The 31st International Technical Conference on Circuits/Systems,
Computers and Communications (ITC-CSCC 2016)

The Discrimination method for Favorability of Facial expression using Convolutional Neural Networks

Masataka Itosu[1] Satoshi Endo[2] Naruaki Toma [2] Koji Yamada [2] Yuhei Akamine [2]
[1]Graduate School of Information Engineering, University of The Ryukyus,
[1] Senbaru, Nishihara, Okinawa 903-0213, Japan
[2]School of Information Engineering, University of The Ryukyus,
E-mail : [1]k168586@ie.u-ryukyu.ac.jp, [2]{endo,tnal,koji,yuhei}@ie.u-ryukyu.ac.jp

Abstract: Visual information like gesture and facial expression is important factor in human communication. Study on the facial expression and impression pointed out that the beholder feels difference of the impression from subtle difference of the facial expression even as the same person. From these results, we considered that it is possible to subdivide the facial expression by human impression. In this paper, we develop the Convolutional Neural Network(CNN) model that discriminate tha favorability of facial expression. We focus on the smile that it is one of the typical facial expressions. The reason for choosing the smile from various facial expressions, people easily identify the feature of smile among the facial expressions. In general, smile is high recognition rate in facial expression recognition. In this study, we deal the 2 classification problem that is divided into good smile class and bad smile class. In addition, we analyze the face parts that is required for discriminate the smile by CNN. We clarify the important face parts for favorability of facial expression by performing the experiment. As a result, we obtained the smile impression recognition rate of certain level performance and this model estimate the good and bad smile by eyes and mouth.

Keywords—**Convolutional Neural Network, Facial expression recognition, Favorability of facial expression,**

1. Introduction

According to the Mehrabian's rule[1], there are three elements for message transmission in human communication such as language information, auditory information and visual information. In these elements, the visual information occupies 55% for message transmission. This result suggest that visual information is most important factor in the interpersonal communication. Inoue et al.[2] described about, smile is a good impression for beholders and they may feel different impressions by open mouth smile and close mouth smile. It is depend on their age or gender. Iguchi et al.[3] and Sugawara et al.[4] described about the relation between the impression and smile featuer that is shape and position of the smile face parts. The paper concluded that the position and shape of the smile parts influence the smile impression. From these results, we considered that it is possible to subdivide the facial expression by human impression. For example, the smile is divided into good smile and bad smile by impression(Fig.1).

Recently, Convolutional Neural Network(CNN) makes top-class performance in image recognition field. For instance, CNN achieved overwhelming accuracy in the big data challenge of Imagenet Large Scale Visual Recognition Challenge(ILSVRC) 2012. As the reason that CNN achieved this result, it is considered to extract features automatically. In this paper, we discuss about the discrimination method for favorability of facial expression using CNN. We focus on the smile in this study, because smile is easily identified to machine and people among the facial expressions.

We proposed the following discrimination method:

Step.1 :Prepare for collect the smile image.
Step.2 :Collect the smile image from face image group.
Step.3 :Give a score to smile image group.
Step.4 :Create the dataset of smile impression.
Step.5 :Training the CNN model using dataset of smile impression.

In step 5, we implement the 2 classification problem to CNN model and define the Good and Bad label. As another experiment, we analyze the face parts that is required for discriminate the smile by CNN. We carried out the experiment to clarify the important face parts for favorability of facial expression.

Figure 1. Defined the Good and Bad smile.

2. Related work

Research of the facial expression recognition has the long history and it is one of the typical image recognition problem. One of the facial expression recognition research using Deep Neural Network(DNN) by Nishime et al.[5], used Auto-encoder and Denoising Auto-encoder. It was obtained an emotion recognition rate about 68%. According to the paper of Victor et al[6]., they applied Support Vector Machine(SVM) and CNN as the classifier system for facial expression recognition problem. In their paper, CNN achieved recognition rate about 65% and this rate means about 6% improved against SVM. But in their paper, they didn't discuss about the subtle difference of the same facial expressions such as good smile and bad smile. Accordingly, we carried out the experiments of learning the subtle difference of the same facial expressions.

