2025 IEEE International Conference on Mechatronics and Automation (ICMA) | 979-8-3315-1426-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICMA65362.2025.11120812

A Classification Model based on SEB-CNN for P300 Signal Recognition of P300 Speller

Zhaohua Lu, Guanqun Li, Xiaowei Liu*
School of Computing and artificial intelligence
Shandong University of finance and economics, Jinan 250014, P.R. China
* Correspondence: zhaohualu@sdufe.edu.cn

Qi Li
School of Computer Science and Technology, Changchun University of Science and Technology, Changchun 130022, P.R. China
liqi@cust.edu.cn

***Abstract* - The P300 speller is a canonical application of brain-computer interface technology, where the critical challenge lies in improving the classification performance of P300 signals - particularly under conditions of limited stimulus repetition. This study proposes an enhanced convolutional neural network architecture that incorporates Squeeze-and-Excitation Blocks (SEB-CNN) to address this challenge. Building upon the conventional CNN framework (C-CNN), our approach introduces two key innovations: (1) spatial convolution mechanisms for feature extraction and (2) temporal attention mechanisms for sequence modeling. This dual mechanism design significantly enhances the model's adaptive capability and cross-subject generalization performance when processing P300 EEG signals. Experimental results demonstrate that the SEB-CNN model achieves markedly superior classification accuracy compared to traditional CNN approaches, with particularly notable improvements observed in low-repetition stimulus conditions (p<0.01).**

Index Terms - Brain Computer Interface, P300 wave, P300 spelling system, CNN, Squeeze-and-Excitation Block.

I. INTRODUCTION

Brain-computer interface (BCI) is a direct communication system that enables communicate between the brain and the external environment without the need for peripheral nerves or muscles. It measures brain activity and records EEG signals through electrodes placed in the cerebral cortex or implanted within the cerebral cortex. Then it provides communication or environmental control for the user by analyzing the recorded EEG signals to send messages or commands to computers or other devices [1-3]. BCI technology enables patients with impaired neural functions to operate computers and specially designed electronic devices, thereby enhancing their quality of life and mitigating the adverse effects of disability. At present, BCI technology demonstrates vast potential across multiple domains, including fundamental and applied neuroscience, assistive technologies for patients with neurological impairments, medical rehabilitation, monitoring of mental and consciousness states, geriatric care, assistive control in extreme environments, and so on.

The character spelling system is a typical application in BCI. It provides users a way to engage with the outside world through "mental" typing by designing different evoked paradigms. The P300 spelling paradigm is one of the earliest and most classic character spelling paradigms in BCI system. P300, a well-known event-related potential (ERP) in brain-computer interfaces [4], exhibits heightened sensitivity to rare stimuli (Oddball paradigm). Research indicates that the amplitude of the P300 waveform increases inversely with the target stimulus probability [5]. It typically emerging in the central region of the parietal lobe about 300 milliseconds after the stimulus is presented, and the waveform is positively deflected. The P300 spelling system (P300 speller) is to apply the Oddball event that can induce the P300 potential to the paradigm design, that is, to set two different presentation probabilities for the randomly presented stimulus (character or group of characters), and the subject responds to the small probability stimulus (target stimulus) (e.g., count the number of times the target appears silently), and the P300 component typically emerges approximately 300 milliseconds following the presentation of a target stimulus., and by detecting the P300 potential and identifying its corresponding target stimulus, the target character can be output[6].

In the P300 speller system, Signal processing and recognition performance of P300 EEG signals are important factors affecting the spelling performance of characters spelling. The accuracy of signal classification is critical to the correct output of the target character. However, EEG signals are susceptible to a variety of factors, such as noise interference, individual differences, and signal acquisition conditions. In order to improve the classification accuracy of P300 EEG signals, researchers have investigated various signal recognition approaches, from traditional machine learning to deep learning, the latter has demonstrated significant promise and has achieved remarkable success in the fields of image, video, audio, and text analysis [7-8]. The most common model for the recognition of P300 EEG signals in the character spelling paradigm is convolutional neural network (CNN), and standard CNN models for P300 classification utilize distinct convolutional layers to learn time-domain and spatial-domain representations of EEG signals, and has achieved good classification performance [9-10]. However, the parameters of complex CNN networks are excessive, and the model needs a large dataset for training, otherwise it will fall into overfitting and local optimality.

