Fully Convolutional Networks to Detect Clinical Dermoscopic Features

Jeremy Kawahara and Ghassan Hamarneh

Medical Image Analysis Lab, Simon Fraser University, Canada

Abstract. We use a pretrained fully convolutional neural network to detect clinical dermoscopic features from dermoscopy skin lesion images. We reformulate the superpixel classification task as an image segmentation problem, and extend a neural network architecture originally designed for image classification to detect dermoscopic features. Specifically, we interpolate the feature maps from several layers in the network to match the size of the input, concatenate the resized feature maps, and train the network to minimize a smoothed negative F1 score. Over the public validation leaderboard of the 2017 ISIC/ISBI Lesion Dermoscopic Feature Extraction Challenge, our approach achieves 89.3% AUROC, the highest averaged score when compared to the other two entries. Results over the private test leaderboard are still to be announced.

1 Introduction

In order to diagnose melanoma from benign lesions, dermatologists often rely on using melanoma-specific related image cues to help their diagnosis. For example, the 7-point checklist [1] is a scoring system that checks for the presence of visual cues (e.g., the presence of streaks) in skin lesions, assigns a numerical score that, if exceeded, indicates the skin lesion is likely melanoma. This is used to help give dermatologist an objective criteria on which to base their diagnosis.

Many groups have studied how to detect and classify clinical dermoscopic features from dermoscopy. Celebi et al. [2] detected the blue-whitish veil in dermoscopy images using decision trees as pixel classifiers. Mirzaalian et al. [3] used the tubular properties of streaks to extract discriminative features that detect streaks. Fabbrocini et al. [4] detected the entire 7-point dermoscopic features using a variety of approaches for each criteria, which included segmentation, extracting color and texture-based features, and training a Logistic Tree Model.

Furthering the research of clinical dermoscopic feature detection, the International Skin Imaging Collaboration (ISIC), in conjunction with the IEEE International Symposium on Biomedical Imaging (ISBI), hosts a skin lesion analysis challenge [5]. This challenge gives an open standardized dataset and metrics to benchmark different approaches on. While this ISIC/ISBI challenge comes in three parts, in this work, we will focus on *Part 2 - clinical feature detection.*

Previous work has shown pretrained convolutional neural networks (CNN) to be useful for skin lesion classification [6,7,8,9], and segmentation [9]. Here we detect clinical features by reformulating the problem as a segmentation task, and finetuning a CNN to detect pixels that contain the studied clinical features.

2 Methods

Given an image x, a corresponding superpixel labelling mask s, and the labelling l for each superpixel in s, our task is, for the i-th superpixel s_i, predict the label l_i. Each label l_i has the following four (potentially overlapping) classes associated with it: Pigment Network; Negative Network; Milia-like Cysts, and Streaks. These are represented as binary vectors of length four (e.g., $l_i = [1, 0, 0, 1]$ indicates the ith superpixel contains pigment network and streaks).

Superpixels to segmentations Rather than treating this as a superpixel labelling problem, we instead model this as a segmentation task. We convert the superpixels s and corresponding labels l into a 3D volume m, where the first two spatial dimensions correspond to the spatial dimensions of x (Fig. 1 *top row*), and the third dimension contains four channels corresponding to the four classes. This allow us to capture the spatial dependencies among the superpixels, and to efficiently leverage pretrained convolutional neural network (CNN).

Segmentations to superpixels While our CNN produces segmentations/pixel predictions (Fig. 1 *bottom left*), our final task is to label superpixels. We convert the predicted segmentation mask $\hat{m}$ back to a predicted superpixel labelling $\hat{s}$ (Fig. 1 *bottom*) by assigning to the i-th superpixel of the c-th class the average predicted probabilities in the corresponding locations in the predicted mask,

$$\hat{s}_i^c = \frac{1}{J} \sum_{j \in L(s_i)} \hat{m}^c(j) \tag{1}$$

where $\hat{m}^c(j)$ is the predicted probability of the c-th class in the j-th spatial location, and $L(s_i)$ returns all J spatial indices corresponding to superpixel s_i.

CNN architecture We modify VGG16 [10], a convolutional neural network, pretrained over ImageNet [11], to act as a pixel-wise classifier. Specifically, we remove the fully-connected layers, and convert the network such that the responses/feature maps throughout the network (from blocks 1-5) are resized to match the size of the input and then concatenated (similar to the proposed architecture in [12]). A $1 \times 1 \times 4$ convolutional filter is added to this concatenated block, which forms our output (i.e., segmentation) for each of the 4 classes. A softmax function is applied element-wise to scale the output between 0 and 1.

Smoothed F1-score loss function The class labels are heavily imbalanced in favour of the background, and even among the classes, some class labels occur much more frequently than others. Thus in order to encourage the CNN to be sensitive to all classes of clinical features, we train the CNN to minimize a modified F1-score:

$$\ell_{F1}(\hat{m}, m) = 1 - \frac{1}{4} \sum_{c=1}^{4} \left(\frac{2\mathrm{TP}(\hat{m}^c, m^c)}{2\mathrm{TP}(\hat{m}^c, m^c) + \mathrm{FP}(\hat{m}^c, m^c) + \mathrm{FN}(\hat{m}^c, m^c) + eps} \right) \tag{2}$$

