.0 framework. The bounding box was cropped along the ground truths of the corresponding organ. Specifically,
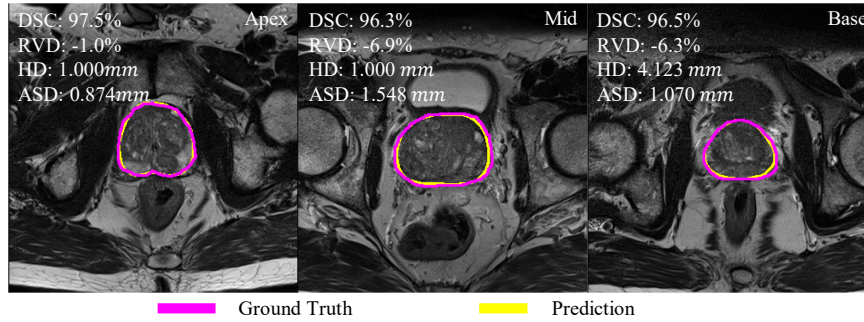

Figure 6: Qualitative results on PROMISE12 test dataset. Magenta contours are the ground truths and the yellow contours are the automatic segmentation results from the proposed Surface-GCN.

given the medical image with its ground truths, we first found the center point and the side length of the rectangle bounding box by using OpenCV libraries. Then, during the training, we randomly scaled the side length 1.2 to 1.4 times. Last, in the test time, we fixedly scaled the side length 1.3 times. Due to the limits of the GPU memory, the batch size was set as 3 during the training process. In both automatic and interactive modules, the entire network was trained with Adam optimizer. For training the prostate MR images, the learning rate was set as 3×10^{-5}, decreasing by $0.9\times$ after every 2 epochs. To alleviate the overfitting problem, data augmentation operations including flipping and rotating were introduced in the training set. For the evaluation of the abdominal multi-organ CT image segmentation task, 9-fold cross-validation was used for a fair comparison. The learning rate was set as 4×10^{-5} decreasing by $0.9\times$ after every 3 epochs.

III.C. Automatic Segmentation Results for Prostate MR Image

Three subjects of qualitative segmentation results from PROMISE12 test dataset are presented in Fig. 6. Noting that the prostate in apex and base regions of the entire MR volume is not easy to segment, all prostate MR subjects were divided into three parts for fair: the apex subregion (30%), mid subregion (40%), and base subregion (30%), respectively. The apex and base parts of the prostate are relatively smaller than the mid part in MR images. Segmenting the apex and base parts of the prostate is more challenging. Simply evaluating the average score of prostate segmentation performance among different methods maybe is not a fair comparison. The magenta contours are the ground truths manually labeled by the radiologists, and the yellow contours are the prediction results from the proposed Surface-GCN. The Dice Similarity Coefficient (DSC), Relative Volume Difference (RVD), Hausdorff Dis-

Table 2: Results on PROMISE12 test dataset compared with other state-of-the-art methods. Results of statistical analysis between two models are also presented in this table. If the p value is lower than 0.05, it is believed that the result is statistically significant.

	Type	Year	DSC↑ (%)	RVD (%)	HD↓ (mm)	ASD↓ (mm)	DSC p value	RVD p value	HD p value	ASD p value
FCN[40]	auto	2015	82.37	6.07	19.64	2.39	1e-32	8e-13	1e-31	2e-4
U-Net[11]	auto	2015	84.71	2.40	15.92	1.89	9e-30	2e-6	3e-28	0.3876
V-Net[12]	auto	2016	85.29	3.49	16.78	2.02	5e-29	1e-8	4e-29	0.1153
PSPNet[39]	auto	2017	75.49	4.71	24.58	2.89	4e-38	1e-10	5e-35	1e-8
DeepLabv3+[41]	auto	2017	86.45	-6.17	23.08	2.20	4e-28	2e-9	5e-34	8e-3
Grab-Cut[23]	semi-auto	2004	78.41	12.22	21.52	2.93	5e-36	7e-20	5e-33	7e-9
Superpixel[42]	semi-auto	2016	87.03	4.28	5.04	2.14	2e-26	5e-10	2e-4	0.0227
ExtremeCut[43]	semi-auto	2018	90.78	-3.44	10.94	1.93	5e-19	4e-31	1e-21	0.2843
Ours	semi-auto	2021	**94.49**	**-1.13**	**3.92**	**1.85**	———	———	———	———

