Fusion of Deep Convolutional Neural Networks and Brain Visual Cognition for Enhanced Image Classification

Xintao Li[1, *], Hongyan Guo[2]
[1]College of Innovation and Entrepreneurship, Henan Open University, Zhengzhou 450046, China
[2]School of Information Engineering and Artificial Intelligence, Zhengzhou Vocational University of Information and Technology, Zhengzhou 450046, China
*Email of Corresponding Author: lxt5168@163.com

Keywords: deep convolutional neural network, brain, visual cognition, intelligent computing model, image classification

Received: December 9, 2024

The brain visual system is one of the core centers for human perception of external information. How to establish the brain visual cognitive system to classify and process image information is a key matter in the area of human-computer connection. In order to improve the accuracy of computer vision image classification, a fusion intelligent computing model based on deep convolutional neural network and brain visual cognition is proposed. This model simulates the visual processing mechanism of the human brain and uses brain computer interface technology to extract electroencephalogram signals, thereby achieving efficient classification and processing of image information. When designing an image classification model based on DCNN, a long short-term memory network structure is introduced to extract time series features of electroencephalogram signals. In order to enhance the classification accuracy of the model, attention mechanism and occlusion independent neural response methods are also applied to improve the accuracy of capturing the correlation information between brain response and image features. The results show that the prediction accuracy of the research model reaches 93.54% and 94.03% in the V4 visual region and L0 visual region, respectively. The highest accuracy on facial visual images reaches 95.46%, while the lowest accuracy on animal visual images is 91.57%. By introducing the long short-term memory module, the loss value of the model decreases from 0.26 to 0.21, with a reduction of 19.23%. In addition, ablation experiments show that by introducing attention mechanisms and occlusion independent neural responses, the final classification accuracy is improved to 93.94%. In summary, the research on the fusion intelligent computing model grounded on deep convolutional neural networks and brain visual cognition effectively improves the accuracy of image classification and demonstrated its potential in the field of intelligent computing.

Povzetek: Predstavljen je inteligentni model za razvrščanje slik, ki združuje globoke konvolucijske nevronske mreže (DCNN) in možgansko vizualno kognicijo preko EEG signalov.

1 Introduction

With the rapid prosperity of artificial intelligence, human-computer interaction has turned into a trend in the current research field. Brain computer interface (BCI), as a cutting-edge scientific research direction, is gradually becoming a meaningful bridge in the area of human-computer connection. The visual system of the human brain has evolved over millions of years and possesses extremely efficient visual processing capabilities. Through multi-level visual processing mechanisms, the brain can quickly and accurately understand complex visual information [1]. When external objects are transmitted to the visual center of the brain through the visual organs, the brain quickly recognizes, classifies, and understands these visual information, thereby forming cognition of the object or scene [2]. BCI can interpret visual cognition of the brain by recording and analyzing electroencephalogram (EEG) signals [3]. The Deep Convolutional Neural Network (DCNN) in computer vision technology has attracted much attention due to its outstanding performance in image processing tasks [4]. sHowever, despite the excellent performance of computer technology in image classification, computers still cannot fully replace the precise image recognition and classification capabilities of the human brain in complex and diverse open environments with interference and occlusion [5]. So, the challenge currently facing the field of computer vision is figuring out how to empower artificial intelligence systems to more effectively mimic human brain cognition and attain precise image classification in intricate scenarios. Therefore, in this context, research innovatively combines the powerful computing power of DCNN with the cognitive characteristics of the brain's visual system, and constructs an intelligent computing model based on the fusion of DCNN and brain visual cognitive information, in order to achieve accurate image classification in complex backgrounds.

The research objectives include designing and implementing an intelligent computing model based on DCNN and EEG signal fusion to improve the performance of image classification in interference and occlusion environments. The research aims to explore how the model simulates the visual recognition process of the human brain, especially for accurate image classification in complex backgrounds. The research hypothesis is that by combining the visual feedback and image features of the brain, intelligent computing models can simulate the visual recognition process of the human brain, thereby improving the accuracy of classification results. The preset results demonstrate that by introducing visual cognitive information from the brain, the model can mimic the cognitive process of the human brain in actual visual tasks, providing new ideas and directions for the integration of BCIs and intelligent systems.

The research content mainly includes four sections. The second section provides a survey of the current study status of visual EEG picture classification and DCNN around the world. The third section conducts research on intelligent computing models that integrate DCNN and brain visual cognition. The first section proposes the design of a picture sorting model grounded on the fusion of DCNN and brain visual cognition information. The second section designs an intelligent computing model based on the fusion of DCNN and brain visual cognitive information. The fourth section validates the intelligent computing model that integrates DCNN with brain visual cognition.

2 Related works

The visual cognitive ability of the brain can recognize, classify, and understand visual information. In recent years, research on visual interpretation based on monitoring the neural response of the brain during visual cognition has gained the eyesight of numerous professionals and savants. Gao et al. raised an attention-based parallel multi-scale Convolutional Neural Network (CNN) model to improve the accuracy of decoding EEG aroused potentials. The model used two parallel convolutional layers to extract temporal features and utilized attention mechanisms to weight features at different times. The outcomes revealed that the model effectively reformed the interpreting ability of ocular aroused potentials under complex conditions [6]. Ahirwal et al. proposed a new channel selection technique that could identify and characterize harmful emotions aiming to raise the precision of emotion sorting of EEG signals. This technique extracted three forms of characteristics from EEG cues: time-domain characteristics, frequency-domain characteristics, and entropy based characteristics, and used Support Vector Machines (SVM) and artificial neural networks to classify emotions based on the extracted features. The outcomes showed that this way effectively optimized the sorting behaviour [7]. Komolovait et al. raised a way of using CNN combined with stable-state ocular aroused potentials to gain interpretable characteristics from rough EEG cues in order to improve the effectiveness of brain activity data in classifying visual stimuli. This method also introduced generative adversarial networks and variational autoencoders to produce composite EEG cues. The results showed that the method was effective [8]. Kumari et al. proposed a multi-channel EEG movement sorting model to improve the precision of EEG movement sorting. The model utilized CNN to extract descriptive emotional state characteristics from EEG signals and generates two-dimensional images to represent these features. The outcomes revealed that the overall precision of this model reached 83.04% [9].

DCNN occupies a momentous position in EEG picture sorting tasks. Santamaria-Vazquez et al. raised a sorting model grounded on different control signals to extract complex features from EEG data for classification. The model used DCNN for time calibration of BCIs and integrated modules for detection of event-related potentials. The outcomes revealed that the command decoding accuracy of this way improved by 16.0% [10]. Yıldırım et al. raised a novel deep one-dimensional CNN monitoring model to optimize the precision of EEG monitoring. The model utilized machine learning techniques to automatically identify regular and aberrant EEG signals, and classified EEG signals using an end-to-end structure. The outcomes revealed that this way was feasible [11]. Miao et al. raised a multi-layer CNN model using a DCNN structure to raise the classification precision of EEG pattern identification algorithms. The model utilized prior knowledge and complex parameter adjustments to extract spatial frequency features. The outcomes showed that this way had good classification capability [12]. Li et al. proposed a way of using DCNN combined with continuous wavelet transform to enhance the identification rate of limbs action image EEG cues. This method mapped the limbs action image EEG cues to time-frequency image signals using continuous wavelet transform, and input the image signals into the CNN structure to collect characteristics and classify them. The outcomes revealed that this way effectively raised the recognition rate [13]. In recent years, the combination of BCI and DCNN has become an important research direction in the analysis of EEG and brain visual neural activity signals. The detailed progress of BCI is as follows: Tang X et al. proposed an end-to-end BCI method based on CNNs, which directly extracts spatiotemporal features from EEG signals and classifies them. The results showed that this method could achieve higher classification accuracy than traditional manual feature extraction methods, especially in motion imagination tasks and various emotional state classification tasks [14]. In addition, Kawala Sterniuk et al. reviewed over 50 years of using BCIs and concluded that BCI not only enables brain control, but also opens the door for regulating the central nervous system through neural interfaces, demonstrating the potential applications of this technology [15]. The research on integrating BCI and DCNN will provide a more solid foundation for the popularization and application of BCI technology. The comparative summary table is shown in Table 1.

