Liver Segmentation in Magnetic Resonance Imaging via Mean Shape Fitting with Fully Convolutional Neural Networks

Qi Zeng[1(✉)], Davood Karimi[1], Emily H. T. Pang[2], Shahed Mohammed[1], Caitlin Schneider[1], Mohammad Honarvar[1], and Septimiu E. Salcudean[1]

[1] Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada
qizeng@ece.ubc.ca
[2] Vancouver General Hospital, Vancouver, BC, Canada

Abstract. In this work, we propose a novel learning-based segmentation technique for delineating liver volumes in magnetic resonance images. The method utilizes the shape prior of the liver for improved accuracy. Instead of labeling the tissue via binary classification, our method completes the segmentation by deforming a label template of the liver average shape based on the learned image features. The average shape of the liver we used is estimated from a large set of expert-labeled computed tomography images. A fully convolutional neural network (FCN) is trained to maximize the overlap between the deformed liver label template and the ground truth segmentation. The proposed method is validated with 51 T2-weighted liver image volumes and achieves an average Dice coefficient of 95.2% with a mean Hausdorff distance of 20.0 mm. Compared to the results obtained with a standard FCN-based method, a three-fold improvement of the Hausdorff distance is observed, indicating the substantial gains achieved by incorporating the shape prior.

1 Introduction

Over the last two decades, magnetic resonance imaging (MRI) has became a standard tool for the diagnosis of chronic liver diseases. Quantitative imaging techniques, such as magnetic resonance elastography (MRE) and proton density fat fraction (PDFF) have shown their potential to be the non-invasive gold standards for assessing hepatic fibrosis and steatosis [14]. Segmenting the liver in images with relatively high soft tissue contrast, such as T1 or T2-weighted MRI, is essential for analyzing the MRE and PDFF imaging data. The segmentation will be used to guide diagnosis in fused multi-parametric images. Figure 1 shows an example of patient data where the segmented T2-weighted image of the right liver lobe was used as background for MRE and PDFF overlays. In addition, automated liver segmentation has the potential to streamline pre-treatment

D. Shen et al. (Eds.): MICCAI 2019, LNCS 11765, pp. 246–254, 2019.
https://doi.org/10.1007/978-3-030-32245-8_28

planning for liver oncologic surgery or transplantation, to provide a more accurate means for diagnosing hepatomegaly or lobar redistribution, and to facilitate the follow-up of patients undergoing portal vein embolization.

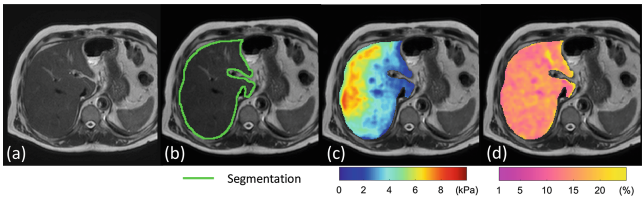

Fig. 1. Multi-parametric liver MRI example. (a) a T2-weighted slice, (b) liver tissue segmentation, (c) the MRE overlay presents liver tissue shear modulus E (unit is Kilopascal (kPa)), and (d) PDFF overlay presents liver tissue fat to water fraction (%).

Deep learning-based methods using fully convolutional neural networks (FCNs) represent the state of the art in medical image segmentation. These methods are effective in learning the complex mapping from the image domain to the segmentation domain using a cascade of convolutions, up-sampling, and down-sampling operations [9,16]. Existing FCN-based techniques commonly take the image intensity as the only source of information, ignoring other prior knowledge of the shape or size of the organ of interest. Recently, Milletari *et al.* proposed a FCN-based technique that utilizes such prior knowledge to segment 2D cardiac ultrasound images [10]. The idea was to predict the coefficients of a principal component analysis (PCA)-based shape model from the learned features to carry out the segmentation. Because this is a difficult task, the FCN architecture included a separate branch to refine the segmentation by focusing on patches around the predicted key-points. Al Arif *et al.* applied a similar shape-aware FCN technique to segment lateral cervical X-ray images and proposed to map the shape prior of cervical vertebra with signed distance functions [1]. Karimi *et al.* later developed a similar end-to-end shape model-based FCN technique to segment prostate in 3D MRI [6]. This study demonstrated the challenges of achieving FCN's convergence when predicting the coefficients of a 3D shape model and reported major difficulties in improving the segmentation performance. Moreover, the above studies focused mostly on segmenting relatively simple objects which have limited inter-subject variability. For 3D segmentation of large and complex organs, such as the liver, similar techniques might yield conservative results.

