Signal, Image and Video Processing
https://doi.org/10.1007/s11760-020-01772-7

ORIGINAL PAPER

Tympanic membrane segmentation in otoscopic images based on fully convolutional network with active contour loss

Van-Truong Pham[1,2] · Thi-Thao Tran[1,2] · Pa-Chun Wang[3,4] · Men-Tzung Lo[2]

Received: 14 January 2020 / Revised: 22 July 2020 / Accepted: 25 August 2020

Abstract
This paper presents a method to automatically segment tympanic membranes (TMs) from video-otoscopic images based on the deep learning approach. The paper introduces a hybrid loss function combining the Dice loss and active contour loss to the fully convolutional network. By this way, the proposed model takes into account the Dice similarity and the desired boundary contour information including the contour length as well as regions inside and outside the contour during learning. The proposed loss function is then applied to the fully convolutional network for tympanic membrane segmentation. We evaluate the proposed approach on TMs data set which includes 1139 otoscopic images from patients diagnosed with and without otitis media. Experimental results show that the proposed deep learning model achieves an average Dice similarity coefficient of 0.895, a mean Hausdorff distance of 19.189, and average perpendicular distance of 6.429, that outperforms other state-of-the-art methods.

Keywords Image segmentation · Active contour model · Deep learning · Otitis media · Tympanic membranes segmentation

1 Introduction

Otitis media is a common ear infection among pediatric population, and might lead to severe life-threatening otological or intracranial complications [1]. Otitis media (OM) is defined as any inflammation of the middle ear and can be classified clinically as acute otitis media (AOM), otitis media with effusion (OME) and chronic otitis media (COM). AOM is caused by active bacterial infection leading to congestion or purulence in the middle ear. OME is a consequence of AOM characterized by accumulation of fluid in the middle ear space after acute stage. COM is the chronic stage of AOM that involves perforation, retraction, or collapse of tympanic membrane following active bacterial infection within the middle ear space [2]. OM can incur significant economic impacts to the society, especially in countries with a vast territory where access to appropriate medical care is not readily available [3]. In fact, OM diseases are related to abnormality of the tympanic membrane (TM) [4, 5]. Establishment of OM diagnosis relies on the identification of morphological or color changes on TM.

The tympanic membrane, also known as the eardrum, is a thin membrane dividing the ear canal from the middle ear. The abnormality of TM, if not early detected and treated, has various consequences for patients including hearing loss and major infection [5]. Therefore, it is necessary to detect the TM abnormality for early OM diagnosis, particularly in children. In TMs image analyzing, the areas of TMs determined from acquired image frames via segmentation procedure of TM play an important role for further steps in pediatric otitis media diagnostic process [6]. From the segmented area of TM, key parameters for OM diagnosis such as TM size, texture, geometry, color distribution could be derived [7].

In fact, TM segmentation is a nontrivial task since TM images normally are taken from video-otoscopic images [5] which present irregular illumination, i.e., leaving some image regions brighter or darker than the average color of a given

✉ Thi-Thao Tran
thao.tranthi@hust.edu.vn

✉ Men-Tzung Lo
mzlo@ncu.edu.tw

1 School of Electrical Engineering, Hanoi University of Science and Technology, Hanoi, Vietnam

2 Department of Biomedical Sciences and Engineering, National Central University, Taoyüan, Taiwan

3 Department of Otolaryngology, Cathay General Hospital, Taipei, Taiwan

4 School of Medicine, Fu Jen Catholic University, Taipei, Taiwan

Published online: 24 September 2020

structure [6]. These characteristics along with the low contrast of anatomical structure boundaries suffer difficulties for TMs automatic segmentation tasks. In addition, due to effects during acquisition process, TM images are normally with intensity inhomogeneity. Moreover, there are hidden areas existent on the TM images taken from video-otoscopic images, which also make TM segmentation to be difficult task [6]. There are number of methods developed to segment TM in literature. Hsu et al. [8], and Ibekwe et al. [9] used computer mouse to select a set of points around interested areas in order to define the boundaries of interested objects before segmenting the TMs. However, it is difficult to exactly obtain the desired TM boundaries since one needs to imprecisely manipulate computer mouse around the interested regions. Therefore, it might lead to errors. In the work of Comunello et al. [6], TM boundaries are defined by semi-automatic tympanum procedure in which one manually adjusts the minor and major axes of an initial ellipse defined by the user. Xie et al. [10] employ a snake-a parametric active contour model (ACM) to segment TM images. However, since the snakes uses gradient information such as image edges to guide the curves, results are not adequate if images are with weak boundaries. Shie et al. [2] performs ACM on images extracted from video-otoscope to segment TMs. Though revealing advantages such as giving subpixel accuracy, ACM generally relies on initial contours. In addition, one might get unsatisfactory results if input images are with weak boundaries, inhomogeneous, or corrupted by noise.

On the other hand, in image segmentation field, deep learning (DL) methods have shown excellent performances [11–13], especially fully convolutional neural networks (FCNs) and variants [11, 14–16]. The DL approach has been applied for various segmentation problems such as skin lesion [12], tumor [17], cardiac MR image [18], myocardium [19], mitotic event [20], and histopathological image segmentation [21]. For training the neural network in image segmentation, Dice loss is commonly used [15, 22]. The Dice loss is used to minimize the mismatch or maximize the overlap between the prediction map (segmentation) and the ground truth (reference). Though being commonly used, the Dice loss lacks boundary information and suffers difficulties in handling highly unbalanced segmentations, i.e., size of object is much smaller than background size. Recently, motivated by ACM, Chen et al. [23] proposed the active contour (AC) loss that considers the area inside as well as outside the interested region and the boundary information. However, while training from scratch, AC loss often suffers from poor convergence if the current segmentation is far from the reference.

