This version of the contribution has been accepted for contribution, after peer review but is not Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available at http://dx.doi.org/10.1007/978-3-031-13321-3_6. Use of this Accepted Version is subject to the publisher's Accepted Manuscript term of use https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms.

From Garment to Skin: The visuAAL Skin Segmentation Dataset *

Kooshan Hashemifard[0000−0001−5086−3064] and Francisco Florez-Revuelta[0000−0002−3391−711X]

Department of Computing Technology, University of Alicante, San Vicente del Raspeig, 03690, Spain
{k.hashemifard,francisco.florez}@ua.es

Abstract. Human skin detection has been remarkably incorporated in different computer vision and biometric systems. It has been receiving increasing attention in face analysis, human tracking and recognition, and medical image analysis. For many human-related recognition tasks, using skin detection cue could be a proper choice. Despite the vast area of usage and applications for skin detection, not many large or reliable skin detection datasets are available, and many of the existing ones, are originally created for other tasks such as hand tracking or face analysis. In this paper, we propose a methodology for extracting skin pixels from garment segmentation and recognition datasets. This is achieved by using deep learning methods to generate automatic skin label masks from them by exploiting human body and hair segmentation and provided garment masks. Following this approach, a large human skin segmentation dataset is introduced. A validation set is also manually segmented in order to evaluate the accuracy of the output skin masks. Finally, usual methods for skin detection and segmentation are evaluated on this new dataset.

Keywords: Skin segmentation · Dataset.

1 Introduction

Padilla et al. [22] proposed a privacy-by-context approach to provide privacy in video data, particularly in active and assisted living applications. The context is given by a number of variables: (i) the observer; (ii) the identity of the person (to retrieve the privacy profile); (iii) the closeness between the person and observer (e.g., relative, doctor or acquaintance); (iv) appearance (dressed?); (v) location (e.g., kitchen); and (vi) ongoing activity or detected event (e.g., cooking, watching TV, fall). The automated recognition of these variables is a requirement to be able to provide privacy appropriately. Among these, appearance recognition, i.e.

* This work is part of the visuAAL project on Privacy-Aware and Acceptable Video-Based Technologies and Services for Active and Assisted Living (https://www.visuaal-itn.eu/). This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 861091.

determining the degree of nudity of a person, is one of them. Nudity estimation requires skin segmentation.

Skin classification is the act of separating skin pixels (or regions) in an image from non-skin ones which could be background pixels or body pixels covered by clothes [19]. Skin detection has been used in human biometric applications such as face analysis [21] and medical image analysis [3], as a preprocessing or validation step or to find the location of human beings and their body parts [29]. Skin segmentation is also an important task in other applications including content retrieval, robotics, sign language recognition, and human computer interaction [19]. For many human-related recognition and prediction tasks, using skin detection cue could be beneficial, since it can be insensitive to variables such as pose, rotation or facial expression.

Deep learning has been applied for semantic segmentation, which can also be exploited for skin segmentation. However, in order to perform adequately, deep learning methods usually rely on large datasets. The lack of large skin segmentation datasets is still a serious issue. Most of the datasets for this task are either small, noisy or suffering from low quality images, some are borrowed from other tasks such as hand tracking, face detection and face recognition and others are unavailable for public use. In addition to the dataset size, a dataset should cover a variety of poses and nonlinear illuminations, aging, makeups, complex backgrounds, different skin characteristics and colors, and also camera variations. All these problems have led many methods to not be experimented based on standard datasets. Therefore, many papers are validated based on random collection of personal or online public images [19]. However, the fact that developing a large-scale manually segmented dataset can be costly and time consuming cannot be ignored. The manual annotation process for segmentation is very demanding and labor-intensive [9]. As an example, annotating a single image in the Cityscapes dataset costs more than 1.5 hours [4] and it will not be much less for a skin dataset due to variations of clothing or hair styles.

