Lesion Attributes Segmentation for Melanoma Detection with Deep Learning

Eric Z. Chen *
Dana-Farber Cancer Institute
Boston, MA, USA
zhang_chen@dfci.harvard.edu

Xu Dong
Virginia Tech
Blacksburg, VA, USA
xu14@vt.edu

Junyan Wu
Cleerly Inc
New York City, New York, USA
mylotarg1989@gmail.com

Hongda Jiang
East China University of Science and Technology
Shang Hai, China
814473689@qq.com

Xiaoxiao Li
Yale University
New Haven, CT, USA
xiaoxiao.li@yale.edu

Ruichen Rong
UT Southwestern Medical Center
Dallas, TX, USA
ruichenrong@gmail.com

Abstract

Melanoma is the most deadly form of skin cancer worldwide. Many efforts have been made for early detection of melanoma. The International Skin Imaging Collaboration (ISIC) hosted the 2018 Challenges to help the diagnosis of melanoma based on dermoscopic images. In this paper, we describe our solutions for the task 2 of ISIC 2018 Challenges. We present two deep learning approaches to automatically detect lesion attributes of melanoma, one is a multi-task U-Net model and the other is a Mask R-CNN based model. Our multi-task U-Net model achieved a Jaccard index of 0.433 on official test data, which ranks the 5th place on the final leaderboard. The code for our solutions is publicly available.

1 Introduction

Skin cancer is one of the most common cancers worldwide and more than one million skin cancers have been diagnosed in the United States each year. Melanoma is the most dangerous form of skin cancer, which causes over 9,000 deaths each year [Group et al., 2013]. Melanoma in the late stage can often spread to other parts of the body and thus it is difficult to treat and usually can be fatal. However, melanoma in the early stage is treatable and the majority can be cured. Many efforts have been made to detect melanoma in the early stage. Dermoscopy is one noninvasive method commonly used in the healthcare to examine pigmented skin lesion. It can generate high-resolution images of the lesion regions on the skin. To diagnose melanoma, it still requires the dermatologist to evaluate the images based on several skin lesion patterns. The common dermoscopic attributes are pigment network, negative network, streaks, milia-like cysts, and globules [Mishra and Celebi, 2016]. Automatic detection of those skin lesion attributes can be a tremendous help for early melanoma diagnosis.

Towards this goal, the International Skin Imaging Collaboration (ISIC) hosted the 2018 Challenges to diagnose melanoma automatically based on dermoscopic images. The 2018 challenges include

*To whom correspondence should be addressed

Preprint. Work in progress.

three tasks. The first task is to predict lesion segmentation boundaries in dermoscopic images. The second task is to predict the locations of five skin lesion patterns (i.e., dermoscopic attributes) in dermoscopic images, which are pigment network, negative network, streaks, milia-like cysts, and globules. The third task is to predict seven disease categories in dermoscopic images, which includes melanoma, melanocytic nevus, basal cell carcinoma, actinic keratosis / Bowen's disease, benign keratosis, dermatofibroma and vascular lesion. The three tasks mimic the steps for lesion analysis performance by dermatologists in the clinic settings. From the previous years' competitions, it has been reported that far less participation in the second task of lesion pattern prediction than in other tasks [Codella et al., 2018]. It seems that the second task is the most difficult one and this motivated us to focus on this problem instead of the other two.

Deep learning, especially convolutional neural network (CNN), has been widely applied to solve many problems in computer vision. Various CNN based models have been developed for object classification and detection, such as VGG [Simonyan and Zisserman, 2014], ResNet [He et al., 2016], Inception [Szegedy et al., 2017], DenseNet [Huang et al., 2017], and more [Liu et al., 2017]. For image segmentation, FCN [Long et al., 2015], U-Net [Ronneberger et al., 2015], Mask R-CNN [He et al., 2017] are some of the most used deep learning models. Many deep learning based approaches have been applied for melanoma detection [Mishra and Celebi, 2016]. However, most of those researches applied deep learning models to either classify skin diseases [Haenssle et al., 2018, Yu et al., 2018] or segment whole lesion regions [Zhang, 2017], which are corresponding to the task 1 and task 3 in ISIC 2018 Challenges. To our knowledge, currently there is no deep learning based approach has been applied to segment specific lesion attributes as in the task 2 of ISIC 2018 Challenges.

In this paper, we describe our solutions for the task 2 of ISIC 2018 Challenges. We present two deep learning approaches to automatically detect lesion attributes of melanoma. One is a U-Net based approach and the other is a Mask R-CNN based approach. In the U-Net model, we replaced the encoder part of the U-Net with a pretrained VGG16 network [Shvets et al., 2018]. In the middle layer and the last layer of the U-Net, we added two classification heads to classify the empty masks versus the non-empty masks. Our multi-task U-Net model achieved a Jaccard index of 0.433 on official test data, which ranks the 5th place on the final leaderboard. The code for our solutions is publicly available at https://github.com/chvlyl/ISIC2018.

