Structure Boundary Preserving Segmentation for Medical Image with Ambiguous Boundary

Hong Joo Lee, Jung Uk Kim, Sangmin Lee, Hak Gu Kim, Yong Man Ro
Image and Video Systems Lab, School of Electrical Engineering, KAIST, South Korea
{dlghdwn008, jukim0701, sangmin.lee, hgkim0331, ymro}@kaist.ac.kr

Abstract

In this paper, we propose a novel image segmentation method to tackle two critical problems of medical image, which are (i) ambiguity of structure boundary in the medical image domain and (ii) uncertainty of the segmented region without specialized domain knowledge. To solve those two problems in automatic medical segmentation, we propose a novel structure boundary preserving segmentation framework. To this end, the boundary key point selection algorithm is proposed. In the proposed algorithm, the key points on the structural boundary of the target object are estimated. Then, a boundary preserving block (BPB) with the boundary key point map is applied for predicting the structure boundary of the target object. Further, for embedding experts knowledge in the fully automatic segmentation, we propose a novel shape boundary-aware evaluator (SBE) with the ground-truth structure information indicated by experts. The proposed SBE could give feedback to the segmentation network based on the structure boundary key point. The proposed method is general and flexible enough to be built on top of any deep learning-based segmentation network. We demonstrate that the proposed method could surpass the state-of-the-art segmentation network and improve the accuracy of three different segmentation network models on different types of medical image datasets.

1. Introduction

For many medical image processing applications, it is an important key to success for the models to correctly segment anatomical structures in the medical image domain [24, 17]. However, it is challenging to obtain accurate segmentation results because of the ambiguity of structure boundary, heterogeneous texture and the uncertainty of the segmented region without domain knowledge. Even experts have differences slightly in their delineation depending on their experience and skill [18].

In the era of deep learning, many convolutional neural networks (CNNs)-based segmentation approaches have been proposed to accurately segment the target object both in the natural and medical image domain [15, 31, 21, 33]. Fully Convolution Network (FCN) [15] with skip-layer to preserve spatial localization information was proposed for semantic segmentation. Inspired by FCN, U-Net [21] was proposed for utilizing the context information of the higher layers to predict precise output by combining higher resolution features with the up-sampled feature. Zhao et al.[33] integrated FCN and Conditional Random Fields in a unified segmentation framework for brain tumor segmentation. These deep learning-based medical image segmentation methods did not explicitly take into account the difficulties in the medical image such as ambiguous structure boundaries and heterogeneous textures. Figure 1 shows cases where U-Net failed to segment the target regions in ultrasound image. To deal with the ambiguous structure boundary issue, a few approaches have been reported [6, 11]. These methods enforce inter-pixel dependency to recover boundary details. However, they need manual parameter tuning as post-processing, which is labor-intensive tasks, and the results are affected by parameter tuning.

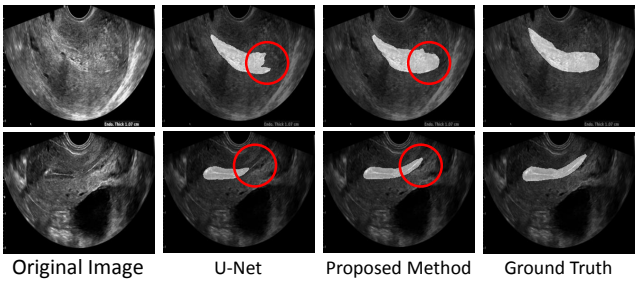

Figure 1. Results of automatic segmentation methods on medical images with the ambiguous structure boundary and heterogeneous texture image. Note that these segmentation results by U-Net have failed in preserving structure boundary.

To overcome this limitation, interactive or semi-automatic segmentation methods have been proposed where ambiguous structure boundary is dealt with interactively during test time. The interactive segmentation methods em-

ploy user inputs such as points, bounding boxes to segmentation network [30, 29, 19]. Wang et al. [29] proposed a deep learning-based interactive segmentation approach in medical image domain. They encoded structure boundary information of target object manually; the network could predict sophisticated structures. However, the interaction approaches require user interaction time and specialized domain knowledge.

In this paper, we focus on tackling the following two segmentation problems raised by medical image domain. First, most of the medical images in application contain ambiguous boundaries because of poor image quality and heterogeneous textures. Unlike objects in the natural image, there might be no salient structure boundary due to the low resolution. Second, it is difficult to automatically predict the correct target region without knowledge of experts such as melanocytic lesions in ultrasound image.

To tackle the aforementioned problems, we propose a novel fully automatic medical image segmentation framework that preserves structure boundary of target region. Firstly, we propose a novel boundary key point selection algorithm. The algorithm automatically selects key points that best fit to target object region. These points put on the structure boundary of the target object. Then, the points are encoded to the segmentation network with a novel structure boundary preserving model, named as Boundary Preserving Block (BPB). It allows segmentation model to further exploit the structure boundary information. In the proposed BPB, the boundary key point map is estimated from visual features. For embedding the experts knowledge in the fully automatic segmentation model, we introduce a novel structure boundary information-based discrimination network named Shape Boundary-aware Evaluator (SBE) in an adversarial manner without the user interaction. In training stage, it tries to evaluate how much structure boundary of the segmentation map is well preserved by using the key point map. Thus, the proposed SBE can give feedback to the segmentation network on the predicted region, based on the ground-truth region marked by experts. In addition, the proposed method mentioned above is general and flexible enough to be applicable to any automatic medical image segmentation models. The flexibility of the proposed method allows any segmentation models to more precisely segment target region by integrating BPB and SBE.

To conclude the introduction, we outline the major contributions of this work as follows.

- We proposed a novel boundary key point selection algorithm that best fit the target region. The selected key points putting on the structure boundary of target region are encoded through the BPB with boundary key point map generator.