Recent CNN architectures increasingly incorporate attention mechanisms. Attention mechanism is a method that simulates the way human attention is distributed, which can enable selective emphasis on relevant input features, thereby enhancing the ability of the neural network to process the input data. The attention mechanism allows the model to focus on discriminative features while suppressing irrelevant information for classification. which can be introduced into multi-scale and multi-branch networks. Hu et al. [11] proposed the Squeezing-Excitation Attention Module (SENet) to automatically learn channel weights, which significantly improved the effect of feature extraction. Park et al. [12] introduced both Bottleneck (BAM) and Convolutional Block (CBAM) Attention Modules[13], which effectively combine spatial and channel attention. In [14], an attention-enhanced pyramid network was proposed for efficient segmentation. (EPSANet) to improve the classification performance without significantly increasing the computational burden. Barmpas et al. [15] proposed a deep learning framework based on dynamic convolution, which uses causal inference to characterize all possible distribution shifts caused by differences between subjects in motor imagination tasks, attention networks learn differences between different subjects, and each convolutional block in CNN uses a parallel trainable convolutional layer, and its deep model achieves better generalization performance in cross-subject experiments. It can be seen that the use of dynamic convolution and attention mechanism to deal with individual differences of subjects and ensure the integrity of features has good research value for improving the decoding performance of EEG signals.

In view of this, this paper proposes an improved CNN model that merges channel attention mechanism and multi-scale feature extraction, which is denoted as SEB-CNN. The model incorporated a Squeeze-and-Excitation Block (SEB) [16], which computes channel-wise attention weights via global average pooling followed by two FC layers, which were then multiplied by the original features. By dynamically adjusting the importance of feature channels, the network can automatically learn to "pay attention" to key features and suppress redundant information, thereby improving the performance of the model.

II. Methods

A. Experimental data

The experimental data were obtained from an EEG dataset recorded during audio-visual bimodal P300 spelling experiment and the single-vision P300 spelling experiment. The spelling paradigm is described fully in reference [17]. Both the audiovisual bimodal and single-visual spelling paradigms originate from the regional flashing paradigm. [18]. The output of the target character in the regional flashing paradigm consists of a group region and a subregion flashing (a total of six group regions and six subregions), which first locates the region block where the target character is located, and then locates the single subregion block where the target character is located. Flashes randomly once every 6 group regions and 6 subregions as a stimulus overlay, and flashes 10 times for 6 group regions and 6 subregions (10 overlays) to output a target character. Therefore, 120 responses from the group region and sub-region flashes should be identified 20 times to determine the target character. In the AV and V spelling experiments, each subject produced 5 words (5 characters per word), totaling 25 characters. This resulted in 25,210 P300 and 251,010 non-P300 instances in the dataset.

The EEG data were recorded at 250 Hz via 31 electrodes. Since ERPs primarily occur within 0–800 ms post-stimulation, each trial was segmented into 200 time points (0–800 ms), yielding a 200×31 data matrix (time × channels) per stimulus. Thus, the EEG data for each subject consisted of 500 target vectors (200×31) and 2500 non-target vectors (200×31).

B. Classical CNN model for P300 signal classification

The P300 EEG signal detection model based on Convolutional Neural Network (CNN) has significant advantages, which extracts spatiotemporal features through multi-layer convolutional kernels, captures the event-related potential waveforms of P300 within 0-800ms in the time dimension, and models the multi-electrode collaborative response in the spatial dimension, overcoming the limitations of traditional methods relying on artificial feature design. Combined with batch normalization and dropout strategies, EMG noise and baseline drift interference can be effectively suppressed, and the detection robustness in the environment with low signal-to-noise ratio is improved. Compared with other machine learning models, CNN has stronger local feature extraction ability, and can effectively capture the local time-domain features of P300[19].

