Feature Learning from Facial Expression Images Using Convolutional Neural Networks

Taiki Nishime[1†] Satoshi Endo[2‡] Naruaki Toma[3‡] Koji Yamada[4‡] Yuhei Akamine[5‡]
[†]Graduate school of Information Engineering, University of The Ryukyus
[‡]School of Information Engineering, University of the Ryukyus
1 Senbaru, Nishihara, Okinawa 903-0213, Japan
E-mail : [1]taiki_one@eva.ie.u-ryukyu.ac.jp, {[2]endo, [3]tnal, [4]koji, [5]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 recognition accuracy of about 58%; two classes (Happiness, Surprise) recognition score was about 70%. We also confirmed that CNN has learned the feature to recognize facial expression from the images. This paper details these experiments and investigations regarding the feature 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 voice and text is increasing because of developing mobile phones and Internets. Thus, it is considered that indirect communication via some devices has increased more than direct communication. "UNMASKING THE FACE" by Paul Ekman and W.V. Friesen described that facial expressions is a close connection with the emotions[1]. As the reason, it is natural to think that we can recognized your happiness if you smiling. Many 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 training. As of now, FACS label can only be given by experts or trained individuals. As a results, The only experts using easily FACS labels to facial expression recognition.

The previous studies on facial expression recognition can be classified into two categories; the FACS based method[2] or the feature learning method[3]. In the FACS based method, they first extracted feature from AUs, then they recognized facial expression from facial images using these extracted feature and Support Vector Machine. In contrast the feature learning method, most of study about recognized facial expression is using Convolutional Neural Networks (CNN)[4]. But, these 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, we carried out facial expression recognition using CNN. In addition, we analyze the feature that was learned by CNN from facial expression.

2. Convolutional Neural Networks

Convolutional Neural Networks (CNN) 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 these 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

In this section, we explain about preprocessing, CNN settings and result of facial expression recognition result.

3.1 FER-2013 Dataset

We have selected Facial Expression Recognition 2013 dataset [5](FER-2013 dataset). FER-2013 dataset 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 pixel and converted to grayscale. Table 1 present the details of the dataset. Facial expression we focused on Anger(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	4097	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[6]. In GCN, subtract by mean and divide by dispersion for each dataset images. By GCN preprocessing, 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.

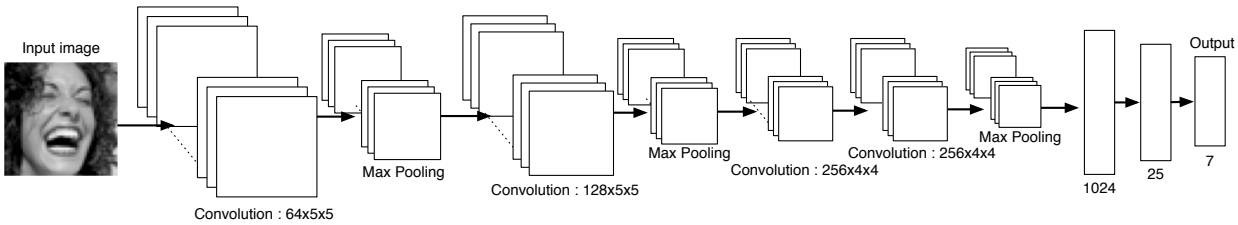

Figure 1. CNN structure: input, convolutional, max-pooling, fully connection, classification layers

3.3 CNN settings

Fig 1 shows CNN model that used in this experiment. Arrows in Fig 1. is shown weights, and number of under each boxes is shown unit number (number of feature map x height size x width size). The number of input units setting to same as number of input image pixels. The number of output unit setting to same as the 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.

		Corrected class						
		An	Di	Fe	Ha	Sa	Su	Ne
Predicted class	An	45.92	0.0	11.58	7.51	21.45	2.57	10.94
	Di	37.5	0.0	14.28	5.35	25.0	1.78	16.07
	Fe	10.08	0.0	37.9	5.84	26.2	7.25	12.7
	Ha	4.24	0.0	2.79	76.2	6.92	2.23	7.59
	Sa	10.71	0.0	13.32	7.65	51.14	1.68	15.46
	Su	3.85	0.0	11.32	4.09	3.37	72.04	5.3
	Ne	7.9	0.0	7.9	8.56	23.39	1.64	50.57

Table 2. Confusion matrix: model performance (in percent)

4. Feature Analysis

In this section, we discuss the features obtained from facial expression images.

