DEEP LEVEL SET WITH CONFIDENCE MAP AND BOUNDARY LOSS FOR MEDICAL IMAGE SEGMENTATION

Yaoyue Zheng, Zhang Chen, Xiaojian Li, Xiangyu Si, Liangjie Dong and Zhiqiang Tian

School of Software Engineering, Xi'an Jiaotong University, Xi'an 710049, China
yaoyz105@gmail.com, 1900938761@qq.com, xiaojianli@stu.xjtu.edu.cn,
sixiangyu@126.com, liannedlj@stu.xjtu.edu.cn, zhiqiangtian@xjtu.edu.cn

ABSTRACT

Level set method is widely used for image segmentation. Recent work combined traditional level set method with deep learning architecture for image segmentation. However, it is limited when dealing with medical images because of the blurred edges and complex intensity distribution, which leads to the loss of spatial details. To address this problem, we propose a deep level set method to refine object boundary details and improve the segmentation accuracy. We integrate augmented prior knowledge into inputs of CNN, which can make the level set evolution result has more accurate shape. In addition, to consider the spatial correlation of the object, we combine a boundary loss with deep level set model for preventing the reduction of details. We evaluate the proposed method on two medical image data sets, which are prostate magnetic resonance images and retinal fundus images. The experimental results show that the proposed method achieves state-of-the-art performance.

Index Terms— Boundary loss, confidence map, deep level set, medical image segmentation

1. INTRODUCTION

Image segmentation is associated with many application scenarios, which requires dense predictions. That is, each pixel receives an output classification. In Particular, image segmentation is a highly relevant task in medical image analysis that expects highly accurate segmentation results, but at the same time objects have blurred edges and complex intensity distribution. Therefore, one of the main research subjects in medical image segmentation is how to acquire accurate segmentation result with these limitations.

Over the past years, researchers have explored different segmentation methods and highlighted the capabilities of Convolutional Neural Networks (CNNs) that achieved state-of-the-art performances. Most studies focused on inferring a class label for each pixel independently [1–4], which may come short at yielding compact and uniform predictions. Level set is a traditional method that widely used in image segmentation, which can solve this problem. With the level set formulation, the segmentation problem can be seen as an energy minimization problem. However, traditional level set method was usually used as post-processing tool [5]. A sensible solution is to combine deep learning architecture with level set method. Recent works [5–7] proposed deep level set methods for image segmentation, which combine the traditional level set method with deep learning architecture in an end-to-end fashion. However, these methods may produce segmentation result with inaccurate boundaries due to the features of medical images mentioned above. In addition, some studies [8, 9] explored the effect of prior knowledge in segmentation tasks by treating prior knowledge as an additional input to CNN, and they demonstrated its effectiveness. In a recent work [7], the authors combined deep level set method with prior knowledge to get more accurate segmentation result. Although this method have achieved outstanding performance on natural image segmentation, but was still insufficiently accurate for clinical use because of the loss of object boundary details.

In this work, we aim to overcome the aforementioned problems and improve the medical images segmentation performance. We propose a deep level set model combined with augmented prior knowledge and boundary loss to refine object boundary details and further improve the segmentation accuracy. We utilize the confidence map (CM) as an augmented prior knowledge and combine CM with input image to feed into the CNN, which can make the level set evolution result has more accurate shape. In addition, we use a boundary (BD) loss to overcome the drawbacks of the loss function used in traditional segmentation method, which calculates the loss at each pixel independently. The boundary loss can make the segmentation result has more boundary details. The proposed method can be trained in an end-to-end manner. An overview of the proposed method is shown in Fig. 1.

