 doi:10.20944/preprints202210.0366.v1

Article

Image Segmentation for Human Skin Detection

Marcelo Leite [1]*, Wemerson Delcio Parreira [1]*, Anita Maria da Rocha Fernandes [1] and Valderi Reis Quietinho Leithardt[2,3]

[1] Master Program in Applied Computer Science, School of Sea, Science and Technology, University of Vale do Itajaí. (M.L, A.M.R.F and W.D.P); Email: marceloleite@edu.univali.br; {parreira, anita.fernandes}@univali.br
[2] VALORIZA, Research Center for Endogenous Resources Valorization, Instituto Politécnico de Portalegre, 7300-555, Portalegre, Portugal. (V.R.Q.L.); E-mail: valderi@ipportalegre.pt
[3] COPELABS, Universidade Lusófona de Humanidades e Tecnologias,1749-024 Lisboa, Portugal
* Correspondence: M.L. and W.D.P. {marceloleite@edu.univali.br, parreira@univali.br}

Abstract: Human skin detection is the main task for various human-computer interaction applications. For this, several computer vision-based approaches have been developed in recent years. However, different events and features can interfere in the segmentation process, such as luminosity conditions, skin tones, complex backgrounds, and image capture equipment. In digital imaging, skin segmentation methods can overcome these challenges or part of them. However, the images analyzed follow an application-specific pattern. In this paper, we presented an approach that uses a set of methods to segment skin and non-skin pixels in images from uncontrolled or unknown environments. Our main result is the ability to segment skin and non-skin pixels in digital images from a non-restrained capture environment. Thus, it overcame several challenges, for instance, lighting conditions, compression, and scene complexity. By applying a segmented image examination approach, we determine the proportion of skin pixels present in the image, considering only the objects of interest, i.e., the people. In addition, this segmented analysis can generate independent information regarding each part of the human body. The proposed solution produced a dataset composed of a combination of other datasets present in the literature, which enabled the construction of a heterogeneous set of images.

Keywords: skin segmentation, skin detection, computer vision, digital image processing

1. Introduction

Due to the natural process of technological evolution, demands that require digital recognition from digital image processing and computer vision are increasingly appearing, such as people and movement identification, and human skin recognition, among others. Of these, skin segmentation is an important step to enable several computer vision-based applications such as: facial expression recognition [1], nudity detection [2,3] or child pornography[4], body motion tracking, gesture recognition[1], skin disease diagnostics, among other human-computer interaction (HCI) applications.

Methods commonly used in the skin segmentation problem, in general, are situation or application specific [5,6]. In the case of face recognition, the methods segment skin and non-skin pixels based on face detection [7]. While, in applications of medicinal interest, e.g. abdominal region, they tend to solve the problem by actually considering the existence of elements corresponding to the abdomen in the examined image [8].

In these situations, the entire examination is applied in a controlled environment – the input image has previously known aspects, considering lighting conditions, capture equipment and objects present. There are approaches suggested by several authors, which seek to segment the skin pixels according to each particular one in an optimal way [9,10]. For example, in an application related to biometrics, the response time for the user is fundamental, together with an acceptable accuracy for the expected result of this application [11]. Otherwise, in an application with medical purposes, the processing speed tends not

 doi:10.20944/preprints202210.0366.v1

to represent so much relevance, otherwise, the accuracy becomes fundamental, due to the characteristics of the analysis, which primarily aims to increase the accuracy of the result and not the prediction time. Therefore, considering the application and its requirements, one approach may be more suitable than another in segmenting skin and non-skin pixels.

There are several problems – depending on the approach – when the input images have no given pattern, either because of difficulty in segmenting the elements belonging to the background of the objects of interest or the quality. The quality is affected by several factors such as image capture equipment, image illumination conditions during the capture process, and sampling process during digitization [5]. Therefore, this work aims to contribute to improving the image examination process in the forensic field, in which the input images may contain any of the above-mentioned aspects. Although the focus of the work is not related to the algorithms used for the treatment of image data privacy, we consider it to be also an important factor, so rules and definitions will be used according to the following works [12], [13], [14], [15] and [16].

Furthermore – considering the possibility of sharing forensic examination target material through internet usage applications – these images suffer from data compression, which causes the loss of relevant image information, reducing its quality [17]. During compression processing, quantization and sampling methods eliminate similar information from image pixels, severely affecting skin regions. Skin regions have very discrete pixel variations, and these form the texture of the skin. However, this texture ends up being compromised by the compression processing removing from it the texture information [17]. Therefore, applications with a controlled environment, considering images with previously known quality standards, allow approaches and techniques based on skin texture recognition to present the expected results [6,18]. However, these methods do not deliver the same results when examining images that have the skin texture information corrupted by the sampling and quantization process and the hard compression processing performed by sharing applications.

Figure 1 presents the change of an image after sharing the image through the message exchange application. Part of the pixel information was eliminated, once the input image is significantly modified after the compression process. This process induces a loss of results qualities, mainly in applications based on the image with discrete textures or of small skin regions. Therefore, texture analysis is a non-viable method for forensic applications [19].

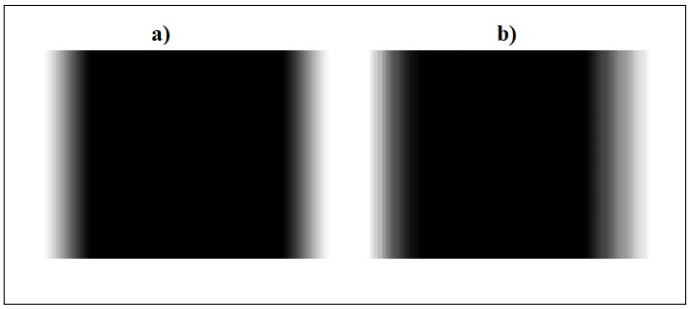

Figure 1. Example of the result of an image submitted to the compression process during transfer by a message transfer application: a) original transferred image; and b) received image with the compression result.

In this work, we propose a solution for skin and non-skin pixels detection in digital images, even with compromised images by events related to capturing and storage. In the following sections, we have reviewed the background of the computational vision algorithm (Section 2). In Section 3, we describe the skin detection process. We present and discuss our results in Section 4. Lastly, conclude this work (Section 5).

 doi:10.20944/preprints202210.0366.v1

2. Background

2.1. Detection of the regions of interest

Considering that the objective of the use of this approach is to segment which pixels of the image are skin or non-skin, as well as quantify the proportion of skin displayed, it was determined as object of interest of analysis the people present in the image, naming these objects as RI (region of interest). The material examined can contain images with complex backgrounds, or even one or more RI's, and for this the first step is to segment these RI's.

In this first step, the task has the essential characteristic of object detection, and to implement this task the proposed approach used an implementation of the *Mask R-CNN* neural network architecture. This is a convolutional neural network, considered the state of the art regarding image segmentation processing and object detection, including instance-based segmentation [20]. Mask R-CNN was developed based on Faster R-CNN, a convolutional neural network based on regions of interest, which favors the goal of this task. During this task, by submitting the examined input image to get the prediction from the network, the coordinates of the IR's present in the image in the form *Bounding Boxes* – Bounding Boxes is obtained. In Figure 2 a conceptual diagram of the operation of this architecture is presented, where in a first phase the input image is submitted a layer responsible for extracting the features and characteristics of the input image. Then the feature descriptors are evaluated by a layer responsible for defining the proposed regions of interest, in other words, possible objects. These objects are then classified, delimited by coordinates and also represented by means of masks.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2022 doi:10.20944/preprints202210.0366.v1

