Ultrasound Image Segmentation: A Deeply Supervised Network With Attention to Boundaries

Deepak Mishra, Santanu Chaudhury, Mukul Sarkar, and Arvinder Singh Soin

Abstract*—*Objective: **Segmentation of anatomical structures in ultrasound images requires vast radiological knowledge and experience. Moreover, the manual segmentation often results in subjective variations, therefore, an automatic segmentation is desirable. We aim to develop a fully convolutional neural network (FCNN) with attentional deep supervision for the automatic and accurate segmentation of the ultrasound images.** ***Method:*** **FCNN/CNNs are used to infer high-level context using low-level image features. In this paper, a sub-problem specific deep supervision of the FCNN is performed. The attention of fine resolution layers is steered to learn object boundary definitions using auxiliary losses, whereas coarse resolution layers are trained to discriminate object regions from the background. Furthermore, a customized scheme for downweighting the auxiliary losses and a trainable fusion layer are introduced. This produces an accurate segmentation and helps in dealing with the broken boundaries, usually found in the ultrasound images.** ***Results:*** **The proposed network is first tested for blood vessel segmentation in liver images. It results in $F1$ score, mean intersection over union, and dice index of 0.83, 0.83, and 0.79, respectively. The best values observed among the existing approaches are produced by U-net as 0.74, 0.81, and 0.75, respectively. The proposed network also results in dice index value of 0.91 in the lumen segmentation experiments on MICCAI 2011 IVUS challenge dataset, which is near to the provided reference value of 0.93. Furthermore, the improvements similar to vessel segmentation experiments are also observed in the experiment performed to segment lesions.** ***Conclusion:*** **Deep supervision of the network based on the input-output characteristics of the layers results in improvement in overall segmentation accuracy.** ***Significance:*** **Sub-problem specific deep supervision for ultrasound image segmentation is the main contribution of this paper. Currently the network is trained and tested for fixed size inputs. It requires image resizing and limits the performance in small size images.**

***Index Terms*—Ultrasound image segmentation, convolutional neural network, sub-problem specific deep supervision.**

I. Introduction

ULTRASOUND (US) images are routinely acquired in clinical activities, for example, diagnosis, intraoperative and postoperative observations, and surgery planning [1], [2]. In many of these applications, an accurate delineation or segmentation of different anatomical structures is desirable. In particular, blood region segmentation is used in vessel size (diameter) measurement for liver surgery planning [3], lumen area measurement for degree of vessel stenosis calculation [4] and subchorionic haemorrhage severity estimation [5]. In general, such segmentation is manually performed by the clinicians [6], which reduces objectiveness of the diagnosis and also makes the process labour intensive [7], [8]. Therefore, an automatic framework for blood region segmentation in US images is desirable.

US images are probably the most difficult medical images for automatic segmentation [9]. The challenges are acoustic shadows, poor contrast, and presence of speckle. The missing and incomplete structural boundaries further increase the difficulties. Some of these artifacts can be seen in the images shown in Fig. 1, where liver US images and manually marked corresponding blood vessel regions are shown in the top and bottom rows, respectively. These are conventional B-mode US images which are acquired using convex array type US transducer. As can be seen, very few pixels ($<2\%$) in the images belong to blood regions. Apart from that, variations in size and shapes of the vessels, along with imaging artifacts, are some other bottlenecks which make the segmentation problem complex.

Blood vessel segmentation provides critical information for the procedures like partial hepatectomy and hepatocellular carcinoma therapies [3]. In general, Computed Tomography (CT) images are used for this purpose. However, an accurate segmentation of vessel regions in US images can provide an alternate and help in avoiding harmful radiation associated with CT. Blood vessel segmentation in US images has been extensively studied. Thresholding is commonly used for this purpose due to its simplicity and low echogenicity of the blood regions [10]. However, it fails in presence of artifacts like acoustic shadows or signal dropout. These issues are resolved in circle or ellipse fitting

Manuscript received December 5, 2017; revised March 18, 2018, July 3, 2018, September 6, 2018, and October 5, 2018; accepted October 14, 2018. Date of publication October 22, 2018; date of current version May 20, 2019. *(Corresponding author: Deepak Mishra.)*

D. Mishra is with the Department of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi 110016, India (e-mail: deemishra21@gmail.com).

S. Chaudhury is with the Department of Electrical Engineering, Indian Institute of Technology Delhi, and also with the Central Electronics Engineering Research Institute.

M. Sarkar is with the Department of Electrical Engineering, Indian Institute of Technology Delhi.

A. S. Soin is with the Medanta Hospital.

Digital Object Identifier 10.1109/TBME.2018.2877577

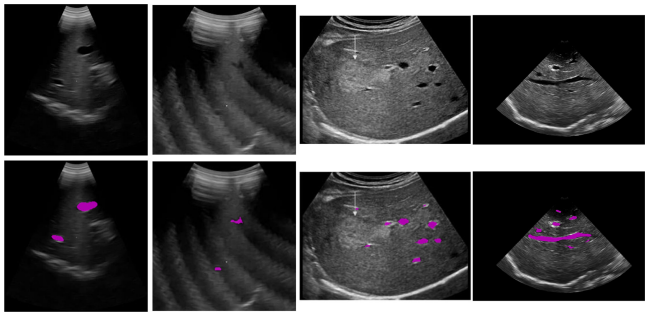


Fig. 1. Examples of blood vessel segmentation. Variations in shapes and sizes of different vessels along with signal loss and speckle make the task challenging. First row: Real US liver images. Second row: Manually segmented vessel regions.

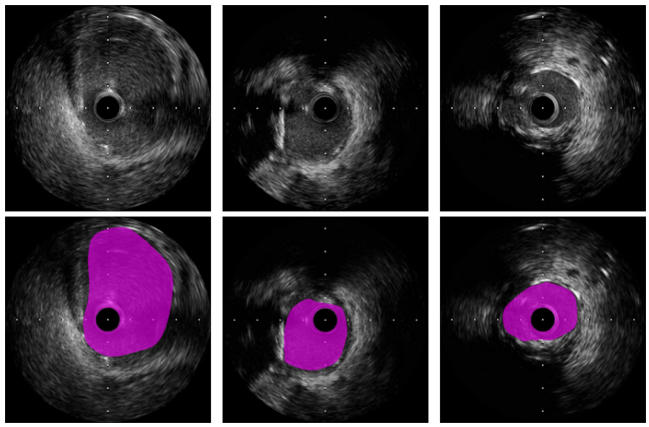


Fig. 2. Examples of lumen segmentation. First row: IVUS liver images. Second row: Manually segmented lumen regions.

approaches by taking vessel anatomy into account [11]–[13]. However, these are semi-automatic approaches, which require manual intervention and often fail in presence of deformation.

Frangi *et al.* [14] developed an automatic approach for tubular structure detection in medical images using pixel intensity information and vessel shapes. Frangi filter performs vessel segmentation using eigenvalue analysis of the Hessian matrix calculated at every pixel of the image. Similar to thresholding, Frangi filter is also unable to handle artifacts like shadowing. Feature learning approaches are also popular for vessel region segmentation in US images. For example, in [15] the input images are first converted into dip images by measuring intensity dips along the beamlines. These dip images are used as the training set to calculate vessel probability density function. The approach is susceptible to misclassification of non-blood pixels and results in inaccurate boundaries.

Blood vessel size measurement using the segmentation of conventional B-mode US images provides useful information for applications like liver surgery. However, such measurements are not sufficient for diagnosis of diseases like atherosclerosis. It requires measurement of the degree of vessel stenosis [4]. The degree of vessel stenosis is generally estimated using the ratio between blood (lumen) area and vessel cross-sectional area obtained from Intravascular US (IVUS) images. Some examples of lumen region segmentation in IVUS images of coronary arteries are shown in Fig. 2. IVUS images are acquired with the insertion of a catheter inside the vessels. The images contain a circular black region of a fixed radius at the center which represents catheter. Lumen region is found surrounding the catheter. Low contrast at the lumen border and catheter shadows are the major challenges faced in lumen segmentation.

Some well-known techniques for lumen region segmentation in IVUS images include active contours [16], [17], K-means clustering [18], fuzzy clustering [19], combination of morphological operations and thresholding [20]. All these approaches are sensitive to noise and need an efficient despeckilng of US images prior to segmentation. Further, these approaches are also dependent on system initialization.

Machine learning techniques are potential alternatives to the classical US image segmentation approaches [21]–[23]. However, performance of the conventional learning techniques is limited due to handcrafted features. Deep neural networks with millions of trainable parameters, for example convolutional neural network (CNN), are the better options [24]. CNNs have achieved phenomenal success in various computer vision tasks [25]–[28]. CNNs have also been used in CT and MRI image segmentations [29]–[31], however, their use in US image segmentation is limited due to less anatomical details and scarcity of the annotated training data.

