2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)
April 4-7, 2018, Washington, D.C., USA

DEEP MULTI-TASK AND TASK-SPECIFIC FEATURE LEARNING NETWORK FOR ROBUST SHAPE PRESERVED ORGAN SEGMENTATION

Chaowei Tan[1], *Liang Zhao*[2], *Zhennan Yan*[2], *Kang Li*[1,3], *Dimitris Metaxas*[1], *Yiqiang Zhan*[2]

[1] Department of Computer Science, Rutgers University, Piscataway, NJ, USA
[2] Siemens Healthcare, Malvern, PA, USA
[3] Department of Industrial and Systems Engineering, Rutgers University, Piscataway, NJ, USA

ABSTRACT

Fully convolutional network (FCN) has shown potency in segmenting heterogeneous objects from natural images with high run-time efficiency. This technique, however, is not able to produce continuous, smooth and shape-preserved regions consistently due to complex organ structures and occasional weak appearance information commonly observed in various anatomical structures in medical images. In this paper, we propose a deep end-to-end network with two task-specific branches to ensure continuousness, smoothness and shape-preservation in segmented structure without additionally sophisticated shape adjustment, e.g., dense conditional random fields. The novelties of the proposed method lie in three aspects. First, we formulate the organ segmentation as a multi-task learning process that combines both region and boundary identification tasks, which can alleviate spatially isolated segmentation errors. Second, we use boundary distance regression to ensure the smoothness of the segmented contours, instead of formulating boundary identification as a classification problem [1]. Third, our deep network is designed to have a "Y" shape, i.e., the first half of the network is shared by both region and boundary identification tasks, while the second half is branched for each task independently. This architecture enables the task-specific feature learning for better region and boundary identification, and offers information for segmentation refinement based on a fusion scheme using energy functional. Extensive evaluations are conducted on a variety of applications across organs and modalities, e.g., MR femur, CT kidney, etc. Our proposed method shows better performance compared to the state-of-the-art methods.

Index Terms— Deep end-to-end network, multi-task and task-specific learning, shape preserved organ segmentation.

1. INTRODUCTION

Fully convolutional network (FCN) [2] has been highlighted as a fundamental segmentation approach for anatomy delineation in medical images. It exploits the deep convolutional neural networks (DCNN) for coarse-to-fine inference and makes a prediction at every pixel. Without manually setting handcrafted features, DCNN has the ability to learn a hierarchical representation of raw input data. However, FCN is limited for lower-level tasks requiring precise localization, e.g., semantic segmentation, since the DCNN-based inferences inside FCN build invariance to spatial transformations and provide only abstraction of spatial details. In Fig. 1, FCN is more likely to have predicted outliers due to the high variability of organic shapes and low-contrast imaging quality in medical images. It may not produce a continuous segmented object with smooth boundary.

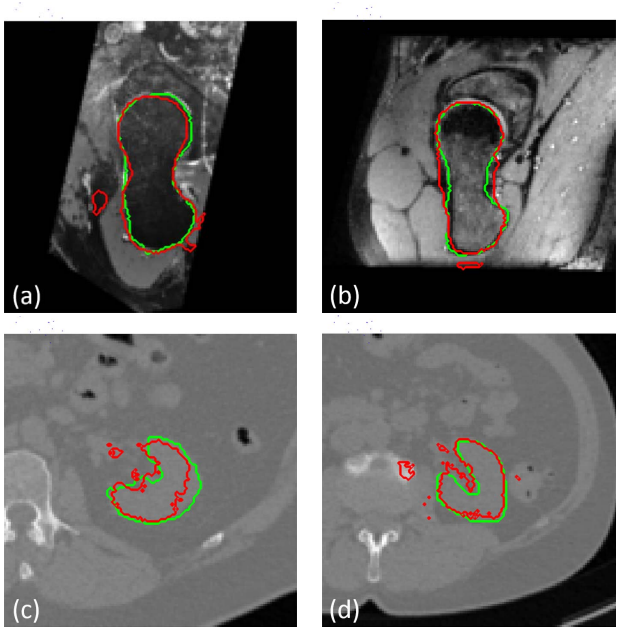

Fig. 1. (a)-(d) show the ground truth (green lines) and segmented contours (red) using FCN method for two MR bone and two CT kidney cases. The FCN-based results exist spatially isolated errors and smoothless segmented boundaries.