We summarize our contributions as follows: i) We present the confidence map as an augmented prior knowledge, which can make the level set evolution result align with the object boundary more accurately. ii) We add a boundary loss to supervise the training process jointly with the cross entropy (CE) loss, which can introduce spatial correlation to

978-1-7281-1331-9/20/$ 31.00

deep level set model. iii) Compared with the state-of-the-art method, the proposed method improves the Dice similarity coefficient (DSC) score 0.4% and 0.2% on the Prostate Magnetic Resonance (MR) [10] images and 2D Retinal Fundus Glaucoma (REFUGE) [11] images, respectively. Ablation study and result visualization further demonstrate its effectiveness.

2. RELATED WORK

2.1. Prior Knowledge for Image Segmentation

Human-assisted segmentation methods are very important for reliable extraction of the regions of interest (RoI) because user interactions can provide supervisory signal to guide segmentation networks. The usual practice is to provide some prior knowledge before the segmentation, such as the location of the foreground and background, which can help CNNs perform the segmentation task.

In this field, researchers have devoted significant efforts to simplify user interaction. GrabCut [12] successfully simplified user interaction by dragging a rectangle around the desired object. Extreme Clicking [13] used a novel approach to replace the traditional bounding box method. The bounding box was obtained by clicking on the left-most, right-most, top and bottom pixels of objects. Different with this strategy, DEXTR [8] took the four extreme points as a form of guidance for deep neural network to implement segmentation task and demonstrated its effectiveness. In a recent work [9], the authors made a further improvement by integrating augmented prior knowledge for inputs of CNN. By using prior knowledge, the network can predict more accurate segmentation results [8].

2.2. Deep Level Set Method for Segmentation

Level Set Method is presented by Osher and Sethian [14] for front propagation. The idea behind level set method is to imbed a curve within a surface, that is, transform the low-dimensional curve evolution problem to a high-dimensional surface evolution problem. In the level set framework, the curve is updated iteratively by moving along the descent of the level set energy. By minimizing the level set energy, the final zero level set can be obtained as the object boundary.

The application of level set method in image segmentation has become increasingly popular. Most previous methods used the level set evolution as a post-processing tool [5]. They are time consuming and rely on a good initialization. In addition, they cannot train the model end-to-end. To solve these problems, Tang et al. [5] integrated level set model with FCNs and iteratively fine-tuned the FCNs by using level set module. Marcos et al. [6] employed CNN to predict parameters of the active contour models that rely on a good curve initialization. Wang et al. [7] let CNN to predict initial level set function and evolution parameters by combining CNN with

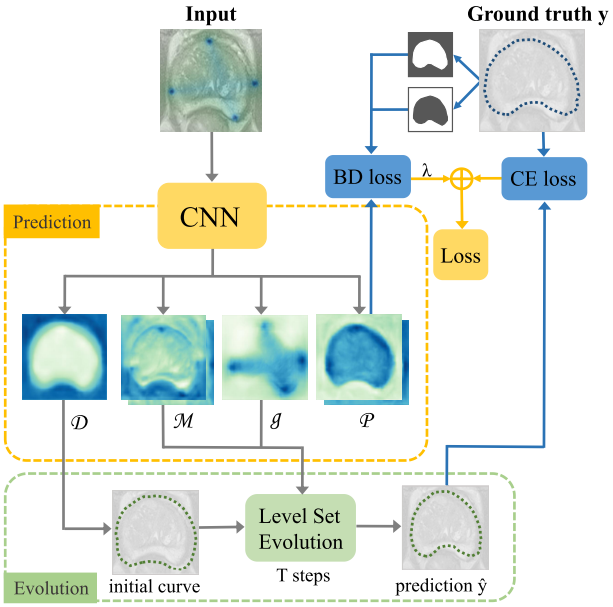

Fig. 1. Overview of the proposed method.

level set method. Therefore, the CNN and level set module can be trained end-to-end. However, for the medical images, the level set evolution parameters predicted by CNN are not accurate enough and become a bottleneck of the segmentation accuracy.

