ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

# Deep Semantic Segmentation in Skin Detection

Daniela Cuza[1], Andrea Loreggia[2], Alessandra Lumini[3], Loris Nanni[1]

1- University of Padova - Dept of Information Engineering
Padova - Italy

2- University of Brescia - Dept of Information Engineering
Brescia - Italy

3- University of Bologna - Dept of Computer Science and Engineering
Bologna - Italy

**Abstract**. Deep semantic segmentation is a task that identifies objects and their boundaries in images, to do that a classification task is performed at the pixel level to tag whether a pixel belongs to an object. In skin detection, areas of images are classified as skin or non-skin regions. In this work, we report a short survey of the recent literature covering the task to help researchers in selecting the most suitable method for their application and to expand the knowledge about the available datasets for this topic. A compact empirical evaluation comparing recent models and a new ensemble model is reported.

## 1 Introduction

In this work, we focus on skin detection, a binary classification task that aims at identifying, in images or videos, areas representing skin or non-skin regions. Possible applications of this task are employed for hand gesture recognition, or in clinical practice to identify skin cancer such as early stages of melanoma or other diseases. Skin cancer is one of the most frequent and dangerous types of cancer worldwide, making its identification one of the primary focuses of researchers. Recently, it was also adopted to improve the performance of face recognition systems and sign language recognition frameworks. The interest in skin detection exploded during the last few years. In particular, with the advent of deep learning architecture and attention-based mechanisms, the trend in applying such techniques to skin detection grows exponentially.

Ensembles are formed by sets of classifiers whose predictions are aggregated to set up the prediction of the system. For a specific task, these groups of individual classifiers are usually trained on the same set of data to make each model generalize differently in the training space. To reach good performance, it is important to enforce some kind of diversity among the different models in the ensemble. In this work, we ensure diversity by employing different loss functions and by combining different deep learning models developed to address the semantic segmentation problem. In particular, we adopt two recent models, namely PVT Transformer and HardNet, and apply various loss and data augmentation methods.

We provide a short survey about the recent literature about the topic, consisting in novel models developed to resolve the task as well as available datasets

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

useful to train deep semantic segmentation systems. We shall conclude reporting an empirical evaluation that compare recent systems with a novel ensemble method[1].

## 2 Related Work

In the following paragraphs, we shall introduce recent advancements in this area. Due to the lack of space, consider it as a non-exhaustive list of recent works.

Recently, a deep learning architecture has been proposed for the treatment of low resolution grayscale images, in particular those obtained using the SPAD array camera [1]. Furthermore, the network is created for the distinct issue of facial skin segmentation. The colorization network proposed in [2] is slightly adjusted for the specific application and then the fine tuning method is used. In order to deal with facial skin segmentation problem, a specific dataset is introduced [1] that is created by adapting and merging MUCT and Helen, two already existing datasets, thus obtaining 6000 facial grayscale images, with the associated skin labeling mask. For a comprehensive study about skin cancer detection, we point the reader to a recent survey on skin cancer detection using Deep Learning Techniques [3].

Convolutional Neural Networks (CNNs) still play an important role in the scope of skin detection, as reported in two recent papers: OR-Skip-Net [4], which consists of a fully convolutional network with outer residual paths from the encoder to the decoder, and a skin detection CNN model [5] that includes three convolution layers, a down-sampling layer, a flatten layer and some fully connected layers. The Skinny network [6] is based on U-Net architecture but it presents main advantages over some architectural components such as inception modules and dense blocks to better profit from both local and global pixel descriptions. An innovative approach to dealing with skin detection problems is based on the zero-sum game theory model [7] modeling the classification problem with two players (i.e., skin and non-skin pixels) competing one each other. The approach divides an image into small regions or patches to be analyzed and classified by a set of classifiers. If all classifiers agree on the prediction for a certain patch, then it is accordingly classified, otherwise, a conflicting patch results from the analysis. The set includes classifiers based on color space thresholding and on artificial neural networks. This method allows for decreasing the non-skin regions detected as skin.

Among the many techniques proposed for skin detection, the first to be introduced is the rule-based, such as thresholding-based methods. The idea behind the color space thresholding is to identify pixels with a certain color, in our case those that have a skin color. In [8] skin thresholds for two of the most popular color models, HSV and YCbCr, are introduced. HSV has three components, hue, saturation, and value representing the cylindrical coordinate of an RGB color model. Instead, YCbCr is composed of the luma, chrominance

[1] All resources are available online at `https://github.com/LorisNanni`

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

blue, and chrominance red components, that can be easily calculated from the RGB values.

