© Universiti Tun Hussein Onn Malaysia Publisher's Office

IJIE

Journal homepage: http://penerbit.uthm.edu.my/ojs/index.php/ijie
ISSN : 2229-838X e-ISSN : 2600-7916

The International Journal of Integrated Engineering

Automatic Face and Hijab Segmentation Using Convolutional Network

Dina M. Madkour[1,*], Ahmed Madani[1], Mohamed Waleed Fakhr[1]

[1] Computer Engineering Depart,
Arab Academy for Science, Technology and Maritime Transport, Cairo, EGYPT

*Corresponding Author

DOI: https://doi.org/10.30880/ijie.2019.11.07.008
Received 30 May 2019; Accepted 30 July 2019; Available online 10 August 2019

Abstract: Taking pictures and Selfies are now very common and frequent between people. People are also interested in enhancing pictures using different image processing techniques and sharing them on social media. Accurate image segmentation plays an important role in portrait editing, face beautification, human identification, hairstyle identification, airport Surveillance system and many other computer vision problems. One specific functionality of interest is automatic face and veil segmentation as this allows processing each separately. Manual segmentation can be difficult and annoying especially on smartphones small screen. In this paper, the proposed model uses fully convolutional network (FCN) to make semantic segmentation into skin, veil and background. The proposed model achieved an outperforming result on the dataset which consists of 250 images with global accuracy 92% and mean accuracy 92.69.

Keywords: Convolutional neural network (CNN), Convolution, image segmentation, skin Segmentation, veil (Hijab) segmentation.

1. Introduction

Image segmentation is the process of partitioning an image into multiple segments (Regions), by assigning label to every pixel in an image where the pixels with the same label have common characteristics. Automatic hair and face segmentation is still an open problem as it plays an important role in human identification, image editing, enhancement and many other computer problems. In Islamic nations, a big percentage of females wear veil (Hijab), which makes it difficult to make accurate segmentation for the image. Many segmentation algorithms were in general object recognition but had a small role in the problem of face recognition. To make segmentation for the four classes it is required to make a per pixel segmentation labels. Pervious methods were used for face and hair segmentation only. To our knowledge, this is the first work that use CNN for veil segmentation.

A brief review on previous work in image segmentation will be discussed. In [1] a network was trained to segment multiple facial regions, including the entire face but it is slow. The method of [2] outperformed [3] on COFW as well as reported real-time processing speeds by using a deconvolution neural network. Some Other approaches have used CNN for semantic segmentation in which each pixel is labeled with the class of its enclosing object or region. Feng et al. [8] trained a CNN to make classification for dataset of 50 images from 5 different C. elegans embroys movies, while Ciresan et al. [9] used deep artificial neural network (DNN) as a pixel classifier on the publicly-available dataset provided by the organizers of the ISBI 2012 EM Segmentation Challenge. Also Farabet et al. [10] uses a multiscale convolutional neural network to extract dense feature vectors that encodes regions of multiple sizes centered on each pixel on 3 different datasets (Stanford Background, SIFR Flow, Barcelona), in [11] Pinheiro et al. used Recurrent convolutional neural network for scene labeling, the system is trained in an end to end manner on Stanfrod Background and SIFT flow datasets. Hariharan et al. [12] uses CNN to extract features on image regions. While Gupta et al. [13]

*Corresponding author: dina.mostafa.madkour@hotmail.com

Propose a decision forest approach that classifies pixel in the detection window. While Huang et al. [14] trained CRF on images from the LFW dataset to build a skin, hair and background classifier. Recently, Kae et al. [15] proposed a model using CRF to capture local appearance features and restricted Botlzmann machines to model global shapes.

In this paper, will discuss convolution neural network (CNN) Fig. 1 and include the methodology to make face, veil, and background segmentation in section 2, then in section 3 will talk about the dataset used, and discuss the results in section 4, and finally section 5 is conclusion and future work.

