GENERATIVE ADVERSARIAL NETWORKS TO SEGMENT SKIN LESIONS

Saeed Izadi, Zahra Mirikharaji, Jeremy Kawahara, and Ghassan Hamarneh

Medical Image Analysis Lab, School of Computing Science, Simon Fraser University, Canada

ABSTRACT

The accuracy of skin lesion segmentation has increased in recent years, thanks to advances in machine learning techniques and a large influx of dermoscopy images. However, there is still room for improvement as there exist many considerable challenges mainly due to the large variability in the appearance of lesions (i.e., shape, size, texture, and occlusions). In this work, we present a novel approach for skin lesion segmentation through leveraging generative adversarial networks. Our approach consists of two models: a fully convolutional neural network designed to synthesize an accurate skin lesion segmentation mask (the segmenter), and a convolutional neural network that distinguishes between synthetic and real segmentation masks (the critic). Our experimental results on 1300 images from the DermoFit dataset show that incorporating a critic network to complement a fully convolutional segmenter, like UNet, increases segmentation accuracy.

1. INTRODUCTION

Malignant melanoma is a common cancer and it is estimated the number of new cases in the United States has increased to 87,110 within the past year [1]. Since early detection of melanoma is critical to improve survival rate, there is a need to expedite the diagnosis process through automating the analysis of dermoscopic images.

Melanoma segmentation refers to the task of localizing and delineating the boundary of lesion and segmenting it from surrounding normal skin regions. However, this task is not trivial as melanoma is subject to many challenging variability in appearance such as size, shape, and texture. Furthermore, melanoma potentially has fuzzy boundaries such that the contrast between the lesion and its surroundings may be unclear. Furthermore, within the image, irrelevant distracting artifacts may be present, such as hairs, vessels, air blobs, medical gauze, and light reflection, over the lesion surface, which makes the segmentation task more difficult and error-prone.

Many efforts have been devoted to automatic pixel-level segmentation of skin lesions. Initial investigations are mostly built on color, luminance, and texture as discriminative factors in lesion localization. For instance, Li et al. used a selection of low-level features like color and depth along with classical edge detectors into contour-based approaches [2]. However, hand-designed approaches may not generalize well over unseen images.

With recent advances in deep learning techniques, specifically deep convolutional neural networks (DCNN), significant performance improvements have been witnessed in the field of computer vision; e.g. image classification [3], object detection [4] and semantic segmentation [5]. DCNNs have also found their way into the medical image analysis domain such that several models have been designed and introduced particularly to improve medical image analysis accuracy [6, 7]. For skin lesion segmentation, Yuan et al. used a 19-layer convolutional-deconvolutional network to train an automatic end-to-end lesion segmentation model [8]. They proposed to utilize the Jaccard distance as their loss function to eliminate the need of sample re-weighting due to imbalance between foreground and background pixels. Local and global information were also used by Jafari et al. in two parallel network branches for more accurate delineation of the lesion region [9].

Deep learning techniques based on generative models, known as generative adversarial networks (GANs) [10], have further pushed the state of the art in some domains. Generally, GANs perform a minmax game between two players, namely a *generator* and a *discriminator* network, referred to as a *segmenter* and *critic*, respectively, in our work. Given a training dataset, the generator/segmenter attempts to synthesize outputs that match the ground truth segmentations, while the discriminator/critic is responsible for distinguishing between synthetic and real outputs. Training these two networks in an adversarial fashion results in two strong models after stabilization.

Inspired by Pan et al. [11] and Luc et al. [12], we propose to use a generative adversarial network to segment skin lesions. In the domain of medical images, other works have used GANs to segment aggressive prostate cancer [13] and brain regions [14] from MRI images. In this paper, we aim to practically examine the role of critic network in improving the performance of an existing model. To this end, we use a fully convolutional segmentation model and augment it with a critic neural network model. Once trained appropriately, we show that including the critic into our network increases the quality of segmentations produced by the segmenter model, compared to the case of a critic-free network architecture. We evaluate our model on the DermoFit skin lesion dataset.

