Emotion Mate: User Emotion Reactive AI Robot Using CNN Image Classification Model

SEUNGHEE HAN[1], JAEMIN CHO[2], AND GEON KIM[3]

[1] *20170719, School of Electrical Engineering, Korea Advanced Institute of Science and Technology(KAIST), Daejeon 34141, Korea*
[2] *20160633, School of Computing, Korea Advanced Institute of Science and Technology(KAIST), Daejeon 34141, Korea*
[3] *20170057, Department of Biological Sciences, Korea Advanced Institute of Science and Technology(KAIST), Daejeon 34141, Korea*
[*] *Team Number: 31*
[*] *Project corresponds to Option 1 (Apply the existing models or techniques to your own problem)*

In this project, we choose option 1 for our subject. (Apply the existing models or techniques to your own problem)
This report contains information of our interactive emotion reactive AI robot service. It explains about how our team design new CNN model, how we train the model using which data sets, and how we construct emotion reactive AI robot system using trained image classification model and handmade robot based on Arduino and OpenCM 9.04 board.

Our team's training/inference codes, pre-trained models and demo video are available at link below:

https://github.com/shhan1755/CS470-Final-Project

1. INTRODUCTION

Recently, with an increasing dependence of computers in our lives, people spends much bigger part of their day for computer. Especially for those who use computers as their task like students major in school of computing, this time will be longer. It is difficult to communicate with others during using the computer, so that the increase in the use time of the computer leads to a decrease in the time of communicating with other people. In addition, such a decrease in emotional exchange makes people's feelings impoverished.

Here, we thought that if people can exchange emotions with others like pets while they were focusing on computers, it would make a warmer time. However, in the case of real animals, people need a lot of money and time to take care of them, so It can be a psychological and economic burden for them. It is especially difficult to raise a pet for those who live in dormitory like students in our school.

Therefore, we thought that if there is a robot that can interact with people using computer by grasping the emotions from their faces, like rejoicing together when people are happy and comforting them by the side when people are sad, it can warm people's hearts.

In this report, we developed a service that use a webcam to capture the face of a person using a computer, use the CNN deep learning model to recognize the emotions in the captured picture, and interact with them by reacting with motion and facial expressions through the robot.

Section 2 describes about the overall system structure. Section 3 describes about our image classification model construction, section 4 describes about our reactive robot construction, and section 5 describes about our full service implementation. Lastly, section 6 contains conclusion and contributions of our project.

2. SYSTEM OVERVIEW

Figure 1 is the whole appearance of the service that we made in this report.
As you can see in figure 1, our system is consist of webcam on the monitor, computer, and robot hardware. First, webcam captures picture of person who is using computer, and our software crop only the part with a face. Then, it recognizes the emotions in cropped facial pictures using the CNN model. Lastly, with the recognized emotion, robot hardware interact with person.

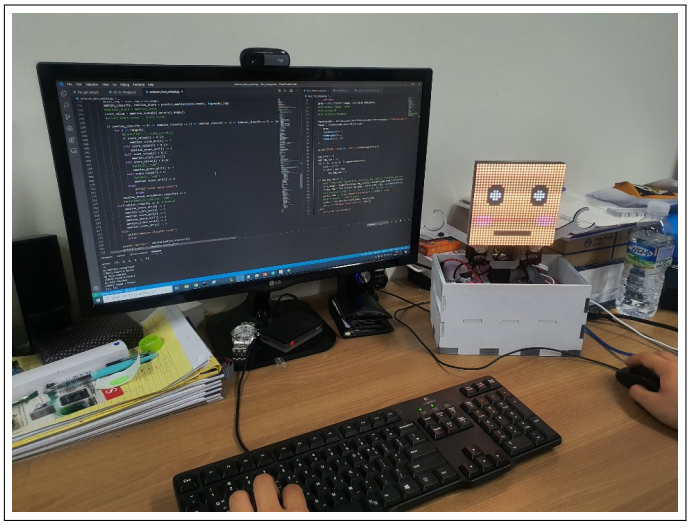

Fig. 1. Appearance of service.

