Combining Convolutional and Recurrent Neural Networks for Human Skin Detection

Haiqiang Zuo, Heng Fan, Erik Blasch, *Senior Member, IEEE*, and Haibin Ling, *Member, IEEE*

***Abstract*—Skin detection from images, typically used as a preprocessing step, has a wide range of applications such as dermatology diagnostics, human computer interaction designs, and etc. It is a challenging problem due to many factors such as variation in pigment melanin, uneven illumination, and differences in ethnicity geographics. Besides, age and gender introduce additional difficulties to the detection process. It is hard to determine whether a single pixel is skin or nonskin without considering the context. An efficient traditional hand-engineered skin color detection algorithm requires extensive work by domain experts. Recently, deep learning algorithms, especially convolutional neural networks (CNNs), have achieved great success in pixel-wise labeling tasks. However, CNN-based architectures are not sufficient for modeling the relationship between pixels and their neighbors. In this letter, we integrate recurrent neural networks (RNNs) layers into the fully convolutional neural networks (FCNs), and develop an end-to-end network for human skin detection. In particular, FCN layers capture generic local features, while RNN layers model the semantic contextual dependencies in images. Experimental results on the COMPAQ and ECU skin datasets validate the effectiveness of the proposed approach, where RNN layers enhance the discriminative power of skin detection in complex background situations.**

***Index Terms*—Convolutional neural networks (CNNs), recurrent neural networks (RNNs), skin classification, skin detection, skin segmentation.**

I. Introduction

SKIN detection is the process of discriminating skin and nonskin pixels in a digital image, which plays an important role in multifarious applications ranging from face detection [1], gesture recognition [2], dermatology diagnostics [3], driver fatigue detection [4], human computer interaction [5], [6] to a variety of computational health informatics [7].

Skin detection is a challenging problem and has drawn extensive attention in the literature. In many of the popularly used color spaces, there is a significant overlap between the skin and nonskin pixels. For example, numerous objects in the background such as wall, wood, and cloth could have similar color to different types of human skin. It is difficult to determine whether a single pixel is skin or nonskin without considering the neighboring pixels. Besides, the performance of skin detection is also influenced by a variety of other factors (uneven illumination, camera characteristics, subject ethnicity, age, gender, so on) [8]. In the past few years, numerous approaches have been proposed to address these issues. Recent studies have focused on different color spaces (e.g., RGB [9], YCbCr [10], CIE-XYZ [11], HSV [12], and SKN [13]), feature extraction (such as color [14], texture [15], and spatial distribution [16]), and classification methods (Bayesian classifier[9], Gaussian mixture model [17], support vector machine[18], neural networks [19], random forest [20], so on). To design an efficient skin color classifier, traditional hand-engineered approaches require extensive work by domain experts.

Convolutional neural networks (CNNs) have recently achieved remarkable results for a variety of computer vision tasks, such as image classification and object detection [21]. Now, there is an increasing interest in pixel-wise prediction [22]–[27] using deep learning architectures instead of handcrafting methods. However, CNN-based architectures are not good at modeling the relationship between pixels and their neighbors. Recently, Zheng *et al.* [27] introduced conditional random fields at the final layer of CNN to refine coarse predictions.

Recurrent neural networks (RNNs) have shown great success in many natural language processing [28] tasks. Unlike CNNs, which assume that all inputs and outputs are independent of each other, the outputs of RNNs depend on previous computations. Researchers have enthusiastically applied RNNs to computer vision applications, such as scene labeling [29]–[37]. On the basis of the foregoing studies, in this letter we introduce RNN layers into fully convolutional neural networks (FCNs), and develop an end-to-end network for human skin detection. Specifically, we use FCN layers to capture generic local features and RNN layers to model the semantic contextual dependencies in images. Our framework is illustrated in Fig. 1.

The remainder of this letter is organized as follows. Section II introduces the framework of our proposed human skin detection algorithm. Experimental results are demonstrated in Section III. Finally, conclusions are presented in Section IV.