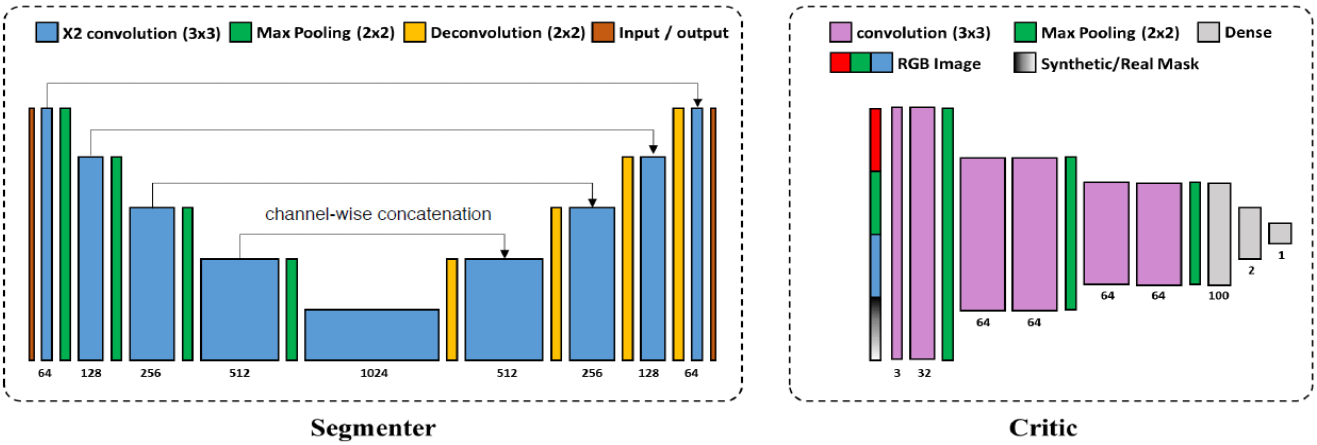

Fig. 1. The schematic of the proposed UNet-Critic model for skin lesion segmentation. The error in the critic is backpropagated through the segmenter to make it produce more realistic segmentation masks.

2. METHOD

Our goal is to accurately segment the skin lesions from their surroundings, independent of the diversity in their appearance and without any manual intervention. To do so, we frame the problem as a binary dense labeling task: Given a dermoscopic image, we aim to predict either "lesion" or "background" labels for each pixel.

Given an existing fully convolutional segmentation model, i.e. segmenter, which synthesizes probabilistic segmentation masks, we propose to design and employ a DCNN with a single output node, i.e. critic, to distinguish between the synthetic segmentation masks and real ground truth. By importing the feedback from the critic into the segmenter, the latter learns to produce more plausible lesion segmentations. A stabilization state occurs when the segmenter synthesizes segmentation masks that the critic is unable to differentiate from ground truth lesion segmentations. We hypothesize that by training these two networks adversarially, the competitive atmosphere will lead to a segmenter that produces more accurate lesion segmentations. Our experimental results demonstrate that adding the critic network to the segmenter model leads to improvements compared to increasing the complexity of the architecture and/or the design. Fig. 1 depicts the schematic of our proposed model.

2.1. Segmenter

We use UNet [6] as our base segmenter model. This model has an encoder-decoder architecture to transform an RGB image into a segmentation mask, while connecting the feature maps from earlier to later layers. These so called skip connections deliberately leverage the precise localization cues captured in earlier layers to produce finer boundaries in the resultant segmentation mask.

Let $I_{rgb} \in \mathbf{I}$ denote an input image, $\mathcal{T} \in \mathbf{T}$ be the ground truth mask, and $\mathcal{M} \in \mathbf{M}$ be the synthetic segmentation mask. Each pixel i in the segmentation mask $\mathcal{M} = \{m_i, i = 1, \ldots, N\}$ takes a value in the range $L = [0, 1]$, and each pixel in $\mathcal{T} = \{t_i, i = 1, \ldots, N\}$ takes a value from the set $\{0, 1\}$. Given an input image, I_{rgb}, and a set of learned parameters, θ_s, the conditional probability of a label assignment $\mathcal{M}$ is:

$$P(\mathcal{M}|I_{rgb}; \theta_s) = \sigma\left(\psi_{\theta_s}(I_{rgb})\right) \tag{1}$$

where $\sigma(\cdot)$ is the *sigmoid* activation function applied to the neural network model's output layer $\psi_{\theta_s}(.)$. We use binary cross-entropy as the loss function to train the segmenter network, which is computed as follows:

$$\mathcal{L}_{\theta_s} = -\frac{1}{N}\sum_{i=1}^{N}\left[t_i \log(m_i) + (1 - t_i)\log(1 - m_i)\right] \tag{2}$$

where t_i and m_i are the predicted and true labels for each pixel, respectively.