Traditional CNN structures typically consist of input layers, multiple convolutional layers, and fully connected layers. The model employs convolutional layers for spatiotemporal feature learning, while classification is achieved through a fully connected layer with nonlinear activations. The classical CNN(C-CNN) used for EEG signal recognition of P300 uses different convolution kernels to extract temporal and spatial features respectively, and uses LeakyReLU as the activation function for classification, and the output layer outputs the predicted probability value corresponding to the corresponding target character recognition [20].

C. SEB-CNN model for P300 signal classification

At present, based on the traditional CNN model, there are still many shortcomings in EEG signal recognition, such as spatiotemporal feature coupling, category imbalance, training instability, etc., so in view of the problems of insufficient spatiotemporal feature coupling and insufficient category sensitivity of traditional convolutional neural network (CNN) in EEG decoding tasks, this paper proposes an improved model that integrates channel attention mechanism and multi-scale feature extraction (SEB-CNN). A Squeeze-and-Excitation Block (SEB) is added to the model structure to generate channel weights through global average pooling and two fully connected layers, which are then multiplied by the original features. By dynamically adjusting the importance of feature channels, the network automatically learns to "pay attention" to key features and suppresses redundant information, thereby improving the performance of the model. Dynamic learning rate scheduling adjusts the learning rate according to the verification loss, avoids falling into local optimum, and improves the convergence effect.

The model optimizes the CNN model to improve the P300 signal detection performance by the following design:

1) The lightweight channel attention module is introduced to dynamically enhance the spatiotemporal feature response of the target-related electrodes, and only a few parameters (two fully connected layers) are added, which can significantly improve the model's ability to model the channel relationship

2) We construct a multi-branch temporal convolution structure, which is compatible with neural response modes with different latency;

3) The joint dynamic learning rate scheduling and hierarchical regularization strategy were designed to improve the training stability.

The proposed model architecture is illustrated in Fig. 1.

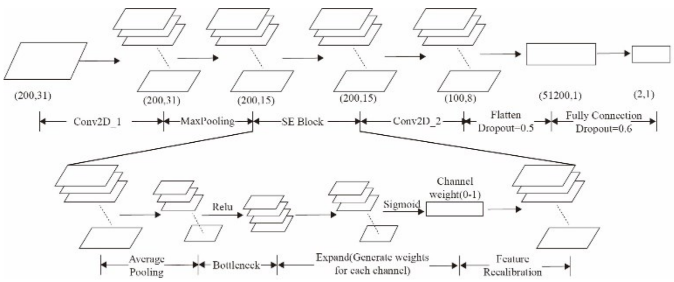

Fig.1. The SEB-CNN model structure for P300 EEG signal classification in P300 speller

The specific design of the model is as follows:

1) Data preprocessing

In the preprocessing phase, SMOTE oversampling was added to alleviate the problem of uneven data distribution and prevent model overfitting by adding P300 sample data by oversampling the problem of imbalance between P300 and non-P300 data. The oversampling formula is as follows:

$$x_{new} = x_i + \lambda \cdot (x_j - x_i) \quad (1)$$

Where: x_i: minority samples, x_j: randomly selected neighboring samples, $\lambda \in [0,1]$

2) Input layer

The input layer dimension is 200×31 (200 time sampling points, 31 channels)

3) The First Convolutional Layer

The first convolutional layer uses 32 6×6 convolutional kernels to extract spatiotemporal features, and the convolutional kernels are set to step 1 to maintain the spatial dimension (output 200×31).

The formula is as follows:

$$x_{out}(i,j,c) = \sum_{di=0}^{5}\sum_{dj=0}^{5}\sum_{k} x_{in}(i+di, j+dj, k) \cdot W(di,dj,k,c) + b(c) \quad (2)$$

Where: (H, W, Cin) represents the input feature map dimension, $(H, W, 32)$ represents the output feature dimension, W is the convolution kernel weight $(6 \times 6 \times C_{in} \times 32)$, and b is the bias term. And use batch normalization to accelerate convergence with the following formula:

$$\hat{x} = \frac{x-\mu}{\sqrt{\sigma^2+\epsilon}}, y = \gamma \cdot \hat{x} + \beta \quad (3)$$

Where, μ is defined as the mean of the batch data, σ^2 is the variance of the batch data, and γ and β are the learnable scaling and offset parameters, respectively.

