

Skin Segmentation at the Pixel Level Using Fully Convolutional Neural Network

Harsha B K¹, Indumathi G²

¹Department of ECE, CMR Institute of Technology (VTU RC), Bengaluru, Karnataka, India

²Department of ECE, Cambridge Institute of Technology, Bengaluru, Karnataka, India

Abstract: Modeling skin segmentation is an important activity in applications involving skin detection as a pre-processing element in biometrics, recognition of human activity and many more. There are many challenges faced in the segmentation of human skin pixels like background blend, changing illumination, varying skin races. Therefore, there is a need for proper pre-processing step to remove these redundancies before processing the image for the given application. The pre-processing is achieved by skin segmentation using fully convolutional neural network in this work. The segmentation is evaluated using four datasets and performance parameters are measured.

Keywords: Human skin segmentation, semantic segmentation, Convolutional Neural Network

1. Introduction

Detecting the regions containing skin pixels is the goal of systems incorporating skin segmentation as pre-processing step. This pre-processing helps in improved performance of the systems such as face recognition, emotion recognition, gesture recognition, content relative image filtering and so on. The process of skin segmentation in images involves identification of skin-colored pixels. All those applications that includes skin pixel identification requires skin segmentation as its primary task. Convolutional neural networks have gained importance with advancements in the neural network field and all the skin detection tasks are adaptive because of the same.

In the arena of artificial intelligence that is focused towards human and machine interface, skin segmentation and recognition algorithms are being paid more attention. The human machine interface systems include smart mirrors, activity identifier and recording sub system, recording the movement of individuals and multiple population, crowd sensing. Though skin segmentation is a pre-processing step, the performance greatly influences the further process. This pre-processing activity is challenging because the skin of humans includes various races. One of the most popular yet really interesting scenario is the crowd recognition and careful monitoring of antisocial activity identification.

In the field of sports, identifying the activities of players are incorporated in applications such as recognizing player actions during a tennis match. This difficulty is exacerbated by structural differences and thus making the process a complicated task. When there are multiple subjects in a frame, the problem becomes significantly more complicated. As a result, when skin segmentation is used, convolutional neural network approach serves better than the traditional algorithms. In this paper, segmentation of skin is presented along with the challenges and related concerns.

2. Literature Survey

Skin detection is widely used in image processing and visual computing for a wide range of applications. Many skin detection studies employ a variety of techniques. These methods are classified into three types: pixel classification based on thresholding algorithms, conventional approaches and the neural network approaches. These procedures use different image channels to segment the skin regions and non-skin parts in the image. In the algorithms based on thresholding methods, skin pixel identification is performed based on color models used and main concept being the threshold. Threshold values are selected relative to the values those are nearer to the skin pixels.

Based on the color models used, threshold value is used to identify if a pixel is similar to skin color or it belongs to non-skin color. But there are various challenges to identify skin pixels. Some of the basic checks posed are background blend, foreground misclassification and so on.

[1] discusses an eye tracker that is developed to improve performance irrespective of the illumination in which the image is captured. In [2] an algorithm is proposed to identify skin colored pixels in the scenario of background blend. In this method, performance is increased significantly by reduced errors. Information regarding weights and selection criteria is discussed in [3]. Various color models are discussed to be used in skin segmentation in [4]. In this work, discussion is on different color models used to detect skin color pixels. The work concentrates on increased performance.

Because skin detection is an important step in image pre-processing, [5] discusses an algorithm that makes use of a grouping method to make sets of same kind of pixels from the input image. Regardless of ethnicity, the technique proves to provide good performance with identification of skin-colored pixels from the input images. The technique also behaves well in reference to the variation in lighting condition. [6] discusses about a net designed to improve regional segmentation and marks the improvement in the

accuracy parameter. The designed net contributes to increase in the sensitivity level. [7] discusses about an object detection algorithm that is effective in detecting small and occluded areas using different scales.

Skin segmentation plays a vital role in many of the medical applications that deals with identification of skin diseases such as lesion detection. [5] discusses a technique for segmenting regions of skin in input images containing humans with the aid of segmentation algorithms. The authors state that they provide competitive satisfaction towards performance. [8] discussed and implemented a method for classifying pixels to belong to skin under changes related to different lighting scenarios.