tance (HD), and Average Symmetric Surface Distance (ASD) scores are also labeled on each segmentation result. The prostate segmentation results on the entire PROMISE12 test set are shown in supplemental material. Our proposed method concludes a DSC of 94.49±1.09%, an RVD of -1.13±3.27%, an HD of 3.92±1.52mm, and an ASD of 1.85±0.76mm on the entire gland. A set of segmentation results of relevant state-of-the-art methods[11,12,23,39,40,41,42,43] for PROMISE12 test dataset are listed in the Table 2. The reported results were obtained by using the application of available code in public and experiments were conducted in the same environments. It can be seen that our method outperforms the other methods in the DSC, RVD, HD, and ASD scores. Furthermore, the statistical analysis of results was conducted using paired t-test. All the p values are also listed in Table 2. The p value that is lower than 0.05 can also prove that the result is statistically significant.

III.D. Automatic Segmentation Results for Abdomen Datasets

Two subjects of qualitative segmentation results are shown in Fig. 7. Blue, green, orange, and cyan contours denote the segmentation results of the left kidney, gallbladder, spleen, and esophagus, respectively. The magenta contours denote the corresponding ground truths. The DSC, 95%HD, and ASD scores are labeled on each result. It can be seen that our method performs accurate segmentation results in four organs. More quantitative results and the segmentation overlaid on the entire CT are in supplemental material.

In Table 3, our method is compared with other state-of-the-art abdominal segmentation methods[12,13,15,16,44,45,46,47,48] in the DSC, 95%HD, and ASD scores. All the methods are deep-learning-based except Shimizu *et al.*[47]. Segmentation results of VoxResNet[44], V-Net[12],

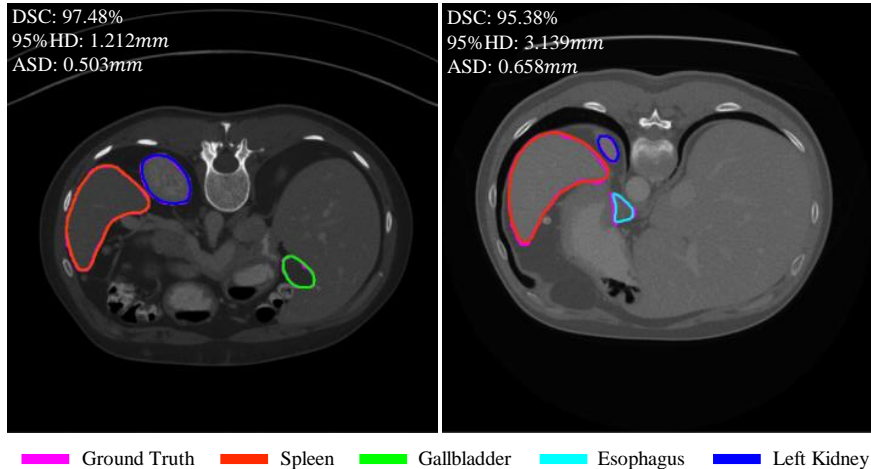| | Ground Truth | | Spleen | | Gallbladder | | Esophagus | | Left Kidney |

Figure 7: Qualitative results for abdominal segmentation. Orange, green, cyan, and blue contours are segmentations of the spleen, gallbladder, esophagus, and left kidney, separately. Magenta contours are the corresponding ground truths.

