International Journal of Recent Engineering Research and Development (IJRERD)
ISSN: 2455-8761
www.ijrerd.com || Volume 04 – Issue 08 || August 2019 || PP. 92-96

Convolutional Neural Network Approach on Face Illustrations

Özgür Deniz Günseli[1], Semih Utku[2]

Computer Engineering Department, Dokuz Eylul University, Izmir, Turkey

[2]Computer Engineering Department, Dokuz Eylul University, Izmir, Turkey

Abstract: Machine learning and artificial intelligence are included in all areas of our lives with the increase in technological possibilities. Artificial intelligence technologies are used in many different fields with high success rates. Machine learning and artificial intelligence methods perform face detection and face analysis in many different areas. As face detection and face recognition technologies develop, the prevalence of usage increases accordingly. Department stores and businesses use emotion analyzes to track the satisfaction of their employees and customers. These solutions perform emotion analysis by detecting the faces of employees and customers. However, in these solutions, data is usually processed on company servers or on cloud. In many ways, such solutions pose problems with the confidentiality of personal data. In this study, we propose an emotion analysis method with Convolutional Neural Network on anonymized face illustrations.

Keywords: Emotion recognition, machine learning, convolutional neural network, deep learning

I. INTRODUCTION

Face detection and face recognition techniques have been used for a long time. However, with the development of artificial intelligence and machine learning methods and the increase in hardware opportunities, the success rate increases rapidly. Especially with the development of GPU technology and Deep Learning techniques in recent years, success rates have exceeded human perceptions. With the Artificial Neural Network and Convolutional Neural Network approaches, face detection and recognition methods have become common technologies.

Conventional face recognition methods consist of three stages. In the first step, the faces in the picture or video are detected. In the second step, feature extraction process is performed on detected faces. In the final stage, the classification is made over the extracted features. These three operations can be performed at a single point or distributed at different points. There are many different approaches for each stage.

In the Convolutional Neural Network approach, the above mentioned feature extraction and classification stages are performed in a single stage. In model training and classification procedures, the visual is given as a whole. Thanks to the development of GPU processors in recent years, very complex model trainings and very high success rates can be achieved with CNN. These success rates outperform conventional approaches. For this reason, it is widely used for the classification of visual data.

However, the high processing power requirements of Convolutional Neural Networks restrict this approach to operating on low processing hardware. For this reason, CNN often works on central server processing of data collected from endpoints with low processing power. This approach creates problems in terms of data collected from public fields. Therefore, in this study, we propose an emotion recognition method based on CNN-based and anonymized data.

II. RELATED WORKS

There have been many studies in the past regarding facial emotion recognition. For facial emotion analysis, first of all, standardized emotion definitions should be performed. In his study, Ekman identified six basic emotions[1], [2]. In later studies[2]–[4], he expanded the basic 6 emotions to 11 different emotions. Plutchik carried out a study involving 8 different basic mental states and their different levels and interactions[5]. Ekman and Friesen, in their studies in between 1967 and 1969, stated that there is a relationship between emotions and muscles in the face[1], [6], [7]. In 1978, they carried out a study called "Facial Action Coding System"[8]. In this study, they stated that different psychological conditions create certain facial expressions on the lips, eyes, eyebrows and cheeks. For each special expression that appears on the face, they have defined the definition of "Action Unit". 68 different landmarks are used in the identification of Action Units. Many different studies have been conducted on this study. Studies have been carried out to extract these features from visual data using different methods.

In recent years, many face detection and emotion detection studies have been carried out using Convolutional Neural Network methods. L. Haoxiang and Lin performed face detection using the Convolutional Neural Network[9]. In their study, they worked on the Annoted Facial Landmarks in the Wild[10] dataset with a size of 48x48 pixels. They used a 2-tier Convolutional Neural Network architecture with 5x5 kernel sizes and 64 features. Omkar M. Parkhi and Andread Vedaldi[11] performed Face Recognition with Convolutional Neural Network using the Labeled Faces in the Wild[12] and Youtube Faces[13] datasets. Ali Mollahosseini, David Chan and Mohammad H. Mahoor conducted a study with Convolutional Neural Network on 7 different data sets[13].They performed facial expression recognition in this study. They used 175K samples for training, 56K samples for testing and 64K samples for validation in 7 different data sets. They trained the Convolutional Neural Network in 200 Epochs. They got 94.7±0.8% accuracy on MultiPIE dataset, 77.6±2.9% accuracy on MMI dataset, 55.0±6.8% accuracy on DISFA dataset, 76.7±3.6% accuracy on FERA dataset, 47.7%1.7± accuracy on SFEW dataset, 93.2±1.4% accuracy in CK+ data set and 66.4±0.6% accuracy on FER2013 dataset.