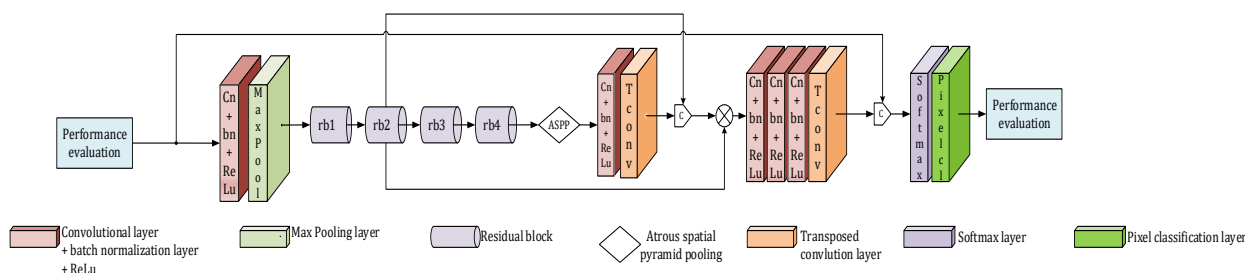
[9] discusses about the color spaces utilized for segmentation of skin pixels based on threshold value. Limiting factor in the detection of skin tone is the background blend. [10] discusses a technique on the basis of color model including the features extracted from the texture information of the input image. The method has proved its efficiency in terms of increased accuracy. [11] orients the study towards a technique based on segmentation based on palm print to improve accuracy.

[12] discusses a model based on the venture of two techniques to improve skin detection performance, to deal

with the problem happening by changes in illumination causing similar skin colored background. [13] discusses a method that improves the skin detection that uses a neural network to detect skin and body. Nevertheless, the present methods on machine learning or older neural networks they all have some or other limitations in terms of performance under specific lighting conditions.

3. System Design

Skin segmentation is difficult due to the cluttered background, objects at multiple scales, and small and deformable objects. Due to common challenges in both, we treated skin segmentation as a semantic segmentation problem in the work presented. As a result, we adapted the well-known DeepLabv3+ architecture, which is state-of-the-art in semantic segmentation, for skin segmentation. The DeepLabv3+ architecture's inherent nature is tailored to scenes with cluttered backgrounds. The image is first subjected to residual learning to address a few skin-segmentation challenges, such as color similarity between foreground and background, and skin reflectance variations due to lighting conditions. We begin by shredding residual blocks in the proposed SSS-Net, keeping only four residual blocks.



There are two underlying reasons for reduced residual blocks: First, to preserve small details that would otherwise be lost in the repetitive convolution process. Previous work supports reducing the number of layers to preserve object and semantic information for semantic segmentation. Second, the number of parameters and computational load should be reduced. Downscaling is not used in the residual learning process to preserve feature information. Due to the deformable nature and limited extent of skin regions, contextual information is critical in skin segmentation. As a result, we use Atrous Spatial Pyramid Pooling (ASPP) to capture image context at various scales.

ASPP has proven to be robust in detecting objects at multiple scales as well as efficient in mitigating background clutter in object recognition scenarios due to its intuitive local feature processing and subsequent fusion. Because of the multi-scale contextual information contained in them, spatial pyramids have been successfully used for dense prediction tasks. We should mention that the ASPP, if not carefully designed, has the potential to miss small skin regions. As a result, we experimented with different dilation rates so that our filters could cater to both small and large skin regions. In our experiments, we discovered that limiting the number of residual blocks to four preserves vital

semantic information of regions that are then passed through the specially designed ASPP module without resulting in the loss of small objects.

4. Experimental results

The network designed was tested on Pratheepan dataset, Compaq, SFA and Schummage datasets those are publicly available. The datasets used for the skin task and obtained outputs are shown in Figure 4.1, Figure 4.2, Figure 4.3, Figure 4.4 and Figure 4.5.



Figure 4.1: Images from Pratheepan dataset [14]single image



Figure 4.2: Images from Pratheepan dataset [14]community image



Figure 4.3: Images from Compaq dataset [15]



Figure 4. 4.: Images from SFA dataset [16]



Figure 4. 5: Images from Schummage dataset [17]

The obtained output is as tabulated in Table 1.

Table 1: Performance parameters obtained from the proposed network

Parameters \ Dataset	Pratheepan	SFA	Compaq	Schummage
Accuracy	0.97688	0.96091	0.96038	0.96680
Precision	0.94480	0.96834	0.65944	0.90909
Recall	0.94190	0.87882	0.53357	0.90909
Specificity	0.98986	0.99093	0.98394	0.97406
FPR	0.02344	0.00907	0.01606	0.02594
FNR	0.01810	0.12118	0.46643	0.34858
PWC	0.02312	0.03909	0.03962	0.03320
F1	0.93669	0.92003	0.86277	0.80340