3. IMAGE CLASSIFICATION MODEL CONSTRUCTION

A. Data Set Preparation

Our purpose was to train the CNN model to classify the facial emotion by 7 classes ('Anger', 'Disgust', 'Fear', 'Happy', 'Sad', 'Surprise', and 'Neutral'). To achieve our goal, we need plenty of data set which is pair of human faces image representing their emotion. Therefore, we collected facial expression data set from Kaggle facial expression recognition challenge, "Challenges in Representation Learning: Facial Expression Recognition Challenge".

From this data set, we made 28,709 training set, 3,589 validation set, and 3,589 test set. The facial image of data set is gray-scale and 48x48 size matrix whose pixels have a value between 0 and 255. Our pre-process steps are as follows. At first, from the 48x48 matrix, create image matrix of pillow library convert elements' value of matrix between 0 and 1. Then, these pillow matrices are converted to tensor format for CUDA calculation. To generalize our CNN model for arbitrary facial images. We collected four additional emotion recognition data set, 1) The Japanese Female Facial Expression database(JAFFE) 2) Facial expression recognition data set for training 'EmoPy' which is python toolkit to predict emotions. 3) Facial expression images of Korean celebrities such as actors, singers, and Youtubers. Then, we changed such images to gray-scale and 48x48 matrix to coincide data format to basic data set.

B. Model Design

We constructed CNN model to recognize 7 classes emotion expression based on VGG-19 which consists of 13 convolution layers and 3 fully connected layers with 5 max pooling layers and takes input as RGB image of size 224x224x3. However, our input image size is 48x48x1 so that it occur input size inconsistency and over pooling problem. Therefore, we revised VGG-19 to fit our data set to CNN model (Figure 2). We reduced some max pooling layers to maintain width and height of hidden layers and decreased the number of convolution layers and fully connected layer to make model simple for real time calculation. Moreover, we added dropout and batch normalization to prevent over-fitting and modified the number output of fully connected layer to seven, the number classes of emotion we categorized.

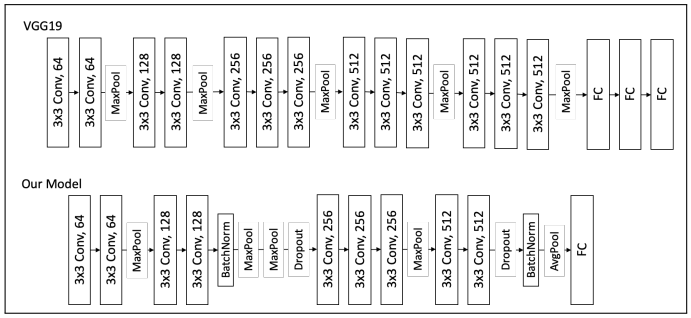

Fig. 2. Structure of CNN model.

C. Model Training

We trained our model using about 29,000 training set and about 3,700 validation set by dividing these data set into mini-batches of size 32. And, 'Adam' is chosen as an optimizer and 'Cross entropy' as loss function. During training, we tested learning degree and accuracy at that time of model, after each training epoch. If accuracy of validation set at epoch exceeded best accuracy, we saved weight of model and new best accuracy into checkpoint file.

D. Model Evaluation

To evaluate our emotion recognition model, we predicted emotions of test set using our model. Then, result of prediction is evaluated by two standard. First is accuracy of model by comparing prediction to true emotion of each data. We calculated accuracy of each class and overall classes by following equations.

$$Acc(c) = \frac{\#\ of\ correct\ predictions\ of\ class\ c}{\#\ of\ test\ set\ whose\ class\ is\ c}$$

$$Acc(overall) = \frac{\#\ of\ correct\ predictions}{total\ \#\ of\ test\ set}$$

Overall accuracy is calculated at about 63%. Interestingly, there is a big difference between accuracy of each class. Accuracy of 'Happy' and 'Surprise' classes exceeded overall accuracy and they were greater than 80%. On the other hand, 'Disgust', 'Fear', 'Sad', 'Neutral' classes were less than overall accuracy. Especially, 'fear' class has intensely bad accuracy of about 40% (Figure 3).

