HUMAN SKIN SEGMENTATION USING FULLY CONVOLUTIONAL NEURAL NETWORKS

Chang-Hsian Ma and Huang-chia Shih

Human-Computer Interaction Multimedia Laboratory,
Department of Electrical Engineering, Yuan Ze University, Chungli, Taiwan
E-MAIL: hcshih@saturn.yzu.edu.tw

ABSTRACT

In recent years, skin segmentation has attracted much of attention from computer vision field. Normally, researchers use a simple pre-trained model or define a fixed threshold in color space to deal with skin segmentation. However, it is highly possible to failure in many conditions. In addition, convolutional neural network (CNN) has achieved great success in computer vision. This paper we present a fully convolutional neural network method in skin segmentation. A hand-crafted skin dataset has provided in this study. In the experiment, we attempt many CNN structures to determine the best one. According to the experimental result, we obtained a considerable result in three well-known skin datasets.

Index Terms—*Convolutional neural network, deep learning, skin segmentation, human skin dataset.*

1. INTRODUCTION

Recent years many research applications with skin segmentation have been presented such as in hand gesture recognition [1], medical image analysis [2], and human-computer interaction [3]. Skin segmentation serves as a pre-processing procedure among them. However, it is challenging due to complex background, diverse light conditions, and cluttered environment. To deal with skin segmentation, we are unable to use a straightforward method such as predefined threshold in every color model. In practice, there are many variations. For example, people would wear various color clothes, the illumination condition changes in different location, and the shadow causes many false detections. In addition, the human races also affect many variances in our task.

In 2012, Krizhevsky et al. presented AlexNet [4] which outperformed all the teams in ILSVRC 2012. The convolutional neural network has become a cutting-edge technique in many computer vision applications. By using machine learning method, we do not need to design a precise algorithm, instead by the machine learning. Therefore, we can address on collecting the training data and modifying the system model.

It exists various skin segmentation applications [5]-[7], which use different color space such as YCbCr and HSI to reduce the effect from illumination condition. It can only valid in certain conditions. As aforementioned, the conventional algorithm is unable to detect skin pixel perfectly in complex environment. In this study, we applied convolutional neural network (CNN) to solve the skin segmentation problem. The convolutional layer can automatically determine the features from image in the training phase, and rebuilt where the skin regions located on the image. Although CNN is a powerful model in many applications, it needs a considerable training data in training process. Unfortunately, it is difficult to find as many public skin datasets, as well as the resolution among them is low. Therefore, we have prepared many skin image with ground truth to enlarge our training set. In this paper we proposed a fully convolutional neural network to achieve skin segmentation. Overall, it received 94.99% accuracy on Pratheepan skin dataset [8].

The remainder of this paper is organized as follows. Section 2 presents the detail of the proposed framework. Experimental results are provided in Section3. Finally, we discuss our findings and draw our conclusion in Section 4.

2. SYSTEM FRAMEWORK

This study, the CNN is used for modeling the skin segmentation framework. The convolution kernel can be defined as:

$$\mathrm{h}[m,n] = f\left(\sum_{i=0}^{s}\sum_{j=0}^{s} x[i,j]h[m+i,n+j]+b\right) \quad (1)$$

where $[m, n]$ denotes the output position, s denotes the kernel size, b denotes the bias, and f denotes the activation function, which used ReLU, expect the output layer. In output layer we use sigmoid function because that we need the output result as the skin confidence map. The max pooling kernel defined as:

$$\mathrm{h}[m,n] = \max\left[x_{m*s,n*s},\cdots,x_{(m+1)*s,(n+1)*s}\right] \quad (2)$$

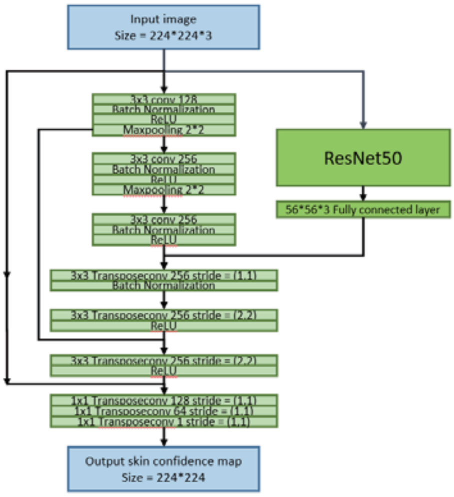

Fig. 1. Framework of the convolutional neural network

Max pooling can reduce image size but losing image information. It can reduce much of time in training process and let our model more stable.