A novel abdominal skin dataset created from Google images is presented in [9]. It consists of 1400 manually segmented images representing the abdomen of various ethnic groups. In particular, 700 images characterize people with dark skin and 700 images characterize people with light skin. Furthermore, some images represent people with higher body mass indices and some represent people with tattoos.

HarD-Net (Harmonic Densely Connected Net-work) [10] [2] is a model influenced by Densely Connected Networks. The pros of this model is that it keeps low the memory consumption by decreasing most of the connection layers at the DenseNet level, in order to reduce the costs of concatenation. In addition, the input / output channel ratio is equalized thanks to the increase of the channel width of the layers (consequently to the increase of its connections).

PVT (Pyramid Vision Transformer) [11] [3] is a pure convolution-free transformer network. The PVT aims to acquire a high-resolution representation starting from a fine-grained input. The computational cost of the model are decreased by a progressive pyramidal shrinkage, accompanied by the depth of the model. A spatial-reduction attention (SRA) layer is introduced to an additional reduction of the computational complexity of the system.

## 3 Empirical Evaluation

We performed an empirical evaluation comparing existing models with different versions of the ensemble. All the models are trained on short training sets from the ECU dataset [12]. In particular, we used the first 2000 images for the training phase, leaving them out from the test set. The testing phase is performed on 10 different datasets that are reported in Table 1.

### 3.1 Data Augmentation

In our tests, we adopted two different data augmentation approaches: DA1, base data augmentation consisting in horizontal and vertical flip, 90 degrees rotation. DA2 is a more articulated procedure. We refer to [22] for an exhaustive description of the two approaches. Some artificial images (DA2 approach) contain only background pixels, to discard them we simply delete all the images where there are less than 10 pixels that belong to the foreground class.

### 3.2 Experimental Results

In order to outline the importance of deep learning techniques, we compare the results of the ensemble with the performance reported in a recent survey [23]. Table 2 shows the results of different methods developed for addressing the task of skin detection. From SegNet down to Ensemble, the boost in the

[2] https://github.com/james128333/HarDNet-MSEG - Last access on June 30th, 2022
[3] https://github.com/DengPingFan/Polyp-PVT - Last access on June 30th, 2022

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

Table 1: Datasets for Skin Segmentation. ECU dataset is split in 2000 images for training and 2000 for test set. For ECU, we considered the subset of images that were not used in the training phase.

| **Tag** | **Name** | **#Samples** | **Ref.** | **Available** |
|---|---|---|---|---|
| CMQ | Compaq | 4675 | [13] | Ask Authors |
| HGR | Hand Gesture Recognition | 1558 | [14] | Yes |
| MCG | MCG-skin | 1000 | [15] | Ask Authors |
| PRT | Pratheepan | 78 | [16] | Yes |
| SFA | SFA | 1118 | [17] | Ask Authors |
| SCH | Schmugge dataset | 845 | [18] | Yes |
| VMD | Human activity recognition | 285 | [19] | Yes |
| ECU | ECU Face and Skin Detection | 2000 | [12] | N/A |
| UC | UChile DB-skin | 103 | [20] | Ask Authors |
| VT | VT-AAST | 66 | [21] | Ask Authors |

Table 2: Performance (Dice=F1-score) in the skin detection problem. Best performance in bold.

