Star Shape Prior in Fully Convolutional Networks for Skin Lesion Segmentation

2 authors, including:

Ghassan Hamarneh
Simon Fraser University
491 PUBLICATIONS   **6,338** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project   Diffusion MRI Analysis View project

Project   Connectome analysis of preterm infants View project

Star Shape Prior in Fully Convolutional Networks for Skin Lesion Segmentation

Zahra Mirikharaji, Ghassan Hamarneh

School of Computing Science, Simon Fraser University, Canada
{zmirikha, hamarneh}@sfu.ca

Abstract. Semantic segmentation is an important preliminary step towards automatic medical image interpretation. Recently deep convolutional neural networks have become the first choice for the task of pixelwise class prediction. While incorporating prior knowledge about the structure of target objects has proven effective in traditional energy-based segmentation approaches, there has not been a clear way for encoding prior knowledge into deep learning frameworks. In this work, we propose a new loss term that encodes the star shape prior into the loss function of an end-to-end trainable fully convolutional network (FCN) framework. We penalize non-star shape segments in FCN prediction maps to guarantee a global structure in segmentation results. Our experiments demonstrate the advantage of regularizing FCN parameters by the star shape prior and our results on the ISBI 2017 skin segmentation challenge data set achieve the first rank in the segmentation task among 21 participating teams.

1 Introduction

Skin cancer is the most common type of cancer in the world. Early detection of skin cancer can increase the five year survival rate of patients from 18% to 98% [1]. While skin cancer can be detected by visual examination, distinguishing malignant from non-malignant lesions is a challenging task. In recent years, computer aided diagnosis has been widely leveraged in automated assessment of dermoscopy and clinical images to assist dermatologists evaluation. Semantic segmentation, the task of labeling each image pixel with the class label of its surrounding object, is generally the first step toward the automatic understanding of images. Remarkable variations in the appearance of healthy and unhealthy skin, including color, texture, lesion shape and size originating from image acquisition and inter- and intra-class variation, complicates the skin lesion segmentation problem.

For decades, since the seminal work of Kass et al. [13], energy functional minimization techniques were the most popular approaches to solve image segmentation problems [15]. Imaging artifacts and variability in the appearance of image regions make the data fidelity term insufficient to achieve robust segmentation results. Therein, the segmentation that minimizes a weighted sum of unary (data) and regularization energy functional terms is sought. Incorporating

prior knowledge about the structure of target object in the objective function to regularize plausible solutions with anatomically meaningful constraints have been widely leveraged to obtain more reliable delineations [10,17]. Active shape models (ASM) was one of the pioneering works to incorporate shape priors into deformable models [9]. To effectuate the shape prior, ASM and many other shape-encoding segmentation methods required an estimate of the object pose (i.e., the orientation, scale, and location of the target object in the image) [11,22]. Some examples of priors which have been utilized in energy optimization based segmentation methods are shape models, topology preservation, moment constraints and geometrical and distance interaction between image regions.

Recently deep fully convolutional networks have achieved significant success in the task of semantic segmentation. Hierarchical extraction of features followed by skip connections and up-sampling operations was first introduced by Long et al. in an end-to-end trainable framework [14]. Despite the success of FCNs, they have indicated clear limitations in the dense per-pixel prediction task. Consecutive spatial pooling and striding convolutions in FCNs reduce the initial image resolution and lead to loss of the image fine structures. Some techniques have been proposed to address these limitations of FCNs. Learning multiple deconvolutional layers and concatenating low-level fine features with high-level coarse features through skip connections are commonly used to retrieve low-level visual features [20]. Dilated convolutional has also been introduced to aggregate multi-scale contextual information without losing image resolutions [24]. Although pixel-wise prediction benefits from these resolution enlarging techniques, they are only capable to partially recover detailed spatial information.

In the context of fully convolutional networks, leveraging prior information about the target object structure in the segmentation model has not been widely studied. By optimizing individual pixel level class predictions in the FCNs loss function, independent class labels are assigned to image pixels without considering high-level label dependencies. There have been some efforts towards structured prediction and leveraging meaningful priors into deep learning frameworks. Deeplab-CRF and CRF-RNN employ probabilistic graphical modeling either as a post processing step or by implementing recurrent layers in FCNs to enforce assigning similar labels to pixels with similar color and position and further improve the object boundaries [6]. Recently BenTaieb et al. proposed a new loss function to encode the geometrical and topological priors of containment and detachment in an end-to-end FCN framework [2,27]. To leverage the shape prior in segmentation models, Chen et al. learn a shape constraint by a deep Boltzmann machine and then employ the learned prior in a variational segmentation method [5]. In addition, training convolutional auto-encoder networks to learn anatomical shape variations has demonstrated improvements in the robustness of FCN segmentation models [18,19].