- In the proposed framework, we employ boundary key point information automatically without the user interaction. To this end, we trained the segmentation network in an adversarial way with SBE. The evaluator gives feedback to segmentation network whether given segmented region coincidences with boundary key points or not.

- The proposed method can be generalized to different segmentation models. To evaluate the generalization ability of the BPB and SBE, we integrate our approach with three recent state-of-the-art segmentation models, U-Net, FCN, Dilated-net [31]. We demonstrate that the proposed method improves the prediction performance with statistical significance.

2. Related Works

2.1. Automatic Medical Image Segmentation

Recently, Deep Convolutional Neural Networks (DCNNs) have shown great success both in natural image and medical image domain [21, 15, 31]. Fully Convolution Network (FCN) [15] is one of the most widely used segmentation networks both on natural image and medical image. The FCN consists of consecutive convolution and max-pooling layers. To preserve spatial localization information, it used skip-layer. Many medical image segmentation network used FCN for medical image segmentation [22, 7, 26, 34, 32]. Roth et al. [22] applied FCN networks cascaded way for medical image segmentation. Vorontsov et al. [26] used two types of FCNs for liver and liver lesion segmentation. In addition to FCN, U-Net [21] shows superior performance in medical image segmentation. U-Net utilized the encoder features to decoder features by skip connections. Since the encoder feature information is transferred to decoder, it shows comparable performance in medical image segmentation. Inspired by U-Net, many deep learning-based automatic segmentation networks were proposed [8, 16, 14, 9, 12]. Dalm et al. [9] proposed 2 and 3 consecutive U-Nets for breast mass segmentation. Although these approaches have achieved reasonable segmentation results in medical image segmentation, it still has problems for preserving boundary.

2.2. Interactive Medical Image Segmentation

In general, interactive segmentation shows superior performance than conventional segmentation method [25]. In medical image segmentation, it shows great performance [29, 19, 27, 20] since it encodes the experts knowledge to segmentation network with several interactions. Rajchl et al. [20] trained the CNNs network by employing user-provided inputs. By providing specific region information, the segmentation prediction performance was improved. Wang et al. [29] proposed deep learning-based interactive segmentation method. They employed structure

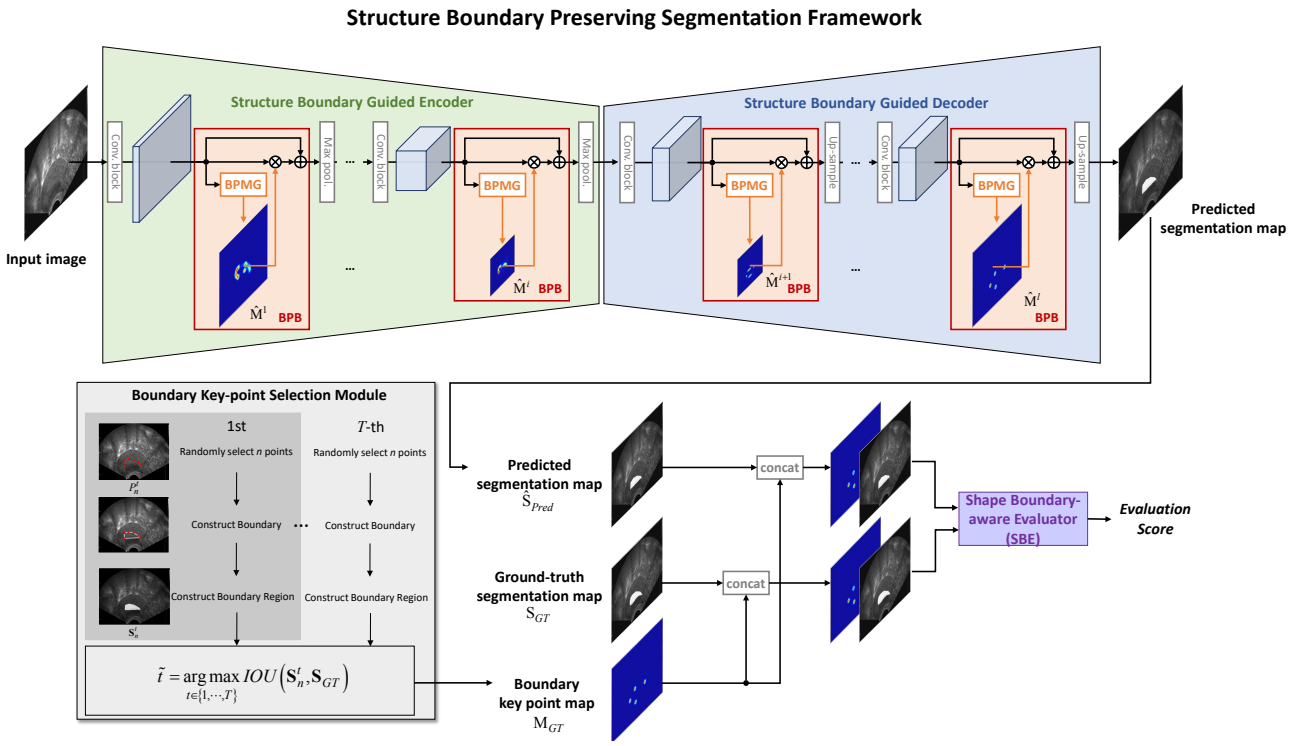

Figure 2. The overview framework for structure boundary preserving segmentation. The network consists of segmentation network with the boundary preserving block (BPB) and the shape boundary-aware evaluator (SBE). Note that the boundary preserving blocks (BPB) are integrated into segmentation network for preserving structure boundary. The shape boundary-aware evaluator (SBE) evaluates the predicted segmentation map, which is used in training stage only.

boundary information to the segmentation network. Different from a previous interactive method, they reduced user interactions by employing a refinement network. These approaches showed superior results by employing user interaction. However, they still need to interact with the experts at inference time.

