

Validation Research on the Application of Depth-wise Separable Convolutional AI Facial Expression Recognition in Non-pharmacological Treatment of BPSD

Xiangyu Liu*

Doctoral Course, Graduate School of Human Sciences, Waseda University, Japan

*Corresponding author: Xiangyu Liu, dukeedison@fuji.waseda.jp

Abstract: One of the most obvious clinical reasons of dementia or The Behavioral and Psychological Symptoms of Dementia (BPSD) are the lack of emotional expression, the increased frequency of negative emotions, and the impermanence of emotions. Observing the reduction of BPSD in dementia through emotions can be considered effective and widely used in the field of non-pharmacological therapy. At present, this article will verify whether the image recognition artificial intelligence (AI) system can correctly reflect the emotional performance of the elderly with dementia through a questionnaire survey of three professional elderly nursing staff. The ANOVA (sig.=0.50) is used to determine that the judgment given by the nursing staff has no obvious deviation, and then Kendall's test (0.722**) and spearman's test (0.863**) are used to verify the judgment severity of the emotion recognition system and the nursing staff unanimously. This implies the usability of the tool. Additionally, it can be expected to be further applied in the research related to BPSD elderly emotion detection.

Keywords: Depth-wise separable convolution; Emotion; BPSD; Dementia; Nursing

Publication date: July 2021; **Online publication:** July 31, 2021

1. The role and status of emotion recognition in the study of dementia

The term "emotion" is defined as "a person's attitude and experience of whether objective things meet their own needs or not." At the same time, in the current research on the interaction between emotion and the brain as humans understand it, emotion is the brain's prediction of the external reaction. It is an important tool to influence and reflect brain activity in another dimension. Richard pointed out that emotions are processed by the limbic system, which can more directly observe and measure people's emotions and needs [1]. In addition to human voices and words, facial expressions are used to convey what they want to express. Similarly, in the study of dementia, we also need to pay attention to the real psychological activities and emotional needs of people with dementia as reflected by facial expressions.

One of the most obvious clinical reasons of dementia or BPSD is the lack of expression of emotions, the increased frequency of negative emotions, and the impermanence of emotions. Observing the reduction of BPSD in dementia patients, through emotions can be considered effective and widely used in the field of non-pharmacological therapy. At present, this article will verify whether the image recognition AI system may correctly reflect the emotional performance of the elderly with dementia through a questionnaire survey of professional elderly nursing staff.

2. Current challenges of emotion recognition

There are currently three main methods for measuring emotions, the emotional facial motion coding system,

electromyography (EMG), and automatic face recognition [2]. Karsten pointed out that the first method is inefficient, as the time required to analyze the fixed images one by one according to the specific use of the facial action coding system (EMFACS) method is too tedious. In previous studies, the emotional facial action coding system can objectively quantify the expression of emotions in normal people. Professionally trained code researchers are used to score objective images. This process requires researchers to manually perform this process, which is relatively inefficient. Compared with emotion recognition, the time required to use this method is one of the reasons why there are few expression studies. The disadvantage of the EMG system is its technical complexity and its limitations in the experimental environment. For example, sensors must be attached to the face to detect the movement of facial muscles. Therefore, it is impossible to use this system to check a person's facial expression in a natural social environment. At the same time, Karsten points out that automatic face recognition may completely replace the first two methods.

3. Improve efficiency through depth-wise separable convolution

In 2017, Google used a new computing structure for convolutional neural networks-depth-wise separable convolution in MobileNet to reduce the computational cost derived from the CNN model [3]. This method mainly splits the original convolution calculation method into two parts, depth-wise convolutions and pointwise convolutions, to reduce the number of calculations without affecting the output structure [4]. We can simply understand that using fewer system resources to successfully compute AI deep learning tasks that previously required a lot of computing resources to achieve, so that AI deep learning can adapt to more lightweight scenarios to meet more types of needs [5]. With the help of AI, we are able to use open-source tools to quantify and analyze the mood swings of elderly people with dementia in social settings.