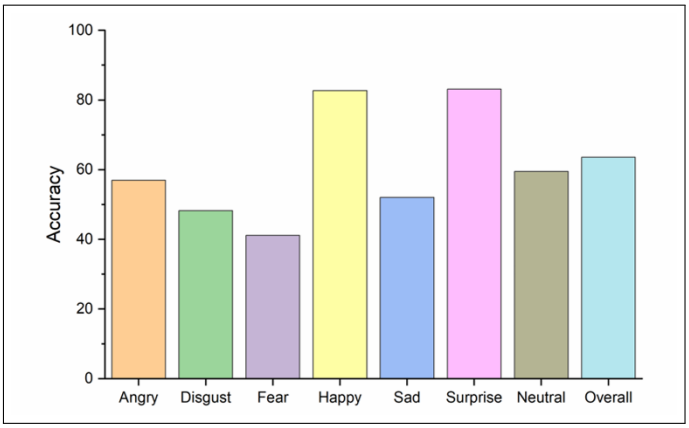

Fig. 3. Accuracy test of CNN model.

Next is AUC score of ROC curve. we calculated true positive rate (TPR) and false positive rate (FPR) of each class (Figure 4, order : (a) Anger, (b) Disgust, (c) Fear, (d) Happy, (e) Neutral, (f) Sad, (g) Surprise). As similar to accuracy test result, AUC score of 'Fear' class was worse than any other class and 'Happy' and 'Surprise' had highest score (Table 1).

The result of accuracy test did not look good but it was an inevitable consequence. There is a convincing reason of this result that each emotion does not have independent and distinct human facial feature, and it make emotion classification difficult and even people are not easy to distinguish the emotions. For example, facial expression 'Anger' and 'Disgust' is very similar and 'Fear' and 'Sad' also have similar expression.

4. REACTIVE ROBOT CONSTRUCTION

For our robot to make reaction for the emotion of person, we divided reaction to two parts, motion and simple facial expression. Driving part is responsible for motion of robot and Facial expression part is responsible for simple facial expression.

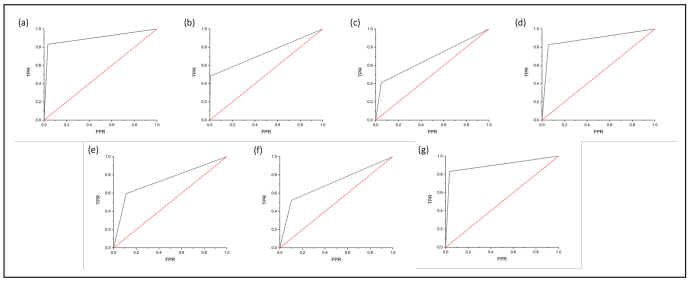

Fig. 4. ROC curve of each emotion class.

Table 1. AUC score of ROC curve for each class.

emotion	Anger	Disgust	Fear	Happy
score	0.75	0.74	0.68	0.89
emotion	Neutral	Sad	Surprise	
score	0.74	0.71	0.90	

A. Driving part Design

Driving part is consist of body part, arm part, controller module, and power module as you can see in figure 5 (a). We choose OpenCM 9.04 board for controller module and Dynamixel AX-18 models for motors for body and arm parts. When OpenCM board get signal from computer, it moves motors with power from SMPS(Switching Multi Power Supply) module. Four motors are used for moving body part of robot and two motors are used for moving arm parts of robot. With the combination of six motor movements, robot can represent various motions. This driving part is combined with the facial expression part, discussed later, to implement a variety of interactive actions.

B. Facial Expression Part Design

Facial expression part is consist of LED matrix module, controller module, and power module as you can see in figure 5 (b). We choose Arduino Mega board for controller module. LED matrix module is powered by SMPS(Switching Multi Power Supply) module, which is shared with driving part. When Arduino board get signal from computer, it changes the led color arrangement of LED matrix module. With the arrangement of LED matrix module, it can express various facial expressions.