Table 1: Comparison summary table

Study	Method	Advantages	Limitations	Missing features
Gao et al. [6]	Parallel multi-scale CNN based on attention	Improved the decoding performance of visual evoked potentials	Still affected by noise in complex environments, requires processing a large amount of temporal features	Failed to effectively combine spatial and temporal features in the brain's visual cognitive process; cannot adapt to complex environmental visual information processing
Ahirwal et al. [7]	EEG-based emotion classification model, combining SVM and artificial neural networks	Improved emotion classification accuracy	Focuses mainly on emotion classification, lacks deep classification and processing of visual information	Cannot process complex visual information and its complex relationship with emotions
Komolovaitė et al. [8]	Steady-state visual evoked potentials combined with CNNs	Effectively improved visual stimulus classification	Poor robustness to signal noise, high complexity in training generative adversarial networks	Failed to effectively combine visual cognitive mechanisms; limited to static visual stimulus processing
Kumari et al. [9]	Multi-channel EEG-based emotion classification model	Achieved an average accuracy of 83.04%	Focuses on emotion classification, mainly uses image feature representations, lacks handling of more complex scenarios	Cannot handle complex image classification tasks, especially multi-class image recognition
Santamaria-Vazquez et al. [10]	Classification model based on different control signals using DCNNs	Increased command decoding accuracy by 16.0%	Relies heavily on event-related potential detection, may face difficulties in decoding complex EEG data	Lacks adaptability to dynamic EEG signals, unable to combine spatial and temporal features
Yıldırım et al. [11]	EEG monitoring model based on deep one-dimensional CNNs	Provides a feasible classification method	Focuses on normal vs abnormal EEG signal classification, lacks ability to handle complex visual tasks	Cannot effectively process multi-class or dynamically changing visual information
Miao et al. [12]	EEG pattern recognition based on multi-layer DCNNs	Shows good classification performance	Mainly focuses on spatial frequency feature extraction, may be limited in handling complex dynamic tasks	Lacks comprehensive capture of dynamic EEG data or multi-dimensional features of visual information
Li et al. [13]	Classification of left/right hand motor imagery EEG signals combined with continuous wavelet transform and DCNNs	Significantly improved recognition rate	Relies on signal preprocessing, suitable for specific tasks	Cannot process EEG signals related to visual tasks, sensitive to environmental noise

In summary, although existing methods have made some progress in EEG classification tasks, they have certain limitations in handling complex dynamic tasks, enhancing robustness, and adapting to multiple tasks. The research combines the visual cognitive mechanism of the brain with DCNNs and Long Short-Term Memory (LSTM) networks to design a fusion intelligent computing model. This model can more comprehensively capture the spatial and temporal features of EEG signals, solving the problems of lack of adaptability and poor adaptability to complex environments in existing methods. It has higher classification accuracy and wide application prospects.

3 Intelligent computing model integrating DCNN and brain visual cognition

Research receives EEG information through BCI, combines voxel encoding and improved DCNN model to achieve image classification, and uses LSTM to collect temporal characteristics of EEG cues. Attention mechanism is utilized to raise the accuracy of image feature extraction, and the correlation between brain response and image features is enhanced by masking irrelevant neural responses.

3.1 Design of image classification model based on DCNN and brain visual cognitive information

Neuroscience research has found that the human brain achieves complex cognitive processing through parallel information exchange between dorsal and ventral streams in visual activities [16]. Abdominal flow is a pathway that connects the primary sensory cortex with the temporal and prefrontal regions, primarily responsible for recognizing visual and auditory stimuli and mapping basic information to higher-level semantic concepts [17]. The dorsal flow is responsible for spatial information and motion control. The activity of brain neurons triggered by visual stimuli is called EEG signals, and BCIs can record and measure these signals through biometric technology to reflect the brain's response to behavior. The core area of ventral flow includes the primary visual cortex, ventral intermediate cortex, ventral lower temporal cortex, and other regions. The ventral lower temporal cortex is particularly closely related to complex visual recognition and is the main functional area for object and face recognition. When the brain receives visual stimuli, it stimulates the cortical regions in the ventral stream, transforming simple visual features into higher-level cognitive concepts. For instance, visual information is initially processed by the primary visual cortex and then passed through intermediate areas, ultimately being mapped to the inferior temporal cortex within the ventral stream, where intricate functions like object recognition and color discrimination take place. The dorsal flow is mainly responsible for processing spatial information, motion perception, and action control. Dorsal flow helps the brain perform functions such as object localization, motion tracking, and hand eye coordination through connections with the parietal lobe, motor cortex, and other areas. Therefore, given the core role of ventral flow in image classification tasks, research focuses on analyzing the brain signal response of ventral flow to better understand the process of visual feature extraction and semantic comprehension. The encoding framework for ventral response based on brain visual cognition is shown in Figure 1.

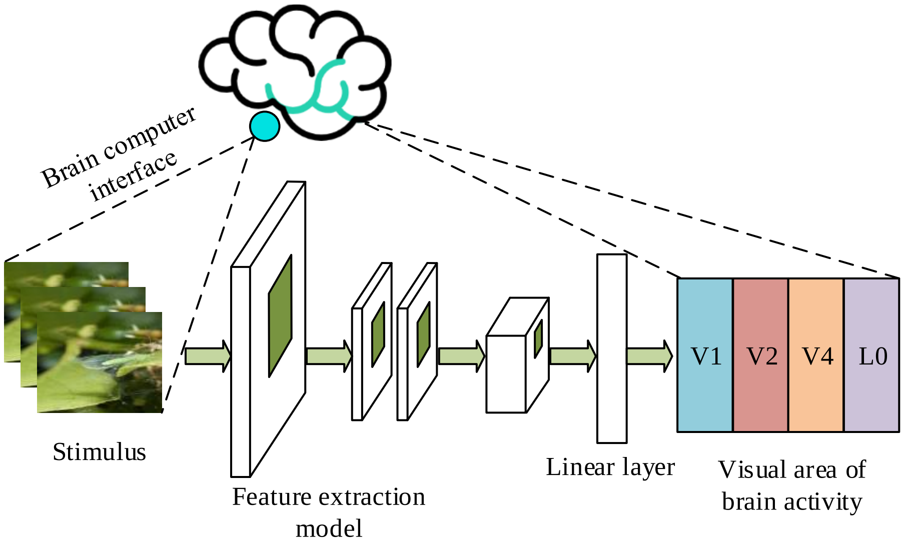


Figure 1: A coding framework for ventral response based on brain visual cognition

