Deep Learning for hand segmentation in complex backgrounds

Adrian-Stefan Ungureanu, Shabab Bazrafkan and Peter Corcoran, *Member, IEEE*
Center for Cognitive, Connected & Computational Imaging, College of Engineering & Informatics,
NUI Galway, Galway, Ireland
E-mails:{ a.ungureanu1, s.bazrafkan1, peter.corcoran}@nuigalway.ie

Abstract--**This paper presents a Deep Learning segmentation approach for hand segmentation in gray level images with cluttered backgrounds where standard techniques cannot be used. Two networks were trained with a database of hand images derived from widely used palmprint image databases, Hong Kong Polytechnic University (HKPU) and Chinese Academy of Science (CASIA). The image dataset is augmented with complex**

providing promising results.

I. Introduction

There are many applications which require the correct extraction of hands from the background. Biometrics, which aims at improving security using the user's behavioral or physiological traits, is one of the most relevant. Within the category of physiological biometrics of the hand, among hand contour or hand geometry [1], palmprint biometrics has seen recent advancements and its transition to the consumer electronics devices [2]. The main advantage presented by palmprints on smartphones is the use of the rear camera and sufficient processing power to deliver quick and efficient authentication. The fact that it does not require extra hardware provides a further advantage, making it suitable for deployment on older devices.

This paper presents a novel approach of hand segmentation which in turn allows the successful extraction of palmprints in cluttered backgrounds, which define the consumer electronics' environment. Two different deep learning networks are presented. These networks accept hand images as input and generate the segmentation map at output. The results for both networks are presented.

II. Literature Review

The main approaches used for extracting hands to be used for palmprint recognition fall into 2 categories - thresholding against a black background [3], [4], typical for grayscale images, and using a predefined or adaptive model for skin segmentation [5], [6]. While the former category imposes significant constraints on the user, the latter does not take into consideration the fact that pixels' values irrespective of the color space used, cannot guarantee the accurate detection of skin in complex backgrounds [7].

III. Image Dataset and Augmentation of Data

Throughout the training and testing of the segmentation with DL, 2 standard databases of palmprint grayscale images from the Chinese Academy of Science (CASIA) [4] and Hong Kong Polytechnic University (HKPU) [8] were employed. The total number of palmprint images amounts to 12,980. Augmentation of these images included the replacing of the simple background with complex images with various texture

The complete augmentation flow is organized as follows:

1. Resize palmprint images to 135x180 pixels. This resizing is required from the initial size of CASIA images (640x480 pixels) and HKPU images (384x284 pixels) because DL is using low resolution images.
2. Threshold the palmprint images to remove the black background (>30 for CASIA and >15 for HKPU).
3. Determine the biggest blob and extract it. This will count as the ground truth (GT) of the hand.
4. Introduce backgrounds behind the biggest blob, which is the hand:
   a. Using 100 possible backgrounds, choose 4 random images.
   b. Combine the hand region with each background.
   c. Determine a boundary of the hand and dilate it with a spherical structuring element of radius 1.
   d. Define a Gaussian filter of size 5x5 and $\sigma = 1$.
   e. Apply the Gaussian filter on the region defined in 4.c) in the images obtained at 4.b).

After augmentation the number of available images (with ground truths) increased to 77,880. A number of backgrounds used for augmentation are presented in Fig. 1, whereas Fig. 2 contains the final augmented images of a palmprint from CASIA database.

IV. Deep Neural Network for Hand Segmentation

A. Network Design

Two different neural networks are designed for the hand segmentation, as described below:

1) U shaped network

The network has 13 layers and the starting kernel size for the input is 3x3pixels, its size being further increased until the center of the network, where the kernel size is 15x15 pixels. From the center towards the end of the network the kernels are gradually getting smaller back to 3x3 pixels for the output layer. The number of channels is set in such a way that the network has the same number of parameters as a SegNet basic

The research work presented here was funded under the Strategic Partnership Program of Science Foundation Ireland (SFI) and co-funded by SFI and FotoNation Ltd. Project ID: 13/SPP/I2868 on Next Generation Imaging for Smartphone and Embedded Platforms.