Fig. 1 shows the proposed CNN framework. Considering ResNet [9] and Inception [10], we use 3x3 size convolution kernel to make our model more deeper. In addition, we use 1 ×1 convolution as output layer instead of fully connected layer. Because that if we use fully connected layer before convolutional layer, we need a large amount of weights to connect it. If there are considerable weights, it is highly possible to meet overfitting problem in training process. Using skip (or highway) scheme to prevent the gradient vanishing problem on deep neural network. Therefore, we assign the network input image to the bottom layer, and the output of the first hidden layer connect to the input of the penultimate layer.

Because that the skin dataset is difficult to obtain, we use ResNet50 convolutional layer with transfer learning. With ResNet50 pre-training in ImageNet, the convolutional layer can catch the feature comprehensively. In addition, let our model not only focus on training data. In order to train our network, the loss function L can compute by

$$L = \frac{1}{m * n} \sum_{i=0}^{m} \sum_{j=0}^{n} (y_{i,j} - Y_{i,j})^2 \qquad (3)$$

where m and n denote the column and row of image coordinate, $y_{i,j}$ denotes the corresponding output pixel value, and $Y_{i,j}$ denotes the binary label of the skin ground truth.

3. EXPERIMENTAL RESULTS

A. Experimental Setup

In this study, the algorithm was trained and tested on fully convolutional network, and obtains the image feature with ResNet50. The computer we used is with NVIDIA GeForce GTX 1080 Ti and with 11GB GPU memory. The training time costs around 2 hours with 20 epochs. The input image size would not have same shape, thus we crop it into patches with 224×224. To evaluate the result, the result image will reconstruct and sum up the scores among the patches. The training optimizer we applied is named "Adam" [11]. The parameters we setup are as follows, learning rate=0.001, beta_1=0.9, beta_2=0.999, and epsilon=1e-08. Finally, we add batch normalization [12] trick to improve out result.

Table. 1. Evaluation results with Pratheepan datasets

Pratheepan dataset				
Methods	*Accuracy*	*Precision*	*Recall*	*F-measure*
Bayesian [15]	0.8237	0.6881	0.8972	0.7788
FSD [16]	0.8255	0.8077	0.6851	0.7414
LASD [17]	0.8361	0.7954	0.8275	0.8111
FPSD [18]	0.8419	0.7387	0.8991	0.8070
DSPF [19]	0.8521	0.7543	0.8436	0.7964
SPSD [20]	0.8782	0.7659	**0.9328**	0.8412
Patch-VGG [21]	0.9299	0.8563	0.8750	0.8655
Patch-NiN [21]	0.9334	0.8802	0.8972	0.8886
Image-VGG [21]	0.9313	0.8577	0.9069	0.8816
Image-NiN [21]	0.9484	**0.9003**	0.8912	**0.8957**
Ours	**0.9499**	0.8480	0.8981	0.8678

Fig. 2. Experimental results in Pratheepan datasets. (1) the input image. (2) the corresponding ground truth. (3) the results of the network output. (4) final results by setting threshold.

B. Dataset

We collected three public skin datasets, including ECU with 4000 skin images [13], SFA with 1000 images [14], and Pratheepan with 80 images. Those skin datasets have different illumination conditions and multi-races, in cluttered background and with occlusion. To improve the system performace, we prepared more than 50 high quality skin images to add on the training dataset. During training phase, we used all the skin datasets except of Pratheepan dataset, either to validation datasets. We used the Pratheepan dataset to test our model, because of the highest resolution among datasets. Table I shows the comparison result with the state-of-the-art systems.