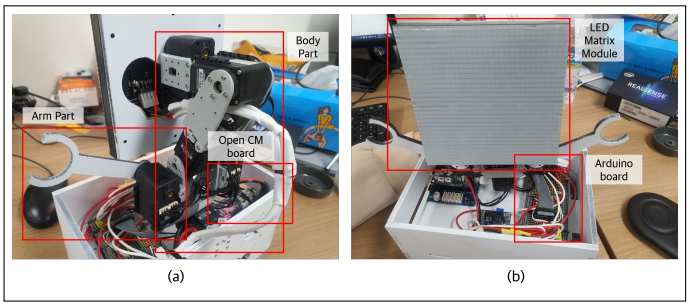

Fig. 5. (a) Driving part. (b) Facial expression part.

C. Hardware Implementation

By combining motions from driving part and facial expressions from facial expression part, our robot can express various reactions. Figure 6 shows the completed emotion reactive AI robot hardware.

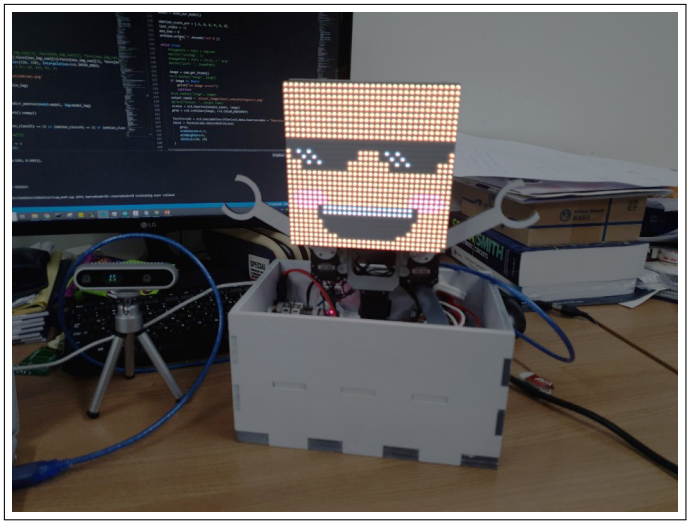

Fig. 6. Completed robot hardware.

Also, since our robot's hardware part is divided into two parts and they each have controller modules, our robot has two software. One is for OpenCM board and other one is for Arduino board. You can check out our codes for OpenCM board and Arduino board on our github link above.

5. SERVICE IMPLEMENTATION

By connecting CNN image classification model and reactive above described above with webcam and our face recognition software, we implemented full service system.

A. Face Recognition Software

To capture the picture of person using computer, we choose webcam because webcam can be installed on monitor as you can see in figure 7 (a) and we can easily get data from it. We captured picture with webcam at regular intervals, After get picture of person, we used OpenCV library to capture the crop only the part with a face as you can see in figure 7 (b). If it finds the face, it takes the picture to the next level. If not, it just waits for next picture from webcam.

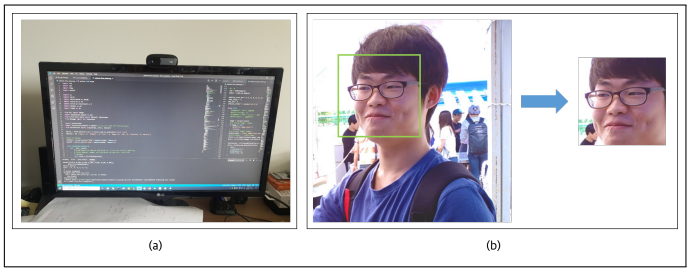

Fig. 7. (a) Webcam. (b) Face recognition software.

B. Emotion Classify Software

If face recognition software finds the face, emotion classify software classify the emotion of person from the picture come from last level using CNN image classification model. To do this, it

does pre-process to the facial image and push it to the CNN image classification model as an input. CNN model works with model structure and checkpoint which we trained with data set as we announced above. With the output from model, it uses softmax function to regularize the outputs, and use this data to recognize the emotion from picture.