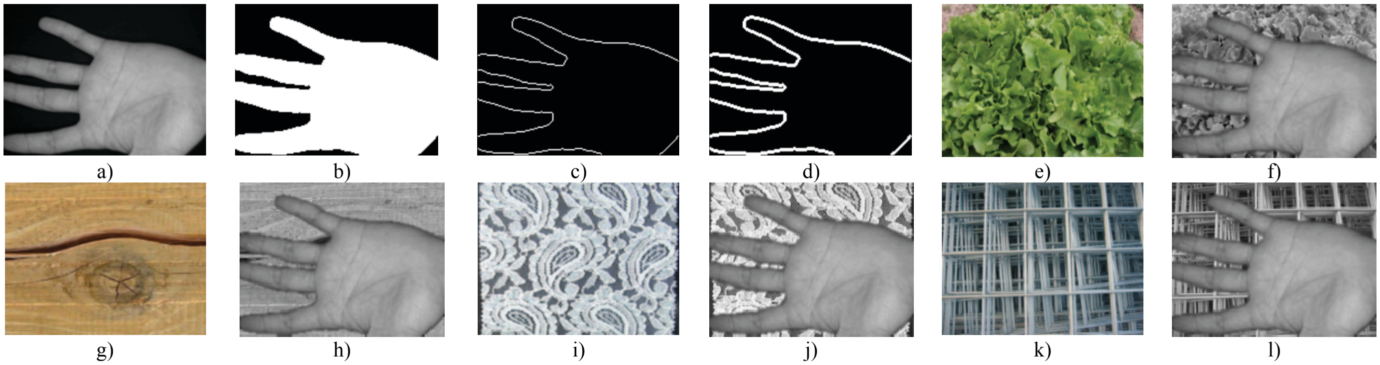

Fig. 1. Process flow for the database augmentation step: a) original image from CASIA, b) ground truth mask; c) boundary of the ground truth; d) dilation of the boundary; e), g), i) and k) backgrounds used for the database augmentation; f), h), j) and l) combine the information from (a) with the randomly selected backgrounds and smoothen the surface defined in (d) with a Gaussian filter.

network.

2) SegNet basic

This network is described in [10]. It is an auto encoder shaped model. Four layers in encode and four layers in decoder part. In the encoder, each layer is followed by a max-pooling layer and the decoder contains an un-pooling layer after convolutions, wherein the location of the indices for un-pooling is taken from the encoder network.

In any of these two networks the input of the network is the hand image with the random background and the output is forced to generate the segmentation map.

B. Training

The training procedure was same for both networks. 70% of the input data was used for training, 20% for validation and 10% for testing. The "Mean Binary Cross-Entropy" function was used to define the loss for each batch. The Nestrov Momentum method was used in order to minimize the loss function. Learning rate was set to 0.01 and momentum to 0.9. Batch normalization technique was used after each convolution layer and ReLU non linearity was used in every layer except the last layer which was taking advantage of sigmoid function.

V. Results and Discussion

Both networks have been trained and tested on the same database described in section III. The results on the test set for these networks are shown table I. The U shaped network gives slightly better results than the SegNet basic.

The simulations show promising results in using the deep learning technique for hand segmentation task in non-constraint conditions.

More discussions on the network architecture, training and different measures will be presented in the extended version of the paper.

TABLE I

Segmentation Results for Two Different Networks

Measure\Network	U shaped network	SegNet basic
Accuracy	**99.7%**	99.51%
Sensitivity	**99.72%**	99.55%
Specificity	**99.65%**	99.42%
Precision	**99.72%**	99.54%
Negative Prediction Value	**99.66%**	99.43%
F1 Score	**99.72%**	99.55%
Matthew Correlation Coefficient	**99.38%**	98.98%
Informedness	**0.9937**	0.9898
Markedness	**0.9938**	0.9898
False Positive Rate	**0.34%**	0.57%
False Negative Rate	**0.27%**	0.44%
False discovery Rate	**0.27%**	0.45%

References

[1] A. de Santos Sierra, C. Sánchez-Ávila, A. Mendaza Ormaza, and J. Guerra Casanova, "An approach to hand biometrics in mobile devices," *Signal, Image Video Process.*, vol. 5, no. 4, pp. 469–475, Nov. 2011.

[2] H. Javidnia, A. Ungureanu, C. Costache, and P. Corcoran, "Palmprint as a smartphone biometric," in *2016 IEEE International Conference on Consumer Electronics (ICCE)*, 2016, pp. 463–466.

[3] W. Jia, R. X. Hu, J. Gui, Y. Zhao, and X. M. Ren, "Palmprint recognition across different devices," *Sensors* (Switzerland), vol. 12, no. 6, pp. 7938–7964, 2012.

[4] A. I. T. C. A. of Sciences, "CASIA palmprint database, http://biometrics.idealtest.org/." .

[5] M. Choraś and R. Kozik, "Contactless palmprint and knuckle biometrics for mobile devices," *Pattern Anal. Appl.*, vol. 15, no. 1, pp. 73–85, 2012.

[6] A. Ungureanu, H. Javidnia, C. Costache, and P. Corcoran, "A review and comparative study of skin segmentation techniques for handheld imaging devices," in *2016 IEEE International Conference on Consumer Electronics (ICCE)*, 2016, pp. 530–531.

[7] A. Albiol, L. Torres, and E. J. Delp, "Optimum color spaces for skin detection," *Proc. 2001 Int. Conf. Image Process*. (Cat. No.01CH37205), vol. 1, no. 3, pp. 122–124, 2001.

[8] The Hong Kong Polytechnic University, "PolyU palmprint database, http://www.comp.polyu.edu.hk/~biometrics/."

[9] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, and A. Vedaldi, "Describing textures in the wild," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3606–3613.

[10] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: {A} Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," CoRR, vol. abs/1511.0, 2015.