3. Proposed Structure Boundary Preserving Segmentation

Figure 2 shows the overview of the proposed boundary preserving segmentation framework. It consists of segmentation network with Boundary Preserving Block (BPB) and Shape Boundary-aware Evaluator(SBE). The BPB and SBE use boundary key points that are selected from our novel key point selection algorithm to preserve the structure boundary of the target object. In the BPB, the key points map is generated and used to refine the input feature. Then, the segmentation network predicts structure boundary preserved segmentation map. In the SBE network, it gives feedback to segmentation network whether given segmentation map coincides with boundary key point map or not. After the segmentation network and SBE are trained in an adversarial way, only the segmentation network is used for inferring.

Algorithm 1: Boundary key point selection algorithm

Input: Total number of iterations T, number of boundary key points n, ground truth segmentation map $\mathbf{S}_{GT}$
Output: Boundary key Points $\tilde{P}$
Initialize $IOU_{best} = 0$
for $t = 1, 2, \cdots, T$ **do**
  Randomly select N points
  $P_n^t \leftarrow \{(x_1^t, y_1^t), (x_2^t, y_2^t), \cdots, (x_n^t, y_n^t)\}$
  $\mathbf{S}_n^t \leftarrow c(P_n^t)$
  $IOU_t \leftarrow IOU(\mathbf{S}_n^t, \mathbf{S}_{GT})$
  if $IOU_t > IOU_{best}$ **then**
    $IOU_{best} \leftarrow IOU_t$
    $\tilde{P} \leftarrow P_n^t$
  end
end
Return: $\tilde{P}$

Followings are the detail explanations of the proposed structure boundary preserving segmentation framework.

3.1. Boundary Key Point Selection Algorithm

To obtain the boundary key points that best fit the ground-truth segmentation map and represent structure

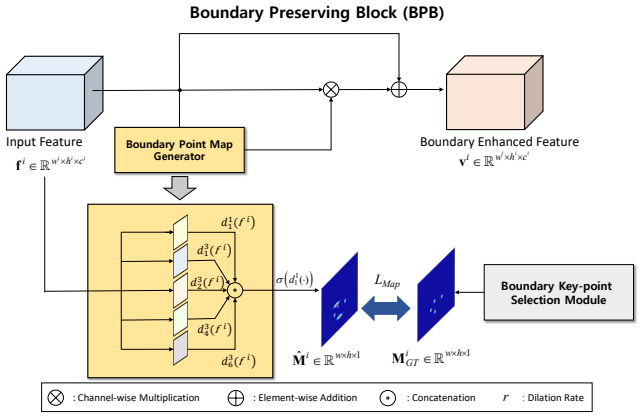

Figure 3. Detailed framework for training boundary preserving block (BPB). The ground truth boundary key point map is generated from boundary key point selection algorithm (see Section 3.1).

boundary of the target region, we devise a novel boundary key point selection algorithm. First, we obtain the boundary of the target object from the ground-truth segmentation map using conventional edge detection algorithm [5]. On the boundary of the target object, we randomly select n points. Let $P_n^t = \{(x_1^t, y_1^t), (x_2^t, y_2^t), \cdots, (x_n^t, y_n^t)\}$ denotes randomly selected n points set, where t denotes the number of trials (i.e., iterations). Then, we construct the boundary region, S_n^t, by connecting the n points at t-th iteration. To obtain the ground-truth boundary key point map, we measure the amount of overlap region between the boundary region $\mathrm{S}_n^t = c(P_n^t)$ and the ground-truth segmentation map S_{GT} by calculating Intersection Over Union (IOU). $c(\cdot)$ is the function to construct the boundary region from the selected point set P_n^t. Finally, the boundary points which lead to the highest value of IOU are selected as the structure boundary key point. The selected key points are written as

$$\tilde{P} = P_n^{\tilde{t}}, \text{where } \tilde{t} = \underset{t \in \{1,\cdots,T\}}{\arg\max}\ \mathrm{IOU}\left(\mathrm{S}_n^t, \mathrm{S}_{GT}\right). \quad (1)$$

The selected key points set $\tilde{P}$ is converted to 2-dimension boundary point map. To allow the tolerance of the key points position in the training phase, we generated a disk on each key point. Let x be the 2-D key points position in the image. Let $D(y) = \{x : \|x - y\| < R\}$ be a disk of radius R centered around y. We set up a binary classification task as $p(x) = 1$ if x is in $D(y)$, otherwise $p(x) = 0$. Therefore, we regarded the key points map as 2-dimensional probability map and use cross-entropy to train the model. These points map is used for training segmentation network and SBE. Algorithm 1 and Figure 2 describe the detail of the proposed boundary key-point selection algorithm.

3.2. Boundary Preserving Block (BPB)

As shown in Figure 3, the segmentation network with Boundary Preserving Blocks (BPBs) predicts the segmentation map. The proposed BPBs include the boundary points map generator to estimate the boundary key point map. By generating the boundary key point map, the segmentation network encodes structure boundary information without interactions. The boundary points map generator uses feature $\mathrm{f}^{\mathrm{i}} \in \mathbb{R}^{\mathrm{w^i} \times \mathrm{h^i} \times \mathrm{c^i}}$ as input, where w^i, h^i and c^i denote the width, height, and channels of the visual feature at the i-th convolutional block, respectively. From the input visual feature f^{i}, the generator estimates the boundary key point map, $\hat{\mathrm{M}}^i \in \mathbb{R}^{w^i \times h^i \times 1}$. Then, the generated boundary key point map $\hat{\mathrm{M}}^i$ is used for preserving structure boundary information of f^{i} similar to residual attention scheme [28]. The structure boundary information preserved feature $\mathrm{v}^{\mathrm{i}} \in \mathbb{R}^{\mathrm{w^i} \times \mathrm{h^i} \times \mathrm{c^i}}$ can be written as

$$\mathrm{v}^i = \mathrm{f}^i \oplus (\mathrm{f}^{\mathrm{i}} \otimes \hat{\mathrm{M}}^{\mathrm{i}}), \quad (2)$$

where $\oplus$ and $\otimes$ denote element-wise addition and channel-wise multiplication, respectively. In the multiplication process, $\hat{\mathrm{M}}^i$ is broadcasted for each channel number. The structure boundary preserved feature, v^i is fed into (i+1)-th convolution block.