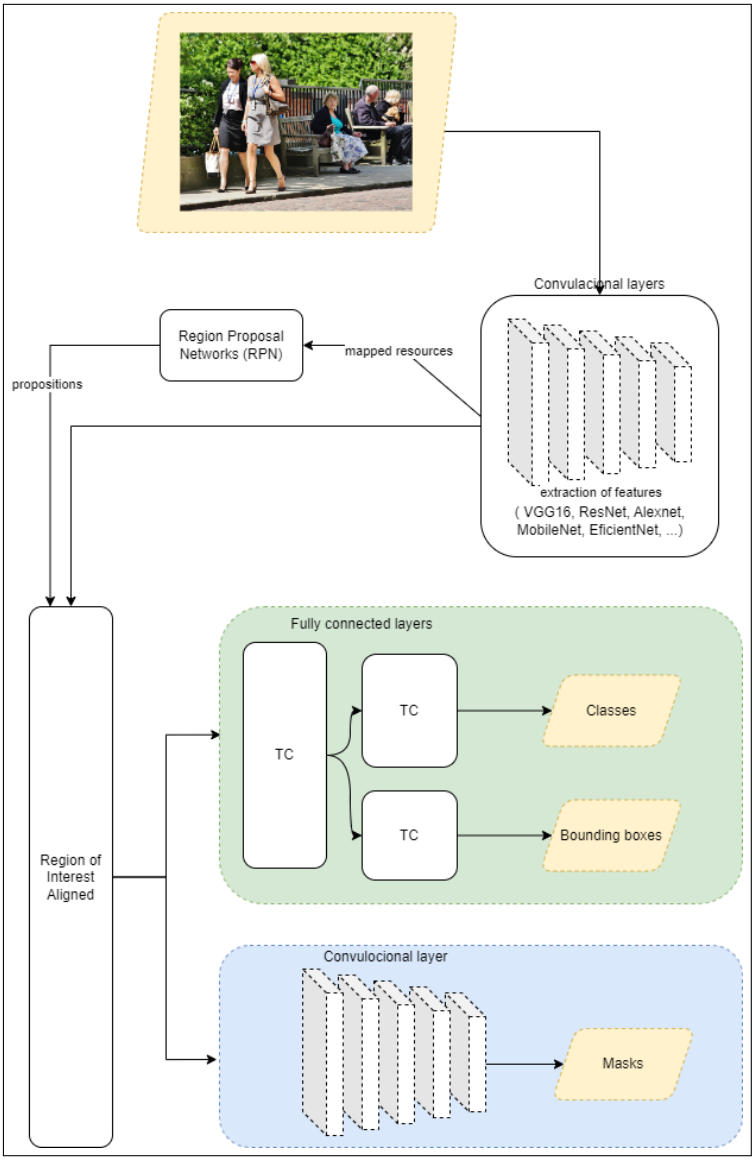

Figure 2. Architecture Diagram *Mask R-CNN*.

For the development of the experiments, it was used the implementation of [21] of Mask R-CNN architecture, and also its pre-trained model, considering for the purpose of this work only objects of class *person* that is the IR objective explored in this task. After the execution of the evaluation of the input image by the implementation of the neural network, is result obtained the delimiting coordinates of the objects of interest. In Figure 3 this result is presented from an examined example image, which has a few people present within a complex scene.

2.2. Segmentation of body parts

This work, in addition to presenting the segmentation of skin and non-skin pixels, evaluates each segmented RI from the previous step to determine the proportion of skin on display. In approaches proposed by other authors, especially when dealing with specific applications, such as biometrics applications, the determination of the proportion rate of skin present in the image [1,4,22,23] is not included in the studies.

Considering the forensic application, for example, the proposed approach needs to estimate the proportion of visible skin in the examined image. Digital images can contain large regions of visible skin due to the size of the image without generating forensic interest, or they can contain small regions of skin – due to the total size of the image – and present

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2022 doi:10.20944/preprints202210.0366.v1

Figure 3. Example of a processed image after submission of the region of interest recognition process.

objects of forensic interest. Thus, this algorithm needs to present the complete region of the human body present in the image, as well as have this region segmented by body part. For this task a layer was implemented in the processing using the BodyPix neural network model [24].

BodyPix is a body segmentation model built under the TensorFlow framework, and taking as input an image containing a human body, uses a pre-trained neural network to separate the parts of a body into segments as specified in Table 1. Among the variety of use of this model is the use intended to remove human objects from an evaluated image. Therefore, in this work – especially in the task of determining the proportion of visible skin – it is necessary to extract the visible parts of the human body from the RI´s, and thus calculate the proportion of skin as a ratio of the area of each visible part.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2022 doi:10.20944/preprints202210.0366.v1

Table 1. Index table with the identifier of each body part detected by the Bodypix model.

#	Body part	Description of body part
0	left_face	
1	right_face	
2	left_upper_arm_front	
3	left_upper_arm_back	
4	right_upper_arm_front	
5	right_upper_arm_back	
6	left_lower_arm_front	
7	left_lower_arm_back	
8	right_lower_arm_front	
9	right_lower_arm_back	
10	left_hand	
11	right_hand	
12	torso_front	
13	torso_back	
14	left_upper_leg_front	
15	left_upper_leg_back	
16	right_upper_leg_front	
17	right_upper_leg_back	
18	left_lower_leg_front	
19	left_lower_leg_back	
20	right_lower_leg_front	
21	right_lower_leg_back	
22	left_feet	
23	right_feet	

In this work, the BodyPix model is used to segment the major parts of the human body and thus determine the total pixels in the image corresponding to each human body part. However, because the original segmentation was very detailed, i.e. containing regions for each limb, the body parts were grouped into larger segments than were available from the original model. Parts such as legs (right and left) and feet, were grouped as the class "legs", to consider as a single class all visible pixels in the image examined belonging to legs. Similarly, the other body regions were grouped as described in Table 2.

Table 2. Grouping rules for BodyPix model key parts.

Body Part	Grouped according to Table 1
face	0,1
arm	2,3,4,5,6,7,8,9,10,11
torso	12,13
leg	14,15,16,17,18,19,20,21,22,23

After associating the images to each part of the body, it is possible to measure more precisely the amount of pixels pertaining only to the human body. Therefore, it is possible to calculate the proportion of visible skin in relation to the total pixels that actually belong to the person being evaluated and not under the total pixels of the image, which generates distorted values and is less useful for the purpose of application of forensic objective. In Figure 4 an example of such segmentation is presented. The human body present in the input image (a) is subdivided into four parts that can be examined individually.

For forensic applications, this individual analysis is important, because by allowing one to know the content of each human body part, in each person present in the examined image, it is possible to efficiently and accurately identify the objects of interest and discard the insignificant ones. For example, an image containing only one hand, that hand in

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2022 doi:10.20944/preprints202210.0366.v1

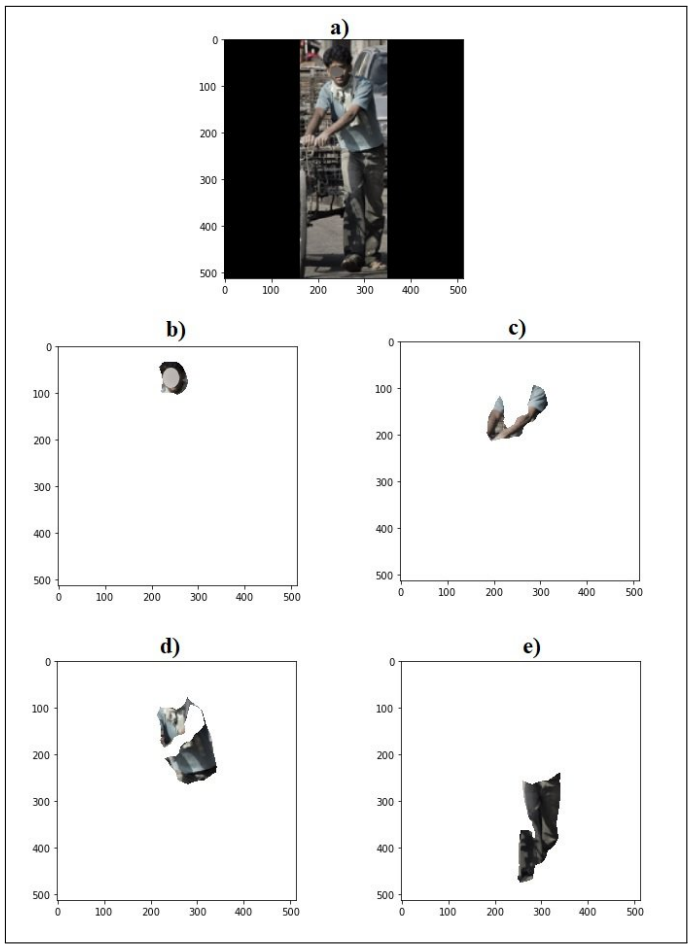

Figure 4. Example of segmentation of human body parts. (a)input image, (b)face, (c)arm, (d)torso and (e)leg.

proportion to the size of the input image may correspond to a large portion in relation to the total size of the image, confusing the analysis because an image with only this fragment would be irrelevant for the evaluation.

3. Skin Detection

3.1. Segmentation

Most approaches on the subject of skin segmentation are based on a particular color space [2,4,23,25], or on fusion components from different color spaces [3,6,26], however, they are evidently limited to color images only. In general, skin color occupies a limited range of values in different color spaces[27], and detection consists of determining an optimal model for the thresholds belonging to the skin and non-skin group[6]. In addition, some studies, such as [26–28], still try to analyze other features such as segmenting regions by texture type, however, this approach needs images with previously known features, because it is necessary to locate the descriptors corresponding to texture, which need to be preserved.

