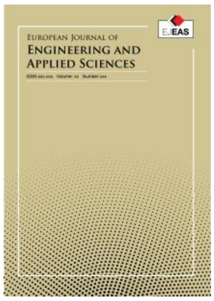
European Journal of Engineering and Applied Sciences



ISSN: 2651-3412 (Print) & 2667-8454 (Online)
Journal homepage: http://dergipark.gov.tr/EJEAS
Published by Çorlu Faculty of Engineering, Tekirdağ Namık Kemal University

European J. Eng. App. Sci. 4(1), 14-17, 2021

Research Article

Emotion Detection from Facial Expression Using Different Feature Descriptor Methods with Convolutional Neural Networks

Fatih Altekin [1,] , Hasan Demir [1*,]

[1] **Electronic and Communication Engineering, Çorlu Faculty of Engineering, Tekirdağ Namık Kemal University, Tekirdağ, Turkey**

Received: 10.01.2021 *Accepted: 16.01.2021*

Abstract: In this article, image processing techniques to detect facial emotion expressions are examined. Studies done to detect facial expression are given in stages. The success of the convolutional neural networks (CNN) method in emotional expression has been investigated. A set of 981 CK + pictures containing human faces in 7 emotion categories was used. The success rates when using HOG, LBP, Wavelet feature of images and the original state of the images in the data set were compared.

Keywords: *Convolutional Neural Network (CNN), Histogram of gradients (HOG), Local Binary Pattern (LBP) Wavelet Transform, Facial Emotion Expression Recognition, CK+datasets*

Farklı Öznitelik Tanımlayıcı Yöntemlerini Evrişimsel Sinir Ağları ile Kullanarak Yüz İfadesinden Duygu Tespiti

Özet: Bu makalede, yüzdeki duygu ifadelerini tespit etmek için imge işleme teknikleri incelenmiştir. Yüzdeki duygu ifadelerini tespit etmek için yapılan çalışmalar aşamalar halinde verilmiştir. Evrişimsel sinir ağları (CNN) yönteminin duygu ifadeleri tespitindeki başarısı ele alınmıştır. 7 duygu katogorisinde, insan yüzleri içeren 981adet imgeden oluşan CK+ imge seti kullanılmıştır. Veri setindeki imgelerin orijinal hali ve HOG, LBP ve İmgelerin dalgacık dönüşümü özniteliklerinin kullanıldığı durumlardaki başarı oranları karşılaştırılmıştır.

Anahtar kelimeler: *Evrişimsel sinir ağları, gradyanların histogramı, yerel iki örüntüler, dalgacık dönüşümü, yüz duygu ifadeleri tanımlama, CK+ veriseti*

Corresponding author.
E-mail address: hdemir@nku.edu.tr (H. Demir)

1. Introduction

In this article, facial emotion detection has been made on images. In order to determine the facial area, important points were determined from the facial regions such as eyes, eyebrows, nose and mouth. Face features are extracted using HOG, LBP and Wavelet transform methods.

Finally, facial emotion analysis was performed using the CNN Classification algorithm and 7 basic emotions such as ’anger’, ’contempt’, ’disgust’, ’fear’, ’happy’, ’sadness’, ’surprise’ were tried to be identified.

In their study, Bayrakdar et al. [1] examined the studies in the literature on facial expression recognition and included some methods used in the analysis of facial expressions. As a result of the rapid development of these methods and human-machine interaction, they concluded that human emotional expressions can be detected by computers. They concluded that fast and clear analysis of facial expressions and their accurate recognition is very important for many software systems in different application areas.

Viola and Jones [2] in their study found a machine learning approach for physical object detection that can process images extremely quickly and achieve high detection rates. First, a new image display called "Integral linage" was designed, which enables the features used by the detector to be calculated very quickly. Next, a learning algorithm based on AdaBoost that can select a small number of critical visual features from a larger cluster and provide a highly effective classroom efficiency was developed. Finally, a method has been implemented to combine increasingly complex classifiers in a "cascade" that allows the background regions of the image to be discarded rapidly when calculating on object-like regions. As a result of the study, the rate of detecting the face region of the system is given by a comparative way.

Lucey et al. [3] examined the Cohn-Kanade (CK) database published in 2000 to automatically detect individual facial expressions. It was concluded that the CK database was used for both action units (AU) and emotion perception. It is stated that the use of random subsets of the original database makes difficult meta-analysis. An expanded Cohn-Kanade (CK +) database has been developed to address these and other concerns. In this database, it was observed that the number of rankings increased by 22% and the number of projects increased by 27%.

Image can be defined as a two-dimensional function. The height of the image is one of these dimension and the width is another dimension [4].

Image processing is a technique that enables operations such as increasing the clarity of the image, emphasizing or changing certain features on the image, detecting any object in the image, on images created by any image capture device.

Mathematically, a two-dimensional matrix is obtained by image digitization. By these matrices, processes such as image enhancement, object detection from the image can be performed.