2.2. Critic

We augment the segmenter network with a DCNN that receives a dermoscopic image and either a synthetic or real lesion segmentation mask as inputs, and attempts to distinguish between the two cases. In particular, synthetic or real segmentation masks are concatenated to the RGB image along the channel dimension, and are assigned a label of 0 (indicating a synthetic image) or 1 (indicating a real image). The new 4-channel image is fed into a set of convolutional and max-pooling operations (Fig. 1), and the final single node learns to predict the true binary labels. The critic network contains six 3×3 convolutional layers, three max-pooling, and three linear layers, all using ReLU activation functions, except for the

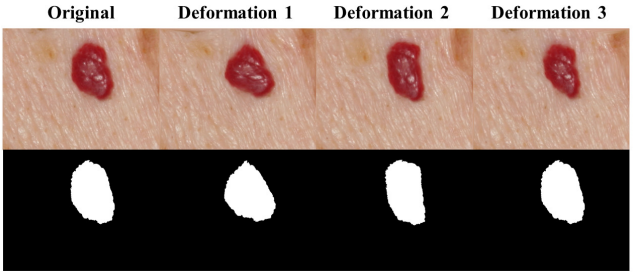

Fig. 2. Results of elastic deformation on skin lesions and their corresponding segmentation masks.

final layer which uses the sigmoid function. Batch normalization [15] is also used after every convolution operation.

As above, let $I_{rgb} \in \mathbf{I}$ be an input image and $\mathcal{S} \in \{\mathcal{M}, \mathcal{T}\}$ be either the synthetic or real segmentation mask. After both inputs (I_{rgb} and $\mathcal{S}$) are concatenated along the channel dimension, it can take the label value of $L = \{0, 1\}$. Once fed into the network, the conditional probability of a label assignment y is:

$$P(y|I_{rgb}, \mathcal{S}; \theta_c) = \sigma\left(\psi_{\theta_c}(I_{rgb}, \mathcal{S})\right) \tag{3}$$

where θ_c denotes the parameters in the critic network, and $\psi_{\theta_c}(\cdot)$ refers to the output of the critic network. Similar to the segmenter model, we use the binary cross-entropy as the loss function for training the critic network, denoted as $\mathcal{L}_{\theta_c}$.

2.3. Training

Optimizing the proposed framework proceeds by alternating between training the segmenter to produce synthetic segmentation masks while keeping the critic fixed, and training the critic using the synthetic and real segmentation masks while the segmenter is fixed. The error in the critic must be back-propagated through the segmenter in order for the segmenter to learn how to produce segmentations that can fool the critic. This is performed through adding the pixel-wise binary cross-entropy error in the segmenter with that of critic. Thus, the final loss function for updating the segmenter is as follows:

$$\mathcal{L}_{\text{final}} = \mathcal{L}_{\theta_s} + \lambda \mathcal{L}_{\theta_c} \tag{4}$$

where $\lambda = 0.2$ is the coefficient to balance the effect of critic error value. In other words, the coefficient is set to encourage the learning rate of both networks to be comparable in value, in order to reduce training instability.

3. EXPERIMENTAL RESULTS

We validate our proposed model on the DermoFit dataset [16], which contains 1300 high quality focal skin lesions. The dataset contains lesions from ten different disease categories and encompasses various potential challenges in lesion appearances, which complicates the segmentation task. In addition to category annotation, there exists a binary segmentation mask for each image. To the best of our knowledge, we are the first to benchmark an automatic lesion segmentation approach on this challenging dataset.

Table 1. Quantitative Results. Bold numbers indicate the best performance.

	Method	Dice	Jacc.	Sens.	Spec.	Acc.
A	U-Net [6]	0.887	0.781	**0.906**	0.955	0.936
B	UNet-Critic	**0.898**	**0.812**	0.891	**0.971**	**0.942**

Data Augmentation. In addition to horizontal flipping, vertical flipping, and rotations, we apply a set of elastic deformations to each image to generate synthetic lesions with different geometric shapes. Fig. 2 shows a set of newly generated lesions using DeformIt [17]. We enlarge the size of the training set by a factor of $\sim$60.

Implementation Details. We divide the original dataset into a training (80%) and test set (20%), and augment the training examples with the aforementioned mask-consistent deformations. The segmenter network is trained for 10 epochs ($\sim$60K iterations). Since the segmenter and critic networks are trained alternately in the adversarial setting, we double the number of epochs in the *UNet-Critic* model for fair comparison. We use SGD with momentum and weight decay regularization to train both networks. All hyper-parameters (e.g., learning rate, λ) of our model are selected on 20 percent of the unaugmented training set via grid search. Training takes $\sim$35 hours on a machine with one Titan X (Pascal) GPU using Lasagne [18].

Quantitative Results Table 1 presents the resulting quantitative metrics from our proposed model and the competing approach. We see that by incorporating a CNN critic, the segmentation performance of UNet is improved. We also highlight this work as a substitution approach to other works that use more complicated architectures, where this additional training may improve overall model performance.