Part of the code of this tool refers to the code in GitHub: https://github.com/oarriaga/face_classification. The detection output results are divided into seven expressions: normal, angry, contempt, disgust, fear, happy, and sad. This article uses data sets containing 28709 training samples, 3859 validation data sets, and 3859 test samples, a total of 35887 images containing seven categories of angry, disgusted, fearful, happy, sad, surprised, and normal. The image resolution is 48×48 (pixels). Most of the images in this dataset have rotations on planes and non-planes, and many images are blocked by occludes such as hands, hair, and scarves. The training model we use is integrated on the FER2013 data set (another relatively complete classification database). The classification accuracy rate is 86%. With the help of such tools, we have completed the basic environment construction (hereinafter referred to as Emo-Rec Application), and then we will actually apply it to the experimental data of BPSD patients with dementia to further explore the actual usability of this tool.

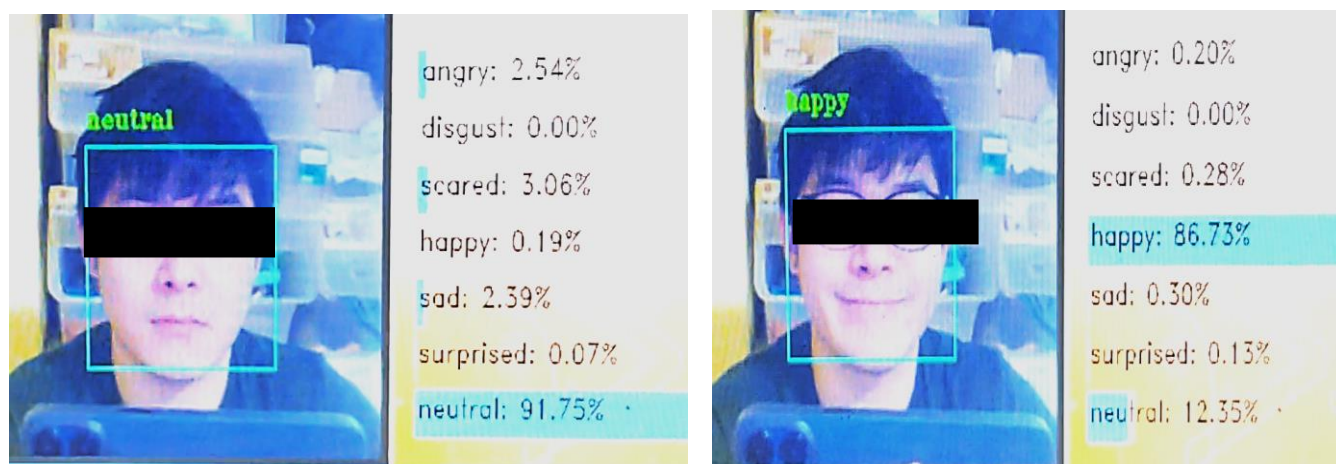


Figure 1. The Emo-Rec Application

4. Design

Reminiscence Group Therapy (RGT) usually takes about 15 minutes for every intervention. If the time point is accurately quantified, the observation result will be affected by the shooting angle and motion state. Therefore, it is unrealistic to identify emotions and to extract data through the quantitative time of each experiment. In the end, we found that we can divide it into three obvious stages: before, during, and after therapy, that can objectively reflect the emotions of the subjects at the time. Since the video data collected through the RGT experiment can be directly obtained in the Emo-Rec Application, we first select the time and record the time in the current video, score, and record the current picture. After that, we interviewed three professional nursing staff who had experience in nursing elderly people with dementia and used questionnaire surveys to rate the uniform time point of data obtained by the Emo-Rec Application (**Figure 2.**)

Afterward, the returned questionnaire was used to verify whether the opinions given by the three experts were consistent. If they are consistent, compare their average results with the results of the Emo-Rec Application to verify whether the Emo-Rec Application can be competent for emotion recognition of elderly people with dementia.

