

# FACIAL COMPONENT SEGMENTATION USING CONVOLUTIONAL NEURAL NETWORK

Gozde YOLCU<sup>1,2</sup>, Ismail OZTEL<sup>1,2</sup>, Serap KAZAN<sup>1</sup>, Cemil OZ<sup>1</sup>, Filiz BUNYAK<sup>2</sup>

<sup>1</sup>Sakarya University, Department of Computer Engineering, Sakarya TURKEY

<sup>2</sup>University of Missouri-Columbia, Department of Electrical Engineering and Computer Science, MO, USA

{gyolcu, ioztel, scakar, coz}@sakarya.edu.tr, bunyak@missouri.edu

**Abstract:** Facial components are important for many face image analysis applications. Facial component segmentation is a challenging task due to variations in illumination conditions, pose, scale, skin color etc. Deep learning is a novel branch of machine learning, very efficient in solving complex problems. In this study, we developed a deep Convolutional Neural Network (CNN) to automatically segment facial components in face images. The network has been trained with face images in Radboud face database. Training labels have been created using Face++ SDK. The developed CNN produces a segmentation mask where mouth, eyes, and eyebrows components of the face are marked as foreground. We have focused on these components because they can include very important information for facial image analysis studies such as facial expression recognition. The segmentation success rate of the study is 98.01 according to average accuracy.

**Keywords:** Convolutional Neural Network, Deep Learning, Facial Feature Extraction, Facial Image Segmentation

## Introduction

Face image analysis tasks such as facial expression recognition, face recognition, face verification etc. have many practical uses in security (Moniruzzaman & Hossain, 2015), traffic (Maralappanavar et al., 2016), and healthcare (Bevilacqua et al., 2011) fields. Facial feature extraction is a fundamental but challenging step in face image analysis because of variations in illumination conditions, head pose and scale, skin color, and other factors such as occlusions, complex backgrounds, etc.

Two main categories of facial feature study and description are holistic description and local feature-based description (Li et al., 2017). Holistic approaches treat and investigate faces as a whole. Principle component analysis (PCA) (Chengjun Liu, 2004) linear discriminant analysis (LDA) (Linsen Wang et al., 2015) and independent component analysis (ICA) (Kwak & Pedrycz, 2007), are common methods for holistic face description. Local feature-based approaches analyze and describe a face in terms of its parts/components rather than as a whole. That can be beneficial for some situation, such as occlusion (Li et al., 2017).

Detection and interpretation of facial components are difficult tasks because of variations in shape, size, appearance, and relative positions of the facial components. Recently, numerous facial feature extraction methods have been proposed. Fang et al. (Fang et al., 2017) proposed a novel partial differential equation based method for facial feature learning. Perakis et al. (Perakis et al., 2014) provided a novel generalized framework of fusion methods for landmark detection. Gong et al. (Gong et al., 2017) presented a new feature descriptor for heterogeneous face recognition. Das et al. Ding et al. (Ding et al., 2014) introduced a new color balloon snake model for face segmentation in color images.

Deep learning catches the attention in machine learning and computer vision area because of its outstanding performance. In deep learning approaches, features are learned automatically and complex connections of the data can be resolved (Krizhevsky et al., 2012). Recently deep learning methods have been very popular in face analysis studies such as facial age estimation (Liu et al., 2017), facial beauty prediction (Xu et al., 2017), face detection (Triantafyllidou & Tefas, 2016), face recognition (Zeng et al., 2015), and facial expression recognition studies (Zhang et al., 2017). For instance, Liu et al. (Liu et al., 2017) presented a group-aware deep feature learning approach that learns discriminative face representation for facial age estimation, Ding et al. (Ding & Tao, 2015) proposed a deep learning framework to jointly learn face representation using multimodal information, Mukherjee et al. (Mukherjee & Robertson, 2015) presented a CNN based model for human head pose estimation in low-resolution multi-modal RGB-D data, Fan et al. (Fan & Zhou, 2016) presented a CNN structure for facial landmark localization.

