Feature Acquisition From Facial Expression Image Using Convolutional Neural Networks

Taiki Nishime
Graduate school of Information Engineering, University of Ryukyus
Nishihara, Okinawa, Japan

Satoshi Endo, Koji Yamada, Naruaki Toma, Yuhei Akamine
School of Information Engineering, University of Ryukyus
Nishihara, Okinawa, Japan
taiki_one@eva.ie.u-ryukyu.ac.jp, endo@ie.u-ryukyu.ac.jp, koji@ie.u-ryukyu.ac.jp
tnal@ie.u-ryukyu.ac.jp, yuhei@ie.u-ryukyu.ac.jp

Abstract

In this study, we carried out the facial expression recognition from facial expression dataset using Convolutional Neural Networks (CNN). In addition, we analyzed intermediate outputs of CNN. As a result, we have obtained a emotion recognition score of about 58%; two emotions (Happiness, Surprise) recognition score was about 70%. We also confirmed that specific unit of intermediate layer have learned the feature about Happiness. This paper details these experiments and investigations regarding the influence of CNN learning from facial expression.

Keywords: facial expression recognition, convolutional neural networks, deep learning, feature learning

1. Introduction

Facial expression recognition is important to non verbal communications among the people. Now, opportunities to communicate using either voice or text is increasing because of developing mobile phones and Internets. Thus, it is considered that communication method via some devises has increased than face to face communication. "UNMASKING THE FACE" by Paul Ekman and W.V. Friesen [1] described that facial expressions have a closely connection with the emotions. For this reason, It is natural to think that we can recognized your happiness if you smiling. Mainstream approaches in facial expression recognition use Facial Action Coding System (FACS) labels. FACS was designed to help facial expression recognition with resolve each expression into several Action Units (AUs). FACS labels approaches need to learn from FACS manual and test. As of now, FACS label can only be given by experts or trained individuals. As a results, not everybody using easily FACS labels to facial expression recognition.

The previous studies on facial expression recognition can be classified into two categories: FACS based method [2] or feature learning method [3]. In FACS based method, they first extracted feature from AUs, then they recognized facial expression from facial images using a these extracted feature and Support Vector Machine. In contrast to feature learning method [3], recognized facial expression using Convolutional Neural Networks (CNN) [4].

© The 2016 International Conference on Artificial Life and Robotics (ICAROB 2016), Jan. 29-31, Okinawa Convention Center, Okinawa, Japan

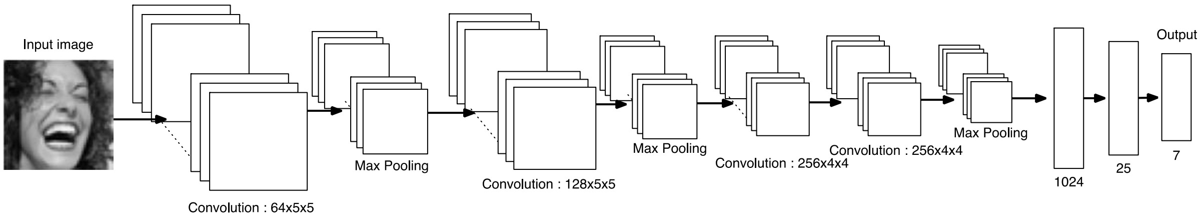

Fig 1. CNN structure: consist of three convolutional layers, three pooling layers, two fully connected layers, and output layer

But, this study is not discussed in CNN model that has finished learning, and there was no argument about learning the feature of CNN.

In this study, firstly, we carried out facial expression recognition using CNN. In addition, we analyze the intermediate outputs of CNN that obtained from facial expression images.

2. Convolutional Neural Networks

Convolutional Neural Networks is a type of feed-forward artificial neural networks that consist of convolutional layers, pooling layers, fully connected layers and output layer. Convolutional layers compute product-sum of image and weight. Pooling layers compute the max value of a particular feature over a region of the image. These convolutional layer and pooling layer were repeated for every such layer. Fully connected layers applied at the end of or after repeated each layer. Fully connected layers is the same as regular multilayer perceptron. By propagating these each layer, CNN was feature extracted from input images.

3. Experiments

3.1. *FER-2013 Dataset*

We have selected Facial Expression Recognition 2013 (FER-2013) dataset [6]. FER-2013 was created by Pierre Luc Carrier. This dataset was created using the Google image search API to search for images of faces that match a set of 184 emotion-related keywords like “blissful”, “enraged” etc. Each images included dataset is cropped around a face, and cropped images were then resized to 48x48 pixels and converted to grayscale. Table. 1. present the details of the dataset. Facial expression We focused on are Angry(An), Disgust(Di), Fear(Fe), Happiness(Ha), Sadness(Sa), Surprise(Su) and Neutral(Ne).

	An	Di	Fe	Ha	Sa	Su	Ne	Total
Training	3993	436	4096	7212	4828	3171	4692	28698
Test	466	56	496	895	653	415	607	3588

Table 1. Detail of FER-2013 dataset

3.2. *Preprocessing*

We preprocess the data using Global Contrast Normalization (GCN) and ZCA whitening [5]. In GCN, subtract by mean and divide by dispersion for each dataset images. By GCN preprocessing, if the value range of the input is normalized from -2 to 2, and can be aligned to that range, even if there is a different axis scales. Natural image is characterized by strong correlation with neighboring pixels. ZCA whitening has function to erase such correlation.