To the best of our knowledge, none of the existing works incorporates a star shape prior as a regularization term in the loss function of FCNs trained in an end-to-end fashion. The star shape prior was first introduced in the context of image segmentation by Veksler, where it was encoded as a regularization term

into the cost function formulation of a graph-based (discrete) image segmentation approach [21]. Later, Chittajallu et al. incorporated three types of shape constraints including star shape prior into a Markov random field based segmentation model and applied their method to non-contrast cardiac computed tomography scans [7]. Yuan et al. extended the star shape prior to 3D objects and applied it to prostate magnetic resonance images [25]. Nosrati et al. derived a star shape prior in a continuous variational formulation and applied it to segmenting overlapping cervical cells [16]. Although the star shape prior clearly improved results for a variety of target objects, one limiting requirement of Vekslers approach and its variants, however, is the assumption that the center of foreground objects is known (e.g. provided by user interaction).

We aim to harness the powerful proven capabilities of deep learning in automatically extracting learnt (i.e., not hand-crafted) pixel-driven image features (i.e., likelihood) and augment it with demonstrably useful shape priors without requiring the knowledge of the target object pose. We propose to encode the star shape prior into the training of fully convolutional networks to improve segmentation of skin lesions from their surrounding healthy skin. Our idea is to formulate the star shape prior in the loss function of FCN frameworks to penalize non-star shape segments in prediction maps and preserve global structures in the output space. Integration of the star shape prior in the loss function makes it possible to train the whole FCN framework in an end-to-end manner. In contrast to Veksler's work and its variants, our approach to star shape prior in a deep learning setting not only eliminates the need for manually setting object centers, but also alleviates, at inference time, the computationally intensive optimization associated with the energy minimizing approaches. Our experimental results illustrate how imposing the shape prior constraint in deep networks refines skin lesion segmentation in comparison to using a single pixel level loss in FCNs.

2 Methodology

Our goal is to leverage the star shape prior into the learning process of an FCN to generate plausible segmentation maps (e.g. skin lesions) from their surrounding background without requiring additional training, user interaction, pre- or post-processing.

FCN's pixel-wise loss In FCNs, given a set of N training images and their corresponding ground truth segmentations, $\{(X(i), Y(i)); i = 1, 2, \ldots, N\}$, the deep network learns to take unseen image samples and generate a segmentation probability map, the same size as the input images that assigns a semantic label to each pixel. Learning the deep network parameters θ, is performed by maximizing the a posteriori probability of giving the true label to each image pixel given the input image. Maximizing the a posteriori probability is usually replaced by minimizing its negative log-likelihood function as a cost function L:

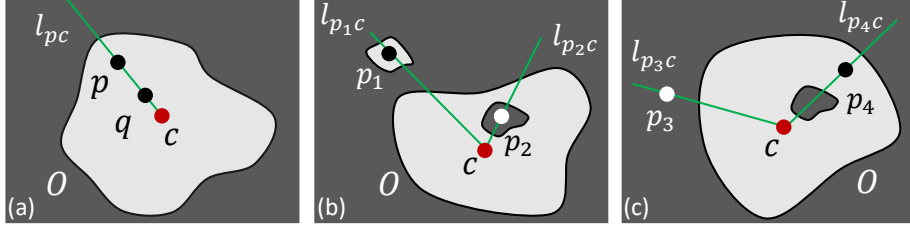$$\theta^* = \arg\min_{\theta} L(X, Y; \theta). \tag{1}$$

Fig. 1: (a) Star shape object O w.r.t. the supplied object center c (*red dot*). (b) Examples of the star shape constraint violation. (c) Examples of cases where conditions (i) and (ii) in (4) are required.