The available medical image segmentation networks are mostly inspired from fully convolutional neural network (FCNN) [32]. FCNN is a semantic segmentation network. A conventional CNN contains convolutional, pooling, and fully connected layers. In FCNN, fully connected layers are replaced by convolutional layers and each pixel is classified into its predicted class. Further, skip connections are included to improve the prediction by using intermediate outputs. Recently, in [33], FCNN has been evaluated on cardiac US images and a better accuracy than the conventional approaches like database-guided segmentation [34] has been achieved.

FCNN is extended to develop U-net [35] which is considered as the state-of-the-art medical image segmentation network [36]. It contains contractive and expansive paths created using the combination of convolutional layers with pooling and upsampling layers. Skip connections are used to concatenate the features from contractive and expansive path layers. Extending further the work on FCNN, stacked FCNN modules are used in [36] wherein the prediction from initial FCNN modules is refined by the later ones.

FCNN's were designed for natural images having completely different characteristics from US images. With several artifacts limiting the information content of US training data, deep networks like FCNN and its variants become hard to train. A possible solution to the problem is deep supervision [37], which provides an optimal way of training the network to its potential [38]. Deeply supervised nets (DSNs) are well suited for structured input-output applications like image segmentation [39]. DSN uses auxiliary branches containing a small number of parameters to supervise the training of intermediate layers of the core network. These auxiliary layers try to predict the desired

outcome from intermediate layer outputs. In general, the auxiliary layers use the same loss function as the final output layer of the network. Auxiliary predictions are fused at the end to get the final output.

In this work, a sub-problem specific DSN is proposed for blood region segmentation in US images. In general, segmentation can be divided into two sub-problems: discrimination of object regions from background and identification of the spatial extent of objects or delineation of object boundaries [40]. Accordingly, in this work, the objectives of auxiliary layers are defined based on input-output characteristics of the corresponding convolutional layers of the core network. Core layers with fine resolution outputs are used to learn fine details like object boundaries. On the other hand, coarse resolution outputs are combined with auxiliary layers for object region discrimination. A weighted fusion of auxiliary outputs results in the desired segmentation.

DSNs have been used in the past for medical image segmentations. Chen *et al.* [41] have used DSN for neuronal structure segmentation. Their work has been extended in [42] for gland segmentation in histology images. Different auxiliary layers are used to simultaneously learn the details of objects and their contours. These layers receive a common input from coarse resolution convolutional layers situated near the end of the core network and use only higher level dense features for prediction. The benefits of combining region and boundary information are well known [43], [44]. However, fine details like object boundaries require deep supervision of fine resolution layers [39], [45]. Boundary definitions at the coarse layers become incoherent. Further, with feature deficient images like US, the supervision of layers with fine resolutions becomes more critical. Also burdening the same convolutional layers to learn features for multiple tasks naturally increases the training complexity. The proposed sub-problem specific DSN provides following advantages for US image segmentation:

1) Fundamentally, US imaging works on the acoustic impedance differences of different tissues. The appearances of different tissues remain consistent with scaling of the images, for example vessel and surrounding parenchyma. Thus, the regions with different gray level intensities can be identified even in downscaled images. Accordingly, in the proposed network, the desired structures are detected using auxiliary layers connected to the coarse resolution core layers.
2) Subtle boundaries of anatomical structures are best visualized at fine resolutions. Downscaling of the images reduces delineation accuracy. Hence, extending deep supervision to direct the attention of fine resolution layers towards object boundaries complements the object detection performed by subsequent layers.
3) In the proposed network an auxiliary layer connecting input to the output performs operations analogous to filtering and thresholding. It makes use of tissue appearances and homogeneity and provides low level feature details during the fusion with higher level features produced by other auxiliary layers.

The proposed network is tested and comprehensively evaluated for blood vessel segmentation in conventional B-mode liver US images. It results in mean intersection over union (mIoU) value of 0.83, whereas the values resulted in by Frangi filter, U-net, and feed-forward CNN are 0.57, 0.81, and 0.72, respectively. Similarly, the dice index value resulted in by the proposed network is 0.79, whereas the values resulted in by Frangi filter, U-net, and feed-forward CNN are 0.25, 0.75, and 0.60, respectively. The experiments are then extended to segment the lumen region in IVUS image dataset introduced in IVUS challenge, MICCAI 2011. Further, to see the generalization capability of the proposed network, experiments are performed to segment focal liver lesion (FLL) in Contrast-Enhanced US (CEUS) images and the results are compared with recent approaches.

The rest of the paper is organized as follows. Section II presents the proposed DSN with the implementation details. The experimental results on different datasets are presented in Section III. Finally, the paper is concluded in Section IV.

II. Proposed Methodology

Performance of deep neural networks is dependent on the amount and quality of datasets. Therefore, the dataset details are first presented in following subsection.

A. Datasets

Due to the unavailability of annotated 3D dataset, the proposed network is tested on 2D US images. The network can be extended to 3D images similar to the work reported in [46]. The ground truth segmentations are marked by the volunteers based on the guidelines provided by an expert radiologist. Later the marked segmentation are reviewed by the radiologist for coarse corrections and boundary delineation.

1) Vessel Segmentation Dataset: The liver contains blood vessels of variable shape, size, and appearance. Therefore, the liver images form a reasonable dataset for vessel segmentation experiment. A dataset comprising of 350 liver US images is used. It contains 281 images acquired using GE Voluson US imaging system with convex array transducers. The images are acquired with different frequency and field of view.[1] The remaining 69 images are obtained from the database[2] publicly available for non-commercial usage.

2) IVUS Challenge, MICCAI 2011 Dataset: The challenge introduced two datasets. Dataset A is a single frame 2D image dataset whereas dataset B is a multi-frame dataset created from the full in-vivo pullbacks of human coronary arteries [47]. In this work, dataset A is used for the experiments. It contains 77 images obtained from a digital 40 MHz IVUS scanner. As per the challenge, the dataset was divided into training and test sets containing 20 and 57 images, respectively. Manually annotated ground truths for the targeted lumen regions are provided by multiple observers. The intra-observer variability is also provided to be used as a reference.

3) Lesion Segmentation Dataset: This set is created using the liver images obtained from the SYSU-FLL-CEUS dataset [48], [49]. The SYSU-FLL-CEUS dataset contains

[1]The complete study has been approved by the Institutional Review Board of Medanta Hospital, Gurgaon, India.

[2]http://www.ultrasoundcases.info/Cases-Home.aspx

CEUS videos of different types of liver lesions. In this work, Focal Nodular Hyperplasia (FNH) videos are considered. These videos show lesions with large variations in size, location, and enhancement patterns. Total 152 images from different FNH videos representing all three phases are used in lesion segmentation experiment. The lesions are easily distinguished in arterial phases and are moderately visible in portal phases. However, as the enhancement fades out in late phases, contrast between the lesion and surrounding tissues becomes poor. This makes the lesion and surrounding tissues indistinguishable. An accurate segmentation of the lesion in late phases will increase the utility of the CEUS.

B. Network Architecture

Complete architecture of the deeply supervised FCNN used in this work is shown in Fig. 3. The core network is similar to VGG-16 [26]. The proposed network contains a series of convolutional and maxpooling layers, supported by the auxiliary side layers. Each convolutional layer output is batch normalized and passed through the leaky rectified linear units (ReLU). The outputs of ReLU non-linearities are fed as input to the subsequent convolutional or pooling layer. The auxiliary side layers facilitate deep supervision of the network. The attention of side layers E_0 to E_3 is steered towards the object boundary definitions whereas the other side layers O_4 and O_5 are focused to identify the regions belonging to the desired object (blood region). A weighted fusion of the side layer predictions provides the final segmentation. More details about the fusion layer (H) are presented in Section II-C. L_{e_i} is the loss responsible for boundary definitions and is used by the layer E_i. Similarly, L_{o_i} represents the loss used by the layer O_i to detect the object. The loss term L_{o_f}, associated with the fusion layer H, provides the desired segmented output. All these losses are class balanced cross-entropy losses:

$$L_l\left(\mathbf{W}, \mathbf{w}^l\right) = -\beta \sum_{j \in Y_+} \log P\left(y_j = 1 | X; \mathbf{W}, \mathbf{w}^l\right) - (1-\beta) \sum_{j \in Y_-} \log P\left(y_j = 0 | X; \mathbf{W}, \mathbf{w}^l\right) \quad (1)$$

where L_l is the loss associated with the layer l. X and Y form a sample pair used for training of the network. $y_j \in \{0, 1\}$ is the label of the jth pixel. Y_+ and Y_- represent the set of pixel indices belonging to the positive and negative classes, respectively. For loss L_{e_i}, Y_+ represents the set of indices belonging only to the boundaries of the objects, whereas for loss L_{o_i}, it represents the set of indices covered by the entire object. $\mathbf{W}$ and $\mathbf{w}^l$ represent the parameters of the core network and the parameters of the lth auxiliary layer, respectively. Further, $\beta = \frac{|Y_-|}{|Y|}$ is a parameter used to balance the effect of skewed positive and negative classes. Similar to (1), L_{o_f} is defined as:

$$L_{o_f}(\mathbf{W}, \mathbf{w}, \mathbf{h}) = -\beta \sum_{j \in Y_+} \log P(y_j = 1 | X; \mathbf{W}, \mathbf{w}, \mathbf{h}) - (1-\beta) \sum_{j \in Y_-} \log P(y_j = 0 | X; \mathbf{W}, \mathbf{w}, \mathbf{h}) \quad (2)$$

where $\mathbf{h}$ represents the parameters of fusion layer. The probabilities in both (1) and (2) are obtained using sigmoid function. The complete loss $\mathfrak{L}$ is given by:

$$\mathfrak{L} = L_{o_f}(\mathbf{W}, \mathbf{w}, \mathbf{h}) + \sum_{i=0}^{3} L_{e_i}(\mathbf{W}, \mathbf{w}^{E_i}) + \sum_{i=4}^{5} L_{o_i}(\mathbf{W}, \mathbf{w}^{O_i}) \quad (3)$$