4.1 Analysis method

The analysis method is divided into three steps. In this analysis, we use correctly classified test images by CNN.

1. Divide the input images into 16 analysis area.(Fig 2-1)
2. Select analysis area from 16 areas and mask each pixel of select area. (As mask processing, we initialized each pixels of analysis area to 0.0, 0.2, 0.4, 0.6, 0.8 and 1.0)(Fig 2-2)
3. Using masked image as input of CNN, we examine the change of output units that represent each facial expressions.(Fig 2-3)

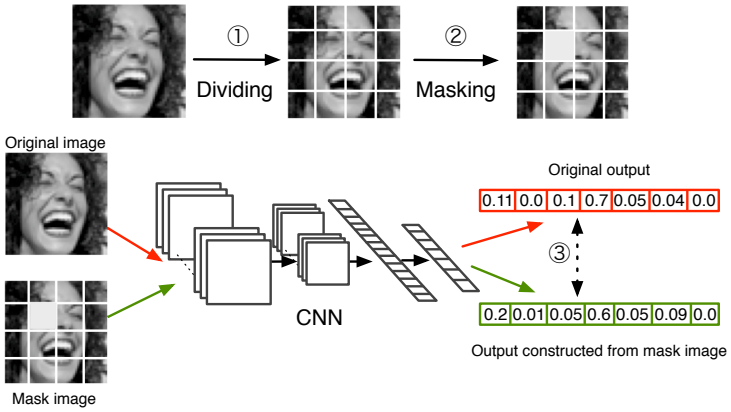

Figure 2. Overview of analysis method. Examine the change of output in before and after the mask processing.

Note that when examining the change of output, we focus on the only output value representing specific facial expression. In below graphs, each line also shows the mean output value representing the facial expression.

4.2 Result of feature analysis

First, we describe the result of happiness that was obtained highest accuracy. Fig 3 shows example of happiness analyzed in the above figure. To easily describe, we assign section number to each analysis area.

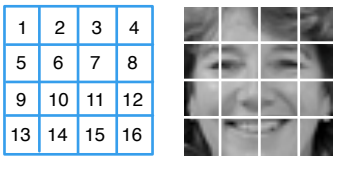

Figure 3. Example of analysis image.

Fig 5, 6, 7, 8 and 9 shows the analysis results. These line graph is shown change in outputs unit representing each class.