3. METHODS

3.1. Overview of the Proposed Method

An overview of the proposed method is shown in Fig. 1. Given an input image, we can click on the top, bottom, left-most and right-most parts of the RoI to get four extreme points. Moreover, we can get two intersecting segments by pairing and connecting the four extreme points. Each pixel in the image has a distance from the intersecting line segments. We can use the distances to generate a confidence map, which assigns a confidence score to each pixel in the image domain. Finally, we combine the extreme points with the confidence map together as an additional channel and concatenate it with the input image to feed into the network.

The network structure is shown in Fig. 2. ResNet-101 is used as the encoder of CNN. Feature maps will be extracted based on different scales and concatenated to feed into four prediction branches separately. Through the prediction branches, CNN predicts three level set evolution elements, which are initial curve, external energy $\mathcal{M}$ and internal energy g. The initial curve is represented by a truncated signed distance function [7] $\mathcal{D}$, where the zero distance set represent the initial curve. $\mathcal{M}$ and g is used to perform the level set evolution. After T steps of evolution, the prediction result $\hat{y}$ is derived from the initial curve. Then the cross-entropy (CE)

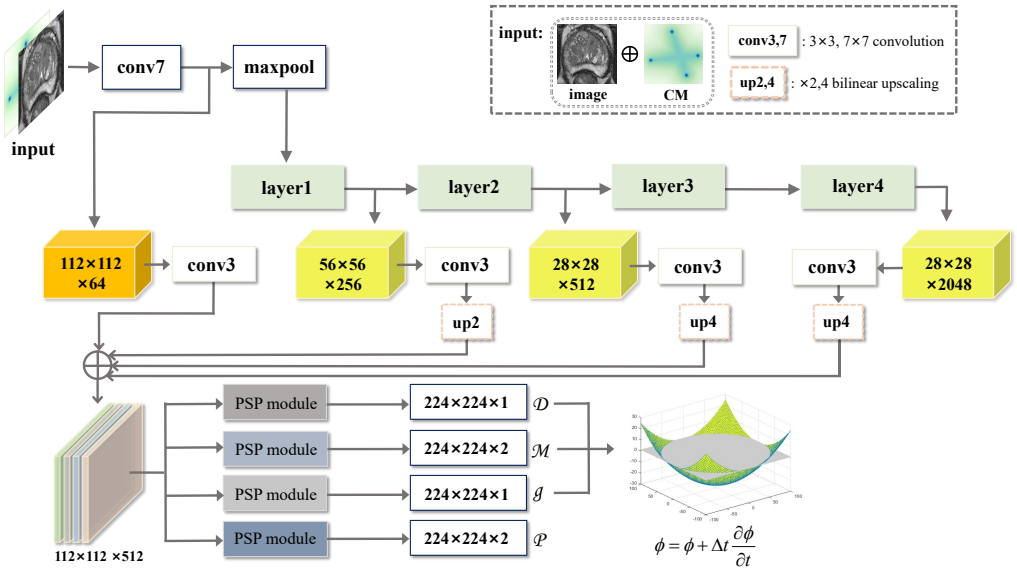

Fig. 2. Overview of the CNN network.

loss is calculated by using the prediction $\hat{y}$ and ground truth y. In addition, the fourth branch of the network predicts probability maps $\mathcal{P}$. The boundary (BD) loss is calculated by using $\mathcal{P}$ and binary masks which obtained by y. Finally, the CNN parameters are updated by backpropagation.

3.2. Confidence Map

In order to improve the segmentation performance, we expand the prior knowledge based on the extreme points. The main idea is to use the four extreme points to generate a confidence map [9]. The confidence map can be used as an augmented prior knowledge to help the training process.

