Tissue segmentation in volumetric laser endomicroscopy data using U-net and a domain-specific loss function

Joost van der Putten[1], **Fons van der Sommen**[1], **Maarten Struyvenberg**[2], **Jeroen de Groof**[2]
Wouter Curvers[3], **Erik Schoon**[3], **Jaques Bergman**[2], **Peter H.N. de With**[1]
[1]VCA Research Group, Eindhoven University of Technology, 5612 AP, Eindhoven, the Netherlands
[2]Academic medical center, 1105 AZ, Amsterdam, the Netherlands
[3]Catharina Hospital, 5623 EJ, Eindhoven, the Netherlands
j.a.v.d.putten@tue.nl

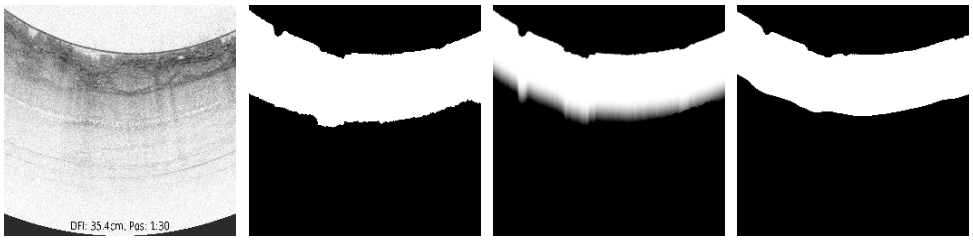
Abstract

Volumetric Laser Endomicroscopy (VLE) is a promising balloon based imaging technique for detecting early neoplasia in Barrett's Esophagus. Especially Computer Aided Detection (CAD) techniques show great promise compared to doctors, who cannot reliably find disease patterns in the VLE signal. However, the relevant tissue has to be segmented in order for these systems to function properly. At present, tissue segmentation has to be done manually and is therefore not scalable for full VLE scans of 1,200 $\times$ 4,096 $\times$ 2,048 pixels. Furthermore, the current CAD methods cannot use the VLE scans to their full potential as only a small section is selected while an automated system can delineate the entire image. This paper explores the possibility of automatically segmenting relevant tissue for VLE scans using a convolutional neural network. The contribution of this work is threefold. First, this is the first tissue segmentation algorithm for VLE scans. Second, we introduce a weighted ground truth that exploits the signal to noise ratio characteristics of the data. Third, we compare our algorithm segmentations against two additional VLE experts. The results show that our approach is on par with the experts and can therefor be used as a preprocessing step for further classification of the tissue.

1 Introduction

Esophageal adenocarcinoma (EA) is a form of cancer whose incidence is rising dramatically in the Western world. Patients with Barrett's Esophagus (BE) have an increased risk for developing dysplasia and eventually EA. Hence, it is crucial that dysplasia associated with BE is detected at an early stage. Currently, BE patients are monitored through periodic endoscopic surveillance and biopsies. However, the current methods have some drawbacks such as diagnostic uncertainty of white-light endoscopy, sampling error, and ambiguous samples in histology. Volumetric Laser Endomicroscopy (VLE) is a novel technique that has the potential to significantly contribute to the early detection of dysplsia. A recent investigation into CAD for VLE analysis showed promising results, with an Area Under Curve (AUC) between 90%-93% [1]. However, in order to use the proposed methods, the tissue under examination has to be extracted, which is currently done by manual cropping. This technique is limited, as not all valuable information is extracted from the images. Additionally, this process is not scalable, since full VLE scans consist of 1,200 slices of 4k $\times$ 2k pixels per patient.

1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands.



(a) (b) (c) (d)

Figure 1: (a) VLE image. (b) Basic ground truth. (c) Weighted ground truth. (d) Prediction output of the CNN, using the weighted ground truth to train the U-net.

In this research we propose the first tissue segmentation algorithm for VLE images, using U-net [2] with an adapted domain-specific loss function. The results are compared against expert delineations to show similar assessor intervariability.

2 Methods

In order to segment the VLE images, we implement an end-to-end learning strategy. First, the ground truth needs to be determined prior to using the images to train a CNN. The various experiments are then evaluated against both general and domain-specific metrics.

2.1 Ground truth labeling

Two different sets of ground truth data were used for the training and evaluation of the algorithm. The first ground truth was manually delineated by a CAD expert with more than three years experience in VLE research, which will be further referenced as the basic ground truth. An example of a VLE image and its ground truth are shown in Figure 1a and Figure 1b, respectively. The strength of the VLE signal weakens considerably over the depth, thereby decreasing the signal-to-noise ratio in the lower parts of the images. Additionally, research has indicated that incident cancer is most discriminative in the top layers of tissue [3]. Hence, the top layers are more important than the lower layers. For this reason, a weighted ground truth was implemented as well. The uppermost border is copied from the first basic ground truth masks and then an additional region below that border is added to the mask, reducing the mask weight for lower layers. An example is shown in Figure 1c.