フォームの左上隅にビデオファイル名がマークされています。フォームにマークされたビデオの時点での写真の高齢者がポジティブかどうかを判断してください（1〜20の整数評価を入力してください。評価基準は：とても幸せ：1、とても悲しい：20）。

The upper left corner of the form is marked with the video file name. Please judge whether the elderly in the picture at the time of the video marked in the form are **positive**. (please give an integer evaluation between 1-20, the evaluation standard is: very happy: 1, very sad : 20)

20201204	be-1	be-2	be-3	du-1	du-2	du-3	ed-1	ed-2	ed-3
Time	00:10	00:18	01:09	06:42	10:28	11:40	14:27	14:53	15:43
A									
B									
C									
D									
E									
F									
G									
H									

Figure 2. Questionnaire surveys

5. Result

The three experts observed in the experimental data of the same subject showed different scoring tendencies (**Table 1.**), but there was no statistically significant difference (**Table 2.**).

Table 1. Different scoring tendencies

	N	M	Std	p	95% CL	
					LL	UL
P1	258	10.3566	3.47710	0.21468	9.9303	10.7829
P2	258	10.6822	3.38912	0.21100	10.2667	11.0977
P3	258	10.6395	3.37735	0.21026	10.2255	11.0536
Total	774	10.5594	3.41346	0.12269	10.3186	10.8003

Table 2. Statistically significant difference

	Sum of S	df	MS	F	Sig.
Between Groups	16.158	2	8.709	0.693	0.500
Within Groups	8990.609	771	11.661		

We used a one-way ANOVA test to determine whether the scores of the three experts are consistent, and concluded that there is no significant difference (sig.= 0.5, P>0.05). It can be considered that the scores given by the three experts in the 258 image scores are relatively consistent and credible.

6. Compare the consistency of manual scoring and Emo-Rec Application scoring

We used a coefficient-weighted method to calculate the percentage of emotions given by the Emo-Rec Application, and obtained a score on the same scale as the questionnaire, and compared it. At the same time, in order to exclude the influence of other emotions on the happy-sad-neutral dimension, other emotions are marked with the same coefficient as neutral. Due to the two sets of data do not conform to the normal distribution (**Table 3.**). The tests of Kendall and Spearman are used to verify the consistency of the two evaluation methods (**Table 4.**).

$$\text{Emo-Rec Application point} = (\text{anger} * 10) + (\text{disgust} * 10) + (\text{scare} * 10) + (\text{happy} * 1) + (\text{sad} * 20) + (\text{surprise} * 10) + (\text{neutral} * 10)$$

Table 3. Lilliefors significance correction

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	Df	Sig.
P123	0.152	258	0.000	0.937	258	0.000
EmoApp	0.140	258	0.000	0.952	258	0.000

Table 4. Kendall and Spearman test

Test			P123	EmoApp
Kendall's	P123	CC	1	.722**
	EmoApp	CC	.722**	1
Spearman's	P123	CC	1	0.863**
		CC	0.863**	1
	EmoApp	N	258	258

**Correlation is significant at the 0.01 level (2-tailed).

In summary, it can be considered that the Emo-Rec Application and the results reflected by the nursing staff have strong consistency. It can be expected that more efficient AI systems will be used in this type of field in the future to obtain statistical data with more reference value.

7. Discussion

Different from the expensive emotion recognition software in the past, Emo-Rec Application can easily quantify facial expression data through statistics and help nurses and researchers intuitively feel the non-pharmacological therapy for the elderly with dementia. The effect brought about was compared with the previous emotion quantification tools used to verify the effects of non-pharmacological therapy, the Emo-Rec Application improves the detection efficiency while ensuring the same accuracy, and can provide reference data in more dimensions (fright, fear, disgust, etc.). and also look forward to more possibilities for researches in this field ^[6].

To be precise, the survey does not reflect emotions in other dimensions belonging to the experience of professional nursing staff, thus we can verify the usability of the tool in the happy and sad dimensions. But through this application, we get other kinds of emotions, and in future research, we can analyze other emotions to get a more complete comparison dimension. For example, surprise and disgust are more detected in the early robot contact experiments, and relatively reduced in the later period; we noticed that some elderly people with missing teeth are more likely to reflect angry emotions, and elderly people who need more care (NPI>30) Mood fluctuations are greater than those of the elderly with a light nursing burden (NPI<30) ^[7].