Inspired by the DL and ACM, in this work, we proposed an approach for automatic segmentation of TMs. The approach is based on the FCN structure proposed by Tran in [15] that has fewer parameters compared to other FCN-based models. To train the network, we proposed a loss function combining AC loss with Dice loss. By this way, we can take advantages of both AC and Dice losses and mitigate their shortcomings. For AC loss, we can consider the boundary information and the region inside as well as outside the interested region. It also preserves the importance of the region overlap between segmentation and reference by using the Dice loss. The network while trained by the proposed loss can obtain better performances than using AC loss or Dice loss solely.

To the best of our knowledge, this is the first study that employs the fully convolutional network architecture to segment TM from video-otoscopic images. The contributions of this work are listed as: (i) propose a new loss function combining region fitting inside and outside the segmentation, contour length, and region overlap for training the neural network. (ii) propose a fully automatic algorithm for TM segmentation. (iii) evaluate the TM segmentation performances by neural network algorithms on a data set of otoscopic images.

The remainder of this paper is organized as follows. Section 2 details database characteristic and the work environment. Section 3 briefly reviews the AC model. The FCN architecture integrated with AC for TM segmentation is presented in Sect. 4. The implementation and experimental details are given in Sect. 5. The discussions are drawn in Sect. 6. We end this paper with a brief conclusion in Sect. 7.

2 Materials

2.1 Database description

Based on the Institutional Review Board approval from Cathay General Hospital (No. CGH-P103040), a database of 1139 OM otoscopic images from children aged 6 months to 12-year-old, captured through digital otoscope (Karl Storz, Tullingen, Germany) by otologists, were retrospectively reviewed. Among them, 630 images are diagnosed with pediatric OM and 509 images are diagnosed normal. The pediatric OM(s) are characterized by the hyperemic change, bulging, or perforation of tympanic membrane, which represent early stage, suppurative stage, and spontaneous perforation of AOM respectively; and also by the presence of purulence or effusion in the tympanic cavity, representing suppurative stage or subacute stage of AOM or OME. The images were all optimal without obstruction of cerumen, and tympanic membrane can be visualized.

2.2 Work environment

Experiments were performed using a fully convolutional network architecture [11] based on the Keras library. Some results from Python package are exported to MATLAB workspace for visualization. The computer used was equipped with an Intel (R) Xeon (R) processor E5-1620 at 3.5 GHz with 64 GB RAM. The graphics processing unit used was a Geforce RTX 2080 model with 8 GB of RAM.

3 Related works

3.1 Region-based active contour models

In the last two decades, various models for image segmentation have been proposed such as, graph cuts [24], fuzzy clustering [25], and active contours models (ACMs) [26–28]. Among them, ACMs have been shown to be a promising approach [29–32].

Let Ω be a bounded open subset of $\Re^2$, let $I : \Omega \rightarrow \Re$ be a given gray level image, and let a curve $C \subset \Re$ divide the image plane into two regions, inside(C) and outside(C). Chan and Vese [29] proposed the energy functional defined as:

$$E(C, c_1, c_2) = \int_{\text{inside}(C)} (I(\mathbf{x}) - c_1)^2 \mathrm{d}\mathbf{x} + \int_{\text{outside}(C)} (I(\mathbf{x}) - c_2)^2 \mathrm{d}\mathbf{x} + \mu \int_0^{\text{Length}(C)} \mathrm{d}s \tag{1}$$

where $\mu > 0$ is an arbitrary parameter, ds is the Euclidean element of the length of curve C; and c_1, c_2 represents the intensity means of two regions inside(C) and outside(C).

Though CV model performs good segmentation results, it slowly converges. It also might get trapped into local minima. To address this, Bresson et al. [33] proposed to use total variation energy, *TV,* to re-define the energy functional of CV model as

$$E(u, c_1, c_2) = \mu TV_g(u) + \lambda \int_\Omega r(\mathbf{x}, c_1, c_2) u \mathrm{d}\mathbf{x} \tag{2}$$

where $r(\mathbf{x}, c_1, c_2) = (I(\mathbf{x}) - c_1)^2 - (I(\mathbf{x}) - c_2)^2$; u is a characteristic function 1_Ω; $TV_g(u) = \int_0^{\text{Length}(C)} g|\nabla u| \mathrm{d}s$ is the total variation energy; λ and μ are weighting parameters.

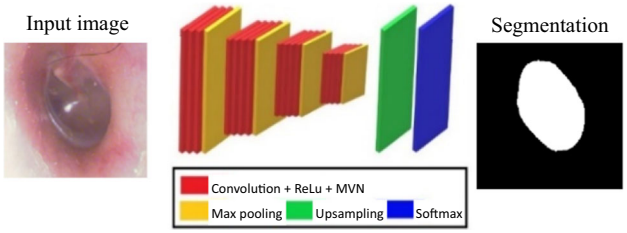

Fig. 1 Basic structure of the fully convolutional network for TM image segmentation

Equation 2 can be rewritten as:

$$E(u, c_1, c_2) = \mu \int_0^{\text{Length}(C)} g|\nabla u| \mathrm{d}s + \lambda \int_\Omega \left((I(\mathbf{x}) - c_1)^2 - (I(\mathbf{x}) - c_2)^2 \right) u \mathrm{d}\mathbf{x} \tag{3}$$

Note that, u is a characteristic function valued between 0 and 1; g is the weight function, normally chosen as an edge indicator function during implementation process.

