Hybrid Deep Learning for Reflectance Confocal Microscopy Skin Images

Parneet Kaur, Kristin J. Dana
Department of Electrical and Computer Engineering
Rutgers University, NJ, USA
Email: parneet@rutgers.edu, kdana@ece.rutgers.edu

Gabriela Oana Cula, Catherine Mack
Johnson & Johnson
Skillman, NJ, USA
Email: GCula@its.jnj.com, cmack3@its.jnj.com

***Abstract*—Reflectance Confocal Microscopy (RCM) is used for evaluation of human skin disorders and the effects of skin treatments by imaging the skin layers at different depths. Traditionally, clinical experts manually categorize the images captured into different skin layers. This time-consuming labeling task impedes the convenient analysis of skin image datasets. In recent automated image recognition tasks, deep learning with convolutional neural nets (CNN) has achieved remarkable results. However in many clinical settings, training data is often limited and insufficient for CNN training. For recognition of RCM skin images, we demonstrate that a CNN trained on a moderate size dataset leads to low accuracy. We introduce a hybrid deep learning approach which uses traditional texton-based feature vectors as input to train a deep neural network. This hybrid method uses fixed filters in the input layer instead of tuned filters, yet superior performance is achieved. Our dataset consists of 1500 images from 15 RCM stacks belonging to six different categories of skin layers. We show that our hybrid deep learning approach performs with a test accuracy of 82% compared with 51% for CNN. We also compare the results with additional proposed methods for RCM image recognition and show improved accuracy.**

I. Introduction

Reflectance Confocal Microscopy (RCM) is a non-invasive technology that is used to diagnose and study skin cancer, skin aging, pigmentation disorders and skin barrier function by measuring thickness of different skin layers [1], [2]. The effect of skin treatments that modify the proliferation processes within the skin can also be measured through thickness changes. RCM images individual skin layers at different depths by the optical sectioning property of the composite lenses and apertures. A laser is used as the monochromatic light source and tissue penetration is wavelength dependent. As the wavelength increases, the penetration depth also increases; however, higher wavelengths result in tissue damage. A typical penetration depth is 200-250 μm for a wavelength of 800-1024 nm [3]. The light passes through a beam splitter, a focusing lens and is focused on a small tissue spot of skin (few microns) [4]. Each skin layer has different cellular structures causing variation in light reflection, refraction, absorption and transmission. The reflected light passes through an objective lens, a pinhole filter and is imaged at the photo-detector. Captured images of each skin layer have an observable image texture that is unique for each skin layer as shown in Figure 1.

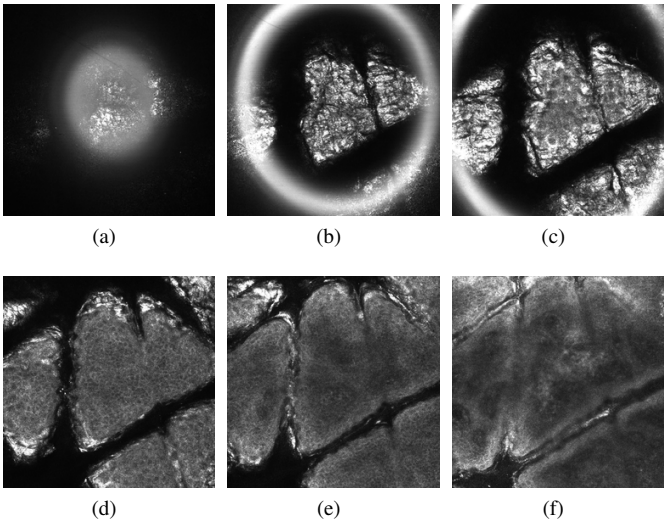
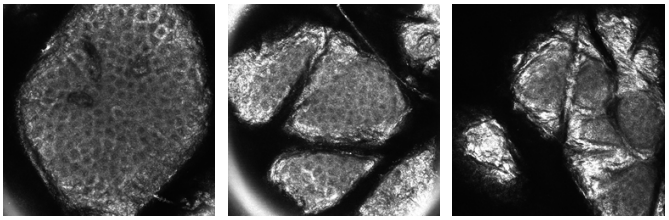


Fig. 1: **Skin layers in an RCM image stack.** (a) Outside epidermis (OE). (b) stratum corneum (SC). (c) stratum granulosum (SG). (d) stratum spinosum (SS). (e) stratum basale (SB). (f) portions of the papillary dermis (PD).