A skin detection process mainly needs two decisions, as presented in Figure 5. The first choice is to determine the appropriate color space in which it is easiest to discriminate skin pixels. Which is a key factor in the design of a skin detector, considering the computational cost of transforming between color spaces [29]. The second decision is to determine the skin segmentation model that will be used to classify a given pixel, being skin or non-skin. Thus, the system must be intelligent and flexible enough to handle various aspects, such as different skin tones or lighting conditions, reducing false negative rates by not identifying a skin pixel as non-skin [7]. In addition, it should be sensitive enough to differences between

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2022 doi:10.20944/preprints202210.0366.v1

classes – background objects with skin color – to reduce false positive rates, i.e., identifying a non-skin pixel as skin [27].

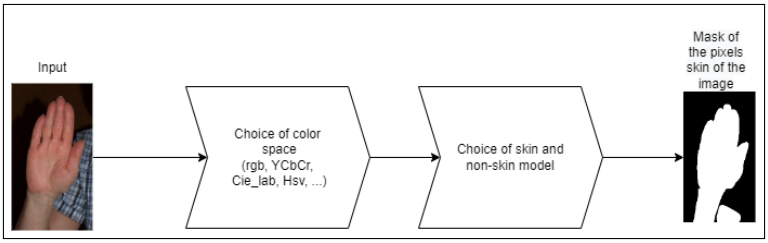

Figure 5. Conceptual diagram of the skin segmentation process.

These choices aim to overcome the main challenges of skin recognition, the definition of an optimal model of skin and non-skin pixels for different skin types [28] and numerous lighting conditions [5,7,9]. Furthermore, the accuracy of skin recognition is affected by pixels of skin-like colors in the image background or clothing. These pixels generate false positives and hinder accurate skin recognition. This problem can be accentuated, when we apply the recognition process in small regions containing skin, for example, a person's hands. Since, the treatment enforced to suppress false negatives can also suppress small true-regions of skin [27].

An example is shown in Figure 6, where there are areas of clothing pixels that are confused with the values of regions belonging to skin tones in the color space. The big question surrounding the recognition of skin and non-skin, is to determine an optimal model for this classification, neither being rigid in the threshold values for skin tones nor being flexible, which would increase the various false positives [9,22].

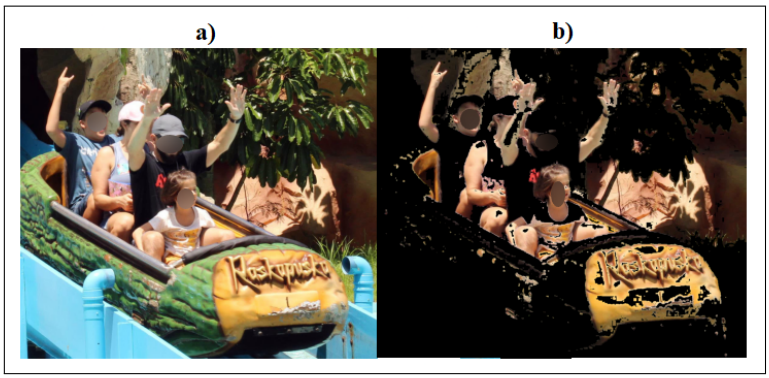

Figure 6. Image with skin-like pixels in clothing and background regions: a) original image and b) image without no-skin pixels

In general, the proposed approaches seek to balance recognition accuracy and required computational power, since human-computer interaction applications need low computational cost coupled with high accuracy [3,4].

The most common strategies for defining skin models can be separated into three approaches. Explicitly defined regions [2,3], where the boundaries are fixed by a set of rules. Parametric methods [7,30] and non-parametric methods [9,26], which are based on histograms applied, or not, to Gaussian models and elliptical boundaries. And finally, there are machine learning-based approaches, which through training of suitable neural network model, enables the identification of skin and non-skin pixel [31].

According to Gonzalez and Woods [32] the fundamental problem of segmentation is the division of an image into regions that satisfy certain conditions set to achieve the end goal. In skin segmentation the problem becomes more complex because the goal is not exactly the detection of an object. It involves examination of image features, such as color

regions of a given color space in color images and intensity levels in non-color images. In general, the process of segmenting skin and non-skin pixels can be represented by

$$G_{(x,y)} = \begin{cases} 0 & \text{se } f_{(x,y)} \leq T \\ 1 & \text{se } f_{(x,y)} > T \end{cases} \quad (1)$$

where the conclusion G is comparison of the boundary $T \in \mathbb{R}$ to the result of the function f, and this represents the function applied by the approach taken in solving the problem.

3.2. Challenges

Considering the classification of pixels in an image as skin and non-skin as a typical image segmentation problem, the adopted approach must have the ability to overcome some difficulties. Depending on the expected result for the final application of the skin identification process, some challenges may substantially affect this result. The following are the main ones:

(i) Lighting conditions: The illumination applied at the moment of image capture directly impacts the result, especially in approaches that use color space analysis to determine if the pixel belongs, or not, to a certain region classified as skin. The lighting variation in the environment where the image was captured creates one of the most difficult difficulties for this operation. Computer vision must have the ability to adjust to this variation, recognizing the same object in different lighting conditions in the same way that human vision has this ability.

(ii) Ethnicity: Ethnic skin tone features represent a difficulty during the classification process because the enlargement of the skin region causes an increase in the occurrence of false positives. Some approaches adopt as a measure to overcome this difficulty the elaboration of a skin pixel model based on the previous detection of the face of the image object, and from the resulting map classifies the image pixels as skin and non-skin. Although this way can work around some cases of skin tone variation, it has no efficient application when there is no face in the image and when there are several people in the same image. Another factor that may make this technique unfeasible is the increased computational consumption as well as the increased processing time.

(iii) Background: The accuracy of skin detection, in the process of segmentation, is severely affected when the background is complex and contains textures and colors similar to the skin region in the color space. The increase in false positives makes skin detection in certain images impractical. Some authors consider this situation a probabilistic problem and propose the definition of a skin probability map for each pixel of the image. Although the result can significantly reduce the occurrence of false positives, there is an increase in the need for computational power, which can make the approach unfeasible for certain applications.

(iv) Characteristics of the scanned image: Applications in which the scanned images have variable characteristics impact the result obtained. Images that have multiple objects, small objects, and objects in different perspectives, represent increased difficulty in the segmentation process. The process needs to have the ability to detect skin without being able to use features such as skin texture. In computer vision, the ability to detect an object is usually related to a set of pre-known characteristics of the object, such as size, position, and quantity, thus generating a certain restriction for the result. It happens that in some final application demands, the condition of the objects in the image is not known and the computer vision system needs to have the same capacity as the human vision to obtain the result waiting independently of the characteristics of the objects in the examined image.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2022 doi:10.20944/preprints202210.0366.v1

3.3. Color Space

Skin detection can be considered a binary pixel classifier with two classes, skin and non-skin [33]. The classification achievement depends on an appropriate feature set for capturing essential elements of these pixels [7]. Each application needs the appropriate color model (RGB, HSV, YCbCr, and others) [9]. When we consider the accuracy, the computational cost required for transformation, and the separability of the pixels with skin colors as the principal decision point [7]. Colorimetry, computer graphics, and video signal transmission standards have given rise to many color spaces with different properties. We have observed that skin colors differ more in luminance intensity than in chrominance due to lighting variation. Therefore, it is common to apply a linear or nonlinear transform in the RGB color space. This process changes the input original space into another color space with independent components, eliminates in the classification process the luminance component analysis, and preserves the chrominance components [5]. Colorspace is made up of two components, chrominance, and luminance [5], and can be subdivided into four groups[33,34]:

(i) Basic – *RGB, CIE-XYZ*: The RGB color model consists of 3 color channels, R is red, G is green and B is blue. Each channel has a value ranging from 0 to 255. This color space was originally developed for the old CRT(Cathodic Ray Tube) [34] monitors. Due to the model mixing the luminance and chromatic components it is not always the most suitable color model for classifying pixels as skin and non-skin, as the variation in illumination greatly affects the accuracy of the result obtained from the classifier. The CIE (Commission Internationale de l'Eclairage) system describes color as a luminance component Y, and two additional components X and Z. According to Kakumanu et al. [5] the *CIE - XYZ* values were constructed from psychophysical experiments and correspond to the color matching characteristics of the human visual system.

(ii) Orthogonal – *YCbCr, YDbDr, YPbPr, YUV, YIQ*: The common goal among these color models is to represent the components as independent as possible, reducing redundancy between their channels, unlike basic models [4,34]. The luminance and chrominance components are explicitly separated, favoring their use in skin detection applications, and that in this case discards the luminance component [5].

(iii) Perceptive – *HSI, HSV, HSL, TSL*: The RGB color space does not directly describe perceptual features of color, such as hue (H), saturation (S), and intensity (I), and many nonlinear transformations are proposed to map RGB to perceptual features. The HSV space defines the property of a color that varies in the passage from red to green as Hue, the property of a color that varies in the passage from red to pink as Saturation, and the property that varies in the passage from black to white as Intensity/Brightness/Luminance/Value [5]. HSV can be a very good choice for skin detection methods because the transformation from RGB to HSV is invariant to high intensity in white lights, ambient light, and surface orientations relative to the light source.