II. Problem Statement

RNNs are traditionally specialized in modeling temporal dependencies but recently have shown accuracy improvement for spatial counterparts [36], [37]. In detail, conventional chain-structured RNNs are not applicable to images where spatial

Manuscript received November 30, 2016; revised January 14, 2017; accepted January 15, 2017. Date of publication January 17, 2017; date of current version February 8, 2017. This work was supported in part by the U.S. National Science Foundation under Grant 1449860 and Grant 13505210, in part by the National Natural Science Foundation of China under Grant 61103056 and Grant 61528204, and in part by the China Scholarship Council. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Edmund Y. Lam.

H. Zuo is with the Department of Chemical Equipment and Control Engineering, China University of Petroleum, Qingdao 266580, China and also with the Department of Computer and Information Sciences, Temple University, Philadelphia, PA 19122 USA (e-mail: zhqupc@upc.edu.cn).

H. Fan and H. Ling are with the Department of Computer and Information Sciences, Temple University, Philadelphia, PA 19122 USA (e-mail: hengfan@temple.edu; hbling@temple.edu).

E. Blasch is with the Air Force Research Laboratory, Rome, NY 13441 USA (e-mail: erik.blasch@rl.af.mil).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2017.2654803

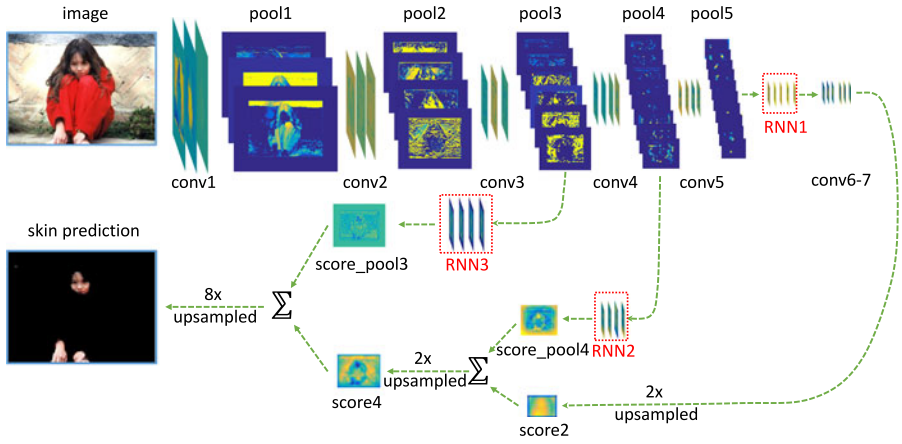

Fig. 1. Framework of human skin detection with FCN and RNN. For simplicity, some intermediate layers such as ReLU are omitted.

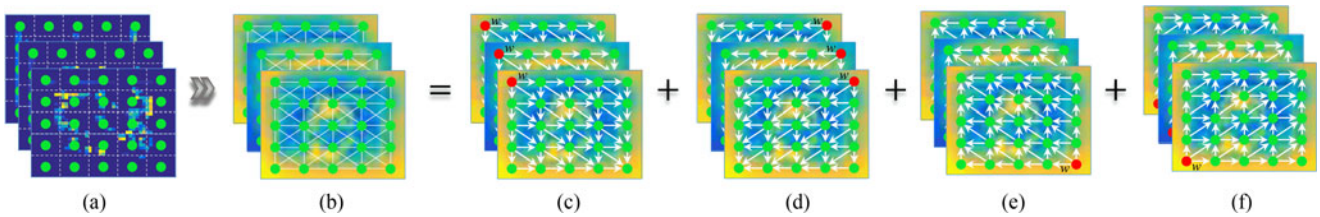

Fig. 2. Structure of RNNs. The undirected cyclic graph is decomposed into four directed acyclic graphs. The green solid dots represent image units, red solid dots are start points in directed acyclic graphs, and w denotes the weight.

relationships among pixels are not available to neighbors in the chain. When RNNs are transformed to the visual domain, undirected cyclic graphs are used to model the interactions among image units. However, undirected cyclic graph cannot be directly used to process the acyclic structure data. The topological structure of an undirected graph is approximated by the combination of four directed acyclic graphs, each with a different starting point, as shown in Fig. 2.

