Paper Reading No.12

Shape-Aware Complementary-Task Learning for Multi-Organ Segmentation

Sheng Lian

August 2019

1 Brief Paper Intro

- ***Paper ref:*** MLMI workshop on MICCAI 2019, https://arxiv.org/abs/1908.05099
- ***Authors:*** See Fig 1.

Fernando Navarro[⋆1], Suprosanna Shit[⋆1], Ivan Ezhov[1], Johannes C. Paetzold[1], Andrei Gafita[3], Jan Peeken[2], Stephanie E. Combs[2], and Bjoern H. Menze[1]

[1] Department of Informatics, Technische Universität München, Germany
[2] Department of Radiotherapy, Klinikum rechts der Isar, Germany
[3] Department of Nuclear Medicine, Klinikum rechts der Isar, Germany
fernando.navarro@tum.de

Figure 1: authors' brief intro.

- ***Paper summary:*** For solving multi-organ segmentation task, authors proposed a complementary-task learning method utilizing different organs' shape prior. The shape prior is come from i) distance map regression and ii) contour map detection to explicitly encode the geometric properties of different organ.

- ***Reading motivation:***

  According to the ICLR paper [1], the CNNs are inherently biased towards texture information over the shape of an object (This paper is interesting that the title conclude all the key views of the author). However, human's organs have discriminative shape feature, this is not well addressed by conventional CNNs. In this solution, authors use a same network to accomplish different tasks, generate multi-model maps, and the different kinds of maps focus on different aspects of organ segmentation tasks. Let's see how this paper do.

2 Backgrounds

3 Contributions/highlights

-
-

4 Methods

In this paper, authors hold the hypothesis that learning shape-prior as a complementary-task improves the performance of the segmentation task. The two complementary tasks for shape prior learning in this paper is i) inferring geometric shape properties of an organ using distance map, and,

ii) detecting the exterior contour of an organ using binary edge-map of each organ.

The first part, distance transformation, aims to help the network to accurately localize the organ from the learned anatomical prior. While the second part, organ contour learning, penalizes for boundary miss-classiffcation.

The different kinds of tasks are shown in Fig.2.

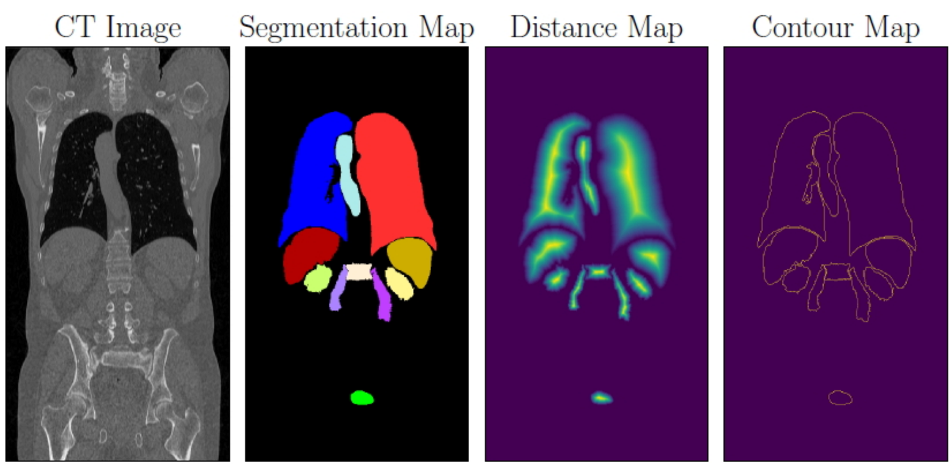

Figure 2: Target generation for complementary-tasks.

4.1 Learning distance-transform

This part can be regarded as a combination of Euclidean distance map and Gaussian heat-map regression. The distance transform of a shape is to calculate the distance of each non-zero point in the image from its nearest zero point. This part learns the geometric properties of shapes.