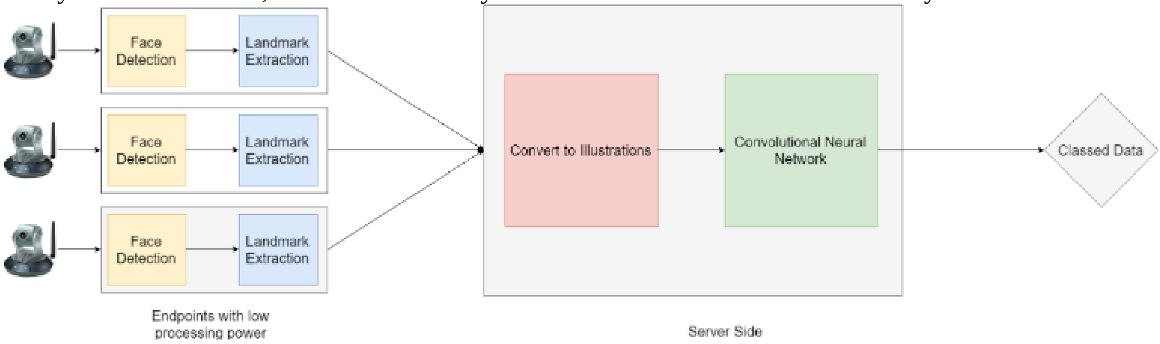


Figure 1: System Architecture

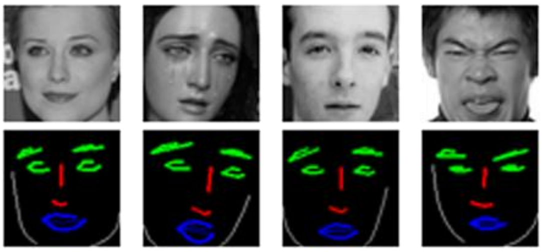


Figure 2: Original and processed images

When we examine the studies that do not use Convolutional Neural Network as a classification method, there are many different approaches for feature extraction and classification stages. S. Bashyal and G. K. Venayagamoorthy used Gabor Filters in their studies[14]. M. Nazir, Z. Jan, and M. Sajjad have performed feature extraction with Histogram of Oriented Gradient method[15]. S. L. Happy and A. Routray used the Local Binary Patterns method[16]. For classification, Support Vector Machine and K-Nearest Neighbor were used in these studies.

III. METHOD

In the proposed method, a hybrid solution of the conventional approach and the Convolutional Neural Network approach is presented. In our study, human face detection and feature extraction are performed by conventional methods. Theclassification process is provided by Convolutional Neural Network approach.

In the first stage of our study, face detection and feature extraction are performed in low processing power units. At this terminal, 68 facial landmarks are separated for each detected face. These separated landmarks are transferred to the server where the classification process will be performed over the internet. Thus, the data generated from the detected faces are purified from the personal data. In addition, the resulting data size is considerably smaller than the original image size.

International Journal of Recent Engineering Research and Development (IJRERD)
ISSN: 2455-8761
www.ijrerd.com || Volume 04 – Issue 08 || August 2019 || PP. 92-96

