LaTeX Supplementary Material for CVPR

Anonymous CVPR submission

Paper ID 23

1. Results on the ECU dataset

1.1. Quantitative results

The main paper set the threshold of predicted probability to be 0.5 to classify the image pixels as skin or non-skin pixels and convert the sigmoid output from the deep learning model to binary. To make the experimental testing results more convincing, we illustrate the precision-recall curve [4] in Figure 1.

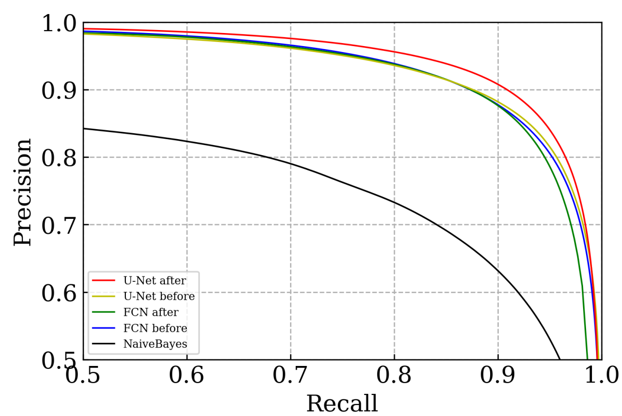


Figure 1. Precision-recall curve from testing experiments on the ECU dataset.

1.2. Qualitative results

This section will supply more output results from the skin detection systems mentioned in the main paper. We illustrate another six examples in Figure 3. The first row is a girl wearing a skin color-like cloth. The second and third rows contain backgrounds that have similar colors as the people in the image. The fourth row is a girl with brown cloth. The fifth row is a baby with strong lights on his head. The sixth row contains multiple people in various poses and skin colors These challenging conditions make other methods fail or perform poorly. The three baseline methods all fail to classify the skin color-like ground in the first and second rows. U-Net (B) works better, but there is still some false positive noise in rows 1 to 5. Moreover, it fails to detect the people on the right in the last row. In contrast, our approach overcame most of the difficulties mentioned above and produced accurate and robust results. Compared with results before color augmentation, models with color augmentation make less false positive and false negative judgments. For example, the FCN (A) does not detect the baby's hair as skin pixels in the fifth row, and it does not make noise as FCN (B) does in the second row. In this part of view, our method outperforms the traditional skin segmentation methods, and color augmentation helps deep learning methods work better.

2. Results on the RFW dataset

2.1. Qualitative results

We demonstrate more results from the RFW dataset in Figure 4. The three traditional skin segmentation methods still misclassify the color-like background to be skin pixels. For example, the background of the door is classified as skin areas in the third row by the three baseline methods. On the opposite, in the second row, glasses covered area is not classified as skin areas. What's more, the skin area covered by other items is detected as non-skin pixels. Compared with results after color augmentation, models without color augmentation are more likely to make false-positive judgments. Moreover, in the darker skin group, models after color augmentation can detect more skin pixels. For example, the result from FCN (A) has less false positive noise than that from FCN (B) in the first row. U-Net (A) detects more skin pixels on the man's head in the fourth row.

2.2. Skin/face ratios

In the main paper, we propose a new method, skin/face ratio, to evaluate the performance of the skin segmentation system with the RFW dataset. It refers to the number of detected skin pixels inside the face area. Although the level of this indicator can reflect the ability of the detection system, larger values do not mean better prediction entirely. In Figure 2, we extract the skin/face values from various groups of predictions to make it more convincing. In this section, we

Table 1. Kullback–Leibler divergence between the standard probability distribution and that from estimated methods. Results are from U-Net before and after color augmentation with different groups and the whole RFW dataset.

	Cau	Asian	Ind	Afr	Overall
Before Aug.	0.32	0.14	0.15	0.19	0.16
After Aug.	0.05	0.03	0.06	0.22	0.05

plot skin/face ratio curve to evaluate the performance of the skin segmentation models. The skin/face ratio curve refers to the probability distribution of the skin/face values from the results.

First, we plot skin/face ratio curve using the annotated ECU dataset and its corresponding ground truth, which will be regarded as a sample or a standard (blue). Then, we plot the exact curve of the results from U-Net before and after color augmentation with RFW dataset. The curves are shown in Figure 2. We calculate Kullback–Leibler divergence (D_{KL}) to measure the difference between the standard probability distribution and that from estimated methods. We expect the resulting curve from a better model to be more relevant to the standard curve, that is, has smaller D_{KL} to the standard distribution.

The D_{KL} values are listed in Table 1. It demonstrates that model after color augmentation is more relevant to the standard distribution in Caucasian, Asian, and Indian groups since they have the smaller D_{KL}. This also happens in the whole RFW dataset. However, for the African group, the model before color augmentation has a better performance.