The ReLU activation function was selected to enhance nonlinearity, and the formula for the ReLU is as follows:

$$f(x) = max(0, x) \quad (4)$$

4) Squeeze-and-Excitation Block (SEB)

A Squeeze-and-Excitation Block (SEB) is added between the two convolutional layers, which is a lightweight channel attention module that dynamically enhances the spatiotemporal feature response of target-related electrodes.

(i)Global Average Pooling

Squeeze compresses the spatial dimension through global average pooling, compresses the spatial dimension (H×W) of the feature map into a channel description vector to capture global information. The formula for global average pooling is as follows

$$z_c = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} x_c(i,j) \quad (5)$$

Where: (H, W, C) represents the input feature map dimension: C is the output vector, Each z_c represents the global information for the c-channel

(ii) The First fully connected layer (Squeeze)

We employ the first fully-connected layer reduces channel dimensionality and the formula is following:

$$h = W_1 z + b_1, W_1 \in R^{C/8 \times C}, b_1 \in R^{C/8} \quad (6)$$

Where, W_1 is the weight matrix, b_1 is the bias vector, and activated with the $ReLU$ activation function

The output channel mean vector (length 32) is compressed to 1/8 of the original channel number (32/8=4) through the first fully connected layer.

(iii) The Second fully connected layer (Excitation)

The second fully connected layer is used to recover the number of channels with the following formula:

$$s = W_2 h_{activated} + b_2, W_2 \in R^{C \times C/8}, b_2 \in R^{C} \quad (7)$$

The Second fully connected layer recovers the spatial dimension, uses the two fully connected layers to learn the nonlinear relationship between the channels, generates channel weights (scaling factors), emphasizes important channels, and suppresses redundant channels.

The weights are then normalized to [0,1] using the Sigmoid function, and the final result is the channel weight vector: $s_{scale} \in R$. Reshape the weight vector s_{scale} to (1, 1, *C*) to match the dimensions of the input feature plot X, and then multiply it channel by channel as follows:

$$\hat{X}(i,j,c) = X(i,j,c) \cdot s_{scale}(c), \forall i,j,c \quad (8)$$

The final output feature map is $\hat{X} \in R^{H \times W \times C}$, In this way, the model can dynamically adjust the importance of each channel, enhancing key features and suppressing redundant information.

5) The second convolutional layer

The second layer of the convolutional layer uses 64 6×6 convolution kernels, sets the step size to 2, realizes down sampling in time and space, and outputs data in 100×7. The formula is as follows:

$$x_{out}(i,j,c) = \sum_{di=0}^{5}\sum_{dj=0}^{5} \sum_{k=1}^{C_{in}} x_{in}(2i+di, 2j+dj, k) \cdot W(di,dj,k,c) + b(c) \quad (9)$$

W: Convolutional kernel weights （$6 \times 6 \times C_{in} \times 64$）

b: bias entry

The features are further refined through batch normalization and ReLU.

6) Flatten Layer

The Flatten Layer facilitates the expansion of these features into 51,200 dimensional vectors. The subsequent fully-connected layer facilitates the mapping of high-level features through a 128-cell fully-connected layer. Hierarchical overfitting suppression is achieved through the implementation of high dropout rates of 50% and 60%.

7) Output layer

The output layer uses SoftMax to output the binary classification probability and the Softmax activation function formula is as follows:

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (10)$$

$$x^5 = \sigma(\sum_{i=1}^{i\leq D_n}(x^4 \cdot W^5 + b^5)) \quad (11)$$

x^5 represents the final output and the P300 signal classification result is defined as:

$$E(X) = \begin{cases} 1 & x^9 \geq 0.5 \\ 0 & x^9 < 0.5 \end{cases} \quad (12)$$

Where, label "1" denotes target-evoked EEG signals, while label "0" indicates EEG signals from non-target responses.

