Predicting emotions using channel attention mechanism on fMRI signal data

Hong-Kun, Lin[1][0009-0007-2101-1525],Chia-Yen, Lee[2*][0000−0002−7477−782X]

[1] National Yang Ming Chiao Tung University, Taiwan
[2] National United University, Taiwan
windspeak122@gmail.com

Abstract. fMRI applications in emotion recognition can assist doctors in better assessing patients' emotions, thereby aiding in the development of personalized treatment plans. This study incorporating an SE block into the first layer of the CNN model to automatically focus on brain regions related to emotion classification among 246 fMRI signals (SE method). The results for the Class classification task achieved a Precision, Sensitivity, and F1-score of 0.73, 0.73, and 0.73, respectively, and for the Level task, they were 0.23, 0.32, and 0.26, respectively. Additionally, in this study, directly using brain regions previously identified as emotion-related as model input resulted in Precision, Sensitivity, and F1-scores of 0.49, 0.48, and 0.47, respectively, and for the Level task, scores of 0.04, 0.07, and 0.05 were achieved. The results indicate that employing the SE method to focus the model on brain regions associated with emotions enhances performance in emotion recognition tasks, achieving an error rate of 0.4213 in the second phase of the ICBHI challenge.

Keywords: fMRI, CNN, SE block.

1 Introduction

Functional magnetic resonance imaging (fMRI) offers high spatial resolution and whole-brain coverage, enabling the identification of different brain regions and the activity patterns of these regions under various emotional states. Additionally, its non-invasive diagnosis makes it more widely accepted by the general public.

The application of fMRI in emotion recognition can assist doctors in better assessing patients' emotions. For example, it can be used to diagnose mental illnesses related to emotional dysregulation, such as depression, anxiety disorders, and post-traumatic stress disorder. It can also be applied in emotional assessments for individuals with limited language abilities, lie detection, and more, thereby aiding doctors in developing personalized treatment plans. Additionally, fMRI-based emotion recognition can improve human-computer interaction, enabling machines or AI to better understand and respond to human emotions, making interactions with machines more human-like.

Today, many studies have used fMRI to understand the brain regions involved in processing emotions. For example, Zald et al. identified the amygdala as the central structure for emotion processing [1], while Lee et al. suggested that the left dorsolateral prefrontal cortex (A9/46d) and anterior superior temporal sulcus (aSTS) are highly correlated with emotions [2]. In recent studies, with the rise of AI, research combining fMRI with CNNs for emotion recognition has gradually emerged. For instance, Liu et al. developed an attention-based deep neural network combined with fMRI images for real-time monitoring of patients with depression [3], and Gu et al. achieved nearly 90% accuracy in binary classification of positive and negative emotions using fMRI combined with CNNs, though slightly lower than that of linear regression methods, it still maintains high accuracy [4].

However, few studies have explored whether using emotion-related brain regions as CNN inputs can enhance emotion classification capabilities. This study prospectively develops a deep learning model architecture that focuses on signals from brain regions associated with emotions, specifically targeting 246 brain regions related to emotion recognition tasks. The performance of using signals from brain regions highly correlated with emotions [2] as CNN model inputs is compared.

The model architecture and code for this study are available at https://github.com/BMS410Wind/ICBHI2024.

2 Materials and Methods

2.1 Dataset acquisition

This study uses data sourced from the 2024 International Conference on Biomedical and Health Informatics, Mind-Reading Emotions Challenge. The challenge provides a dataset consisting of 20 participants, with each participant contributing 30 trials of 25-second data collected from three synchronized sources: 1. Preprocessed fMRI data (including 246 brain regions), 2. Photoplethysmography (PPG) data, and 3. Respiratory data. However, this study focuses on the signals from the various fMRI brain regions for emotion classification and, therefore, does not utilize the PPG and respiratory data provided in the challenge.

Each trial was collected while the participant watched an emotion-provoking video clip of three predefined classes (positive valence, negative valence, or neutral valence). After each trial, participants rated the valence of their emotions on a 9-level scale, ranging from -4 (maximum negative valence) to 4 (maximum positive valence). Positive valence is associated with emotions such as happiness, joy, and contentment, making experiences appealing and desirable. In contrast, negative valence involves emotions such as sadness, anger, and fear, which make experiences unappealing and undesirable.

2.2 Methods

This study uses the Squeeze-and-Excitation block (SE block) [5] to enable the model to automatically focus on brain regions relevant to the classification task, and compares this approach with using the emotion-related brain regions identified by Lee et al. [2] as inputs for the emotion classification model. The dataset is split into training, validation, and testing sets in an 8:1:1 ratio.

2.2.1 Data Preprocessing

Since the signals measured by fMRI have different value ranges, they are normalized to a range between 0 and 1 (Equation 1). In deep learning models, calculations often involve matrix multiplication, and if the input data has a wide range, it can lead to numerical instability, such as gradient explosion or vanishing gradient problems. Additionally, when different features have varying magnitudes, the model may be biased toward features with larger values, potentially leading to imbalanced learning outcomes. By normalizing the data, the values of different features become similar, which helps the model learn from all features more equivalently.

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

2.2.2 Baseline CNN model

The baseline CNN model in this study integrates residual connections [6], Batch Normalization [7], and SE block to enhance the model's performance and stability. Residual connections are used to alleviate the vanishing gradient problem in deep networks, Batch Normalization helps accelerate convergence and improve the model's generalization ability, and the SE block enhances the model's focus on important features by adaptively adjusting the weights of channels. The detailed model architecture can be found in the GitHub link provided in Chapter 1.