As shown in Figure 1, in the ventral response encoding framework based on brain visual cognition, the brain activity caused by visual stimuli can be obtained through the BCI, and the stimulus image can be input into the feature extraction model. After nonlinear calculation, the feature space of the image can be obtained. Then, these features are used to predict the voxel space of the visual region through linear layers. The voxel encoding model transforms human-readable data into a format that machines can store, facilitating the achievement of either shared encoding across various visual regions or unique encoding for specific visual areas. This process aids in pinpointing the regions within the brain's visual cortex that are responsible for processing visual information. [18]. Voxel encoding converts brain activity into a feature space, enabling precise association between cognitive responses and visual stimuli. This mapping helps to reveal the roles of different brain regions in visual information processing, thereby enhancing the accuracy of image classification tasks. The EEG signals capture the electrical activity of the cerebral cortex, which can be mapped to specific regions of the brain through modeling techniques such as source localization, in order to infer activity responses in different areas. This type of method can correlate the spatiotemporal patterns of EEG signals with voxels in functional neuroimaging data. There may be some common neural response patterns between multiple visual regions. These shared response patterns can be captured in voxel encoding models, revealing how these regions collectively respond to the same visual stimuli. For example, in image classification tasks, certain visual regions may exhibit similar neural activity responses to the same visual features, so voxel encoding can reflect the similarity and interactivity between these regions as a shared encoding pattern. By combining the results of brain visual cognition and image classification, complementary information exchange and expression can be achieved, thereby obtaining a more comprehensive joint representation. DCNN can automatically learn image feature representations by combining convolutional and pooling layers, which helps extract abstract features from data. Therefore, the study adopts DCNN to extract image features and designs a picture sorting model grounded on the fusion of DCNN and brain visual cognitive information, as shown in Figure 2.

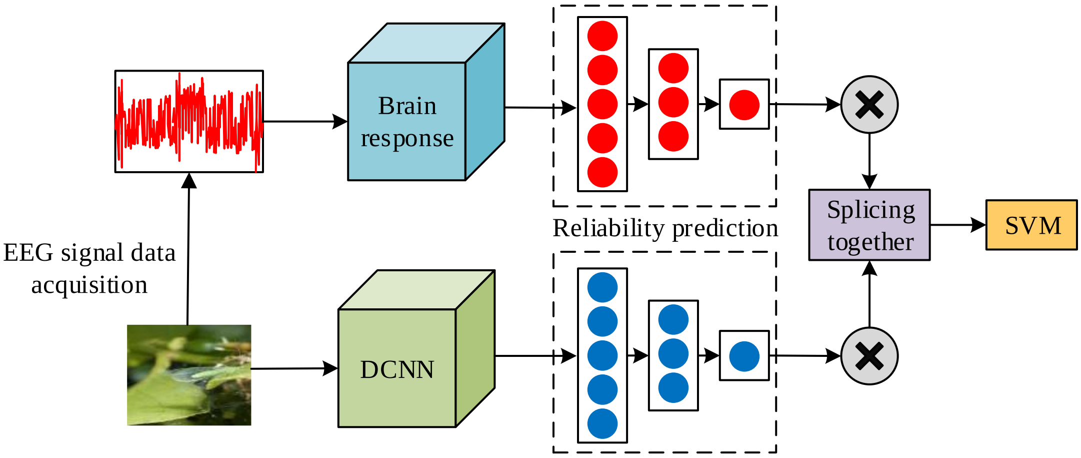


Figure 2: Image classification model grounded on information fusion

As represented in Figure 2, the image classification model based on information fusion mainly includes three parts: characteristic collection structure, characteristic reliability prediction structure, and brain computer information fusion classification structure. The brain response data utilizes the ventral response encoding framework to extract semantic features, while image data is extracted through the DCNN structure for image characteristic collection. Next, the extracted characteristics are input into the feature reliability prediction structure for reliability calculation, and then the fusion weights of image features and brain response characteristic are automatically adjusted. Finally, the fused features are input into the SVM for classification. EEG signals will undergo denoising processing after acquisition, such as bandpass filtering, independent component analysis, signal normalization, and other preprocessing operations to ensure signal quality. Subsequently, EEG signals will be synchronized with the presentation time of visual stimuli to ensure accurate matching between brain responses and image features at each moment. The loss function for reliability prediction is shown in equation (1).

$$L_{MSE} = \sum_{n=1}^{n} (d_p^{'} - d_{f_b}^{'})^2 / n \tag{1}$$

In equation (1), L_{MSE} means the loss function of reliability prediction, $d_p^{'}$ represents the feature reliability prediction value, $d_{f_b}^{'}$ represents the classification sensitivity index of brain response features, and n represents the batch size. The fusion weights of image features are shown in equation (2).

$$w_v = d_{f_v}^{'} / (d_{f_b}^{'} + d_{f_v}^{'}) \tag{2}$$

In equation (2), w_v represents the fusion weight of image features, and $d_{f_v}^{'}$ represents the classification sensitivity index of image features. The fusion weight of brain response features is shown in equation (3).

$$w_b = d_{f_b}^{'} / (d_{f_b}^{'} + d_{f_v}^{'}) \tag{3}$$

In equation (3), w_b represents the fusion weight of brain response features. The math description for the fusion feature is represented in equation (4).

$$f_F = (w_b \times f(b)) concat (w_v \times f(v)) \tag{4}$$

In equation (4), f_F represents fusion features, $f(b)$ represents brain response features, and $f(v)$ represents image features. Due to the fact that EEG signals are collected over a continuous period of time and have time-series characteristics, there is a continuity relationship between the signals at each moment and those before and after [19]. However, although existing feature extraction models perform well in many application scenarios, they often do not fully consider the temporal dependencies in time series data. Especially when it comes to traditional models like CNNs, while they excel at extracting features from images and static data, they can fall short when it comes to capturing temporal information and dynamic signal changes in time-series data, such as EEG signals. In response to this situation, the study uses an LSTM structure to extract time series features of EEG signals. The architecture for extracting brain response features based on time series is shown in Figure 3.

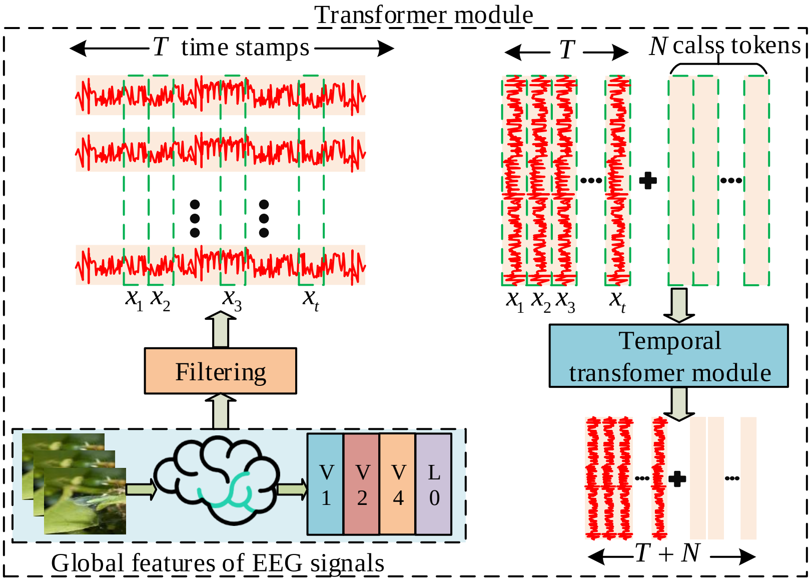


Figure 3: Architecture for extracting brain response features based on time series

As shown in Figure 3, the brain response feature extraction architecture based on time series uses Transformer module to extract global features of EEG signals on time series, and embeds absolute positions to maintain the order of the model. Before inputting positional encoding, the classification identification bits are concatenated with the time series, and then mapped through linear transformation to raise the diversity of characteristic collection. In research models, LSTM is mainly used to integrate brain response data collected from BCIs. The integration process is as follows: Firstly, the brain response signals collected from the BCI system are preprocessed, such as denoising and normalization, to obtain clean time-series data. Then, these preprocessed brain response data are used as inputs for the LSTM network. LSTM networks can capture temporal dependencies in data and learn neural response patterns of the brain at different time points. Next, through the time-dependent modeling of LSTM, the output data contains the gradual response patterns of the brain to visual stimuli throughout the entire image processing process. Finally, the temporal response of the brain is processed by LSTM and combined with image features extracted by DCNN. The calculation for the forget gate of LSTM structure is shown in equation (5).