Specifically, each image layer can be represented by a directed cyclic graph $G = \{V, E\}$, where $V = \{v_i\}_{i=1}^{N}$ denotes the vertex set and $E = \{e_{ij}\}$ is the edge set, where e_{ij} represents an edge from v_i to v_j. The structure of a hidden layer h follows the same topology as G. A forward propagation sequence can be seen as traversing G from a start point following a certain path. For vertex v_i, its hidden layer $h^{(v_i)}$ is expressed as a nonlinear function over current input $x^{(v_i)}$ and summarization of hidden layers of all its predecessors. Following [36], we decompose an undirected graph G^u into four directed acyclic graphs $G^u = \{G_d\} = \{G_{\mathrm{SE}}, G_{\mathrm{SW}}, G_{\mathrm{NW}}, G_{\mathrm{NE}}\}$ in southeast, southwest, northwest, and northeast context propagation directions, respectively. Then, forward and backward propagations are applied independently to each subgraph and the corresponding hidden layer h_d is obtained. The summation over all independent hidden layers yields the output layer o. Specifically, these operations can be mathematically formulated by

$$\begin{cases} h_d^{(v_i)} = \phi\left(U_d x^{(v_i)} + \sum_{v_j \in P_{G_d}(v_i)} W_d h_d^{(v_j)} + b_d\right) \\ o^{(v_i)} = \sigma\left(\sum_{G_d \in G^u} V_d h_d^{(v_i)} + b_o\right) \end{cases} \tag{1}$$

where U_d, W_d, and V_d are weight matrices, b_d and b_o are bias terms, $P_{G_d}(v_i)$ denotes the direct predecessor set of v_i in G_d, and $\phi(\cdot)$ and $\sigma(\cdot)$ are element-wise nonlinear activation functions.

TABLE I
CONFUSION MATRIX OF BINARY SKIN DETECTION PROBLEM

	Predicted as skin	Predicted as nonskin
Ground-truth skin	True positive (TP)	False negative (FN)
Ground truth Nonskin	False positive (FP)	True negative (TN)

Next, we integrate the RNN into the fully convolution network [24]. Since the fully connected layers of traditional CNNs throw away spatial information, these layers are converted into convolution layers by setting kernels that cover their entire input regions. Doing so casts them into fully convolutional networks. In order to provide further precision, skips (shortcut connections) are added to combine the final prediction layer with lower layers. By fusing predictions from pool3, at stride 8, with a $2\times$ upsampling of predictions fused from pool4 and conv7, the final segmentation FCN-8s network is built.

We insert the aforementioned RNN layer after pool3, pool4, and pool5, respectively, in the FCN-8s network. The directed acyclic graphs of RNN layers directly take the outputs of preceding pooling layer as input. The dimensions of RNN layers are set equal to channels of corresponding preceding pooling layers. We change the output channel dimension of all the score and final upsampling layers to two to predict scores for binary output (0 for nonskin and 1 for skin). To train the network, the loss L is computed with

$$L = -\frac{1}{N} \sum_{v_i \in G^u} \sum_{j=0}^{1} \log(y_j^{(v_i)} Y_j^{(v_i)}) \tag{2}$$