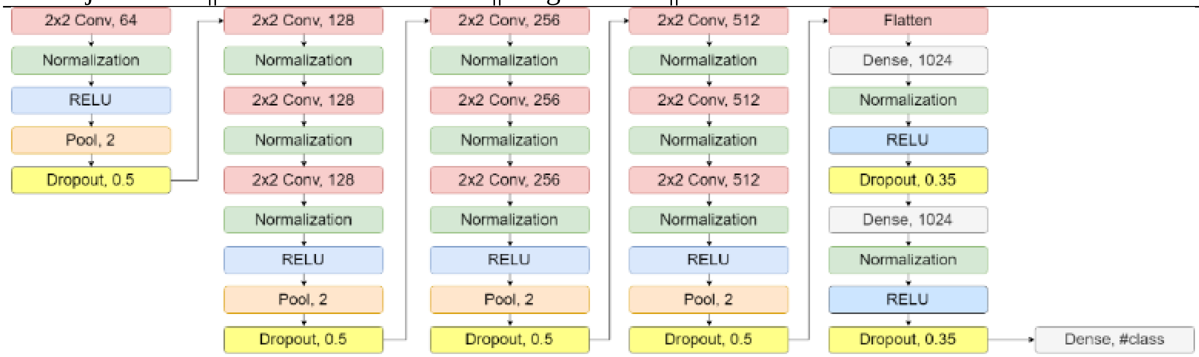
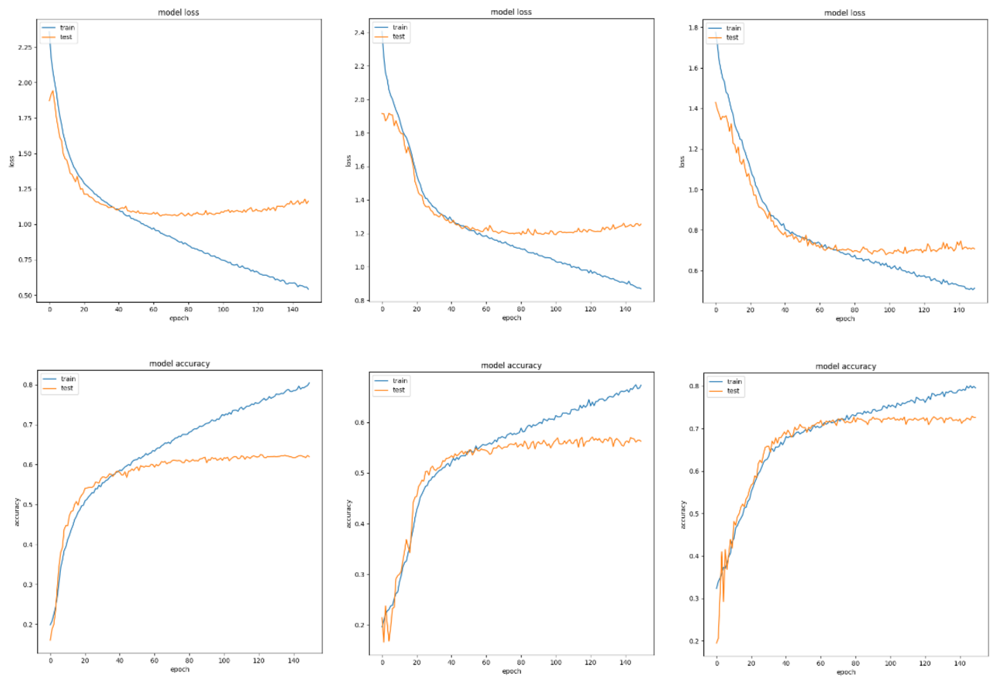


Figure 3: CNN Architecture

The separated facial landmarks are translated into illustrated images on the server. For each face, eyes, eyebrows, nose, lips and jaw as the image again 100x100 pixel size is created.

The images are generated from the facial landmarks which are collected from the terminals. Convolutional Neural Network is used on the server for the classification of images. In this way, the data collected from the terminals with low processing power is processed on the server with Convolutional Neural Network. The general structure of the system is given in Figure 1.

In our study, we create a network of 4 Convolutional Layers. We performed Batch Normalization after each Convolution process. We added dropout to the end of each Convolutional Layer to reduce overfitting. We added 2 dense layers to the end of the Convolutional Layers. Each dense layer consists of 1024 neurons. After each dense layer there is Relu activation and dropout. We used the FER-2013 database[17] to train our CNN model. The FER-2013 database contains 35,887 images from 7 classes. We trained our model primarily on the FER-2013 database, which we did not make any image processing on it. Then we trained the same model on the pictures that we made face extraction and illustrate with 68 facial landmarks. Finally, we trained our model only with angry, happy, natural and sad emotion images that is processed and illustrated. We performed 150 Epochs in each training process.



a) b) c)

Figure 4: Loss and accuracy values for each dataset a) Original FER2013 Dataset b) Processed FER2013 Dataset c) 4 Selected Emotion Dataset

IV. RESULT

The FER-2013 data set used to train our model has a state-of-the-arts value of 75.2% and human accuracy was 68±5%. In the Kaggle competition held in 2013, the first four success rates were 71.16%, 69.26%, 68.82% and 67.48%[17]. The indicated studies were performed with a specific test data set. In our study, we selected 20% of whole data and used for testing. In our tests with the proposed CNN architecture, we achieved an accuracy of 62.49% in our training, which includes raw FER-2013 data.

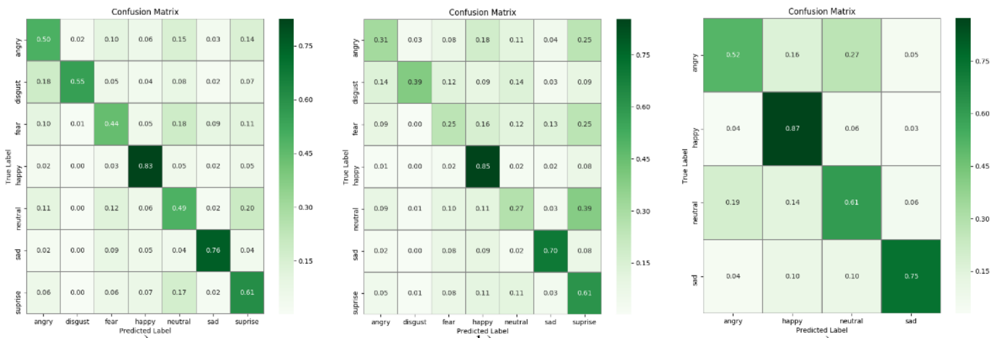
In the analyzes we performed on all processed data set, our model performed classification with 57.02% accuracy. The difference between our classification value which was lower than the state-of-the-arts value and the accuracy value of our model in the standard pictures was calculated as 5.47%. The accuracy value obtained in processed images of 4 different emotions was 72.90%.