Recent researches employ a new strategy called multi-task learning for organ segmentation. The main task of this strategy is to optimize target extraction by leveraging auxiliary information from a set of correlated tasks (e.g., background

978-1-5386-3636-7/18/$31.00 ©2018 IEEE

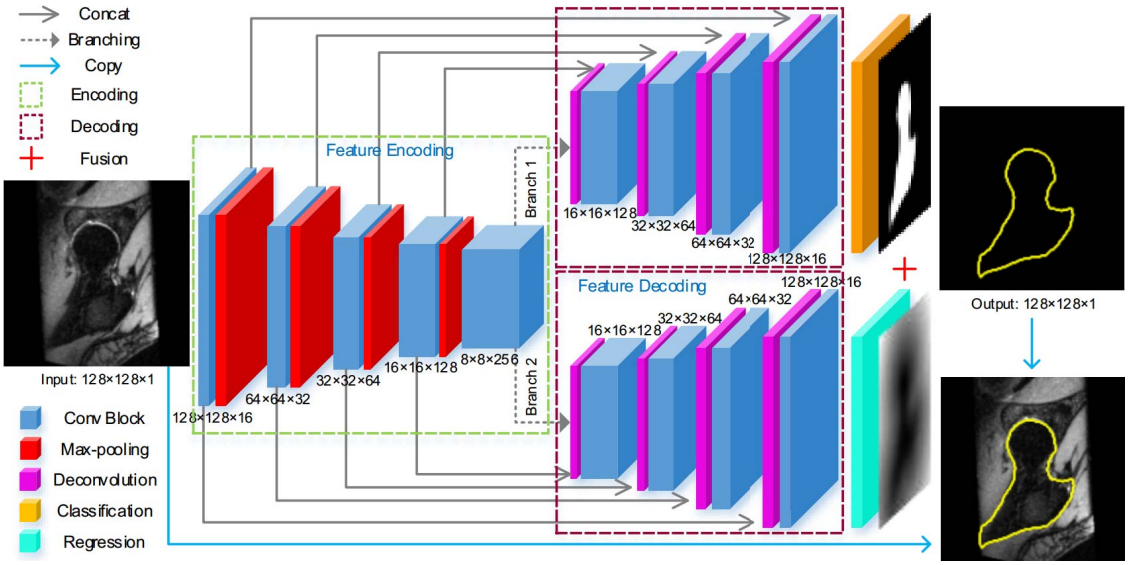

Fig. 2. Proposed deep multi-task and task-specific feature learning network. The left half is encoding for the abstraction of multi-level contextual feature. The right half has two decoding branches for the task-specific learning. Each blue convolution block includes two convolutional layers with filter size of 3×3 and zero-padding of 1. Parametric rectified linear unit and batch normalization are also adopted in all convolutional and deconvolutional layers.

classification or bounding box investigation). Chen et al. [1] studied a multi-task deep representation that combined region and boundary classifications to obtain continuous tissue description. In order to alleviate the spatially isolated segmentation errors in Fig. 1, we formulate the organ segmentation as a multi-task network consisting of two parallel end-to-end branches, as shown in Fig. 2. Each task involves two symmetrical parts, i.e., encoding and decoding. The first task is a conventional fully convolutional network for the inference of organ probability map. The second task is a novel deep regression network (DRN) that regresses the distance constraint information of organic boundary. During the distance regressing, it produces continuously numerical constraints of spatial information. Thus the DRN obtains better potential to preserve the smoothness of boundary, comparing with the discrete classified labels obtained by boundary classification.