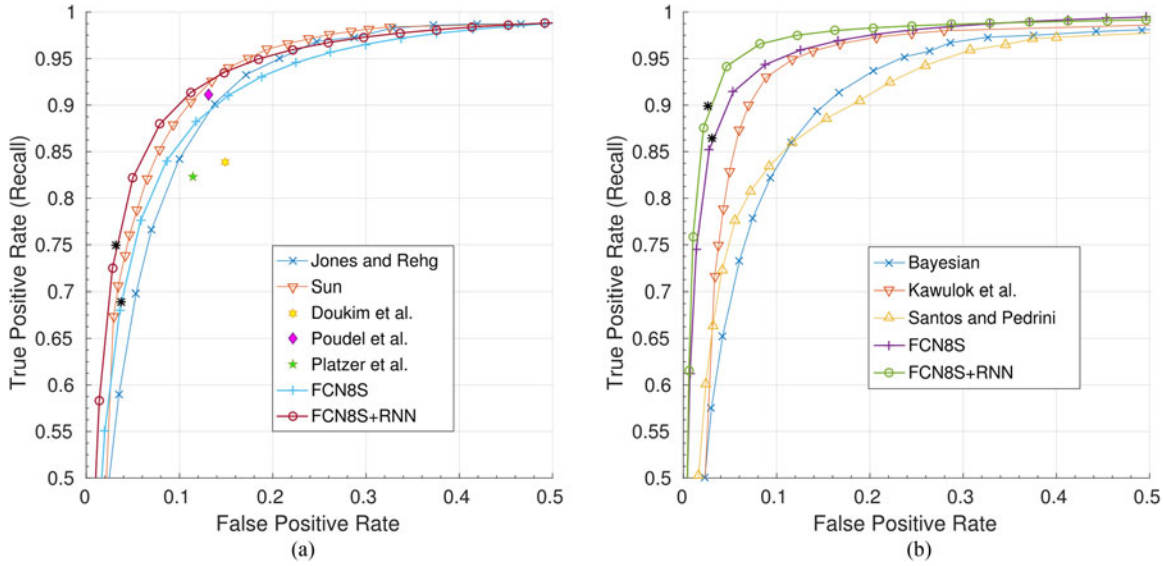

Fig. 3. ROC curves of skin detection on the COMPAQ (a) and ECU (b) datasets. The asterisk "*" represents the optimal operating point for corresponding curve. (a) COMPAQ (b) ECU.

where N denotes the number of image units, $j = 0$ means the negative class, $j = 1$ means the positive class, $y^{(v_i)}$ represents class likelihood vector, and $Y^{(v_i)}$ is the binary label indicator vector for the image unit at v_i. Fig. 1 demonstrates the overall framework of our human skin detection algorithm.

III. Experiments and Analysis

A. Experimental Setup

The algorithm is trained and tested with MatConvNet [38] under Matlab environment on a NVIDIA GeForce GTX TITAN X with 12 GB GPU memory. The networks can take input of arbitrary size and produce detection result with the same size of input. During the experiment, all images and their corresponding ground-truth labels are resized to 360×480 pixels by bilinear interpolation. It is still an open problem on what color space is the best for skin detection, and some authors have demonstrated that the performance is somewhat independent of the color space [8]. Therefore, in this letter we use the common RGB color space.

We initialize and update the FCNs with the publicly available pretrained weights of the FCN-8s network, which was trained on the PASCAL VOC 2011 dataset and Microsoft COCO by Zheng *et al.* [27]. The weights of RNN layers are initialized with normally distributed random numbers. We train the network by stochastic gradient descent with momentum. The implementation uses a fixed learning rate of 10^{-7}, momentum of 0.9, and doubled learning rate for biases. Results are reported after 20 training epochs. At test phase, it takes about 55 ms to process an image with size 360×480 pixels on our computer.

B. Dataset

In order to facilitate evaluation and comparison, two widely used and publicly available skin datasets (COMPAQ [9] and ECU [39]) are used in this letter. The COMPAQ dataset was released by Jones and Rehg [9] in 2002 which contains 13 634 images (4670 skin images and 8964 nonskin images). Only 4670 skin images are used in our experiments, and among them 976 images are randomly selected as the testing set, and the remaining are taken as the training set. The ECU dataset was developed by Phung *et al.* [39] in 2005, which consists of 4000 images. The images are divided into 1000 images for testing and 3000 images for training. Both datasets are provided with ground-truth skin masks representing skin regions and have versatile attributes (skin-like background, uneven illumination, camera characteristics, ethnicity, age, gender, so on).