From Figure 2, we find that there is a peak at point '0' for the model before augmentation, which does not appear in the standard curve. This peak indicates that the model does not detect any skin pixels from the face area, which is incorrect. After color augmentation, the model works well, and this peak disappears. This explains a lot why skin/face ratio increase after color augmentation. Although the level of this indicator can reflect the ability of the work

3. Results from cross dataset experiments

In this section we illustrate additional results of grayscale images from the ECU dataset and our self-made drastic dataset. FCN (B) can detect only a small area of skin pixel in grayscale image but U-Net (B) fail to detect any skin pixels. On the other hand, both models without color augmentation can hardly detect a single skin pixel for the images with unconstrained illumination. Improvements are obvious after color augmentation is applied to the network. Models with color augmentation work well and correctly detected skin pixels for both grayscale images and images with unconstrained illumination.

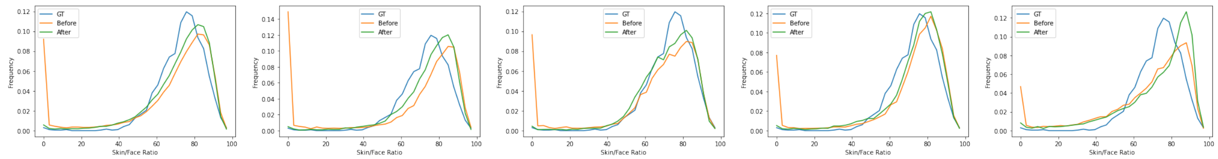
References

[1] Djamila Dahmani, Mehdi Cheref, and Slimane Larabi. Zero-sum game theory model for segmenting skin regions. *Image and Vision Computing*, 99:103925, 2020. 3, 4

[2] Michael J. Jones and James M. Rehg. Statistical color models with application to skin detection. *International Journal of Computer Vision*, 46(1):81–96, 2002. 3, 4

[3] S. Kolkur, D. Kalbande, P. Shimpi, C. Bapat, and J. Jatakia. Human skin detection using RGB, HSV and YCbCr color models. In *Proceedings of the International Conference on Communication and Signal Processing*, pages 324–332. Atlantis Press, 2016/12. 3, 4

[4] Xuebin Qin, Zichen Zhang, Chenyang Huang, Chao Gao, Masood Dehghan, and Martin Jagersand. Basnet: Boundary-aware salient object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 1



(a) Overall (b) Caucasian (c) Asian (d) Indian (e) African

Figure 2. Skin/face ratio distributions curves for Overall RFW dataset (a) and the four different races in RFW dataset (b to e). Blue line refers to the sample distribution curve we get from the annotated ECU dataset. Orange and green line refer to the distribution from testing results before and after color augmentation.

Figure 3. Additional results on the ECU dataset, by various skin segmentation methods including Kolkur *et al*. [3], Dahmani *et al*. [1], Jones *et al*. [2], FCN before (B) and after (A) augmentation, and U-Net before (B) and after (A) augmentation (Columns 2 to 8). Input and ground truth are shown in column 1 and 9.