The advantage of training the side layers for boundary definitions can be understood by considering the derivatives of different loss terms for these layers. If $Y^{(l)}$ represents the output of layer l, then

$$\frac{\partial \mathfrak{L}}{\partial Y^{(E_i)}} = \frac{\partial L_{o_f}}{\partial Y^{(E_i)}} + \frac{\partial L_{e_i}}{\partial Y^{(E_i)}} \quad (4)$$

Accordingly,

$$\frac{\partial \mathfrak{L}}{\partial \mathbf{w}^{E_i}} = \frac{\partial \mathfrak{L}}{\partial Y^{(E_i)}} \frac{\partial Y^{(E_i)}}{\partial \mathbf{w}^{E_i}} \quad (5)$$

$$= \left(\frac{\partial L_{o_f}}{\partial Y^{(E_i)}} + \frac{\partial L_{e_i}}{\partial Y^{(E_i)}} \right) \frac{\partial Y^{(E_i)}}{\partial \mathbf{w}^{E_i}} \quad (6)$$

Hence, weight update has a dedicated term to learn the features of object boundaries which is not the case with coarse resolution side layers O_4 and O_5. The coarse resolution layers learn dense features and are therefore trained to identify object regions in the image. On the other hand, the fine resolution side layers, E_0 to E_3, are trained to learn both the boundaries and the location of the objects. The effect of (6) is extended to the core network layers connected to the side layers which results in the deep supervision of all the layers. The proposed network is named as DSN-OB.

Further, to understand the effect of the sub-problem specific deep supervision, a network is designed by replacing all the E_i auxiliary layers with O_i in DSN-OB and named as DSN-O. Accordingly, the loss of DSN-O is defined as:

$$\mathfrak{L} = L_{o_f}(\mathbf{W}, \mathbf{w}, \mathbf{h}) + \sum_{i=0}^{5} L_{o_i}(\mathbf{W}, \mathbf{w}^{O_i}) \quad (7)$$

Thus, unlike DSN-OB, all the side layers in DSN-O are trained to learn the object definitions and perform the complete segmentation. Similarly, a smaller version of DSN-O, named as DSN-Os, is trained to understand the effect of the predictions from the input layer. It does not contain the O_0 auxiliary layer and the loss is defined as:

$$\mathfrak{L} = L_{o_f}(\mathbf{W}, \mathbf{w}, \mathbf{h}) + \sum_{i=1}^{5} L_{o_i}(\mathbf{W}, \mathbf{w}^{O_i}) \quad (8)$$

Input layer of the networks are with the finest resolution, therefore, the effect of the absence of O_0 layer is interesting from network performance analysis point of view. DSN-Os is equivalent to the conventional DSN architectures [38], [39].

C. Fusion

Fusion layer (H) is an important part of the proposed network. Instead of an arithmetic [41] or an engineered sum [50], fusion layer weights are trained to result in a linear combination of the auxiliary outputs. These weights are learned

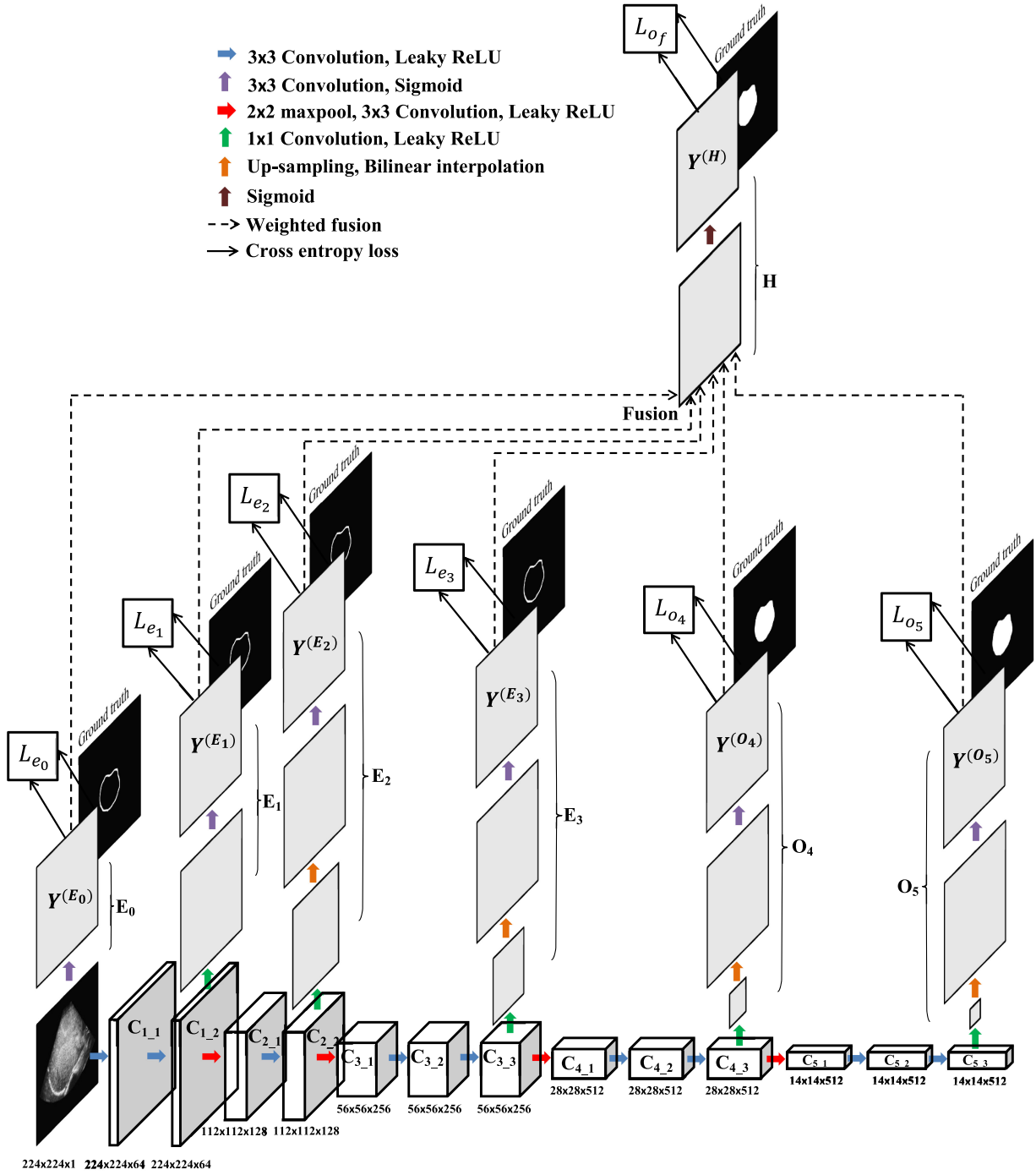


Fig. 3. Complete architecture of the DSN-OB network. The convolutional layers from C_{1_1} to C_{5_3} form the core network and deeply supervised by the auxiliary layers. The auxiliary layers represented by E_i's are used to learn the boundary definition and the layers represented by the O_i's are focused on the object region discrimination from the background. This sub-problem specific deep supervision results in the desired segmentation.

along with other network parameters. The final output of the network $Y^{(H)}$ can be represented using fusion layer weights $\mathbf{h} = \{h_0, h_1, h_2, h_3, h_4, h_5\}$ as:

$$Y^{(H)} = \sigma(\mathcal{H}) \tag{9}$$

$$\mathcal{H} = h_0 Y^{(A_0)} + h_1 Y^{(A_1)} + \cdots\cdots + h_5 Y^{(A_5)} \tag{10}$$

Here $\sigma(.)$ represents sigmoid function and A_i is a general notation used to represent the ith auxiliary layer with an analogous loss term L_{a_i}. Analogous to (4), we may write,

$$\frac{\partial \mathfrak{L}}{\partial Y^{(A_i)}} = \frac{\partial L_{o_f}}{\partial Y^{(H)}} \frac{\partial Y^{(H)}}{\partial Y^{(A_i)}} + \frac{\partial L_{a_i}}{\partial Y^{(A_i)}} \tag{11}$$

$$\frac{\partial \mathfrak{L}}{\partial Y^{(A_i)}} = \frac{\partial L_{o_f}}{\partial Y^{(H)}} \sigma'(\mathcal{H}) h_i + \frac{\partial L_{a_i}}{\partial Y^{(A_i)}} \tag{12}$$

As explained later in Section II.D.2, the effect of L_{a_i} is reduced at later training stages and the auxiliary layers only observe a

scaled gradient of the final output loss with the scale factor of respective $\sigma'(\mathcal{H})\mathrm{h}_i$'s. Analogously, in later training stages, the auxiliary outputs are scaled to obtain the final output. Learning of the fusion layer weights introduces flexibility in the network to automatically select the best features for generating the dilated segmentation and use the remaining features for refinement.