In addition, the training model utilizes a joint dynamic learning rate scheduling approach with Stochastic Gradient Descent (SGD) with Momentum for learning rate optimization, as depicted by the following parameter update formula:

$$v_t = \gamma \cdot v_{t-1} + \eta \cdot \nabla_\theta J(\theta) \quad (13)$$

$$\theta_{t+1} = \theta_t - v_t \quad (14)$$

γ: momentum factor（0.9）

η: learning rate（10^{-3}）

III. Result

A. Model evaluation metrics

We assessed model performance through standard metrics: accuracy (Acc), precision (P), recall (R), and F1-score, and they are defined by:

$$Acc = \frac{TP+TN}{TP+FP+TN+FN} \quad (15)$$

$$P = \frac{TP}{TP+FP} \quad (16)$$

$$R = \frac{TP}{TP+FN} \quad (17)$$

$$F1 - score = 2\frac{P\times R}{P+R} \quad (18)$$

Classification Accuracy (Acc) measures the ratio of correct predictions (true positives + true negatives) to total samples, Precision quantifies the fraction of correctly identified positives among all instances classified as positive, and Recall (R) is the proportion of positive samples detected from true positives, where P and R affect each other. If the classification model works well, both metrics will rise at the same time, but it is often difficult to optimize them at the same time, and the other will decrease when one is high, so the F1-score is chosen as their composite metric.

B. Experimental results

The performance of the P300 signal recognition algorithm is the key to the character recognition of the P300 character speller, and the C-CNN and improved SEB-CNN models were used to evaluate the classification accuracy of P300 signal from the audiovisual bimodal (AV) speller and single visual (V) speller, respectively.

In both the AV and V spelling paradigm experiments, EEG signal recordings were obtained from nine participating individuals. For every participant across each experimental condition, a complete dataset of 25 characters spellings was acquired. To establish robust model evaluation, the collected data was strategically partitioned such that 20 characters trials were allocated for training, while the remaining 5 character were reserved exclusively for testing the classification performance.". According to the design of the regional experimental paradigm, the spelling of each character requires 10 flashes of target and non-stimulus, so the data of the training set and the test set are divided into 10 periods, corresponding to 1 stimulus flashing (1 sequence) to 10 cumulative flashings (10 sequences). In the stimulus presentation protocol, each flash event generates 240 training samples (20 characters × 12 flashing units, comprising 6 group regions and 6 subregions) and 60 test samples (5 characters × 12 flashing units). This sampling ratio persists throughout the experimental design. For practical P300 speller applications, enhancing classification accuracy under minimal stimulus sequences represents a critical performance metric that directly impacts system usability, therefore when comparing the classification performance of C-CNN and SEB-CNN, we focused on the performance indicators when stimuli were superimposed 1, 2, 3, 4, 5, and 10 times. The results of the comparison are shown in Fig. 2.

Fig. 2. shows the average accuracy, average precision, average recall and average F1-score of the EEG data of 9 subjects in the AV paradigm and the V paradigm under different superimposed times (sequences). The graphical representation utilizes distinct lines-style to denote different performance metrics across four subplots: mean classification accuracy (upper left), mean precision values (upper right). mean recall rates (lower left) and average F1-scores (lower right) for P300 signals from AV and V paradigms based on the SEB-CNN model, respectively. The red dotted line and the blue dotted line represent the four performance indicators on two EEG data based on the C-CNN model.

As illustrated in the figure, the proposed model (solid line) demonstrates significant improvements in mean accuracy, precision, recall, and F1-score for P300 signal classification compared to the C-CNN baseline (dotted line) on the sequence from AV paradigm and V paradigm from 2 to 10.