2. Methodology

The input is an image of size 250x250 entered to a convolutional neural network (CNN), consists of 91 layers, (convolution layers, ReLU layers, Max Pooling layers, Max Unpooling layers, soft max layer, and pixel classification layer), to make segmentation for the image into 3 classes (face, veil (Hijab) and background). The output of each layer of the network is entered to the next layer until reach the output layer that gives output image of the same size the input image. The output is the segmented representation of the input image.

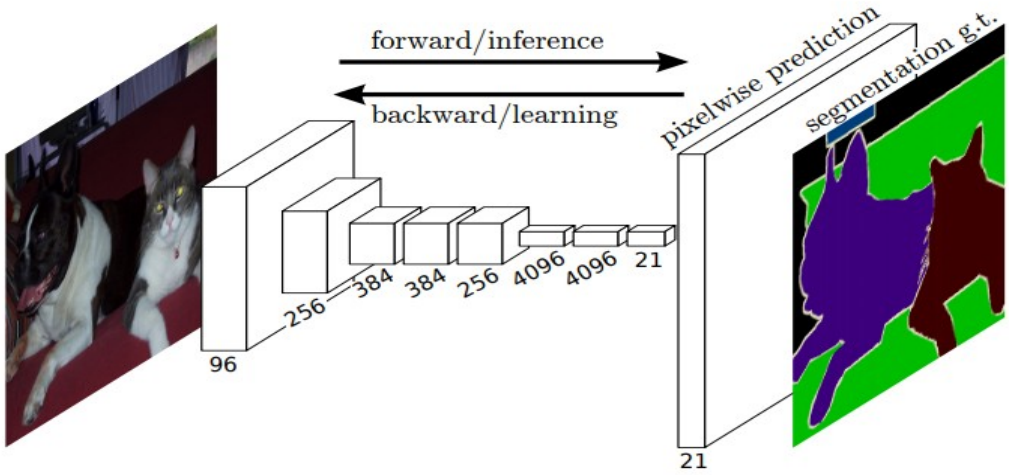

Fig. 1 - Fully convolutional networks can efficiently learn to make dense predictions for per-pixel tasks like semantic segmentation [4]

2.1 Convolutional neural network (CNN)

Computers see in a different way than people do, their world consists only of numbers. Each image is represented by 2D array of numbers known as pixels. Convolutional neural network needs to see many images before it is able to make prediction for new images.

Convolutional neural networks have different architecture than regular neural networks as shown in Fig. 2. Regular neural networks, Fig. 2 Left, consists of input layer (Image), followed by a series of hidden layers. Every layer consists of a set of neurons where each layer is fully connected to all neurons in the layer before, and finally the output layer that represent the prediction [18].

There are many deep learning approaches that are developed for different purposes, such as object detection, classification and segmentation. CNNs are the most commonly applied to image segmentation and classification [6]. Convolutional neural networks, Figure (2) Right, are a bit different where the layers are organized in three dimensions: width, height, and depth [19]. Every layer of a CNN transforms the 3D input volume to a 3D output volume of neurons activations. In this example, the red input layer holds the image, so its width and height would be the dimensions of the image, and the depth would be three (Red, Green and Blue channels). The neurons in one layer are not connected to all the neurons in the next layer but to a small region only. Finally, the output layer will be reduced to a single vector of probability score organized along the depth dimension. Convolutional neural networks perform a series of convolutions and pooling operations during feature detection and extraction [17]. CNNs generate more discriminative representations compared to traditional methods based on handcrafted features [5],[20].

2.2 Convolution

Convolution is one of the main blocks of CNN; it refers to the mathematical combination of two functions to produce a third function. In CNN, convolution is performed on the input image data with the use of a filter (Kernel) to produce a feature map, the output of the convolution will be passed through an activation function (ReLU activation function) to make non-linear output.