(iv) Uniforms – *CIE-Lab, CIE-Luv*: According to [5], perceptual uniformity describes how two colors differ in appearance to a human observer. However, perceptual uniformity in these color spaces is obtained at the expense of heavy computational transformations. In these color spaces, the calculation of luminance (L) and chromaticity (ab or uv) is obtained by a nonlinear method of mapping the coordinates XYZ.

3.4. Methods

Our approach is based in four task, namely:

(i) Explicit boundaries: In this method a small region in a given color space is selected, where the pixels that belong to that region are considered skinned. This approach is one of the simplest and most widely used, although it has many limitations regarding

 doi:10.20944/preprints202210.0366.v1

the challenges already discussed in this article. Also known in the literature as *thresholding*, it seeks to determine a threshold for considering pixels belonging to skin. Many approaches improve their results by applying conversion techniques from the RGB color space to other color spaces where one can work with the chrominance and luminance values in separate ways. As an example, Basilio [2], converts the RGB color space to *YCbCr* and considers a *thresholding* of $80 \leq Cb \leq 120$ and $133 \leq Cr \leq 173$.

(ii) Histogram based: A skin and non-skin color model is obtained by training with a training dataset, where skin and non-skin pixels are identified. After obtaining the global histogram based on this data set, the color space is divided into two classes of pixels (skin and non-skin). This approach is widely used [1,23] as it shows better results on varied condition images and needs low computational power for its execution. Buza, et al [26] used in their work a histogram-based approach as the basis for their hybrid implementation, which ultimately classified skin and non-skin pixels using a clustering algorithm *k-means*.

(iii) Neural Networks: Neural networks play an important role in research related to skin segmentation, especially the MLP (MultiLayer Perceptron) model. Given a dataset of skin and non-skin samples, and the determination of learning parameters, the net adjusts synaptic weights according to the expected training result. After obtaining the network training it is possible to classify skin and non-skin pixels in an analyzed image. For example, to overcome the problems related to ethnic differences, capture conditions such as lighting and object organization, [31] used in his work a neural network architecture composed of convolutional layers followed by a deep learning layer for classification of skin and non-skin pixels.

The final task of the proposed solution is the segmentation of the skin and non-skin pixels present in the examined image for each part of the body previously segmented. In this step, the objective is to know the proportion of visible skin concerning the total number of pixels corresponding to the examined part, thus being able to present an accurate indicator of the skin content in the image.

Because it was necessary to determine a color space for this process, the shape that best represented the color hues, especially for a human observer, was chosen by us. Therefore, the H and S components, related to the hue from the HSV model, were mixed with the a and b components of the hue coordinates from the CIE-Lab model. This way we modified the skin tone components to reduce the impact of false positives in images with complex backgrounds – containing skin-like elements. At the same time, we remove components responsible for the pixel brightness intensity.

We applied the U-Net neural network model to determine the skin pixel map. In this case, we elaborate on a heterogeneous dataset – based on the mixture of several datasets collected during the literature review – for network training. It was necessary to assemble this dataset because each one available in the literature – related to skin segmentation – was directed to its respective application. For example, an image dataset for a face recognition application contains only face examples, other applications only contain abdomen examples, and so on. Then by collecting these datasets, we compiled for the network training a new, totally heterogeneous.

3.5. Dataset assembled

The dataset of this project was collected, organized and normalized, containing the union of 7 dataset's collected during the literature review. The images and their respective *ground-truth's* were set to the same image format (JPEG) and were randomly distributed so that the dataset has diversity of images so that similar images are not clustered together.

The collection was based on the following datasets:

(i) SFA [35]

(ii) Abdomen [8]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2022 doi:10.20944/preprints202210.0366.v1

(iii) FSD [36]

(iv) VDM [37]

(v) Pratheepan [9]

(vi) TDSD [38]

(vii) HGR [39]

After the compilation, a new dataset was created, containing 8,209 varied images with their respective ground-truth's. This way, the dataset became diversified, as shown in Figure 7, containing images of close-ups, of body parts with exposed skin areas, other images with complex backgrounds considering several lighting conditions.

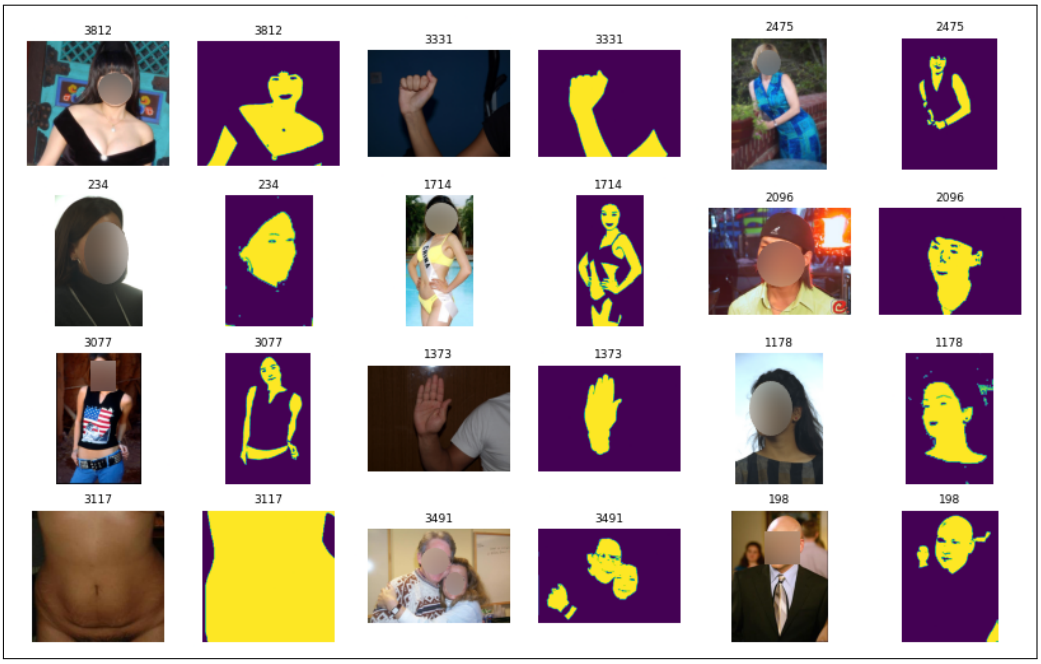

Figure 7. Examples of the images included in the proposed dataset.

U-Net[40] is a neural network model proposed for semantic segmentation and consists of a contraction process followed by an expansion process, as demonstrated in Figure 8. It is a widely used architecture for image segmentation problems, especially in biomedical applications. In this proposal, it was trained based on the aforementioned dataset, with the objective of segmenting skin and non-skin pixels.

It was presented in 2015 and which won the cell tracking challenge of the International Symposium on Biomedical Imaging (ISBI) of the same year in several categories. It was considered, on that occasion, a revolution in deep learning. It uses the concept of fully convolutional networks for this approach. The intention of U-Net is to capture both context and location characteristics and this process is successfully completed by the type of architecture built. The main idea of the implementation is to use successive contracting layers, which are immediately followed by the sampling survey operators to achieve higher resolution results on the input images. The goal of the U-Net architecture is to recognize the features and localization of the examined context, and for this, the main idea is to use successive contraction layers, followed by a decoder composed of expansion layers.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2022 doi:10.20944/preprints202210.0366.v1

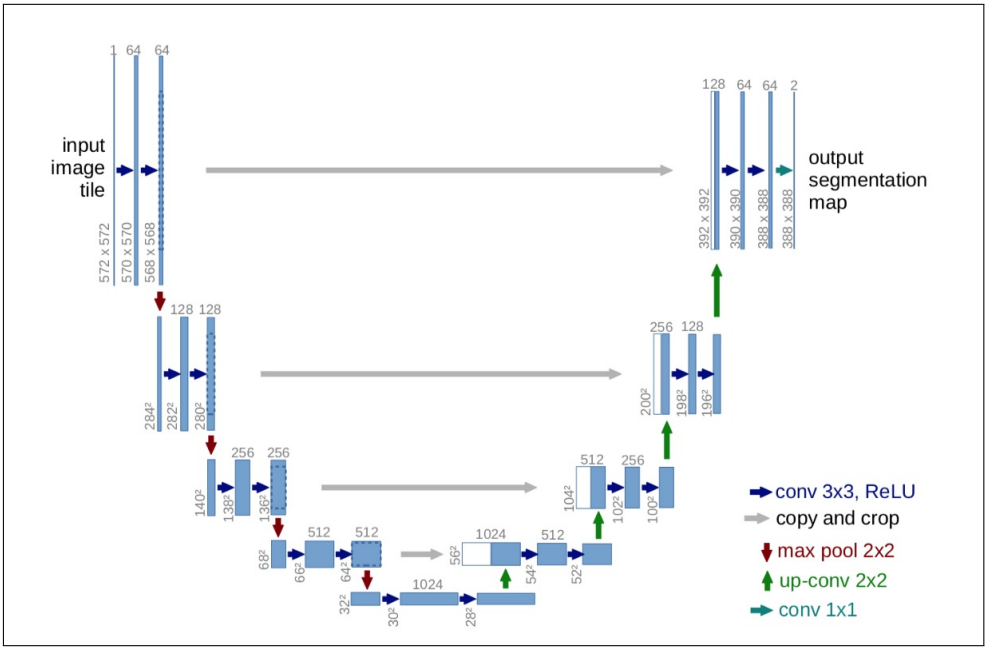

Figure 8. Diagram demonstrating the construction of the U-Net neural network architecture.