2.2 Convolutional neural network (CNN)

A two-class CNN similar to the one described by Ronneberger *et al.* [2] was used to segment the tissue of the VLE scans. The original snapshots have a resolution of 1,342 $\times$ 1,024 pixels, which were resized to 256 $\times$ 256 pixels for computational efficiency. The final layer has a sigmoid activation function, so that the output represents the probability whether a pixel belongs to the segmented tissue or not. For the loss function, the Dice Similarity Coefficient (DSC) was used. Generally, the DSC is calculated using two binary masks. However, a threshold is necessary in order to binarize the masks, making the DSC non-differentiable, which is required by the back-propagation algorithm. These conflicting demands were solved by implementing the DSC with parameter y as the ground truth, $\hat{y}$ being the prediction and $y, \hat{y} \in [0, 1]$, as follows:

$$DSC = \frac{2(y \cdot \hat{y})}{(y + \hat{y})}.$$

2.3 Evaluation

The applied data set for this research consisted of 137 histopathologically matched VLE snapshots. Fourfold cross-validation was performed to calculate the metrics over the entire data set. Prior to calculating the evaluation metrics, the weighted ground truth and the predicted segmentation were first binarized with the Otsu threshold to facilitate fair comparisons. The predicted tissue segmentation

was compared to the ground truth by calculating the binary DSC and a custom metric, called balloon distance.

The balloon distance is calculated by first taking the difference between the binarized weighted ground truth and the prediction. Since a correct upper boundary is critical for further classification tasks, only the wrongly classified pixels in the top region are considered. We have defined the balloon distance as follows:

$$\text{Balloon distance} = 1 - min(N_z(\text{difference mask})/1000, 1),$$

where N_z denotes the sum of the non-zero elements in the difference mask. Since the balloon in the resized ground truth has an approximate height of 3 or 4 pixels and spans the entire width of 256 pixels, the image was rescaled with a value of 1000 to normalize the results. This yields in a metric where unity indicates that the top part of prediction aligns perfectly with the top part of the ground truth, and zero indicates a total mismatch between the ground truth and the prediction.

3 Experimental Results

The results of our experiments are shown in Table 1, where we employed 10,000 epochs of training for all experiments. At the left side of the table, both the DSC and the Balloon Distance (BD) were evaluated, using the predictions obtained from the model generated from the corresponding ground truth type. An example of a prediction made by the algorithm is shown in Figure 1d. For this example, the network was trained with the weighted ground truth DSC as a loss function. Two additional assessors with multiple years of VLE experience also annotated the original VLE scans. Table 1 (right) shows how well Assessor 2 and 3 score with respect to other assessors.

	Basic model		Weighted model		Assessor 2		Assessor 3	
	DSC	BD	DSC	BD	DSC	BD	DSC	BD
Assessor 1	0.95	0.88	0.97	0.88	0.96	0.83	0.96	0.83
Assessor 2	0.95	0.83	0.97	0.83	-	-	0.96	0.80
Assessor 3	0.95	0.85	0.97	0.85	-	-	-	-

Table 1: Left: DSC and balloon distance (BD) comparison of the different assessors and the weighted ground truth and basic ground truth predictions, obtained with their respective model. Right: different assessor annotations scored against each other.

4 Conclusions

Training a network with a weighted ground truth increases the DSC by approximately 2%, resulting in tissue segmentations that are within the intervariability between different assessors. However, using a weighted groundtruth has no significant effect on the ability of the model to precisely delineate the topmost layer (BD metric) of the relevant VLE tissue. Additionally, the different assessors score lower on both metrics in relation to each other, compared to the network predictions when trained with their respective ground truths. This is especially true for the BD metric. This indicates that the algorithm predictions are just as valid as expert tissue segmentations. In this work, we have proposed the first tissue segmentation algorithm for VLE scans, as well as a domain-specific loss function that exploits the signal-to-noise characteristics of VLE scans. Interestingly, our type of data exploitation and characteristics learning is of generic nature and can therefore also be implemented for other similar modalities, such as ultrasound and other optical coherence tomography applications.

References

[1] F. van der Sommen et al, "Predictive features for early cancer detection in Barrett's esophagus using volumetric laser endomicroscopy," Comput Med Imaging Graph, 2018, [in press].

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," pp. 1–8, 2015.

[3] A. F. Swager et al., "Computer-aided detection of early Barrett's neoplasia using volumetric laser endomicroscopy," Gastrointest. Endosc., vol. 86, no. 5, pp. 839–846, 2017.