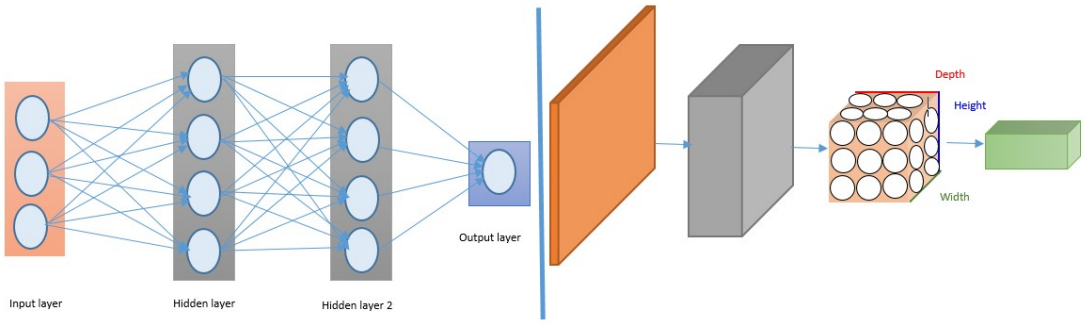

Fig. 2 - Left: A regular 3-Layer Neural Network. Right: A CNN arranges its Neurons in a 3- dimension (Width, Height, and depth).

2.3 Pooling

After convolution layer a pooling layer is added between CNN layers, its function is to reduce dimensionality to reduce the number of parameters and computations in the network. It decreases the feature map size while at the same time keeps the significant information. There are several non-linear functions to implement pooling but max pooling, Fig. 3, is the most common used. It partitions the input image into a set of non-overlapping rectangles and, for each such sub-region, outputs the maximum. Max un-pooling Fig. 3 use the switch variable created by pooling layer to place each activation value to its original location.

To use CNN there are four important hyper parameters to consider

- The kernel size
- Filter count
- Stride
- Padding

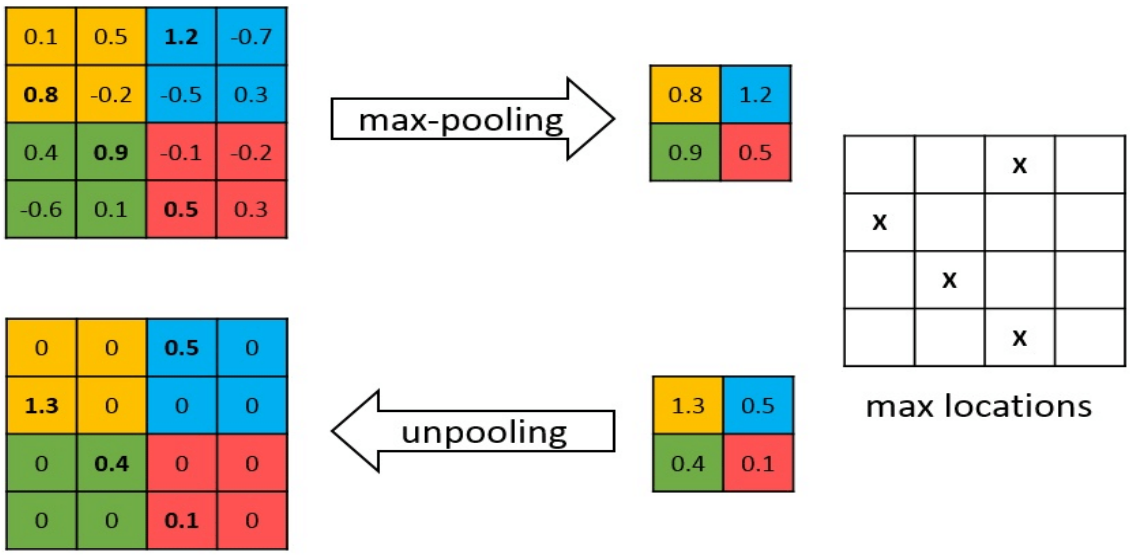

Fig. 3 - Pooling and un-pooling layers.

3. Data Set

The network was evaluated on a dataset of 250 face images for girls wearing veil. Each image is of size 250x250, with large variance in background, skin color, head pose, veil shape and colour. The dataset is divided into two parts, first part is the training set which consists of 150 images, samples of training dataset are shown in Fig. 4 and the second part is the testing set which consists of 100 image. All images were manually labeled as skin, background and veil to generate the ground truth.