Therefore, the main aim of this work is to propose a methodology to create larger skin segmentation datasets, by exploiting available garment datasets in conjunction with deep learning methods for human body and hair segmentation in order to automatically generate ground truth skin labels for images. This tackles some of the problems mentioned above in order to build large human skin datasets. Following this approach, a new skin segmentation dataset (the visuAAL Skin Segmentation dataset) is introduced. A study on different preprocessing and postprocessing steps is done to evaluate the results, lower the segmentation noise and achieve the most precise masks possible. These evaluations are carried out employing a portion of randomly selected and manually segmented images from the dataset. Finally, usual and state-of-the-art algorithms for semantic segmentation and skin detection are evaluated on this dataset.

The remaining of this paper is organized as follows. Section 2 presents a brief review of existing human skin datasets. Section 3 proposes a methodology for generating skin annotations automatically. In Section 4 the details of the dataset, evaluation metrics and quality assessment are discussed, and Section 5

Table 1: Comparison between skin datasets.

Dataset	Year	Number of images	Annotation Quality
Compaq	2002	13,640	Imprecise
TDSD	2004	554	Imprecise
ECU	2005	6,000	Precise
Schmugge	2007	845	Imprecise
MCG	2011	1,000	Imprecise
HGR	2012	1,558	Precise
Pratheepan	2012	78	Precise
SFA	2013	1,118	Precise

is dedicated to baseline specification and measuring the performance of existing methods on the new dataset. Finally, Section 6 presents some conclusions and future work.

2 Skin Detection and Segmentation Datasets

While in skin detection it is very common for researchers to collect images and use their own datasets, there also exist popular human skin datasets. Though these datasets may not follow the same protocols (i.e., some considered eyebrows and lips as skin and some excluded them) or could be noisy, they still provide an opportunity to evaluate and compare the methods. Also, as mentioned before, some of them are not directly skin datasets but originally developed for face recognition and hand tracking tasks. Table 1 presents a summary of the main details of this datasets. Next, some of the main available datasets are described.

Compaq Dataset [14]. Compaq is one of the largest datasets including 13,640 images divided into two groups of skin and non-skin. Skin ground truths are semi-manually labeled. This dataset roughly contains 1 billion pixels in total, including more than 80 million skin pixels. Compaq has been widely used by the research community. However, low quality images and noisy labels are considerable issues.

ECU Dataset [24]. ECU consists of 4,000 high quality color images with high accuracy ground truth which are segmented manually for face detection and skin segmentation. Different skin types, backgrounds, and illuminations make this dataset more diverse. It contains 4.9 million skin pixels and 13.7 million non-skin pixels.

SFA Dataset [2]. This dataset is collected based on FERET [23] and AR [20] face images datasets. It includes 1,118 images. Though the ground truth is precise (lips, eyebrows and eyes are excluded), it mostly consists of face images only.

Fig. 1: Samples from the FashionPedia dataset.

Schmugge Dataset [27]. Schmugge et al. collected a general and diverse dataset consisting of 845 images with nearly 5 million skin pixels and 13.7 million non-skin pixels. The ground truths are generated in a semi-supervised way and are noisy.

HGR Dataset [15]. This dataset was collected by Kawulok et al. for hand gesture recognition and includes 1,558 images with different sizes, backgrounds and conditions.

MCG Dataset [12]. MCG-skin includes 1,000 images randomly collected from the web and social media. Images cover a variety of different backgrounds, skin colors and races, and illumination conditions with diverse quality and resolution. In this dataset, eyes, lips and eyebrows are labeled as skin.

TDSD Dataset [36]. It consists of 554 images randomly picked from the web with over 24 million skin pixels and 75 million non-skin pixels. Skin ground truths are labeled manually using Photoshop.

Pratheepan Dataset [31]. The images in this dataset are randomly collected from the web. The images were captured with a range of cameras using different color enhancement techniques and under different illuminations. Though the dataset is diverse in terms of background and lighting and the ground truths are very precise, it only consists of 78 images and is mostly used as a benchmark dataset for evaluation purposes.