3.2.1 Boundary Key Point Map Generator

Figure 3 shows the architecture of the proposed boundary key point map generator. As shown in Figure 3, the proposed boundary key point map generator is designed to estimate the boundary key point map considering various receptive fields. In our boundary key point map generator, by employing the dilated convolution[31], the generator can effectively encode/decode the features with a various range of receptive fields. Let $d_r^s(\cdot)$ denote dilated convolution function with a dilation rate r and filter size of $s \times s$. The encoded feature maps $d_r^s\left(\mathrm{f}^i\right)$ with various receptive fields are concatenated across channel-wise and projected into the boundary key point map space. The generated boundary key point map $\hat{\mathrm{M}}^i$ can be written as

$$\hat{\mathrm{M}}^i = \sigma\left(d_1^1\left\{\left[d_1^1(\mathrm{f}^i), d_1^3(\mathrm{f}^i), d_2^3(\mathrm{f}^i), d_4^3(\mathrm{f}^i), d_6^3(\mathrm{f}^i)\right]\right\}\right) \quad (3)$$

where σ denotes a sigmoid function.

The boundary key point map generator is optimized by minimizing the cross-entropy loss between the estimated boundary key point map $\hat{\mathrm{M}}^i$ and ground-truth boundary key point map M_{GT}^i. The objective function for the boundary key point map generation is defined as

$$L_{Map}^i = -\mathrm{M}_{GT}^i \cdot \log \hat{\mathrm{M}}^i - (1 - \mathrm{M}_{GT}^i) \cdot \log(1 - \hat{\mathrm{M}}^i). \quad (4)$$

3.3. Shape Boundary-aware Evaluator (SBE)

To embed experts knowledge in the fully automatic segmentation models, we propose a novel structure boundary information-based discrimination network named as Shape Boundary-aware Evaluator (SBE). The SBE gives feedback to the segmentation network by using the boundary key point map. To this end, the SBE takes the boundary key point map and segmentation image (predicted or the ground-truth) as input. The boundary key point map and the given segmentation image are concatenated across the channel-wise and fed into the SBE network. Then, the SBE is to evaluate whether the segmentation results are consistent with the boundary key point map or not. Given the ground-truth segmentation map S_{GT} and boundary key point map, the SBE provides a high evaluation score. On the other hand, with the poorly predicted segmentation map and boundary key point map, the SBE provides a low evaluation score since the poorly predicted segmentation map is not consistent with the boundary key point map. To this end, we trained the SBE network with the following loss.

$$\begin{aligned} L_{SBE} = &- \log\left(D\left(\mathrm{S}_{GT}; \mathrm{M}_{GT}\right)\right) \\ &- \log\left(1 - D(\hat{\mathrm{S}}_{Pred}; \mathrm{M}_{GT})\right). \end{aligned} \tag{5}$$

where $D(\cdot)$ denotes the SBE function that projects the input segmentation map and boundary key point map onto the evaluation score. $\hat{\mathrm{S}}_{Pred}$ denotes the predicted segmentation result. Detailed structure of the SBE is described in supplementary material (Supplementary A.2).

3.4. Training Segmentation Network

To train the segmentation network including the proposed BPBs and the proposed SBE in an adversarial way, we employ three types of loss functions. The first one is a segmentation loss function to reduce the difference between the ground-truth segmentation map and the predicted segmentation map. It is defined as

$$\begin{aligned} L_{Seg} = &- \mathrm{S}_{GT} \cdot \log\left(\hat{\mathrm{S}}_{Pred}\right) \\ &- (1 - \mathrm{S}_{GT}) \cdot \log\left(1 - \hat{\mathrm{S}}_{Pred}\right). \end{aligned} \tag{6}$$

The second one is a key point map loss which is devised for the proposed key point map generation (see Eq. 4). The last one is a boundary aware loss considering the back-propagation of the SBE. The boundary aware loss is designed to allow the segmentation network to achieve accurate segmentation results by considering boundary key points. It can be written as

$$L_{BA} = - \log\left(D\left(\hat{\mathrm{S}}_{Pred}; \mathrm{M}_{\mathrm{GT}}\right)\right). \tag{7}$$

By reducing the boundary aware loss, the segmentation network predicts structure boundary preserved segmentation result.

Finally, the total loss function for training segmentation network can be defined as

$$L_{Total} = L_{Seg} + L_{BA} + \sum_{i=1}^{l} L_{Map}^{i}, \tag{8}$$

where l denotes the total number of BPBs in the segmentation network and L_{Total} denotes the segmentation training loss.

4. Experimental Results

4.1. Datasets

We conduct experiments to verify our proposed structure boundary preserving method on two medical image segmentation datasets. The first one is PH2 [31]+ISBI 2016 [10] Skin Lesion Challenge dataset and the other one our own Transvaginal Ultrasound (TVUS) dataset.

The PH2 + ISBI 2016 dataset is a publically available dataset for evaluating the skin lesion segmentation. ISBI 2016 Skin Lesion Challenge dataset includes 900 skin lesion images with different image size. PH2 dataset includes 200 dermoscopic images. For the training, we use 900 images which are from the ISBI 2016 dataset. Then, for the testing, we use 200 images which are from the PH2 dataset. We follow the experimental protocols in [3].