2 Methods

2.1 Data

We downloaded the task 2 dataset from ISIC 2018 Challenges website [Tschandl et al., 2018]. For each skin image, this task is aimed at detecting five lesion attributes (pigment network, negative network, streaks, milia-like cysts, and globules). The dataset consists of 2,594 images and 12,970 ground truth masks (five masks for each skin image) as training data. The skin images are RGB images in JPG format and the masks are grayscale images in PNG format. The five masks are corresponding to five lesion attributes in each skin image. Note that some of the attributes may not present in skin images, therefore the corresponding masks are empty. There are 100 images provided as validation data and 1,000 images as test data. No ground truth masks are provided for validation data and test data during the competition.

2.2 U-Net based approach

Based on the original U-Net [Ronneberger et al., 2015], we replace the encoder part with a pretrained VGG16 [Iglovikov et al., 2018]. The pretrained weights are from ImageNet. In the skip connections, the feature maps from the encoder part are concatenated with the corresponding feature maps in the decoder part. The output of the U-Net is a five channel mask. We define the loss function for pixel wise segmentation as a combination of binary cross entropy loss L and Jaccard index J [Shvets et al., 2018].

$$loss_1 = L - logJ \tag{1}$$

We observed that for each skin images, not all lesion attributes present in the masks. Therefore, we also add two classification heads to classify the empty masks versus the non-empty masks. Since the last layer of our U-Net is a convolutional layer and the output is five feature maps, we add a global max pooling layer to this last convolutional layer and use this to classify the empty and non-empty

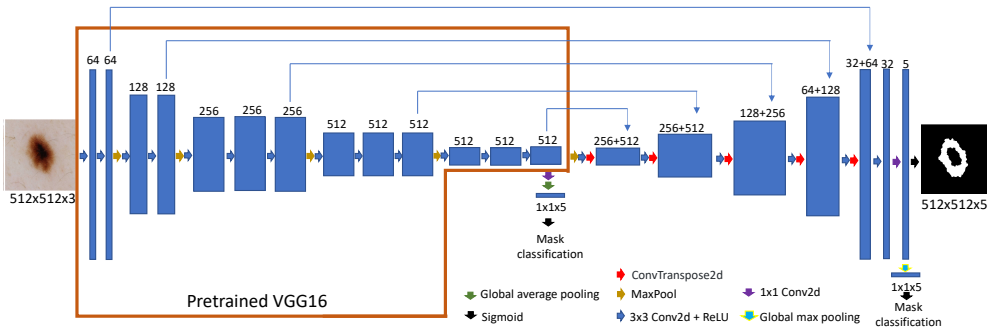

Figure 1: Network structure using in U-Net based approach. The encoder part of the U-Net is replaced with a pretrained VGG16 network [Shvets et al., 2018]. We also add two classification heads to the network, where one is added to the middle layer with 1×1 convolution and global average pooling layers and the other is added to the last layer with a global max pooling layer.

masks for five attributes. This classification head will act as a regularizer. Since if the truth mask is empty and our network predicts a non-empty mask, after global max pooling, the classification loss will be large. Therefore, this classification head will force the network to predict an empty mask if the true mask is empty. We also adapt the similar idea in Inception [Szegedy et al., 2017] and add another classification head to the middle layer with a 1×1 convolutional layer and a global average pooling layer. The network structure is shown in Figure 1.

The final loss is a combination of pixel wise segmentation loss $loss_1$, classification loss using the middle layer $loss_2$ and classification loss using the last layer $loss_3$. We used binary cross entropy for $loss_2$ and $loss_3$. We assigned weights 0.5 to the last two losses.

$$loss = loss_1 + 0.5 \times loss_2 + 0.5 \times loss_3 \tag{2}$$

The raw skin images have various sizes, from 540×722 to 4499×6748. We first resize all the skin images and corresponding masks to 512×512. Then for each skin image, we concatenate the five corresponding grayscale masks into a five-channel mask. Each channel is a 512×512 mask. Both skin images and mask images are divided by 255 to scale the values in [0,1].

We first freeze all the encoder layers and only train the decoder layers for 50 epochs. Then we train the whole network for another 250 epochs. We use a learning rate of 0.01 with Adam optimizer and then decrease it by 0.8 when validation loss does not improve for 5 epochs during training. During training, we augment the skin images and masks by random flipping, rotation, scaling as well as randomly adjusting brightness and saturation. We use cutoff 0.3 to transform the predicted probability into the binary mask. The whole analysis is implemented in Pytorch.