3. Proposed method

3.1 Convolutional Neural Network(CNN)

Artificial neural network is modeled on biological neuron in the human brain and deep learning is constituted by further stacking the previous neural network. CNN is one of the deep learning that is modeled on visual cortex. Generally CNN is composed of convolutional layer, pooling layer and fully connected layer. Convolutional layer outputs the feature map by filter processing to input image. Pooling layer compute the local area of feature map and it provides a form of translation invariance. Unit of these layers are connected locally and sparsely. Fully connected layer is located after several these layers. In this study, we use the CNN network shown Fig.2 as the CNN A and B. At the experiment, we use this CNN model that it won the championship by ILSVRC2012[7].

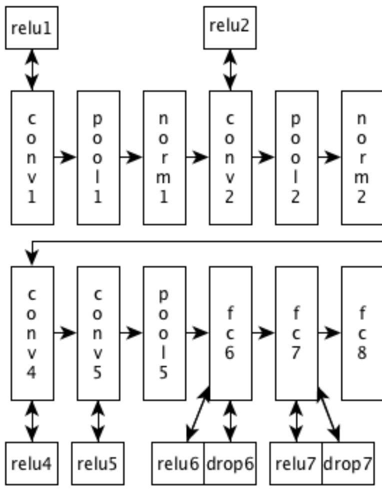

Figure 2. Structure of the CNN network.

3.2 Proposed algorithm

We proposed following methods to training the CNN model that can discriminate for favorability of smile as shown in Fig.3. We proposed the method that is divided into 5 steps. In step 1, labeled facial expression dataset is divided into smile and other facial expressions, CNN A is trained the smile and other facial expressions using the dataset. In step 2, the non labeled data group is identified by CNN A and we obtain the data that is estimated smile. In step 3, we take a survey to give a score to smile face image data group that create with step 2. In step 4, we define the high threshold and low threshold to create the smile impression datasets. When score of smile image is high threshold or more, smile image is assigned to the Good label. When the score is low threshold or less, it is assigned to the Bad label. In this way, we obtain the high quality dataset that contains good impression and bad impression smile data. In step 5, training the CNN B that identifies the good smile and bad smile, using dataset of step 4.

Step.1 :Training the CNN A that identifies smile and other facial expressions.
Step.2 :Identify the non labeled data group to extraction the smile face image by CNN A.
Step.3 :Give a five scale score by questionnaire to smile face image data group that create with Step 2.
Step.4 :Create the labeled dataset kind of good smile and bad smile by referring questionnaire score.
Step.5 :Training the CNN B that identifies the good smile and bad smile, using dataset of step 4.

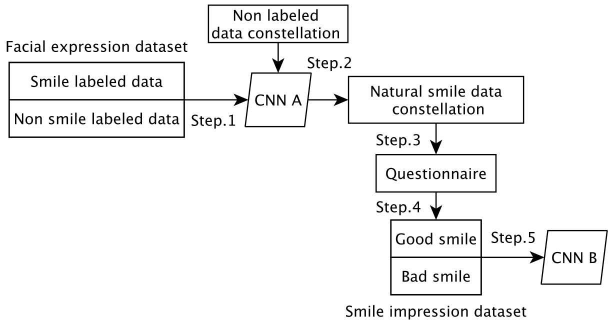

Figure 3. Details of the proposed method.

4. Experiment

4.1 Experiment Overview

We carried out an experiment to evaluate our proposed method and to evaluate the performance of CNN B. We used following dataset.

- The Japan Female Facial Expression Database(JAFFE)[8]
- Cohn-Kanade database(CK)[9]
- Montreal Set of Facial Displays of Emotion(MSFDE)[10]
- Karolinska Directed Emotional Faces Database(KDEF)[11]
- Happy People Images(HAPPEI)[12]

Table 1 shows the number of facial images included in each datasets(JAFFE, CK, MSFDE and KDEF). These datasets is used for learning of CNN A at step 1. In step2, we use non labeled data group that contains 7425 facial images. The questionnaire of Step 3 is 5 score evaluation and it is conducted to 5 people(4 men and 1 woman). In step 4, we set the high threshold to 4.3 and low threshold to 2.7. The smile impression datasets is divided into half. We use them as the training data and the test data. In this experimant, we measure the accuracy of the CNN B using 2 fold cross validation.

dataset	hap	sad	sup	ang	dis	fea	neu	total
JAFFE	40	34	41	26	39	8	25	213
CK	69	28	83	44	57	24	0	305
MSFDE	31	32	0	32	32	32	0	159
KDEF	138	130	134	134	133	136	129	934
total	278	224	258	236	261	200	154	1611

Table 1. The number of data set.