Through the extreme points we can get the following information suitable for most situations: 1) Four extreme points can be paired and connected to form two intersecting line segments s_1, s_2. 2) The intersection point o of the two line segments is more likely belong to the RoI. 3) The point x in the image Ω have a distance from the line segments, we can assume that the smaller the distance between x and the two line segments, the more likely it is belong to the RoI [9], and vice versa. The confidence map can be calculated as the following formula:

$$d_{s1}(x) = \frac{dist(x, s1)}{\sigma_{s1}},\ d_{s2}(x) = \frac{dist(x, s2)}{\sigma_{s2}}, \tag{1}$$

where $dist(x, s)$ is the distance from point x to line segment s. σ_s approximates the variance along segment s.

$$\begin{aligned} d_1(x) &= min\{d_{s1}(x), d_{s2}(x)\}, \\ d_2(x) &= \sqrt{d_{s1}(x)^2 + d_{s2}(x)^2}, \end{aligned} \tag{2}$$

where $d_1(x)$ and $d_2(x)$ measure the Chebyshev and Mahalonobis distance of the point x to line segment s_1 and s_2. The confidence map can be obtained by:

$$CM(x) = \frac{1}{1 + d_1(x) + d_2(x)}. \tag{3}$$

3.3. Deep Level Set with Boundary Loss

In the level set framework, curve $\mathcal{C}$ can be represented as the (zero) level line crossing the level set function $\phi(x, y)$:

$$\mathcal{C} = \{(x, y) | \phi(x, y) = 0\}. \tag{4}$$

The segmentation of the image is computed by locally minimizing an appropriate energy function. The key is to evolve the initial boundary $\mathcal{C}$ in the direction of the negative energy gradient, which is done based on evolution of the level set function $\phi(x, y, t)$:

$$\frac{\partial \mathcal{C}}{\partial t} = \frac{\partial \phi}{\partial t} = -|\phi| F, \tag{5}$$

where F is the force along the normal direction of the cruve.

The traditional level set method cannot train end-to-end with CNN. Inspired by deep level set [7], we propose an end-to-end segmentation method, as shown in Fig. 1. The first branch predicts a truncated signed distance map $\mathcal{D}$, and the initial curve for the level set evolution can be represented by the zero distance set of $\mathcal{D}$. The second branch predicts motion maps $\mathcal{M}$ based on the image gradient, which can promote the curve to the desired position. The third branch predicts a modulation map $g \in [0, 1]$ to keep curvature in the real sharp corners. Moreover, The fourth branch predicts probability map for each class to further refine boundary details and improve the prediction accuracy. Weighted binary cross entropy loss

Table 1. Ablation study on prostate MR dataset.

Model	DSC (%)	RVD (%)	HD (mm)	ABD (mm)
Baseline	96.0 ± 0.7	-5.8 ± 1.7	7.4 ± 3.6	1.2 ± 0.2
+CM	96.4 ± 0.5	-3.7 ± 1.7	7.3 ± 3.9	1.1 ± 0.2
+BD	96.2 ± 0.6	-3.4 ± 1.9	6.9 ± 3.0	1.1 ± 0.2
+CM+BD	96.4 ± 0.5	-2.9 ± 1.9	6.8 ± 3.1	1.0 ± 0.2

was employed by the previous work [7], which used to monitor the evolution results of the curve:

$$L_{ce} = -\Bigg(\sum_{(i,j)} w_p Y(i,j) log H(\hat{\phi}_T(i,j)) + w_n(1 - Y(i,j)) log(1 - H(\hat{\phi}_T(i,j))) \Bigg), \quad (6)$$

where w_p and w_n are the weights of the foreground and background classes, respectively. Y is ground-truth and $H(\hat{\phi})$ is evolution result. H is the Heaviside function.

