International Journal of Advanced Culture Technology Vol.6 No.1 60-67 (2018)
https://doi.org /10.17703/IJACT.2018.6.1.60

IJACT 18-1-8

A Study on the Facial Expression Recognition using Deep Learning Technique

Bong Jae Jeong[1], Min Soo Kang[2], Yong Gyu Jung[2]

[1]Department of Medical IT Marketing, Eulji University, Korea
[2]Department of Medical IT, Eulji University, Korea
E-mail: bongjaejeong@naver.com, {mskang, ygjung}@eulji.ac.kr

Abstract

In this paper, the pattern of extracting the same expression is proposed by using the Android intelligent device to identify the facial expression. The understanding and expression of expression are very important to human computer interaction, and the technology to identify human expressions is very popular. Instead of searching for the symbols *that users often use, you can identify facial expressions with a camera, which is a useful technique that can be used now. This thesis puts forward the technology of the third data is available on the website of the set, use the content to improve the infrastructure of the facial expression recognition accuracy, to improve the synthesis of neural network algorithm, making the facial expression recognition model, the user's facial expressions and similar expressions, reached 66%. It doesn't need to search for* symbols*. If you use the camera to recognize the expression, it will appear* symbols *immediately. So, this service is the* symbols *used when people send messages to others, and it can feel a lot of convenience. In countless* symbols*, there is no need to find* symbols*, which is an increasing trend in deep learning. So, we need to use more suitable algorithm for expression recognition, and then improve accuracy.*

Keywords: *Deep Learning, Tensor Flow, CNN, Facial Expression Recognition, Android Intelligent Phone.*

1. Introduction

Human communication includes not only by speaking or hearing but also non-verbal cues such as hand gestures and facial expressions which are used to express feelings. According to Albert Mehrabian's law, a person receives 55% of the image from the other, 38% of the hearing, and 7% of the language and visual communication takes up a large part of the image. Psychologists Ekman and Friesen studied the commonality of human facial expressions on different cultural backgrounds and proposed six representative facial expressions. To understand people's emotions, researchers have searched in terms of psychology, physiology, image processing, pattern recognition, machine vision, artificial intelligence, etc., and research results play an important role in various fields.

In this paper, I conducted a study to extract appropriate symbols for each facial expression by using facial expression recognition technology. ‘FER2013’ data provided in Kaggle is used to make a facial expression recognition model. This data contains total 35887 images and has 7 different expressions (Angry, Disgust, Fear, Happy, Sad, Surprise, Neutral). CNN (Convolutional Neural Network) which is used often for image

Manuscript Received: February 15, 2018 / Revised: February 27, 2018 / Accepted: March 10, 2018
Corresponding Author: mskang@eulji.ac.kr
Tel:+82-31-740-7153
Dept. of Medical IT, Eulji University, Korea

recognition is used to make a model with 'FER2013' data.

Based on this research result, an appropriate research on extracting the correct symbol using the trained model applied Android intelligent device camera will help advance the uncomfortable issues of the Android intelligent device.

2. Related Study

2.1 Deep Learning

Deep learning is defined as a set of machine learning algorithms that attempt to combine high-level abstractions (summarizing core content or functions in large amounts of data or complex data) through a combination of several non-linear transformation techniques, in a big frame, it is a field of machine learning that teaches computers how people think. When there is any data, it is expressed in a form that the computer can understand (for example, in the case of an image, pixel information is represented by a column vector) and to apply this to learning, many studies (how to make better expression techniques and how to model them) are processed. Because of these efforts, various deep learning techniques such as deep neural networks, convolutional deep neural networks, and deep belief networks have been applied to fields such as computer vision, speech recognition, natural language processing, and voice/signal processing.

2.2 Tensor flow

It is a machine learning library that opened as open source in Google. Deep learning and machine learning are offered in various functions to make it easier for the public to use. You can write operations using Python, a high-level programming language. Most other languages are supported, but most are Python related. Even though it was not so long ago, Tensor flow has been used in various fields.

Both the regular and GPU-accelerated versions are available. The regular version can run on any computer, and the GPU-accelerated version works much faster because it uses GPGPU to perform large-scale operations quickly. However, since it uses NVIDIA's GPGPU language, CUDA, it cannot be used without an NVIDIA graphics card.

2.3. CNN (Convolutional Neural Network) Algorithm

Convolutional Neural Network is a type of multi-layer perceptron designed for minimal preprocessing. CNN is composed of one or several layers of convolutional hierarchy and general artificial neural networks layered on top of it and utilizes additional weighting and pooling layers. This structure allows CNN to fully utilize the input data of the two-dimensional structure. Compared to other deep learning structures, CNN offers good performance in both video and audio applications. CNN can also be trained through standard back propagation. CNN is easier to train than other feedforward artificial neural network techniques and has the advantage of using fewer parameters.