D. Training

Network performance is stochastically optimized using gradient descent. Due to the limited available hardware capacity (Nvidia GeForce GTX 980 GPU, 4 GB), a small minibatch size of 10 images is used for training. An initial learning rate of 0.0001 is used with the momentum of 0.9 for vessel and lesion segmentation experiment. In lumen segmentation experiment, an order of lower learning rate with same momentum value is used due to a comparatively smaller size of the training set. Learning rate is reduced by a factor of 3 at the training error plateaus. The networks are trained for 500 epochs in all experiments. The complete training takes about 25 hours on the GPU in vessel segmentation experiment and comparatively smaller time in the other two experiments. Apart from batch normalization, L_2 norm of the network weights $(\mathbf{W}, \mathbf{w}, \mathbf{h})$ with a decay factor of 0.0001 is used during the optimization process. The success of DSNs depends on the deep supervision provided by the auxiliary layers, therefore, instead of using pretrained weights of any existing architecture, the proposed networks are trained from scratch and the weights are initialized as in [51]. All the networks are built on top of Theano [52].

1) Data Preprocessing: During training, the input size is fixed at $224 \times 224 \times 1$. The height and width of the images have substantial variation. Therefore, larger side of the images is first cropped to make height and width equal, then the images are resized to 224×224 using bicubic interpolation or downsampling, depending on their original sizes. Further, considering the limited size of the available dataset, it is necessary to use techniques like augmentation. In this work three types of augmentations, horizontal flipping, random linear translation, and gamma augmentation, are applied to increase the size of the training set. Vertical flipping cannot be applied due to the conical field of view of US transducers. The IVUS images do not have any such limitations, therefore, two additional augmentations, vertical flipping and image transpose, are used.

2) Class Balancing and Refinement: As in (1) and (2), the parameter β is used for class balancing. Its value is calculated after data preprocessing. Further, as reported in [45], a gradual reduction of the effect of the side layer loss terms improves learning. Therefore, the training objective $\mathfrak{L}$ is modified as:

$$\mathfrak{L} = L_{o_f}(\mathbf{W}, \mathbf{w}, \mathbf{h}) + \left(1 - \frac{t}{T}\right)\left(\sum_{i=0}^{5} L_{a_i}\left(\mathbf{W}, \mathbf{w}^{A_i}\right)\right) + \lambda\left(||\mathbf{W}||_2 + ||\mathbf{w}||_2 + ||\mathbf{h}||_2\right) \tag{13}$$

where the third term is the L_2 norm of the network weights, used for regularization. λ is the weight decay parameter. t and T are current and total number of training epochs. As the value of t approaches T, the effect of $L_{o_f}(\mathbf{W}, \mathbf{w}, \mathbf{h})$ dominates and the network parameters are trained for the desired output. However, this makes the loss $\mathfrak{L}$ a function of the total number of iteration, which reduces the capability of fine-tuning and reusing the trained weights for transfer learning. Instead of using total number of iterations as the value of T, an intermediate value can be used as a solution to the problem. This requires further modifications of the objective function as:

$$\mathfrak{L} = L_{o_f} + \max\left\{0, \left(1 - \frac{t}{T_s}\right)\right\}\left(\sum_{i=0}^{5} L_{a_i}\right) + \lambda\left(||\mathbf{W}||_2 + ||\mathbf{w}||_2 + ||\mathbf{h}||_2\right) \tag{14}$$

where $T_s < T$. Experiments with different values of T_s are performed which include two special cases of $T_s = T$ and $T_s = 0$. Here $T_s = T$ is the conventional way of downweighting the auxiliary losses [45] whereas $T_s = 0$ represents the condition in which the auxiliary losses are absent. Although with $T_s = 0$, the deep supervision of the core network layers due to auxiliary losses is not present, the core network still experiences an implicit deep supervision due to the presence of auxiliary branches. This can be understood by considering the loss gradient for core network layers. In a general scenario the loss gradient is given as:

$$\frac{\partial \mathfrak{L}}{\partial Y^{(C_{i\_j})}} = \frac{\partial L_{o_f}}{\partial Y^{(C_{(i+1)\_1})}} \frac{\partial Y^{(C_{(i+1)\_1})}}{\partial Y^{(C_{i\_j})}} + \frac{\partial L_{o_f}}{\partial Y^{(H)}} \sigma'(\mathcal{H})\mathrm{h}_i \frac{\partial Y^{(A_i)}}{\partial Y^{(C_{i\_j})}} + \frac{\partial L_{a_i}}{\partial Y^{(A_i)}} \frac{\partial Y^{(A_i)}}{\partial Y^{(C_{i\_j})}} \tag{15}$$

where C_{i_j} represents the last convolutional layer of ith stage in the core network, to which auxiliary branch A_i is connected. The first term in (15) represents the propagation of supervision through core network convolutional layers, which is prone to vanishing gradient. On the other hand, the last two terms appear due to auxiliary branches and represent a direct supervision of core network layers. Thus, even in absence of the third term, which comes from auxiliary losses, the network is deeply supervised due to the second term. In complete absence of deep supervision, the network turns into a vanilla CNN in which, the supervision propagates only through a single feed-forward path (CNN-ff).

3) Cross-Validation: Five-fold cross-validation is used to verify the performance of the proposed network in blood vessel segmentation. The complete dataset is divided into five random non-overlapping sets of equal sizes. Images of four sets are used for training and the remaining set is used for validation. Similarly, in lesion segmentation, three-fold cross-validation is used. For lumen segmentation in IVUS images, the training and test sets as per the MICCAI 2011 challenge are used.

III. Experimental Results

This section begins with the comparison of DSN-O and DSN-Os networks for blood vessel segmentation. This helps in understanding the advantage of the input layer predictions in US image segmentation. Further, the analysis of the DSN-OB network performance and its comparison with DSN-O and some

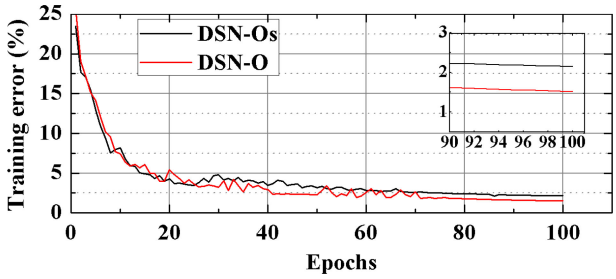


Fig. 4. Training behaviour of DSN-O and DSN-Os. The plot clearly shows that DSN-O converges faster as compared to DSN-Os.

well-known vessel segmentation approaches is presented. The lumen segmentation in IVUS images and lesion segmentation in CEUS images are also described.

A. Comparison of the DSN-O and DSN-Os

The difference between DSN-O and DSN-Os is the auxiliary layer (O_0), which makes a direct connection between input and fusion layer. The reduction in training error with increasing epochs for the two networks is shown in Fig. 4. A similar training behaviour is observed for all the five-fold validation experiments performed for vessel segmentation. DSN-O shows a faster convergence as compared to DSN-Os network.

Outputs of the two networks trained with $T_s = 300$ are compared using $F1$ score, dice index and mIoU. $F1$ score is useful for object detection and is defined as:

$$F1 = \frac{2PR}{P+R},\ P = \frac{N_{TP}}{N_{TP}+N_{FP}},\ R = \frac{N_{TP}}{N_{TP}+N_{FN}} \quad (16)$$

where P and R represent the precision and recall. N_{TP}, N_{FP}, and N_{FN} are the count of true positives, false positives, and false negatives, respectively. If a segmented vessel object in the network output has more than 50% area overlap with its corresponding object in the ground truth, it is considered as a TP otherwise FP [42]. A ground truth vessel object which does not have at least 50% area overlap with any of the segmented objects is considered as FN. The ground truth corresponding to a segmented object is the object in the annotated image having a maximum area overlap with the segmented object.