3 Approach

As mentioned in Section 1, due to the difficulties of creating large skin segmentation datasets and overwhelming annotation process, a common alternative

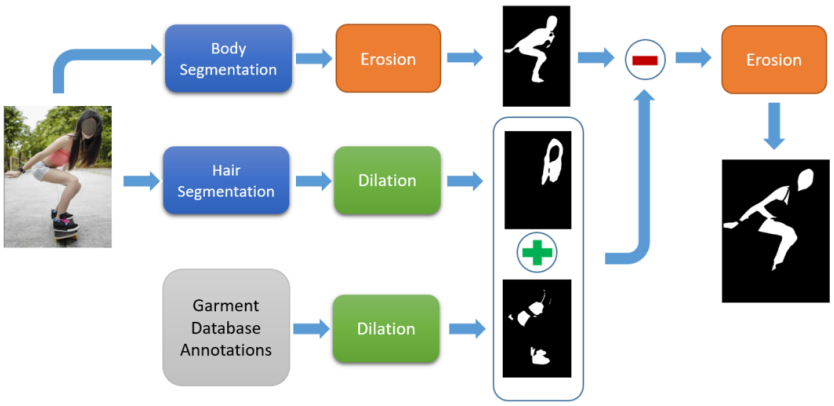


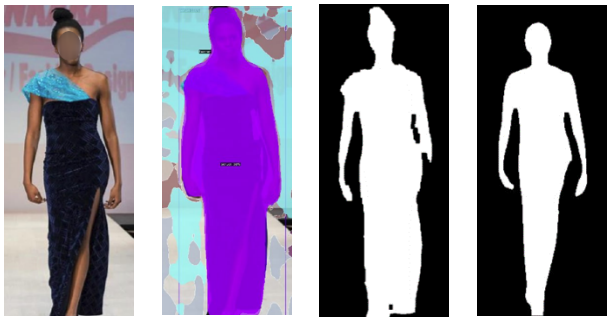
Fig. 2: Pipeline to obtain skin masks from garment masks.

approach is to adapt datasets from homogeneous tasks. In this paper, we propose a methodology to take advantage of available fashion and garment datasets by using their provided clothing annotations, for instance, the masks provided by the FashionPedia [13] dataset are shown in Figure 1. The approach proposed in this work is to obtain the skin areas as the inverse of those masks, as they provide a close result for skin segmentation. However, hair areas also need to be removed from the human body area. This approach is shown in Figure 2.

There are several accurate methods for human body segmentation. Some of them have been tested in this work, namely DensePose [8], Mask R-CNN [11] and MediaPipe framework [18]. DensePose empirically works better than the other ones in abnormal or twisted body poses or camera rotations, since DensePose is specifically developed for human body segmentation and it is the one used in this paper. Mask R-CNN is a general segmentation method for predicting multiple object classes simultaneously. Furthermore, as shown in Fig. 3, extra margins add unwanted background that cannot get modified or erased easily from the output. In contrast, DensePose boundaries are much more precise. MediaPipe, even though it has a dedicated body segmentation model and performs acceptable in normal situation, its performance drops drastically with complex backgrounds or unusual body poses.

Although DensePose usually excludes most parts of the hair area from the body mask, a method is required to remove all remaining hair areas, for instance, hair covering torso areas. For this purpose, a PSPNet model [34] is trained on the Figaro dataset [30], a hair segmentation dataset with 1,000 images and manual precise ground truth. Using a pyramid pooling module, PSPNet can take object relationships into account, which leads to increase in accuracy for irregular hair styles and colors.

The three obtained masks (body area, clothing, and hair) are processed using mathematical morphology operators in order to remove noise and improve the results. An study to select these operators is presented in Section 4.2.



(a) (b) (c) (d)

Fig. 3: Body segmentation results. (a) Original image, (b) Mask R-CNN, (c) MediaPipe and (d) DensePose.

