Facial Expressions Identification System

Dae Yon Hwang

Electrical & Computer Engineering at the University of Toronto

Introduction

The advancements in computer vision, computing resources and machine learning make our world richer, efficient and innovatory. One of great achievements in machine learning area is facial expression recognition (FER). Robot, computer and many other devices can recognize the person's feeling from his face and use this information to boost or relax his emotion.
In this project, I build the facial expressions identification system using CNN and compare the results between three types of data. One of them is the facial image without background information and others are made by additional cropping on former one.

Methods

Table 1. Distribution of Expressions in Data

	Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral
Number of data	30	29	32	31	31	30	30

- Data -

Japanese Female Facial Expression (JAFFE) data [1] is used for training and testing the network. The data has 213 images with 7 facial expressions and each image is 256x256 pixel size with 8 bit precision in gray scale values. As table1, the distribution of 7 expressions are almost uniform, thus data augmentation is not needed to control the class imbalance problem.

- Facial Detector and Landmarking -

Facial detector and landmarking are applied to build the three types of input data. There are two steps to do localization: First step is to detect the face in the image (facial detector) and second step is to detect the key facial structures from ROI of face (facial landmarking).

Algorithms

- Data Preparation -

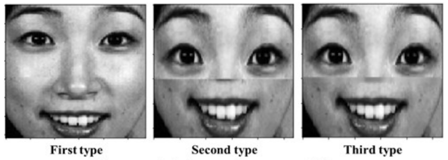

Figure 1. Three Types of Data

As figure 1, first type is the cropped image from eyebrow to mouth. Second type is the image from first type, but except for nose part. Due to reduced image size, it requires less execution time with similar accuracy because the facial expression is mainly determined by the remaining parts. Third type data is built by blurring the unrefined connection around middle part of second type image.

- Preprocessing -

Figure 2. Preprocessing Steps

- Network -

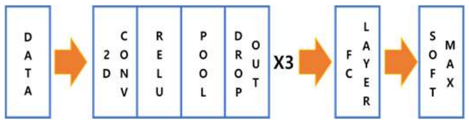

Figure 3. Structure of Network

- Performance Measurement -

Table 2. Confusion Matrix

	True Data	False Data
Predicted as True	True Positive (TP)	False Positive (FP)
Predicted as False	False Negative (FN)	True Negative (TN)

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \qquad False\ Acceptance\ Rate(FAR) = \frac{FP}{FP + TN}$$

$$Precision = \frac{TP}{TP + FP} \qquad False\ Rejection\ Rate(FRR) = \frac{FN}{TP + FN}$$

Results

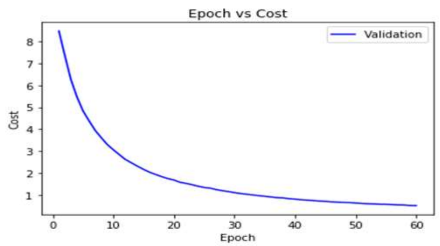

Figure 4. Epoch vs Cost curve

Figure 4 explains the cost curve of validation set in cross validation, depending on the number of epoch. The validation cost curve gives clear decreasing shape without any increasement, meaning there is no overfitting.

Table 3. Performance Comparison

	First type image	Second type image	Third type image
Average Accuracy	0.86	0.88	0.88
Average Precision	0.87	0.88	0.89
Average FAR	0.14	0.11	0.12
Average FRR	0.12	0.11	0.1
Average Time Consumption (sec)	1687	742	734

Table 3 shows the system performance of three types of data (figure 1). From table 3, we can build more stabilize and reliable system using second and third type images. Especially, they require less than half computational time than first type data, meaning they are more suitable for real time processing. In addition, third type of data gives almost same performances as second type, meaning CNN can overcome the unrefined connection area in second type image.

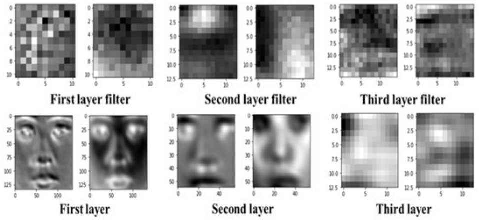

Figure 5. Feature Visualization

Conclusions

Shallow-CNN to build the facial expressions identification system

Test three types of data and find the room for improvements

Later, build the deeper CNN to get better performance and interpretation

Bibliography

1. Lyons, Michael, et al. "Coding facial expressions with gabor wavelets." Proceedings Third IEEE international conference on automatic face and gesture recognition. IEEE, 1998.
2. Kazemi, Vahid, and Josephine Sullivan. "One millisecond face alignment with an ensemble of regression trees." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014
3. Ijjina, Earnest Paul, and C. Krishna Mohan. "Facial expression recognition using kinect depth sensor and convolutional neural networks." 2014 13th International Conference on Machine Learning and Applications. IEEE, 2014.
4. Lopes, André Teixeira, et al. "Facial expression recognition with convolutional neural networks: coping with few data and the training sample order." Pattern Recognition 61 (2017): 610-628.
5. Cowen, Alan S., and Dacher Keltner. "Self-report captures 27 distinct categories of emotion bridged by continuous gradients." Proceedings of the National Academy of Sciences 114.38 (2017): E7900-E7909.
6. Brunelli, Roberto, and Tomaso Poggio. "Face recognition: Features versus templates." IEEE transactions on pattern analysis and machine intelligence 15.10 (1993): 1042-1052.
7. Suk, Myunghoon, and Balakrishnan Prabhakaran. "Real-time mobile facial expression recognition system-a case study." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2014.

printed by MegaPrint Inc. www.postersession.com