3.2 Fully convolutional network (FCN)

In this sub-section, we detail the pipeline of the FCN architecture [15] as the segmentation framework for TMs. The basic structure of the network is presented in Fig. 1. It includes 15 convolution layers, 3 max pooling layers, upsampling layers and a softmax layer.

The network is divided into two main parts, contracting path and expanding path. The contracting path consists of 3×3 convolution layers with zero padding to preserve the spatial structure of the feature map and 3×3 max pooling layers with stride 2. Each convolution layer (Conv) is followed by a rectified linear unit (ReLU) and a mean–variance normalization (MVN). The expanding path consists of 3×3 convolution-transpose layers with stride 2, which are used to reconstruct the spatial structure of image.

4 Proposed method

4.1 Loss function

The loss function for the network is proposed as:

$$\text{Loss} = L_{\text{AC}} + \gamma_{\text{Dice}} L_{\text{Dice}} \tag{4}$$

where L_{AC} is the active contour loss, L_{Dice} is the Dice loss, and γ_{Dice} is the weighting parameter for the Dice loss term.

Inspired by the active contour approach in [23, 33], we proposed the AC loss as:

$$L_{\text{AC}} = \frac{E_{\text{AC}}}{W \times H} \tag{5}$$

where W and H are respectively the width and height of the input image; E_{AC} is the AC energy, defined as:

$$E_{\text{AC}} = \mu\text{Length} + \lambda\text{Region} \tag{6}$$

in which

$$\text{Length} = \int_C |\nabla u| \mathrm{d}s \tag{7}$$

with C is the curve dividing the image plane into inside(C) and outside(C); and u [0,1] is the segmentation.

$$\text{Region} = \int_\Omega \Big((v - c_1)^2 - (v - c_2)^2\Big) u \mathrm{d}\mathbf{x} \tag{8}$$

where Ω is the image domain, v [0,1] is the reference; c_1 and c_2 are respectively the intensity means of two image regions, inside(C) and outside(C).

In pixel-wised way, Eqs. 7 and 8 can be written as:

$$\text{Length} = \sum_{i=1}^{W}\sum_{j=1}^{H} \sqrt{\left|(\nabla u_{x_{i,j}})^2 + (\nabla u_{y_{i,j}})^2\right| + \varepsilon} \tag{9}$$

where x and y are horizontal and vertical directions, respectively. $u_{x_{i,j}}$ and $u_{y_{i,j}}$ are indexes of pixel locations in those directions of the segmentation.$\varepsilon > 0$ is a parameter to avoid zero for the square root operation.

$$\begin{aligned}\text{Region} = &\left|\sum_{i=1}^{W}\sum_{j=1}^{H}\Big(u_{i,j}(v_{i,j} - c_1)^2\Big)\right| \\ &+ \left|\sum_{i=1}^{W}\sum_{j=1}^{H}\Big((1 - u_{i,j})(v_{i,j} - c_2)^2\Big)\right|\end{aligned} \tag{10}$$

where $u_{i,j}$, $v_{i,j}$ are indexes of pixel locations in the segmentation u, and the reference v. Following the ACM presented in [33], c_1 and c_2 are computed as:

$$c_1 = \frac{\sum_{i=1}^{W}\sum_{j=1}^{H} v_{i,j}u_{i,j}}{\sum_{i=1}^{W}\sum_{j=1}^{H} u_{i,j}}, \quad c_2 = \frac{\sum_{i=1}^{W}\sum_{j=1}^{H} v_{i,j}(1 - u_{i,j})}{\sum_{i=1}^{W}\sum_{j=1}^{H} (1 - u_{i,j})} \tag{11}$$

The Dice loss is defined as:

$$L_{\text{Dice}} = 1 - \text{Dice}(u, v) \tag{12}$$

where

$$\text{Dice}(u, v) = \frac{2\Big(\sum_{i=1}^{W}\sum_{j=1}^{H}(u_{i,j}v_{i,j}) + \varepsilon_s\Big)}{\sum_{i=1}^{W}\sum_{j=1}^{H}(u_{i,j}) + \sum_{i=1}^{W}\sum_{j=1}^{H}(v_{i,j}) + \varepsilon_s} \tag{13}$$

with ε_s, often chosen as 1, is the smooth factor used for numerical stability.

The proposed loss function can be rewritten as:

$$\text{Loss}_{\text{AC}-\text{Dice}} = \mu L_{\text{Length}} + \lambda L_{\text{Region}} + \gamma_{\text{Dice}} L_{\text{Dice}} \tag{14}$$

with

$$L_{\text{Length}} = \frac{\text{Length}}{W \times H}, \quad L_{\text{Region}} = \frac{\text{Region}}{W \times H} \tag{15}$$

It is noted that Length and Region in Eqs. 9 and 10 computed over the whole image domain are generally large, even thousand times larger than L_{Dice}, especially at the beginning of training. Therefore, tuning the parameters for loss terms might be complicated and relying on image size. To tackle this problem, we divided E_{AC} by the total pixels, $W \times H$, as in Eqs. 5, and 15. By this way, L_{Length} and L_{Region}, are not far from 1, the maximum value of L_{Dice}. Hence, it is simpler for tuning weighting parameters when combining AC and Dice losses.