4 Case Example: the visuAAL Skin Segmentation Dataset

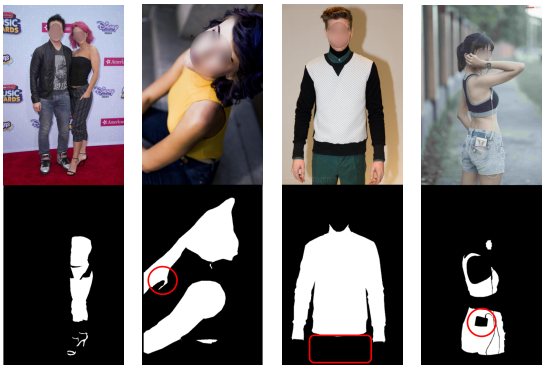
In order to show the effectiveness of our methodology regarding the creation of datasets for skin segmentation, it is tested on a garment dataset. Next, fashion and garment datasets suitable for our work are explained. Then, a study is carried out on the quality of the results by defining proper evaluation metrics. Furthermore, error sources and ways to address them are discussed.

4.1 Fashion and Garment Datasets

There are several available fashion datasets with cloth labels, bounding box, pose and attributes from which few provide segmentation masks for each cloth parts too. A complete list of datasets is shown in [13]. Only FashionPedia [13], DeepFashion2 [7], Runway2Realway [32], and Modanet [35] have main garment and accessories segmentation at the pixel level. FashionPedia is the newest, consisting of 48,825 of everyday and celebrity event high quality images from different genders and skin colors, with exhaustive cloth segmentation and fine-grained attributes. It has the largest set of labeled cloth categories, and then fewer unlabeled cloth parts which could be mistakenly assigned as skin. To form the complete garment mask, all garment masks for a given image can be extracted and accumulated together. Therefore, We choose to validate our methodology using the FashionPedia dataset.

4.2 Dataset Quality Assessment

Putting aside the garment ground truth masks provided in FashionPedia, which can be assumed reliable, body segmentation and hair segmentation could impact the accuracy of the obtained skin masks. In order to measure the quality and



(a) (b) (c) (d)

Fig. 4: Different error sources of the dataset. (a) Wrong detected body, (b) noisy body boundaries, (c) missed annotation in FashionPedia, (d) object covering body mask.

performance of the proposed methodology, 100 images were randomly selected and manually segmented using the CVAT open source image annotation tool [28].

The evaluation is carried out by comparing these manual ground truths with the masks generated by our algorithm. In semantic segmentation and skin detection, evaluation metrics quantify the difference between ground truth and output segmentation and the performance is usually evaluated using statistical measures including: Precision (P), Recall (R), F1-score (F1), Correct Detection Rate (CDR), a.k.a Accuracy, Intersection over Union (IoU), a.k.a Jaccard Index, and Dice Similarity Coefficient (DSC).

Given that in most images, the background is larger than the area covered by the person, causing class imbalance between skin and non-skin pixels, TN and CDR are not suitable metrics for skin segmentation evaluation. On the other hand, Precision and Recall and their integration (F1-score) are designed to deal with skewed datasets. IoU and DSC (which are highly correlated) have been proven to be more useful for segmenting small objects in an image and could be proper choices [6].

Figure 4 shows common error sources when generating the skin masks. First, DensePose might detect wrong body areas when several persons are in the image (since FashionPedia only provides garment mask for the main person in it). To address this issue, we compare the body area obtained with DensePose and the garment mask obtained from FashionPedia. If both overlap, the body and the garment masks belong to the same person. Otherwise the image is removed from the dataset.

The second issue is related to the accuracy of the body and hair segmentation. Though DensePose is much more accurate comparing to other methods, there are still cases in which body masks include some background pixels around body boundaries. This issue can also happen in hair removal as the segmentation

Table 2: Morphology filters effect on the visuAAL Skin Segmentation Dataset.

Body Erosion	Garment Dilation	Hair Dilation	Skin Erosion	Precision	Recall	F1-score	CDR
7	7	7	7	96.05%	85.94%	90.71%	98.58%
7	5	5	5	95.72%	88.07%	91.74%	98.72%
5	**1**	**3**	**1**	**93.45%**	**93.05%**	**93.25%**	**98.92%**
5	1	1	7	93.11%	93.27%	93.19%	98.90%
3	1	3	3	92.28%	94.49%	93.37%	98.92%
3	1	1	7	91.94%	94.72%	93.31%	98.91%
1	1	1	1	90.44%	95.8%	93.04%	98.85%
0	1	3	0	90.78%	95.56%	93.11%	98.86%