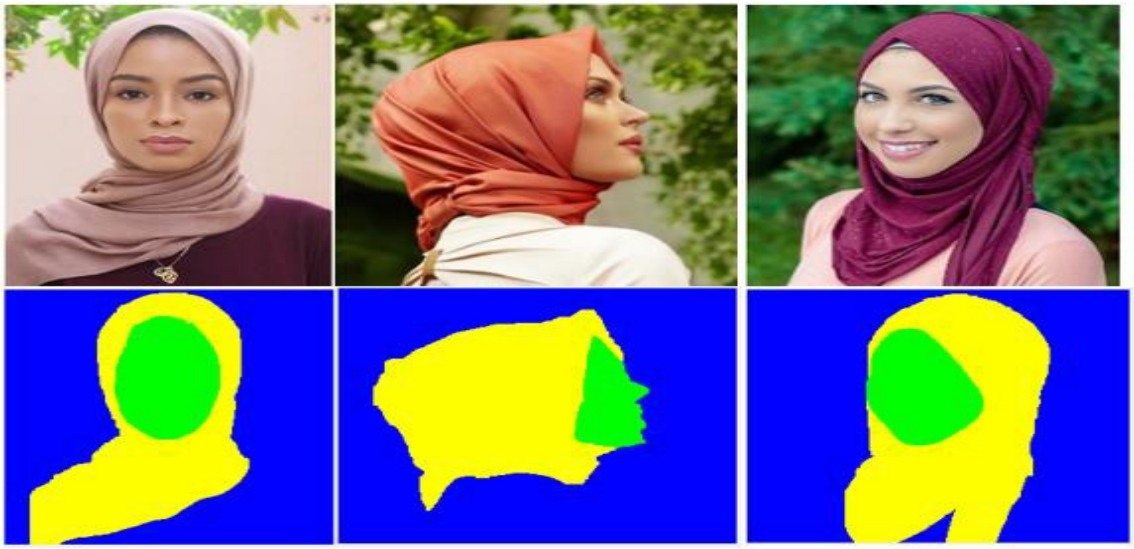

Fig. 4 - Samples of Training Set images and their corresponding Ground truth

4. Results

The proposed model is evaluated on a task of segmenting images on the created dataset to make segmentation of each image as skin, veil (Hijab) and background. The proposed model achieved global accuracy 92% and mean accuracy 92.69%. In [4] the FCN-8 achieved mean accuracy 51.7% on SIFT Flow dataset, while in [10] the pixel accuracy on the Stanford Background 78.8%, and 78.5% on Sift Flow, and finally 67.8% on Barcelona dataset. The training and testing were implemented on 2.8GHz 7 core processor, Nvidia GeForce GTX 1060. The training phase was done on 150 image it took 74 minutes to complete training, and average time 3.4 seconds on testing an image for segmentation. We reported two Class Metrics, Accuracy and Intersection over union (IoU), and three Data Set Metrics, Global Accuracy, Mean Accuracy, Mean IoU.

Class Metrics:

1. Accuracy: The ratio of correctly classified pixels to the total number of pixels in that class, according to the ground truth.

$$cc = \frac{P}{P + FN} \quad (1)$$

$$cy =$$

2. IoU: The ratio of correctly classified pixels to the total number of ground truth and predicted pixels in that class.

$$= \frac{P}{P + FP + FN} \quad (2)$$

Where TP is the number of true positives, FP is the number of False Positive and FN is the number of false negatives.

Dataset Metrics:

1. Global Accuracy: Ratio of correctly classified pixels to total pixels, regardless of class.
2. Mean Accuracy: Ratio of correctly classified pixels in each class to total pixels, averaged over all classes
3. Mean IoU: Average intersection over union (IoU) of all classes.

Table 1 shows the class metrics results, while Table 2 shows the dataset metrics and finally Table 3 shows the confusion matrix where Each table element (i,j) is the count of pixels known to belong to class i but predicted to belong to class j.