4. CONCLUSION

In this paper we used a convolutional neural network architecture to achieve skin segmentation. Using skip scheme to prevent the gradient vanishing problem on deep neural network. A hand-crafted skin dataset has provided in this study. According to the experimental result, the proposed skin segmentation method obtained convincing result in three public skin datasets.

ACKNOWLEDGMENT

This research is supported by the Ministry of Science and Technology, Taiwan, under Grant MOST 107-2221-E-155-031-MY2.

REFERENCES

[1] S. Reshna and M. Jayaraju, "Spotting and Recognition of Hand Gesture for Indian Sign Language Recognition System with Skin Segmentation and SVM" IEEE WiSPNET 2017 conference.

[2] H. Alquran1, I. A. Qasmieh, A. M. Alqudah, S. Alhammouri, E. Alawneh, A. Abughazaleh, F. Hasayen. "The Melanoma Skin Cancer Detection and Classification using Support Vector Machine," in *Proc. of IEEE Jordan Conference on AEECT,* 2017.

[3] N. Vo, Q. Tran, T. B. Dinh, T. B. Dinh, Q. M. Nguyen. "An Efficient Human-Computer Interaction Framework Using Skin Color Tracking and Gesture Recognition," in *Proc. of IEEE RIVF*, 2010.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Proc. NIPS'12*, Dec. 2012, pp 1097-1105.

[5] Y. F. A. Gaus and F. Wong, "Hidden Markov Model - based Gesture Recognition with Overlapping Hand-Head/Hand- Hand Estimated using Kalman Filter," in *Proc. ISMS'12*, Feb 2012, pp. 262–267.

[6] K. B. Shaika, P. Ganesan, V.Kalist, B.S. Sathish, J. M. M. Jenithab, "Comparative Study of Skin Color Detection and Segmentation in HSV and YCbCr Color Space," in *Proc. of ICRTC,* 2015.

[7] S. L. Phung, A. Bouzerdoum, D. Chai, "Skin segmentation using color pixel classification: analysis and comparison." *IEEE TPAMI*, vol. 27, no. 1, January 2005.

[8] P. Yogarajah, J. Condell, K. Curran, A. Cheddad, and P. McKevitt, "A dynamic threshold approach for skin segmentation in color images," in *Proc. of IEEE ICIP*, 2010, pp. 2225–2228.

[9] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition." *arXiv*:1512.03385

[10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, "Going Deeper with Convolutions," *arXiv*:1409.4842.

[11] D. Kingma, J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv*:1412.6980v8.

[12] S. Ioffe, . Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *arXiv*:1502.03167.

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

[14] J.P.B. Casati, D. R. Moraes, and E. L. L. Rodrigues. "SFA: A Human Skin Image Database based on FERET and AR Facial Images," in *Proc. of WVC*, 2013.

[15] M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection," *International Journal of Computer Vision*, vol. 46, no. 1, pp. 81–96, 2002.

[16] W. R. Tan, C. S. Chan, P. Yogarajah, and J. Condell, "A fusion approach for efficient human skin detection," *IEEE Transactions on Industrial Informatics*, vol. 8, no. 1, pp. 138–147, 2012.

[17] I. Hwang, S. H. Lee, B. Min, and N. I. Cho, "Luminance adapted skin color modeling for the robust detection of skin areas," in *Proc IEEE ICIP*, 2013, pp. 2622–2625.

[18] M. Kawulok, "Fast propagation-based skin regions segmentation in color images," in *Proc. of IEEE FG*, 2013, pp. 1–7.

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

[20] I. Hwang, Y. Kim, and N. I. Cho, "Skin detection based on multi-seed propagation in a multi-layer graph for regional and color consistency," in *Proc. of IEEE ICASSP*, 2017.

[21] Y. Kim, I. Hwang, N. I. Cho, "Convolutional neural networks and training strategies for skin detection" in *Proc. of IEEE ICIP*, 2017.