RCM captures the cellular details at a spatial resolution that is comparable to histopathology, which is an invasive, painful and time consuming procedure. Using RCM, a series of images are acquired at the same position, from epidermis through upper dermis and are collectively called a stack. The epidermis is divided into four sub-layers: stratum corneum (SC), stratum granulosum (SG), stratum spinosum (SS) and stratum basale (SB) ((Figure 1(b)-(e)). Portions of papillary dermis are also imaged (Figure 1(f)). In addition, certain images are categorized as outside the dermis (OD) (Figure 1(a)).

Traditionally, these high resolution stack images are labeled based on visual observation by clinical experts. This manual labeling requires significant time and also results in labeling variability depending on the expertise of the clinical grader. We propose an automated method based on detecting skin features and training a mutilayer neural network. Our method is a hybrid of classic methods in texture recognition called *texton-based recognition* and recent methods of *deep learning*. Our results indicate that this hybrid approach gives higher



(a) (b) (c)

Fig. 2: **Intra-Class variation**. Images of the same class from different RCM stacks have similar texture but the overall structure can change significantly and therefore the recognition problem is quite challenging.

recognition rates than CNN deep learning methods and other recent texture recognition methods. Moreover, in situations where moderate size image sets are available (thousands as opposed to millions) the results are significantly better (higher accuracy) and faster than the competing approaches.

A key component in CNN research is the ability to obtain optimal filters with automatically derived features. However, the lowest level of these networks look very similar to traditional filter banks. The CNN methods were developed for the task of general image recognition [5], [6]. For RCM labeling, the task is texture recognition where the number of classes is small and the image variation is subtle. Texture recognition methods have traditionally relied on multiscale, multi-orientation, gradient-like features called textons that are paralleled in the human visual system [7], [8], [9], [10], [11], [12]. Texton-based features have the advantage that they are computed from fixed weight filters and need not be trained. More recently texture recognition has been addressed using deep learning and CNN [13]. We develop a hybrid approach that uses texton-based features as input to train a deep neural network in order to gain the advantages of both traditional and recent texture recognition frameworks.

Our dataset comprises of 1500 skin images from 15 RCM stacks, each image belongs to one of the skin layers as shown in Figure 1 where labels were obtained by clinical skin experts. Figure 2 shows images of the same class but from different RCM stacks (one stack corresponds to the images obtained by varying the depth through one skin sample). Notice that the texture within each class is similar, but the overall structure can change significantly and therefore the recognition problem is quite challenging.

We demonstrate that our hybrid deep learning approach performs with a test accuracy of 81.73%. In most stacks, mislabeling is observed between adjoining layers as the images transition from one skin layer to another. These transitions may be ambiguous during manual labeling as well.

II. Related Work

Computational analysis of RCM skin images has been used for automatic detection of tumors [14], detecting malignant features for superficial spreading melanoma [15] and skin aging assessment [16]. Texture of RCM images has also been analyzed to identify melanocytic skin lessions at dermal-epidermal junction [17] using Speeded Up Robust Features (SURF) to capture the texture of localized features and use Support Vector Machine (SVM) classifier to distinguish between them.

Automated methods to categorize RCM stack images into skin layers has been recently proposed in [18], [19], [20], [21]. Delineation of dermal-epidermal junction is presented in [18] but is limited to categorizing dermis and epidermis skin layers based on the difference in their contrast. The sub-layers in epidermis are not identified. To categorize each image in the RCM stack as one of the skin layers, a fully unsupervised texton-based approach is presented in [19]. A texton library is created by using a filter bank and then projecting the filter response to a lower dimensional subspace using principal component analysis. The texton histogram of images are projected to a lower dimensional subspace and clustered into five skin layers using k-means clustering results. A high correlation is reported between the ground truth and obtained labels. However, only three stacks are used in these experiments for evaluation.