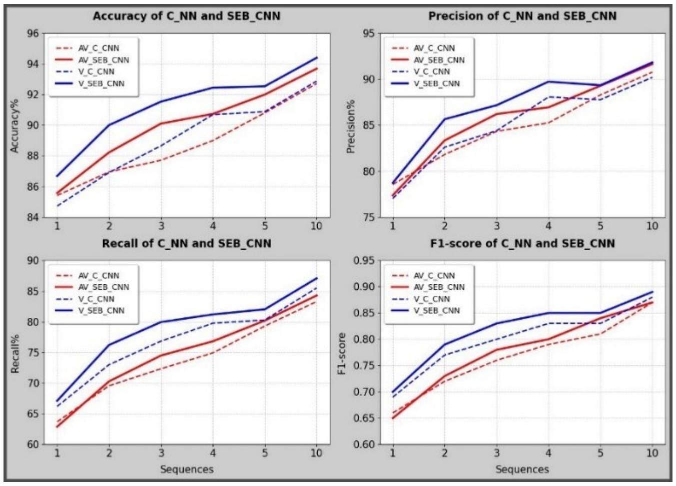

Fig. 2. The average accuracy, average precision, average recall, and average F1-score of P300 EEG signal based on the C-CNN and SEB-CNN models in AV speller and V speller

The most obvious increase is the classification accuracy. In terms of the number of superimpositions, the accuracy of P300 signal classification based on the SEB-CNN model increases significantly more than that of other superimpositions when the superposition is 2, 3 and 4 times.

Table I shows the classification accuracy of P300 EEG signals at superimposed 3 times in AV paradigm and at 2 times in V paradigm for 9 subjects. The comparisons in Table I reveal the SEB-CNN's accuracy advantage in P300 EEG signal classification relative to the C-CNN model when 9 subjects from the V spelling paradigm are superimposed 2 times, and the accuracy of P300 EEG signal classification for 8 subjects from the AV spelling paradigm is greater than that of the C-CNN model when stacking 3 times, and the P300 classification accuracy of S1 in the AV paradigm is 6% higher than that of the P300 classification based on C-CNN. In the V paradigm,

TABLE I
COMPARATIVE CLASSIFICATION ACCURACY OF P300 EEG SIGNALS: EVALUATING C-CNN VS. SEB-CNN MODELS ACROSS AV AND V SPELLER PARADIGMS IN A 9-SUBJECT STUDY

Acc% / Subjects	AV on 3 sequences		V on 2 sequences	
	C-CNN	SEB-CNN	C-CNN	SEB-CNN
1	**85**	**91**	89.17	89.2
2	85.33	86.1	90.83	93.3
3	89.44	92.2	87.5	89.2
4	86.67	90.6	85.83	87.5
5	90	93.8	**82.22**	**89.2**
6	88.75	90.6	89.17	90.8
7	85	87.2	86.67	92.5
8	90	89.4	90.83	93.3
9	89.17	90	80	85
Avg	**87.71**	**90.10**	**86.91**	**90.00**

the P300 classification accuracy of S5 based on SEB-CNN was 6.98% higher than that of P300 based on C-CNN.

We evaluated the significance of classification accuracy differences between C-CNN and SEB-CNN models across V and AV paradigms using paired t-tests (Tables II-III). Table 2 demonstrates that in the V paradigm, SEB-CNN achieves significantly superior P300 classification accuracy compared to C-CNN ($P<0.05$) across 2-10 stimulus repetitions, with peak significance at 3 repetitions ($P<0.005$). Consistent with these findings, Table 3 demonstrates that the SEB-CNN model also achieves significantly superior classification accuracy for P300 EEG signals in the AV paradigm compared to the C-CNN model across 2–10 stimulus repetitions ($P < 0.05$). Notably, the performance difference reaches its highest statistical significance during the 3rd and 4th times superimposing ($P < 0.01$).

TABLE II
T-TEST OF CLASSIFICATION ACCURACY OF V-P300 EEG DATA BY C-CNN AND SEB-CNN MODELS

	Superposition times					
(C-CNN, SEB-CNN)	1	2	3	4	5	10
t	-1.22	-4.02	-5.86	-2.50	-2.95	-4.80
P	0.2557	**0.0039**	**0.0004**	**0.0371**	**0.0184**	**0.0013**

TABLE III
T-TEST OF CLASSIFICATION ACCURACY OF AV-EEG DATA BY C-CNN AND SEB-CNN MODELS

	Superposition times					
(C-CNN, SEB-CNN)	1	2	3	4	5	10
t	-0.14	-2.39	-3.60	-4.22	-2.10	-2.48
P	0.890	**0.044**	**0.007**	**0.003**	**0.069**	**0.038**