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{5}$$

In equation (5), f_t means the output of the forget gate, W_f means the weight of the forget gate, σ means the Sigmoid activation function, b_f represents the offset term of the forget gate, x_t represents the input signal at time t, and h_{t-1} represents the output signal at time $t-1$. The unit update calculation is shown in equation (6).

$$C_t = f_t \cdot C_{t-1} + i_t \cdot C_t \tag{6}$$

In equation (6), C_t means the cell condition at time t, C_{t-1} represents the cell condition at time $t-1$, and i_t represents the output matrix of the input gate. The calculation for the output gate is shown in equation (7).

$$o_t = \sigma\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right) \tag{7}$$

In equation (7), o_t represents the output gate, W_o means the weight of the output gate, and b_o means the offset term of the output gate. The output features are shown in equation (8).

$$h_t = o_t \cdot \tanh(C_t) \tag{8}$$

In equation (8), h_t represents the output feature. The unique gating mechanism of LSTM can effectively handle the problem of long time intervals and delays in time series, and can discard and store large-span information in EEG data, thus better encoding EEG signals.

3.2 Design of intelligent computing model based on DCNN and brain visual cognitive information

The study simulates the connectivity and classification patterns of biological brain neurons, exploring the connection between picture features and brain reactions. DCNN has demonstrated significant capabilities in image feature extraction. By combining convolutional and pooling layers, it can automatically learn multi-level abstract feature representations of images, effectively capturing low-level and high-level features in images. However, despite DCNN's high efficiency in feature extraction, the image features it extracts still struggle to fully explain the brain's response patterns. This is because the visual cognitive process of the brain not only relies on low-level visual features of images, but also involves complex high-level semantic information processing, perceptual integration, and interaction with other cognitive processes such as memory and emotion. The features extracted by DCNN mainly focus on significant visual information in the image, but these features often lack sufficient high-level semantic depth and are difficult to fully integrate with the complex responses of the brain in visual cognition. Therefore, the picture characteristics extracted by DCNN are difficult to fully explain the representation information of brain response, and there are some modality specific expressions between the two, making it difficult to deeply explore their deep correlations [20, 21]. In response to this limitation, a network structure based on DCNN is studied to construct an intelligent computing model for the brain. The brain response is used as supervised information for images, and difficult to interpret high-level semantic information is transferred to the DCNN model to achieve more accurate mining of the brain's visual cognitive response. The intelligent computing model structure grounded on data fusion is represented in Figure 4.

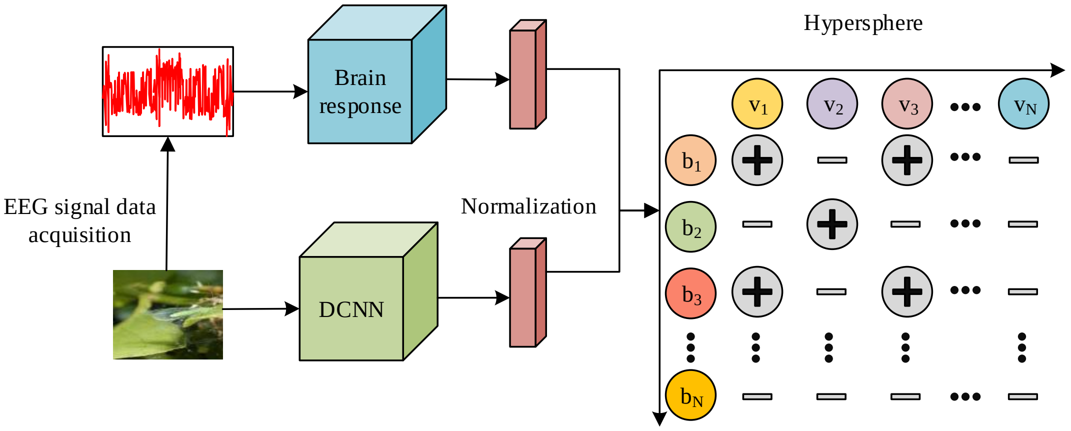


Figure 4: Intelligent computing model structure grounded on data fusion

As represented in Figure 4, in the intelligent computing model framework based on information fusion, characteristic extraction is first constructed on the cognitive response data of the brain to visual images collected by the BCI. Then, the DCNN structure is applied to collect characteristics from the input picture. After normalizing the two extracted features separately, the fused features are mapped onto an N-dimensional sphere. Subsequently, based on the normalized features, a set of positive and negative samples are constructed, and the InfoNCE loss function is used for calculation, thereby achieving the transfer of correlated information between the two feature maps. The math description of the InfoNCE loss function is represented in equation (9).

$$L_i = -log \frac{exp(S(z_i, z_i^+)/\tau)}{\sum_{j=0}^{N} exp(S(z_i, z_j)/\tau)} \qquad (9)$$

In equation (9), L_i represents the InfoNCE loss function, τ represents the temperature coefficient, z_i represents the image representation corresponding to the input data x_i, $S(z_i, z_j)$ represents the cosine similarity between image representations, $S(z_i, z_i^+)$ represents the alignment characteristics during hypersphere mapping, and N represents the total amount of positive and negative samples. The calculation for the comparative loss is represented in equation (10).

$$L_i' = -log \frac{\sum_{j=0}^{m} exp(S(f(v_i), f(b_j^+))/\tau)}{\sum_{k=0}^{n} exp(S(f(v_i), f(b_k^-))/\tau)} \qquad (10)$$

In equation (10), L_i' represents the contrastive loss, m means the amount of positive samples, n means the amount of negative samples, $f(v_i)$ means the mapped image features, $f(b_j^+)$ represents brain response features of the same category as the image features, and $f(b_j^-)$ represents brain response features of different categories from the image features. In intelligent computing models based on the fusion of DCNN and brain visual cognitive information, the DCNN structure may be affected in classification accuracy due to irrelevant information. To address this issue, research is being conducted to improve the DCNN structure by incorporating attention mechanisms. The DCNN feature extraction model grounded on attention mechanism is represented in Figure 5.