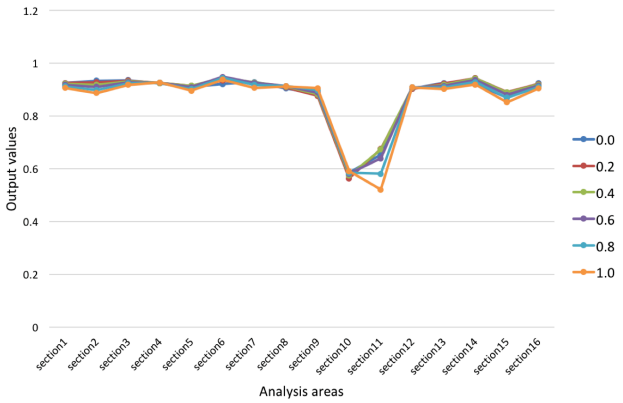

Figure 4. Change to Happiness class value

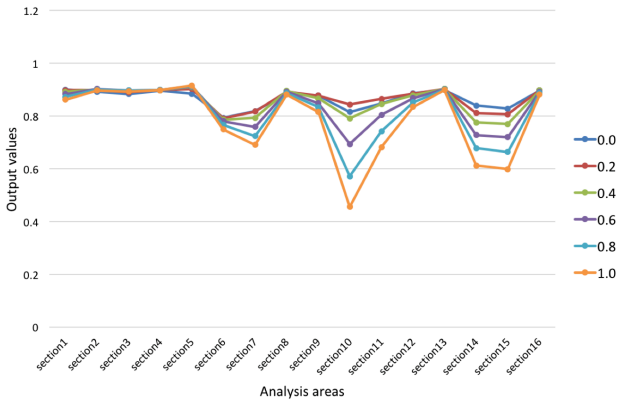

Figure 5. Change to Surprise class value

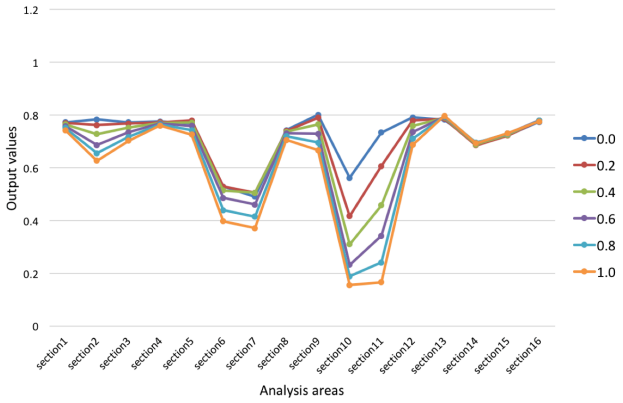

Figure 6. Change to Anger class value

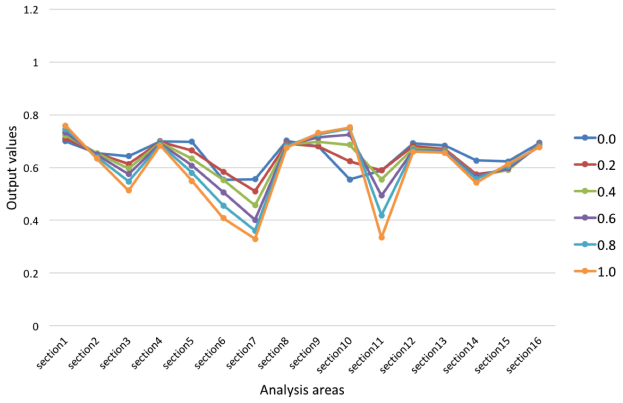

Figure 7. Change to Sadness class value

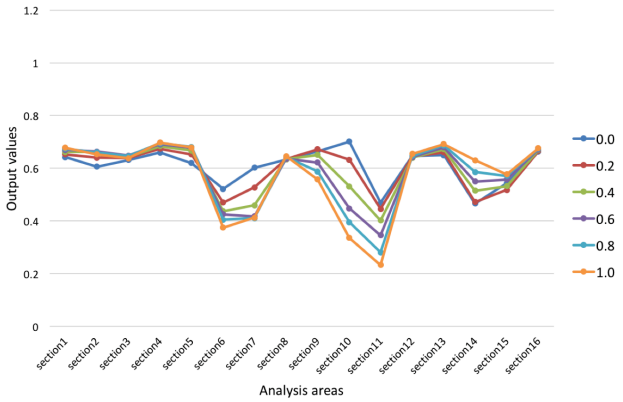

Figure 8. Change to Fear class value

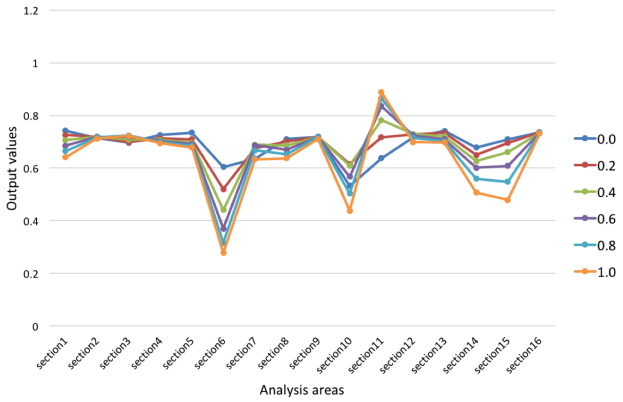

Figure 9. Change to Neutral class value

In Fig 4, horizontal axis shows each analysis area and vertical one shows the output unit mean values. From this graph, it can be seen that value of output unit is lower than other units when analysis area is 10 and 11. As shown in Fig 4, the mouth of the most analyzed images are located in area 10 and 11, these areas have closely connected with recognition of Happiness. From these results, recognition of the Happiness by CNN is conceivable that features of the mouth are important.