Bag-of-features representation of images has been used in computer vision for tasks such as object or scene recognition [22], [23]. Local interest features of iconic patches are used to build a dictionary and each image is represented by frequency of each visual word in the dictionary. In [20], the authors apply a bag-of-features approach and use logistic regression classifier. A representative dictionary is built by combining hierarchical and k-means clustering for normalized 7×7 patches. In [21], conditional random fields are followed by structured SVM as a classifier instead of logistic regression and shows improved performance. For our approach, we instead use a deep-learning framework that is generally more powerful in discrimination when compared to the SVM classifier. We also empirically demonstrate that the filter-based texton histogram is a better feature to classify RCM skin images compared to patch-based features.

Perceptual attributes have been used in describable texture database where each image is assigned several attributes based on perception, inspired by the human vision [13]. In [24] perceptual attributes are used to categorize macroscopic skin textures. We extend this method by using the attribute histograms to train another classifier which can be used to label RCM images. However, we observe that even though this approach provides pixel level attribute labels, the test image accuracy using attribute histograms for training a neural network on the RCM data is relatively low (74.94%). Further details about this method are given in Secion III-C.

Convolutional Neural Networks (CNNs) have been used successfully for several computer vision tasks such as image classification, video analysis and object recognition [13], [6], [25]. CNNs learn a hierarchy of features for classification automatically from a large set of input images (in millions). We demonstrate that a CNN trained on a moderate size dataset results in low test accuracy. The CNN architecture

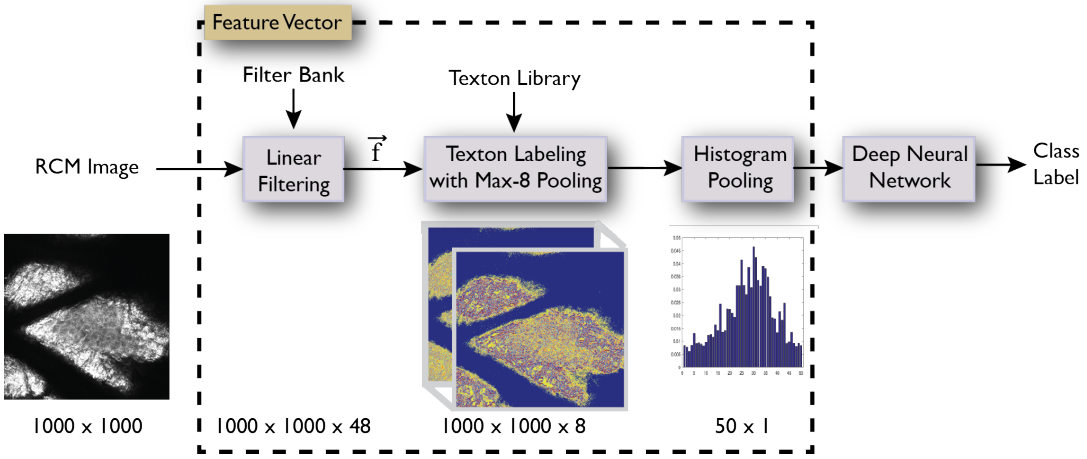


Fig. 3: **Hybrid Deep Learning**. Texton-based feature vectors are obtained by using a pre-built texton library. Patches (5×5) centered at each pixel of the RCM image are labeled with 8 closest textons in the texton library. These texton labels are pooled to obtain a texton histogram which is used as input to train a deep neural network with multiple layers.

we use is presented in Section III-D. In RCM skin images, the differences in the textures of various categories are very subtle (Figures 1) and the datasets are relatively small, which makes skin classification a challenging problem.

III. Methods

A. Imaging

The acquisition of RCM stacks is done using a Vivascope 1500 (Lucid Technologies, Rochester, NY, USA) using 785 nm laser illumination. The image stack was collected up to the depth of 100μm, at a step size of 1μm. The dimensions of the images are 1000×1000 pixels.We collected a dataset consists of 15 stacks, with 100 images in each stack. Each image is labeled by a human dermatology expert as one of the following skin layer category as illustrated in Figure 1: Outside epidermis (OE), stratum corneum (SC), stratum granulosum (SG), (d) stratum spinosum (SS), (e) stratum basale (SB), (f) portions of the papillary dermis (PD).