The second dataset is Transvaginal Ultrasound (TVUS) dataset. The TVUS dataset is a private database collected for our experiments. It consists of 3,360 transvaginal ultrasound images and the corresponding endometrium segmentation maps. The endometrium segmentation maps are annotated by three expert gynecologists. For the evaluation, we conduct five-fold cross-validation. Therefore, we use 2,688 images for training and 672 images for testing at each fold.

4.2. Implementation Details

We optimize our model using ADAM optimizer [13] with learning rate 0.0001 both in the segmentation network and SBE network. We train the networks from scratch with randomly initialized weights with 8 input batches. For every iteration, we train the segmentation network 8 times and the SBE network 3 times to train two networks in an adversarial manner.

To make boundary key point maps, we select 6 points and perform the boundary key point selection algorithm 40000 times (n=6, T=40000). Beforehand, we train the network, we create the boundary key point maps in advance.

To verify the proposed method, we integrate our method to segmentation networks. U-Net [21], FCN [15], and

Table 1. Dice and Jaccard coefficient comparison of our approach and six different approaches on PH2 + ISBI 2016 Challenge dataset.

Method	Dice Coefficient	Jaccard Coefficient
SCDRR [4]	86.00	76.00
JCLMM [23]	82.85	-
MSCA [2]	81.57	72.33
SSLS [1]	78.38	68.16
FCN [15]	89.40	82.15
Bi et al. (2017) [3]	90.66	83.99
FCN+BPB+SBE (Our method)	**91.84**	**84.30**

Table 2. Dice and Jaccard coefficient comparison of our approach and conventional segmentation network on TVUS dataset.

Method	Dice Coefficient	Jaccard Coefficient
U-Net [21]	82.30	70.38
FCN [15]	81.19	69.12
Dilated-Net [31]	82.40	70.36
Park et al. (2019) [18]	82.67	70.46
Dilated-Net+BPB+SBE (Our method)	**83.52**	**71.58**

Dilated-Net [31]. Detailed structures of the integrated networks are described in a supplementary material (Supplementary A.1).

4.3. Quantitative Evaluation

To demonstrate the advantage of our structure boundary preserving segmentation framework, we compare our method with other methods. As for evaluation metrics, we utilize a Dice Coefficient Score (DCS) and a Jaccard Coefficient Score (JCS).

Table 1 quantitatively compares our method with six state-of-the-art segmentation methods on PH2+ISBI 2016 dataset (SCDRR [4], JCLMM [23], MSCA [2], SSLS [1], FCN [15], Bi et al [3]). As seen in Table 1, our method achieves 91.84% in DCS and 84.30% in JCS. By integrating BPB and SBE into FCN network, it improves 2.44%, 2.15% in DCS and JCS, respectively. Also, our method shows 1.18% higher accuracy than recent state-of-the-art skin lesion segmentation method [3]. Overall, our approach achieves better performance than other methods.

To further evaluation of the proposed method, we conduct more experiment on challenging TVUS dataset. Table 2 shows the comparison with conventional segmentation networks and a state-of-the-art endometrium segmentation method. As seen in Table 2, our method achieves 83.52% in DCS and 71.58% in JCS respectively. By integrating BPB and SBE into Dilated-Net, it improves 1.12%, 1.22% in DCS and JCS, respectively. Also, our method shows 0.85% higher accuracy than recent state-of-the-art endometrium segmentation method [18]. These results are

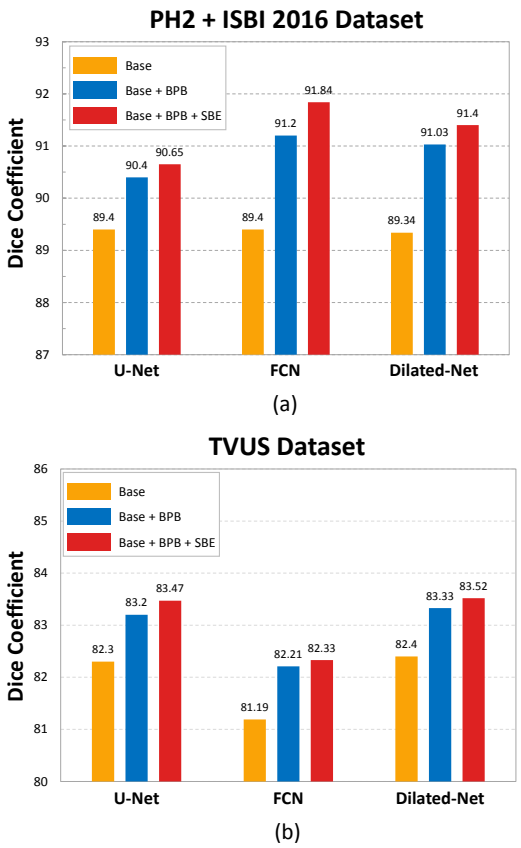

Figure 4. Ablation results on PH2+ISBI 2016 (a) and TVUS (b) datasets. The blue bars are the DCS of the baseline segmentation network (U-Net, FCN, Dilated-Net). The red bars are the DCS that training the baseline network with BPB. The green bars are the DCS that training the baseline network with BPB and SBE.

attributed to the efficient encoding of the structure boundary through BPB and SBE.

4.4. Ablation Studies for Generalization

One of the main contributions is that our proposed BPB and SBE can be general and flexible enough to be applicable to any automatic medical image segmentation models. To verify that, we integrate BPB and SBE into three different segmentation network; U-Net, FCN, Dilated-Net.

Figure 4 (a) shows the ablation studies results on PH2+ISBI 2016 dataset. Firstly, by adopting the BPB to baseline segmentation network, the DCS increase by 1.0%, 1.8%, and 1.69% on U-Net, FCN, and Dilated-Net, respectively. In addition, as SBE is additionally adopted, DCS is further increased by 0.25%, 0.64%, and 0.37% on U-Net, FCN, and Dilated-Net, respectively.