Zhou[45], DenseVnet[16] and Shimizu[47] are from Gibson's paper[16]. Segmentation accuracy of different organs by using these methods changes dramatically because the boundary information is hard to capture on small organs. Our method performs DSC scores of 95%, 91%, 95%, and 88%, 95%HD scores of $2.0mm$, $2.2mm$, $2.4mm$, $2.0mm$, and ASD scores of $0.9mm$, $1.0mm$, $1.1mm$, and $0.9mm$ for the left kidney, gallbladder, spleen, and esophagus respectively. It can be seen that our method reaches better performance on 4 organs, especially for the esophagus. The shape of the esophagus changes drastically and irregularly. By using the proposed method the segmentation performance on two small organs (esophagus and gallbladder) is similar to two large organs (left kidney and spleen). Our method reaches similar accuracy in DSC metric with[48] for the gallbladder. Our method reaches similar accuracy both in DSC metric and in 95%HD metric with[16] for the spleen. However, our method achieves a high score for the rest of the organs. Compared to other deep-learning-based methods, the advantage of our method is that the shape of the target organ is utilized to efficiently propagate image feature information.

III.E. Evaluation of Loss Functions

To analyze the effectiveness of proposed loss functions, a set of networks with different combinations of L_E, L_V, and L_S were conducted to experiment with segmentation performance. Experimental results are shown in Table 4, where α-balanced denotes that the L_E and L_V are calculated by α-balanced CE. Concretely, "Fixed" denotes that the α_e and the α_v are set

Table 3: Segmentation results of left kidney, gallbaldder, spleen, and esophagus on abdomen datasets compared with other state-of-the-art methods.

		DSC↑ (%)					95%HD↓ (mm)			
Method	Year	Left Kidney	Gallbladder	Spleen	Esophagus	Method	Left Kidney	Gallbladder	Spleen	Esophagus
VoxResNet[44]	2018	89	69	90	65	VoxResNet[44]	4.2	5	3.8	6
V-Net[12]	2016	90	64	92	65	V-Net[12]	3.7	7.5	3.6	6.5
Zhou[45]	2016	91	65	92	43	DenseVNet[16]	3.1	4.6	**2.4**	5.6
Multi-scale[48]	2018	87	**91**	89	59	Ours	**2.0**	**2.2**	2.4	**2.0**
DenseVNet[16]	2018	93	73	**95**	71			ASD↓ (mm)		
UNet++[13]	2018	91	60	89	65	Method	Left Kidney	Gallbladder	Spleen	Esophagus
Larsson[46]	2018	91	62	93	66	VoxResNet[44]	1.6	1.9	1.8	2.2
Shimizu[47]	2007	88	77	91	37	V-Net[12]	1.3	2.8	1.2	2.4
Zhang[15]	2020	91	68	91	69	DenseVNet[16]	3.1	4.6	2.4	5.6
Ours	2021	**95**	**91**	**95**	**88**	Ours	**0.9**	**1.0**	**1.1**	**0.9**

as 1.0 and are fixed during training the network. In contrast, "Adaptive" denotes that the α_e and the α_v are changed by Eq. (3) and Eq. (4), respectively. It can be seen that the proposed MBASM module increase the DSC accuracy from 91.25% to 94.49%. Through supervising the proposed SFE-Net by L_E and L_V, the segmentation DSC score is improved from 91.25% to 93.06%. Furthermore, by using the proposed mini-batch adaptive α-balanced CE, the performance is increased from 93.06% DSC score to 94.49%.

Table 4: Loss comparison on PROMISE12 test dataset.

Loss Type	α-balanced	DSC↑ (%)	HD↓ (mm)	ASD↓ (mm)
L_S	None	91.25	4.33	2.24
L_V+L_S	Fixed	92.91	3.99	2.01
$L_E+L_V+L_S$	Fixed	93.06	3.95	2.01
$L_E+L_V+L_S$	Adaptive	**94.49**	**3.90**	**1.85**

III.F.   Evaluation of SFE-Net

To test the feature extraction ability of the proposed SFE-Net, Table 5 shows several CNN structures. 152-SFE-Net (details are presented in Table 1) increases more convolutional layers than 101-SFE-Net. PPM[39] block and ASPP[24] block are introduced to increase the reception field of the encoder network. It can be seen that our method shows robustness about different SFE-Net structures.