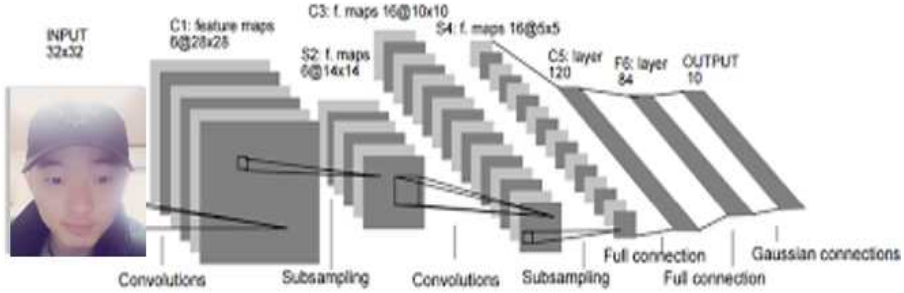

Figure 1. CNN Algorithm Structure

Recently, Convolutional Deep Belief Network(CDBN) has been developed in deep learning, which is structurally very similar to the existing CNN. It can take advantage of the two-dimensional structure of a Figure 1 and at the same time, it can take advantage of pretraining in Deep Belief Network(DBN). CDBN provides a generic structure that can be used for a variety of image and signal processing techniques and is used in several benchmark results for standard image data such as CIFAR.

2.4. Face Recognition

Face Detection is a field of computer vision, which is a technique that tells the location of a face in an image. The algorithmic basic structure of face detection is defined by Rowley, Baluja, and Kandae (1995After generating a pyramid image to detect faces of various sizes, it is moved by one pixel and whether the corresponding area of a specific size (for example, 20x20 pixels) is a face or not is detected by a Neural Network, Adaboost, Support Vector Machine and so on.

2.5. Harr

Harr is a feature that uses the brightness difference between regions in the image, and there are various basic features, and it is a method to extract the features of the object by combining these features with various sizes and positions. The feature value for the basic feature is calculated by subtracting the brightness sum of the black portion from the brightness sum of the image pixels corresponding to the white portion of the feature. The identification of the object using the feature uses whether the brightness difference of the calculated region is larger or smaller than the threshold value given to the feature. Instead of using one feature, many features are used in combination. Even if the same kind of feature is considered as a different feature according to the position and size in the object, it is possible to combine features close to infinity. In the case of a human face, hair, eyebrow, pupil, and lips have a characteristic difference in brightness, which makes them suitable for applying the Harr feature. The Harr feature basically maintains the geometric information of the object and has characteristics that can cover the shape change of the object and some position changes to some extent. However, there is a disadvantage in that it is difficult to detect when the object rotates due to the contrast change of the image, the change of the image brightness due to the change of the direction of the light source. The Harr feature improves the detection accuracy by using various preprocessing methods because it has the disadvantage of detecting faces with similar features.

3. Body

There are various methods to read people's mind or feelings which is very important for human life. Communication has a crucial effect on many field in human life such as business, commerce, and relationship. Emotion is one of the best way to understand humans' individual feelings. Without a single word spoken, quite exact feelings or moods of a person can be understood by just looking at the person's facial expression. I have studied the method to analyze and predict emotion through a camera. The model presented in this paper is a model that extracts symbols according to the user's facial expressions using the camera installed in Android intelligent device. First of all, I made a model that predicts the emotion from a image with a high accuracy. With the developed model, emotion prediction by a camera was processed.

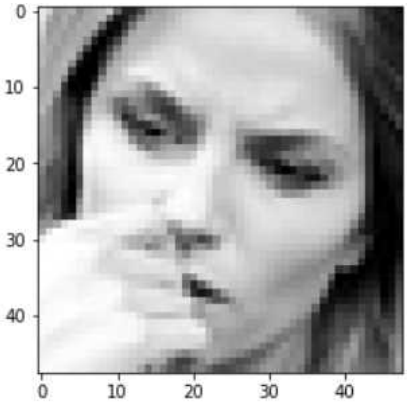

Figure 2. Data for Training Model

Figure 2 is an image referencing emotion of sorrow. Numerous amount of these images are used to train the CNN model. I used 35887 images and 7 facial expressions (anger, disgust, horror, joy, sorrow, surprise, neutral) of the photo data provided at Kaggle website and applied the CNN algorithm with Tensor flow to create a face recognition model. I used 16GB Intel i7-7700HQ CPU 2.80GHz with laptop performance and a high-level programming language called Python to train the facial expression recognition model.