4.2 Data augmentation and training

We augment the data for training process by performing some affine transformations techniques like rotation, vertical and horizontal flipping.

The network was trained on a GPU-based machine using Keras framework based on Tensorflow. The optimal algorithm used is Stochastic Gradient Descent (SGD) with Momentum with the weight of moment equal to 0.9. For the loss, the parameter for length term, μ, is chosen as 1, as in Bresson et al. [33]. For the region, as investigated by Chen et al. [23], we get satisfied results with$\lambda > 2$. So, we set as 5 to emphasize the importance of the region compared to the length term. While keeping the two parameters fixed as $\mu = 1$, $\lambda = 5$, we tuned γ_{Dice}. We tried γ_{Dice} with increasing values from 0.1 to 2 and saw that the results tend to be worse if $\gamma_{\text{Dice}} > 1$ so we select $\gamma_{\text{Dice}} = 0.5$. Finally, the parameters are set as $\gamma_{\text{Dice}} = 0.5$,$\lambda = 5$, and$\mu = 1$ for all experiments. The network is trained with the maximum number of epochs is chosen as 170. Using the SGD scheme, during training process, the loss including Dice and AC loss terms decreases until convergence. The Dice loss term is minimized when the similarity between segmentation and reference is maximized. For AC loss term, the contour length is minimized when the contour best preserves its regularity

and smoothness. The region fittings in AC loss are minimized when the overlap between object of segmentation and background of reference is smallest, and concurrently the overlap between background of segmentation and object of reference is minimized. The weight of best performance measured on validation set is chosen to predict the segmentation maps of test set.

5 Experiments

5.1 Performance metrics

To evaluate the performances of the automatic segmentation, we compare its accuracy with the ground truth (manual segmentation). Particularly, we used Dice similarity coefficient (DSC), the Intersection over Union (IoU), the Hausdorff distance (HD), and average perpendicular distance (APD) metrics. The DSC measures the similarity between automatic and manual segmentations, computed as follows

$$\mathrm{DSC} = \frac{2\mathbf{S}_{\mathrm{am}}}{\mathbf{S}_{\mathrm{a}} + \mathbf{S}_{\mathrm{m}}} \tag{16}$$

where $\mathbf{S}_{\mathrm{a}}$, $\mathbf{S}_{\mathrm{m}}$, and $\mathbf{S}_{\mathrm{am}}$ are, respectively, the automatically delineated region, the manually segmented region, and the intersection between the two regions.

The Intersection over Union (IoU) is used to measure the similarity between two sets, defined as:

$$\mathrm{IoU} = \frac{\mathbf{S}_{\mathrm{am}}}{\mathbf{S}_{\mathrm{a}} + \mathbf{S}_{\mathrm{m}} - \mathbf{S}_{\mathrm{am}}} \tag{17}$$

The Hausdorff distance is used to calculate the errors between the automatically extracted boundary, A, and the manually segmented boundary, B, defined [34], as

$$HD(A, B) = \max\{h(A, B), h(B, A)\} \tag{18}$$

where $h(A, B) = \max_{a \in A} \min_{b \in B}\{\mathrm{dist}(a, b)\}$, and dist(a, b) is the Euclidean distance between points a and b.

The APD measures the distance from automatically segmented contour to the corresponding contour manually drawn by experts, averaged over all contour points. The smaller the APD, the better the matching of automatically segmented contour and manual contour.

5.2 Experimental results

We applied the proposed model to the dataset including 1139 TMs images, including training (800), validation (112), and test (227) sets. All images were resized to $192 \times 192 \times 3$ pixels. During training process, segmentations of validation set

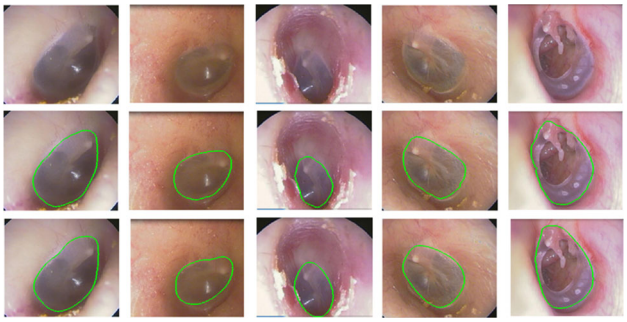

Fig.2 Representative results by the FCN with the proposed loss function. First row: input images; second row: result by proposed approach; last row: ground truth

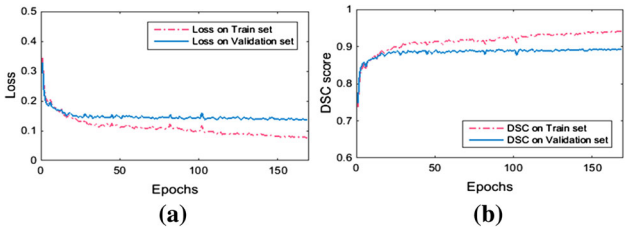

Fig. 3 The Loss versus epochs (**a**), and the DSC score versus epochs (**b**) of the proposed approach

are predicted and evaluated with ground truths. Since a segmentation with a high DSC is analogous to a high IoU, and often low HD as well as APD values, we pick DSC as the metric to select the weight, as in typical CNN-based image segmentation methods. The weight of best DSC for the validation data is chosen and used for predicting maps in the test set. For evaluation, results on test set with chosen weight are assessed using DSC, IoU, HD and APD metrics.