4.2 Learning organ contour

This part explicitly learn the organ contour as the second complementary-task alongside the distance map regression.

4.3 Loss function

The loss function consist of three parts from different task: segmentation, distance regression, and contour detection. We set $p_l(x)$ and $g_l(x)$ as the predicted probability and ground truth, in location x of class l, respectively.

Here, both the segmentation loss and contour loss can be defined as a combination of cross-entropy loss and Dice loss, which goes as:

$$\mathcal{L}_{seg}, \mathcal{L}_{contour} = \underbrace{-\sum_{x} g_l(x) \log p_l(x)}_{\text{Cross-Entropy Loss}} - \underbrace{\frac{2\sum_x p_l(x) g_l(x)}{\sum_x p_l^2(x) + \sum_x g_l^2(x)}}_{\text{Dice Loss}} \tag{1}$$

When the $l =$ number of organs $+1$, this equation means the segmentation loss. While when the $l = 2$, this equation means the contour loss.

For distance map, the loss function is

$$\mathcal{L}_{dist} = -\frac{1}{n} \sum_{x} (g(x) - p(x))^2 \tag{2}$$

where the $p(x)$ is the estimated distance map of pixel xl and $g(x)$ is the ground truth.

The overall loss function is

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{seg}} + \mathcal{L}_{\text{contour}} + \mathcal{L}_{\text{dist}} \tag{3}$$

4.4 Network

The overall architecture of the proposed network is shown in Fig. 3. The input is a CT slice. The network is similar U-Net model with three branches diverging at the end of the last up-convolution. The outputs of the network are the segmentation map, the distance map, and the contour map.

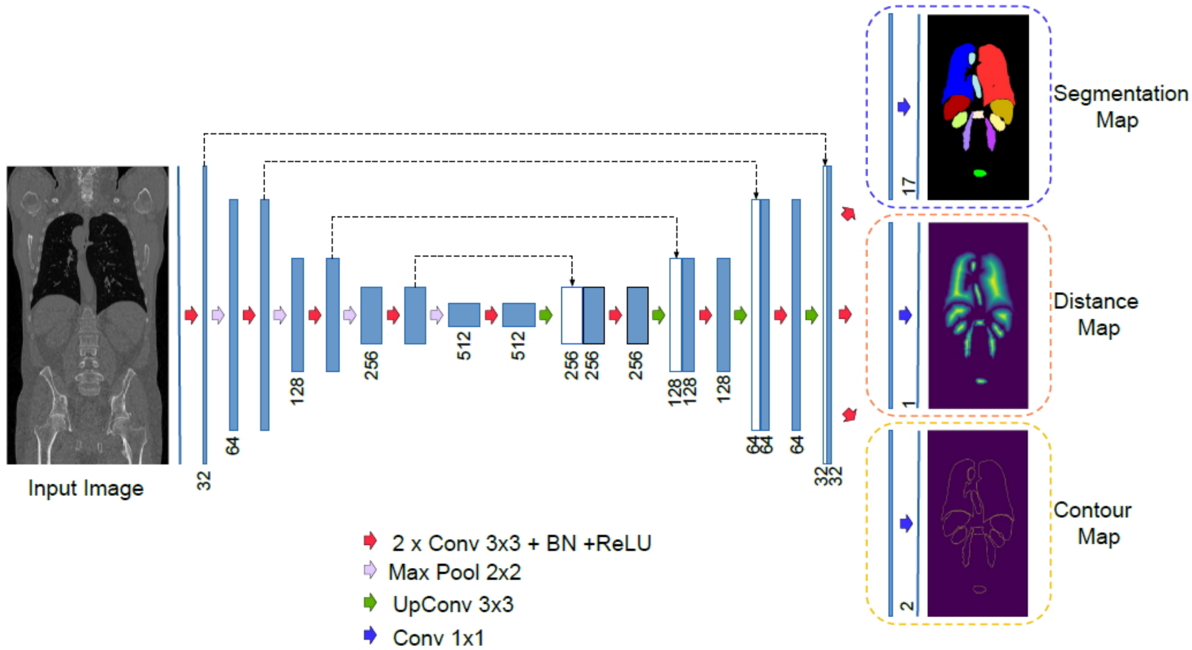

Figure 3: Network architecture for complementary-task learning.