C. Experimental Results

Skin detection is a binary classification problem, in which the skin pixels are labeled as positive and nonskin pixels as negative. The confusion matrix is provided in Table I. A plethora of sophisticated evaluation metrics have been used for measuring the performance of skin detection algorithms. For comparison with other methods, in this study we employ the following metrics: the receiver operating characteristics (ROC) and respective area under curve (AUC) and equal error rate (1-EER).

In the test mode, the final output layer of the FCN8S+RNN network is a softmax layer of channel two, which represents the probability of each pixel to be a nonskin or skin pixel. A threshold is set to determine whether a pixel is skin or nonskin. If the output skin probability exceeds the threshold value, it is judged as skin, otherwise it is determined as nonskin. The ROC curve is created by plotting the true positive rate against the false positive rate at various threshold settings. Each threshold value produces a different point in ROC space. Fig. 3 demonstrates ROC curves of different skin detection algorithms on the COMPAQ and ECU datasets. Some authors (Doukim *et al.* [41], Poudel *et al.* [42] and Platzer *et al.* [43]) used fixed thresholds and thus their results correspond to only a point in the ROC space. The "*" on the ROC curves indicates optimal operating point that is calculated by moving the straight line with slope S from the upper left corner of the ROC plot down and to the right, until it intersects the ROC curve[45]. The slope S is calculated by

$$S = \frac{\text{Cost}(P|N) - \text{Cost}(N|N)}{\text{Cost}(N|P) - \text{Cost}(P|P)} * \frac{TN + FP}{TP + FN} \tag{3}$$

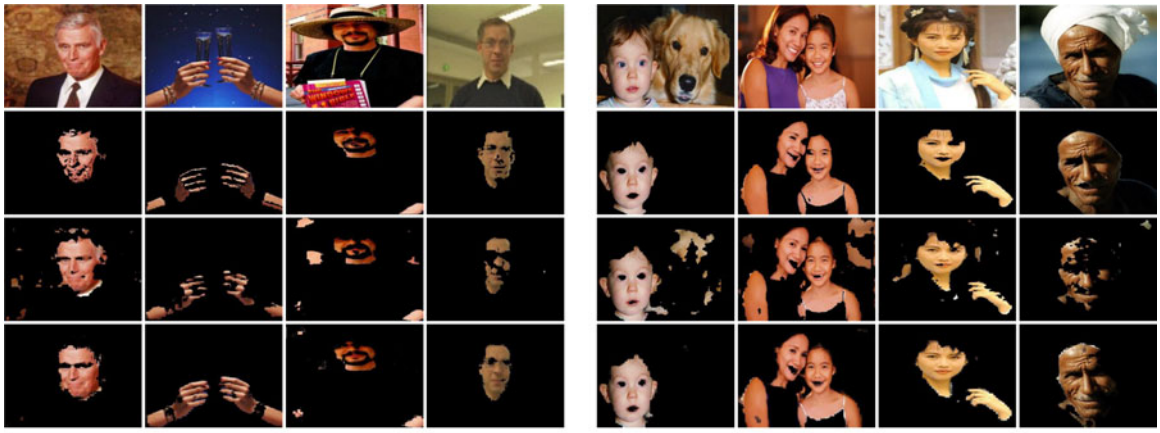

Fig. 4. Sample results of skin detection on the (a) COMPAQ and (b) ECU datasets. The first row shows the original images, and the second row gives the corresponding ground truth. The third row corresponds to the detected skin pixels using FCN8S without RNN layers, and the last row corresponds to the skin detection results using the proposed FCN8S+RNN framework.