4. Results

The set of tasks, explained in the previous sections of the proposed solution, form a system. The image submitted to the system is initially processed by the task of recognition and delimitation of the regions of interest, that is, people present in the image, which finally, if they exist, segments them from the image. Then each segmented region of interest is submitted to the recognition and segmentation of the visible body parts. Each recognized body part is then processed by the next task, which produces the segmentation of the skin and non-skin pixels present in the analyzed body region. Finally the image is then reconstructed, identifying the mask of the pixels recognized as skin in the image, the number of instances of objects classified as person, and the ratio of skin pixels to non-skin pixels – the proportion of visible skin –, for each person recognized in the image. In Figure 9 we presented the macro diagram of the proposed solution.

To evaluate the results, some images and their respective ground truth's were submitted to processing as described. In Figure 10 the result obtained after processing is presented. In the "Predicted Mask" column the neural network prediction of each human body part is displayed, which is presented reassembled according to the input image. In Table 3, the data extracted from each processing step are presented, to then allow to calculate more precisely the estimate of skin presence in the examined image. Analytically, in Table 4 the individualized result per part of each image of the experiment is presented. It is possible to verify the results obtained for each examined part, knowing the ability of skin recognition, in specific applications, according to the part of the body as an objective.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2022 doi:10.20944/preprints202210.0366.v1

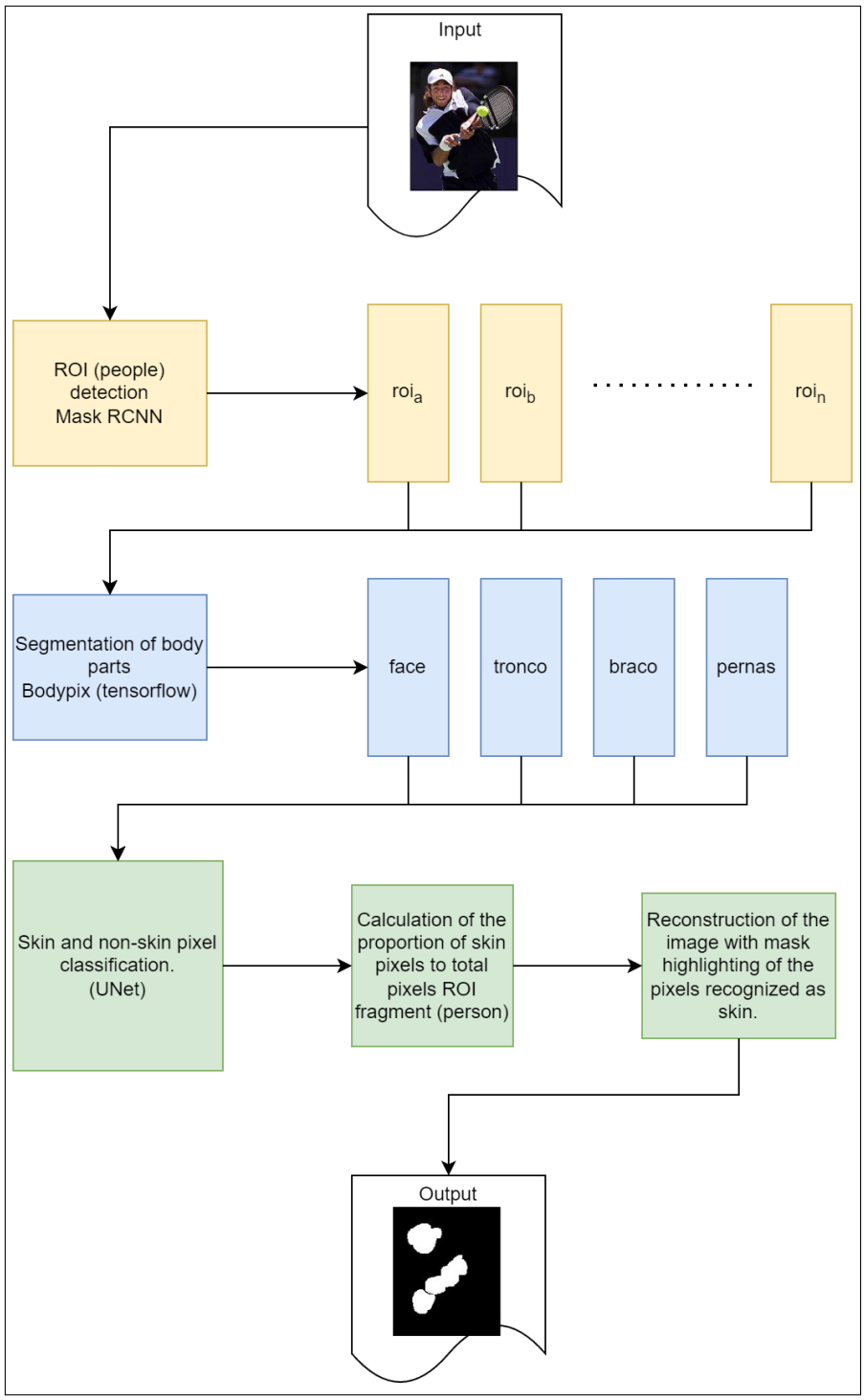

Figure 9. Diagram showing the flow of processing tasks of the proposed solution

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2022 doi:10.20944/preprints202210.0366.v1

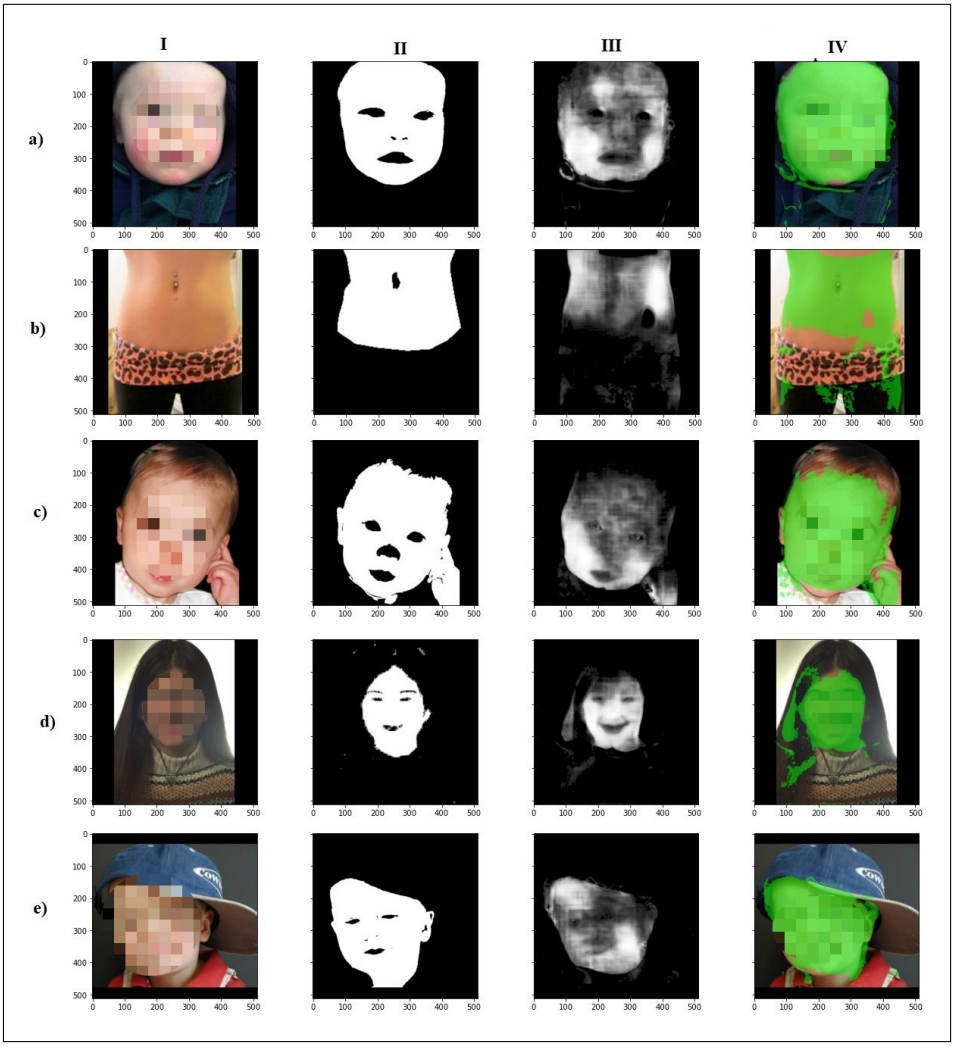

Figure 10. Results obtained from processing. I) corresponds to input, II) ground truth, III) prediction and IV) prediction mask overlaid on input