In this work, we propose a new way of incorporating shape prior in FCN based segmentation techniques, and we demonstrate the performance of our proposed method for liver segmentation in T2-weighted MRI images. The shape prior of the liver is extracted from a set of images with expert-provided liver segmentation as the estimated average liver shape. Our method produces the liver segmentation of a test image by deforming this average shape. A FCN is

trained to estimate the deformation mapping between the average liver shape and the test image volume in the form of a dense deformation field (DDF). Opting for such a free-form deformation ensures that the model has sufficient flexibility to accommodate complex shapes and to better tackle the challenging regions such as the liver left lobe and the inferior right lobe where the image contrast is low and inter-subject variability is high.

2 Materials and Methods

Data Preparation. The data used in this work consists of 51 T2-weighted MRI scans of 15 healthy and 10 patient volunteers who participated in a liver MRI study approved by the University of British Columbia Clinical Research Ethics Board. Images were acquired with an Achieva 3T scanner (Philips Inc., Best, Netherlands) using a T2-weighted turbo spin echo (T2W-TSE) sequence with a SENSE Torso XL receiving coil. Details of the scan settings are as follows: TR = 1400 ms, TE = 80 ms, and flip angle = 90°. The reconstructed transverse slices were $432 \times 432 \times 25$ with a voxel size of $0.7 \times 0.7 \times 6\,\text{mm}^3$. Preprocessing steps included: (1) N4 bias field correction [15], and (2) re-sampling to obtain isotropic voxels of size $1.4 \times 1.4 \times 1.4\,\text{mm}^3$, resulting in an image size of $224 \times 224 \times 96$. For each volume, the liver was first manually segmented by an experienced research assistant. Then, a clinical radiologist edited this segmentation to obtain the ground truth. 40 out of the 51 image volumes (80%) were randomly selected for the training and cross-validation of the proposed method. The remaining 11 volumes (20%) were used as test data and remained unseen to the training and parameter tuning.

Segmentation with Mean Shape Fitting. Consider a training dataset consisting of N images $I = \{I_i\}_{i=1}^N$ and corresponding ground-truth segmentation masks $Y = \{Y_i\}_{i=1}^N$, where $I_i \in \mathbb{R}^3 \rightarrow \mathbb{R}$ and $Y_i \in \mathbb{R}^3 \rightarrow \{0, 1\}$, where 0 denotes the background and 1 denotes the foreground (liver). Our proposed segmentation method is based on deforming a "mean liver shape", in our case a binary label template, denoted with Z, to the desired segmentation of the image at hand.

The mean shape used in this work was created using the computed tomography (CT) liver images publicly available through the liver tumor segmentation challenge (LiTS) [2]. Liver surface meshes of the 131 volumes in this dataset were extracted as an atlas of the liver shape. We aligned the atlas and estimated the mean shape (Z) using a group-wise non-rigid point cloud registration algorithm [12]. The computed mean shape surface mesh was converted into a binary label mask as shown in Fig. 2 where the grid size and spacing were set to be the same as in our dataset described above.

In our proposed method, predicting the segmentation of an image amounts to estimating a "best fit" deformation field that deforms the liver mean shape to the liver tissue volume in the image. To achieve this, we train our FCN to estimate a 3D displacement vector for each voxel of the liver mean shape template, denoted by $u_i \in \mathbb{R}^3 \rightarrow \mathbb{R}^3$. The resulting free-form dense deformation field (DDF) gives

the model complete flexibility in representing both local and global non-rigid deformations. To ensure efficient gradient back-propagation for FCN training, our DDF is implemented by linear operations such as grid warping and re-sampling as suggested in [4]. The estimated deformation vectors u_i are first added to the reference grid Cartesian coordinates of the mean shape template in $\mathbb{R}^3$. The deformed label mask is then computed via trilinear interpolation using the updated grid coordinates to produce the final segmentation mask.