| Method | PRT | MCG | UC | CMQ | SFA | HGR | SCH | VMD |
|---|---|---|---|---|---|---|---|---|
| GMM | 0.581 | 0.688 | 0.615 | 0.600 | 0.789 | 0.658 | 0.595 | 0.130 |
| Bayes | 0.631 | 0.694 | 0.661 | 0.599 | 0.760 | 0.871 | 0.569 | 0.252 |
| SPL | 0.551 | 0.621 | 0.568 | 0.494 | 0.700 | 0.845 | 0.490 | 0.321 |
| Cheddad | 0.597 | 0.667 | 0.649 | 0.588 | 0.683 | 0.767 | 0.571 | 0.261 |
| Chen | 0.540 | 0.656 | 0.598 | 0.549 | 0.791 | 0.732 | 0.571 | 0.165 |
| SA1 | 0.613 | 0.664 | 0.567 | 0.593 | 0.788 | 0.768 | 0.482 | 0.199 |
| SA2 | 0.693 | 0.755 | 0.663 | 0.645 | 0.771 | 0.806 | 0.594 | 0.156 |
| SA3 | 0.709 | 0.762 | 0.625 | 0.647 | 0.863 | 0.877 | 0.586 | 0.147 |
| DYC | 0.599 | 0.680 | 0.663 | 0.618 | 0.569 | 0.616 | 0.613 | 0.275 |
| SegNet | 0.730 | 0.813 | 0.802 | 0.737 | 0.889 | 0.869 | 0.708 | 0.328 |
| U-Net | 0.787 | 0.779 | 0.713 | 0.686 | 0.848 | 0.836 | 0.671 | 0.332 |
| DeepLab | 0.875 | 0.879 | 0.899 | 0.817 | 0.939 | 0.954 | 0.774 | 0.628 |
| Vote1 | 0.717 | 0.754 | 0.670 | 0.666 | 0.737 | 0.849 | 0.625 | 0.269 |
| Vote2 | 0.811 | 0.816 | 0.81 | 0.772 | 0.854 | 0.949 | 0.700 | 0.481 |
| Vote3 | 0.812 | 0.841 | 0.829 | 0.773 | 0.902 | 0.950 | 0.714 | 0.423 |
| Vote4 | 0.879 | 0.878 | 0.897 | 0.819 | 0.944 | 0.967 | 0.776 | 0.620 |
| Hardnet | 0.913 | 0.880 | 0.900 | 0.809 | 0.951 | 0.967 | 0.792 | 0.717 |
| PVT | 0.920 | 0.888 | 0.925 | 0.851 | 0.951 | 0.966 | 0.792 | 0.709 |
| Ensemble | **0.927** | **0.894** | **0.932** | **0.868** | **0.954** | **0.971** | **0.797** | **0.767** |

performance is mostly related to the introduction of deep learning and attention-based techniques. In Table 2, Methods Vote$x$ refer to fusion of handcrafted methods and deep learning approaches proposed in [23], both Hardnet and PVT

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

have been trained using Adam optimizer and DA1 data augmentation.

Here, we tested several combinations for the ensemble (based on fusions of HardNet and PVTs varying the data augmentation and the loss functions), and all of them are reporting higher performance with respect to previous models. Here we report only the ensemble method that performs the best. It refers to the aggregation of individual prediction using the sum rule between two HardNet trained using SGD, two HardNet trained using Adam optimizer, and two PVTs. For each pair of segmentators, one model has been trained using DA1 and the other using DA2 data augmentation. The performance of the ensemble is better than the performance of the single individual classifiers used as baselines. These results support the evidence that different individual classifiers can generalize differently in the training space leading to an improvement in the final performance of the ensemble [24]. This can be achieved by either varying the type of data augmentation or the kind of individual classifiers.

## 4 Conclusion

Deep semantic segmentation plays an important role in many disciplines. In this work, we have focused on skin detection. In particular, we have revised the recent literature in this area. We compared several approaches and we suggest a protocol for testing that is easily reproducible. That is use a single dataset for training and then testing on many other different datasets. Compared to a previous survey, the new suggested method performs much better suggesting ensemble as a promising technique to deal with this kind of task.

## References

[1] Marco Paracchini, Marco Marcon, Federica Villa, and Stefano Tubaro. Deep skin detection on low resolution grayscale images. *Pattern Recognition Letters*, 131:322–328, 2020.

[2] Federico Baldassarre, Diego González Morín, and Lucas Rodés-Guirao. Deep koalarization: Image colorization using cnns and inception-resnet-v2. *arXiv preprint arXiv:1712.03400*, 2017.

[3] Mehwish Dildar, Shumaila Akram, Muhammad Irfan, Hikmat Ullah Khan, Muhammad Ramzan, Abdur Rehman Mahmood, Soliman Ayed Alsaiari, Abdul Hakeem M Saeed, Mohammed Olaythah Alraddadi, and Mater Hussen Mahnashi. Skin cancer detection: A review using deep learning techniques. *International Journal of Environmental Research and Public Health*, 18(10), 2021.

[4] Muhammad Arsalan, Dong Seop Kim, Muhammad Owais, and Kang Ryoung Park. Or-skip-net: Outer residual skip network for skin segmentation in non-ideal situations. *Expert Systems with Applications*, 141:112922, 2020.