Table 3. Results obtained from the experiment

Image	Total	Face Skin	% Skin	Total	Arm Skin	% Skin	Total	Torso Skin	% Skin	Total	Leg Skin	% Skin	Acc	TP	TN	FP	FN
a	116977	117803	101%	141	0	0%	56320	2286	4%	0	0	0%	0.910980	97570	141238	22400	936
b	0	0	0%	0	0	0%	118713	88376	74%	53854	12243	23%	0.843704	81796	139376	18823	22149
c	125831	103231	82%	10498	9583	91%	20886	3808	18%	0	0	0%	0.894405	101125	133338	15192	12489
d	75892	46195	61%	0	0	0%	61497	4586	7%	0	0	0%	0.915478	31745	208242	18956	3201
e	106837	76174	71%	0	0	0%	22889	3044	13%	0	0	0%	0.886963	51493	181019	27553	2079

The values shown correspond to the amount of pixels

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2022 doi:10.20944/preprints202210.0366.v1

Table 4. Results obtained from the experiment performed segmented by part

		Prediction			Matrix confusion				Metrics			
Image	**Part**	**Total**	**Skin**	**% Skin**	**TP**	**TN**	**FP**	**FN**	***acc***	***recall***	***precision***	***fscore***
a	face	16977	117803	1.007061	101645	143746	16158	595	0.936092	0.99418	0.862839	0,92
	arm	141	0	0	0	262144	0	0	1	0	0	0
	torso	56320	2286	0.040589	12	258309	2274	1549	0.985416	0.007687	0.005249	0,01
	leg	0	0	0	0	262144	0	0	1	0	0	0
b	face	0	0	0	0	262144	0	0	1	0	0	0
	arm	0	0	0	0	262144	0	0	1	0	0	0
	tronco	118713	88376	0.7444509	81702	159677	6674	14091	0.92078781	0.8529015	0.9244817	0.8872502
	leg	53854	12243	0.227336	0	249901	12243	0	0.953296	0	0	0
c	face	125831	103231	0.820394	93432	150837	9799	8076	0.931812	0.9204398	0.905076	0.912693
	arm	10498	9583	0.912840	7650	251356	1933	1205	0.988029	0.863918	0.798288	0.829808
	torso	20886	3808	0.182323	924	251939	2884	6397	0.964595	0.126212	0.242647	0.166052
	leg	0	0	0	0	262144	0	0	1	0	0	0
d	face	75892	46195	0.608694	36665	212350	9530	3599	0.949917	0.910615	0,793701	0,85
	arm	0	0	0	0	262144	0	0	1	0	0	0
	torso	62497	4586	0.07338	225	255367	4361	2191	0.975006	0.093129	0.049062	0.06
	leg	0	0	0	0	262144	0	0	1	0	0	0
e	face	106837	76174	0.712993	66586	185566	9588	404	0.961884	0.993969	0.87413	0.93
	arm	0	0	0	0	262144	0	0	1	0	0	0
	torso	22889	3044	0.13299	1092	253790	1952	5310	0.972298	0.170572	0.358739	0.23
	leg	0	0	0	0	262144	0	0	1	0	0	0
Total	**face**	**425537**	**343403**	**0.806987**	**298328**	**954643**	**45075**	**12674**	**0.955941**	**0.959248**	**0.86874**	**0.91**
	arm	**10639**	**9583**	**0.900743**	**7650**	**1299932**	**1933**	**1205**	**0,997606**	**0.863919**	**0.798289**	**0.83**
	torso	**281305**	**102100**	**0.362951**	**83955**	**1179082**	**18145**	**29538**	**0.963621**	**0.739737**	**0.822282**	**0.78**
	leg	**53854**	**12243**	**0.227337**	**0**	**1298477**	**12243**	**0**	**0.990659**	**0**	**0**	**0**

The values shown correspond to the amount of pixels.

In Figure 11 another example is presented where more than one person is present in the input image. As can be seen, there are three people in the figure, and they are segmented to be examined individually by the process presented in the proposed solution.

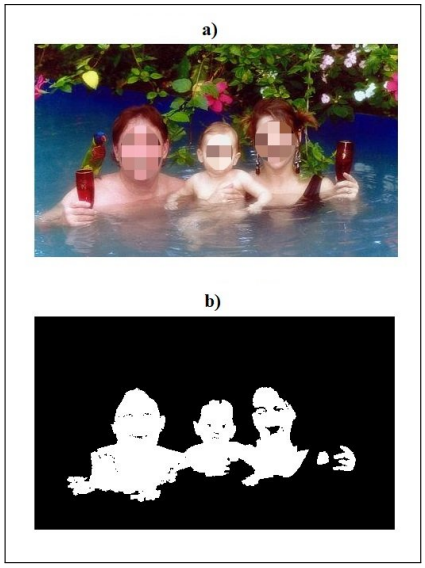

Figure 11. Input image for the example considering the presence of more than one person in the scene. a) input and b) ground truth

The results shown in Figure 12 present the segmented way of examining the process. The first and second lines show the data for the first segmented person, the first line shows the segmentation of the body part (face, arm, torso and leg) and the second line shows the prediction of the neural network corresponding to the segmentation of the skin and non-skin pixels for each part.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2022 doi:10.20944/preprints202210.0366.v1

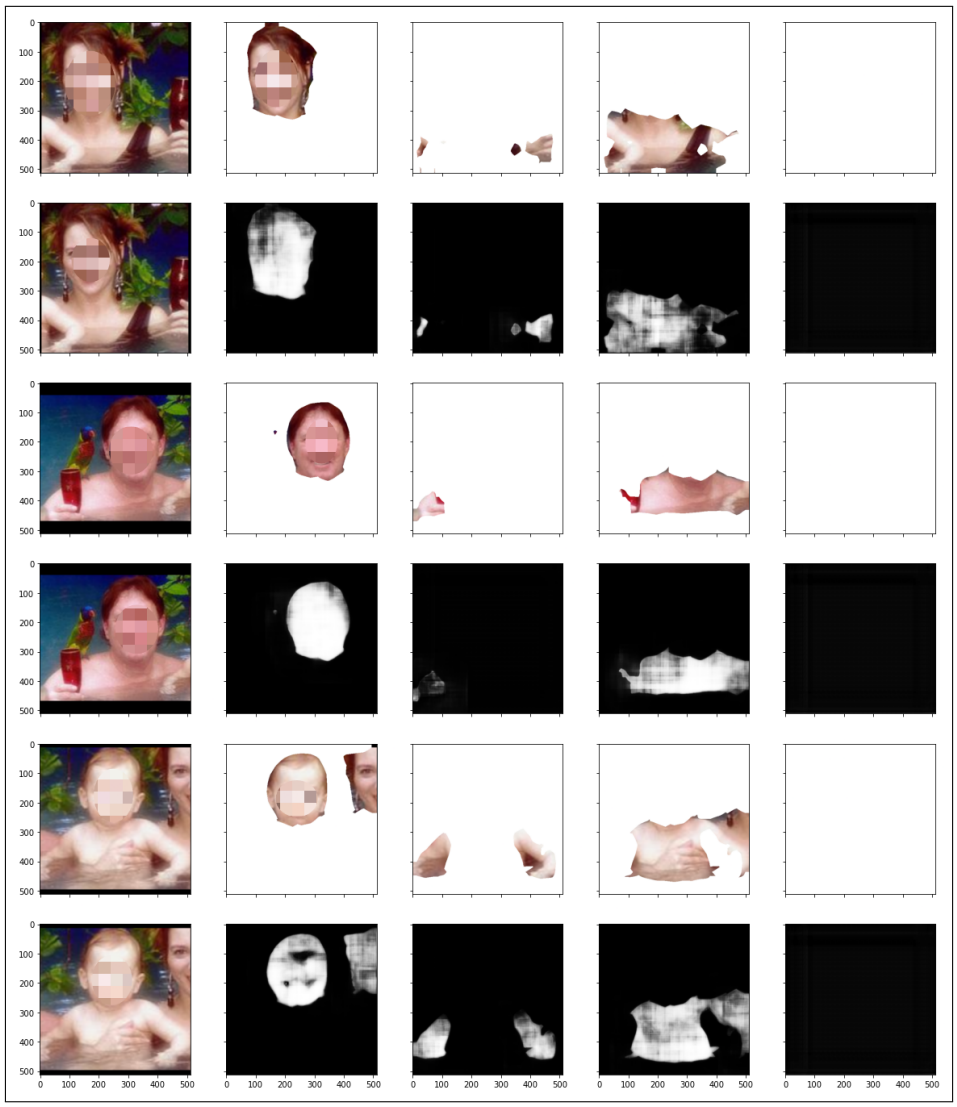

Figure 12. Processing results containing more than one person and showing each body part segment