Dice index is generally used to evaluate the segmentation accuracy. Segmentation is seen as a pixel level classification where each pixel is classified as object (vessel) or background (non-vessel). Accordingly, the dice index is defined as:

$$D(G,S) = \frac{2|G \cap S|}{|G|+|S|} \quad (17)$$

where G and S are the sets of the pixels classified as object in ground truth and segmented output obtained from the trained networks, respectively.

The third metric mIoU is similar to the dice index. For binary segmentation (vessel and non-vessel), it is defined as:

$$\text{mIoU} = \frac{1}{2}\left(\frac{n_{TP}}{n_{TP}+n_{FP}+n_{FN}} + \frac{n_{TN}}{n_{TN}+n_{FN}+n_{FP}}\right) \quad (18)$$

TABLE I
EVALUATION METRIC VALUES OBSERVED FOR DSN-OS AND DSN-O FOR BLOOD VESSEL SEGMENTATION DATASET

Network	F1 score	Dice index	mIoU
DSN-Os	0.68	0.72	0.78
DSN-O	**0.73**	**0.74**	**0.80**

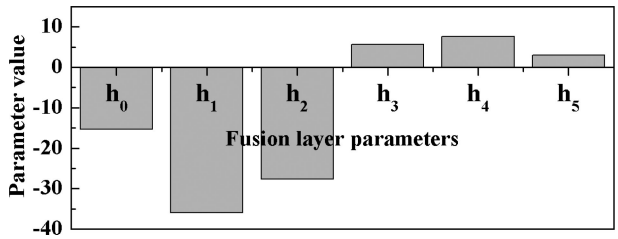


Fig. 5. Fusion layer parameters (**h**) observed from the trained DSN-O network.

where n_{TP} and n_{TN} are the total number of correctly classified object and background pixels, respectively. n_{FP} are the total number of background pixels incorrectly classified as object pixels. Similarly, n_{FN} are the total number of object pixels incorrectly classified as background pixels.

Average values of these evaluation metrics observed from the outputs of the two networks are shown in Table I. Substantial difference in $F1$ score values shows the improvement in the vessel object detection accuracy. For further analysis, the fusion layer parameters (**h**) observed from the trained DSN-O network are shown in Fig. 5. A large non-zero value of h_0 indicates considerable information gains from the O_0 layer.

Sample vessel segmentation outputs observed from the validation sets are shown in Fig. 6. The segmented blood vessel regions are in magenta and ground truth regions are highlighted using their boundaries in green. It is clear that the outputs of DSN-O network (Fig. 6, second row) have less shape mismatch with the ground truth as compared to the outputs of DSN-Os network (Fig. 6, first row). Further, comparing the outputs shown in the first column, it can be seen that the DSN-O also shows comparatively less false positives and less false negatives. The auxiliary layer in DSN-O connecting input and fusion layer provides the domain knowledge to the fusion layer and helps in improving the performance.

B. Comparison of the DSN-OB and DSN-O

Analogous to Section III-A, the DSN-OB and DSN-O networks are compared here. The difference between the two network is in training objective of the auxiliary layers. As described in Section II, unlike DSN-O, the fine resolution auxiliary layer in DSN-OB network are trained to focus on object boundaries. As observed during the comparison of DSN-Os and DSN-O, the connection between input and fusion layer improves the network performance, therefore, DSN-OB also contains an analogous auxiliary layer.

DSN-OB is trained for different values of the hyperparameter T_s. Observations for the first validation set in five-fold cross-validation are summarized in Fig. 7. The evaluation metrics, dice

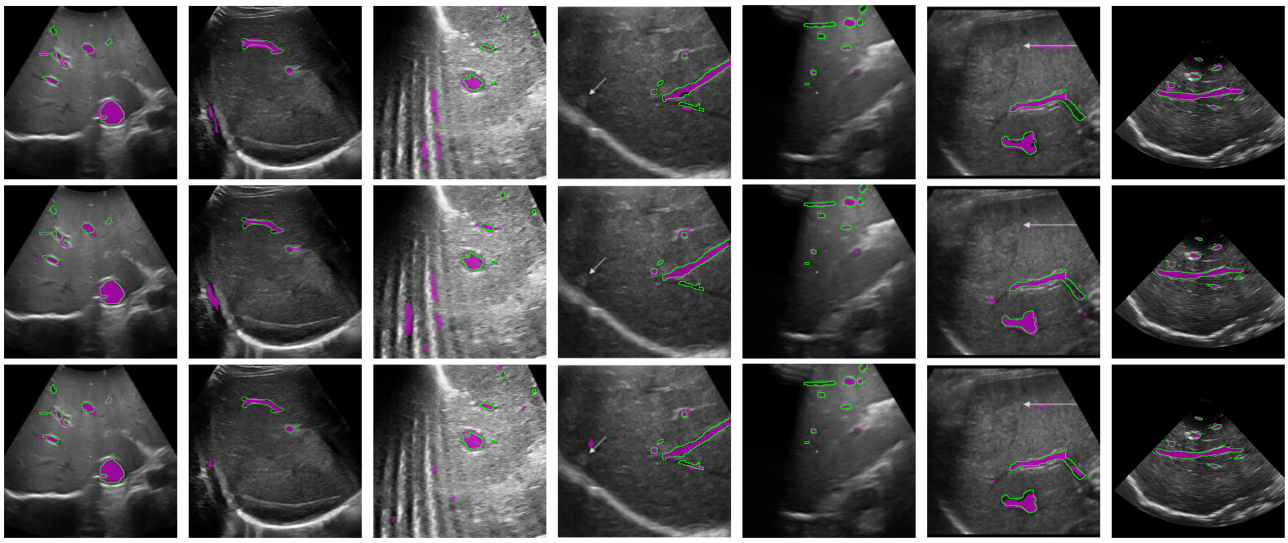


Fig. 6. Comparison of segmentation performances of the networks. The outputs obtained using DSN-Os, DSN-O, and DSN-OB are shown in the first, second, and third row, respectively. Segmented regions are shown in magenta and ground truth region boundaries are shown by green contours. The performance of DSN-O is better than DSN-Os, however, the best results are shown by DSN-OB.

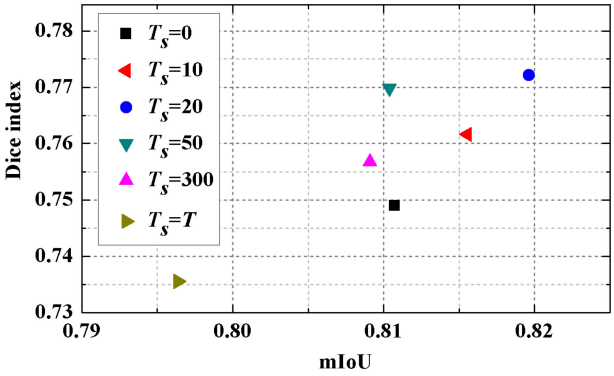


Fig. 7. Observations of the experiments performed with DSN-OB for different values of T_s.

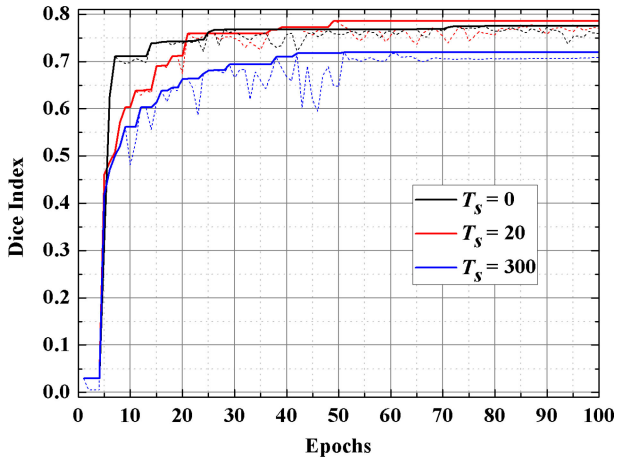


Fig. 8. Dice index variation for validation set in the experiments performed with DSN-OB for different values of T_s. Dotted lines show dice index values measured at every epoch whereas solid lines show the best value of dice index observed over all previous epochs.

index and mIoU, which are useful to measure the segmentation accuracy, are considered. The best performance is observed for $T_s = 20$.