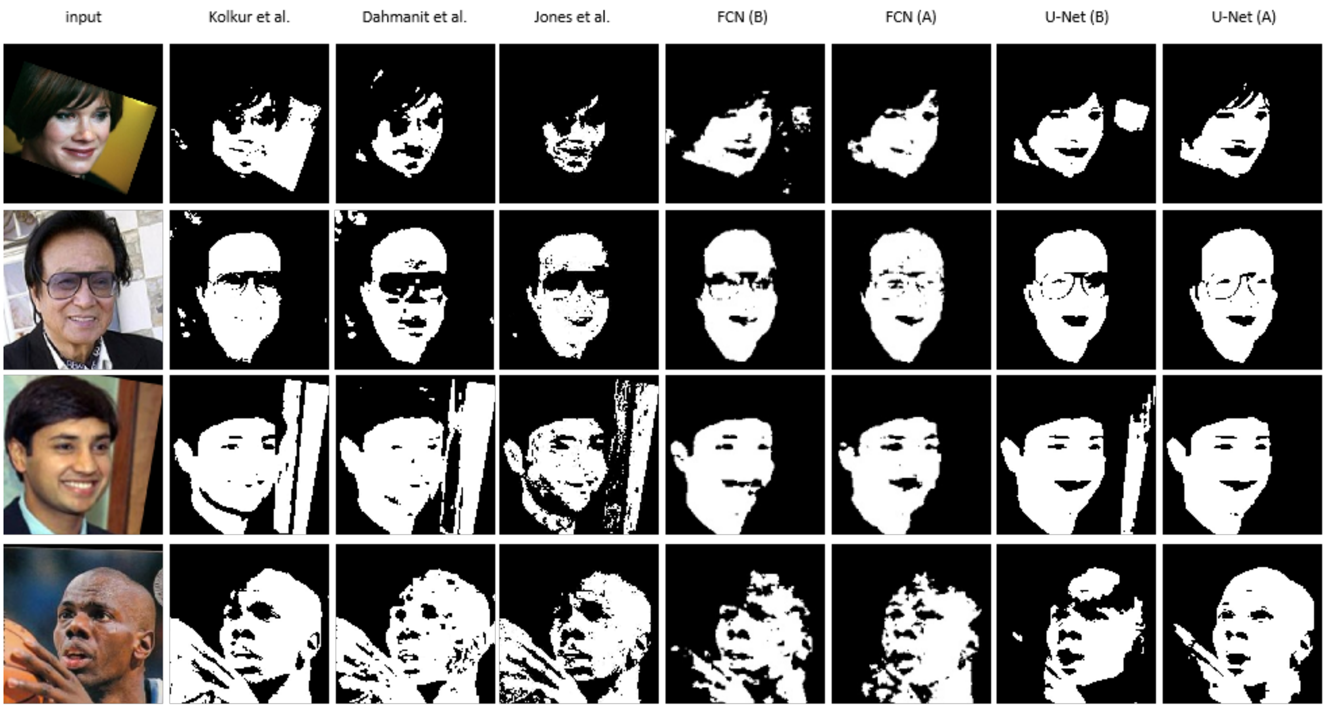
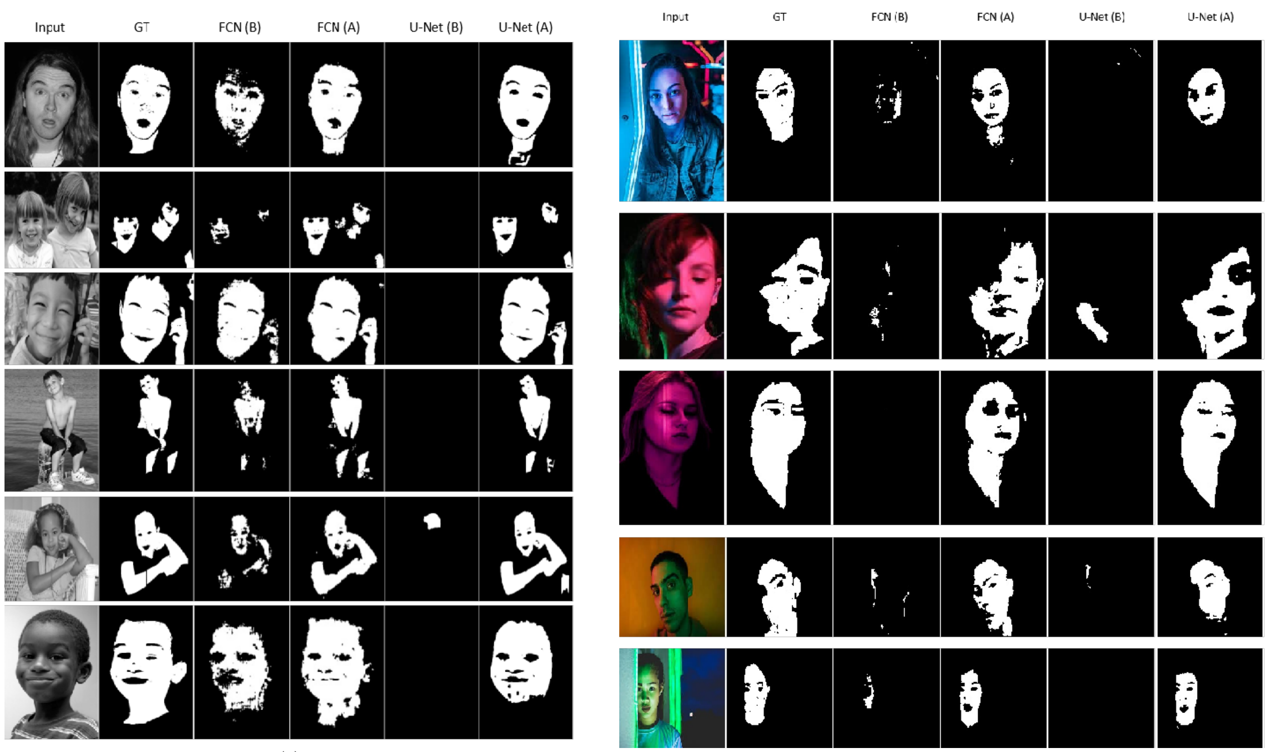


Figure 4. Testing results on the RFW dataset, by various skin segmentation methods including Kolkur *et al*. [3], Dahmani *et al*. [1], Jones *et al*. [2], FCN before (B) and after (A) augmentation, and U-Net before (B) and after (A) augmentation (Columns 2 to 8). Input are shown in column 1. Result shown for different races: Caucasian, Asian, Indian, and African (Row 1 to 4).



(a) (b)

Figure 5. Testing results on our self-made dataset (a) and grayscale images from the ECU dataset (B) by deep learning models FCN and U-Net. The label (B) on the top of the images refers to the results from the model before color augmentation. In comparison, label (A) refers to the model with color augmentation. Input images and ground truth are shown in columns 1 and 2 in each group.