This paper presents a deep learning approach to automatically segment facial components in face images. Viola & Jones face detection algorithm (Viola & Jones, 2001), has been used for face detection and cropping in the database images. We have developed a deep convolutional neural network (CNN) for segmentation of facial components in cropped face images. The network has been trained with face images and corresponding binary facial component masksmarking mouth, eyes, and eyebrows regions of the face. Training masks have been created using Face++ toolbox (Face++, 2017), postprocessing, and visual inspection. These specific components were selected because our ultimate goal is facial expression recognition and mouth, eyes, and eyebrows play and important role in facial expression formation (Lin Zhong et al., 2012).

The rest of the article is structured as follows. Section 2 gives details of the system. Section 3 presents the experimental results. Finally, Section 4 compiles the results of the study and makes suggestions for future studies.

### Proposed System

The proposed processing pipeline consists of three steps: (1) face cropping, (2) training data generation, and (3) development, training, and testing of a convolutional neural network (CNN) architecture. Block diagram of the proposed pipeline is shown in Figure 1. Training and testing images are first cropped by Viola&Jones face detection algorithm (Viola & Jones, 2001). Cropped images are then used for training data generation and facial component segmentation as follows.

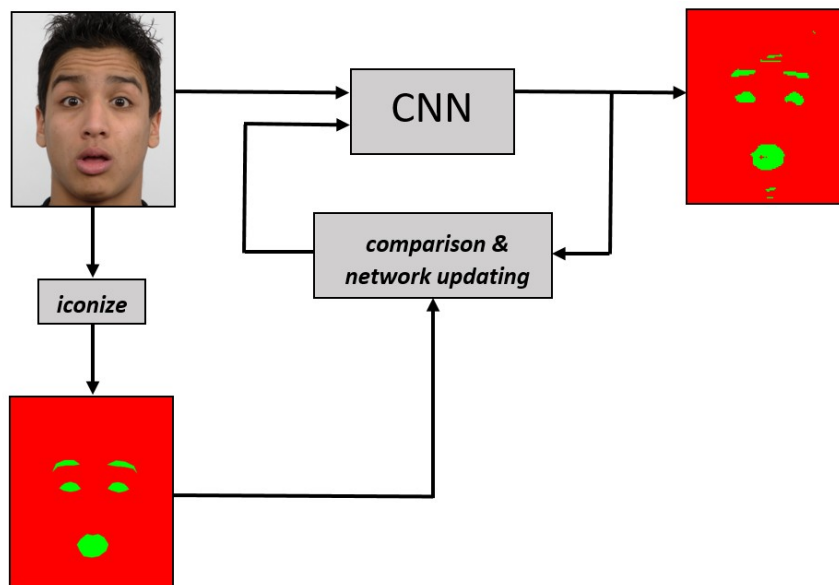


Figure 1. Proposed System

#### a) Training Data Generation

We have created training masks for facial component segmentation using Face++ SDK (Face++, 2017). The Face++ SDK detects 83 keypoints on human face. With the help of these keypoints, we have created close shapes on human eyes, eyebrows and mouths by polygon fitting to the detected landmark points. Finally, the input images have been iconized. These steps can be seen in Figure 2.