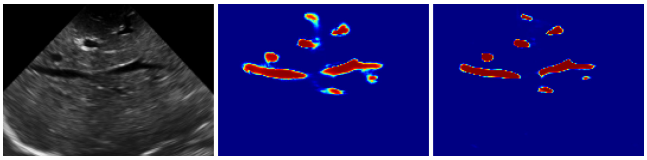
For further analysis, dice index variation with increasing epochs for the validation set is plotted in Fig. 8. Dotted lines show dice index values measured at every epoch whereas solid lines show the best value of dice index observed over all the previous epochs. As can be seen, DSN-OB with $T_s = 20$ results in the best performance. The dice index value slowly increases in the beginning and reaches the best value at the end, which shows the regularization effect of the auxiliary losses. In contrast, DSN-OB output with $T_s = 0$ only asymptotically meets the output of $T_s = 20$. The dice index curve for $T_s = 0$ shows rapid growth in the beginning, however, remains lower than $T_s = 20$. This suggests that in absence of auxiliary losses, the network has a possibility to get stuck into local minima. However, large value of T_s results in underfitting due to a large amount of regularization, for example the output of $T_s = 300$, shown in Fig. 8.

In subsequent experiments $T_s = 0$, $T_s = 20$, and $T_s = 300$ are considered for DSN-OB. The experiments are also performed with CNN-ff (no deep supervision) for completeness.

Sample outputs obtained using DSN-OB ($T_s = 20$) are shown in the bottom row of Fig. 6. For good quality and noise-free images (first column of Fig. 6), improvements in DSN-OB outputs are restricted to the boundaries when compared to DSN-O. However, for images corrupted by noise or having deceptive object appearances (second and third column of Fig. 6) DSN-OB outputs contain comparatively less false positives. Although no considerable difference in the convergence rate of the two networks is observed, there is a noticeable difference in the shape similarity of the DSN-OB segmented outputs with ground truths. These improvements are summarized in Table II, which shows average values of the evaluation metrics observed

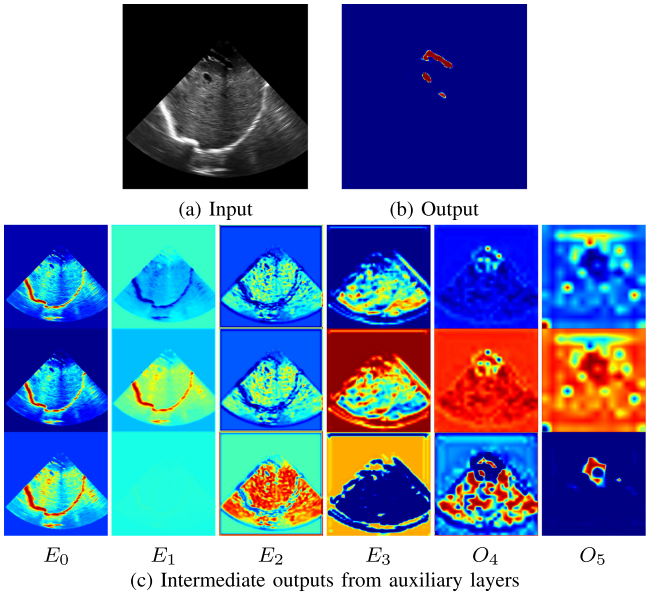
TABLE II
EVALUATION METRIC VALUES OBSERVED FROM THE OUTPUTS OF DIFFERENT NETWORKS FOR BLOOD VESSEL SEGMENTATION DATASET

Network	F1 score	mIoU	Dice index
Frangi filter	0.37	0.57	0.25
U-net	0.74	0.81	0.75
CNN-ff	0.57	0.72	0.60
DSN-O	0.73	0.80	0.74
DSN-OB (eq_fusion)	0.08	0.49	0.01
DSN-OB ($T_s = 0$)	**0.83**	0.83	0.78
DSN-OB ($T_s = 20$)	0.82	**0.83**	**0.79**
DSN-OB ($T_s = 300$)	0.77	0.81	0.76



(a) Input (b) DSN-O output (c) DSN-OB output

Fig. 9. Comparison of the output probability maps observed from DSN-O and DSN-OB networks. Clearly the boundaries in the output of DSN-OB are sharper than DSN-O.



(a) Input (b) Output

(c) Intermediate outputs from auxiliary layers

Fig. 10. (a) Input image. (b) DSN-OB output. (c) Intermediate outputs from the auxiliary layers. First row: Up-sampled output of the core network layer. Second row: Application of learned 3×3 convolutional filter on first row images. Third row: Application of sigmoid on second row images.

in five-fold cross-validation. DSN-OB shows considerable improvements over other networks for all values of T_s, which reflects the advantage of sub-problem specific deep supervision. Table II also shows the performance of DSN-OB where equal weighted fusion (eq_fusion) of auxiliary outputs is performed with $T_s = 0$. In eq_fusion, instead of learning the fusion layer parameters (h_i), all h_i's are assigned an equal value. It results in a poor performance which shows the importance of fusion layer training.

Further, the output probability maps for a test image observed from DSN-O and DSN-OB are shown in Fig. 9. DSN-OB shows comparatively better and sharper vessel boundaries. The advantage of using fine resolution layers to learn boundary definitions is clearly observed.

Fig. 10(a) shows an image where the vessel at the center of the image is partially visible due to the shadowed region. Fig. 10(b) shows that DSN-OB is able to detect the vessel accurately. To understand its learning behaviour, intermediate outputs from auxiliary layers are shown in Fig. 10(c). As discussed in Section II-B, the multichannel input received from a core network layer is first reduced to single channel and up-sampled by the auxiliary layers. The up-sampled outputs for the test image are shown in first row of Fig. 10(c). These up-sampled outputs are passed through 3×3 convolutional filters and sigmoid non-linearities which are shown in second and third row of Fig. 10(c), respectively. The presence of shadowed region is clearly visible in the outputs of initial layers, E_0 and E_1. However, the subsequent layers outputs show that the network learns semantic features of blood vessels and suppresses the shadowed region. The fusion layer parameters h_4 and h_5 have positive values whereas the parameters h_0 to h_3 are negative. Accordingly, the auxiliary outputs of O_4 and O_5 show dilated segmentation which are refined by the outputs of the remaining auxiliary layers (E_0 to E_3).

The complexity of the segmentation due to variation in size, shape, and appearances of the vessels is observed from the experimental outputs. Nevertheless, DSN-OB shows a performance close to the ground truth segmentation. This shows that training the fine resolution layers for the object boundaries can certainly benefit the object segmentation in the US images.

C. Comparison of the Proposed and Existing Approaches

Performance of DSN-OB is further compared with Frangi filter [14] and U-net [35]. Parameters of the Frangi filter are tuned using grid search. For U-net, a publicly available implementation[3] is used and trained analogous to the proposed network.

Observations from validation sets for the considered approaches are included in Table II. In Frangi filter, implementation parameters are tuned for a randomly selected image from a set and same parameter values are used for remaining images. Frangi filter enhances the vessels and suppress the remaining region. Its outputs are binarized using thresholding to obtain the segmented outputs. The sample variations are not taken into account by the filter, which results in poor performance. On the other hand, U-net shows good performance in terms of all evaluation metrics. U-net, being a deep learning approach, is able to learn data variability along with the desired objectives. Table II also shows that the performances of U-net and DSN-O are comparable to each other, however, DSN-OB results are comparatively better.

[3]https://github.com/zhixuhao/unet

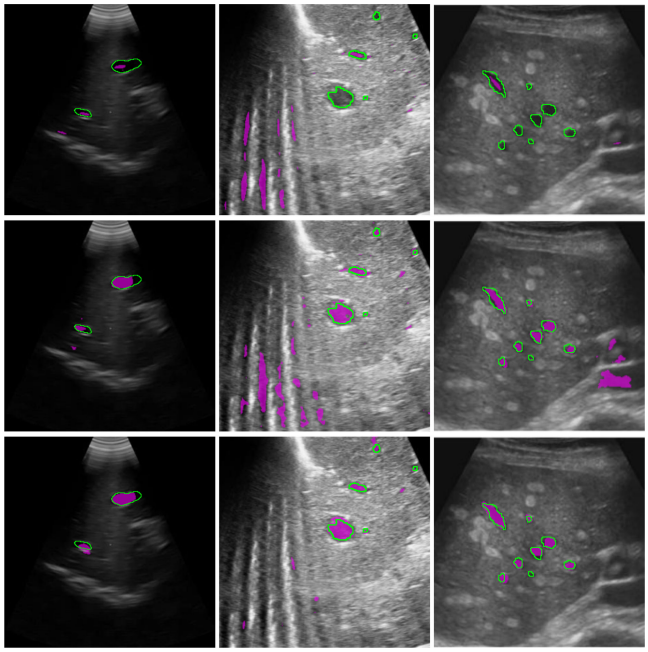


Fig. 11. The outputs obtained using Frangi filter, U-net, and DSN-OB are shown in the first, second, and third row, respectively. Segmented regions are shown in magenta and ground truth region boundaries are shown by green contours. Clearly the proposed DSN-OB network shows the best performance with least false negatives and false positives.