model may fail to detect all the hair pixels. To address and reduce these errors, image processing operations and morphology filters are exploited. A study has been done over 100 manually segmented images using grid search, to find appropriate filter sizes (between 0 and 9). The best results are presented in Table 2. We have employed the one in bold, as it gets good results in terms of F1-score and CDR, and precision and recall are balanced. For the body mask obtained with DensePose, erosion filters showed the best effect on the extra unwanted margins. For garment and hair masks, dilation filters are used to fill the small holes and cover errors in the boundaries between skin and cloth parts. Though these filtering may cause losing some amount of the actual skin pixels, they exclude the non-skin parts from skin masks which is much more important for the purpose of this dataset.

The third issue is due to wrong mask labels provided in the FashionPedia dataset. Although it includes very accurate garment annotations, there are images in which one part of clothing, for instance pants, are skipped or missed. In these cases, since there are no correct garment labels to get subtracted from body mask, those body parts would be assigned to skin class in the output. Although these cases are rare, they cannot be detected in an automatic labeling.

The last case are small clothing parts which classes are not included in the FashionPedia dataset such as necklaces and wristbands, and external objects which may exist in the body boundaries such as cups or phones. Since these parts are blocking the garment, they are not included in the garment annotations. Then, the pixels belonging to them will be added to the skin masks. These cases are also rare and the objects are mostly small. Therefore, they are ignored in this work, considering them as some amount of noise of our dataset.

4.3 The visuAAL Skin Segmentation Dataset

Applying the proposed methodology to the FashionPedia dataset, allows the creation of a new large-scale skin segmentation dataset, named the visuAAL Skin Segmentation Dataset. It contains 46,775 high quality images divided into a training set with 45,623 images, and a validation set with 1,152 images, from

Fig. 5: Samples from the visuAAL Skin Segmentation Dataset.

Table 3: Traditional skin detection methods evaluation on the visuAAL Skin Segmentation Dataset.

Method	Precision	Recall	F1-score	CDR	DSC
Histogram	42.11%	74.14%	53.72%	89.72%	53.71%
Watershed	28.78%	82.90%	42.73%	82.12%	42.72%
Adaptive Threshold	38.72%	60.93%	47.35%	89.10%	47.35%

which 100 images have been annotated manually. The dataset is diverse, covering different indoor and outdoor backgrounds, skin tones and body poses. From the FashionPedia dataset, non-skin and garment-only images and images containing multiple persons have been removed. Comparing to Table 1, it includes many more images than previous datasets for skin segmentation, and the annotations can be assumed relatively precise. Some samples of the visuAAL Skin Segmentation dataset with the corresponding skin masks are presented in Figure 5. The dataset is available at https://github.com/visuAAL-ITN/SkinSegmentationDataset.

5 Baseline

Several existing methods for skin detection and semantic segmentation are tested on this new dataset to evaluate their performance. Some well-known traditional skin detection along with more recent deep learning segmentation architectures have been implemented. The experimented traditional methods are inspired by Adaptive Threshold [25], Histogram-based [33] and Watershed [16] methods. The poor performance of these methods showed in Table 3, can implies that they are highly biased on the reported datasets.

There are many works using semantic segmentation based methods as a backbone of their approach. Due to the lack of standard evaluation of these methods, we evaluated the most successful semantic segmentation architectures on our dataset, so that they can be used as a baseline to improve. These architectures

Table 4: Semantic Segmentation Methods Evaluation on the visuAAL Skin Segmentation dataset.

Method	Precision	Recall	F1-score	CDR	DSC
FCN	70.34%	84.88%	76.80%	97.11%	74.40%
SegNet	80.71%	80.12%	80.29%	97.75%	78.82%
UNet	82.66%	85.34%	83.83%	98.01%	82.38%
HLNet	76.50%	79.86%	78.01%	97.43%	76.08%
DSNet	85.80%	85.08%	85.40%	98.35%	84.14%

include standard FCN [17], SegNet [1], UNet [26], HLNet [5] and DSNet [10]. Table 4 shows the results with these deep models.