Figure 3. CNN Model Structure

The figure 3 shows the structure of CNN model. The parameters of the convolutional layer consist of a series of learnable filters. It is common to place a pooling layer periodically in the middle of the convolution layers. What the pooling layer does is reduce the spatial size of the representation to reduce the number of parameters and computation in the network. Many kinds of normalization layers have been proposed for real brain inhibition mechanisms and so on. To carry out the research of this paper, I made a facial recognition model with Python code. The structure of CNN in this paper is based on the paper of Deep Learning using Linear Support Vector Machines. The CNN algorithm code was written using the tensor flow library and the model was learned. When the image was input from seven facial expressions (anger, dislike, horror, joy, sorrow, surprise, neutral) of a person, the accuracy of how accurately the expression was predicted through the learned

CNN model was output.

The Python code in Table 1 is the Pseudo code of the tensor flow CNN model in Figure 3 and confirms the learning and accuracy of the model created with fer 2013 data.

Table 1. Tensor flow CNN Model Pseudo Code

Algorithm 1: Tensorflow CNN Model Pseudo Code

Input : image, resource for training and testing
Train_label, label of each different emotion for training
Test_label, label of each different emotion of testing
 	
Output : The accuracy of trained model
 	
FUNCTION Training Phase
 	
for (all training labels)
if (Train_label <- outputclass)
Train_label = 1;
Else
Train_label = -1;
for all testing labels
if (test label <- outputclass)
test label = 1;
Else
test label = -1;
for all training points
activation = dot product of weight vector and training image
if(activation > 0)
CurrentOutput = 1;
else
CurrentOutput = -1;
if(CurrentOutput != Y(m,1))
W = W + dot product of Train_label and Train Image
 	
FUNCTION Testing Phase
 	
for all testing phase
TestImageActivation = dot product of Weight vector and Test Image
if(TestImagesActivation > 0)
TestOutput = 1;
Else
TestOutput = -1;
if(TestOutput <- Test_label(z,1))
accuracy = accuracy + 1;

4. Results

The research in this paper uses ‘fer2013’ dataset provided by Kaggle and applied 35887 pictures to CNN algorithm and trained emotion recognition model.

```
training_accuracy / validation_accuracy => 0.12 / 0.18 for step 98600
training_accuracy / validation_accuracy => 0.12 / 0.18 for step 98700
training_accuracy / validation_accuracy => 0.14 / 0.18 for step 98800
training_accuracy / validation_accuracy => 0.14 / 0.18 for step 98900
training_accuracy / validation_accuracy => 0.20 / 0.18 for step 99000
training_accuracy / validation_accuracy => 0.12 / 0.18 for step 99100
training_accuracy / validation_accuracy => 0.14 / 0.18 for step 99200
training_accuracy / validation_accuracy => 0.14 / 0.18 for step 99300
training_accuracy / validation_accuracy => 0.14 / 0.18 for step 99400
training_accuracy / validation_accuracy => 0.12 / 0.18 for step 99500
training_accuracy / validation_accuracy => 0.16 / 0.18 for step 99600
training_accuracy / validation_accuracy => 0.16 / 0.18 for step 99700
training_accuracy / validation_accuracy => 0.14 / 0.18 for step 99800
training_accuracy / validation_accuracy => 0.14 / 0.18 for step 99900
training_accuracy / validation_accuracy => 0.10 / 0.18 for step 99999
```

Figure 4. Training Accuracy

As shown in figure 4 shows the training accuracy and the validation accuracy of emotion recognition. 100000 repetition learning was performed which took around 35 hours, and the highest model training accuracy was 66%.

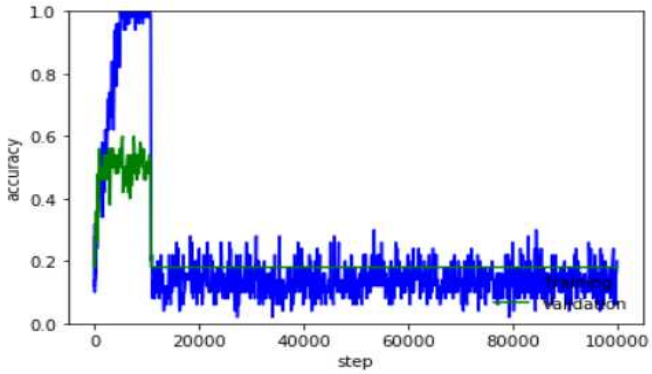

Figure 5. Accuracy Graph

As shown in Figure 5 100,000 repetition learned data accuracy is shown as a graph and shows 66% of accuracy. By looking at the graph, longer the model gets trained the accuracy decreases. I looked up for the number of steps which showed the highest accuracy of validation of emotion recognition. Around 7500 to 10700 steps of training showed 66% of accuracy.