Segmented outputs from Frangi filter, U-net, and DSN-OB are respectively shown in the top, middle, and bottom rows of Fig. 11. The images shown have considerable variation in speckle extent and shapes of the vessels. The first column shows the images with low speckle extent whereas the other two columns show images with relatively higher speckle extent. In addition, the images in the first column are affected by signal fallout. Similarly, the images in the second column contain artifact generated due to environmental interference. Further, the images in the third column have low contrast and contain numerous cysts. The proposed DSN-OB network shows a robust performance against all these challenges. It also shows fewer false positives and more accurate boundaries as compared to U-net and Frangi filter. The performance can be further improved by post-processing, for example, morphological hole filling. However, such experiments are not conducted here.

D. Lumen Segmentation

Lumen segmentation is the second blood region segmentation experiment considered in this work. MICCAI 2011 IVUS challenge dataset is used in the experiment. The networks are trained using the training set comprising of 20 images. The evaluation metrics used in the challenge do not include any of the metrics used in the blood vessel segmentation experiment. However, the Jaccard measure (JM), which was used as one of the criteria in the challenge, can be used to calculate the dice index (D) as:

$$D = \frac{2JM}{1 + JM} \tag{19}$$

TABLE III
DICE INDEX VALUES OBSERVED FROM THE OUTPUTS OF THE DSN-OB FOR THE LUMEN SEGMENTATION IN MICAAI 2011 IVUS CHALLENGE DATASET

Network	Dice index
Intra-observer variability (reference)	0.93
U-net	0.87
CNN-ff	0.88
DSN-OB ($T_s = 0$)	**0.91**
DSN-OB ($T_s = 20$)	**0.91**
DSN-OB ($T_s = 300$)	0.87

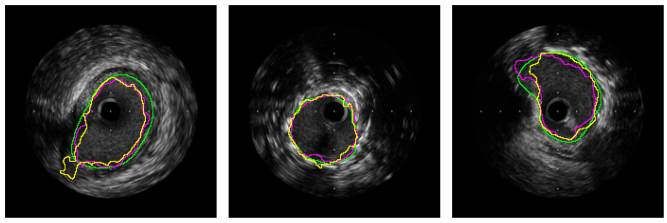


Fig. 12. Boundaries of the lumen regions in the test images from IVUS dataset, green curve shows the ground truth and magenta and yellow curves represent the DSN-OB and U-net outputs, respectively.

The average dice index values observed on the test set comprising of 57 images is shown in Table III. Although the training set is very small, DSN-OB shows a performance close to the reference.

Boundaries of the lumen regions segmented in test images are shown in Fig. 12. Green curve shows the ground truth and magenta and yellow curves represent the outputs of DSN-OB and U-net, respectively. The lumen regions in all the images are connected to the catheter region (lying at the center of the image), therefore, the segmented region disconnected from the center of the image is discarded. The first and second column outputs in Fig. 12 show that DSN-OB is able to segment lumen even in the presence of the catheter shadows.

E. Lesion Segmentation

Apart form blood region segmentation, lesion segmentation experiment is performed to evaluate the generalization capability of the proposed network. All CEUS images considered in this experiment contain one lesion. The low contrast of the images acquired in late phases is the major challange. The lesion in such images becomes barely distinguishable from the surrounding tissues. The performance of the segmentation approaches in these samples brings out the difference in evaluation metric values. In this experiment, DSN-OB is compared with U-net and the superpixel energy minimization based lesion segmentation approach [53] in three-fold cross-validation. The observed evaluation metric values are listed in Table IV. Variation in T_s does not have any considerable effect on DSN-OB performance. However, the values observed from DSN-OB are better than the other two approaches. The difference is also observed in segmented outputs shown in Fig. 13. The second and fourth column of Fig. 13 are the examples of late phase images. Both, U-net and the approach in [53], in this phase show sub-optimal performances.

TABLE IV
EVALUATION METRIC VALUES OBSERVED FROM THE OUTPUTS OF DIFFERENT APPROACHES FOR LESION SEGMENTATION DATASET

Network	F1 score	mIoU	Dice index
Wang et al. [53]	0.64	0.75	0.68
U-net	0.89	0.91	0.91
CNN-ff	0.97	0.92	0.92
DSN-OB ($T_s = 0$)	**0.99**	**0.94**	**0.94**
DSN-OB ($T_s = 20$)	0.98	**0.94**	**0.94**
DSN-OB ($T_s = 300$)	0.96	0.93	**0.94**

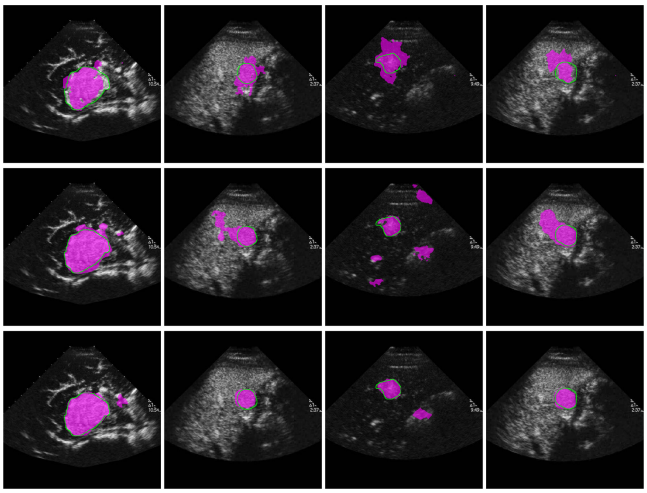


Fig. 13. Segmented outputs of the approach in [53], U-net and DSN-OB are shown in the first, second, and third row, respectively. Segmented regions are shown in magenta and ground truth region boundaries are shown by green contours. The better performance of the proposed network can be appreciated from the outputs in second and fourth column.

F. Discussion

There are several observations made during the experiments, which are discussed in this section. The most important observation is the utility of the sub-problem specific deep supervision. In this work, the final objective of blood region segmentation is divided into two sub-problems, boundary detection and discrimination of object regions from background. The fine resolution layers are trained to better learn the boundaries and show the desired performance improvement. Hence, deeply supervising the CNN layers depending on their input-output characteristics can provide a performance boost. Further, the lesion segmentation experiment shows that DSN-OB can be easily adapted for a general purpose US image segmentation by simply changing the training data.

DSN-OB has half the number of trainable parameters as compared to U-net, however, DSN-OB shows a better segmentation outputs. This gain is achieved from the auxiliary side layers which are missing in U-net. Moreover, DSN-OB network quickly learns the desired objective without using any advanced optimization techniques like AdaGrad [54] or Adam [55]. The sub-problem specific deep supervision results in a fast convergence of the network. It increases the discriminative capability of the intermediate layers without overfitting.

For small datasets with low complexity problems, for example, lumen segmentation, the network gives good performance even in absence of the auxiliary losses. The auxiliary connections without the auxiliary losses provide sufficient deep supervision to solve the problem. However, for a complex problem, like vessel segmentation, auxiliary losses are necessary for the desired performance.

There is a lack of smooth segmentation of vessel wall in images with size smaller than 224 × 224, for example Fig. 6, column five. However, it is not the case with all images, for example, Fig. 9 and 10 show an accurate segmentation. Since the network is trained and tested with 224 × 224 input size, the small size images are up-sampled using interpolation before the segmentation operation, which leads to some inaccuracies. Further, there are some open problems, which can be considered to take this work forward. For example, identification of the best position in the core network upto which the auxiliary layers should be trained for the boundaries, however, this has not been explored in our work.

IV. CONCLUSION

In this work, a DSN architecture suitable for US image segmentation is described. The auxiliary side layers in the network are trained to focus on different sub-problems of the final objective. Fine resolution layers try to learn the accurate boundary definitions and coarse resolution layers look for the complete object regions. The deep supervision of the network based on the input-output characteristics of the layers results in an improved learning. The network performs better as compared to the training of the auxiliary layers with an identical objective. Further, the experiments show that an auxiliary layer connecting the input and fusion layer helps in better learning of the US image characteristics and results in an improved performance.

The proposed network is compared with existing deep networks as well as classical US image segmentation approaches. It shows a better performance in terms of $F1$ score, dice index and mIoU. Although the proposed method has focused on blood region segmentation, it shows equally good performance on lesion segmentation which suggests that it can be extended to other applications as well.

ACKNOWLEDGMENT

Authors would like to acknowledge AKAB Healthcare Pvt. Ltd. to provide support for the study.

REFERENCES

[1] J. A. Noble and D. Boukerroui, "Ultrasound image segmentation: A survey," *IEEE Trans. Med. Imag.*, vol. 25, no. 8, pp. 987–1010, Aug. 2006.
[2] S. Petroudi *et al.*, "Segmentation of the common carotid intima-media complex in ultrasound images using active contours," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 11, pp. 3060–3069, Nov. 2012.
[3] S. Esneault *et al.*, "Liver vessels segmentation using a hybrid geometrical moments/graph cuts method," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 2, pp. 276–283, Feb. 2010.
[4] E. Brusseau *et al.*, "Fully automatic luminal contour segmentation in intracoronary ultrasound imaging-a statistical approach," *IEEE Trans. Biomed. Imag.*, vol. 23, no. 5, pp. 554–566, May 2004.
[5] S. M. Norman *et al.*, "Ultrasound-detected subchorionic hemorrhage and the obstetric implications," *Obstetrics Gynecology*, vol. 116, no. 2, Part 1, pp. 311–315, 2010.