Traditional cross-entropy loss function cannot consider the global information of the object [15]. In this paper, we aim to let the network to learn more discriminative features and improve the segmentation performance. Inspired by [15], we take the spatial correlation into account by combine a boundary loss with the weighted binary cross entropy loss as follows:

$$L = L_{ce} + \lambda L_{bd}, \quad (7)$$

where λ is boundary loss weight. Specifically, the boundary loss function is defined as follows:

$$L_{bd} = \sum_{c \in C} \Bigg(\int_{\Omega_c} |Y_c(x,y) - t_{ci}|^2 H(\hat{Y}_c^*(x,y)) dxdy + \int_{\Omega_c} |Y_c(x,y) - t_{co}|^2 (1 - H(\hat{Y}_c^*(x,y))) dxdy \Bigg), \quad (8)$$

where Y_c and $\hat{Y}_c^*$ are ground-truth and predicted contour for class $c \in C$. $\hat{Y}_c^*$ is a shifted probability map derived from CNN prediction, that is, $\hat{Y}_c^* = \hat{Y}_c - 0.5 \in [-0.5, 0.5]$. t_{ci} and t_{co} are constant functions that represent average intensity inside and outside of the boundary in binary ground truth Y_c.

During the level set evolution process, the initial contour evolves iteratively based on the predicted parameters. After T evolution steps, the final level set function (i.e. segmentation result) is obtained. Then we calculate the weighted sum of the boundary loss and the cross-entropy loss, and CNN parameters can be updated by backpropagation.

4. EXPERIMENTS AND RESULTS

4.1. Datasets and Evaluation Metrics

We evaluate our method on the Prostate MR [10] images and 2D REFUGE [11] images. For prostate data, we use 3795 images and divide it into 70/10/20 for training, validation, and testing, respectively. For REFUGE data, we use 1200 public images from official annotation and divide it according to the official settings. In addition, we crop the REFUGE images to resolve category imbalance problem.

We utilize the following quantitative measures to evaluate our method: Dice similarity coefficient (DSC), relative volume difference (RVD), Hausdorff distance (HD), and average boundary distance (ABD). For 2D REFUGE images, RVD score is not taken into account.

4.2. Implementation Details

We use a modified ResNet-101 as the CNN encoder. The last fully-connected layer is removed and PSP module [3] is used as the prediction branch.

In the training phase, random noise is added to the ground-truth annotation to get extreme points used for training. We follow DEXTR [8] to place a 2D Gaussian around each of the extreme point to get a heatmap. Each training process executes 100 epochs. The training batch size is 10. Initial learning rate is 3e-4 and decayed by 0.3 every 10 epochs. The level set evolution phase uses $T = 5$ both in the training and testing.

4.3. Experimental Validation

To evaluate the effects of confidence map and boundary loss, we take DELSE [7] as baseline. An ablation experiment is performed based on the prostate MR images, which shown in Table 1. The CM aims to improve the performance on the RoI , while BD aims to improve the performance on the contour of the RoI, which is reflected in HD metric.

We compare our method with other state-of-the-art methods under identical and unbiased settings. The experimental results are shown in Table 2 and Table 3, which are reported as mean $\pm$ standard deviation. The proposed method achieves the best performance on both the Prostate and REFUGE datasets (average results of two classes).

Table 2. Compare with other methods on prostate dataset.

Model	DSC (%)	RVD (%)	HD (mm)	ABD (mm)
PSPNet [3]	75.5 ± 9.4	4.7 ± 20.5	24.6 ± 15.3	2.9 ± 0.7
FCN [1]	82.4 ± 5.6	6.1 ± 10.6	19.6 ± 19.8	2.4 ± 0.7
U-Net [2]	84.7 ± 6.5	$\mathbf{2.4} \pm 8.0$	15.9 ± 6.9	1.9 ± 0.4
V-Net [16]	85.3 ± 6.8	3.5 ± 8.8	16.8 ± 6.6	2.0 ± 0.5
DeepLabv3+ [4]	86.5 ± 5.1	-6.2 ± 7.1	23.1 ± 19.1	2.2 ± 0.4
GrabCut [12]	78.4 ± 15.6	12.2 ± 51.5	21.5 ± 11.3	2.9 ± 1.7
DEXTR [8]	95.8 ± 0.6	$-6.5 \pm \mathbf{1.5}$	7.6 ± 3.7	1.2 ± 0.3
DELSE [7]	96.0 ± 0.7	-5.8 ± 1.7	7.4 ± 3.6	$1.2 \pm \mathbf{0.2}$
Ours	$\mathbf{96.4} \pm \mathbf{0.5}$	-2.9 ± 1.9	$\mathbf{6.8} \pm \mathbf{3.1}$	$\mathbf{1.0} \pm \mathbf{0.2}$

Table 3. Compare with other methods on REFUGE dataset. CUP and DISC are two different categories.