TABLE II
PERFORMANCES OF DIFFERENT SKIN DETECTION ALGORITHMS ON THE COMPAQ DATASET

Method	AUC(%)	1-EER(%)
Jones and Rehg [9]	94.2	88.0
Sun [40]	95.2≈	88.8≈
FCN8S	94.78	88.21
FCN8S+RNN	**95.93**	**90.18**

≈These values are computed from the ROC curves listed in corresponding paper, not available in the original papers.

TABLE III
PERFORMANCES OF DIFFERENT SKIN DETECTION ALGORITHMS ON THE ECU DATASET

Method	AUC(%)	1-EER(%)
Bayesian [39]	93.9≈	87.8≈
Kawulok *et al.* [16]	95.8≈	92.0≈
Santos and Pedrini [44]	93.8≈	87.3≈
FCN8S	97.73	93.08
FCN8S+RNN	**98.10**	**94.80**

≈ These values are computed from the ROC curves listed in corresponding paper, not available in the original papers.

where $\mathrm{Cost}(I|J)$ is the cost of assigning an instance of class J to class I. In the experiment, the costs for correct classification $\mathrm{Cost}(N|N)$ and $\mathrm{Cost}(P|P)$ are set to zero, and $\mathrm{Cost}(P|N)$ and $\mathrm{Cost}(N|P)$ are set to 0.5 meaning that both positives are misclassified as negatives and negative are misclassified as positives are equally important. The sample images shown in Fig. 4 are reported under the corresponding optimal operating point derived from ROC curves. In our experiments, higher false positive rate (FPR) is not desirable due to the imbalance between skin and nonskin pixels. Typically, the number of nonskin pixels is many times larger than skin pixels. A high FPR will result in a significant reduction in precision and accuracy. Therefore, the performance in the far left-hand side of the ROC graphis more dominating.

In general, the curve at the top left of the ROC space has better performance which has higher TP rate and lower FP rate. It can be seen in Fig. 3, Tables II and III that the performances of the classifier using the deep learning methods are superior to most of the traditional hand-engineered methods and the proposed fully convolutional and RNNs (FCN8S+RNN) algorithm does the best job of discriminating skin and nonskin pixels on both the COMPAQ and ECU datasets. The overall performances on the COMPAQ dataset are worse than those on the ECU dataset because the COMPAQ contains a large number of low-quality images (saved in low-resolution gif format) and the semiautomatic ground truth is inaccurate.

Tables II and III present the performances of different skin detection algorithms on the COMPAQ and ECU datasets, respectively. Note that it is difficult to make a comparison of existing alogrithms, as they either use different evaluation metrics, or use nonpublic datasets, or use arbitrary operating points (or thresholds). Some values marked with ≈ are computed from the ROC curves listed in corresponding paper, not available in the original papers.

Fig. 4 demonstrates examples of skin detection results on the COMPAQ and ECU datasets. The penultimate row is the skin detection result of using FCN8S [24] algorithm, and the last row is the result of our method by combining FCN and RNN (FCN8S+RNN) algorithm. Fig. 4 shows that the FCN8S+RNN can effectively improve the stability of skin detection algorithm under complex background by introducing RNN layers.

IV. CONCLUSION

This study presents an end-to-end network for human skin detection by integrating recurrent neural layers into FCNs. RNN layers are employed to model the semantic spatial dependencies between image pixels. Experimental results show that our proposed FCN and RNN algorithm outperforms standard methods on both the COMPAQ and ECU skin datasets. RNN layers effectively improve the stability of skin detection algorithm under complex backgrounds.

REFERENCES

[1] W. Chen, K. Wang, H. Jiang, and M. Li, "Skin color modeling for face detection and segmentation: A review and a new approach," *Multimedia Tools Appl.*, vol. 75, no. 2, pp. 839–862, 2016.

[2] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: A survey," *Artif. Intell. Rev.*, vol. 43, no. 1, pp. 1–54, 2015.