Notice that our method does not use the commonly applied shape atlas representation, such as a statistical shape model (SSM). Instead, only the registration between the mean shape and the input image volume was considered. The main challenge of incorporating additional shape modes in our method is

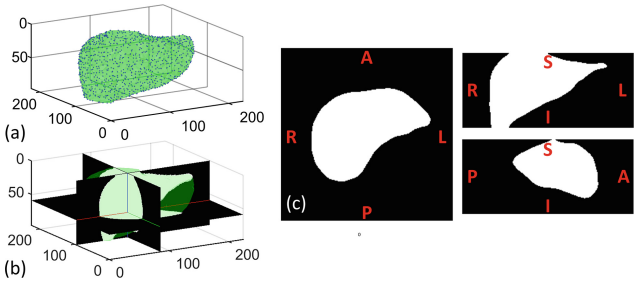

Fig. 2. Mean liver shape computed from the LiTS data: (a) the mean shape surface mesh (green) with the surface point cloud (blue), (b) the 3D orthogonal slices of the mean shape binary label template, and (c) the label template 2D views. (Color figure online)

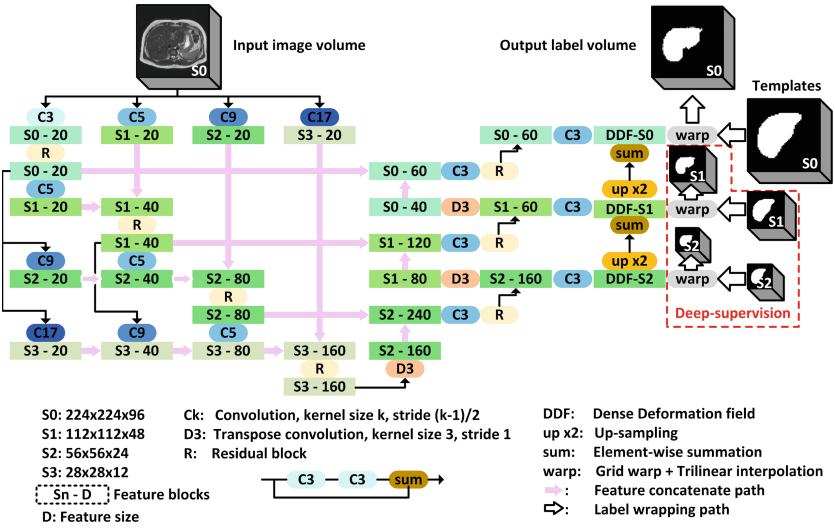

Fig. 3. Schematic representation of the FCN architecture used in this work. (Color figure online)

how to manage the convergence of the DDF for multiple modes in a consistent manner. Computing multiple DDFs for multiple shape modes is theoretically feasible, but hard to implement due to its high computational burden.

Network Design. Figure 3 shows the details of our FCN architecture. The overall design follows the V-net architecture [9] and consists of contracting and expanding paths of convolutional filters. In the contracting branch we compute features with different fields-of-view using 3D kernels of increasing sizes ($\{3^3, 5^3, 9^3, 17^3\}$) and strides ($\{1, 2, 4, 8\}$). This resulted in feature maps at four different scales $\{s0, s1, s2, s3\}$ with 20 features at each scale. Following the idea of feature reuse and dense connections proposed in [5], these feature maps are re-sized with additional convolution filters and forwarded to all coarser layers via concatenation paths. Residual blocks are also employed to improve the gradient back propagation [3]. In the expanding branch, feature maps are up-sampled via transpose-convolutions and concatenated with features from the contracting branch. At the scale of $\{s0, s1, s2\}$, feature maps go through a final convolutional layer to compute DDFs. Coarser DDFs are then up-sampled and added to the finer DDFs sequentially to improve the gradient flow. All convolution operations in this network are followed by a ReLU activation function [11].