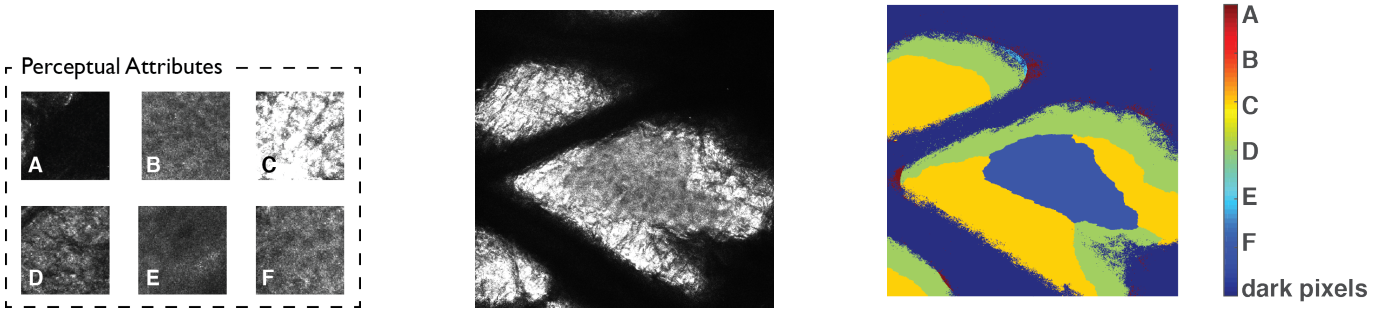
B. Hybrid Deep Learning

Our hybrid deep learning approach combines the unsupervised texton-based approach with supervised deep neural networks. It consists of the following layers as shown in Figure 3:

1) *Convolution layer:* We use a fixed-weight filter bank with 48 filters. These filters include 36 first and second order derivative of Gaussain filters (6 orientations, 3 scales each), 8 Laplacian of Gaussain filters and 4 Gaussian filters ([8]). Each pixel is filtered over a 5×5 region and represented by a 48-dimensional vector.
2) *Texton labeling with Max-8 Pooling:* Patches (5×5) centered at each pixel of the skin image are labeled using a pre-built texton library. The texton library is obtained by clustering the 48-dimensional filter outputs of a random sampling of skin images over 5×5 region into T clusters. For our experiments we use k-means clustering with $T = 50$ clusters. The texton library needs to be built only once. Using the texton library, 48-dimensional filtered output of each pixel is mapped to its 8 closest textons from the texton library. Each pixel is associated with 8 nearest cluster centers resulting in 8 texton maps for each RCM image. We associate each pixel to 8 clusters instead of 1 so that the effect of the similar filter responses being assigned to neighboring textons is nullified in next layers [26].
3) *Histogram Pooling:* The textons labels from the 8 texton maps are pooled together by weighing them their distance. For each pixel, p, the histogram bin h corresponding to its texton labels is updated as:

$$h(t) = h(t) + \left[1 - \frac{d(i)}{\sum_{i=1}^{8} d(i)}\right], \qquad (1)$$

where $d(i)$ is the distance of the i^{th} closest texton to the filtered output at pixel p, t is the i^{th} closest centroid in the texton library for the filtered output at pixel p. Texton labels corresponding to dark pixels are ignored and improve the classifier performance.
4) *Deep Neural Network:* We use a feed-forward deep neural network, which consists of an input layer, two hidden layers and an output layer [27]. The input and output layers have 50 and 6 neurons, respectively. The two hidden layers have 40 and 10 neurons, respectively. Tan-sigmoid function is used for activation of the neurons. The network parameters were tuned empirically using the training data. Adding more hidden layers or neurons in each hidden layer does not improve the performance for this dataset. An advantage of using the deep neural network is that it returns probability estimates for all the class instead of a single class label, which can also be easily converted to class labels using a linear transfer



(a) (b) (c)

Fig. 4: **Attributes-based approach**. (a) Perceptual Attributes. Exemplars of typical visual appearance (150×150 patches) in RCM skin images. (b) Input RCM skin image. (c) A patch centered at each pixel is labeled as one of the perceptual attributes.