Similarly, Figure 4 (b) shows the results on TVUS dataset. By applying the BPB to baseline segmentation network, the DCS is increased by 0.9%, 1.02%, 0.93% Dice coefficient on U-Net, FCN, and Dilated-Net, respectively.

Table 3. Statistical significance analysis of performance improvements by paired t-test on PH2+ISBI 2016 dataset.

Baseline Network	Mean difference ± Standard Error	95% CI	*p*-value
U-Net [21]	1.22±0.21	[0.51, 1.35]	p<0.0001
FCN [15]	2.17±0.28	[1.41, 2.72]	p<0.0001
Dilated-Net [31]	2.03±0.25	[1.34, 2.46]	p<0.0001

Table 4. Statistical significance analysis of performance improvements by paired t-test on TVUS dataset.

Baseline Network	Mean difference ± Standard Error	95% CI	*p*-value
U-Net [21]	0.92±0.13	[0.66, 1.17]	p<0.0001
FCN [15]	1.62±0.28	[1.07, 2.17]	p<0.0001
Dilated-Net [31]	0.90±0.11	[0.62, 1.1]	p<0.0001

In addition, by applying the SBE, it is increased by 0.27%, 0.12%, and 0.19% respectively.

For the further analysis that the performance improvement is statistical significance, we conduct paired t-test. It provides the statistical evaluation and qualification procedure of the segmentation models. We compare performance improvement between baseline network and baseline network+BPB+SBE. The results are shown in Table 3 and Table 4. As shown in the tables, the performance improvement is statistical significance with p-value < 0.0001. These results indicate that our proposed structure boundary preserving segmentation framework could be applicable to segmentation model effectively and improve the performance with statistical significance.

4.5. Effect of Multiple BPBs

In this section, we investigate the effect of multiple BPBs. To verify this, we place the BPB in 5 ways. Table 5 shows the performance comparison along with the number of BPBs on U-Net. The experiment is conducted on TVUS dataset. The 5 variants are as follows:

- **Encoder (front)**: A BPB in the first layer of U-Net encoder.
- **Decoder(end)**: A BPB in the last layer of U-Net decoder.
- **Center (1)**: A BPB in the center of U-Net
- **Center (3)**: 3 BPBs after 8 convolution layers
- **Center (6)**: 6 BPBs after 4 convolution layers. It is our proposed structure.

As shown in the Table 5, when we place the BPB the first layer of U-Net encoder the performance decreased to 82.15%. Since the first layer of the network extracts low level feature, it is hard to predict key-point. Therefore, the BPB can't work well. In the case of Decoder (end), the performance increase. However, since the structure boundary information can't preserve during the encoding, the performance improvement is marginal. With the multiple BPBs from center layer (Center (1) $\sim$ Center (3)), they show as the number of BPBs increases, the performance increases. These results can be interpreted that it is important to preserve structure boundary during the encoding and decoding. Therefore, it would be better to use multiple BPBs.

Table 5. Performance changes along with the number of BPBs on U-Net.

Method	Dice Coefficient
Encoder(front)	82.15
Decoder(end)	82.43
Center (1)	82.47
Center (3)	82.66
Center (6)	**83.20**

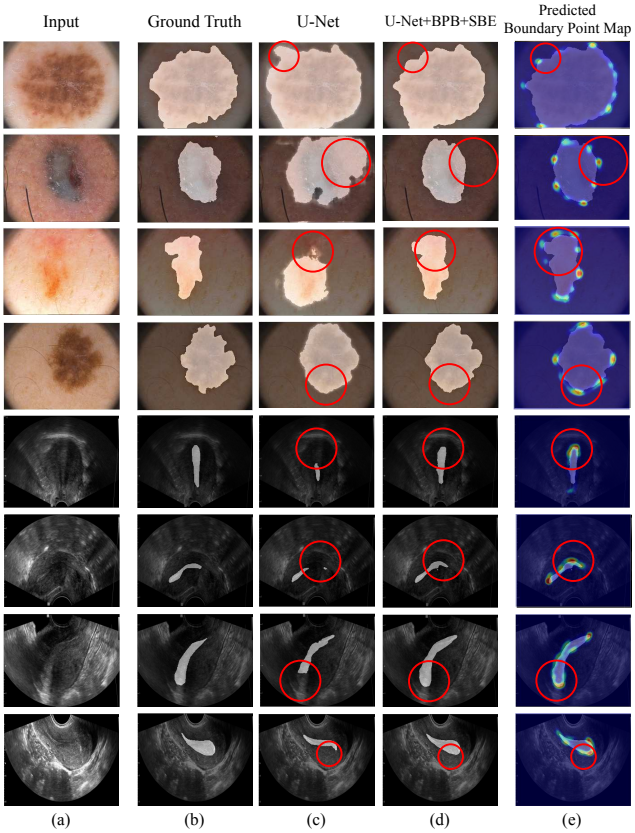

Figure 5. Segmentation results comparison of U-Net and U-Net+BPB+SBE method on PH2+ISBI 2016 (1-3 rows) and TVUS (4-6 rows). (a) is the original images, (b) is the ground truth segmentation images, (c) is the results of the U-Net, (d) is the U-Net+BPB+SBE and (e) is the visualization results of the generated key point map.