[3] G. Schaefer, R. Tait, and S. Y. Zhu, "Overlay of thermal and visual medical images using skin detection and image registration," in *Proc. IEEE Int. Conf. Eng. Med. Biol. Soc.*, 2006, pp. 965–967.

[4] M. S. Devi and P. R. Bajaj, "Driver fatigue detection based on eye tracking," in *Proc. 1st Int. Conf. Emerg. Trends Eng. Technol.*, 2008, pp. 649–652.

[5] D. Chai, S. L. Phung, and A. Bouzerdoum, "Skin color detection for face localization in human-machine communications," in *Proc. 6th Int. Symp. Signal Process. Appl.*, 2001, vol. 1, pp. 343–346.

[6] A. Jaimes and N. Sebe, "Multimodal human–computer interaction: A survey," *Comput. Vis. Image Understanding*, vol. 108, no. 1, pp. 116–134, 2007.

[7] R. Fang, S. Pouyanfar, Y. Yang, S.-C. Chen, and S. Iyengar, "Computational health informatics in the big data age: A survey," *ACM Comput. Surveys*, vol. 49, no. 1, 2016, Art no. 12.

[8] M. R. Mahmoodi and S. M. Sayedi, "A comprehensive survey on human skin detection," *Int. J. Image, Graph. Signal Process.*, vol. 8, no. 5, pp. 1–35, 2016.

[9] M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection," *Int. J. Comput. Vis.*, vol. 46, no. 1, pp. 81–96, 2002.

[10] R. Khan, A. Hanbury, J. Stöttinger, and A. Bais, "Color based skin classification," *Pattern Recognit. Lett.*, vol. 33, no. 2, pp. 157–163, 2012.

[11] J. M. Chaves-González, M. A. Vega-Rodríguez, J. A. Gómez-Pulido, and J. M. Sánchez-Pérez, "Detecting skin in face recognition systems: A colour spaces study," *Digit. Signal Process.*, vol. 20, no. 3, pp. 806–823, 2010.

[12] V. Oliveira and A. Conci, "Skin detection using HSV color space," in *Proc. H. Pedrini, & J. Marques de Carvalho, Workshops of Sibgrapi*, 2009, pp. 1–2.

[13] M. M. Oghaz, M. A. Maarof, A. Zainal, M. F. Rohani, and S. H. Yaghoubyan, "A hybrid color space for skin detection using genetic algorithm heuristic search and principal component analysis technique," *Plos One*, vol. 10, no. 8, 2015, Art. no. pp. e0134828.

[14] M. Z. Osman, M. A. Maarof, and M. F. Rohani, "Towards integrating statistical color features for human skin detection," *Int. J. Comput. Elect. Autom. Control Inf. Eng.*, vol. 10, no. 2, pp. 292–296, 2016.

[15] H. K. Al-Mohair, J. M. Saleh, and S. A. Suandi, "Hybrid human skin detection using neural network and k-means clustering technique," *Appl. Soft Comput.*, vol. 33, pp. 337–347, 2015.

[16] M. Kawulok, J. Kawulok, and J. Nalepa, "Spatial-based skin detection using discriminative skin-presence features," *Pattern Recognit. Lett.*, vol. 41, pp. 3–13, 2014.

[17] M. M. Hasan and P. K. Mishra, "Superior skin color model using multiple of Gaussian mixture model," *Brit. J. Sci.*, vol. 6, no. 1, pp. 1–14, 2012.

[18] Q. Zhu, K.-T. Cheng, and C.-T. Wu, "A unified adaptive approach to accurate skin detection," in *Proc. Int. Conf. Image Process.*, 2004, vol. 2, pp. 1189–1192.

[19] K. Bhoyar and O. Kakde, "Skin color detection model using neural networks and its performance evaluation," in *J. Comput. Sci.*, vol. 6, no. 9, pp. 963–968, 2010.