[6] C. Y. Ahn *et al.*, "Fast segmentation of ultrasound images using robust Rayleigh distribution decomposition," *Pattern Recognit.*, vol. 45, no. 9, pp. 3490–3500, 2012.

[7] Y. Zhao *et al.*, "Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images," *IEEE Trans. Med. Imag.*, vol. 34, no. 9, pp. 1797–1807, Sep. 2015.

[8] X. Zang *et al.*, "Methods for 2-D and 3-D endobronchial ultrasound image segmentation," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1426–1439, Jul. 2016.

[9] J. A. Noble, "Ultrasound image segmentation and tissue characterization," *Proc. Inst. Mech. Eng., Part H, J. Eng. Med.*, vol. 224, no. 2, pp. 307–316, 2010.

[10] W. H. Nam *et al.*, "Automatic registration between 3D intra-operative ultrasound and pre-operative CT images of the liver based on robust edge matching," *Phys. Med. Biol.*, vol. 57, no. 1, pp. 69–91, 2011.

[11] P. Abolmaesumi *et al.*, "Real-time extraction of carotid artery contours from ultrasound images," in *Proc. 13th Symp. Comput.-Based Med. Syst.*, 2000, pp. 181–186.

[12] P. Abolmaesumi *et al.*, "Image-guided control of a robot for medical ultrasound," *IEEE Trans. Robot. Automat.*, vol. 18, no. 1, pp. 11–23, Feb. 2002.

[13] J. Guerrero *et al.*, "Real-time vessel segmentation and tracking for ultrasound imaging applications," *IEEE Trans. Med. Imag.*, vol. 26, no. 8, pp. 1079–1090, Aug. 2007.

[14] A. Frangi *et al.*, "Multiscale vessel enhancement filtering," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Interventation*, 1998, pp. 130–137.

[15] J. M. Blackall *et al.*, "Alignment of sparse freehand 3-D ultrasound with preoperative images of the liver using models of respiratory motion and deformation," *IEEE Trans. Med. Imag.*, vol. 24, no. 11, pp. 1405–1416, Nov. 2005.

[16] M. E. Plissiti *et al.*, "An automated method for lumen and media-adventitia border detection in a sequence of IVUS frames," *IEEE Trans. Inf. Technol. Biomed.*, vol. 8, no. 2, pp. 131–141, Jun. 2004.

[17] A. Vard *et al.*, "An automated approach for segmentation of intravascular ultrasound images based on parametric active contour models," *Australas. Phys. Eng. Sci. Med.*, vol. 35, no. 2, pp. 135–150, 2012.

[18] D. S. Jodas *et al.*, "Automatic segmentation of the lumen region in intravascular images of the coronary artery," *Med. Image Anal.*, vol. 40, pp. 60–79, 2017.

[19] E. Dos Santos *et al.*, "Detection of luminal contour using fuzzy clustering and mathematical morphology in intravascular ultrasound images," in *Proc. 27th Annu. Conf. IEEE Eng. Med. Biol.*, 2005, pp. 3471–3474.

[20] M. C. Moraes and S. S. Furuie, "Automatic coronary wall segmentation in intravascular ultrasound images using binary morphological reconstruction," *Ultrasound Med. Biol.*, vol. 37, no. 9, pp. 1486–1499, 2011.

[21] L. L. Vercio *et al.*, "Assessment of image features for vessel wall segmentation in intravascular ultrasound images," *Int. J. Comput. Assisted Radiol. Surgery*, vol. 11, no. 8, pp. 1397–1407, 2016.

[22] C. Kotropoulos and I. Pitas, "Segmentation of ultrasonic images using support vector machines," *Pattern Recognit. Lett.*, vol. 24, no. 4, pp. 715–727, 2003.

[23] G. Carneiro *et al.*, "Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree," *IEEE Trans. Med. Imag.*, vol. 27, no. 9, pp. 1342–1355, Sep. 2008.

[24] K. Lekadir *et al.*, "A convolutional neural network for automatic characterization of plaque composition in carotid ultrasound," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 48–55, Jan. 2017.

[25] A. Krizhevsky *et al.*, "Imagenet classification with deep convolutional neural networks," *Adv. Neural Inf. Process. Syst.*, vol. 25, pp. 1097–1105, 2012.

[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2015. [Online]. Available: https://arxiv.org/pdf/1409.1556.pdf

[27] K. He *et al.*, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[28] Y. LeCun *et al.*, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[29] J. A. Noble, "Reflections on ultrasound image analysis," *Med. Image Anal.*, vol. 33, pp. 33–37, 2016.

[30] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," vol. 42, pp. 60–88, 2017.

[31] D. Shen *et al.*, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, 2017.

[32] J. Long *et al.*, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

[33] H. Chen *et al.*, "Iterative multi-domain regularized deep learning for anatomical structure detection and segmentation from ultrasound images," in *Proc. Med. Image Comput. Comput.-Assisted Intervention*, 2016, vol. 9901, pp. 487–495.

[34] B. Georgescu *et al.*, "Database-guided segmentation of anatomical structures with complex appearance," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, vol. 2, pp. 429–436.

[35] O. Ronneberger *et al.*, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2015, pp. 234–241.

[36] Y. Zhang *et al.*, "Coarse-to-fine stacked fully convolutional nets for lymph node segmentation in ultrasound images," in *Proc. IEEE Int. Conf. Bioinform. Biomed.*, 2016, pp. 443–448.

[37] C.-Y. Lee *et al.*, "Deeply-supervised nets," *Artif. Intell. Statist.*, pp. 562–570, 2015.

[38] L. Wang *et al.*, "Training deeper convolutional networks with deep supervision," 2015. [Online]. Available: https://arxiv.org/abs/1505.02496.

[39] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1395–1403.

[40] P. A. Miranda *et al.*, "Object delineation by κ-connected components," *EURASIP J. Adv. Signal Process.*, vol. 2008, pp. 1–14, 2008.

[41] H. Chen *et al.*, "Deep contextual networks for neuronal structure segmentation." in *Proc. 13th AAAI Conf. Artif. Intell.*, 2016, pp. 1167–1173.

[42] H. Chen *et al.*, "DCAN: Deep contour-aware networks for object instance segmentation from histology images," *Med. Image. Anal.*, vol. 36, pp. 135–146, 2017.

[43] M. I. Daoud *et al.*, "Accurate and fully automatic segmentation of breast ultrasound images by combining image boundary and region information," in *Proc. 13th IEEE Int. Symp. Biomed. Imag.*, 2016, pp. 718–721.

[44] L.-C. Chen *et al.*, "Semantic image segmentation with task-specific edge detection using CNNs and a discriminatively trained domain transform," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 4545–4554.

[45] I. Kokkinos, "Pushing the boundaries of boundary detection using deep learning," in *Proc. Int. Conf. Learn. Represent.*, 2016. [Online]. Available: https://arxiv.org/pdf/1511.07386.pdf

[46] F. Milletari *et al.*, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis.*, 2016, pp. 565–571.

[47] S. Balocco *et al.*,"Standardized evaluation methodology and reference database for evaluating IVUS image segmentation," *Comput. Med. Imag. Graph.*, vol. 38, no. 2, pp. 70–90, 2014.

[48] X. Liang *et al.*, "Recognizing focal liver lesions in contrast-enhanced ultrasound with discriminatively trained spatio-temporal model," in *Proc. 11th IEEE Int. Symp. Biomed. Imag.*, 2014, pp. 1184–1187.

[49] X. Liang *et al.*, "Recognizing focal liver lesions in CEUS with dynamically trained latent structured models," *IEEE Trans. Med. Imag.*, vol. 35, no. 3, pp. 713–727, Mar. 2016.

[50] G. Ghiasi and C. C. Fowlkes, "Laplacian pyramid reconstruction and refinement for semantic segmentation," in *Proc. Euro. Conf. Comput. Vis.*, 2016, pp. 519–534.

[51] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.

[52] J. Bergstra *et al.*, "Theano: A CPU and GPU math compiler in python," *in Proc. 9th Python Sci. Conf.*, 2010, pp. 1–7.

[53] W. Wang *et al.*, "An automatic energy-based region growing method for ultrasound image segmentation," in *Proc. IEEE Int. Conf. Image Process.*, 2015, pp. 1553–1557.

[54] J. Duchi *et al.*, "Adaptive subgradient methods for online learning and stochastic optimization," *J. Mach. Learn. Res.*, vol. 12, pp. 2121–2159, 2011.

[55] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, 2015.