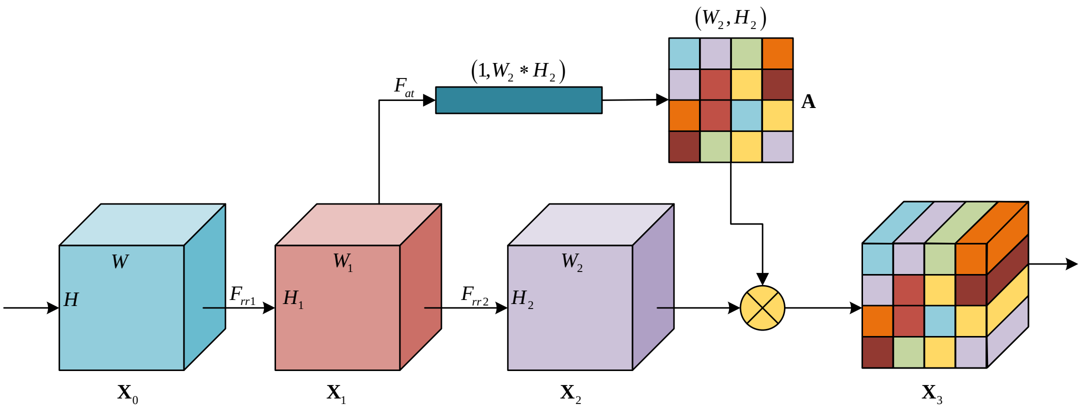


Figure 5: DCNN feature extraction model grounded on attention mechanism

As represented in Figure 5, the DCNN architecture used mainly includes multiple convolutional layers, pooling layers, activation functions, and fully connected layers, aiming to extract multi-level feature representations from images to improve classification performance. The core idea of DCNN is to extract local features from images through a series of convolution and pooling operations, and then add nonlinear transformations through nonlinear activation functions to learn more complex image representations. The convolutional layer, as a fundamental component in the DCNN architecture, can extract local features of the input image through convolution operations. After each convolution operation in each layer, the study will use activation functions to perform nonlinear transformations on the output results. The purpose of the activation function is to introduce nonlinear factors, so that the network can learn more complex mapping relationships. The function of the pooling layer is to downsample the feature map output by the convolutional layer, thereby reducing the spatial size of the feature map while preserving important features. After extracting sufficient local features in the convolutional and pooling layers, the last few layers are usually fully connected layers. The fully connected layer linearly combines the extracted features and generates the final output result through an activation function. The DCNN feature extraction model grounded on attention mechanism adds a parallel attention branch to the initial DCNN structure to learn the importance information of feature map position. This path can correct the activation values of feature maps, reduce the activation values of redundant information, and thus improve the accuracy of image characteristic collection [22, 23]. The feature transformation process is shown in equation (11).

$$\mathbf{X}_1 = F_{rr1}(\mathbf{X}_0) \tag{11}$$

In equation (11), $\mathbf{X}_1$ represents the transformed abstract feature map, $\mathbf{X}_0$ represents the initial feature map, and F_{rr1} represents the downsampling operation. The calculation for position importance is shown in equation (12).

$$\mathbf{A} = F_{at}(\mathbf{X}_1) \tag{12}$$

In equation (12), $\mathbf{A}$ represents positional importance and F_{at} represents fully connected operation. The new feature map obtained by further downsampling the abstract features is shown in equation (13).

$$\mathbf{X}_2 = F_{rr2}(\mathbf{X}_1) \tag{13}$$

In equation (13), $\mathbf{X}_2$ represents the new feature map after further downsampling, and F_{rr2} represents the further downsampling operation. The corrected feature map is shown in equation (14).

$$\mathbf{X}_3 = \sum_{i=1}^{W_2}\sum_{j=1}^{H_2} \mathbf{A}(i, j) \cdot \mathbf{X}_2(i, j) \tag{14}$$

In equation (14), $\mathbf{X}_3$ represents the feature map obtained after attention branch correction, W and H represent the width and height of the characteristic map, and (i, j) represents the feature values on the feature map. When capturing the correlation information between brain visual cognitive responses and image features, some non-correlated neural responses may affect the determination of representation similarity. These "unrelated neural responses" pertain to neural activities that aren't directly tied to visual tasks and might stem from background noise, irrelevant visual cues, or various other bodily influences. For example, the activity of certain regions in EEG signals may be unrelated to the current visual task, and this irrelevant neural activity can lead to misleading similarity judgments when the brain processes visual information. To address this issue, research has been conducted on an intelligent computing model based on the fusion of DCNN and brain visual cognitive information, which devotes to raise the precision of capturing correlated information by masking non-correlated neural responses. In the intelligent computing model based on the fusion of DCNN and brain visual cognitive information, the study aims to add windows of different scales to the extracted image features to mask non-correlated neural responses. The motivation of this method is to better highlight the effective response of the brain to visual information and improve the accuracy of similarity determination between brain visual cognitive responses and image features by reducing or eliminating the influence of irrelevant neural reactions. The visualization process of brain response and image feature correlation information is shown in Figure 6.

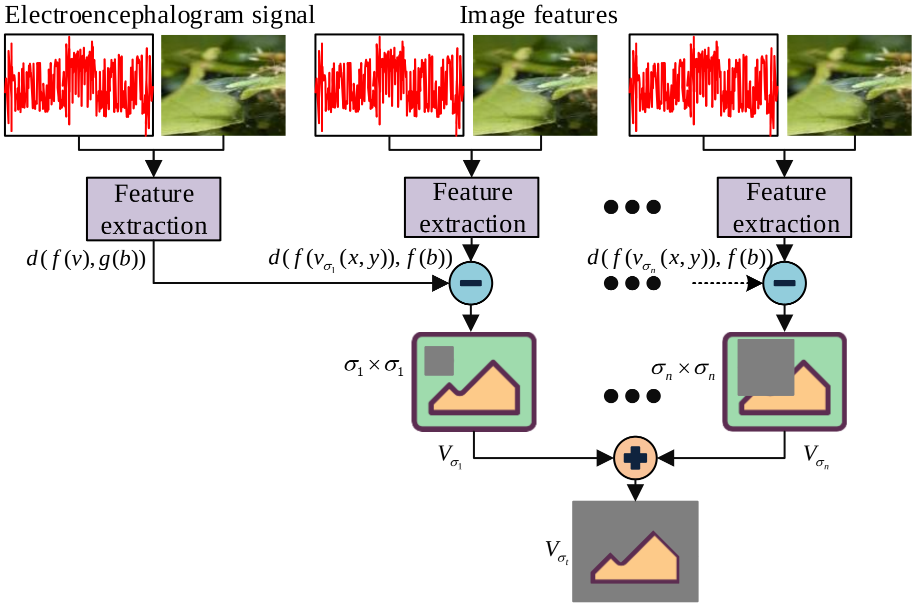

Electroencephalogram signal
Image features
Feature extraction
Feature extraction
Feature extraction
$d(f(v), g(b))$
$d(f(v_{\sigma_1}(x,y)), f(b))$
$d(f(v_{\sigma_n}(x,y)), f(b))$
$\sigma_1 \times \sigma_1$
$\sigma_n \times \sigma_n$
V_{σ_1}
V_{σ_n}
V_{σ_t}

Figure 6: Visualization process of brain response and image feature correlation information

As shown in Figure 6, in the process of visualizing the correlation information between brain response and image features, the correlation features are first represented in the shared representation space and their Euclidean distance is calculated. Then, a window with a scale of $\sigma_i \times \sigma_i$ is added to the extracted image features to mask non-correlated neural responses. Based on the Euclidean distances calculated from various image features, different sizes of occlusion windows are determined. The saliency maps obtained through occlusion at different scales are then combined to create a comprehensive saliency map that encapsulates the relationship between brain responses and image features. The calculation for the significance map is shown in equation (15).

$$V_{\sigma_t} = | d(f(v_{\sigma_t}(x,y)), f(b)) - d(f(v), g(b)) | \quad (15)$$

In equation (15), V_{σ_t} represents the significance map and d represents the Euclidean distance.

4 Validation of an intelligent computing model integrating DCNN and brain visual cognition

After setting up the experimental environment, the behaviour of the image classification model grounded on information fusion was first verified, and then the intelligent computing model based on information fusion was experimentally analyzed.