References

- [1] A. Haria, A. Subramanian, N. Asokkumar, S. Poddar, and J. S. Nayak, "Hand Gesture Recognition for Human Computer Interaction, " *Procedia Comput. Sci.*, vol.115, pp.367–374, 2017, doi: 10.1016/j.procs.2017.09.092.
- [2] B. K. Chakraborty and M. K. Bhuyan, "Skin segmentation using Possibilistic Fuzzy C-means clustering in presence of skin-colored background, " *2015 IEEE Recent Adv. Intell. Comput. Syst. RAICS 2015*, pp.246–250, Jun.2016, doi: 10.1109/RAICS.2015.7488422.
- [3] L. Leng, J. Zhang, J. Xu, M. K. Khan, and K. Alghathbar, "Dynamic weighted discrimination power analysis in DCT domain for face and palmprint recognition, " *2010 Int. Conf. Inf. Commun. Technol. Converg. ICTC 2010*, no. December, pp.467–471, 2010, doi: 10.1109/ICTC.2010.5674791.
- [4] M. Z. Osman, M. A. Maarof, and M. F. Rohani, "Improved Dynamic Threshold Method for Skin Colour Detection Using Multi-Colour Space, " *Am. J. Appl. Sci.*, vol.13, no.2, pp.135–144, Feb.2016, doi: 10.3844/ajassp.2016.135.144.
- [5] E. Hassan, A. R. Hilal, and O. Basir, "Using ga to optimize the explicitly defined skin regions for human skin color detection, " *Can. Conf. Electr. Comput. Eng.*, Jun.2017, doi: 10.1109/CCECE.2017.7946699.
- [6] Y. Zhang, J. Chu, L. Leng, and J. Miao, "Mask-refined R-CNN: A network for refining object details in instance segmentation, " *Sensors (Switzerland)*, vol.20, no.4, 2020, doi: 10.3390/s20041010.
- [7] J. Chu, Z. Guo, and L. Leng, "Object Detection Based on Multi-Layer Convolution Feature Fusion and Online Hard Example Mining, " *IEEE Access*, vol.6, pp.19959–19967, 2018, doi: 10.1109/ACCESS.2018.2815149.
- [8] S. K. Ueng and C. Y. Chang, "An improved skin color model, " *2016 Int. Conf. Appl. Syst. Innov. IEEE ICASI 2016*, vol. D, no. November, 2016, doi: 10.1109/ICASI.2016.7539812.
- [9] K. B. Shaik, P. Ganesan, V. Kalist, B. S. Sathish, and J. M. M. Jenitha, "Comparative Study of Skin Color Detection and Segmentation in HSV and YCbCr Color Space, " *Procedia Comput. Sci.*, vol.57, pp.41–48, 2015, doi: 10.1016/j.procs.2015.07.362.
- [10] H. K. Al-Mohair, J. Mohamad 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, Aug.2015, doi: 10.1016/J.ASOC.2015.04.046.
- [11] M. Z. Osman, M. A. Maarof, and M. F. Rohani, "Improved skin detection based on dynamic threshold using multi-colour space, " *Proc.-2014 Int. Symp. Biometrics Secur. Technol. ISBAST 2014*, pp.29–34, 2015, doi: 10.1109/ISBAST.2014.7013089.
- [12] A. A. Zaidan, N. N. Ahmad, H. Abdul Karim, M. Larbani, B. B. Zaidan, and A. Sali, "Image skin segmentation based on multi-agent learning Bayesian and neural network, " *Eng. Appl. Artif. Intell.*, vol.32, pp.136–150, 2014, doi: 10.1016/j.engappai.2014.03.002.
- [13] T. Xu, Y. Wang, and Z. Zhang, "Towards Independent Color Space Selection for Human Skin Detection, " *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol.7674 LNCS, pp.337–346, 2012, doi: 10.1007/978-3-642-34778-8_31.
- [14] W. R. Tan, C. S. Chan, P. Yogarajah, and J. Condell, "A Fusion Approach for Efficient Human Skin Detection, " *IEEE Trans. Ind. Informatics*, vol.8, no.1, pp.138–147, Feb.2012, doi: 10.1109/TII.2011.2172451.
- [15] M. J. Jones and J. M. Rehg, "Statistical Color Models with Application to Skin Detection, " *Int. J. Comput. Vis.* 2002 461, vol.46, no.1, pp.81–96, 2002, doi: 10.1023/A: 1013200319198.
- [16] 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, " *IX Work. Visão Comput.*, 2013.
- [17] S. J. Schmugge, S. Jayaram, M. C. Shin, and L. V. Tsap, "Objective evaluation of approaches of skin detection using ROC analysis, " *Comput. Vis. Image Underst.*, vol.108, no.1–2, pp.41–51, 2007, doi: 10.1016/j.cviu.2006.10.009.