5.2.1 Performance of the proposed approach

Some representative samples of the obtained results for the TM test data are given in Fig. 2. In this figure, the original TM images, the segmentation results by the proposed method presented in the first and second rows of the figure. The ground truths are also shown in the last rows. As can be seen from this figure, there is a good agreement between the results obtained by the proposed approach and the ground truths.

We also plot the loss function versus epochs of the proposed approach in Fig. 3a. In addition, the evolutions of DSC values versus epochs are also depicted in Fig. 3b. It can be seen from these figures, the proposed approach achieves smooth loss function curves and stable DSC score values during training process.

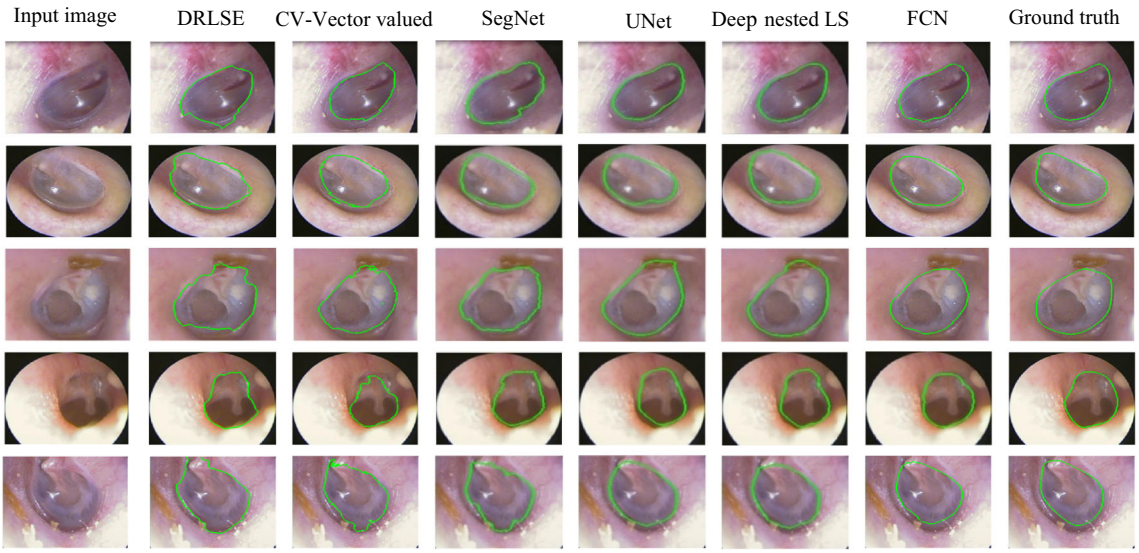

Fig. 4 Comparison between DRLSE, CV-vector valued, SegNet, UNet, deep nested LS, and FCN

Table 1 Result comparison in mean (STD) of the DSC, IoU, HD (in mm), and APD (in mm) for the DRLSE, CV-vector valued, SegNet, UNet, deep nested LS, and FCN

Methods	DSC	IoU	HD	APD
DRLSE [26]	0.868 (0.069)	0.773 (0.099)	23.223 (12.818)	7.824 (4.045)
CV-vector valued [35]	0.870 (0.066)	0.776 (0.098)	22.978 (11.914)	7.872 (4.128)
SegNet [16]	0.879 (0.069)	0.791 (0.102)	22.120 (11.417)	7.870 (4.164)
UNet [14]	0.884 (0.067)	0.799 (0.099)	19.572 (10.668)	6.892 (3.849)
Deep nested LS [18]	0.891 (0.062)	0.809 (0.094)	19.192 (10.209)	6.610 (3.523)
FCN [15]	0.895 (0.062)	0.816 (0.095)	19.189 (11.193)	6.429 (3.862)

5.2.2 Comparison with other approaches

To evaluate the performances of the proposed model, we compared its results with other state of the arts. To this end, we implemented two level set (LS)-based models including the edge-based distance regularized level set method (DRLSE) [26], and Chan Vese's (CV) vector valued [35]. We also reproduced the two architectures, SegNet [16], and UNet [14]. Furthermore, we adapted the deep nested LS by Duan et al. [18] for the TM segmentation problem. We also used the proposed loss function for training three above networks with the same training protocol as the FCN. The segmentation results for test images in the TMs database by above methods are compared with ground truths. Some representative results are shown in Fig. 4. In which, the input images are shown in the first column. The results by DRLSE [26], CV-vector valued [35], SegNet, UNet, Deep nested LS, FCN, and the ground truths are respectively presented in the succeeding columns. As can be seen from second and third columns, results by DRLSE and CV-vector valued methods are not satisfactory since segmented regions are not close to the ground truths. For the remaining approaches, we can obtain satisfied results. Nevertheless, the results by the proposed model are in better agreement with ground truths, especially for the second and fifth images as in Fig. 4. It is also noted that, the DRLSE and the CV-vector valued methods need initial curves nearby the desired boundaries for such segmentations, whereas remaining approaches do not require any curve initialization. To quantitatively assess the segmentations, we report the mean and standard deviation (STD) of the obtained DSC, IoU, HD, and APD of each method when segmenting all test images in the TM database in Table 1.