Table 5: Evaluation of the proposed SFE-Net on PROMISE12 test dataset. Aug. denotes using the augmentation strategies that includes flipping and rotating.

101-SFE-Net	152-SFE-Net	PPM[39]	ASPP[24]	Aug.	DSC↑ (%)	HD↓ (mm)	ASD↓ (mm)
✓					93.85	4.14	2.02
✓		✓			94.03	**3.90**	1.98
✓			✓		93.98	4.18	1.96
✓		✓		✓	94.09	4.02	1.99
✓			✓	✓	**94.49**	3.92	**1.85**
	✓		✓	✓	94.05	4.19	1.96

III.G.   Interactive Segmentation Results

There are two different training strategies for the interactive module. One is to train the interactive module from scratch. The other is to finetune the module based on the pretrained automatic module. We find the latter strategy reaches a better performance and need less training time compared to the prior. We hence used the pretrain automatic module for training the interactive module. Fig. 8 shows interactive segmentation results on abdominal organs. In reality, to imitate radiologist interactions, the model iteratively makes corrections until the prediction cannot be improved. Segmentation results of the left kidney, spleen, esophagus, and gallbladder are presented in order. In Fig. 9, the segmentation accuracy *vs.* interactive pseudo clicks is presented. Following the protocol in[31,32], the model stopped making further corrections after the IOU was greater than a threshold T. Curve-GCN[31] is used to compare. Four abdominal organs are tested at $T = 1.0$, $T = 0.9$, and $T = 0.8$ situations. For the spleen and left kidney, the segmentations could reach higher accuracy with the increase of interactive pseudo clicks. For the esophagus and the gallbladder, interactive pseudo clicks could increase the segmentation accuracy more rapidly. And the proposed method reduces 5-10 clicks to reach more accurate segmentations in four organs than the Curve-GCN. It demonstrates the superiority of the proposed interactive segmentation method that utilizes radiologist experience in medical image slices. More details comparasions between the proposed automatic and interactive segmentation are presented in supplemental material.
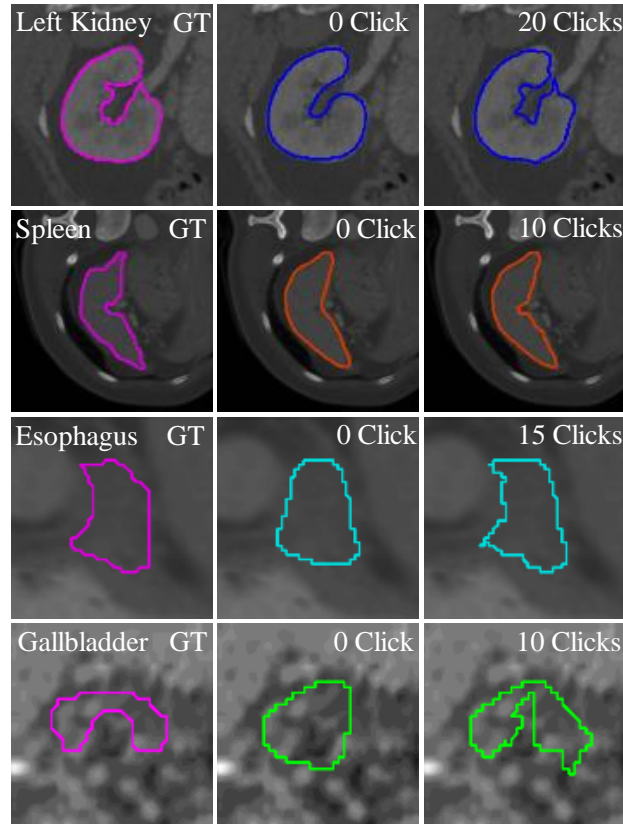
Figure 8: Interactive segmentation results of the spleen, esophagus, left kidney, and gall-bladder. The images are cropped along the bounding boxes of the organs. It can be seen that the expriences of radiologists' corrections are well captured by the proposed method.