According to many findings in this field in previous studies, we can explain the reasons for the emotions of the elderly we currently observe. For example, principles such as PE (person entertainment) can explain the positive emotions of the elderly who are just beginning to come into contact with robots ^[8], as well as surprises and large fluctuations in Performance. Additionally, some studies have found that non-pharmacological treatment of BPSD can make the mood changes of the elderly tend to be positive ^[9]. These can help us better study the various mood changes of elderly patients with dementia in non-pharmacological therapy in future research.

8. Limitation

In the actual application of the Emo-Rec Application, we found that most of the elderly's emotions can be accurately identified in a short time, and the main emotions and other emotions are given. Contrarily, at the current stage of the covid-19 epidemic, people need to wear masks to communicate in some situations. In this case, the Application is relatively difficult to function. Although there are cases of partial occlusion or rotation in the samples learned by the system, but it is still difficult to make it work because masks cover more than half of the face (**Figure 3.**).

The Emo-Rec Application for the elderly with dementia produced by the depth-wise separable convolution technology brings more efficient accuracy. It can increase the training samples to improve the recognition accuracy, and it can also flexibly match the accuracy through different training models ^[10]. However, according to Rachael's discovery, the training database of Asian faces and the database of European and American faces may not be universal in actual use, which will bring certain errors. Using local emotion databases in different cultures and regions will get better results ^[11]. Finally, adjusting the frequency of data collection according to the intervention method and time can also better reflect the results of the intervention experiment.

Ordinary situation:



anger:	0
disgust:	0
scared:	0
happy:	30
sad:	0
surprised:	0
neutral:	70

Wearing a mask:



anger:	?
disgust:	?
scared:	?
happy:	?
sad:	?
surprised:	?
neutral:	?

Figure 3. The Emo-Rec Application for the elderly with dementia produced by the depth-wise separable convolution technology with and without mask.

Disclosure statement

We use Python3.7 to implement the EMO-REC app of this research, mainly based on GITHUB: https://github.com/oarriaga/face_classification. Opencv-python:4.1.0.25; tensorflow:1.13.1.

References

- [1] Cytowic R, 2003, The Man Who Tasted Shapes, April 2003 MIT Press.
- [2] Wolf K, 2015, Measuring facial expression of motion, Dialogues Clin Neurosci. Dec; 17(4): 457–462.
- [3] Andrew G., 2017, Howard, MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, arXiv:1704.04861v1.
- [4] Pröve PL, 2017, An Introduction to Different Types of Convolutions in Deep Learning, <https://towardsdatascience.com/types-of-convolutions-in-deep-learning-717013397f4d>.
- [5] Yoon K, 2014, Convolutional Neural Networks for Sentence Classification, EMNLP, 1746-1751
- [6] Chollet F, 2017, Xception: Deep Learning with Depth-Wise Separable Convolutions, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1251-1258.

- [7] Arriaga O, Valdenegro-Toro M, Plöger P, 2017, Real-Time Convolutional Neural Networks for Emotion and Gender Classification. arXiv preprint arXiv:1710.07557.
- [8] Jain DK, Shamsolmoali P, Sehdev P, 2019, Extended Deep Neural Network for Facial Emotion Recognition. Pattern Recognition Letters, 120: 69-74.
- [9] Rachael EJ, 2012, Facial Expressions of Emotion are Not Culturally Universal, PNAS, 109 (19): 7241-7244
- [10] Heerink M, Krose B, Evers V, Wielinga B, 2010, Assessing Acceptance of Assistive Social Agent Technology by Older Adults: The Almere Model. Int J Soc Robot 2:361–375
- [11] Stafford RQ, et al., 2010, Improved Robot Attitudes and Emotions at a Retirement Home After Meeting a Robot, 19th IEEE International Symposium on Robot and Human Interactive Communication, Principe di Piemonte - Viareggio, Italy.