Similarity Metric and Training Loss. The segmentation method performs the following loss function minimization:

$$\arg\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} -J(T_\theta(Z, I_i), Y_i) \; + \; \gamma ||\nabla u_i||^2 \tag{1}$$

where θ denotes the parameters of the FCN. The first term in the above loss function quantifies the similarity between the ground truth segmentation mask and the predicted segmentation mask, which is estimated by deforming the mean shape label template, Z. This deformation is denoted as $T_\theta(Z, I_i)$ because it is computed by the FCN as a function of the input image, I_i. For the similarity measure we use the Tversky metric [13]:

$$J = \frac{\mid TP \mid}{\mid TP \mid \; + \alpha \mid FP \mid \; + \beta \mid FN \mid} \tag{2}$$

As an extension to the Dice similarity coefficient (DSC), the Tversky metric allows control over the trade-off between false positive (FP) and false negative (FN) by adjusting parameters α and β. For our data, we found that the setting of $\alpha = 0.6$ and $\beta = 0.4$ helps to reduce the number of false positives caused by over-segmentation into the surrounding tissue. The second term in our loss function is the L^2 norm of the DDF's first spatial derivative. This is a regularization term that is necessary to encourage the DDF to be locally smooth and thus avoid excessive local deformations [4]. Such deformations can occur due to spurious image features or image artifacts at the isolated small regions. The regularization

parameter, γ, controls the trade-off between the two loss terms. Figure 4 shows how γ regularizes the DDF. The results reported in this paper were produced with $\gamma = 10^{-2}$, which we empirically found to lead to good results.

Training Strategy. We used 5-fold cross-validation to train an ensemble of 5 networks. The final segmentation mask is produced by averaging the results of the 5 networks. Each network was trained for 500 epochs using the Adam optimizer [7] with a batch size of 1. The initial learning rate was 10^{-5}, which was reduced by 5% after every 20 epochs. During training, three types of data augmentation were used: (1) B-spline local image deformation with an amplitude of 2 mm, (2) addition of Gaussian noise with a maximum amplitude of 10% of the average image intensity, (3) rigid image translation with a maximum of 5 mm in each of the x, y, z directions. As suggested by [8,16], deeply-supervised multi-scale training was introduced to our implementation as highlighted in Fig. 3 with red dashed lines. The lower-resolution ground-truth segmentation masks were generated by down-sampling the ground-truth at the original resolution.

3 Results and Discussion

We compare our method with the standard V-net [9] and report five results for our method: (1) Proposed-OneFCN: one FCN is trained without deep supervision to produce the DDF and the warped mean shape is considered as the final segmentation; (2) Proposed-OneFCN-DS: similar to (1), except that deep supervision at three different scales is used; (3) Proposed-Ensemble: five FCNs are

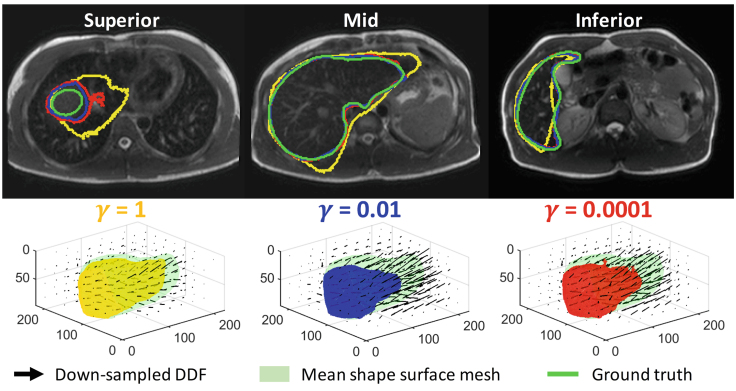

Fig. 4. An example of segmentations produced with $\gamma = \{1, 10^{-2}, 10^{-4}\}$. Top row: segmentation contours for the superior, mid, and inferior slices. Bottom row: 3D graphics to show how the label template is deformed. The DDF obtained with $\gamma = 1$ is too conservative causing the deformed template to fail to accurately delineate the liver boundary. The DDF with $\gamma = 10^{-4}$ is overly aggressive causing unexpected regions of over- and under-segmentation.