4.1 Experiment environment construction

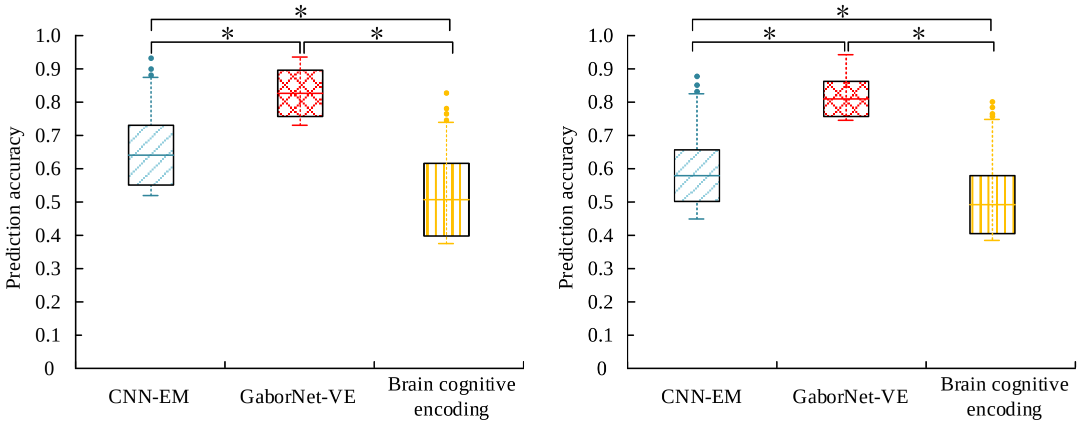
To tesify the effectiveness of the intelligent computing model that integrated DCNN and brain visual cognition, the study first conducted the construction of an experimental environment. The experimental hardware system configuration was as follows: the processor was Intel i7-8700, the GPU was Nvidia GeForce 1080Ti, and the memory was 64 GB DDR4. The experimental model used Python language and was implemented using the PyTorch framework. The experimental parameters were set as follows: batch size was 16, original learning rate was 0.001, Adam optimizer was used during training, output layer size was 40, and the key vector value in the self attention mechanism was 128. The dataset was sourced from the comprehensive evaluation platform Brain Score. This dataset aimed to evaluate the effectiveness and accuracy of computer simulated brain operation models, thus covering response data of primate visual systems. The dataset contains approximately 5000 image stimuli, each corresponding to a recorded brain electrophysiological response data. The stimulus images cover a total of 40 categories, including natural scenes and artificial objects. The number of images in each category is roughly equal to ensure data balance. The size of each image is 224x224 pixels, which can retain sufficient visual information and meet the input requirements of CNN. Any data augmentation techniques used by the research include random cropping, horizontal flipping, random rotation, and color jitter. The above data augmentation techniques can effectively expand the diversity of training data, avoid model overfitting, and improve the generalization ability to various visual stimuli. After preprocessing, the data was separated into a training set and a testing set in a 3:7 ratio. While primarily intended for evaluating brain functional models, the "Brain Score" dataset is well-suited as a data source in this study to verify the efficacy of intelligent computing models that integrate image classification with brain visual cognition, given its abundance of visual stimulus images and corresponding EEG response data. In the experimental design of this study, the evaluation of image classification focuses on guiding the learning and classification of image features through brain response data, rather than simply image classification. The detailed experiment environment configuration and network training parameters are represented in Table 2.

Table 2: Experiment environment configuration and network training parameters

Experimental environment	Configuration	Training parameters	Configuration
CPU	Intel i7-8700	Batch Size	16
GPU	Nvidia GeForce 1080Ti	Initial learning rate	0.001
Memory	64 GB DDR4	Output layer size	40
programming language	Python	Key vector value	128
Frame	PyTorch	Optimizer	Adam

4.2 Performance verification of image classification model based on information fusion

In order to verify the predictive accuracy of ventral response encoding based on brain visual cognition for brain cognitive response, this method was compared and analyzed with other voxel encoding methods, including Convolutional Neural Network Enhancement Model (CNN-EM) and GaborNet Visual Encoding (GaborNet-VE). The accuracy comparison of different encoding methods in different visual regions is represented in Figure 7. From Figure 7(a), within the V4 visual region, the prediction accuracy of the ventral response encoding method based on brain visual cognition was significantly higher than the other two methods. The maximum prediction accuracy of this method reached 93.54%, which was 6.05% and 19.57% higher than the maximum prediction accuracy of CNN-EM and GaborNet-VE, which were 87.49% and 73.97%, separately. From Figure 7(b), the results of visual area L0 show that the maximum prediction accuracy of the ventral response encoding method based on brain visual cognition was 94.03%, which was 11.49% and 18.95% higher than the maximum accuracy of 82.54% and 75.08% of the other two methods, respectively. In addition, the study used paired t-tests to validate the credibility of the results. In the V4 region, the difference in accuracy between the ventral response encoding based on brain visual cognition and CNN-EM reached a statistically significant level (t=4.72, *P*<0.05). The difference in accuracy between ventral response encoding based on brain visual cognition and GaborNet VE also reached a statistically significant level (t=6.88, *P*<0.05). In the L0 region, the accuracy difference between ventral response encoding based on brain visual cognition and CNN-EM reached a statistically significant level (t=5.23, *P*<0.05). Similarly, the accuracy difference between ventral response encoding based on brain visual cognition and GaborNet VE was also statistically significant in the L0 region (t=7.14, *P*<0.05). Ventral response encoding based on brain visual cognition could accurately predict brain cognitive response.



(a) Comparison of Accuracy in V4 Visual Region

(b) Comparison of Accuracy in L0 Visual Region

Figure 7: Comparison of accuracy of different encoding methods （*Indicating P<0.05）

To further testify the performance of the image classification model based on information fusion, a relative unpack was operated on the classification models before and after adding LSTM, as represented in Figure 8. From Figure 8, the loss value of the model before adding LSTM converged to 0.26, while the loss value of the model after inserting LSTM converged to 0.21, with a reduction of 19.23%. This indicated that the classification model incorporating LSTM had better convergence performance.

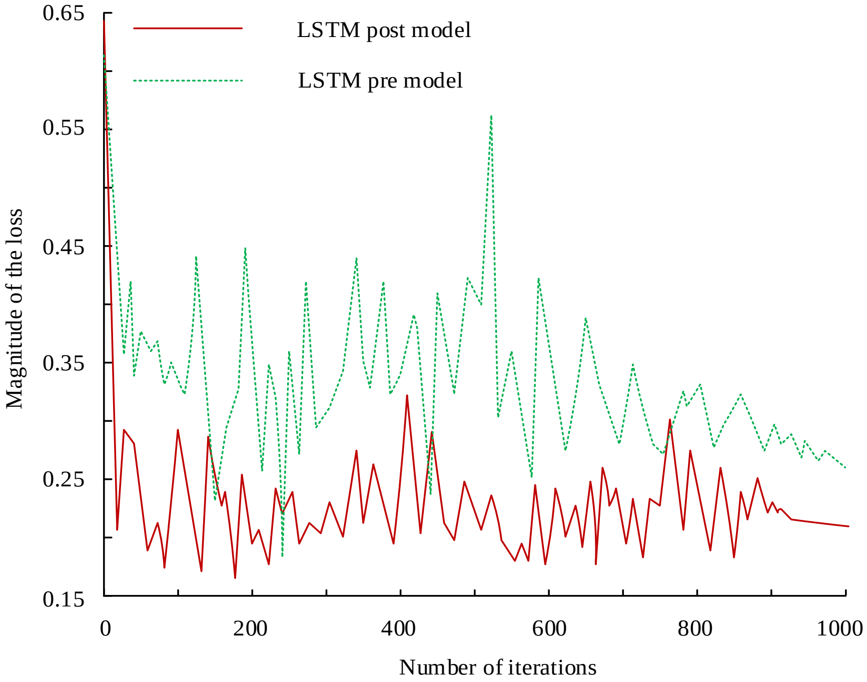


Figure 8: Comparison of training loss values for networks incorporating LSTM Models