After digitizing the image being processed in order to make an emotion analysis from a facial expression, the first step is to detect the face in the image. Then, by determining the features of the face, the basic points that will help us to identify facial expressions such as mouth, nose, eyes, eyebrows are determined. In the next step, the points determined are analyzed and the features of facial expressions are extracted. In the last step, the facial expression recognition process is performed by making the classification process. The stages are shown in Figure 1.1.

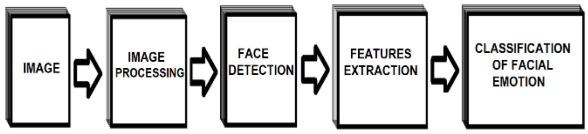


Figure 1.1 Emotion recognition diagram from facial expressions

1.1. Face detection

Face detection is the absence of a facial area in an image. There are many different methods for face detection, such as color space method, Haar Cascade classifiers method, using dlib libraries.

1.2. Face Features Extraction

The next step after face detection is the removal of facial features. The permanent features of the face such as eyes, eyebrows, nose, mouth and facial lines are detected. Feature descriptors are used to extract face features.

The features descriptor is a representation of an image created by taking useful information and eliminating unnecessary information, thus simplifying the image.

Typically, a feature descriptor converts an image of size width, height, 3 (channels) into a feature vector or array of length n.

Histogram of Oriented Gradient (HOG), focuses on the structure or shape of an object. When it comes to edge properties, it is only determined whether the pixel is an edge. And this is done by subtracting the gradient and direction of the edges.

Local Binary Pattern (LBP) is a simple but very efficient texture operator that labels the pixels of an image by mapping the neighborhood of each pixel and views the result as a binary number. Because of its discriminatory power and computational simplicity, the LBP tissue operator is a popular approach in a variety of applications. Traditionally, tissue analysis can be seen as a unifying approach to different statistical and structural models.

Wavelet Transform can be applied to one-dimensional signals as well as two-dimensional signals (pictures, etc.). This is the part that concerns the fields of Computer Vision and Image Processing. It allows to compress the image down to a certain size.

1.3. Facial Expression Classification

It is the last and most important stage in which facial emotion expressions are determined. Emotional expressions are tried to be defined by using the facial features found in the previous stage. Classification is made according to the similarities among the facial features. A wide variety of classification methods have been developed. In this article, we used the Convolutional Neural Networks (CNN) method.

In deep learning, a convolutional neural network (CNN or ConvNet) is a class of deep neural networks most commonly applied to analyze visual images. It has applications in image and video recognition, advisory systems, image classification, medical image analysis, natural language processing, and financial time series.

The convolution network called CNN or ConvNet is a feed forward neural network and is also called multilayer perceptrons (MLPs) [5]. Its main subject is image recognition and image classifications, which are widely used in the current generation. Computer vision begins with reading the input image in the form of strings of pixels.

The Convolution Layer extracts the features of an input image from the training dataset where the core is a small part of an input image for feature classification. The design begins with starting the CNN model by taking an input image (static or dynamic) created by adding a convolution layer, merge layer, flattening layers, and dense layers. Convolution layers are added for better accuracy in large data sets.

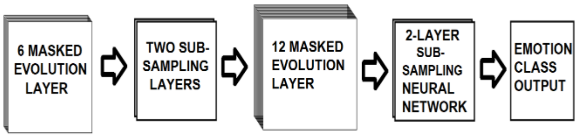


Figure 1.2 Convolutional Neural Network Architecture for Facial Expression Recognition

The architecture of CNN is depicted in Figure 1.2 [6]. It has two convolutional layers and two subsampling layers. The first convolutional layer is known as the undermasked or c1 layer. The next layer is the subsampling layer with two layers (s1). The latter convolutional layer (or c2) has 12 masks. The final subsampling neural network has two layers. The last is a fully linked layer that causes classification.

2. Emotion Detection from Facial Expression with Convolutional Neural Networks

In this article, applications are written using python programming language and numpy, opencv, matplotlib, dlib, pandas, sklearn, tensorflow, keras libraries [7, 8, 9, 10, 11, 12, 13]

CK + image set consisting of 981 facial images was used to recognize emotions from facial expression. First, the data set was created using the data set obtained from the CK + picture set. And this dataset has been trained through the model which we created by convolutional neural networks (CNN). The success rate of the model was calculated by entering test data into the trained model. In addition, the success of the model created by calculating the complexity matrix in detecting each emotion was determined. Since the data in the complexity matrix is 30% test data, it will be 981 * 0.3 = 295, so 295 data labels were obtained.

With the python programs we wrote, HOG, LBP, Wavelet pictures of each picture in the original CK+ picture set were found. The data set created from the pictures obtained was applied to our CNN model.

The CNN model diagram which we have created with the Python programming language is shown in figure 2.1.

The success rates of the facial expression recognition models created by using HOG, LBP, Wavelet and Original pictures by CNN model are given in Table 2.1.

When we examine the accuracy score table, very approximate results are obtained when the CNN model is used with all face feature detection methods or original pictures.

It was determined that the training period of the model was shorter due to the fact that the wavelet pictures were approximately 1/4 smaller than the other pictures.