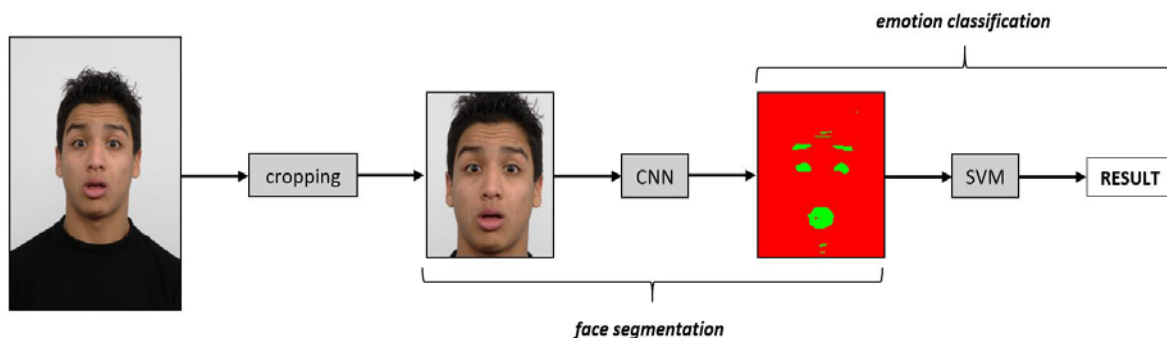
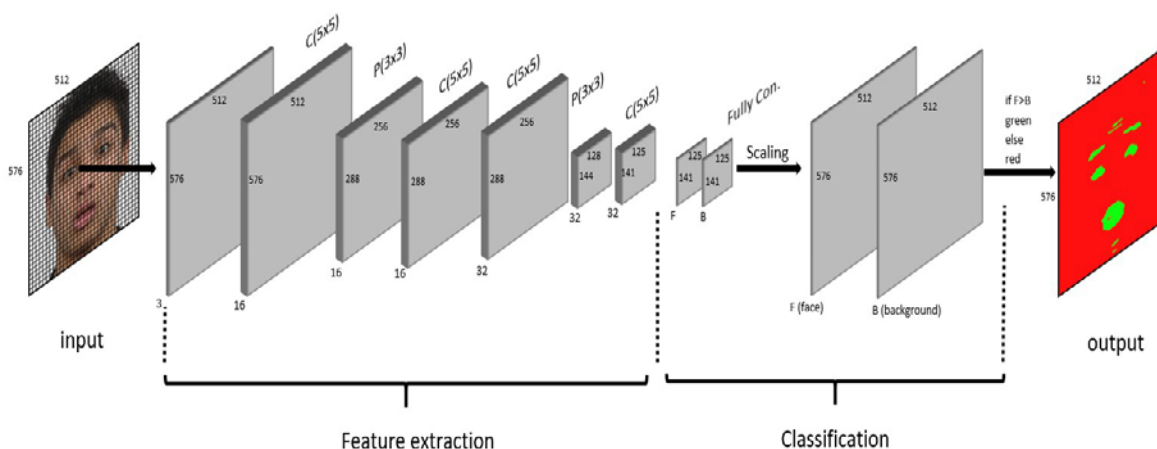


Figure 2. Training data generation steps

**b) The Proposed CNN Architecture**

We have developed a CNN structure to segment mouth, eyes and eyebrows component in face images. Segmentation is done by classifying 16x16 blocks on the image as background versus facial component. The proposed CNN architecture consists of one batch normalization layer, four convolutional layers (two layers with 16 5x5 filters, one layer with 32 5x5 filters and one layer with 32 4x4 filters), two pooling layers, and one fully connected layer. Training is done with 16x16 non-overlapping blocks extracted from the training image. It has been obtained 345600 blocks totally. Blocks are assigned a training data label based on the percentage of pixels from facial components and background classes. Blocks having 80% or more of their pixels from facial components or background classes, are kept for training, remaining 3018 mixed class blocks are removed from training. 7132 facial components blocks and 335450 background class are obtained.

Testing is done by feeding the whole image to efficiently simulate sliding window processing using the sliding window processing method described in (Shelhamer et al., 2016). Figure 3 illustrates the proposed CNN architecture. Table 1 shows kernel size, number of filters, and input and output size for each convolution layer.