In our study, the highest accuracy rate for all three data sets was achieved for happy. In the study conducted on the processed data, a success rate of 85% was achieved and a success rate of 87% was obtained in the data set limited to 4 mental states. For all data sets, the second successful accuracy rate was obtained for sad. The accuracy of 70% on the processed data set and the accuracy of 75% on the data set limited to 4 emotion yielded an accuracy of 76% on the raw data set. 25.04% of the data set in which the study was carried out consisted of happy images and 11.51% consisted of sad images. By increasing the number of sample data describing each mental state, the accuracy rate can be increased to higher levels.

In this study, while creating face shapes from facial landmarks, all face shapes were created with the same pixel size. When performing illustration operations, accuracy can be increased with using different pixel sizes for each face shape.

Table 1: Test Results

Data	*Validation Accuracy*	*Validation Loss*	*State-of-the-arts*
Original FER2013 Data	% 62.49	1.095	% 75.2
Processed FER2013 Data	% 57.02	1.2079	-
4 Selected Emotions	% 72.90	0.7069	-



a) b) c)

Figure 5: Confusion matrix for each dataset a) Original FER2013 Dataset b) Processed FER2013 Dataset c) 4 Selected Emotion Dataset

References

[1] P. Ekman and W. V. Friesen, “Constants across cultures in the face and emotion,” J. Pers. Soc. Psychol., 1971.
[2] P. Ekman, “Universal-Facial-Expressions-of-Emotion,” California Mental Health. 1970.
[3] P. Ekman, “An Argument for Basic Emotions,” Cogn. Emot., 1992.
[4] P. Ekman, “Are There Basic Emotions?,” Psychol. Rev., 1992.
[5] R. PLUTCHIK, “A GENERAL PSYCHOEVOLUTIONARY THEORY OF EMOTION,” in Theories of Emotion, 1980.
[6] P. Ekman and W. V. Friesen, “Head and body cues in the judgment of emotion: a reformulation.,” Percept. Mot. Skills, 1967.
[7] P. Ekman and W. V. Friesen, “The repertoire of nonverbal behavior: Categories, origins, usage, and coding,” in Mouton Classics: From Syntax to Cognition. From Phonology to Text, 2013.
[8] P. Ekman and W. V. Friesen, “Facial Action Coding System: A Technique for the Measurement of Facial Movement,” J. Pers. Soc. Psychol., 1971.
[9] L. Haoxiang and Lin, “A Convolutional Neural Network Cascade for Face Detection,” CVPR, 2015.
[10] M. Köstinger, P. Wohlhart, P. M. Roth, and H. Bischof, “Annotated facial landmarks in the wild: A

large-scale, real-world database for facial landmark localization," in Proceedings of the IEEE International Conference on Computer Vision, 2011.

[11] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," 2015.

[12] E. Learned-Miller, G. B. Huang, A. RoyChowdhury, H. Li, and G. Hua, "Labeled faces in the wild: A survey," in Advances in Face Detection and Facial Image Analysis, 2016.

[13] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in 2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016, 2016.

[14] S. Bashyal and G. K. Venayagamoorthy, "Recognition of facial expressions using Gabor wavelets and learning vector quantization," Eng. Appl. Artif. Intell., 2008.

[15] M. Nazir, Z. Jan, and M. Sajjad, "Facial expression recognition using histogram of oriented gradients based transformed features," Cluster Computing, 2017.

[16] S. L. Happy and A. Routray, "Automatic facial expression recognition using features of salient facial patches," IEEE Trans. Affect. Comput., 2015.

[17] I. J. Goodfellow et al., "Challenges in representation learning: A report on three machine learning contests," Neural Networks, 2015.

Author Profile

Özgür Deniz Günseli received the B.S. degree in Computer Science from Dokuz Eylul University in 2015 and he is attending Master degrees in Computer Science at Dokuz Eylul University. Working as senior software developer in Turkey.

Semih UTKU is an Assistant Professor at the Department of Computer Engineering at Dokuz Eylul University. He received Ph.D. degree
from the University of Dokuz Eylul in 2011. Research interests include information retrieval, machine learning and data mining, database design and applications, distributed system and cloud computing.