4.6. Qualitative Evaluation

To demonstrate that our proposed BPB and SBE preserve structure boundary of the target region during the segmentation, we visualize our segmentation results. Figure 5

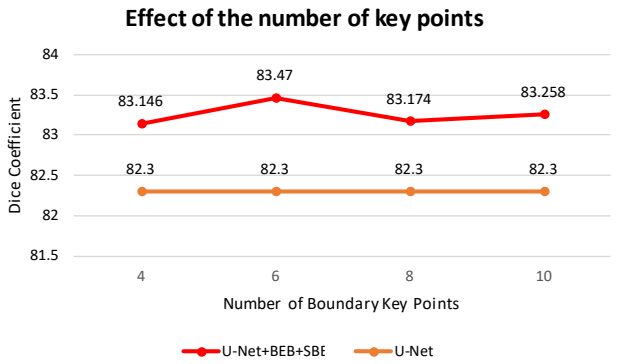

Figure 6. Performance evaluation in accordance with the number of boundary key points on TVUS dataset. It shows the comparison results between U-Net and U-Net + BPB + SBE.

visually compares segmentation results from our proposed method (U-Net+BPB+SBE) and U-Net method. First to the fourth columns are the segmentation results on PH2+ISBI 2016 dataset and Fifth to eighth columns are the segmentation results on TVUS dataset.

As shown in red circles in the Figure 5, our approach preserves structure boundary of target region. For the analysis, we visualize the generated boundary key point maps. The boundary key point maps are generated from the last BPB. They are shown in the Figure 5 (e). The results show that the tendency of preserved boundaries is obtained with a corresponding boundary key point map. Even the images have ambiguous structure boundary or heterogeneous texture, our method preserves structure boundary of target region; on the other hand, segmentation results obtained from U-Net failed to preserve boundary.

Refer to the supplementary materials for more results of U-Net, FCN and Dilated-Net (Supplementary B).

5. Discussion

5.1. Define the Number of Key Points

In our experiment, we used fixed number of key points. To define the number of key points, we scrutinized the performance changes according to the number of points. Figure 6 shows the experiment results on U-Net + BPB + SBE network. We performed this experiment on TVUS dataset. As shown in the Figure 6, we can find that when we use 6 key points, the results show best performance. However, this number of key points is not optimal solution for all objects. For example, in the case of brain tissue which has many folded structures to segment, it requires more key points and iterations. It would be interesting to investigate the relation between the number of key points and object shape. Detailed experiment results are shown in the supplementary material Figure 16.

5.2. Future Work

There are several avenues for future work. We focused on the preserving structure boundary of target object in medical image segmentation. The idea can be further improved by generalizing the method to general object segmentation. The method may vary, but preserving structure boundary points will still be an important issue. Also, it is interesting to adapt our method to recently proposed segmentation network. We describe more experiment results and further discussion in the supplementary material.

6. Conclusion

This paper presents a novel fully automatic segmentation framework for medical image segmentation with an ambiguous boundary. To preserve the structure boundary of the target object in the medical image, we embed structure boundary key points into segmentation network. To this end, we generate key point map through boundary key point selection algorithm. The generated key point map is used for training our novel structure boundary preserving block BPB and SBE. In the BPB, the boundary key point map generator generates boundary key point map. It allows segmentation network to further exploit the structure boundary of the target object. Then, to embed experts knowledge to segmentation network, we train the SBE in an adversarial manner. The proposed SBE tries to evaluate clearly between the predicted segmentation map and the ground-truth segmentation map for given medical image. Experimental results demonstrate that the proposed framework can easily be integrated into various segmentation networks. Then, the proposed method improves accuracy with statistical significance.

Acknowledgment

This work was supported by Institute for Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (No.2017-0-01779, A machine learning and statistical inference framework for explainable artificial intelligence). The corresponding author of this project is Yong Man Ro.

References

[1] Euijoon Ahn, Lei Bi, Youn Hyun Jung, Jinman Kim, Changyang Li, Michael Fulham, and David Dagan Feng. Automated saliency-based lesion segmentation in dermoscopic images. In *2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pages 3009–3012. IEEE, 2015. 6

[2] Lei Bi, Jinman Kim, Euijoon Ahn, Dagan Feng, and Michael Fulham. Automated skin lesion segmentation via image-wise supervised learning and multi-scale superpixel based

cellular automata. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 1059–1062. IEEE, 2016. 6

[3] Lei Bi, Jinman Kim, Euijoon Ahn, Ashnil Kumar, Michael Fulham, and Dagan Feng. Dermoscopic image segmentation via multistage fully convolutional networks. *IEEE Transactions on Biomedical Engineering*, 64(9):2065–2074, 2017. 5, 6

[4] Behzad Bozorgtabar, Mani Abedini, and Rahil Garnavi. Sparse coding based skin lesion segmentation using dynamic rule-based refinement. In *international workshop on machine learning in medical imaging*, pages 254–261. Springer, 2016. 6

[5] John Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 1986. 4

[6] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017. 1

[7] Patrick Ferdinand Christ, Mohamed Ezzeldin A Elshaer, Florian Ettlinger, Sunil Tatavarty, Marc Bickel, Patrick Bilic, Markus Rempfler, Marco Armbruster, Felix Hofmann, Melvin DAnastasi, et al. Automatic liver and lesion segmentation in ct using cascaded fully convolutional neural networks and 3d conditional random fields. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 415–423. Springer, 2016. 2

[8] Mehmet Ufuk Dalmış, Geert Litjens, Katharina Holland, Arnaud Setio, Ritse Mann, Nico Karssemeijer, and Albert Gubern-Mérida. Using deep learning to segment breast and fibroglandular tissue in mri volumes. *Medical physics*, 44(2):533–546, 2017. 2

[9] Mehmet Ufuk Dalmış, Geert Litjens, Katharina Holland, Arnaud Setio, Ritse Mann, Nico Karssemeijer, and Albert Gubern-Mérida. Using deep learning to segment breast and fibroglandular tissue in mri volumes. *Medical physics*, 44(2):533–546, 2017. 2

[10] David Gutman, Noel CF Codella, Emre Celebi, Brian Helba, Michael Marchetti, Nabin Mishra, and Allan Halpern. Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016, hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1605.01397*, 2016. 5