2.3 Mask R-CNN based approach

Mask R-CNN has been recognized as the state-of-art neutral network architecture for the instance segmentation. It utilizes the region based convolution neutral network and achieves final pixel-level segmentation based on the generated bounding box. The success of Mask R-CNN in the medical image field has been demonstrated by some previous work. Here we adapted the Mask R-CNN architecture for the skin attribute detection task in the challenge.

We used the ResNet 101 as the backbone of the network. The backbone weights are initialized by the weights trained in the coco dataset. We added the head layers onto the ResNet 101 and the gradient backpropagation was only done to the head layer during the training. The Adam optimizer was used with the initial learning rate as 0.001 and momentum as 0.99.

We did the training directly using the original images in the provided dataset without any preprocessing because it gave better result compared with training onto the resized images (1024 by 1024). To promote the training, only those image with non-empty attribute masks are used in the training. Under such training strategy, the best Jaccard index we obtained is 0.233.

Table 1: Summary of non-empty masks in the training data

Attributes	Number of non-empty masks	Percentage
Pigment network	1522	58.7%
Negative network	189	7.3%
Streaks	100	2.9%
Milia-like cysts	681	26.3%
Globules	602	23.2%
Total skin images	2594	100%

Table 2: Jaccard index on the official validation and test dataset

Model	Validation	Test
U-Net	0.477	0.433
Mask R-CNN	0.233	NA

2.4 Results

Figure 2 shows an example of several skin images and their corresponding lesion attributes (for each attribute, top rows show the ground truth masks). We noticed that not all attributes present in each skin image. That is, some of the masks are empty for certain skin images. Table 1 shows the summary of non-empty masks for each lesion attribute. 58.7%, 26.3% and 23.2% of the skin images have non-empty masks for Pigment network, milia-like cysts, and globules, respectively. While only 7.3% and 2.9% of the skin images have non-empty masks for negative network and streaks, respectively. Since segmentation is essentially a pixel-wise classification problem, a large number of empty masks increase the number of samples in negative class. Therefore, in the Mask R-CNN based approach, we train the model only using the non-empty masks. In the U-Net based approach, we add two classification heads to classify the empty versus non-empty masks. Figure 2 shows the predicted masks from the U-Net based approach (for each attribute, the bottom rows show the predicted masks). The Jaccard index on the official validation dataset is 0.477 for U-Net based approach and 0.233 for Mask R-CNN based approach (Table 2).

2.5 Discussion

In this paper, we proposed two deep learning based approaches to detect five attributes of skin lesions. We achieved a Jaccard index of 0.477 and 0.233 on the official validation data. The final Jaccard index on the official test data is 0.433 for our multi-task U-Net model, which ranks the 5th place on the final leaderboard. We didn't submit our test result for the Mask R-CNN model.

For our U-Net based model, we tried to add the batch normalize layers to the encoder and decoder parts of the network. However, this does not improve the model performance compared to the model without batch normalization layers. We also tried to replace the encoder part of the U-Net network with pretrained DenseNet. However, the GPU memory usage was too hight and we could only fit one image per patch. The training took a long time and the final results were worse than VGG as the encoder.

We also noticed that the five lesion attributes are highly imbalanced in the training data. For example, 58.7% of the skin images have the pigment network attribute but only 2.9% of skin images have the negative network attribute. Classification of imbalanced data is often difficult. One potential solution is to assign different weighs to each attribute during training. However, we spent some time exploring different weight configurations but it seems to us there was no significant improvement.

References

N. C. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, et al. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the