Next, we compare Happiness with other class. Fig 5, 6, 7, 8 and 9 are shows analysis results. Each line graph shows Surprise, Anger, Sadness, Fear and Neutral. From these results, it is evident that each class value changed by influence of section 6, 7, 10 and 11. The 4 classes (Anger, Sadness, Fear, Neutral) is different from Happiness and Surprise, they are

also influenced by section 6 and 7. Therefore, it is assumed that 4 class is difficult to recognize by using only feature of mouth. Considering each class accuracy in Table 2, it is consider that CNN haven't learned difference between the eyes and mouth of each facial expression.

4.3 Verification the result of analysis

In this section, we perform experiments to verify the result of feature analysis. In previous section, we confirmed that CNN has learned feature on section 6, 7, 10 and 11. From this result, Using the images masked except section 10, 11 and 6, 7, 10, 11, we examine the change of recognition result. The example of masked images is shown Fig 10.

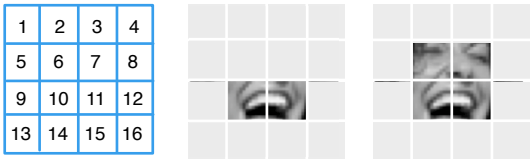

Figure 10. Example of masked images.

The result of experiments shown in Table 3, 4. Table 3 is shows the result of when masked except section 10, 11 (Fig 10-center). Table 4 is also shows the result of when masked except section 6, 7, 10, 11 (Fig 10-right). From these results, compared to the result of Table 2, this accuracy is down, but the accuracy of Happiness is over the 50%. From these results, we confirmed that the CNN need feature of mouth to recognition of Happiness and it was learning the feature of mouth.

		Corrected class						
		An	Di	Fe	Ha	Sa	Su	Ne
Predicted class	An	6.22	0.0	4.29	32.83	8.15	1.28	47.21
	Di	0.0	0.0	5.35	26.78	8.92	1.78	57.14
	Fe	3.22	0.0	5.84	32.66	12.7	1.41	44.15
	Ha	2.68	0.0	2.56	59.77	5.36	0.78	28.82
	Sa	2.6	0.0	5.51	34.91	10.87	1.22	44.86
	Su	5.3	0.0	5.7	24.33	3.85	5.54	55.18
	Ne	2.14	0.0	5.43	26.68	7.9	1.15	56.67

Table 3. Confusion matrix: model performance when masked section 10 and 11 (in percent)

		Corrected class						
		An	Di	Fe	Ha	Sa	Su	Ne
Predicted class	An	32.4	0.0	3.64	38.19	10.94	0.64	14.16
	Di	25.0	0.0	5.35	46.42	10.71	1.78	10.71
	Fe	12.29	0.0	12.29	34.87	17.94	2.01	20.56
	Ha	4.13	0.0	1.67	74.18	6.14	0.22	13.63
	Sa	12.55	0.0	6.12	34.15	27.41	1.22	18.52
	Su	9.15	0.0	8.67	25.78	5.78	25.06	25.54
	Ne	9.39	0.0	4.28	30.47	13.5	1.15	41.18

Table 4. Confusion matrix: model performance when masked section 6, 7, 10 and 11 (in percent)

5. Conclusion

In this paper, We carried out the facial expression recognition from facial expression image using a CNN. As a result, we have obtained an average facial expression recognition score of 57%; two emotions (Happiness, Surprise) recognition score was about 70%. We were evaluated as the feature learned by CNN from input layer. The result of feature analysis and verification suggest that CNN learned the feature representing facial expression, such as mouth and eyes.

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

[6] Alex Krizhevsky, "Learning Multiple Layers of Features from Tiny Images", 2009.