[20] R. Khan, A. Hanbury, and J. Stoettinger, "Skin detection: A random forest approach," in *Proc. IEEE Int. Conf. Image Process.*, 2010, pp. 4613–4616.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[22] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1915–1929, Aug. 2013.

[23] C. Liang-Chieh, G. Papandreou, I. Kokkinos, K. Murphy, and A. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–14. [Online]. Available: https://arxiv.org/abs/1412.7062

[24] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

[25] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1520–1528.

[26] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," 2015. [Online]. Available: https://arxiv.org/abs/1511.00561

[27] S. Zheng *et al.*, "Conditional random fields as recurrent neural networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1529–1537.

[28] I. Sutskever, J. Martens, and G. E. Hinton, "Generating text with recurrent neural networks," in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 1017–1024.

[29] P. H. Pinheiro and R. Collobert, "Recurrent convolutional neural networks for scene labeling," in *Proc. 31st Int. Conf. Mach. Learn.*, 2014, pp. 82–90.

[30] A. Karpathy and L. Fei-Fei, "Deep visual-semantic alignments for generating image descriptions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3128–3137.

[31] W. Byeon, T. M. Breuel, F. Raue, and M. Liwicki, "Scene labeling with LSTM recurrent neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3547–3555.

[32] F. Visin *et al.*, "ReSeg: A recurrent neural network-based model for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2016, pp. 426–433.

[33] S. Bell, C. L. Zitnick, K. Bala, and R. Girshick, "Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2874–2883.

[34] A. Van den Oord, N. Kalchbrenner, and K. Kavukcuoglu, "Pixel recurrent neural networks," in *Proc. 33th Int. Conf. Mach. Learn.*, 2016, pp. 1747–1756.

[35] Z. Zuo *et al.*, "Learning contextual dependence with convolutional hierarchical recurrent neural networks," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 2983–2996, Jul. 2016.

[36] B. Shuai, Z. Zuo, G. Wang, and B. Wang, "Dag-recurrent neural networks for scene labeling," *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 3620–3629.

[37] H. Fan, X. Mei, D. Prokhorov, and H. Ling, "Multi-level contextual RNS with attention model for scene labeling," 2016, [Online]. Available: https://arxiv.org/abs/1607.02537

[38] A. Vedaldi and K. Lenc, "Matconvnet – convolutional neural networks for MATLAB," in *Proc. ACM Int. Conf. Multimedia*, 2015, pp. 689–692.

[39] S. L. Phung, A. Bouzerdoum, and D. Chai, "Skin segmentation using color pixel classification: Analysis and comparison," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 1, pp. 148–154, Jan. 2005.

[40] H.-M. Sun, "Skin detection for single images using dynamic skin color modeling," *Pattern Recognit.*, vol. 43, no. 4, pp. 1413–1420, 2010.

[41] C. A. Doukim, J. A. Dargham, A. Chekima, and O. Sigeru, "Combining neural networks for skin detection," *Signal Image Process., Int. J.*, vol. 1, no. 2, pp. 1–11, 2010.

[42] R. P. Poudel, H. Nait-Charif, J. J. Zhang, and D. Liu, "Region-based skin color detection," in *Proc. Int. Conf. Comput. Vis. Theory Appl.*, 2012, vol. 1, pp. 301–306.

[43] C. Platzer, M. Stuetz, and M. Lindorfer, "Skin sheriff: A machine learning solution for detecting explicit images," in *Proc. 2nd Int. Workshop Security Forensics Commun. Syst.*, 2014, pp. 45–56.

[44] A. Santos and H. Pedrini, "Human skin segmentation improved by saliency detection," in *Proc. Int. Conf. Comput. Anal. Images Patterns*. Berlin, Germany: Springer, 2015, pp. 146–157.

[45] M. H. Zweig and G. Campbell, "Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine," *Clin. Chem.*, vol. 39, no. 4, pp. 561–577, 1993.