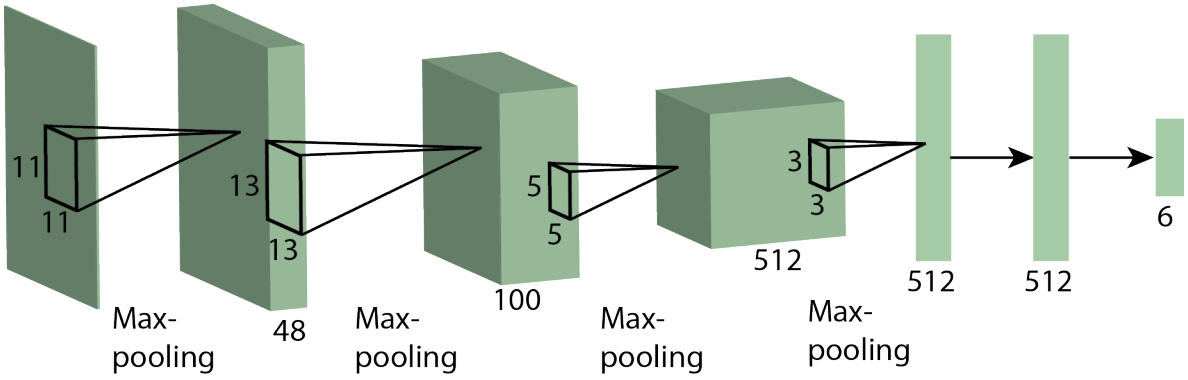


Fig. 5: **Our CNN Architecture** consists of four convolutional layers with kernels of sizes 11, 13, 5 and 3. Each convolutional layer is followed by a max-pooling layer. The last layer of the network is a fully-connected layer with 512 neurons which gives output labels of the test image.

function at the output layer. The texton histogram of training data is used to train the network, which can then be used to categorize the test skin images into skin layers.

C. Attribute-based approach

Perceptual attributes are used in [24] to classify macroscopic skin textures. Figure 4(a) shows exemplars of typical visual appearance in RCM skin images. Patches of size 150×150 are randomly sampled from the training images and used to create a labeled attribute dataset of 40000 training and 20000 test samples. Texton histograms of each attribute patch are used to train a neural network classifier. The trained attribute classifier performs with an accuracy of 97.2% on the test set. For test RCM images, a patch centered at each pixel of the RCM image can be labeled as one these perceptual attributes as shown in Figure 4(c). Integral histograms are used for computational speed of the input feature vectors corresponding to each pixel [28], [29]. We extend this method by using the histograms of the attributes labeled RCM image to train another neural network with an input layer, a hidden layers and an output layer with 6, 20 and 6 neurons, respectively.

D. Convolutional Neural Networks:

A convolutional neural network (CNN) typically consists of multiple convolutional, pooling and rectilinear linear units followed by a fully connected layer [5]. Our CNN architecture as shown in Figure 5 consists of following layers: The input RCM images are resized to 250×250 and given as input to the first convolutional layer which consists of 48 kernels of size $11 \times 11 \times 1$ with a stride of 1 pixel. The filter output is followed by a max-pooling layer with stride of 4 pixels. The second convolutional layer takes the pooled output of first layer as input and filters it with 100 kernels of size 13×13. The third convolutional layers is connected to the pooled output of the previous layer with 512 kernels of size 5×5. The fourth convolutional layer has filter size 3×3. Finally, the fully-connected layers have 512 neurons each with 6 output neurons. We tried different variations of the CNN layers but they resulted in similar performance on the test data.