IV. DISCUSSION AND ANALYSIS

A. *SEB-CNN model analysis*

SEB-CNN adds a Squeeze-and-Excitation Block between two convolutional layers, which has significant advantages in deep learning model design, especially in tasks that require dynamic feature selection and efficient computation (such as EEG signal processing, image classification, etc.). It can adaptively adjust the weights of each channel, amplify key features, suppress noise channels, and can automatically learn contribution weights without manual presets, adapt to individual differences, and be more powerful than traditional neural networks in terms of more discriminative features in learning. In the design of the model, the P300 EEG signals of different individuals were considered in time and space, and the adaptive ability and generalization ability of the model were enhanced by introducing spatial convolution and temporal attention mechanisms. The SE Block realizes channel - wise attention by means of global pooling and two fully - connected layers. It has the advantage of introducing a very small number of additional parameters. Moreover, the computational burden brought by the global average pooling and the fully - connected layers is far lower than that of convolution operations, making them suitable for embedding in mobile or real-time systems (e.g., real-time EEG signal processing). The channel adaptability of SEB-CNN has obvious advantages, SE Block uses the weight vector (between 0~1) generated by Sigmoid, and the SEB-CNN convolutional layer adopts multi-scale time convolution to extract features in space and time, which further improves the performance of the model.

B. Classification performance of P300 EEG signal

In the P300 spelling system, the EEG signals obtained usually have a low signal-to-noise ratio, so the stimulation needs to be repeated multiple times to enhance the robustness of character spelling. Nevertheless, an increase in the number of stimulus repetitions leads to a reduction in spelling speed. Consequently, the critical problem for optimizing the character spelling system lies in how to improve the detection accuracy of P300 with fewer superimpositions. In the AV paradigm and V paradigm, the P300 EEG signal classification results show that the SEB-CNN model achieves markedly superior performance over the C-CNN model, with significantly higher scores in classification accuracy, precision, recall, and F1-measure when the number of superimpositions is 2, 3, 4 and 5 times, which indicates that the P300 EEG signal classification performance of the SEB-CNN-based model is improved when the number of superimpositions is small. However, the classification performance of P300 EEG signals was not significantly improved when superimposed once, and its precision, recall and F1-score even decreased in the AV paradigm. The reason may be that when the number of stimulus superposition is one time, it is greatly affected by noise, and in the AV spelling paradigm, it may also be affected by other inducing components [21-22].

V. Conclusion

This paper introduces a Squeeze-and-Excitation Block-based CNN (SEB-CNN) model integrating channel attention mechanism and multi-scale feature extraction based on the traditional CNN model for P300 signal recognition in the P300 character speller. The model employs global average pooling followed by two fully connected layers to generate channel-wise attention weights. This mechanism enables the network to dynamically recalibrate channel-wise feature responses, automatically emphasizing discriminative features while suppressing less informative ones through learned attention weights, thereby improves the model performance. And uses dynamic learning rate scheduling to adjust the learning rate according to the verification loss, avoid falling into local optimum, and enhances model convergence. Experimental results validate that the proposed method achieves superior classification performance compared to the traditional CNN model on both types of data. We focused on the classification accuracy of the model when the number of stimuli superimposed was small, and the results showed that the classification accuracy of the SEB-CNN model was significantly improved compared with the C-CNN model.

Acknowledgment

This work was financially supported by the Natural Science Foundation of the Shandong Province Program (grant number: ZR2022MF287), and Youth Fund Project of Humanities and Social Sciences Research of the Ministry of Education (grant number 24YJCZH183).

References

[1] L. Carelli, F. Solca, A. Faini et al., "Brain-computer interface for clinical purposes: Cognitive assessment and rehabilitation," *Biomed. Res. Int.*, vol. 2017, Art. no. 1695290, 2017.

[2] U. Chaudhary, N. Birbaumer, and A. Ramos-Murguialday, "Brain-computer interfaces for communication and rehabilitation," *Nat. Rev. Neurol.*, vol. 12, no. 9, pp. 513–525, 2016.