4.5 Experiments

Authors implemented their model on VISCERAL dataset, which consists of 97 CT scans from gold corpus and silver corpus.

They performed four different experiments as Fig 4.

- U-Net: baseline using only $\mathcal{L}_{seg}$.
- U-Net + distance: $\mathcal{L}_{seg} + \mathcal{L}_{dist}$.
- U-Net + contour: $\mathcal{L}_{seg} + \mathcal{L}_{contour}$.
- U-Net + distance,contour: $\mathcal{L}_{seg} + \mathcal{L}_{dist} + \mathcal{L}_{contour}$

Figure 4: Four different experiments implemented in this paper for ablation study.

5 Results

See Fig 5. After adding distance task and contour task, the segmentation accuracy is improved in overall measuring and single organ measuring.

Table 1: Quantitative results: mean and standard deviation of the global dice scores shows that complementary-task learning achieves the best result.

Model	Dice
U-Net	0.8849 ± 0.120
U-Net + distance	0.8868 ± 0.116
U-Net + contour	0.8791 ± 0.118
U-Net + distance,contour	**0.9018 ± 0.116**

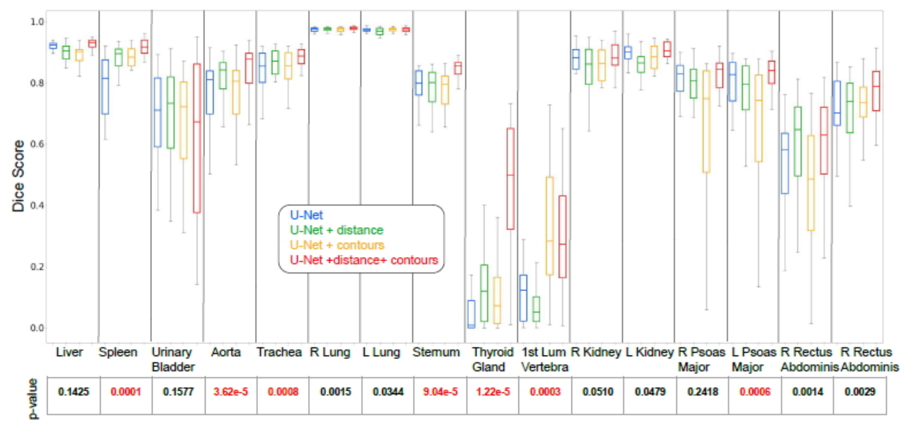

Fig. 4: Box plots of different organs show consistent improvement of the dice score using our proposed model (U-Net+distance+contour) over the baseline model(U-net). We also report the *p-value* obtained using Wilcoxon signed-rank statistical test between our proposed model and the baseline. The statistically significant *p-values* < 0.001 are shown in red.

Figure 5: Quantitative results and box plots of the proposed method.

6 My thoughts

- "Convolutional neural networks are inherently biased towards texture information over the shape of an object." according to [1], this theory is worth consideration when designing CNN network.

- The contour task part, according to this paper, is achieved using contour loss, which looks alike to the segmentation loss. Both of them are combination of cross-entropy loss and dice loss. In my opinion, this loss is essentially a loss of segmentation and is affected by the accuracy of the segmentation. This loss is just a switch from the multi-organ segmentation into a two-class loss. Whether it really uses the contour information, avoids the shortcomings of CNN? This requires strict proof and discussion.

References

[1] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A Wichmann, and Wieland Brendel. Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv preprint arXiv:1811.12231*, 2018.