For binary dense class prediction, a binary cross entropy loss L_{ce} is generally deployed:

$$L_{ce}(X, Y; \theta) = -\sum_{i=1}^{N} \sum_{p \in \Omega} [y_{ip} \log P(y_{ip} = 1|X(i); \theta)$$
$$+ (1 - y_{ip}) \log(1 - P(y_{ip} = 1|X(i); \theta))] \qquad (2)$$

where Ω is the pixel space, y_{ip} is the ground truth label of pixel p in image i and $P(y_{ip} = 1|X(i); \theta)$ is the FCN sigmoid function output indicating the predicted probability of the pth pixel of the ith image being a skin lesion. The pixel-wise binary logistic loss L_{ce} penalizes the deviation of the predicted label for each pixel from its true label.

Star shape regularized loss Assuming c is the center of object O, object O is a star shape object if, for any point p interior to the object, all the pixels q lying on the straight line segment connecting p to the object center c are inside the object (Fig. 1-(a)). This definition of star shape prior holds for a large group of object shapes including convex ones. To incorporate the star shape prior as a new regularization term, we augment the loss function in (2) with a new loss term to penalize line segments that violate the prior (e.g. Fig. 1-(b)) in the prediction maps:

$$L(X, Y; \theta) = \alpha L_{ce} + \beta L_{sh} \qquad (3)$$

where α and β are hyper-parameters setting the contribution of each term in the optimization function, L_{ce} is the binary cross entropy loss and L_{sh} is our star shape prior:

$$L_{sh}(X, Y; \theta) = \sum_{i=1}^{N} \sum_{p \in \Omega} \sum_{q \in l_{pc}} B_{pq}^{i} \times |y_{ip} - P(y_{ip} = 1|X(i); \theta)|$$
$$\times |P(y_{ip} = 1|X(i); \theta) - P(y_{iq} = 1|X(i); \theta)|; \qquad (4)$$

$$B_{pq}^{i} = \begin{cases} 1, & \text{if } y_{ip} = y_{iq} \\ 0, & \text{otherwise} \end{cases} \qquad (5)$$

where l_{pc} is the line segment connecting pixel p to the object center c and q is any pixel incident on line l_{pc}. L_{sh} is trained to assign to all such q pixels a label identical to the label of pixel p as long as (i) p and q have the same ground truth labels ($B_{pq}^i = 1$) and (ii) the difference between the ground truth label and the predicted labels for p is non-zero ($|y_{ip} - P(y_{ip}|X(i);\theta)| > 0$). The 3rd term of (4) determines how labels of pixels internal to the lesion are penalized to ensure star shapes, whereas the first two terms of (4) are designed to allow discontinuities of pixel labels across the ground truth boundary of the lesion and ignore the star shape term when the given label is true. In Fig. 1-(c), $p = p_3$ and $p = p_4$ are examples where the value of $\sum_{q \in l_{pc}} |P(y_{ip}|X(i);\theta) - P(y_{iq}|X(i);\theta)|$ is positive while their assigned labels should not be penalized. Condition (i) chooses a set of pixels q on l_{p_3c} and allows discontinuities between the background (p_3) and foreground assigned labels and, condition (ii) enforces the loss function not to penalize the label assigned to p_4.

In our implementation of 4, instead of penalizing the difference between the predicted probabilities and ground truth labels for all the points on the straight line l_{pc}, we only examine the m closest pixels to p on l_{pc} and compute the loss value per pixel p based on those m predicted probabilities. We also quantize, to a set of d directions, the possible angles of all lines passing through p.   In the training of our deep network, we automatically find the star object center from binary ground truth maps. At inference time, we do not need to supply the center of star objects as prediction maps are achieved by a forward pass through the network whose parameters are already trained to generate segmentations.

3   Experiments

Data description We validated our proposed segmentation approach on dermoscopy data provided by the International Skin Imaging Collaboration (ISIC) at ISBI 2017 *Skin Lesion Analysis Towards Melanoma Detection Challenge* [8]. The data set contains 2000 training, 150 validation, and 600 test images. We first re-scaled all images to 192×192 pixels and normalized each RGB channel by the mean and standard deviation of the training data. To confirm the suitability of adopting the star-shape prior for this task, we calculated the percentage of segmentation mask pixels that violate the star shape definition to be only 0.14% over the whole dataset (0.05% of training, 0.3% of validation, and 0.38% of test image pixels). Fig. 2 shows examples of rare pixels where the star shape constraint is violated.