**Figure 3.**Proposed CNN architecture

**Table 1:** Detailed layer information for the proposed CNN

Layer	Kernel	Filter	Input	Output
Conv1	5x5	16	576x512x3	576x512x16
Conv2	5x5	16	288x256x16	288x256x16
Conv3	5x5	32	288x256x16	288x256x32
Conv4	4x4	32	144x128x32	141x125x32

**Experiments**

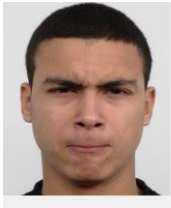


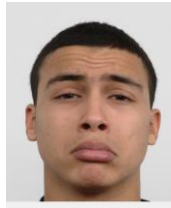
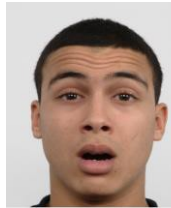
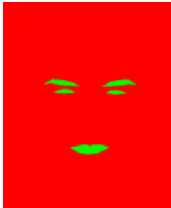
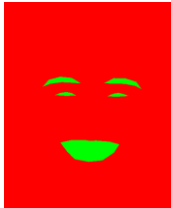
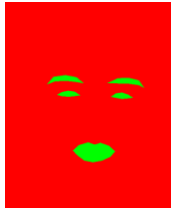
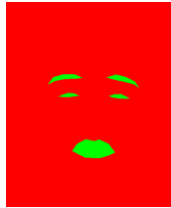
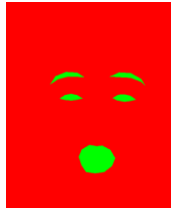




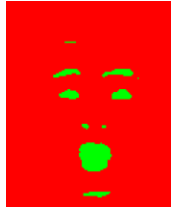
Radboud Face Database (Langner et al., 2010) has been used for training and testing of the proposed network. The images have been resized to 512x576 pixels in order to avoid partial blocks. 335 images have been used for the study (300 for training and 35 for testing).

We evaluated our segmentation result in terms of accuracy as shown in Table 2, where TP, TN, FP, and FN refer to true positives, true negatives, false positives, and false negatives, respectively. As can be seen in the Table 2; our average accuracy is 98.01%. Sample test results can be seen in Table 3.

**Table 2:** Success rate of our study

Index	Equation	Success Rate (%)
Accuracy (Acc)	$\frac{TP + TN}{TP + TN + FP + FN}$	98.0133

**Table 3:** Sample facial component segmentation results.

Expressions	Angry	Happy	Neutral	Sad	Surprised
Original face images					
Training Data					
Segmentation results produced by the proposed system					

### Conclusion

In this study, we presented a deep convolutional neural network approach for facial component segmentation. We focused on eyes, eyebrows and mouth components because these components play an important role in facial analysis studies, such as facial expression recognition and head pose estimation. Our next plan is to extend this work to facial expression recognition.

We anticipate that the iconized images obtained using the proposed pipeline will be useful to reduce data size requirements and privacy concerns.

### Acknowledgements

This research is supported by The Scientific and Technological Research Council of Turkey (TUBITAK) and the Sakarya University Scientifics Research Projects Unit (Project Number: 2015-50-02-039).

### References

- Bevilacqua, V. et al., (2011). A new tool to support diagnosis of neurological disorders by means of facial expressions. In *2011 IEEE International Symposium on Medical Measurements and Applications*. IEEE, pp. 544–549. Available at: <http://ieeexplore.ieee.org/document/5966766/>.
- Chengjun Liu, (2004). Gabor-based kernel pca with fractional power polynomial models for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5), pp.572–581. Available at: <http://ieeexplore.ieee.org/document/1273927/>.
- Ding, C. & Tao, D., (2015). Robust Face Recognition via Multimodal Deep Face Representation. *IEEE Transactions on Multimedia*, 17(11), pp.2049–2058. Available at: <http://ieeexplore.ieee.org/document/7243358/>.
- Ding, X. et al., (2014). Color balloon snakes for face segmentation. *Optik - International Journal for Light and Electron Optics*, 125(11), pp.2538–2542. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0030402613015076>.
- Face++ Cognitive Services. Available at: <https://www.faceplusplus.com/> [Accessed January 8,2017].
- Fan, H. & Zhou, E., (2016). Approaching human level facial landmark localization by deep learning. *Image and Vision Computing*, 47, pp.27–35. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0262885615001341>.