3.3. *CNN settings*

Fig 1. shows CNN model that used in this experiment. Arrows in Fig 1. show weights, and number of under each boxes show unit number. The number of input units setting to same as number of input image pixels. The number of output unit setting to same as number facial expression classes. As the facial expression recognition result, using the maximum value in output layer units.

3.4. *Result*

The results of this experiments shown in Table. 2. We have obtained an recognition score 57.02%; Happiness and Surprise facial expression recognition score was about 70%. In contrast, Fear, Sad and Neutral score was below 52%. Also these recognitions from only image data is seemed to be difficult. We have obtained an Disgust recognition score 0% because of Disgust data was less then other facial expression data.

	An	Di	Fe	Ha	Sa	Su	Ne
An	45.92	0	11.58	7.51	21.45	2.57	10.94
Di	37.5	0	14.28	5.35	25	1.78	16.07
Fe	10.08	0	37.9	5.84	26.2	7.25	12.7
Ha	4.24	0	2.79	76.2	6.92	2.23	7.59
Sa	10.71	0	13.32	7.65	51.14	1.68	15.46
Su	3.85	0	11.32	4.09	3.37	72.04	5.3
Ne	7.9	0	7.9	8.56	23.39	1.64	50.57

Table 2. Confusion matrix of facial expression recognition

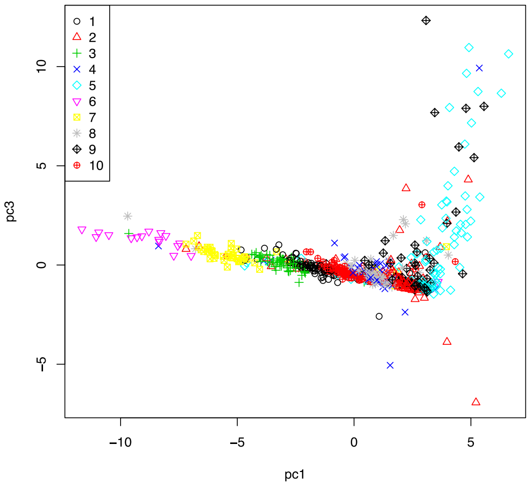

Fig 2. Visualized the 25-dimensional intermediate output of recognized Happiness. The 25-dimensional output was compressed to 2-dimension output by Principal Components Analysis. Plotting first and third principal components. 48.959% is contribution rate, it is show that 48.959% information remained by PCA. Labeled by k-means.

3.5. *Analyzing the intermediate outputs*

As mentioned before, CNN was extracted feature at each layers, and these result of feature extraction was influence on the recognition. Thus, it seems that recognition accuracy is higher, feature extraction is better result. From this reason, we focused on Happiness, and analyzing the intermediate output values. By this analyzing, we reveal that extracted feature from Happiness data.

In this experiment, there are image that recognized 682 Happiness images in 895 labeled images. We are focused on the Happiness images and intermediate output values previous layer for the output layer. Result of compressed intermediate output to two dimension is shown in Fig 2. From Fig 2, It can be seen that some clusters are overlapped with each other like cluster 1 and cluster 10, but independent cluster like clusters 5 and cluster 6 is also exists. As a result, we investigated the difference in the distinguishable clusters 5 and cluster 6, we examine the extracted features from the result of this investigation. Examination method is simple, we checked that whether the intermediate outputs connoting a pattern.

All of 25 units intermediate output of cluster 5 and cluster 6 are shown in Fig 3, and also examples of image and intermediate outputs include in cluster 5 and cluster 6 are shown in Fig 4. As a results, It can be seen that the intermediate outputs of cluster 6 was bigger than cluster 5. As far as we have confirmed detail of outputs in cluster 6, all data was Happiness face, and is commonly mouth was open, as shown in Fig 4. On the other hand, cluster 5 included Happiness face both opened mouth and closed mouth. Most of cluster 5 data was Happiness face with closed mouth. We focus on the intermediate output of each unit in data of cluster 6 (Fig 3), for example, it can be seen that the output value of 21th unit is large. Also in Fig 4, the Happiness face with opened mouth and closed mouth were difference in 21th unit output value. In Fig 4, it is selected some example from cluster 5 and cluster 6. It was also observed that this result of 21th unit output value was the same tendency as Happiness face with opened mouth in other cluster. In these results, the 21th unit output have closure connected in shape of mouth open. We conclude that