Network architecture We exploited two state-of-the-art fully convolutional network architectures to evaluate our proposed new loss: 1)U-Net[20] 2)ResNet-DUC. ResNet-DUC deploys the FCN version of ResNet-152, pretrained on ImageNet as an encoder [12]. Instead of using multiple deconvolutional layers to decode low resolution feature maps into the original image size prediction maps, single Dense Upsampling Convolution (DUC) layer is used to reconstruct fine-
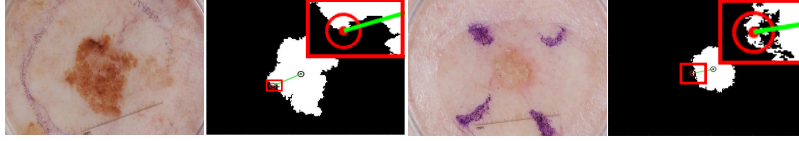
Fig. 2: Examples of skin lesion pixels violating the star shape constraint.

detailed information from coarse feature maps [23]. Furthermore, dilated convolutions are used in the encoder to benefit from multi-scale contextual information from previous layers activations [24].

Implementation   We trained deep networks implemented with the PyTorch library, over mini-batches of size 12. We tuned all hyper-parameters on the validation set. Loss functions are optimized using the stochastic gradient descent algorithm with an initial learning rate of 10^{-4}. The learning rate was divided by 10 when the performance of model on validation data set stopped improving. Momentum and weight decay were set to 0.99 and 5×10^{-5}, respectively. For the implementation of the star shape regularized loss function, $\alpha = 1$, $\beta = 5$, $m = 6$ and $d = 8$. We first trained the deep network with binary cross entropy function for 5 epochs and then fine-tuned the network parameters with the proposed loss function. Training takes 2 days and test takes 1 sec/image on our 12 GB GPU.

Results   We evaluated the performance of U-Net and ResNet-DUC trained with and without the star shape prior. As shown in Table 1, using our shape regularized loss function in the training of U-Net and ResNet-DUC, the Jaccard index is improved by more than 3% (row A vs. B and row C vs. D). We measured the statistical significance of our results by exploring the Jaccard index over the test data. We used the non-parametric Wilcoxon signed rank sum test and found that the results of U-Net and ResNet-DUC with and without incorporation of star shape prior are statistically significantly different at $p < 0.05$.

We compared our proposed method with 21 competing methods participating in the challenge. The ResNet-DUC architecture trained with our star shape regularized loss achieved the first rank based on the challenge ranking metric, Jaccard index. Table 1, rows E, F and G, show results of the first three ranked teams. Although all top three teams used FCNs to perform image segmentation, in contrast to our work, they employed various additional steps like averaging over multiple model results, multi-scale image input as well as pre- and post-processing approaches like inclusion of different color spaces in the input and multi-thresholding. Qualitative results of our proposed approach are presented in Fig. 3. Encoding star shape prior into the loss function results in smoother prediction maps with a single connected component as lesion for most cases.

4   Conclusion

We encoded the star shape prior in the loss function of an end-to-end trainable fully convolutional network to generate more accurate and plausible skin

Table 1: Segmentation quantitative performance. Bold numbers indicate the best performance. All values are in percentages.

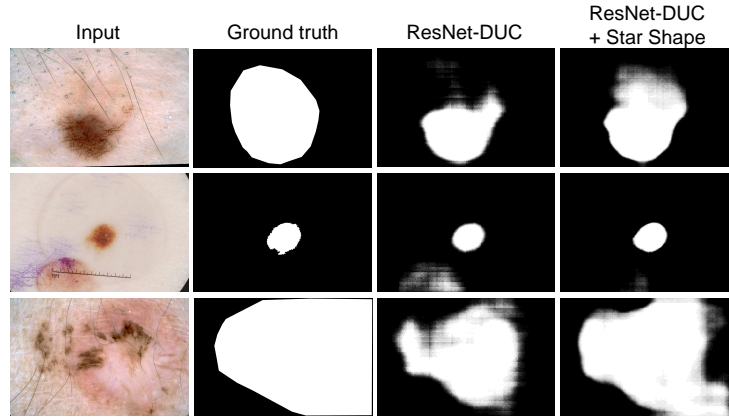|	Method	Jaccard	Dice	Accuracy	Specificity	Sensitivity
A	U-Net [20]	70.5	79.7	91.8	97.8	77.0
B	U-Net + Star Shape	73.3	82.4	92.4	95.3	85.4
C	ResNet-DUC [23]	74.0	83.3	93.00	98.2	80.0
D	ResNet-DUC + Star Shape	**77.3**	**85.7**	**93.8**	97.3	**85.5**
E	Yuan et al. [26]	76.5	84.9	93.4	97.5	82.5
F	Berseth et al. [3]	76.2	84.7	93.2	97.8	82.0
G	Bi et al. [4]	76.0	84.4	93.4	**98.5**	80.2