IV. Experiments and Results

To evaluate the performance of algorithms for RCM skin classification, we perform k-fold cross-validation with $k = 5$ on our dataset. For each iteration, the training consists of 12 RCM stacks and the classifier is tested on three RCM

	Accuracy ± standard deviation	Sensitivity	Specificity	Precision	F-score
Unsupervised texton-based approach [18]	0.5395 ± 0.03	0.5270	0.90	0.5422	0.4883
Patch based bag-of-features approach followed by SVM classifier [20]	0.5413 ± 0.25	0.3811	0.9020	0.5128	0.3721
Patch based bag-of-features approach followed by a Neural Network classifier	0.7993 ± 0.05	0.6918	0.9587	0.6990	0.6842
Convolutional Neural Network	0.5134 ± 0.07	0.20	0.8392	0.60	0.1653
Attribute approach (extension of [23])	0.7120 ± 0.04	0.5564	0.9392	0.5711	0.5520
Texton approach followed by SVM classifier	0.75 ± 0.03	0.6102	0.9482	0.6346	0.6013
Hybrid deep learning	**0.8173 ± 0.05**	**0.7174**	**0.9620**	**0.7236**	**0.7104**

TABLE I: **Comparison of different approaches for RCM skin image classification.** Our proposed hybrid deep learning outperforms other methods.

stacks. Each RCM stack consists of 100 images. The ground truth for each image is obtained through manual labeling by a clinical expert. Each image is classified as one of the six skin layers (Figure 1). The following performance metrics are computed in each fold using the confusion matrix for the test sets: accuracy, sensitivity, specificity, precision and f-score [30]. Table I lists the average performance over five folds. Our proposed hybrid deep learning method outperforms other methods. Using the hybrid deep learning approach in Section III-B, average accuracy on test sets is 0.8173 (or 81.73%). If we replace the multi-layer neural network by SVM classifier, the accuracy reduces to 0.75.

The perceptual attribute based approach described in Section III-C gives pixel level attribute labels. An example of a test image labeled with perceptual attributes is shown in Figure 4. The histogram of attribute labels given as a feature to a multi-layer neural network results in an accuracy of 0.7120. The CNN architecture proposed in Section III-D performs poorly with average test accuracy of 0.5147. We designed the architecture empirically and tried variations of different layers and kernel sizes. The network was implemented in MATLAB using MatConvNet library [27], [31]. The CNNs learn the kernel weights automatically from the labeled training data. Since our datasets is moderate size, it may be difficult for the network to converge and learn the correct filter weights. Further, the higher layers of CNN also learn complex features apart from the texture such as external contour shape, which is not useful for skin texture classification of RCM images.

We also compare our method to the unsupervised texton based approach in [19] which gives an accuracy of 0.5395%. The patch based bag-of-features approach proposed in [21] followed by a SVM classifier and deep neural network give an accuracy of 0.5413 and 0.7993, respectively. In addition to accuracy, the hybrid deep learning method also results in the best sensitivity, specificity, precision and f-score among all the methods.

In most stacks, the mislabeling is observed between adjoining layers as the images transition from one skin layer to another. Table II shows the confusion matrix for all the test stacks. Figure 6 shows examples of mislabeled images in the first column. The correctly labeled images are shown in the second and third columns. Figure 6(a) was labeled as SC by the algorithm but as OE by the human expert. Figures 6(b) and (c) show examples correctly labeled by the algorithm as SC and OE, respectively. Such transitions may be ambiguous to a clinical expert as well.

Output Class \ Input Class	OE	SC	SG	SS	SB	PD
OE	130	25	0	0	0	1
SC	14	67	17	5	0	0
SG	0	22	53	18	0	0
SS	2	1	20	146	37	0
SB	0	0	2	32	89	42
PD	6	0	0	1	33	737

TABLE II: **Confusion Matrix.** Note that the mislabeling is between adjoining skin layers. Also see examples in Figure 6.

V. Conclusions

In this paper we introduce a hybrid deep learning approach for automatically labeling RCM skin images. Texton-based features are obtained using a fixed multi-resolution, multi-orientation filter bank to train a deep neural network. We compare our method with a suite of texture recognition methods and show that it outperforms the state-of-the-art with a test accuracy of 81.73%. We demonstrate that smaller training datasets are insufficient for CNN training and feature extraction is essential is such cases.

Acknowledgment

The authors gratefully acknowledge the support received for this project from Johnson and Johnson Consumer Products Research and Development.