III.H.   Discussion and Limitation

Discussion about contributions. In this paper, we conducted the experiment on prostate MR image segmentation task and abdominal multi-organ CT image segmentation task. For prostate segmentation, the proposed method achieve better results compared to other state-of-the-art methods by using automatic module. For multi-organ segmentation, the proposed method reaches promising results compared to other methods. And 4 tested organs achieve similar accuracy (95% DSC for left kidney, 91% for gallbladder, 95 % for spleen, and 88% for esophagus), which demonstrates that the proposed method could segment various organs simultaneously. The contributions of the proposed method are three-fold:

1. We proposed a GCN-based interactive organ segmentation method that integrates radiologist corrections into residuals on organ surfaces. It directly learns radiologist interaction experience in medical volume to reduce 5-10 interaction clicks for achieving
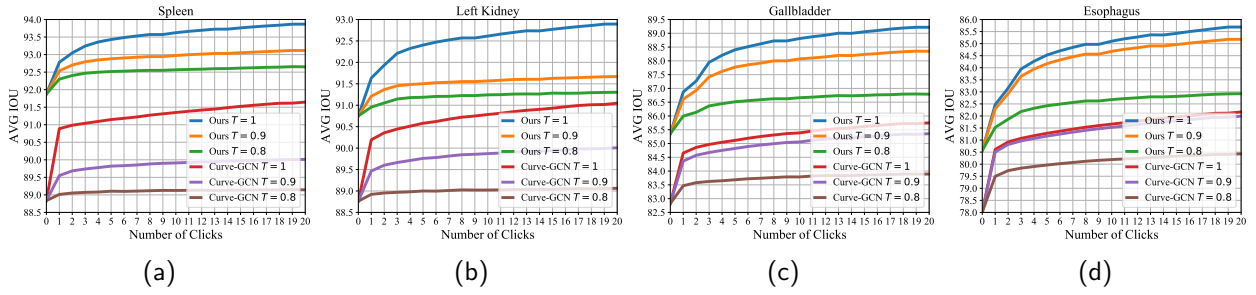
Figure 9: Interactive segmentation IOU scores (%) *vs.* the number of pseudo clicks. (a)-(d) show the results on abdominal multi-organ images at different thresholds T. It can be seen that the proposed method learns more radiologists' interactive experiences and reaches more accurate segmentation results.

the same segmentation accuracy.

2. We designed a Surface Feature Extraction Network to capture the organ surface details and a Mini-batch Adaptive Surface Matching module to help the Surface Feature Extraction Network focus on the organ surfaces. The proposed network and the proposed module further imporves the organ segmentation accuracies.

3. We defined a graph convolution operation to propagate organ surface information. The operation naturally utilizes both intra-slice and inter-slice dependencies in medical volume and establishes connections on two dependencies.

Discussion about single image segmentation. Our method could also segment a single image slice in medical volume. In practice, segmenting medical images slice by slice denotes ignoring the inter-slice dependency information, although we clarify that 2D medical image segmentation is a special situation of 3D volume segmentation. In practice, we need segment medical volume instead of single image slice.

Discussion about the experience learning of radiologist interation. First, given an inaccurate segmentation contour of a target organ by automatic segmentation method, radiologists will re-segment it by revising the contour. We believe this procedure includes the experience of radiologists since there is no ground truth and radiologists modify the segmentation contour with their experience. The experience could be specified as the movement of the contour, which is the shifting of the vertex in our model. Second, our model aims to re-adjusting the contour with the above guidance. When radiologists have revised one vertex on the contour, their experience, the shifting of the vertex, is incorporated into our model

---

III.   EXPERIMENTS   III.H.   Discussion and Limitation

for improving the segmentation. We find that the experience from one specific point could be provided guidance to correct the rest boundary, as discussed in the introduction part there exist intra-boundary and inter-boundary correspondence in medical volume. Third, to imitate the radiologists' procedure, we design the simulation algorithm. Although there is no real radiologist involved,we find they always correct the point with the largest distance or correct one point from the top k points as discussed in supplemental material. We design the algorithm by following this principle.