Table 1. Class metrics evaluation

	Accurac	Io
Skin	95.433%	88.162
Veil	90.612%	79.462
Background	92.049%	85.837%

Table 2. Dataset metrics evaluation

Global Accuracy	Mean	Mean
0.92008	0.92698	0.844

Table 3. Confusion matrix shows number of pixels belonging to class i but predicted to belong to class j

	Skin	Veil	Background
Skin	7.9522e+05	38039	20
Veil	38039	1.9403e+06	1.6713e+05
Background	585	2.5628e+05	2.9739e+06

As shown in in Fig. 5 result on some samples of the test images compared with the ground truth.

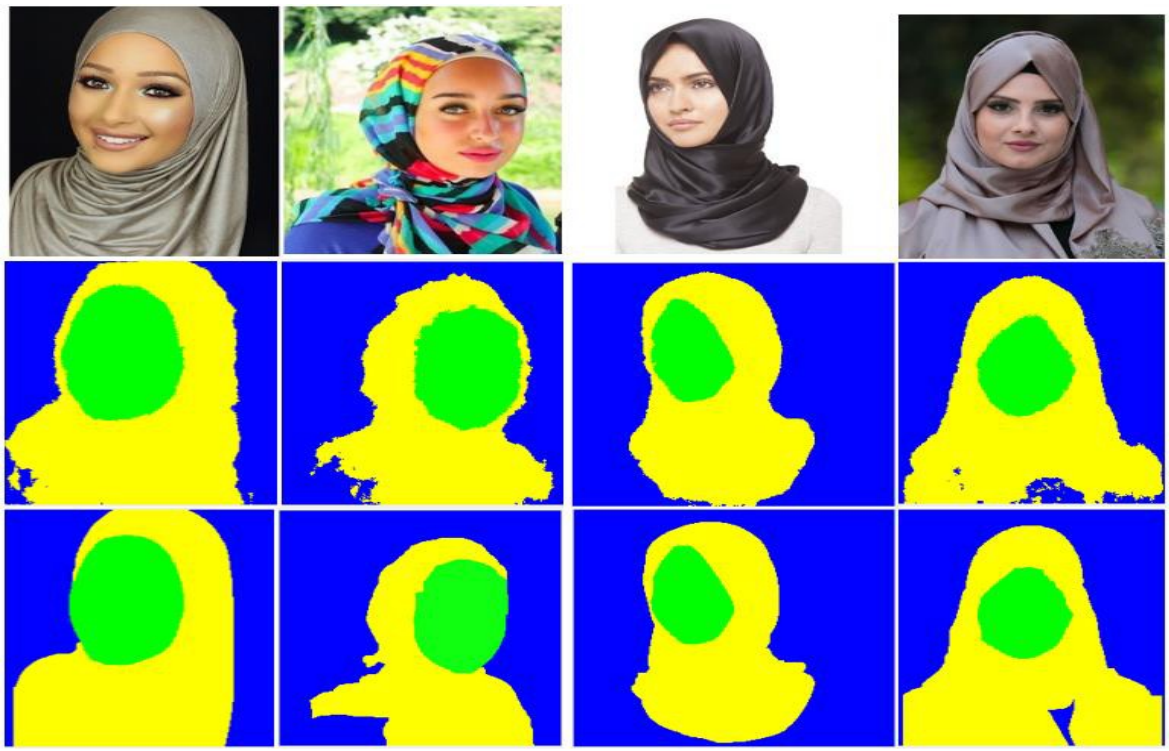

Fig. 5 - Samples of Testing Set images in first row, Results in second row and the ground truth in the third row

5. Conclusion and Future Work

Multiple image segmentation algorithms have been used to analyze images and a wide range of algorithms is being used to carry out the process of image segmentation. The aim of image segmentation is to cluster the pixels of the image into specific regions and classes [19]. Face segmentation is challenging due to the diversity of veil shape or style, head pose, skin color and background. This is the first work to develop a CNN model to automatically segment images for girls wearing veil. The proposed model has reached a good accuracy comparing to other face and hair segmentation previous work. The future research is to enhance accuracy of the network and to expand its border to include hair segmentation. The model proposed can find application in image enhancement and editing.