References

[1] R. Hofmann-Wellenhof, G. Pellacani, J. Malvehy, and H. P. Soyer, *Reflectance confocal microscopy for skin diseases*. Springer Science & Business Media, 2012.

[2] M. Ulrich and S. Lange-Asschenfeldt, "In vivo confocal microscopy in dermatology: from research to clinical application," *Journal of biomedical optics*, vol. 18, no. 6, pp. 061 212–061 212, 2013.

[3] J. L. S. Sánchez-Mateos, C. M. G. del Real, P. J. Olasolo, and S. González, "Reflectance-mode confocal microscopy in dermatological oncology," *Lasers in Dermatology and Medicine*, pp. 285–308, 2011.

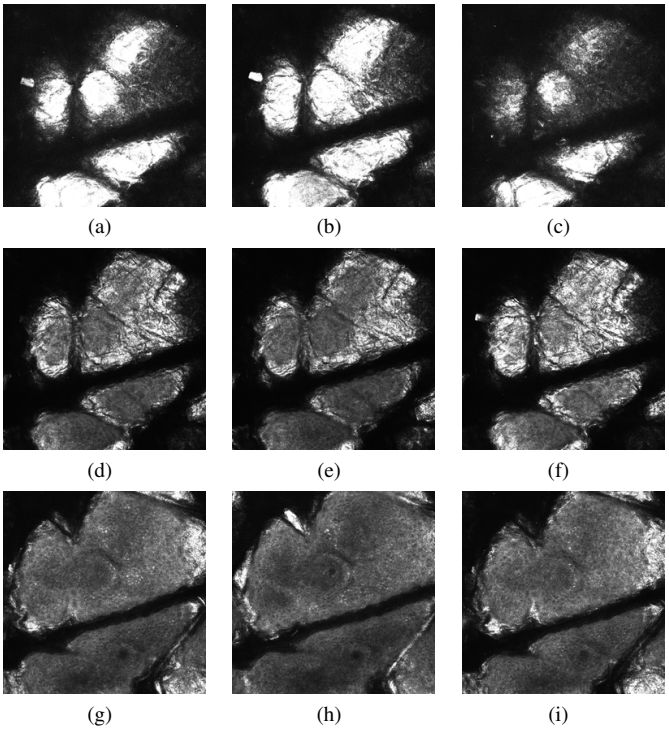


Fig. 6: Ambiguity of appearance in transition regions between skin layers. *First Column:* Examples of mislabeled images. (a) Human label:OE, Automated label:SC. (d) Human label:SC, Autometed label:SS. (g) Human label:SS, Automated label:SB. *Second Column:* Correcty labeled: (b) SC (e) SS (h) SB. *Third Column:* Correcty labeled: (c) OE (f) SC (i) SS.

[4] P. Calzavara-Pinton, C. Longo, M. Venturini, R. Sala, and G. Pellacani, "Reflectance confocal microscopy for in vivo skin imaging," *Photochemistry and photobiology*, vol. 84, no. 6, pp. 1421–1430, 2008.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[6] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2014, pp. 1725–1732.

[7] O. G. Cula and K. J. Dana, "Recognition methods for 3d textured surfaces," *Proceedings of SPIE Conference on Human Vision and Electronic Imaging VI*, vol. 4299, pp. 209–220, January 2001.

[8] T. Leung and J. Malik, "Representing and Recognizing the Visual Appearance of Materials using Three-dimensional Textons," *Int. J. Comput. Vision*, vol. 43, no. 1, pp. 29–44, June 2001.

[9] O. G. Cula and K. J. Dana, "Compact representation of bidirectional texture functions," vol. 1, pp. 1041–1067, December 2001.

[10] C. Schmid, "Constructing models for content-based image retrieval," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 2. IEEE, 2001, pp. II–39.

[11] G. O. Cula, J. K. Dana, P. F. Murphy, and K. B. Rao, "Skin texture modeling," *International Journal of Computer Vision*, vol. 62, no. 1, pp. 97–119, 2005.

[12] M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," *International Journal of Computer Vision: Special Issue on Texture Analysis and Synthesis*, vol. 62, no. 1–2, pp. 61–81, April 2005.