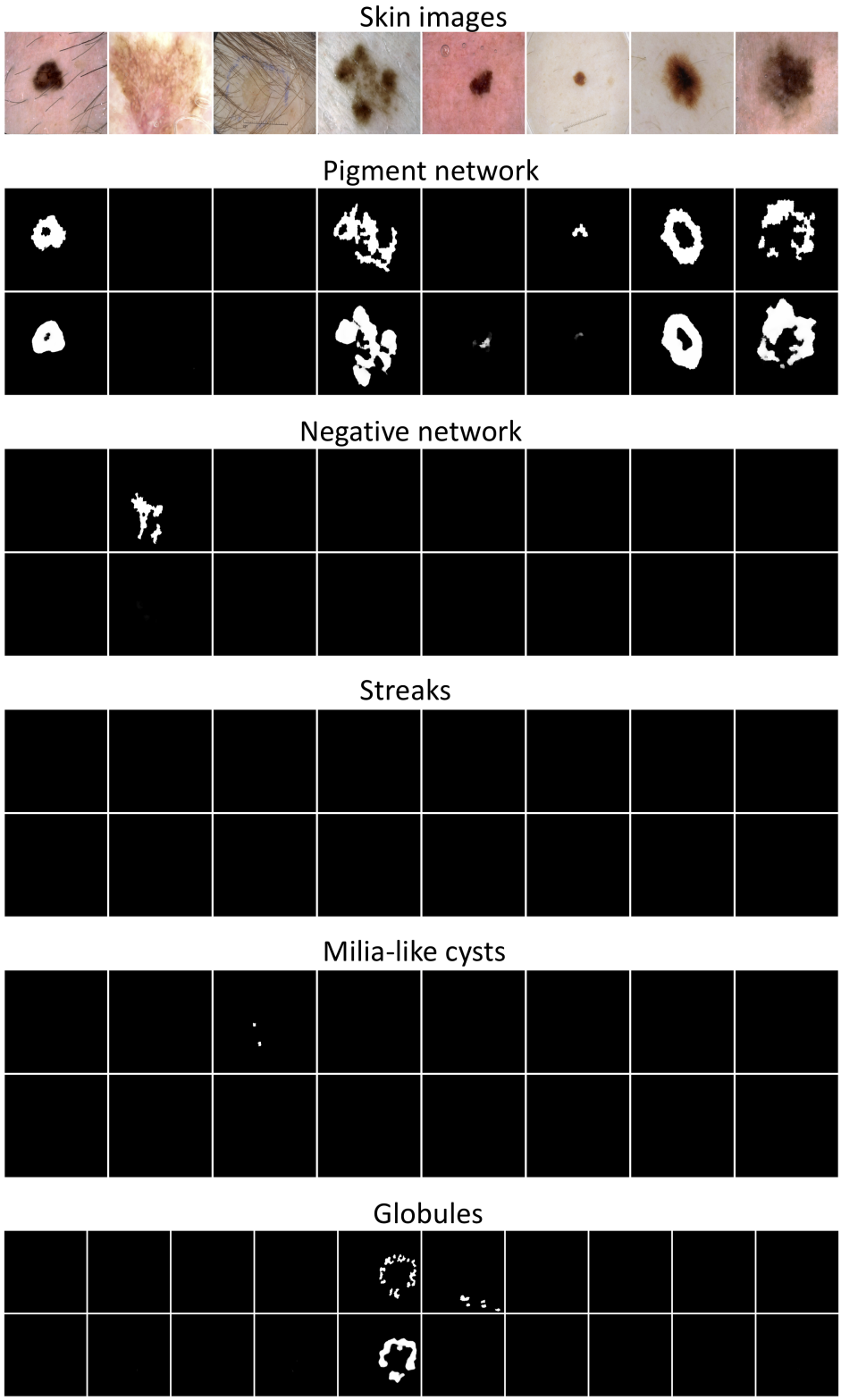

Figure 2: Example of skin images and five corresponding lesion attributes. For each attribute, the top row is the ground truth masks and the bottom row is the predicted masks from U-Net based approach.

international skin imaging collaboration (isic). In *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*, pages 168–172. IEEE, 2018.

U. Group et al. United states cancer statistics: 1999–2010 incidence and mortality web-based report. *US Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute, Atlanta, GA, Report. http://www. cdc. gov/uscs*, 2013.

H. Haenssle, C. Fink, R. Schneiderbauer, F. Toberer, T. Buhl, A. Blum, A. Kalloo, A. Hassen, L. Thomas, A. Enk, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of Oncology*, 2018.

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, pages 2980–2988. IEEE, 2017.

G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *CVPR*, volume 1, page 3, 2017.

V. I. Iglovikov, S. Seferbekov, A. V. Buslaev, and A. Shvets. Ternausnetv2: Fully convolutional network for instance segmentation. *arXiv preprint arXiv:1806.00844*, 2018.

W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi. A survey of deep neural network architectures and their applications. *Neurocomputing*, 234:11–26, 2017.

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*, pages 3431–3440, 2015.

N. K. Mishra and M. E. Celebi. An overview of melanoma detection in dermoscopy images using image processing and machine learning. *arXiv preprint arXiv:1601.07843*, 2016.

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*, pages 234–241. Springer, 2015.

A. Shvets, A. Rakhlin, A. A. Kalinin, and V. Iglovikov. Automatic instrument segmentation in robot-assisted surgery using deep learning. *arXiv preprint arXiv:1803.01207*, 2018.

K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *AAAI*, volume 4, page 12, 2017.

P. Tschandl, C. Rosendahl, and H. Kittler. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data*, 5:180161, 2018. doi: 10.1038/sdata.2018.161.

C. Yu, S. Yang, W. Kim, J. Jung, K.-Y. Chung, S. W. Lee, and B. Oh. Acral melanoma detection using a convolutional neural network for dermoscopy images. *PloS one*, 13(3):e0193321, 2018.

X. Zhang. Melanoma segmentation based on deep learning. *Computer Assisted Surgery*, 22(sup1): 267–277, 2017.