[11] Konstantinos Kamnitsas, Christian Ledig, Virginia FJ Newcombe, Joanna P Simpson, Andrew D Kane, David K Menon, Daniel Rueckert, and Ben Glocker. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical image analysis*, 36:61–78, 2017. 1

[12] Jung Uk Kim, Hak Gu Kim, and Yong Man Ro. Iterative deep convolutional encoder-decoder network for medical image segmentation. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 685–688. IEEE, 2017. 2

[13] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5

[14] Simon Kohl, Bernardino Romera-Paredes, Clemens Meyer, Jeffrey De Fauw, Joseph R Ledsam, Klaus Maier-Hein, SM Ali Eslami, Danilo Jimenez Rezende, and Olaf Ronneberger. A probabilistic u-net for segmentation of ambiguous images. In *Advances in Neural Information Processing Systems*, pages 6965–6975, 2018. 2

[15] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015. 1, 2, 5, 6, 7

[16] Dong Nie, Li Wang, Roger Trullo, Jianfu Li, Peng Yuan, James Xia, and Dinggang Shen. Segmentation of craniomaxillofacial bony structures from mri with a 3d deep-learning based cascade framework. In *International Workshop on Machine Learning in Medical Imaging*, pages 266–273. Springer, 2017. 2

[17] Dong Nie, Li Wang, Lei Xiang, Sihang Zhou, Ehsan Adeli, and Dinggang Shen. Difficulty-aware attention network with confidence learning for medical image segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1085–1092, 2019. 1

[18] Hyenok Park, Hong Joo Lee, Hak Gu Kim, Yong Man Ro, Dongkuk Shin, Sa Ra Lee, Sung Hoon Kim, and Mikyung Kong. Endometrium segmentation on tvus image using keypoint discriminator. *Medical physics*, 2019. 1, 6

[19] Sang Hyun Park, Yaozong Gao, Yinghuan Shi, and Dinggang Shen. Interactive prostate segmentation based on adaptive feature selection and manifold regularization. In *International Workshop on Machine Learning in Medical Imaging*, pages 264–271. Springer, 2014. 2

[20] Martin Rajchl, Matthew CH Lee, Ozan Oktay, Konstantinos Kamnitsas, Jonathan Passerat-Palmbach, Wenjia Bai, Mellisa Damodaram, Mary A Rutherford, Joseph V Hajnal, Bernhard Kainz, et al. Deepcut: Object segmentation from bounding box annotations using convolutional neural networks. *IEEE transactions on medical imaging*, 36(2):674–683, 2016. 2

[21] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 1, 2, 5, 6, 7

[22] Holger R Roth, Hirohisa Oda, Xiangrong Zhou, Natsuki Shimizu, Ying Yang, Yuichiro Hayashi, Masahiro Oda, Michitaka Fujiwara, Kazunari Misawa, and Kensaku Mori. An application of cascaded 3d fully convolutional networks for medical image segmentation. *Computerized Medical Imaging and Graphics*, 66:90–99, 2018. 2

[23] Anandarup Roy, Anabik Pal, and Utpal Garain. Jclmm: A finite mixture model for clustering of circular-linear data and its application to psoriatic plaque segmentation. *Pattern Recognition*, 66:160–173, 2017. 6

[24] Berkman Sahiner, Aria Pezeshk, Lubomir M Hadjiiski, Xiaosong Wang, Karen Drukker, Kenny H Cha, Ronald M Summers, and Maryellen L Giger. Deep learning in medical imaging and radiation therapy. *Medical physics*, 46(1):e1–e36, 2019. 1

[25] Ran Shi, King Ngi Ngan, Songnan Li, and Hongliang Li. Interactive object segmentation in two phases. *Signal Processing: Image Communication*, 65:107–114, 2018. 2

[26] Eugene Vorontsov, An Tang, Chris Pal, and Samuel Kadoury. Liver lesion segmentation informed by joint liver segmentation. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1332–1335. IEEE, 2018. 2

[27] Bo Wang, K Wei Liu, K Marcel Prastawa, Andrei Irima, Paul M Vespa, John D Van Horn, P Thomas Fletcher, and Guido Gerig. 4d active cut: An interactive tool for pathological anatomy modeling. In *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, pages 529–532. IEEE, 2014. 2

[28] Fei Wang, Mengqing Jiang, Chen Qian, Shuo Yang, Cheng Li, Honggang Zhang, Xiaogang Wang, and Xiaoou Tang. Residual attention network for image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3156–3164, 2017. 4

[29] Guotai Wang, Maria A Zuluaga, Wenqi Li, Rosalind Pratt, Premal A Patel, Michael Aertsen, Tom Doel, Anna L David, Jan Deprest, Sébastien Ourselin, et al. Deepigeos: a deep interactive geodesic framework for medical image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 41(7):1559–1572, 2018. 2

[30] Tao Wang, Jian Yang, Zexuan Ji, and Quansen Sun. Probabilistic diffusion for interactive image segmentation. *IEEE Transactions on Image Processing*, 28(1):330–342, 2018. 2

[31] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015. 1, 2, 4, 5, 6, 7

[32] Qihang Yu, Lingxi Xie, Yan Wang, Yuyin Zhou, Elliot K Fishman, and Alan L Yuille. Recurrent saliency transformation network: Incorporating multi-stage visual cues for small organ segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8280–8289, 2018. 2

[33] Xiaomei Zhao, Yihong Wu, Guidong Song, Zhenye Li, Yazhuo Zhang, and Yong Fan. A deep learning model integrating fcnns and crfs for brain tumor segmentation. *Medical image analysis*, 43:98–111, 2018. 1

[34] Yuyin Zhou, Lingxi Xie, Wei Shen, Yan Wang, Elliot K Fishman, and Alan L Yuille. A fixed-point model for pancreas segmentation in abdominal ct scans. In *International conference on medical image computing and computer-assisted intervention*, pages 693–701. Springer, 2017. 2