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 46, 46, 64)	1792
conv2d_1 (Conv2D)	(None, 46, 46, 64)	36928
batch_normalization (BatchNo	(None, 46, 46, 64)	256
max_pooling2d (MaxPooling2D)	(None, 23, 23, 64)	0
dropout (Dropout)	(None, 23, 23, 64)	0
conv2d_2 (Conv2D)	(None, 23, 23, 128)	73856
batch_normalization_1 (Batch	(None, 23, 23, 128)	512
conv2d_3 (Conv2D)	(None, 23, 23, 128)	147584
batch_normalization_2 (Batch	(None, 23, 23, 128)	512
max_pooling2d_1 (MaxPooling2	(None, 11, 11, 128)	0
dropout_1 (Dropout)	(None, 11, 11, 128)	0
conv2d_4 (Conv2D)	(None, 11, 11, 256)	295168
batch_normalization_3 (Batch	(None, 11, 11, 256)	1024
conv2d_5 (Conv2D)	(None, 11, 11, 256)	590080
batch_normalization_4 (Batch	(None, 11, 11, 256)	1024
max_pooling2d_2 (MaxPooling2	(None, 5, 5, 256)	0
dropout_2 (Dropout)	(None, 5, 5, 256)	0
conv2d_6 (Conv2D)	(None, 5, 5, 512)	1180160
batch_normalization_5 (Batch	(None, 5, 5, 512)	2048
conv2d_7 (Conv2D)	(None, 5, 5, 512)	2359808
batch_normalization_6 (Batch	(None, 5, 5, 512)	2048
max_pooling2d_3 (MaxPooling2	(None, 2, 2, 512)	0
dropout_3 (Dropout)	(None, 2, 2, 512)	0
flatten (Flatten)	(None, 2048)	0
dense (Dense)	(None, 512)	1049088
dropout_4 (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 256)	131328
dropout_5 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 128)	32896
dropout_6 (Dropout)	(None, 128)	0
dense_3 (Dense)	(None, 7)	903

Total params: 5,907,015
Trainable params: 5,903,303
Non-trainable params: 3,712

Fig. 2.1 *The CNN model diagram*

Table 2.1. Accuracy score comparison chart

CNN model with different Features	Training accuracy score	Test accuracy score	Training Time (Second)
HOG images features	% 99.27	% 96.27	450.8
LBP images features	% 100.0	% 96.95	460.6
Wavelet images features	% 97.81	% 95.93	313.6
Original images features	% 98.67	% 96.61	460.6

When we created a CNN model through HOG, LBP, Wavelet feature detection methods and original pictures, it was found by using the confusion matrix how accurately 7 facial emotion expressions were predicted from 295 test pictures. And the results are given in table 2.2 comparatively.

Table 2.2. Facial emotion recognition accuracy comparison

CNN emotion predict	**anger**	**contempt**	**disgust**	**fear**	**happy**	**sadness**	**surprise**
HOG features	93.02	100	98.24	100	93.15	96.15	98.59
LBP features	95.12	73.33	98.21	94.73	100	96.29	100
Wavelet features	94.69	100	94.91	90.47	100	85.71	100
Orginal features	94.73	100	98.24	95	100	80.64	98.61

3. Conclusions

In this study, the most successful results in detecting emotion expression with convolutional neural networks (CNN) were obtained by using LBP features. Wavelet features gave the best results in terms of duration.

References

[1] S. Bayrakdar, D. Akgün, and İ. Yücedağ, "Yüz ifadelerinin otomatik analizi üzerine bir literatür çalışması A survey on automatic analysis of facial expressions," pp. 383–398, 2016.

[2] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, 2001.

[3] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression," *2010 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. - Work. CVPRW 2010*, no. July, pp. 94–101, 2010.

[4] R. C. Gonzalez, "Digital Image Processing Third Edition."

[5] R. O. K. Reddy and C. Raghavendra, "Effective Facial Emotion Recognition using Convolutional Neural Network Algorithm," *Int. J. Recent Technol. Eng.*, vol. 8, no. 4, pp. 4351–4354, 2019.

[6] D. Y. Liliana, "Emotion recognition from facial expression using deep convolutional neural network," *J. Phys. Conf. Ser.*, vol. 1193, no. 1, 2019.

[7] C. R. Harris *et al.*, "Array programming with {NumPy}," *Nature*, vol. 585, no. 7825, pp. 357–362, 2020.

[8] G. Bradski, "The OpenCV Library," *Dr. Dobb's J. Softw. Tools*, 2000.

[9] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, 2007.

[10] D. E. King, "Dlib-ml: A Machine Learning Toolkit," *J. Mach. Learn. Res.*, vol. 10, pp. 1755–1758, 2009.

[11] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[12] Martin~Abadi *et al.*, "{TensorFlow}: Large-Scale Machine Learning on Heterogeneous Systems." 2015.

[13] F. Chollet, "keras," *GitHub repository*, 2015. [Online]. Available: https://github.com/fchollet/keras.