		PSPNet [3]	FCN [1]	U-Net [2]	DeepLabv3+ [4]	DEXTR [8]	DELSE [7]	Ours
CUP	DSC (%)	63.2 ± 3.5	84.8 ± 1.3	86.0 ± 1.1	70.2 ± 2.2	95.7 ± 0.2	95.6 ± 0.2	$\mathbf{96.0} \pm 0.2$
	HD (mm)	12.2 ± 1.0	4.0 ± 0.5	3.3 ± 0.6	5.7 ± 1.1	1.4 ± 0.2	1.4 ± 0.2	$\mathbf{1.2} \pm 0.2$
	ABD (mm)	1.6 ± 0.2	0.7 ± 0.05	0.6 ± 0.04	1.0 ± 0.1	$\mathbf{0.3} \pm 0.01$	$\mathbf{0.3} \pm 0.01$	$\mathbf{0.3} \pm 0.01$
DISC	DSC (%)	88.0 ± 0.6	97.1 ± 0.2	$\mathbf{97.7} \pm 0.1$	82.0 ± 0.3	97.5 ± 0.1	97.4 ± 0.1	97.6 ± 0.1
	HD (mm)	7.4 ± 0.4	2.5 ± 0.3	$\mathbf{2.0} \pm 0.5$	9.6 ± 0.6	2.1 ± 0.4	$\mathbf{2.0} \pm 0.4$	$\mathbf{2.0} \pm 0.4$
	ABD (mm)	1.0 ± 0.1	$\mathbf{0.3} \pm 0.01$	$\mathbf{0.3} \pm 0.01$	3.8 ± 0.4	$\mathbf{0.3} \pm 0.01$	$\mathbf{0.3} \pm 0.01$	$\mathbf{0.3} \pm 0.01$
mean	DSC (%)	75.6	91.0	91.9	76.1	96.6	96.5	**96.8**
	HD (mm)	9.8	3.2	2.7	7.6	1.7	1.7	**1.6**
	ABD (mm)	1.3	0.5	0.4	2.4	**0.3**	**0.3**	**0.3**

We evaluate the accuracy of the predicted boundary curve in terms of F-measure. We can calculate the recall and precision of each prediction boundary and get the F-measure score. The experimental results are shown in Fig. 3.

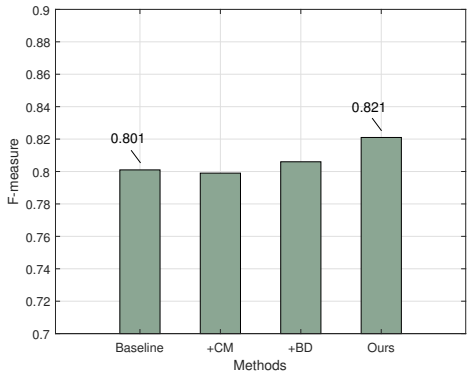

Fig. 3. F-measure on prostate MR dataset.

Visualization of segmentation results on the prostate MR images is shown in Fig. 4. The flaws are fixed by adding confidence map and boundary loss, which can improve the closeness between the level set evolution result curve and the ground-truth.