4.2 Experimental results

In the result of step 1, CNN A achieved recognition rate of 98.33%. In the result of step 2, we could extracted 1052 smile images from the non labeled data group. In the result of step 4, we obtained the smile impression datasets. The datasets consists of 63 good label images, 80 bad label images and reverse image of these images. All images of the datasets total are 286 images. We show examples of the datasets in Fig.4. In the result of step 5, CNN B achieved recognition rate of 70.28% for discrimination of smile favorability. The CNN B accuracy is shown in Fig.5. In Fig.5, row represents the actual class and column represents the predicted class.

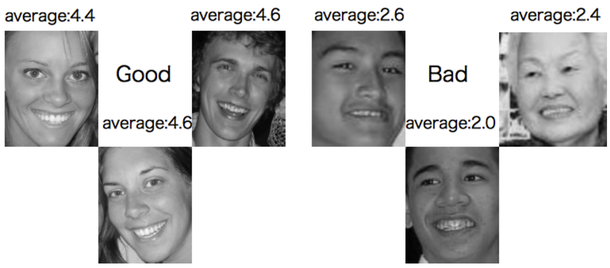

Figure 4. Examples of the smile impression evaluation dataset.

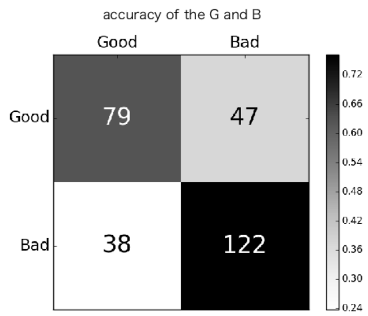

Figure 5. Confusion matrix of CNN B.

We evaluate the performance of CNN B using the evaluation data group as shown in Fig.6. In Fig.6, the horizontal axis represents the questionnaire score and the vertical axis represents the number of images. The evaluation data group is rest of step 4 that creates the smile impression datasets.

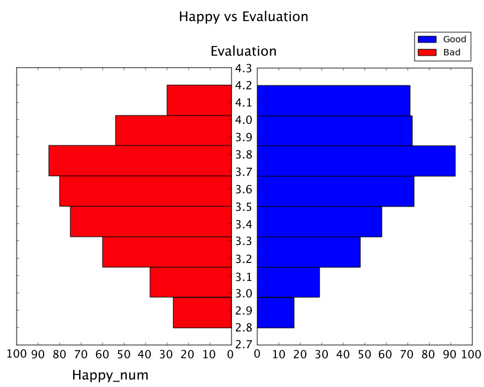

Figure 6. Recognition result of evaluation data group.

From Fig.6, when the questionnaire score of smile images are 3.6 point or less, there are more bad smile images than good smile images. This trend is consistent with the sense of the people. When the score of smile images are 3.7 point or more, there are more good smile images than bad smile images. Especially, when score is 4.1 point, good smile is a large number compared to the bad smile. From this result, CNN B discriminate the good and bad smile to some extent.

As the other experiment, to confirm how CNN B discriminate the good and bad smile, we extract the feature produced by fc8 layer of CNN B network(Fig.2). We collect the 50 images each for the good smiles and bat smiles, after that we extract the feature from these images. In Fig.7 shows that two dimensional plot of these features.

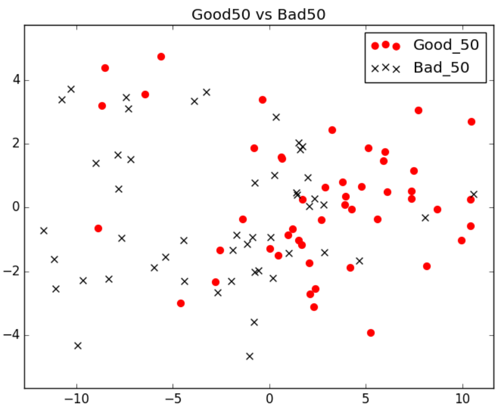

Figure 7. Two dimensional plot of feature amount of 100 images.

From Fig.7, Good smile images are gathered on the right side and Bad smile images are gathered on the left side. From these results, CNN B achieved the certain level performance that discriminate the good and bad smile.

5. Analysis for features of the face parts

We carried out discriminate the hiding each part of face(Fig.8): eyes, nose and mouth. We compared the original recognition result and this experiment recognition result to analyze which parts of the face affect recognition by CNN B. We use dataset of Fig.7 recognition result(Good:57,Bad:43) in this experiment. These recognition results are shown in Fig.9,Fig.10 and Fig.11.