In addition, the boxplots of the four metrics from segmentation results by the above models are also plotted in Fig. 5. From Fig. 5 and Table 1, we could see that the FCN with the proposed loss obtains more accurate results than those by LS-based methods. The FCN also achieves better results than other networks when using the same loss function and training protocol with highest DSC and IoU, and lowest HD and APD values as in Table 1.

5.2.3 Performance of the proposed loss function

To evaluate the performance of the proposed loss, we trained SegNet, UNet, and FCN with other loss functions and compared the results. The loss functions include Dice, AC, binary cross entropy (BCE), and the combinations of Dice with AC, AC with BCE, and Dice with BCE. The quantitative results

Table 2 Results in mean (STD) of the DSC, IoU, HD, and APD for the SegNet, UNet, and FCN for difference losses

Method	Parameters (M)	Loss	DSC	IoU	HD (mm)	APD (mm)
SegNet [16]	29	Dice	0.881 (0.069)	0.794 (0.102)	22.758 (12.709)	8.002 (4.376)
		AC	0.883 (0.070)	0.797 (0.103)	22.622 (12.791)	7.767 (4.372)
		BCE	0.874 (0.080)	0.784 (0.116)	23.875 (13.901)	8.303 (4.716)
		Dice + AC	**0.886** (0.066)	**0.801** (0.099)	**21.913** (11.677)	**7.623** (4.213)
		AC + BCE	0.874 (0.081)	0.784 (0.116)	23.876 (13.901)	8.303 (4.716)
		Dice + BCE	0.868 (0.081)	0.776 (0.117)	24.604 (14.082)	8.915 (4.642)
UNet [14]	31	Dice	0.879 (0.069)	0.790 (0.101)	20.991 (9.931)	7.568 (3.988)
		AC	0.880 (0.072)	0.793 (0.105)	20.407 (10.943)	7.336 (4.022)
		BCE	0.877 (0.073)	0.788 (0.107)	20.438 (11.559)	7.431 (4.053)
		Dice + AC	**0.884** (0.067)	**0.799** (0.099)	**19.572** (10.668)	**6.892** (3.849)
		AC + BCE	0.881 (0.072)	0.794 (0.106)	20.205 (11.871)	7.073 (4.025)
		Dice + BCE	0.880 (0.073)	0.792 (0.106)	21.247 (13.519)	7.305 (4.169)
FCN [15]	11	Dice	0.892 (0.068)	0.812 (0.102)	19.951 (12.042)	6.738 (4.316)
		AC	0.892 (0.067)	0.812 (0.100)	**18.892** (11.021)	6.463 (3.876)
		BCE	0.889 (0.068)	0.807 (0.101)	19.952 (12.381)	6.835 (4.073)
		Dice + AC	**0.895** (0.062)	**0.816** (0.095)	19.189 (11.193)	**6.429** (3.862)
		AC + BCE	0.891 (0.066)	0.809 (0.099)	19.411 (11.236)	6.710 (3.933)
		Dice + BCE	0.889 (0.068)	0.807 (0.101)	20.012 (12.115)	6.866 (3.966)

Note that the FCN in the current study is the structure proposed by Tran [15]. Bold values refer to the best values obtained for the corresponding method

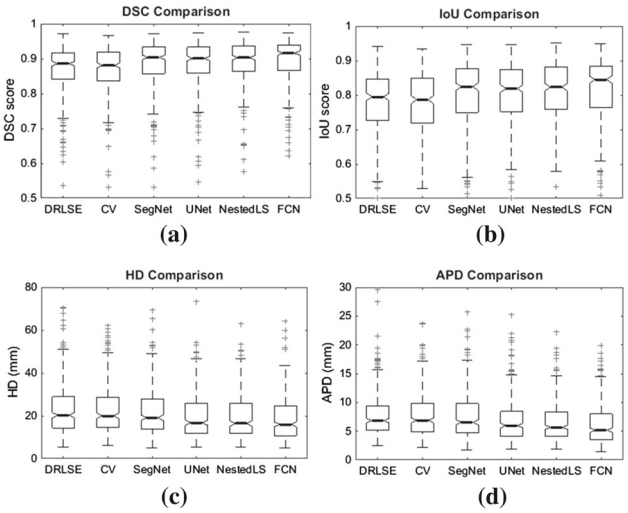

Fig. 5 Boxplots of the **a** DSC, **b**, IoU, **c** HD, **d** APD for methods (left to right in each figure): DRLSE, CV-vector valued, SegNet, UNet, deep nested LS, and FCN

of TM test set for the three networks employing different loss functions are shown in Table 2. As shown in rows 5, 11, and 17 of Table 2, for each network, the combination of the Dice and AC losses give the best performances with almost highest DSC and IoU scores. It also achieves lowest HD and APD values.