To further validate the capability of the image classification model grounded on information fusion, this study compared and analyzed the model with other advanced image classification models, including Feature Weighted Classification (FWC), Residual Network (ResNet), and Visual Geometry Group (VGG). In addition, to ensure the broad applicability and contextualization of the research results, the performance of these models was compared with benchmark test results in the current field of computer vision. The accuracy comparison of different classification models is represented in Figure 9. From Figure 9, in the datasets of various visual images, the accuracy of the picture sorting model grounded on information fusion was the best. On facial vision images, the accuracy of this model was as high as 95.46%, which was an improvement of 6.30%, 5.41%, and 10.03% compared to the accuracy of 89.16%, 90.05%, and 85.43% of FWC, ResNet, and VGG, respectively. This result showed significant advantages compared to the mainstream benchmarks in the current field of facial recognition. In facial recognition tasks, many of the most advanced facial recognition technologies, such as FaceNet and ArcFace, achieved high accuracy on multiple standard datasets such as LFW and CASIA WebFace. For example, FaceNet reported an accuracy of 94.63% on the LFW dataset [6]. Moreover, ArcFace also achieved a recognition rate of nearly 94.51% on the same dataset [7]. However, the above studies all achieved accuracy in interference free environments, while this study still achieved an accuracy of up to 95.46% in actual environments with complex interference and occlusion. Compared with existing benchmark tests, the researched image classification model based on information fusion had more advantages in performance. On animal visual images, the classification accuracy of this model was the lowest at 91.57%, which was 16.31%, 12.03%, and 19.08% higher than the accuracy of 75.26%, 79.54%, and 72.49% of the other three models, respectively. Compared with basic testing image classification tasks, animal classification often faces more complex backgrounds and varying object shapes, which makes this task an important criterion for testing model robustness. Therefore, the significant improvement of the research model in this task indicated that it has stronger generalization ability and adaptability when facing highly complex and dynamically changing visual environments. In summary, the image classification model based on information fusion has demonstrated excellent classification performance in multiple tasks. Moreover, the performance of the research model still has significant advantages compared to benchmark testing.

The reason why facial image recognition had better accuracy than animal image recognition was that facial images had more stable and easily recognizable features compared to animal images. Facial recognition typically fixed structural features and relatively consistent backgrounds, which enabled information fusion based models to fully exploit the effective information in the brain's visual cognitive model, thereby improving recognition accuracy. However, animal images face more complex challenges, including background noise, changes in animal size and morphology, different shooting angles, different species, etc. These factors make classification tasks more complex and varied. Therefore, in terms of recognition accuracy, the performance of facial image classification was better than that of animal image classification.

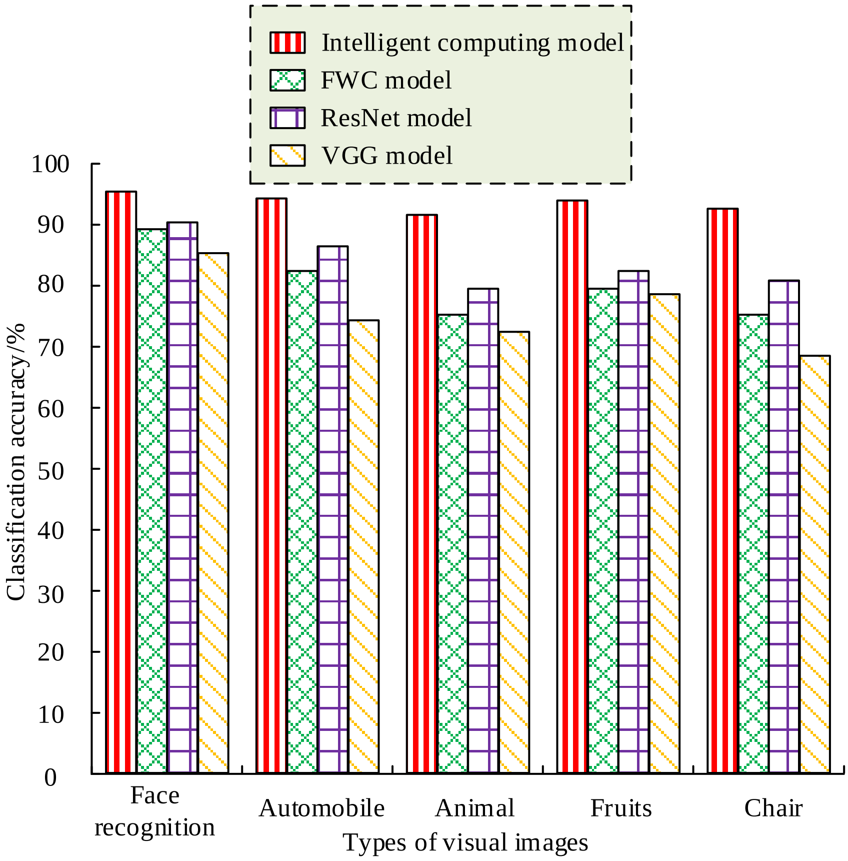
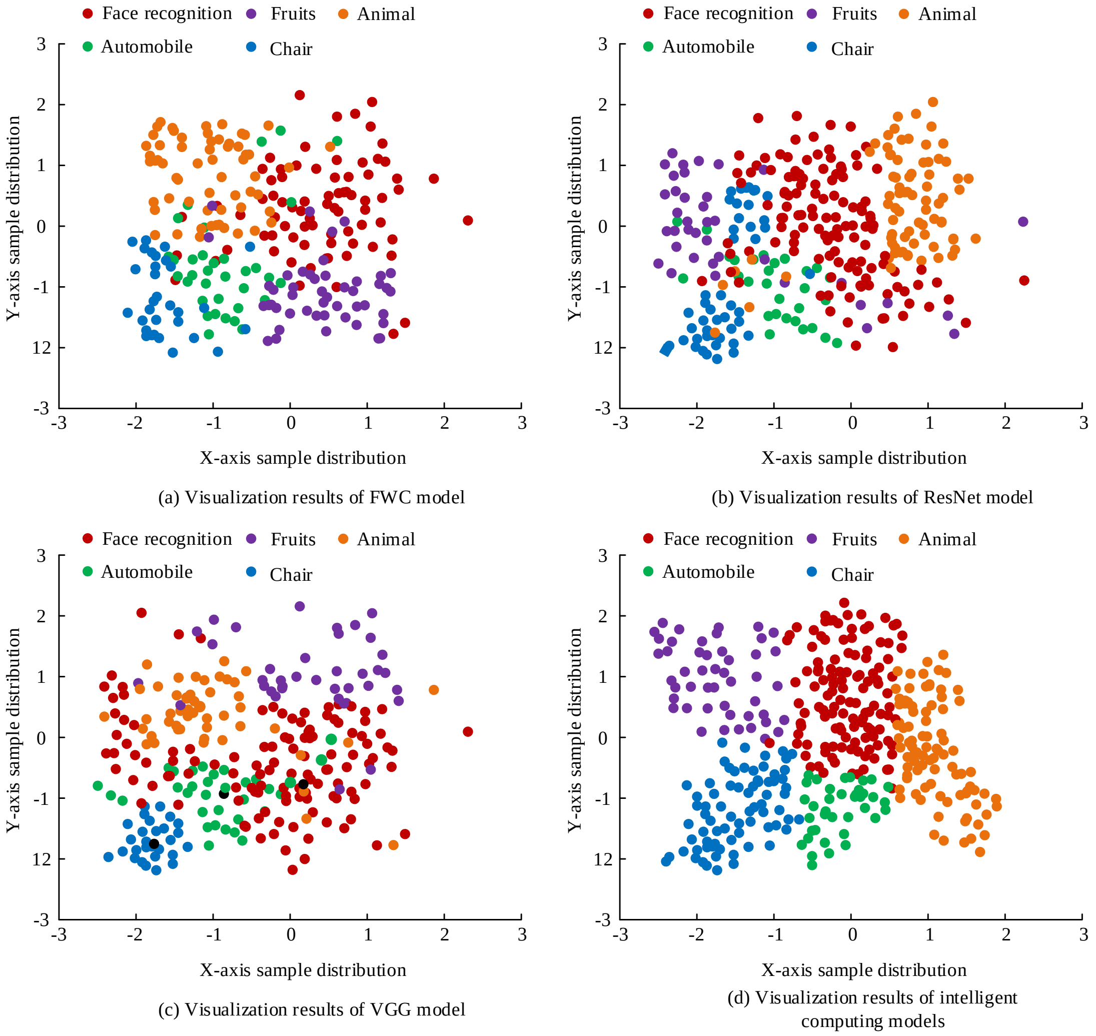