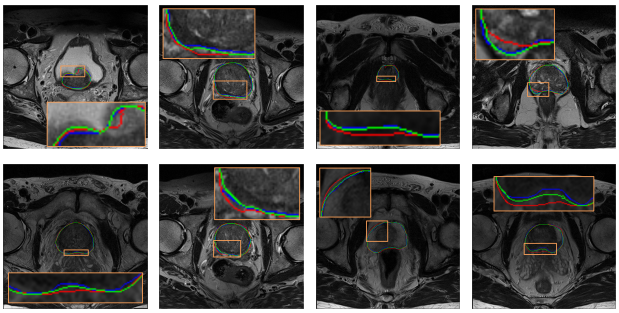

Fig. 4. Result visualization. The blue, red, and green lines represent the ground-truth, the segmentation results of baseline, and our method, respectively.

5. CONCLUSION

This paper presents a medical image segmentation method that combines confidence map and boundary loss with the deep level set model. We use confidence map and boundary loss to make the segmentation results have more accurate shape and boundary details. Compared with the state-of-the-art method, our method improves the DSC score 0.4% and 0.2% on the prostate images and retinal fundus images, respectively. The results visualization on prostate images show that our method can further refine the boundary details. Our

work evaluates the comprehensive effect of the level set and deep learning architecture. In some cases, the intersection point o may deviate from the RoI area. We plan to further study a feasible solution.

Acknowledgement

This work was supported in part by the National Natural Science Foundation of China under grant No. 61876148. This work was also supported in part by the Fundamental Research Funds for the Central Universities No. XJJ2018254, and China Postdoctoral Science Foundation No. 2018M631164.

6. REFERENCES

[1] 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*, 2015, pp. 3431–3440.

[2] 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*. Springer, 2015, pp. 234–241.

[3] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.

[4] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.

[5] Min Tang, Sepehr Valipour, Zichen Zhang, Dana Cobzas, and Martin Jagersand, "A deep level set method for image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 126–134. Springer, 2017.

[6] Diego Marcos, Devis Tuia, Benjamin Kellenberger, Lisa Zhang, Min Bai, Renjie Liao, and Raquel Urtasun, "Learning deep structured active contours end-to-end," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8877–8885.

[7] Zian Wang, David Acuna, Huan Ling, Amlan Kar, and Sanja Fidler, "Object instance annotation with deep extreme level set evolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 7500–7508.

[8] Kevis-Kokitsi Maninis, Sergi Caelles, Jordi Pont-Tuset, and Luc Van Gool, "Deep extreme cut: From extreme points to object segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 616–625.

[9] Shadab Khan, Ahmed H Shahin, Javier Villafruela, Jianbing Shen, and Ling Shao, "Extreme points derived confidence map as a cue for class-agnostic segmentation using deep neural network," *arXiv preprint arXiv:1906.02421*, 2019.

[10] Zhiqiang Tian, LiZhi Liu, and Baowei Fei, "A supervoxel-based segmentation method for prostate mr images," in *Medical Imaging 2015: Image Processing*. International Society for Optics and Photonics, 2015, vol. 9413, p. 941318.

[11] José Ignacio Orlando, Huazhu Fu, João Barbossa Breda, Karel van Keer, Deepti R Bathula, Andrés Diaz-Pinto, Ruogu Fang, Pheng-Ann Heng, Jeyoung Kim, JoonHo Lee, et al., "Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs," *Medical image analysis*, vol. 59, pp. 101570, 2020.

[12] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," in *ACM transactions on graphics (TOG)*. ACM, 2004, vol. 23, pp. 309–314.

[13] Dim P Papadopoulos, Jasper RR Uijlings, Frank Keller, and Vittorio Ferrari, "Extreme clicking for efficient object annotation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 4930–4939.

[14] Stanley Osher and James A Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations," *Journal of computational physics*, vol. 79, no. 1, pp. 12–49, 1988.

[15] Youngeun Kim, Seunghyeon Kim, Taekyung Kim, and Changick Kim, "Cnn-based semantic segmentation using level set loss," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2019, pp. 1752–1760.

[16] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, 2016, pp. 565–571.