Gao et al. [3] proposed a boundary extractor by learning a joint objective function from displacement estimation and organ classification, and these two tasks only share parameters in the final loss. But in Fig. 2, the two task-specific branches in the proposed network share parameters in the encoding process in the first half of Y-shape network, while having their own decoding parameters to represent the features for the classification and regression, respectively. This structure ensures a balanced and sufficient parameters learning to represent the task-specific features for region and boundary identification. During the training, the cumulative loss is optimized by jointly investigating the two losses from each single branch. Adding the distance regression task can effectively regularize the smoothness of segmented boundary and reduce the isolated segmentation errors. Simultaneously, the classification branch is efficient for locating and extracting target. Finally, the proposed method also explores a unified segmentation architecture which incorporates a shape refinement. We combine the inferred organ probability map and the regressed boundary distance map based on a fusion scheme using energy functional. This scheme can efficiently refine organ boundary, and avoid the complex parameter tuning of additionally sophisticated shape adjustment (e.g., dense conditional random fields [4]).

2. METHODOLOGY

2.1. Deep end-to-end network with multi-task learning

In this section, we present a deep end-to-end network branched by two task-specific learning for the organ segmentation. As shown in Fig. 2, the network takes the entire 2D image as input, and the first task is a conventional FCN for the inference of organ probability map, and the second one is the novel DRN regressing the distance constraint information of organic boundary. The main structure of each task is designed as a symmetric way, i.e., encoding-decoding. The two branches share the encoding part, which contains 4 max-pooling layers with stride 2 to obtain 4 different resolutions of raw image. Under each resolution, two convolutional layers are utilized for feature abstraction. This successive encoding allows to obtain multi-size contextual information which is helpful to receive the integral interior structure of tissue and the sufficient background knowledge surrounding it, and then improve the recognition performance. In the decoding part of

each task, we deploy 4 deconvolutional layers in a cascaded way for up-sampling feature maps. Each deconvolutional layer adopts stride 2, so it avoids the usage of large up-sampling factors (16, or 32), and effectively reduces computation and details missing in deconvolution. These deconvolutional operations restore input image's resolution from lower to higher, and finally reach the original size. Each upscaling operation also follows two convolutional layers, playing the same role to abstract features. The design of the proposed network firstly makes the encoder extract high-level abstraction features, and then the two decoders acquire pixel-wise organ probability map and boundary distance map, respectively.

The deep contour-aware network [1] makes its boundary identification branch as classification. This approach provides strong boundary constraints, and can increase the overall segmentation accuracy and reduce spatially isolated errors. However, the classification-based auxiliary task offers discrete boundary labels, which may cause some non-smooth segmented contours. Here, we consider boundary distance regression as a reasonable criteria to produce continuously numerical constraints of spatial information. So we build the boundary identification task by the distance regression, and this criteria performs as complementary cues to the probability map of organ classification, and then it regularizes the trouble of classification outliers.

We define the loss of the classification branch L_{cls} by applying multi-class cross entropy loss to each pixel of the output probability map. In the regression branch, we formulate the loss term L_{dis} based on the following loss:

$$L_{dis} = \frac{1}{2K} \sum_{i=1}^{K} \sum_{\mathbf{x} \in \Omega} w\left(\mathbf{x}\right) \left\| \hat{D}_i\left(\mathbf{x}\right) - D_i\left(\mathbf{x}\right) \right\|_2^2 \qquad (1)$$

where K is the number of classes. For the i-th class, $\hat{D}_i\left(\mathbf{x}\right)$ and $D_i\left(\mathbf{x}\right)$ are the predicted and the ground truth distance maps at pixel location $\mathbf{x} \in \Omega$, where Ω is the image space, and $\subset Z^2$. w is a weight function that gives higher penalty weight to the pixels that are farther away the organ border.

Thus, the objective function of the network is as follows:

$$\arg\min_{\{\mathbf{W}_t\}_{t=1}^{T}} \left\{ L_{cls} + \alpha L_{dis} + \beta \sum_{t=1}^{T} \|\mathbf{W}_t\|_2^2 \right\} \qquad (2)$$

where α and β are the balance weights. T represents the number of tasks. $\mathbf{W}_\mathrm{t}$ denotes the parameters of t-th task.