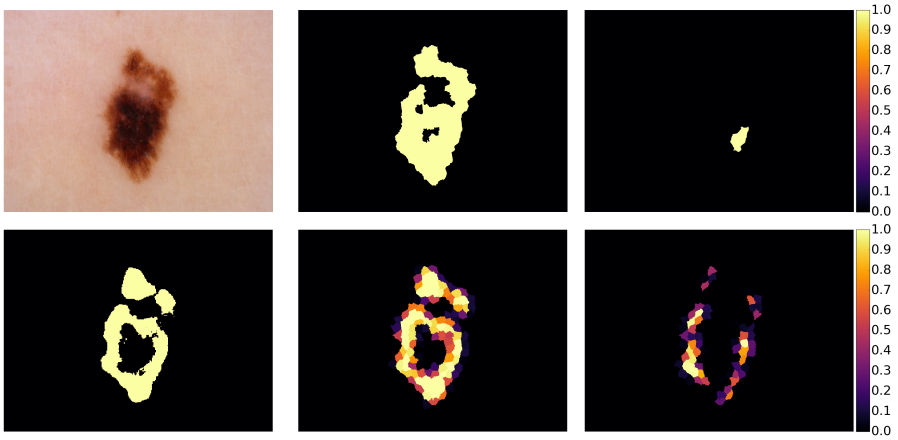

Fig. 1. Superpixels to segmentations, and segmentations to superpixels. (*Top left*) The original image. Expertly annotated pigment-network (*top middle*) and streaks (*top right*) superpixels converted to binary segmentations. (*Bottom left*) Pigment-network pixel-wise predictions from our CNN. Our pigment-network (*bottom middle*) and streak (*bottom right*) pixel-wise predictions converted to superpixels.

where $\hat{m}$ indicates the 3D predicted segmentations, m indicates the ground truth, c indicates a channel corresponding to a particular class labels (m^c is a 2D matrix corresponding to class c), and $eps = 1$ prevents divide by zero errors. TP, FP, and FN, correspond to fuzzy true positives, false positives, and false negatives, as defined in [13], except we sum over pixels. We note that averaging over the classes helps prevent those classes with more positive pixels from dominating the other classes.

Training We train our CNN to minimize Eq. 2. While VGG is trained on images of size 224×224 for classification, we use larger image resolutions of 336×336, which is possible since all our layers are convolutional. We use a mini-batch of size 12 as larger batches exceeded our GPU memory, and stop after only 5 epochs, as more epochs yield segmentations less sensitive to the clinical features. We use a similar training approach and architecture for our segmentation entry (*Part 1*), but minimize a smoothed negative dice score.

3 Results and Conclusions

We trained our network over 1700 images from the ISIC-ISBI skin analysis challenge, and used 300 images to monitor the networks performance with different hyperparameters. The public leaderboard consists of 150 images, with a separate private leaderboard of 600 images.

Our approach, which converted the superpixels to segmentations and trained a CNN finetuned on resized feature maps from early network layers, achieved the highest averaged area under the receiver operator curve (AUROC) over the public validation leaderboard (0.893 vs. 0.848 second place[1]), with the highest AUROC score for pigment network, negative network, and streaks, when compared to the other two entries. Results over the private leaderboard are still to be announced.

Acknowledgments. Thanks to the Natural Sciences and Engineering Research Council (NSERC) of Canada for funding and to the NVIDIA Corporation for the donation of a Titan X GPU used in this research. Thanks to Kathleen Moriarty for helpful discussions and assistance in data preparation.

References

1. Argenziano, G., et al.: Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. Comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. Arch.Dermatol. **134**(12) (1998)
2. Celebi, M.E., et al.: Automatic detection of blue-white veil and related structures in dermoscopy images. CMIG **32**(8) (2008) 670–677
3. Mirzaalian, H., Lee, T.K., Hamarneh, G.: Learning features for streak detection in dermoscopic color images using localized radial flux of principal intensity curvature. MMBIA (2012) 97–101
4. Fabbrocini, G., et al.: Automatic Diagnosis of Melanoma Based on the 7-Point Checklist. In Scharcanski, J., Celebi, M.E., eds.: Computer Vision Techniques for the Diagnosis of Skin Cancer. Springer (2014) 71–107
5. Gutman, D., Codella, N.C.F., Celebi, E., Helba, B., Marchetti, M., Mishra, N., Halpern, A.: Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging 2016. arXiv preprint (2016)
6. Codella, N.C.F., Cai, J., Abedini, M., Garnavi, R., Halpern, A., Smith, J.R.: Deep Learning, Sparse Coding, and SVM for Melanoma Recognition in Dermoscopy Images. In: MICCAI MLMI. Volume 9352. (2015) 118–126
7. Kawahara, J., BenTaieb, A., Hamarneh, G.: Deep features to classify skin lesions. In: IEEE ISBI. (2016) 1397–1400
8. Kawahara, J., Hamarneh, G.: Multi-Resolution-Tract CNN with Hybrid Pretrained and Skin-Lesion Trained Layers. In: MICCAI MLMI, Springer (2016) 164–171
9. Codella, N.C.F., Nguyen, Q.B., Pankanti, S., Gutman, D., Helba, B., Halpern, A., Smith, J.R.: Deep learning ensembles for melanoma recognition in dermoscopy images. IBM Journal of Research and Development **61**(4) (2017)
10. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. ICLR (2015)
11. Russakovsky, O., et al.: ImageNet large scale visual recognition challenge. IIJCV **115**(3) (2015) 211–252
12. Long, J., Shelhamer, E., Darrell, T.: Fully Convolutional Networks for Semantic Segmentation. CVPR (2015)
13. Pastor-Pellicer, J., et al.: F-measure as the error function to train neural networks. In: IWANN. Volume 7902., Springer (2013) 376–384

[1] `https://challenge.kitware.com/#phase/584b0afacad3a51cc66c8e29`