It is worth mentioning that in the skin segmentation task, Precision means the ratio of actual skin pixels to all predicted skin pixels by the model, and Recall is the percentage of detected skin pixels from all the skin pixels appearing in the image. Therefore, depending on the task, the metric priority can be different.

6 Conclusion

This paper presents a methodology for creating large-scale skin segmentation datasets from garment datasets by exploiting state-of-the-art deep learning and semantic segmentation methods. Unlike previous adapted datasets, which are mostly based on face or hand datasets, this dataset is based on the whole human body. This methodology has been validated by creating a new dataset, the visuAAL Skin Segmentation Dataset, which includes more images and in more diverse conditions than previous datasets. The efficiency of the obtained masks is validated by using a small sample of images manually annotated. In the near future, the dataset will also provide the skin areas related to each specific body part. Adding these annotations to the original garment datasets can make them more diverse and useful for different applications in human related segmentation tasks, and appearance and nudity detection. In future works, we aim at addressing the four issues presented in Section 4.2. To create a more precise version of the dataset, computer vision techniques will be used to remove non-skin areas, such as eyes, eye brows and lips. Additionally, object detection and segmentation methods can be exploited in the body boundaries in order to detect wrong or missing labels in the garment dataset. Finally, this methodology has been applied to the FashionPedia dataset. There are other bigger datasets with garment annotations at the pixel level. However, each of them have specific issues that need to be solved before applying the methodology, e.g. masks are not accurate or unlabeled clothes/accessories.

References

1. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pat-

tern analysis and machine intelligence **39**(12), 2481–2495 (2017)
2. Casati, J.P.B., Moraes, D.R., Rodrigues, E.L.L.: Sfa: A human skin image database based on feret and ar facial images. In: IX workshop de Visao Computational, Rio de Janeiro (2013)
3. Codella, N.C., Nguyen, Q.B., Pankanti, S., Gutman, D.A., Helba, B., Halpern, A.C., Smith, J.R.: Deep learning ensembles for melanoma recognition in dermoscopy images. IBM Journal of Research and Development **61**(4/5), 5–1 (2017)
4. Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B.: The cityscapes dataset for semantic urban scene understanding. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3213–3223 (2016)
5. Feng, X., Gao, X., Luo, L.: Hlnet: A unified framework for real-time segmentation and facial skin tones evaluation. Symmetry **12**(11), 1812 (2020)
6. Furtado, P.: Testing segmentation popular loss and variations in three multiclass medical imaging problems. Journal of Imaging **7**(2), 16 (2021)
7. Ge, Y., Zhang, R., Wang, X., Tang, X., Luo, P.: Deepfashion2: A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5337–5345 (2019)
8. Güler, R.A., Neverova, N., Kokkinos, I.: Densepose: Dense human pose estimation in the wild. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7297–7306 (2018)
9. Hao, S., Zhou, Y., Guo, Y.: A brief survey on semantic segmentation with deep learning. Neurocomputing **406**, 302–321 (2020)
10. Hasan, M.K., Dahal, L., Samarakoon, P.N., Tushar, F.I., Martí, R.: Dsnet: Automatic dermoscopic skin lesion segmentation. Computers in Biology and Medicine **120**, 103738 (2020)
11. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 2961–2969 (2017)
12. Huang, L., Xia, T., Zhang, Y., Lin, S.: Human skin detection in images by mser analysis. In: 2011 18th IEEE International Conference on Image Processing. pp. 1257–1260. IEEE (2011)
13. Jia, M., Shi, M., Sirotenko, M., Cui, Y., Cardie, C., Hariharan, B., Adam, H., Belongie, S.: Fashionpedia: Ontology, segmentation, and an attribute localization dataset. In: European conference on computer vision. pp. 316–332. Springer (2020)
14. Jones, M.J., Rehg, J.M.: Statistical color models with application to skin detection. International journal of computer vision **46**(1), 81–96 (2002)
15. Kawulok, M., Kawulok, J., Nalepa, J., Smolka, B.: Self-adaptive algorithm for segmenting skin regions. EURASIP Journal on Advances in Signal Processing **2014**(170), 1–22 (2014)
16. Khaled, S.M., Islam, M.S., Rabbani, M.G., Tabassum, M.R., Gias, A.U., Kamal, M.M., Muctadir, H.M., Shakir, A.K., Imran, A., Islam, S.: Combinatorial color space models for skin detection in sub-continental human images. In: International Visual Informatics Conference. pp. 532–542. Springer (2009)
17. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431–3440 (2015)
18. Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.L., Yong, M.G., Lee, J., et al.: Mediapipe: A framework for building perception pipelines. arXiv preprint arXiv:1906.08172 (2019)