[3] Y. Wang, M. Nakanishi, and D. Zhang, "EEG-based brain-computer interfaces," *Adv. Exp. Med. Biol.*, vol. 1101, pp. 41–65, 2019.

[4] J. Polich and A. Kok, "Cognitive and biological determinants of P300: An integrative review," *Biol. Psychol.*, vol. 41, no. 1, pp. 103–146, Oct. 1995.

[5] S. Sutton, P. Tueting, J. Zubin, and E. R. John, "Information delivery and the sensory evoked potential," *Science*, vol. 155, no. 3768, pp. 1436–1439, Mar. 1967.

[6] L. A. Farwell and E. Donchin, "Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials," Electroencephalogr. *Clin. Neurophysiol.,* vol. 70, no. 6, pp. 510–523, 1988.

[7] S. M. Hossain, H. M. Al-Hammadi, and G. Muhammad, "Automatic fruit classification using deep learning for industrial applications," *IEEE Trans. Ind. Informat*., vol. 15, no. 2, pp. 1027–1034, 2019.

[8] I. Ahmed, S. Din, G. Jeon et al., "Exploring deep learning models for overhead view multiple object detection," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 5737–5744, 2019.

[9] H. Cecotti and A. Graser, "Convolutional neural networks for P300 detection with application to brain-computer interfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 433–445, Mar. 2011.

[10] M. F. Liu, W. Wu, Z. H. Gu, Z. L. Yu, F. F. Qi, and Y. Q. Li, "Deep learning based on batch normalization for P300 signal detection," *Neurocomputing,* vol. 275, pp. 288–297, Jan. 2018.

[11] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," *in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, *Salt Lake City, UT, USA*, 2018, pp. 7132–7141.

[12] J. Park, S. Woo, J. Y. Lee et al., "BAM: Bottleneck attention module," *in Proc. Brit. Mach. Vis. Conf. (BMVC), Newcastle, UK*, 2018, pp. 1–14.

[13] S. Woo, J. Park, J. Y. Lee et al., "CBAM: Convolutional block attention module," *in Proc. Eur. Conf. Comput. Vis. (ECCV), Munich, Germany,* 2018, pp. 3–19.

[14] H. Zhang, K. K. Zu, J. Lu et al., "EPSANet: An efficient pyramid squeeze attention block on convolutional neural network," *in Proc. 16th Asian Conf. Comput. Vis. (ACCV), Macao, China,* 2023, pp. 541–557.

[15] K. Barmpas, Y. Panagakis, S. Bakas et al., "Improving generalization of CNN-based motor-imagery EEG decoders via dynamic convolutions," *IEEE Trans. Neural Syst. Rehabil. Eng.,* vol. 31, pp. 1997–2005, 2023.

[16] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, 2020, DOI: 10.1109/TPAMI.2019.2913372.

[17] Z. Lu, Q. Li, N. Gao, J. Yang, and O. Bai, "A novel audiovisual P300-speller paradigm based on cross-modal spatial and semantic congruence," *Front. Neurosci*., vol. 13, p. 1040, 2019.

[18] R. Fazel-Rezai, S. Gavett, W. Ahmad et al., "A comparison among several P300 brain-computer interface speller paradigms," *Clin. EEG Neurosci.,* vol. 42, no. 4, pp. 209–213, 2011.

[19] G. Dai, J. Zhou, J. Huang, and N. Wang, "HS-CNN: A CNN with hybrid convolution scale for EEG motor imagery classification," *J. Neural Eng.*, vol. 17, no. 1, Art. no. 016025, 2020.

[20] L. Xu, Z. Ma, J. Meng, M. Xu, T. P. Jung, and D. Ming, "Improving transfer performance of deep learning with adaptive batch normalization for brain-computer interfaces," *in Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC),* 2021, pp. 5800–5803.

[21] D. J. Krusienski, M. Grosse-Wentrup, F. Galán et al., "Critical issues in state-of-the-art brain-computer interface signal processing," *J. Neural Eng.,* vol. 8, no. 2, Art. no. 025002, Apr. 2011.

[22] F. Lotte, M. Congedo, A. Lecuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 4, no. 2, pp. R1–R13, Jun. 2007.