For the case of AC loss, we observed that if we train the network from scratch, the performances are poorer than those by using a pretrained weight initialized by using another loss. So, for performances in rows 3, 9, and 15 of Table 2, we have performed a weight initialization step before using AC loss alone. In more details, we trained the network using the commonly used Dice loss for first several epochs, then used the AC loss for succeeding epochs. This can be considered as a 'contour initialization' step for the model using AC loss that takes into account the regions inside and outside the contour of the predicted map. From experiments, we found that it is more efficient if we use the AC loss in combination with other losses, such as Dice or BCE, in order to eliminate the initialization step for obtaining a pretrained weight as in the case of using AC loss solely. Results show that the combination between the AC and Dice losses as proposed achieves the best performances for TM segmentation.

6 Discussion

Otitis media is a common infectious disease and a major cause of hearing impairment. It is shown in many reports that children may have at least one episode of AOM by 3 years of life, and some of them might have multiple recurrences in the years to come [36]. It is also proven that the subsequent OME after AOM, normally affects children between 3 and 7 years old, is a leading cause of preventable conductive hearing loss [7]. Though the prevalence of OM has declined over time globally thank to antibiotics and universal vacci-

nation coverage, OM is still a major public health threat in many countries.

Recently, artificial intelligences have been applied to computer-aided diagnosis in various medical imaging problems. With the advances of Wi-Fi-connected mobile devices and digital otoscope technologies, eardrum images can be captured, uploaded to the server and then analyzed through cloud computation. The analyzed information might assist the parents making decision as whether to bring children for professional helps. Diagnosis of OM relies on identification of the changes over eardrum, meaning: the hyperemic change, bulging, or perforation of TM, which represent early stage, suppurative stage, and spontaneous perforation of AOM respectively; and also by the presence of purulence or effusion in the tympanic cavity, representing suppurative stage or subacute stage of AOM or OME. The target area on the eardrum should be well delineated through the manual segmentation steps or semiautomatic segmentation such as the works in [6, 8, 9].

Considering the DL approach, although UNet and SegNet are widely used, in the current study, we showed that for TM segmentation, the FCN structure by Tran [15] achieves better performance than SegNet and UNet for all common metrics including DSC, IoU, HD, and APD. The advantages of the structure are also shown by having fewer parameters and less training time. The total parameters of the FCN are 11 million, while SegNet has 29 million, and UNet has 31 million parameters. For training from scratch with 170 epochs using an 8 GB NVDIA®GTX 2080 GPU, the FCN takes less than 2 h, while SegNet takes 3.5 h, and UNet takes 2.6 h. At test time on the GPU device, the FCN takes 6.5 ms to segment an image, while SegNet takes 15 ms, and UNet takes 8 ms for the same task.

7 Conclusions

We have presented a deep learning-based approach for fully automated segmentation of tympanic membrane from otoscopic images. The FCN network architecture is trained by the proposed loss function which is formed by combining active contour loss with the commonly used Dice loss function. The approach is applied to a dataset including 1139 TMs images. Segmentation results are also compared with other state of the arts. Experiments show that the proposed approach achieves high performance for all common segmentation metrics including DSC, IoU, HD and APD. In the future work, the proposed approach can be integrated into a computer-aided diagnosis system to classify patients with normal or OM and assist doctors in OM diagnosis.

Acknowledgements This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.05-2018.302.

References

1. Jabarin, B., Pitaro, J., Lazarovitch, T., Gavriel, H., Muallem-Kalmovich, L., Eviatar, E., Marom, T.: Decrease in pneumococcal otitis media cultures with concomitant increased antibiotic susceptibility in the pneumococcal conjugate vaccines era. Otol. Neurotol. **38**(6), 853–859 (2017)
2. Shie, C., Chang, H., Fan, F., Chen, C., Fang, T., Wang, P.: A hybrid feature-based segmentation and classification system for the computer aided self-diagnosis of otitis media. In: Proceedings of Conference IEEE Engineering in Medicine and Biology Society, pp. 4655–4658 (2014)
3. Fang, T., Rafai, E., Wang, P., Bai, C., Jiang, P., Huang, S.N., Chen, Y., Chao, Y., Wang, C., Chang, C.: Pediatric otitis media in Fiji: survey findings 2015. Int. J. Ped. Otorhinolaryngol. **85**, 50–55 (2016)
4. Lieberthal, A., Carroll, A., Chonmaitree, T., Ganiats, T., Hoberman, A., Jackson, M., Joffe, M., Miller, D., Rosenfeld, R., Sevilla, X., Schwartz, R., Thomas, P., Tunkel, D.: The diagnosis and management of acute otitis media. Pediatrics **131**(3), e964–e999 (2013)
5. Jaisinghani, V., Hunter, L., Li, Y., Margolis, R.: Quantitative analysis of tympanic membrane disease using video-otoscopy. Laryngoscope **110**(10 Pt 1), 1726–1730 (2000)
6. Comunello, E., Wangenheim, A., Junior, V., Dornelles, C., Costa, S.: A computational method for the semi-automated quantitative analysis of tympanic membrane perforations and tympanosclerosis. Comput. Biol. Med. **39**(10), 889–895 (2009)
7. Tran, T., Fang, T., Pham, V., Lin, C., Wang, P., Lo, M.: Development of an automatic diagnostic algorithm for pediatric otitis media. Otol. Neurotol. **39**(8), 1060–1065 (2018)
8. Hsu, C., Chen, Y., Hwang, J., Liu, T.: A computer program to calculate the size of tympanic membrane perforations. Clin. Otolaryngol. Allied Sci. **29**(4), 340–342 (2004)
9. Ibekwe, T., Adeosun, A., Nwaorgu, O.: Quantitative analysis of tympanic membrane perforation: a simple and reliable method. J. Laryngol. Otol. **123**(1), e2 (2009)
10. Xie, X., Mirmehdi, M., Richard Maw, R., Amanda Hall, A.: Detecting abnormalities in tympanic membrane images. In: Proceedings of the 9th Medical Image Understanding and Analysis, pp. 19–22 (2005)
11. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431–3440 (2015)
12. Huang, L., Zhao, Y.G., Yang, T.J.: Skin lesion segmentation using object scale-oriented fully convolutional neural networks. SIViP **13**(3), 431–438 (2019)
13. Öztürk, S., Özkaya, U., Akdemir, B., Seyfi, L.: Convolution kernel size effect on convolutional neural network in histopathological image processing applications. In: International Symposium on Fundamentals of Electrical Engineering (ISFEE), Bucharest (2018)
14. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Proceedings of the International Conference on Medical Imaging and Computer-Assisted Intervention, pp. 234–241 (2015)
15. Tran, P.V.: A fully convolutional neural network for cardiac segmentation in short-axis MRI. https://arxiv.org/abs/1604.00494 (2016)
16. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: a deep convolutional encoder–decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(12), 2481–2495 (2017)