19. Mahmoodi, M.R., Sayedi, S.M.: A comprehensive survey on human skin detection. International Journal of Image, Graphics and Signal Processing **8**(5), 1 (2016)
20. Martinez, A., Benavente, R.: The ar face database, cvc. Copyright of Informatica (03505596) (1998)
21. Naji, S., Jalab, H.A., Kareem, S.A.: A survey on skin detection in colored images. Artificial Intelligence Review **52**(2), 1041–1087 (2019)
22. Padilla-López, J.R., Chaaraoui, A.A., Gu, F., Flórez-Revuelta, F.: Visual privacy by context: Proposal and evaluation of a level-based visualisation scheme. Sensors **15**(6), 12959–12982 (2015)
23. Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The feret evaluation methodology for face-recognition algorithms. IEEE Transactions on pattern analysis and machine intelligence **22**(10), 1090–1104 (2000)
24. Phung, S.L., Bouzerdoum, A., Chai, D.: Skin segmentation using color pixel classification: analysis and comparison. IEEE transactions on pattern analysis and machine intelligence **27**(1), 148–154 (2005)
25. Rahmat, R.F., Chairunnisa, T., Gunawan, D., Sitompul, O.S.: Skin color segmentation using multi-color space threshold. In: 2016 3rd International Conference on Computer and Information Sciences (ICCOINS). pp. 391–396. IEEE (2016)
26. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
27. Schmugge, S.J., Jayaram, S., Shin, M.C., Tsap, L.V.: Objective evaluation of approaches of skin detection using roc analysis. Computer vision and image understanding **108**(1-2), 41–51 (2007)
28. Sekachev, B. et al.: opencv/cvat: v1.1.0 (Aug 2020). https://doi.org/10.5281/zenodo.4009388
29. Shaik, K.B., Ganesan, P., Kalist, V., Sathish, B., Jenitha, J.M.M.: Comparative study of skin color detection and segmentation in hsv and ycbcr color space. Procedia Computer Science **57**, 41–48 (2015)
30. Svanera, M., Muhammad, U.R., Leonardi, R., Benini, S.: Figaro, hair detection and segmentation in the wild. In: 2016 IEEE International Conference on Image Processing (ICIP). pp. 933–937. IEEE (2016)
31. Tan, W.R., Chan, C.S., Yogarajah, P., Condell, J.: A fusion approach for efficient human skin detection. IEEE Transactions on Industrial Informatics **8**(1), 138–147 (2011)
32. Vittayakorn, S., Yamaguchi, K., Berg, A.C., Berg, T.L.: Runway to realway: Visual analysis of fashion. In: 2015 IEEE Winter Conference on Applications of Computer Vision. pp. 951–958. IEEE (2015)
33. Zarit, B.D., Super, B.J., Quek, F.K.: Comparison of five color models in skin pixel classification. In: Proceedings International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems. In Conjunction with ICCV'99 (Cat. No. PR00378). pp. 58–63. IEEE (1999)
34. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2881–2890 (2017)
35. Zheng, S., Yang, F., Kiapour, M.H., Piramuthu, R.: Modanet: A large-scale street fashion dataset with polygon annotations. In: Proceedings of the 26th ACM international conference on Multimedia. pp. 1670–1678 (2018)
36. Zhu, Q., Wu, C.T., Cheng, K.T., Wu, Y.L.: An adaptive skin model and its application to objectionable image filtering. In: Proceedings of the 12th annual ACM international conference on Multimedia. pp. 56–63 (2004)