With radiologists' experience, our method could refine the segmentation results inter-actively. In the experiment as presented in Fig. 9, we have tested the effect of radiologist interaction. With one pseudo click, segmentation accuracy could increase about 2% IOU on spleen and left kidney. Specifically for the gallbladder and esophagus, the accuracy increases 3% IOU after one pseudo click. The proposed method with few pseudo clicks achieves accurate segmentation. Moreover, it still reaches a promising result with 7-10 clicks.

Limitation.   In Table 2, for prostate segmentation, there are two different types of method that is used to compare. One is the automatic segmentation method. We select the representative ones (U-Net and V-Net) that are widely used in medical imaging and the ones (FCN[40], PSPNet[39], and DeepLab series[41]) that are common baselines for deep learning methods. The other is the interactive segmentation method. It is hard to compare different interactive methods because different methods interact with humans differently. We select the Grab-Cut[23] method for it has a great impact on interactive methods. Superpixel[42] is a method specifically designed for prostate segmentation. There are few methods designed for the prostate method, so in supplementary material, we reimplement Curve-GCN[31] on our datasets and use it as the baseline.

For abdominal organ segmentation, the disconnected components were removed both in the training step and in the testing step. There are about 2.91%, 1.61%, 2.06%, and 4.91% slices that have disconnected parts for the spleen, left kidney, gallbladder, and esophagus, respectively. The total DSC scores for segmenting the spleen and the left kidney remains similar to the ones in Table 3. The segmentation accuracy for the gallbladder and the esophagus drops by 1% DSC score compared with the results in Table 3.

In the paper, we train and evaluate the proposed method in the same datasets. So the

Last edited $Date$:   III.H.   Discussion and Limitation

proposed method may not perform well while training one dataset and evaluating another dataset collected from different imaging machines.

IV.   Conclusion

In this paper, we proposed a Graph Convolutional Network called Surface-GCN to solve the medical organ segmentation challenge. Organ segmentation is considered a regression problem. Our model delineates the target organ by a closed surface. First, we proposed the SFE-Net and MBASM module to capture the organ surface feature. Second, we designed a GCN module for automatic organ segmentation. The proposed module establishes intra-slice and inter-slice dependency connections in the medical image volume to propagate the organ boundary information. Third, we proposed an interactive segmentation module to learn the radiologist interactive experience. Experiments on the PROMISE12 test dataset and two abdominal multi-organ datasets demonstrate that our method could segment multiple organs and achieves better performance compared with other state-of-the-art methods. In the future, we will evaluate the proposed method on different organs.

As mentioned in the Sec. II., in some slices of medical volume the abdominal organ is split into disconnected components by surrounding tissues. Interestingly it is noted that all the organs are self-connected in space. So in future work, we will extend the proposed method to directly delineate the target organ from medical volume by establishing its boundary surface. Concretely, we will represent the organ boundary as a three-dimension mesh instead of contours in each slice of medical volume. In addition, although our method achieves high DSC scores for automatic segmentation in all organs, the performance of interactive segmentation on small organs is not as good as on relatively large ones. As shown in Fig. 8, the interactive segmentation results of the esophagus and gallbladder reach promising performance but leave a gap compared to other organ results from the proposed method. Organs like the prostate, liver, and spleen can be segmented well by the proposed method. Organs like the esophagus are small organs that maybe it needs the medical image in higher resolution to segment well. Our future work will also focus on promoting the segmentation performance of these small targets.

IV.   CONCLUSION

V. Acknowledgement

VI. Conflicts of interest

The authors have no relevant conflicts of interest to disclose.

References

[1] K. Kasabwala et al., The learning curve for magnetic resonance imaging/ultrasound fusion-guided prostate biopsy, European Urology Oncology **2**, 135–140 (2019).

[2] P. A. Pinto et al., Magnetic resonance imaging/ultrasound fusion guided prostate biopsy improves cancer detection following transrectal ultrasound biopsy and correlates with multiparametric magnetic resonance imaging, Journal of Urology **186**, 1281–1285 (2011).

[3] H. Kobatake, Future CAD in multi-dimensional medical images. - Project on multi-organ, multi-disease CAD system -, Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society **31**, 258–66 (2007).