In addition to the recognized emotion data from the current picture, the recognized emotion data from the previous picture is combined to classify the current person's emotions. To do this, we designed an emotion score model. Each emotion class has own score to classification. After we recognize the emotion with CNN image classification model, clear emotion recognition adds a score of corresponding emotion class, while unclear emotion subtracts a score of corresponding emotion class. In the case of an emotion with a score above a certain point, it means that a person has that emotion and we get this emotion as classified emotion of person. This approach is meaningful to our system because it can compensate for some of the slight errors that software and trained models can bring, and allows our software to consider continuous changes in human emotions. Figure 8 shows an example of how emotion classify software works.

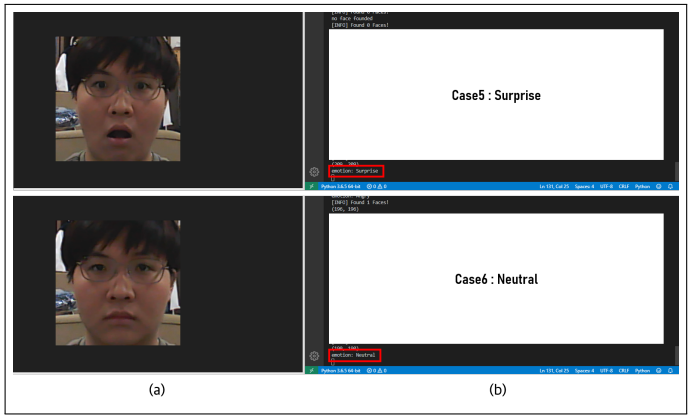

Fig. 8. (a) Cropped facial image. (b) Emotion classification.

C. System construction

We finished our system by connecting face recognition software, emotion classification software, and reactive robot. When person use computer, face recognition software captures picture of person with webcam. Then, emotion classification software classify the person's emotion using the methods mentioned above. When a person's feelings are recognized and triggered, it sends emotion code to robot by serial communication based on table 2 depending on what emotions are recognized. Lastly, Robot interacts with person by creating different gestures and facial expressions based on emotion code.
Figure 9 shows operation algorithm of the whole system and figure 10 shows how the complete service system is utilized.

Table 2. Emotion Code of Each Emotion Classes.

Emotion Code	0	1	2	3
Emotion Class	Angry	Disgust	Fear	Happy
Emotion Code	4	5	6	
Emotion Class	Sad	Surprise	Neutral	