Based on the proposed multi-task network and losses, the objective function will not suffer from the non-trivial solving issue mentioned in [5]. We can easily derive a standard solution for the novel ℓ_2-based regression loss. Moreover, except the two loss layers in the two branches, the rest layers use the same decoding network design, and thus Eq. 2 holds low model complexity and can be effectively solved through, for instance, the stochastic gradient descent (SGD) solver.

2.2. Energy functional based fusion scheme

A further refinement could fuse the predicted organ probability map $\hat{P}$ and boundary distance map $\hat{D}$ to obtain the final segmentation. Here, the fusion scheme minimizes an energy functional $F(\hat{P}, \hat{D})$ based on the Chan-Vese model [6]. The scheme treats $\hat{P}$ and $\hat{D}$ as the optimization target and signed distance function, respectively. Because $\hat{D}$ has already been very close to the real organ, it can be used as a good and straightforward initialization for the fusion. With few iterations for an image, the fusion will finish in 0.02s.

3. EXPERIMENTS

Experimental settings. We validate the proposed method on three datasets. The first synthetic dataset includes 6000 2D simulated images (training: 4000, validating: 1000, testing: 1000). We use 3 types of geometry elements (circle, triangle and square) to construct the toy examples. The target for segmentation is combined by a circle and a triangle, which is initially located in the center of image with a stochastic offset in the x and y directions. In order to simulate complex varieties of shape, the angle, length and direction of the target are randomly set. Each toy image also includes some interferences by randomly placing several squares and circles with various sizes surrounding the target. Meanwhile, heavy Gaussian noises are added to blur all shapes. The second dataset includes 2304 2D magnetic resonance (MR) images (training: 1368, validating: 468, testing: 468). They are from 64 3D MR femur scans with voxel spacing (1mm, 1mm, 1mm). On each 3D femur volume, we rotationally sample 36 slices around the femoral shaft, with 5 degrees interval angle. The third set is built by 107 computed tomography (CT) 3D kidney images. Each kidney image is resampled and cropped, and has the same physical size (20cm$\times$20cm$\times$15cm) with voxel spacing (1mm, 1mm, 1mm). Along the axial direction of each kidney data, we totally sample 16050 2D images with 1mm interval distance (training: 9000, validating: 3000, testing: 4050). Patients randomly used in dataset 2 and 3 are independent from others. All the images are resized to 128×128, and their pixel intensity is linearly normalized in [0, 1].

Two state-of-the-art medical segmentation approaches are evaluated with our method. One is the U-net [7], and the second is the deep contour-aware networks for accurate gland segmentation (DCAN [1]). For validation, dice similarity coefficient (DSC $= \frac{2TP}{2TP+FP+FN}$) and relative error (RE $= \frac{FP+FN}{TP+FN}$) between the ground truth (GT) labels and segmentation results are reported. TP, FP and FN are the number of pixels correctly identified, incorrectly identified and incorrectly rejected respectively. The mini-batch is employed in the training phase, and its size is set to around 80 for each training of the compared methods. We use a momentum of 0.9 and a learning rate initially set as 0.001 (multiplied by a factor of 0.95 every 10,000 iterations).

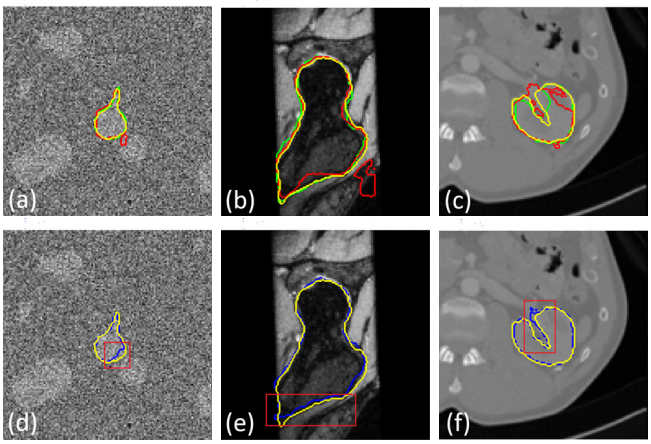

Fig. 3. 2D visual comparisons for simulated data, femur bone and kidney. Green, red, blue and yellow lines are for the GT, U-net, DCAN and proposed method, respectively.