In Table 5 is described the information based on the composition of the pixels of the image, dividing this information by person and part of the person. For example, it is possible to verify that in the person (c) of Figure 13 has in all parts of the body exposed the presence of a high index of skin presence, except for the legs region. With the known information applying the proposed solution, it may allow an image examination application to consider it potentially relevant for further analysis, considering the high index of skin pixels in specific regions of the human body.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 October 2022 doi:10.20944/preprints202210.0366.v1

Table 5. Results obtained from the experiment performed on the image containing more than one person, as shown in the input image Figure 11

Image	Part	Prediction			Matrix confusion				Metrics			
		Total	Skin	% Skin	TP	TN	FP	FN	*acc*	*recall*	*precision*	*fscore*
a	face	59227	61604	1.040134	38913	200540	22691	404	0.912035	0.989725	0,631664	0.77
	arm	8263	9612	1.163258	2196	253337	6416	195	0.974781	0.918444	0.254993	0.4
	torso	68351	65671	0.960791	44283	195871	21388	602	0.916115	0.986588	0.674316	0.8
	leg	0	0	0	0	262144	0	0	1	0	0	0
b	face	45226	46101	1.019347	36495	215987	9606	56	0.963142	0.998468	0.791631	0.88
	arm	6450	1925	0.29845	441	258989	1484	1230	0.989647	0.263914	0.229091	0.25
	torso	47943	37951	0.791586	33437	213509	4514	10684	0.942024	0.757848	0.881057	0.81
	leg	0	0	0	0	262144	0	0	1	0	0	0
c	face	60389	63441	1.050539	53747	198664	9694	39	0.962872	0.999275	0.847197	0.92
	arm	26160	28435	1.086965	18755	233684	9680	25	0.962978	0.998669	0.659574	0.79
	torso	65749	67820	1.031499	49285	193983	18535	341	0.927994	0.993129	0.726703	0.84
	leg	0	0	0	0	262144	0	0	1	0	0	0

The values shown correspond to the amount of pixels

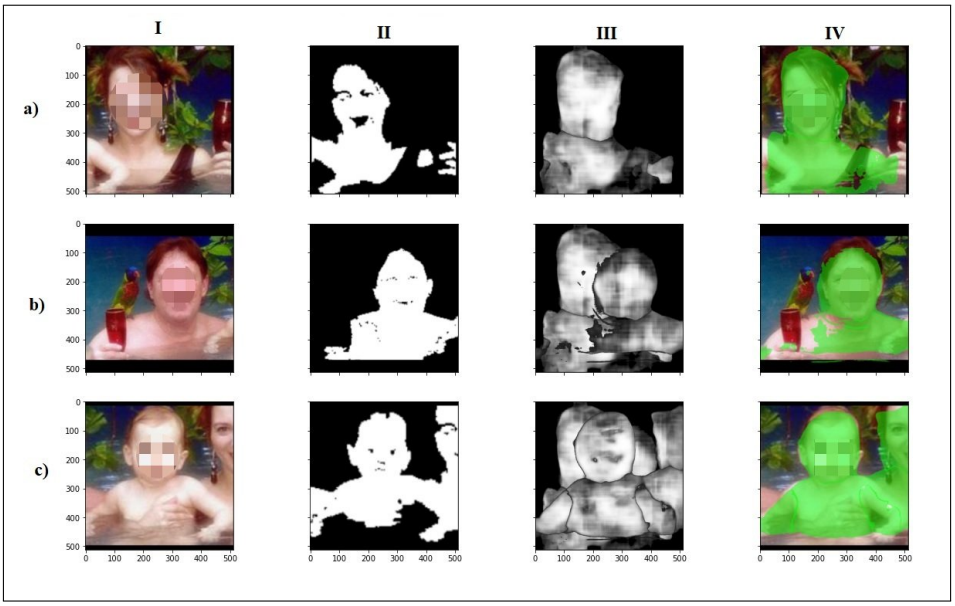

Figure 13. Results of processing the image containing more than one person and showing the final result. I) corresponds to input, II) ground truth, III) prediction and IV) prediction mask overlaid on input

5. Conclusion

Considering the use in applications that need accurate information about the proportion of visible skin in the image, the proposed solution is efficient. We presented, in the experiments of the previous sections, the ability of a solution to recognize the proportion of skin according to the human body part. Our method makes it possible to restrict the recognition of images with visible skin by considering analyzing the relevant human body part, as occurs in cases of forensic image examination. In addition, we also presented a new dataset of images for segmentation, which was generated from the collection of research in the literature review. In future work, we consider the performance analysis of experiments and evaluation of other specific neural network architectures for infrared segmentation, and in the same way, for the segmentation phase of skin and non-skin pixels. An approach may present difficulties in a certain image condition, but it does not mean that this difficulty will result in harm to the final application. In general, we enable control of the capture

process, allowing the simplest approach to present the best result, considering other factors such as time, complexity, and processing power. Another relevant factor to be considered is the level of precision needed, since not all applications need that. For each demand of the final application, there is a suitable approach.

Author Contributions: Conceptualization, M.L, W.D.P., and A.M.d.R.F.; methodology, M.L and W.D.P.; analysis, M.L., W.D.P. and A.M.d.R.F.; writing—original draft preparation, M.L., V.R.Q.L., A.M.d.R.F. and W.D.P.; writing—review and editing, M.L., V.R.Q.L., W.D.P., and A.M.d.R.F. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Fundação para a Ciência e a Tecnologia, I.P. (Portuguese Foundation for Science and Technology) by the project UIDB/05064/2020 (VALORIZA–Research Centre for Endogenous Resource 764 Valorization), and Project UIDB/04111/2020, ILIND–Instituto Lusófono de Investigação e Desenvolvimento, under project COFAC/ILIND/ COPELABS/3/2020.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chakraborty, B.K.; Bhuyan, M.K.; Kumar, S. Fusion-Based Skin Detection Using Image Distribution Model. In Proceedings of the Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing; Association for Computing Machinery: New York, NY, USA, 2016; ICVGIP '16. https://doi.org/10.1145/3009977.3010002.
2. Basilio, J.A.M.; Torres, G.A.; Pérez, G.S.; Medina, L.K.T.; Meana, H.M.P. Explicit Image Detection Using YCbCr Space Color Model as Skin Detection. In Proceedings of the Proceedings of the 2011 American Conference on Applied Mathematics and the 5th WSEAS International Conference on Computer Engineering and Applications; World Scientific and Engineering Academy and Society (WSEAS): Stevens Point, Wisconsin, USA, 2011; AMERICAN-MATH'11/CEA'11, p. 123–128. https://doi.org/10.5555/1959666.1959689.
3. Platzer, C.; Stuetz, M.; Lindorfer, M. Skin Sheriff: A Machine Learning Solution for Detecting Explicit Images. In Proceedings of the Proceedings of the 2nd International Workshop on Security and Forensics in Communication Systems; Association for Computing Machinery: New York, NY, USA, 2014; SFCS '14, p. 45–56. https://doi.org/10.1145/2598918.2598920.
4. de Castro Polastro, M.; da Silva Eleuterio, P.M. NuDetective: A Forensic Tool to Help Combat Child Pornography through Automatic Nudity Detection. In Proceedings of the 2010 Workshops on Database and Expert Systems Applications, 2010, pp. 349–353. https://doi.org/10.1109/DEXA.2010.74.
5. Kakumanu, P.; Makrogiannis, S.; Bourbakis, N. A survey of skin-color modeling and detection methods. *Pattern Recognition* **2007**, *40*, 1106–1122. https://doi.org/doi.org/10.1016/j.patcog.2006.06.010.
6. Nadian-Ghomsheh, A. Hybrid color-texture multivariate Gaussian model for skin detection. In Proceedings of the 2017 10th Iranian Conference on Machine Vision and Image Processing (MVIP), 2017, pp. 123–127. https://doi.org/10.1109/IranianMVIP.2017.8342336.
7. Hassan, E.; Hilal, A.R.; Basir, O. Using ga to optimize the explicitly defined skin regions for human skincolor detection. In Proceedings of the 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), 2017, pp. 1–4. https://doi.org/10.1109/CCECE.2017.7946699.
8. Topiwala, A.; Al-Zogbi, L.; Fleiter, T.; Krieger, A. Adaptation and Evaluation of Deep Leaning Techniques for Skin Segmentation on Novel Abdominal Dataset. In Proceedings of the 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE). IEEE, 2019, pp. 752–759. https://doi.org/10.1109/BIBE.2019.00141.
9. Tan, W.R.; Chan, C.S.; Yogarajah, P.; Condell, J. A Fusion Approach for Efficient Human Skin Detection. *IEEE Transactions on Industrial Informatics* **2012**, *8*, 138–147. https://doi.org/10.1109/TII.2011.2172451.