[13] M. Cimpoi, S. Maji, and A. Vedaldi, "Deep filter banks for texture recognition and segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[14] S. Koller, M. Wiltgen, V. Ahlgrimm-Siess, W. Weger, R. Hofmann-Wellenhof, E. Richtig, J. Smolle, and A. Gerger, "In vivo reflectance confocal microscopy: automated diagnostic image analysis of melanocytic skin tumours," *Journal of the European Academy of Dermatology and Venereology*, vol. 25, no. 5, pp. 554–558, 2011.

[15] D. Gareau, R. Hennessy, E. Wan, G. Pellacani, and S. L. Jacques, "Automated detection of malignant features in confocal microscopy on superficial spreading melanoma versus nevi," *Journal of biomedical optics*, vol. 15, no. 6, pp. 061 713–061 713, 2010.

[16] A. P. Raphael, T. A. Kelf, E. M. Wurm, A. V. Zvyagin, H. P. Soyer, and T. W. Prow, "Computational characterization of reflectance confocal microscopy features reveals potential for automated photoageing assessment," *Experimental dermatology*, vol. 22, no. 7, pp. 458–463, 2013.

[17] K. Kose, C. Alessi-Fox, M. Gill, J. G. Dy, D. H. Brooks, and M. Rajadhyaksha, "A machine learning method for identifying morphological patterns in reflectance confocal microscopy mosaics of melanocytic skin lesions in-vivo," in *SPIE BiOS*. International Society for Optics and Photonics, 2016, pp. 968 908–968 908.

[18] S. Kurugol, K. Kose, B. Park, J. G. Dy, D. H. Brooks, and M. Rajadhyaksha, "Automated delineation of dermal–epidermal junction in reflectance confocal microscopy image stacks of human skin," *Journal of Investigative Dermatology*, vol. 135, no. 3, pp. 710–717, 2015.

[19] E. Somoza, G. O. Cula, C. Correa, and J. B. Hirsch, "Automatic localization of skin layers in reflectance confocal microscopy," in *Image Analysis and Recognition*. Springer, 2014, pp. 141–150.

[20] S. Hames, M. Ardigò, H. P. Soyer, A. P. Bradley, and T. W. Prow, "Automated segmentation of skin strata in reflectance confocal microscopy depth stacks," *bioRxiv*, p. 022137, 2015.

[21] S. C. Hames, M. Ardigo, H. P. Soyer, A. P. Bradley, and T. W. Prow, "Anatomical skin segmentation in reflectance confocal microscopy with weak labels," in *Digital Image Computing: Techniques and Applications (DICTA), 2015 International Conference on*. IEEE, 2015, pp. 1–8.

[22] J. Wu and J. Rehg, "Beyond the euclidean distance: Creating effective visual codebooks using the histogram intersection kernel," in *Computer Vision, 2009 IEEE 12th International Conference on*, 2009, pp. 630–637.

[23] L. Fei-Fei and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2, June 2005, pp. 524–531 vol. 2.

[24] P. Kaur, K. Dana, and G. Cula, "From photography to microbiology: Eigenbiome models for skin appearance," in *BioImage Computing Workshop, Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*. IEEE, 2015.

[25] A. Dosovitskiy, J. T. Springenberg, M. Riedmiller, and T. Brox, "Discriminative unsupervised feature learning with convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2014, pp. 766–774.

[26] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context," *International Journal of Computer Vision*, vol. 81, no. 1, pp. 2–23, 2009.

[27] *MATLAB version 8.5 (R2015a) and Neural Network Toolbox*, The Mathworks, Inc., Natick, Massachusetts, 2015.

[28] F. Porikli, "Integral histogram: a fast way to extract histograms in cartesian spaces," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, June 2005, pp. 829–836 vol. 1.

[29] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," 2008.

[30] T. Fawcett, "An introduction to roc analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.

[31] A. Vedaldi and K. Lenc, "Matconvnet: Convolutional neural networks for matlab," in *Proceedings of the 23rd Annual ACM Conference on Multimedia Conference*. ACM, 2015, pp. 689–692.