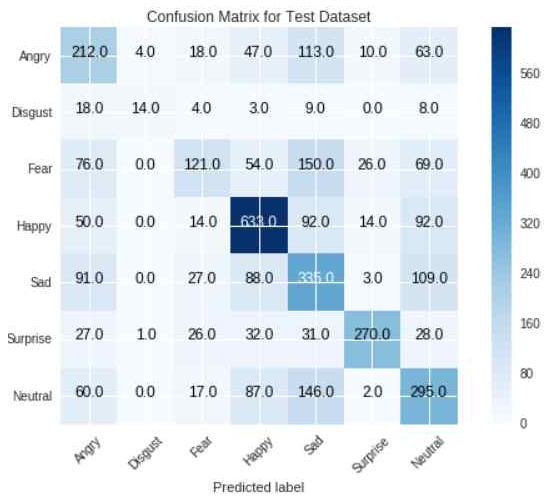

Figure 6. Confusion Matrix

The figure 6 is the confusion matrix for a dataset to test the model. The straight bar represents how accurate the expression is recognized, the higher the better. As shown on the confusion matrix, each expression recognition presents good score.

5. Conclusion

In this research a service to extract the similar symbols as facial expressions through users' facial recognition is proposed. By inputting a picture of the users' facial expression, the model that I built will consider which expression is being made. 35 hours of training the model gave me the conclusion of a quite high expression recognition accuracy. Because of learning various pictures with CNN algorithm, I made a model that recognizes facial expressions with accuracy of 66%. Based on this model, it will be useful in various fields such as creating an app or a situation when recognizing facial expression is necessary by applying it to Android smart device. Also, studying higher accuracy based on this model will improve the performance of the face recognition model.

Since the symbols appear immediately when the camera recognizes facial expressions without having to search for desired symbols, the service will be very convenient to use symbols when people send and receive messages. There is no need to search for symbols among many symbols, and it is necessary to increase the accuracy by using an algorithm that is more suitable for facial recognition because the research in the field of deep learning is increasing.

6. Acknowledgement

"This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (2017R1D1A1B03034411)"

7. Reference

[1] Boyle, G. J. (1998). Review of Arousal Seeking Tendency Scale. In J. C. Impara & B. S. Plake (Eds.), *The thirteenth mental measurements yearbook* (pp. 49-50). Lincoln, NE: Buros Institute of Mental Measurements.

[2] Darwin, C. (1871). *The descent of man, and selection in relation to sex.* London: John Murray.

[3] Ekman, P., & Friesen, W. V. (1977). *The Facial Action Coding System (FACS): A Technique for the Measurement of Facial Action.* Palo Alto: Consulting Psychologists Press.

[4] Mehrabian, A. (1971). *Silent Messages* (1st ed.). Belmont, CA: Wadsworth.

[5] NAMU (2017). *Tenser Flow*. Retrieved May 22, 2017. from https://namu.wiki/w/tenserflow

[6] Pervaiz, A. Z. (2010). Real Time Face Recognition System Based on EBGM Framework. In Computer Modelling and Simulation (UKSim),12th International Conference on 2010, pp.262-266.

[7] Rowley, H. A. Baluja, S., & Kanade, T. (1995). Human face detection in visual scenes. CMUCS-95-158R, Carnegie Mellon University, Retrieved November 22, 1995. from http://www.cs.cmu.edu/˜har/faces.html.

[8] Wikipedia (2017a*). Convolutional Neural Network*. Retrieved May 22, 2017. from https://ko.wikipedia.org/wiki/%EB%94%A5_%EB%9F%AC%EB%8B%9D#.ED.95.A9.EC.84.B1.EA.B3.B1_.EC.8B.A0.EA.B2.BD.EB.A7.9D.28Convolutional_Neural_Network.2C_CNN.29

[9] Wikipedia (2017b*). Deep Learning*. Retrieved May 22, 2017. from https://ko.wikipedia.org/wiki/딥_러닝

[10] Wikipedia (2017c*). Deep Learning*. Retrieved May 22, from https://ko.wikipedia.org/wiki /%EC%96%BC% EA%B5%B4_%EA%B2%80%EC%B6%9C

[11] http://cs231n.stanford.edu/reports/2016/pdfs/022_Report.pdf
[12] http://aikorea.org/cs231n/convolutional-networks/#norm
[13] Gil Levi, Tal Hassner Emotion Recognition in the Wild via Convolutional Neural Networks and Mapped Binary Patterns
[14] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. Int. J. Comput. Vision, 40(2):99–121, 2000.
[15] M. J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, and J. Budynek. The japanese female facial expression (jaffe) database, 1998.
[16] I. Borg and P. J. Groenen. Modern multidimensional scaling: Theory and applications. Springer Science & Business Media, 2005.
[17] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman. Return of the devil in the details: Delving deep into convolutional nets. arXiv preprint arXiv:1405.3531, 2014.