Figure 8. Example of hiding faces.

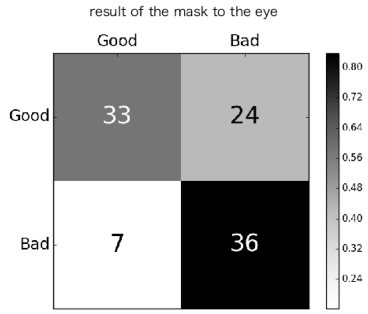

Figure 9. Confusion matrix of hiding eyes.

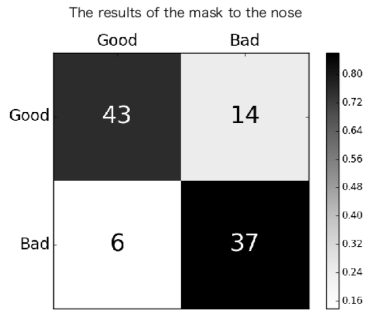

Figure 10. Confusion matrix of hiding nose.

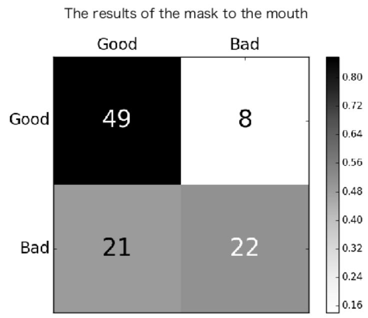

Figure 11. Confusion matrix of hiding mouth.

In the case of hiding eyes, 24 good images are misrecognized as bad smile. That means by hiding eyes area makes to disappear the important feature to discriminate the good smile. The case of hiding nose, CNN B misrecognized the 20 images, but this case is less prone to change the recognition rate compared to the other cases. Thus, nose is not important feature to discriminate the good and bad smile. In the case of hiding mouth, more than half of bad images are misrecognized as good smile. That means by hiding mouth area makes to disappear the important feature to discriminate the bad smile. From these results, recognition by CNN B is more affected by eyes and mouth.

6. Conclusion

In this paper, we performed discrimination for favorability of smile using CNN. As a result of the experiment, we obtained smile impression recognition rate of 70.28% and CNN B has certain level of performance that discriminate the good and bad smile. Recognition by CNN B is influenced by eyes and mouth of smile. From these results, we conclude that it is possible to discriminate for favorability of subtle human facial expression using CNN.

References

[1] Mehrabian Albert,*“Silent messages”* (1971)

[2] INOUE Kiyoko, et al.,*“First Time Impressions Given by Facial Expressions.”* Human Life Science 36(2014):183-194

[3] IGUCHI Takeyoshi, et al.,*“Geometrical Features of the Attractive Smile.”* Konica Minoruta technology report 4 (2007): 91-96.

[4] SUGAHARA Tohru, et al.,*“ANALYSIS OF THE RELATIONSHIP BETWEEN A DIVERSITY OF SMILES AND FACIAL IMPRESSIONS.”* Journal of Japan Society of Kansei Engineering 7.2(2007): 401-407.

[5] NISHIME Taiki, ENDO Satoshi, TOMA Naruaki,*“Emotion Acquisition From Facial Image Using Deep Neural Network.”* Journal of Japanese Society for Artificial Intelligence 29(2015): 1-4.

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

[7] http://image-net.org/challenges/LSVRC/2012/results

[8] M.J.Lyons, S.Akamatsu, M.Kamachi and J.Gyoba.,*“Coding Facial Expressions with Gabor Wavelets”* Proc. of 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp.200-205, (1998)

[9] Kanade, Takeo, Jeffrey F. Cohn, and Yingli Tian.,*“Comprehensive database for facial expression analysis.”* Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on. IEEE, (2000).

[10] Beaupré, Martin G., and Ursula Hess.,*“Cross-cultural emotion recognition among Canadian ethnic groups.”* Journal of Cross-Cultural Psychology 36.3 (2005): 355-370.

[11] Lundqvist, D., A. Flykt, and A. Öhman. The Karolinska Directed Emotional Faces-KDEF. CD-ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, Stockholm, Sweden. ISBN 91-630-7164-9, 1998.

[12] Dhall, Abhinav, et al.,*“Finding happiest moments in a social context.”* Computer VisionACCV 2012. Springer Berlin Heidelberg, 2012. 613-626.