[4] J. Weese and C. Lorenz, Four challenges in medical image analysis from an industrial perspective, Medical Image Analysis **33**, 44–49 (2016), 20th anniversary of the Medical Image Analysis journal (MedIA).

[5] Z. Khan, N. Yahya, K. Alsaih, M. I. Al-Hiyali, and F. Meriaudeau, Recent Automatic Segmentation Algorithms of MRI Prostate Regions: A Review, IEEE Access **9**, 97878–97905 (2021).

[6] Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang, A review of deep learning based methods for medical image multi-organ segmentation, Physica Medica **85**, 107–122 (2021).

7   G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, Densely connected convolutional networks, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2261–2269, 2017.

8   J. Wang et al., Prior-attention residual learning for more discriminative COVID-19 screening in CT images, IEEE Transactions on Medical Imaging **39**, 2572–2583 (2020).

9   Q. Zhu, B. Du, and P. Yan, Boundary-weighted domain adaptive neural network for prostate MR image segmentation, IEEE Transactions on Medical Imaging **39**, 753–763 (2020).

10 Y. Lin, Y. Chi, H. Han, M. Han, and Y. Guo, Multimodal Orthodontic Corpus Construction Based on Semantic Tag Classification Method, Neural Processing Letters (2021).

11 O. Ronneberger, P. Fischer, and T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in *Medical Image Computing and Computer Assisted Intervention Society*, volume 9351, pages 234–241, 2015.

12 F. Milletari, N. Navab, and S.-A. Ahmadi, V-Net: Fully convolutional neural networks for volumetric medical image segmentation, IEEE 3D Vision , 565–571 (2016).

13 Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, UNet++: Redesigning skip connections to exploit multiscale features in image segmentation, IEEE Transactions on Medical Imaging **39**, 1856–1867 (2020).

14 C. Huang, H. Han, Q. Yao, S. Zhu, and S. K. Zhou, 3D U^2-Net: A 3D universal U-Net for multi-domain medical image segmentation, in *Medical Image Computing and Computer Assisted Intervention Society*, edited by D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan, pages 291–299, 2019.

15 L. Zhang et al., Block level skip connections across cascaded V-Net for multi-organ segmentation, IEEE Transactions on Medical Imaging **39**, 2782–2793 (2020).

16 E. Gibson et al., Automatic multi-organ segmentation on abdominal CT with dense V-networks, IEEE Transactions on Medical Imaging **37**, 1822–1834 (2018).

17   T. N. Kipf and M. Welling, Semi-supervised classification with graph convolutional networks, Proceedings of the International Conference on Learning Representations , 1–14 (2017).

18   V. Lempitsky, P. Kohli, C. Rother, and T. Sharp, Image segmentation with a bounding box prior, in *International Conference on Computer Vision*, pages 277–284, 2009.

19   J. Liew, Y. Wei, W. Xiong, S. H. Ong, and J. Feng, Regional interactive image segmentation networks, in *International Conference on Computer Vision*, pages 2746–2754, 2017.

20   N. Xu, B. Price, S. Cohen, J. Yang, and T. Huang, Deep interactive object selection, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 373–381, 2016.

21   G. Wang et al., Interactive medical image segmentation using deep learning with image-specific fine tuning, IEEE Transactions on Medical Imaging **37**, 1562–1573 (2018).

22   S. H. Park et al., Interactive prostate segmentation using atlas-guided semi-supervised learning and adaptive feature selection, Medical Physics **41**, 111715–111715 (2014).

23   C. Rother, V. Kolmogorov, and A. Blake, "GrabCut" - Interactive foreground extraction using iterated graph cuts, in *ACM Transactions on Graphics*, volume 23, pages 309–314, 2004.

24   L. Chen, G. Papandreou, F. Schroff, and H. Adam, Rethinking atrous convolution for semantic image segmentation, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

25   M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, Geometric deep learning: Going beyond euclidean data, IEEE Signal Processing Magazine **34**, 18–42 (2017).