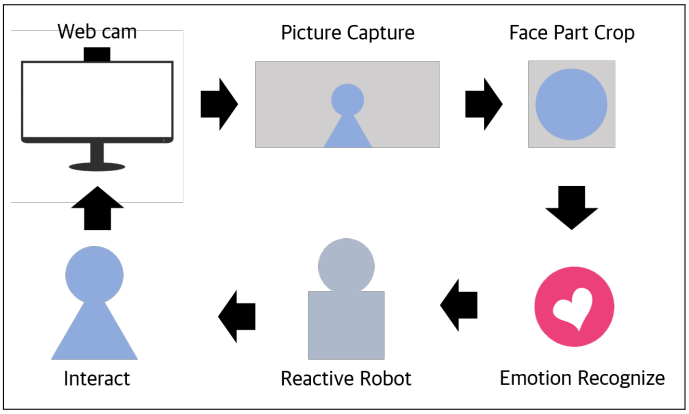

Fig. 9. Operation algorithm of system

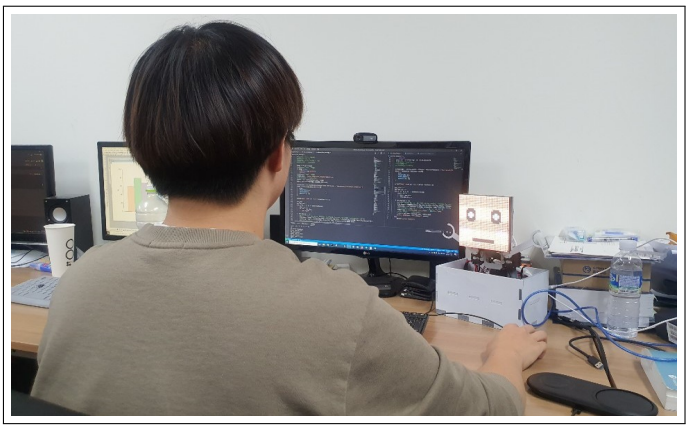

Fig. 10. Utilized complete service system

6. CONCLUSION

In this report, we made user emotion reactive AI robot service system using CNN image classification model. Through out service system, people will be able to receive various interactions. As a result of actual use, it was confirmed that it gives a meaningful response to us while we using computer. Promotion video and trial performance video are available at youtube link below.

Promotion Video
https://www.youtube.com/watch?v=001Ti6mlWBU

Trial Performance Video
https://www.youtube.com/watch?v=K0ksBRtCAI8

Even if our system has good points, it also has limitations. First, webcams often didn't recognize people's faces depending on their surroundings. We think this problem can be solved by using a higher quality webcam or by using additional software to pre-process the pictures from the webcam. Second, our system up to now only consider person's face, and can interact with only face and motion. This problem can be solved by adding additional input/output methods to our robot like mic and speaker.These modifications will bring more scalability to our service. Lastly, in relation to the CNN image classification model, the model did not show a very high accuracy in classifying emotions. This is because there are different emotions, but

there are emotions that look similar through the face. Happy and surprise, for example, have distinct elements of each emotion, but in disgust and sad cases, there is little difference in what appears in the expression. As a result, there were cases that were difficult to classify in human view (In the actual test results, the happy expression was rarely recognized as sad, but the expression of fear was often recognized as sad). Therefore, I think that learning and evaluating classifications with fewer emotions that are clearly visible on the face will give higher accuracy.

We think our system can be a great help for people who need mental healing from their pets, especially who are hard to raise real pets such as those living in dormitories or living alone. In the future, we also plan to make our service to react not also from face picture data but also from human voice data to warm people's hearts using artificial intelligence service. Our service will be able to show the interaction like communicating with real people.

7. CONTRIBUTIONS

The following is about our team's contributions.

- Seunghee Han:

  Data set search and pre-process

  CNN image classification model design

  Emotion reactive robot design

  Emotion reactive robot motion design

  Face Recognition Software develop

  Emotion Classify Software develop

  Write report

  Presentation preparing

  Demo video appearance

- Jaemin Cho:

  Data set search and pre-process

  Emotion reactive robot facial expression design

  Emotion reactive robot motion design

  Write report

  Presentation preparing

  Demo video production

- Geon Kim:

  Data set search and pre-process

  CNN image classification model design

  hyper-parameter tuning on CNN image classification model training

  CNN image classification model verification

  Emotion Classify Software develop

  Emotion reactive robot design

  Write report

  Presentation preparing

8. REFERENCE

1. "Python Face Recognition in Real Time", 2019. [Online].
2. Karen Simonyan & Andrew Zisserman, "Very Deep Convolutional Networks For Large-scale Image Recognition", ICLR 2015, 2015.
3. "thoughtworksarts/EmoPy", GitHub, 2019. [Online].
4. "fer2013", Kaggle.com, 2019. [Online].
5. M. Lyons, M. Kamachi and J. Gyoba, "The Japanese Female Facial Expression (JAFFE) Database", Zenodo, 2019. [Online].
6. "Coding facial expressions with Gabor wavelets - IEEE Conference Publication", Doi.org, 2019. [Online].

9. ACKNOWLEDGEMENT

I would like to thank Prof. Seung Hoon Hong for his hard work on the basics of artificial intelligence for the past semester. I think we learned a lot from this class.
We would also like to thank our TAs for reviewing our project proposal and giving in-depth advice on the direction of the topic.