10. Lumini, A.; Nanni, L. Fair comparison of skin detection approaches on publicly available datasets. *Expert Systems with Applications* **2020**, *160*, 113677. https://doi.org/https://doi.org/10.1016/j.eswa.2020.113677.
11. Chelali, F.Z.; Cherabit, N.; Djeradi, A. Face recognition system using skin detection in RGB and YCbCr color space. In Proceedings of the 2015 2nd World Symposium on Web Applications and Networking (WSWAN), 2015, pp. 1–7. https://doi.org/10.1109/WSWAN.2015.7210329.
12. Suzin, J.C.; Zeferino, C.A.; Leithardt, V.R.Q. Digital Statelessness. In Proceedings of the New Trends in Disruptive Technologies, Tech Ethics and Artificial Intelligence; de Paz Santana, J.F.; de la Iglesia, D.H.; López Rivero, A.J., Eds.; Springer International Publishing: Cham, 2022; pp. 178–189.
13. Apolinário, V.A.; Bianco, G.D.; Duarte, D.; Leithardt, V.R.Q. Exploring Feature Extraction to Vulnerability Prediction Problem. In Proceedings of the International Conference on Disruptive Technologies, Tech Ethics and Artificial Intelligence. Springer, 2023, pp. 79–90.
14. Cesconetto, J.; Augusto Silva, L.; Bortoluzzi, F.; Navarro-Cáceres, M.; A. Zeferino, C.; R. Q. Leithardt, V. PRIPRO—Privacy Profiles: User Profiling Management for Smart Environments. *Electronics* **2020**, *9*. https://doi.org/10.3390/electronics9091519.
15. Martins, J.A.; Ochôa, I.S.; Silva, L.A.; Mendes, A.S.; González, G.V.; De Paz Santana, J.; Leithardt, V.R.Q. PRIPRO: A Comparison of Classification Algorithms for Managing Receiving Notifications in Smart Environments. *Applied Sciences* **2020**, *10*. https://doi.org/10.3390/app10020502.
16. Lopes, H.; Pires, I.M.; Sánchez San Blas, H.; García-Ovejero, R.; Leithardt, V. PriADA: Management and Adaptation of Information Based on Data Privacy in Public Environments. *Computers* **2020**, *9*. https://doi.org/10.3390/computers9040077.
17. TELEGRAPH, T.I.; COMMITTEE, T.C. DIGITAL COMPRESSION AND CODING OF CONTINUOUS-TONE STILL IMAGES. https://www.w3.org/Graphics/JPEG/itu-t81.pdf.
18. Lei, Y.; Xiaoyu, W.; Hui, L.; Dewei, Z.; Jun, Z. An algorithm of skin detection based on texture. In Proceedings of the 2011 4th International Congress on Image and Signal Processing, 2011, Vol. 4, pp. 1822–1825. https://doi.org/10.1109/CISP.2011.6100627.
19. Nian, F.; Li, T.; Wang, Y.; Xu, M.; Wu, J. Pornographic image detection utilizing deep convolutional neural networks. *Neurocomputing* **2016**, *210*, 283–293. SI:Behavior Analysis In SN, https://doi.org/10.1016/j.neucom.2015.09.135.
20. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R.B. Mask R-CNN. *CoRR* **2017**, *abs/1703.06870*, [1703.06870].
21. Abdulla, W. Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow. https://github.com/matterport/Mask_RCNN, 2017.
22. Ma, B.; Zhang, C.; Chen, J.; Qu, R.; Xiao, J.; Cao, X. Human Skin Detection via Semantic Constraint. In Proceedings of the Proceedings of International Conference on Internet Multimedia Computing and Service; Association for Computing Machinery: New York, NY, USA, 2014; ICIMCS '14, p. 181–184. https://doi.org/10.1145/2632856.2632885.
23. Dhantre, P.; Prasad, R.; Saurabh, P.; Verma, B. A hybrid approach for human skin detection. In Proceedings of the 2017 7th International Conference on Communication Systems and Network Technologies (CSNT), 2017, pp. 142–146. https://doi.org/10.1109/CSNT.2017.8418526.
24. Papandreou, G.; Zhu, T.; Kanazawa, N.; Toshev, A.; Tompson, J.; Bregler, C.; Murphy, K. Towards Accurate Multi-person Pose Estimation in the Wild, 2017, [arXiv:cs.CV/1701.01779].
25. Fang, X.; Gu, W.; Huang, C. A method of skin color identification based on color classification. In Proceedings of the Proceedings of 2011 International Conference on Computer Science and Network Technology, 2011, Vol. 4, pp. 2355–2358. https://doi.org/10.1109/ICCSNT.2011.6182445.
26. Buza, E.; Akagic, A.; Omanovic, S. Skin detection based on image color segmentation with histogram and K-means clustering. In Proceedings of the 2017 10th International Conference on Electrical and Electronics Engineering (ELECO), 2017, pp. 1181–1186.
27. Chakraborty, B.K.; Bhuyan, M.K. Skin segmentation using Possibilistic Fuzzy C-means clustering in presence of skin-colored background. In Proceedings of the 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS), 2015, pp. 246–250. https://doi.org/10.1109/RAICS.2015.7488422.
28. Shih, H.C.; Chen, J.Y. Multiskin Color Segmentation Through Morphological Model Refinement. *IEEE Transactions on Emerging Topics in Computational Intelligence* **2021**, *5*, 225–235. https://doi.org/10.1109/TETCI.2019.2892715.
29. Vezhnevets, V.; Sazonov, V.; Andreeva, A. A Survey on Pixel-Based Skin Color Detection Techniques **2004**.

30. Khosravi, S.; Chalechale, A. A hybrid neural network using ICA and CGA for skin detection in RGB images. In Proceedings of the 2016 Al-Sadeq International Conference on Multidisciplinary in IT and Communication Science and Applications (AIC-MITCSA), 2016, pp. 1–6. https://doi.org/10.1109/AIC-MITCSA.2016.7759950.
31. Arsalan, M.; Kim, D.S.; Owais, M.; Park, K.R. OR-Skip-Net: Outer residual skip network for skin segmentation in non-ideal situations. *Pattern Recognition* **2019**.
32. Gonzalez, R.C.; Woods, R.C. *Processamento Digital de Imagens*, 3 ed.; Pearson, 2010.
33. Xiong, W.; Li, Q. Chinese skin detection in different color spaces. 2012, pp. 1–5. https://doi.org/10.1109/WCSP.2012.6542853.
34. Moreira, D.C.; Fechine, J.M. A Machine Learning-based Forensic Discriminator of Pornographic and Bikini Images. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), 2018, pp. 1–8. https://doi.org/10.1109/IJCNN.2018.8489100.
35. 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 Proceedings of the IX workshop de Visao Computational; , 2013.
36. Phung, S.; Bouzerdoum, A.; Chai, D. Skin segmentation using color pixel classification: analysis and comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2005**, *27*, 148–154. https://doi.org/10.1109/TPAMI.2005.17.
37. SanMiguel, J.C.; Suja, S. Skin detection by dual maximization of detectors agreement for video monitoring. *Pattern Recognition Letters* **2013**, *34*, 2102–2109. https://doi.org/https://doi.org/10.1016/j.patrec.2013.07.016.
38. 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 Proceedings of the 12th Annual ACM International Conference on Multimedia; Association for Computing Machinery: New York, NY, USA, 2004; MULTIMEDIA '04, p. 56–63. https://doi.org/10.1145/1027527.1027538.
39. Kawulok, M. Fast propagation-based skin regions segmentation in color images. In Proceedings of the 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2013, pp. 1–7. https://doi.org/10.1109/FG.2013.6553733.
40. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation, 2015. https://doi.org/10.48550/ARXIV.1505.04597.