2.2.3 Deep learning model architecture focusing on signals from emotion-related brain regions.

Since previous studies have shown that only certain fMRI brain region signals have a higher correlation with emotions, the relevance of different brain region signals to the emotion recognition task varies. Therefore, this study uses the SE block to input the 246 brain region signals as different channels(see Fig. 1), allowing the model to focus on emotion-related brain region signals through AI training, thereby enhancing the model's emotion recognition capabilities.

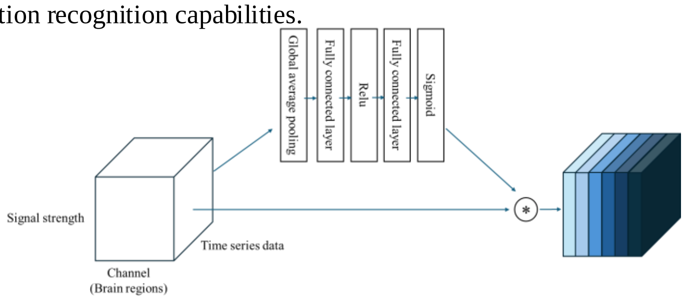

Fig. 1. Focusing on emotion-related brain region signals through the SE block.

2.3 Evaluation methods

This study employs common classification model evaluation metrics, including Precision, Sensitivity, and F1-score. The submission Score (Equation 2) is calculated by uploading the predicted results to ICBHI challenge, and the Score is composed of Level error (Equation 3) and Class error (Equation 4).

$$\text{Score}=0.6*\text{CLASS}_{\text{error}}+0.4*\text{LEVEL}_{\text{error}} \tag{2}$$

$$\text{CLASS}_{\text{error}}=1-\frac{1}{N}\sum_{K=1}^{N}(\text{pred\_class}_k==\text{true\_class}_k\ ?1:0) \tag{3}$$

$$\text{LEVEL}_{\text{error}}=\frac{1}{N}\sum_{K=1}^{N}\left(\frac{|\text{pred\_level}_k-\text{true\_level}_k|}{|\text{pred\_level}_k|+|\text{true\_level}_k|}\right) \tag{4}$$

3 Results

By incorporating the SE block into the first layer of the CNN model for automatic focus on emotion-related brain regions (SE), the Precision, Sensitivity, and F1-score for the Class classification task are 0.73, 0.73, and 0.73, respectively. In contrast, directly using brain regions previously identified as emotion-related as inputs (Emotion-related brain regions, EBR) resulted in Precision, Sensitivity, and F1-score of 0.49, 0.48, and 0.47, respectively, as shown in Table 1 and Table 2. The submission of SE prediction results yielded a first-stage score of 0.5132 and a second-stage score of 0.4213, as shown in Table 3.

Table 1. Class classification task inner test performance (Vanilla: Baseline CNN model, SE: Using SE block in the first layer of the CNN model, EBR: Using emotion-related brain regions as inputs to the CNN model)

Class task	Precision	Sensitivity	F1-score
Vanilla	0.59	0.54	0.52
SE	0.73	0.73	0.73
EBR	0.49	0.48	0.47

Table 2. Level classification task inner test performance

Level task	Precision	Sensitivity	F1-score
Vanilla	0.19	0.25	0.21
SE	0.23	0.32	0.26
EBR	0.04	0.07	0.05

Table 3. SE method Submit Score

SE	Score
First Stage	0.5132
Second Stage	0.4213

4 Discussion

Comparing the SE method with the Vanilla method, the results show improvements of 14%, 19%, and 21% in Precision, Sensitivity, and F1-score, respectively, for the Class classification task, and improvements of 4%, 7%, and 5% in Precision, Sensitivity, and F1-score, respectively, for the Level classification task. The results indicate that adding the SE block to the first layer of the baseline CNN model for automatic brain region focus contributes to better model performance.

When comparing the SE method with the EBR method, the SE method shows a clear advantage in Precision, Sensitivity, and F1-score for both the Class and Level classification tasks. Additionally, the EBR method performs worse in Precision, Sensitivity, and F1-score compared to the Vanilla method in both Class and Level classification tasks.

These experimental results demonstrate that using the SE method to allow the model to adaptively focus on brain regions is more effective for emotion recognition tasks than directly using emotion-related brain regions as inputs for the baseline CNN model.

5 Conclusion

To further enhance the performance of emotion recognition using fMRI combined with CNNs, this study compares the SE and EBR methods. The results show that the SE method outperforms the EBR method used in this study in Precision, Sensitivity, and F1-score for both Class and Level classification tasks. The SE method automatically adjusts channel weights, enabling the model to more effectively focus on important features relevant to emotion classification, thereby improving classification performance.

However, the EBR approach varies across different studies. In the future, by integrating EBR methods from other studies and then performing cross-validation to compare the SE and EBR methods, the accuracy and stability of fMRI-based emotion recognition tasks can be further improved.

References

1. Zald, D. H. (2003). The human amygdala and the emotional evaluation of sensory stimuli. Brain Research Reviews, 41(1), 88-123.
2. Lee, Y., Seo, Y., Lee, Y., & Lee, D. (2023). Dimensional emotions are represented by distinct topographical brain networks. International Journal of Clinical and Health Psychology, 23(4), 100408.
3. Liu, J., Dey, N., Das, N., Crespo, R. G., Shi, F., & Liu, C. (2022). Brain fMRI segmentation under emotion stimuli incorporating attention-based deep convolutional neural networks. Applied Soft Computing, 122, 108837.

4. Gu, J., Luo, X., Gong, X., & Su, C. (2024). Investigating the brain network characteristics of multimodal emotion recognition and its classification applications based on functional connectivity patterns. Biomedical Signal Processing and Control, 96, 106635.
5. Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7132-7141).
6. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
7. Ioffe, S. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167.