References

[1] Saito, S., Li, T., & Li, H. (2016). Real-time facial segmentation and performance capture from rgb input. In European Conference on Computer Vision (pp. 244-261). Springer, Cham.

[2] Ghiasi, G., Fowlkes, C. C., & Irvine, C. (2015). Using Segmentation to Predict the Absence of Occluded Parts. In BMVC (pp. 22-1).

[3] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).

[4] Liu, S., Yang, J., Huang, C., & Yang, M. H. (2015). Multi-objective convolutional learning for face labeling. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3451-3459).

[5] Akkus, Z., Galimzianova, A., Hoogi, A., Rubin, D. L., & Erickson, B. J. (2017). Deep learning for brain MRI segmentation: state of the art and future directions. Journal of digital imaging, 30(4), 449-459.

[6] Qin, S., Kim, S., & Manduchi, R. (2017, July). Automatic skin and hair masking using fully convolutional networks. In Multimedia and Expo (ICME), 2017 IEEE International Conference on (pp. 103-108). IEEE.

[7] Ning, F., Delhomme, D., LeCun, Y., Piano, F., Bottou, L., & Barbano, P. E. (2005). Toward automatic phenotyping of developing embryos from videos. IEEE Transactions on Image Processing, 14(9), 1360-1371.

[8] Ciresan, D., Giusti, A., Gambardella, L. M., & Schmidhuber, J. (2012). Deep neural networks segment neuronal membranes in electron microscopy images. In Advances in neural information processing systems (pp. 2843-2851).

[9] Farabet, C., Couprie, C., Najman, L., & LeCun, Y. (2013). Learning hierarchical features for scene labeling. IEEE transactions on pattern analysis and machine intelligence, 35(8), 1915-1929.

[10] Pinheiro, P. H., & Collobert, R. (2014). Recurrent convolutional neural networks for scene labeling. In 31st International Conference on Machine Learning (ICML) (No. EPFL-CONF-199822).

[11] Hariharan, B., Arbeláez, P., Girshick, R., & Malik, J. (2014). Simultaneous detection and segmentation. In European Conference on Computer Vision (pp. 297-312). Springer, Cham.

[12] Gupta, S., Girshick, R., Arbeláez, P., & Malik, J. (2014). Learning rich features from RGB-D images for object detection and segmentation. In European Conference on Computer Vision (pp. 345-360). Springer, Cham.

[13] Huang, G. B., Narayana, M., & Learned-Miller, E. (2008). Towards unconstrained face recognition. In Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on (pp. 1-8). IEEE.

[14] Kae, A., Sohn, K., Lee, H., & Learned-Miller, E. (2013). Augmenting CRFs with Boltzmann machine shape priors for image labeling. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2019-2026).

[15] David, O. E., & Netanyahu, N. S. (2016). DeepPainter: painter classification using deep convolutional autoencoders. In International Conference on Artificial Neural Networks (pp. 20-28). Springer, Cham.

[16] Hu, W., Huang, Y., Wei, L., Zhang, F., & Li, H. (2015). Deep convolutional neural networks for hyperspectral image classification. Journal of Sensors, 2015.

[17] Balaji, S. A., & Baskaran, K. (2013). Design and development of artificial neural networking (ANN) system using sigmoid activation function to predict annual rice production in Tamilnadu. arXiv preprint arXiv:1303.1913.

[18] Gokmen, T., Onen, M., & Haensch, W. (2017). Training deep convolutional neural networks with resistive cross-point devices. Frontiers in neuroscience, 11, 538.

[19] Akhtar, Nishat, Junita Mohamad Saleh, and Clemens Grelck (2018). Parallel Processing of Image Segmentation Data Using Hadoop. International Journal of Integrated Engineering, vol. 11, no. 1, pp. 74-84.

[20] Sofian, Hannah, et al. (2018). Calcification Detection of Coronary Artery Disease in Intravascular Ultrasound Image: Deep Feature Learning Approach. International Journal of Integrated Engineering, vol. 10, no. 7, pp. 43-57.