Qualitative Results As shown in Fig. 3, our approach leads to more uniform and compact segmentation masks. We note that our proposed model succeeds in filling the holes in the foreground and eliminating the island regions in the background area. For lesions with clearly defined boundaries, both approaches perform similar; however, our proposed model yields higher quality segmentations than the simple UNet model for cases with low contrast, unclear and irregular lesion boundaries. Our experiments verify that the boundary improvements are a result of adding the critic, which explicitly compares the synthetic segmentation mask with the ground truth during training.

4. CONCLUSION

We examined the effect of adding a critic network to an existing skin lesion segmenter model. The critic network receives

the synthetic or real segmentation mask along with the input dermoscopy image and learns to distinguish between these two cases. We then backpropagate the error of the critic into the segmenter training procedure to encourage more realistic segmentation masks. Quantitatively, our proposed approach shows a relative improvement to a state-of-the-art model. Our qualitative results also reveals that including the critic module helps the segmenter to uniformly highlight the interior regions of the lesion and produce fine predictions around boundaries. Our work is also the first to benchmark lesion segmentation over the DermoFit dataset. Future work would evaluate our proposed approach over other skin datasets such as the ISIC dataset [19].

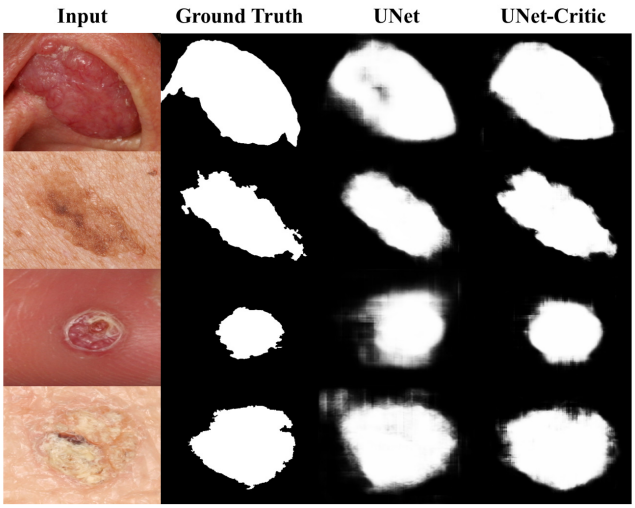

Fig. 3. Qualitative results of UNet-Critic vs. UNet.

5. REFERENCES

[1] R. L. Siegel *et al.*, "Cancer statistics, 2017," *CA: A Cancer Journal for Clinicians*, vol. 67, no. 1, pp. 7–30, 2017. [Online]. Available: http://dx.doi.org/10.3322/caac.21387

[2] X. Li *et al.*, "Depth data improves skin lesion segmentation," *MICCAI*, pp. 1100–1107, 2009.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385

[4] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," in *CVPR*, 2017.

[5] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *preprint arXiv:1703.06870*, 2017.

[6] O. Ronneberger *et al.*, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015, pp. 234–241.

[7] F. Milletari *et al.*, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *3D Vision (3DV)*, 2016, pp. 565–571.

[8] Y. Yuan, M. Chao, and Y.-C. Lo, "Automatic skin lesion segmentation using deep fully convolutional networks with Jaccard distance," *IEEE Transactions on Medical Imaging*, pp. 1876–1886, 2017.

[9] M. H. Jafari *et al.*, "Skin lesion segmentation in clinical images using deep learning," in *ICPR*, 2016, pp. 337–342.

[10] I. Goodfellow *et al.*, "Generative adversarial nets," in *NIPS*, 2014, pp. 2672–2680.

[11] J. Pan *et al.*, "Salgan: Visual saliency prediction with adversarial networks," in *CVPR Scene Understanding Workshop (SUNw)*, 2017.

[12] P. Luc *et al.*, "Semantic segmentation using adversarial networks," *preprint arXiv:1611.08408*, 2016.

[13] S. Kohl *et al.*, "Adversarial networks for the detection of aggressive prostate cancer," *preprint arXiv:1702.08014*, 2017.

[14] P. Moeskops *et al.*, "Adversarial training and dilated convolutions for brain mri segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 2017, pp. 56–64.

[15] S. Ioffe *et al.*, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in *ICML*, vol. 37, 2015, pp. 448–456.

[16] L. Ballerini *et al.*, "A color and texture based hierarchical k-NN approach to the classification of non-melanoma skin lesions," in *Color Medical Image Analysis*, 2013, pp. 63–86.

[17] G. Hamarneh *et al.*, "Simulation of ground-truth validation data via physically-and statistically-based warps," in *MICCAI*, 2008, pp. 459–467.

[18] S. Dieleman *et al.*, "Lasagne: First release." Aug. 2015. [Online]. Available: http://dx.doi.org/10.5281/zenodo.27878

[19] N. C. 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)," *preprint arXiv:1710.05006*, 2017.