26   N. Wang et al., Pixel2Mesh: Generating 3D mesh models from single RGB images, in *Proceedings of European Conference on Computer Vision*, volume 11215, pages 55–71, 2018.

Last edited *Date* :

27   G. Li, M. Muller, A. Thabet, and B. Ghanem, DeepGCNs: Can GCNs go as deep as CNNs?, International Conference on Computer Vision , 9266–9275 (2019).

28   C. Yuksel, S. Schaefer, and J. Keyser, Parameterization and applications of CatmullRom curves, in *Computer-Aided Design*, volume 43, pages 747–755, 2011.

29   H. Fan, H. Su, and L. Guibas, A point set generation network for 3D object reconstruction from a single image, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition , 2463–2471 (2017).

30   Y. Rubner, C. Tomasi, and L. J. Guibas, The earth mover's distance as a metric for image retrieval, International Journal of Computer Vision **40**, 99–121 (2000).

31   H. Ling, J. Gao, A. Kar, W. Chen, and S. Fidler, Fast interactive object annotation with Curve-GCN, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5252–5261, 2019.

32   D. Acuna, H. Ling, A. Kar, and S. Fidler, Efficient interactive annotation of segmentation datasets with Polygon-RNN++, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 859–868, 2018.

33   G. Litjens et al., Evaluation of prostate segmentation algorithms for MRI: The PROMISE12 challenge, Medical Image Analysis **18**, 359–373 (2014).

34   E. Gibson et al., Multi-organ abdominal CT reference standard segmentations, 2018.

35   H. R. Roth, L. Lu, A. Farag, H. C. Shin, J. Liu, E. Turkbey, and R. M. Summers, DeepOrgan: Multi-level deep convolutional networks for automated pancreas segmentation, in *Medical Image Computing and Computer Assisted Intervention Society*, pages 556—-564, 2015.

36   K. Clark et al., The cancer imaging archive (TCIA): Maintaining and operating a public information repository, Journal of Digital Imaging **26**, 1045–1057 (2013).

37   H. R. Roth et al., Data from pancreas-CT, The Cancer Imaging Archive (2016).

38   L. BA, X. Z, I. JE, S. M, L. TR, and K. A, Medical Image Computing and Computer Assisted Intervention Society, 2015.

39 H. Zhao et al., Pyramid scene parsing network, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6230–6239, 2017.

40 J. Long, E. Shelhamer, and T. Darrell, Fully convolutional networks for semantic segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence **39**, 640–651 (2015).

41 L. C. Chen et al., Encoder-decoder with atrous separable convolution for semantic image segmentation, in *Proceedings of European Conference on Computer Vision*, pages 833–851, 2018.

42 Z. Tian, L. Liu, Z. Zhang, and B. Fei, Superpixel-based segmentation for 3D prostate MR images, IEEE Transactions on Medical Imaging **35**, 791–801 (2016).

43 K.-K. Maninis et al., Deep extreme cut: From extreme points to object segmentation, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 616–625, 2018.

44 H. Chen, Q. Dou, L. Yu, J. Qin, and P.-A. Heng, VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images, NeuroImage **170**, 446 – 455 (2018).

45 X. Zhou, T. Ito, R. Takayama, S. Wang, T. Hara, and H. Fujita, Three-dimensional CT image segmentation by combining 2D fully convolutional network with 3D majority voting, in *Deep Learning and Data Labeling for Medical Applications*, pages 111–120, 2016.

46 M. Larsson, Y. Zhang, and F. Kahl, Robust abdominal organ segmentation using regional convolutional neural networks, Applied Soft Computing **70**, 465 – 471 (2018).

47 A. Shimizu et al., Segmentation of multiple organs in non-contrast 3D abdominal CT images, International Journal of Computer Assisted Radiology and Surgery **2**, 135–142 (2007).

48 H. R. Roth et al., A multi-scale pyramid of 3D fully convolutional networks for abdominal multi-organ segmentation, in *Medical Image Computing and Computer Assisted Intervention Society*, pages 417–425, 2018.

Last edited *Date* :