- Fang, C. et al., (2017). Feature learning via partial differential equation with applications to face recognition. *Pattern Recognition*, 69, pp.14–25. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0031320317301449>.
- Gong, D. et al., (2017). Heterogeneous Face Recognition: A Common Encoding Feature Discriminant Approach. *IEEE Transactions on Image Processing*, 26(5), pp.2079–2089. Available at: <http://ieeexplore.ieee.org/document/7812744/>.
- Krizhevsky, A., Sutskever, I. & Hinton, G.E., (2012). ImageNet Classification with Deep Convolutional Neural Networks. In F. Pereira et al., eds. *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc., pp. 1097–1105. Available at: <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>.
- Kwak, K.-C. & Pedrycz, W., (2007). Face Recognition Using an Enhanced Independent Component Analysis Approach. *IEEE Transactions on Neural Networks*, 18(2), pp.530–541. Available at: <http://ieeexplore.ieee.org/document/4118266/>.
- Langner, O. et al., (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion*, 24(8), pp.1377–1388.
- Li, Z. et al., (2017). Multi-ethnic facial features extraction based on axiomatic fuzzy set theory. *Neurocomputing*, 242, pp.161–177. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0925231217303946>.
- Lin Zhong et al., (2012). Learning active facial patches for expression analysis. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 2562–2569. Available at: <http://ieeexplore.ieee.org/document/6247974/>.
- Linsen Wang et al., (2015). Linear discriminant analysis using sparse matrix transform for face recognition. In *2015 IEEE 17th International Workshop on Multimedia Signal Processing (MMSP)*. IEEE, pp. 1–6. Available at: <http://ieeexplore.ieee.org/document/7340852/>.
- Liu, H. et al., (2017). Group-aware deep feature learning for facial age estimation. *Pattern Recognition*, 66, pp.82–94. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0031320316303417>.
- Maralappanavar, S., Behera, R. & Mudenagudi, U., (2016). Driver’s distraction detection based on gaze estimation. In *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, pp. 2489–2494. Available at: <http://ieeexplore.ieee.org/document/7732431/>.
- Moniruzzaman, M. & Hossain, M.F., (2015). Image watermarking approach of criminal face authentication with recovery for detecting exact criminal. In *2015 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*. IEEE, pp. 1–6. Available at: <http://ieeexplore.ieee.org/document/7226020/>.
- Mukherjee, S.S. & Robertson, N.M., (2015). Deep Head Pose: Gaze-Direction Estimation in Multimodal Video. *IEEE Transactions on Multimedia*, 17(11), pp.2094–2107. Available at: <http://ieeexplore.ieee.org/document/7279167/>.
- Perakis, P., Theoharis, T. & Kakadiaris, I.A., (2014). Feature fusion for facial landmark detection. *Pattern Recognition*, 47(9), pp.2783–2793. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0031320314001058>.
- Shelhamer, E., Long, J. & Darrell, T., (2016). Fully Convolutional Networks for Semantic Segmentation. Available at: <http://arxiv.org/abs/1605.06211>.
- Triantafyllidou, D. & Tefas, A., (2016). Face detection based on deep convolutional neural networks exploiting incremental facial part learning. In *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE, pp. 3560–3565. Available at: <http://ieeexplore.ieee.org/document/7900186/>.
- Viola, P. & Jones, M., (2001). Rapid object detection using a boosted cascade of simple features. *Computer Vision and Pattern Recognition (CVPR)*, 1, p.I–511–I–518.
- Xu, J. et al., (2017). Facial attractiveness prediction using psychologically inspired convolutional neural network (PI-CNN). In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, pp. 1657–1661. Available at: <http://ieeexplore.ieee.org/document/7952438/>.
- Zeng, J., Zhai, Y. & Gan, J., (2015). A Novel Sparse Representation Classification Face Recognition Based on Deep Learning. In *2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom)*. IEEE, pp. 1520–1523. Available at: <http://ieeexplore.ieee.org/document/7518453/>.
- Zhang, K. et al., (2017). Facial Expression Recognition Based on Deep Evolutional Spatial-Temporal Networks. *IEEE Transactions on Image Processing*, pp.1–1. Available at: <http://ieeexplore.ieee.org/document/7890464/>.