Fig. 3: Qualitative comparison of ResNet-DUC architecture results with and without star shape prior.

lesion segmentations. In contrast to energy minimization approaches, our proposed framework does not require computationally expensive optimization at inference time nor a user-defined object centre. Our experiments indicated that leveraging the prior knowledge in fully convolutional networks yield convergence to an improved output space. In future works, we will extend to other prior information including but not limited to anatomically meaningful priors in fully convolutional networks trained for other 2D and 3D medical imaging applications.

References

1. Cancer facts and figures 2017. http://www.cancer.org/acs/groups/content/ @editorial/documents/document/acspc-048738.pdf (2017)

2. BenTaieb, A., Hamarneh, G.: Topology aware fully convolutional networks for histology gland segmentation. In: MICCAI. pp. 460–468 (2016)
3. Berseth, M.: ISIC 2017-skin lesion analysis towards melanoma detection. arXiv:1703.00523 (2017)
4. Bi et al.: Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks. arXiv:1703.04197 (2017)
5. Chen et al.: Deep learning shape priors for object segmentation. In: IEEE CVPR. pp. 1870–1877 (2013)
6. Chen et al.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. arXiv:1606.00915 (2016)
7. Chittajallu et al.: A shape-driven MRF model for the segmentation of organs in medical images. In: IEEE CVPR. pp. 3233–3240 (2010)
8. Codella et al.: Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC). arXiv:1710.05006 (2017)
9. Cootes et al.: Active appearance models. IEEE TPAMI 23(6), 681–685 (2001)
10. Cremers et al.: A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. IJCV 72(2), 195–215 (2007)
11. Freedman, D., Zhang, T.: Interactive graph cut based segmentation with shape priors. In: IEEE CVPR. vol. 1, pp. 755–762 (2005)
12. He et al.: Deep residual learning for image recognition. In: IEEE CVPR. pp. 770–778 (2016)
13. Kass et al.: Snakes: Active contour models. IJCV 1(4), 321–331 (1988)
14. Long et al.: Fully convolutional networks for semantic segmentation. In: IEEE CVPR. pp. 3431–3440 (2015)
15. McInerney, T., Terzopoulos, D.: Deformable models in medical image analysis: a survey. MIA 1(2), 91–108 (1996)
16. Nosrati, M.S., Hamarneh, G.: Segmentation of overlapping cervical cells: a variational method with star-shape prior. In: IEEE ISBI. pp. 186–189 (2015)
17. Nosrati, M.S., Hamarneh, G.: Incorporating prior knowledge in medical image segmentation: a survey. arXiv:1607.01092 (2016)
18. Oktay et al.: Anatomically constrained neural networks (ACNNs): Application to cardiac image enhancement and segmentation. IEEE TMI 37(2), 384–395 (2018)
19. Ravishankar et al.: Learning and incorporating shape models for semantic segmentation. In: MICCAI. pp. 203–211 (2017)
20. Ronneberger et al.: U-Net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241 (2015)
21. Veksler, O.: Star shape prior for graph-cut image segmentation. In: ECCV. pp. 454–467 (2008)
22. Vu, N., Manjunath, B.: Shape prior segmentation of multiple objects with graph cuts. In: IEEE CVPR. pp. 1–8 (2008)
23. Wang et al.: Understanding convolution for semantic segmentation. arXiv:1702.08502 (2017)
24. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. arXiv:1511.07122 (2015)
25. Yuan et al.: An efficient convex optimization approach to 3D prostate MRI segmentation with generic star shape prior. PROMISE Challenge, MICCAI (2012)
26. Yuan et al.: Automatic skin lesion segmentation with fully convolutional-deconvolutional networks. arXiv:1703.05165 (2017)
27. Zheng et al.: Conditional random fields as recurrent neural networks. In: ICCV. pp. 1529–1537 (2015)