17. Li, L., Zhao, X., Lu, W., Tan, S.: Deep learning for variational multimodality tumor segmentation in PET/CT. Neurocomputing (2019). https://doi.org/10.1016/j.neucom.2018.1010.1099
18. Duan, J., Schlemper, J., Bai, W., Dawes, J.W., Bello, G.T., Doumou, G., De Marvao, A., O'Regan, D.P., Rueckert, D.: Deep nested level sets: fully automated segmentation of cardiac MR images in patients with pulmonary hypertension. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 595–603 (2018)
19. Avendi, M.R., Kheradvar, A., Jafarkhani, H.: A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. Med. Image Anal. **30**, 108–119 (2016)
20. Öztürk, S., Akdemir, B.: A convolutional neural network model for semantic segmentation of mitotic events in microscopy images. Neural Comput. Appl. **31**(8), 3719–3728 (2019)
21. Öztürk, Ş., Akdemir, B.: Cell-type based semantic segmentation of histopathological images using deep convolutional neural networks. Int. J. Imaging Syst. Technol. **29**(3), 234–246 (2019)
22. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 4th International Conference on 3D Vision (3DV), pp. 565–571 (2016)
23. Chen, X., Williams, B.M., Vallabhaneni, S.R., Czanner, G., Williams, R., Zheng, Y.: Learning active contour models for medical image segmentation. In: Proceeding of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11623–11640 (2019)
24. Boykov, Y., Lee, V.S., Rusinek, H., Bansal, R.: Segmentation of dynamic N–D data sets via graph cuts using Markov models. In: Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), pp. 1058–1066 (2001)
25. Rezaee, M., van der Zwet, P., Lelieveldt, B., van der Geest, R., Reiber, J.: A multiresolution image segmentation technique based on pyramidal segmentation and fuzzy clustering. IEEE Trans. Image Process. **9**(7), 1238–1248 (2000)
26. Li, C., Xu, C., Gui, C., Fox, M.D.: Distance regularized level set evolution and its application to image segmentation. IEEE Trans. Image Process. **19**(12), 3243–3254 (2010)
27. Tran, T.T., Pham, V.T., Shyu, K.K.: Zernike moment and local distribution fitting fuzzy energy-based active contours for image segmentation. SIViP **8**(1), 11–25 (2014)
28. Duan, J., Pan, Z., Yin, X., Wei, W., Wang, G.: Some fast projection methods based on Chan–Vese model for image segmentation. EURASIP J. Image Video Process. **7**, 7 (2014)
29. Chan, T., Vese, L.: Active contours without edges. IEEE Trans. Image Process. **10**(2), 266–277 (2001)
30. He, L., Peng, Z., Everding, B., Wang, X., Han, C., Weiss, K., Wee, W.G.: A comparative study of deformable contour methods on medical image segmentation. Image Vis. Comput. **26**(2), 141–163 (2008)
31. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: active contour models. Int. J. Comput. Vis. **1**(4), 321–331 (1988)
32. Sethian, J.A.: Level Set Methods and Fast Marching Methods. Cambridge University Press, Cambridge (1999)
33. Bresson, X., Esedoḡlu, S., Vandergheynst, P., Thiran, J., Osher, S.: Fast global minimization of the active contour/snake model. J. Math. Imaging Vis. **28**(2), 151–167 (2007)
34. Tohka, J.: Surface extraction from volumetric images using deformable meshes: a comparative study. In: Proceedings of the 7th European Conference in Computer Vision, pp. 350–364 (2002)
35. Chan, T., Sandberg, Y., Vese, L.: Active contours without edges for vector-valued images. J. Vis. Commun. Image Represent. **11**(2), 130–141 (2000)
36. Wang, P., Chang, Y., Chuang, L., Su, H., Li, C.: Incidence and recurrence of acute otitis media in Taiwan's pediatric population. Clinics **66**(3), 395–399 (2011)