Table 1. Quantitative comparisons.

Method	Simulation		Femur Bone		Kidney	
	DSC	*RE*	*DSC*	*RE*	*DSC*	*RE*
U-net	0.90	0.19	0.90	0.19	0.70	0.62
DCAN	0.92	0.15	0.91	0.18	0.83	0.31
Ours	**0.96**	**0.09**	**0.93**	**0.14**	**0.90**	**0.19**

Experimental results. Fig. 3 (a)-(c) show visual comparisons between the proposed method and the U-net model for three cases. Shown by these, the U-net model can locate the position of organ (or target) correctly. Its segmented regions capture most of the correct tissue areas, and get good quantitative measurements. Yet its results suffer from leakages to the surrounding areas, where have similar pixel intensity to the targets. On the other hand, the classification branch of our method is the same as the U-net, but with the additional regression branch and the joint training, the proposed approach can prevent the leakage issue, and thus obtain better total segmentation performance. Since the few visual comparisons may not reflect the overall performance clearly, quantitative comparisons of overlapping accuracies are shown in Table 1.

After showing the effectiveness of the proposed model to prevent leakages, we also visually compare the boundary smoothness of results by the DCAN and proposed method. In order to have a better view, only the segmented contours by the two approaches are plotted for the same cases in Fig. 3 (d)-(f). The DCAN method does not show significant leakage problem, and obtains higher total segmentation accuracy comparing with the U-net in Table 1. However, by considering the boundary smoothness shown in the red boxes, the proposed method achieves better performance. The two methods both utilize multi-task strategy to preserve shape, but in the DCAN, its boundary classification task offers discrete boundary labels which may cause non-smooth boundary. In our regression branch, the boundary distance regression could provide continuously numerical constraints of spatial information during optimizing the regression loss. Hence the proposed method could produce higher smoothness on boundary. Besides the visual comparisons, the overall quantitative measurements between the two approaches are shown in Table 1.

4. CONCLUSION

In the present work, we propose a deep multi-task network for robust shape preserved organ segmentation. The network has a unified architecture to formulate organ segmentation as multi-task learning that combines both region and boundary identification. This multi-task learning with the novel boundary distance regression can alleviate spatially isolated segmentation errors as well as ensure the smoothness of segmented contours. The proposed deep network is designed as a Y shape, bifurcated at the end of the encoding path. Hence the shared encoding and non-shared decoding paths have balanced layers and parameters for each task branch.

5. REFERENCES

[1] H. Chen, X. Qi, L. Yu, and P.-A. Heng, "Dcan: Deep contour-aware networks for accurate gland segmentation," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 2487–2496, 2016.

[2] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, 2015.

[3] Y. Gao, Y. Shao, J. Lian, A. Z. Wang, R. C. Chen, and D. Shen, "Accurate segmentation of ct male pelvic organs via regression-based deformable models and multi-task random forests," *IEEE transactions on medical imaging*, vol. 35, no. 6, pp. 1532–1543, 2016.

[4] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," *arXiv preprint arXiv:1412.7062*, 2014.

[5] I. Kokkinos, "Ubernet: Training auniversal'convolutional neural network for low-, mid-, and high-level vision using diverse datasets and limited memory," *arXiv preprint arXiv:1609.02132*, 2016.

[6] T. F. Chan and L. A. Vese, "Active contours without edges," *Image Processing, IEEE Transactions on*, vol. 10, no. 2, pp. 266–277, 2001.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Springer, 2015.