trained without deep supervision, the final segmentation is obtained by setting the threshold of the average label maps value produced by the five networks at 0.5; (4) Proposed-Ensemble-DS: similar to (3) except that networks are trained with deep supervision; and (5) Proposed-Baseline: we separately trained our FCN to generate the segmentation via the conventional binary labeling approach. The results presented show the performance of a classification approach with a more up-to-date FCN implementation than the V-net. Our evaluation criteria include Jaccard, Dice, false positive (FP), false negative (FN), Hausdorff Distance (HD), and mean surface distance (MSD).

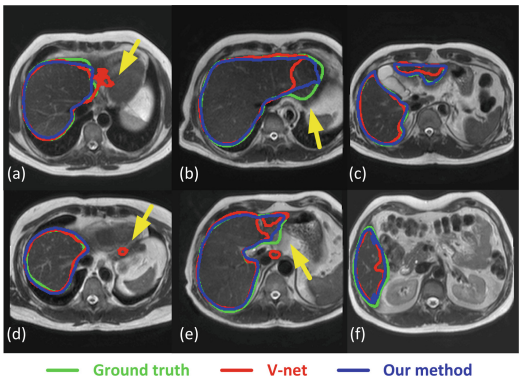

Fig. 5. Examples of results produced by our proposed method and V-net.

Table 1 summarizes the performance of the different methods. Figure 5 presents examples of image slices segmented by V-net and the Proposed-Ensemble technique. From the results, our method outperformed V-net in terms of all performance criteria. In terms of volume labeling accuracy, our method with network ensemble was able to achieved a mean Dice of 95.2%. A paired t-test showed a p-value of 0.003 when comparing the Dice scores achieved by our method with that of V-net. Our method also achieved improvements in terms of reduced surface distance errors. Compared with V-net the HD reported from our method was significantly smaller ($p = 0.04$). The conservative performance

Table 1. The comparison of the proposed method with V-net at different stages

Metrics	Jaccard (%)	DICE (%)	FN (%)	FP(%)	HD (mm)	MSD (mm)
V-net	86.9 ± 2.6	92.9 ± 1.5	3.7 ± 2.3	9.9 ± 2.6	60.3 ± 50.3	1.4 ± 0.3
Proposed-Baseline	88.4 ± 1.9	93.8 ± 1.0	3.2 ± 4.5	4.6 ± 2.7	63.8 ± 18.2	1.5 ± 0.3
Proposed-OneFCN	90.7 ± 1.0	95.1 ± 0.5	3.8 ± 2.8	5.6 ± 3.0	20.2 ± 9.8	1.0 ± 0.2
Proposed-OneFCN-DS	90.4 ± 2.1	94.9 ± 1.2	4.9 ± 4.2	4.9 ± 2.5	24.6 ± 10.2	1.1 ± 0.3
Proposed-ENS	$\mathbf{90.9 \pm 1.8}$	$\mathbf{95.2 \pm 1.0}$	$\mathbf{3.3 \pm 2.7}$	6.0 ± 3.0	$\mathbf{20.0 \pm 12.1}$	$\mathbf{0.9 \pm 0.2}$
Proposed-ENS-DS	90.9 ± 2.4	95.2 ± 1.3	5.5 ± 4.1	$\mathbf{3.8 \pm 2.4}$	23.8 ± 13.9	1.0 ± 0.3

of V-net in terms of a higher HD is due to its inconsistent results at the superior left lobe and the inferior right lobe as show on Fig. 5(a), (d) and (c). In comparison, our method was able to better avoid large segmentation errors, because the mean shape prior provided a good initial segmentation at these regions. When our proposed FCN structure was trained based on the conventional classification approach, it only achieved marginal improvements on the volume overlapping metrics, while no gain on surface distance metrics was observed. This demonstrates again the advantage of using the shape prior in further regulating the segmentation surface distance errors.