[5] Khawla Ben Salah, Mohamed Othmani, and Monji Kherallah. A novel approach for human skin detection using convolutional neural network. *The Visual Computer*, 38(5):1833–1843, 2022.

[6] Tomasz Tarasiewicz, Jakub Nalepa, and Michal Kawulok. Skinny: A lightweight u-net for skin detection and segmentation. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 2386–2390. IEEE, 2020.

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

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

[8] Sajaa G Mohammed, Abdulrahman H Majeed, Enas Kh Ali Aldujaili, and Safa S Hassan. Image segmentation for skin detection. *Journal of Southwest Jiaotong University*, 55(1), 2020.

[9] Anirudh Topiwala, Lidia Al-Zogbi, Thorsten Fleiter, and Axel Krieger. Adaptation and evaluation of deep learning techniques for skin segmentation on novel abdominal dataset. In *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 752–759. IEEE, 2019.

[10] Ping Chao, Chao-Yang Kao, Yu-Shan Ruan, Chien-Hsiang Huang, and Youn-Long Lin. Hardnet: A low memory traffic network. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3552–3561, 2019.

[11] Bo Dong, Wenhai Wang, Deng-Ping Fan, Jinpeng Li, Huazhu Fu, and Ling Shao. Polyp-PVT: Polyp segmentation with pyramid vision transformers. arXiv, 2021.

[12] Son Lam Phung, Abdesselam Bouzerdoum, and Douglas Chai. Skin segmentation using color pixel classification: analysis and comparison. *IEEE transactions on pattern analysis and machine intelligence*, 27(1):148–154, 2005.

[13] Michael J. Jones and James M. Rehg. Statistical color models with application to skin detection. *Int. J. Comput. Vis.*, 46(1):81–96, 2002.

[14] Michal Kawulok, Jolanta Kawulok, Jakub Nalepa, and Bogdan Smolka. Self-adaptive algorithm for segmenting skin regions. *EURASIP Journal on Advances in Signal Processing*, 2014(1):1–22, 2014.

[15] Lei Huang, Tian Xia, Yongdong Zhang, and Shouxun Lin. Human skin detection in images by MSER analysis. In Benoît Macq and Peter Schelkens, editors, *18th IEEE International Conference on Image Processing, ICIP 2011, Brussels, Belgium, September 11-14, 2011*, pages 1257–1260. IEEE, 2011.

[16] Wei Ren Tan, Chee Seng Chan, Yogarajah Pratheepan, and Joan Condell. A fusion approach for efficient human skin detection. *IEEE Trans. Ind. Informatics*, 8(1):138–147, 2012.

[17] Joao Paulo Brognoni Casati, Diego Rafael Moraes, and Evandro Luis Linhari Rodrigues. Sfa: A human skin image database based on feret and ar facial images. In *IX workshop de Visao Computational, Rio de Janeiro*, 2013.

[18] Stephen J Schmugge, Sriram Jayaram, Min C Shin, and Leonid V Tsap. Objective evaluation of approaches of skin detection using roc analysis. *Computer vision and image understanding*, 108(1-2):41–51, 2007.

[19] Juan C Sanmiguel and Sergio Suja. Skin detection by dual maximization of detectors agreement for video monitoring. *Pattern Recognition Letters*, 34(16):2102–2109, 2013.

[20] Javier Ruiz-del-Solar and Rodrigo Verschae. Skin detection using neighborhood information. In *Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FGR 2004), May 17-19, 2004, Seoul, Korea*, pages 463–468. IEEE Computer Society, 2004.

[21] Abdallah S Abdallah, Mohamad Abou El-Nasr, and A Lynn Abbott. A new color image database for benchmarking of automatic face detection and human skin segmentation techniques. In *Proceedings of World Academy of Science, Engineering and Technology*, volume 20, pages 353–357. Citeseer, 2007.

[22] Loris Nanni, Alessandra Lumini, Andrea Loreggia, Alberto Formaggio, and Daniela Cuza. An empirical study on ensemble of segmentation approaches. *Signals*, 3(2):341–358, 2022.

[23] Alessandra Lumini and Loris Nanni. Fair comparison of skin detection approaches on publicly available datasets. *Expert Systems with Applications*, 160:113677, 2020.

[24] Cristina Cornelio, Michele Donini, Andrea Loreggia, Maria Silvia Pini, and Francesca Rossi. Voting with random classifiers (VORACE): theoretical and experimental analysis. *Autonomous Agents and Multi-Agent Systems*, 35(2):22, 2021.