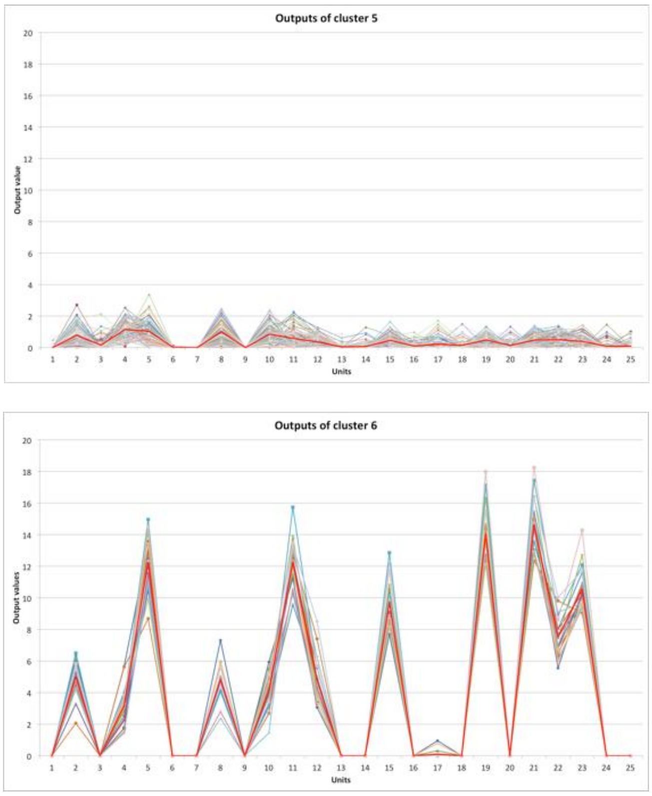

Fig 3. Intermidiate outputs of all 25 units. Upper line graph is shown intermediate output of cluster 5, lower line graph is shown intermediate output of cluster 6. Red bodal line is average value of 25 units. Other lines represent one input reconstructed Happiness facial image.

CNN have learned the feature of facial expression about mouth.

3.6. *Future work*

In this paper, We carried out the emotion recognition from facial expression image using a CNN. As a result, we have obtained an average emotion recognition score of 58%; two emotions (Happiness, Surprise) recognition score was about 70%. From the result of the analyzing intermediate outputs of CNN, we confirmed specific intermediate outputs was likely to be reflected in shape of mouth. In this experiments, though using the simple method that analyze the intermediate outputs pattern, it is considered to effective method that analyzing the extracted feature from input data. In the future work, we will try to analyze larger intermediate outputs than in this experiment. And we revealed that method of feature extracted of deep layer.

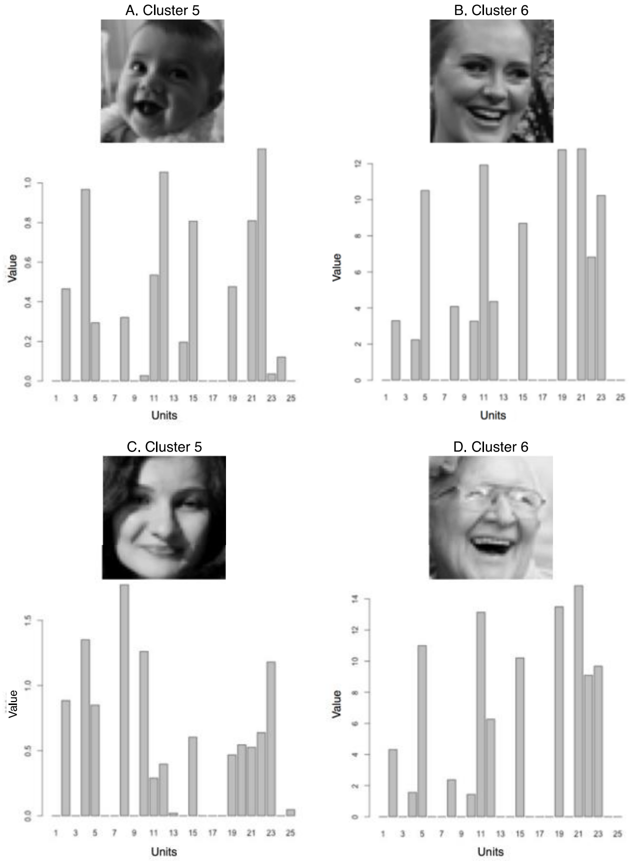

Fig 4. Example of All of 25 units intermediate output of cluster 5 and 6.

References

[1] Paul Ekman, W.V. Frisen, “UNMASKING THE FACE” (1975)

[2] Hiroki NOMIYA, Teruhisa HOCHIN, “Facial Expression Recognition using Feature Extraction based on Estimation of Useful Features” (2011)

[3] VICTOR-EMIL NEAGOE, ANDREI-PETRU BRAR, NICUSEBE, PAUL ROBITU, “A Deep Learning Approach for Subject Independent Emotion Recognition from Facial Expressions”, Recent Advances in Image, Audio and Signal Processing, pp.93-98 (2013)

[4] “Convolutional Neural Networks (LeNet) – DeepLearning 0.1 documentation”, LISA Lab. (2013)

[5] Alex Krizhevsky, “Learning Multiple Layers of Features from Tiny Images”, 2009.

[6] Goodfellow, Ian J, .et al. “Challenges in Representation Learning: A report on three machine learning contest” Neural Information Processing. Springer Berlin Heidelberg, 2013.