Although the use of Tversky metric allowed a higher penalty on the FP, a high FP rate (>5%) was still reported in all cases, indicating that accurately isolating the liver from surrounding tissue background is still a relatively challenging task for learning-based methods. It is also interesting to see that the overall performance of our method had no significant gain from introducing deep-supervision of the network training, except it might have helped to better regulate FP.

The purposed method was implemented using Tensorflow. With an NVIDIA TITAN RTX GPU, the training for 500 echos took approximately 24 h. For a test image volume, each FCN produces a segmentation in 0.8 s.

4 Conclusion

In the context of liver segmentation in 3D MRI T2-weighted image volumes, we proposed a new learning based method that utilizes prior shape information and non-rigid registration to improve the segmentation accuracy. Our method achieved significantly better results than the competing binary classification based method in terms of Dice and HD. Particularly, our method substantially reduced the maximum surface distance errors in the most challenging regions such as the superior left liver lobe and the inferior right lobe. Our technique can be easily extended to segment other more complicated organs when a good image atlas is available. A more detailed comparison study between the proposed method and some of the existing SSM-based techniques is warranted in a future study.

Acknowledgment. This project is funded by Natural Sciences and Engineering Research Council of Canada (NSERC). We deeply appreciate the support from the Charles A. Laszlo Chair in Biomedical Engineering held by Prof. Salcudean.

References

1. Al Arif, S.M.M.R., Knapp, K., Slabaugh, G.: SPNet: shape prediction using a fully convolutional neural network. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11070, pp. 430–439. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_49
2. Bilic, P., et al.: The liver tumor segmentation benchmark (LiTS). arXiv preprint arXiv:1901.04056 [cs.CV] (2019)

3. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE CVPR, pp. 770–778 (2016)
4. Hu, Y., et al.: Adversarial deformation regularization for training image registration neural networks. In: Frangi, A.F., et al. (eds.) MICCAI 2018. LNCS, vol. 11070, pp. 774–782. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_87
5. Huang, G., Liu, Z., ven der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: IEEE CVPR, pp. 2261–2269 (2017)
6. Karimi, D., et al.: Prostate segmentation in MRI using a convolutional neural network architecture and training strategy based on statistical shape models. Int. J. Comput. Assist. Radiol. Surg. **13**(8), 1211–1219 (2018)
7. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 [cs.LG] (2014)
8. Lee, C.Y., Xie, S., Gallagher, P., Zhang, Z., Tu, Z.: Deeply-supervised nets. In: PMLR, vol. 38, pp. 562–570 (2015)
9. Milletari, F., Navab, N., Ahmadi, S.: V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571 (2016)
10. Milletari, F., et al.: Integrating statistical prior knowledge into convolutional neural networks. In: Descoteaux, M., et al. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 161–168. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_19
11. Nair, V., Hinton, G.E.: Rectified linear units improve restricted Boltzmann machines. In: ICML, pp. 807–814 (2010)
12. Rasoulian, A., Rohling, R., Abolmaesumi, P.: Group-wise registration of point sets for statistical shape models. IEEE Trans. Med. Imag. **31**(11), 2025–2034 (2012)
13. Salehi, S.S.M., Erdogmus, D., Gholipour, A.: Tversky loss function for image segmentation using 3D fully convolutional deep networks. In: Wang, Q., et al. (eds.) MLMI 2017. LNCS, vol. 10541, pp. 379–387. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67389-9_44
14. Taouli, B., Ehman, R.L., Reeder, S.B.: Advanced MRI methods for assessment of chronic liver disease. AJR Am. J. Roentgenol. **193**(1), 14–27 (2009)
15. Tustison, N.J., et al.: N4ITK: Improved N3 bias correction. IEEE Trans. Med. Imag. **29**(6), 1310–1320 (2010)
16. Yang, D., et al.: Automatic liver segmentation using an adversarial image-to-image network. In: Descoteaux, M., et al. (eds.) MICCAI 2017. LNCS, vol. 10435, pp. 507–515. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_58