Figure 9: Comparison of accuracy between different classification model

4.3 Performance verification of intelligent computing models based on information fusion

To tesify the ability of intelligent computing models grounded on information fusion, the visualization sample distribution results of different models under various brain visual cognitive image stimuli were compared and studied. The dataset contains 40 categories of images, which are divided into two main categories: natural scenes and artificial objects. Natural scenes include image categories such as faces and animals, while artificial objects include image categories such as cars, fruits, chairs, etc. In the experiment, a combination of these image categories was used to test the classification performance of the model under different visual stimuli. The visualization outcomes of various models are represented in Figure 10. From Figure 10, the sample distribution of FWC and VGG was relatively chaotic, while the sample distribution of ResNet was relatively clear. The ResNet model had a more obvious distinction between facial and animal visual images, but it was more confusing in distinguishing images such as fruits and cars. The intelligent computing model based on information fusion studied exhibited significant classification clarity and good classification performance under all visual image stimuli. This was because images of facial and animal categories were more consistent in natural scenes and were easily distinguishable by models. However, categories such as cars, fruits, and chairs belong to the category of artificial objects, and the visual differences between these categories were significant, posing greater challenges to the model. The intelligent computing model based on information fusion revealed the deep level features of brain response, effectively improving classification accuracy.



(a) Visualization results of FWC model

(b) Visualization results of ResNet model

(c) Visualization results of VGG model

(d) Visualization results of intelligent computing models

Figure 10: Visualization results of different models

To further validate the ability of the intelligent computing model grounded on information fusion, ablation experiments were conducted. The classification accuracy in ablation experiments was calculated based on the precision of image classification tasks, which only reflected the accuracy of image classification results. The ablation experiment results of image classification are shown in Table 3. From Table 3, the classification accuracy of the brain visual cognitive response encoding framework was 81.42%. When the DCNN structure was fused, the accuracy was improved by 5.64%. After adding the LSTM structure, the accuracy was increased to 90.48%. When attention mechanism was added for improvement, the classification accuracy increased by 2.17%. When further optimizing using occluded non-correlated neural responses, the accuracy of the model reached 93.94%. From the above, it can be seen that the addition of the above modules brought benefits to the classification performance of the model, effectively raising the classification accuracy of images.

Table 3: Ablation experiment

Brain Response Coding Framework	DCNN	LSTM	Attention mechanism	Obstructing non correlated neural responses	Accuracy rate/%
√	/	/	/	/	81.42
√	√	/	/	/	87.06
√	√	√	/	/	90.48
√	√	√	√	/	92.65
√	√	√	√	√	93.94

Note："√" indicates the existence of the module； "/" indicates its non existence

5 Discussion

In order to improve the accuracy of computer vision image classification, a fusion intelligent computing model was constructed by simulating the visual processing mechanism of the human brain, using BCI technology to extract EEG signals generated by human visual cognition, and combining DCNN structure. The results showed that after adding LSTM, the convergence of the model was significantly improved, with the loss value decreasing from 0.26 to 0.21, indicating a 19.23% increase in convergence speed. This indicated that LSTM could effectively capture time series features, improve the model's ability to process time-series data, and thus make the model more accurate in learning dynamic information. After incorporating the advantages of LSTM into the model, it could better understand the temporal dependencies in brain activity, resulting in more accurate prediction performance. In addition, compared with other advanced methods, the research method was significantly superior to other methods. For example, although the model studied by Gao Z et al. effectively improved the decoding performance of VEP in complex environments, it still faced the problem of noise interference and failed to effectively integrate spatial and temporal features in the brain's visual cognitive process [6]. The model studied in this article not only considers spatial features but also integrates dynamic temporal information when predicting brain responses in visual regions, significantly improving the accuracy of predictions. In addition, the model proposed by Ahirwal M K et al. achieved good results in emotion classification, but it mainly focused on emotion classification and cannot handle complex visual information and multi-class image classification tasks [7]. The model studied in this article could not only handle complex visual information, but also adapt to the multidimensional features of the brain's visual cognitive process, thus exhibiting a more comprehensive classification and understanding of visual information.

The potential extensions of the research model to other tasks include video analysis, multi-modal data fusion, etc. Video data not only contains spatial information of static images, but also dynamic time series information. Therefore, models based on brain visual cognition can better understand the dynamic changes in videos by integrating spatial and temporal features, especially with the addition of LSTM modules. In the field of multi-modal data fusion, cross modal learning can be achieved by introducing multi-modal neural network structures and combining data from different modalities. For example, in video description generation tasks, visual information of video frames can be combined with speech or text information to generate more accurate and natural descriptions.

The reason why the research method is superior to other methods is that it considers the spatial and temporal characteristics of the brain in the visual cognitive process, while other methods rely more on a single spatial or static feature. In addition, the introduction of LSTM further enhances the model's ability to process temporal information, enabling the model to decode complex dynamic brain signals more accurately. The potential applications of this discovery cover fields such as neuroscience experiments, intelligent medical devices, and brain computer interaction systems. However, this method still has certain limitations. For example, the study only explored the classification of EEG images, so the research results are not comprehensive enough. This aspect can be further improved in the future.

6 Conclusion

In recent years, the introduction of visual cognitive mechanisms in the brain has provided new solutions to the limitations of accuracy and generalization ability of traditional DCNN in processing complex visual information. Research used BCIs to receive EEG information, used voxel encoding models to obtain the expression content of visual images, and combined an improved DCNN structure to construct an efficient image classification model. On this basis, the LSTM structure was further introduced to extract time series features of EEG signals. Attention mechanisms and occlusion independent neural responses were utilized to enhance the accuracy of capturing correlation information between brain responses and image features. The outcomes revealed that the ventral response encoding method grounded on brain visual cognition achieved prediction accuracy of 93.54% and 94.03% in the V4 visual region and L0 visual region, significantly better than the CNN-EM and GaborNet VE methods. In the model validation, after adding the LSTM module, the loss value decreased from 0.26 to 0.21, with a reduction of 19.23%. In terms of image classification capability, the accuracy of the information fusion based model on facial visual images was as high as 95.46%, and the lowest accuracy on animal visual images was 91.57%, both significantly better than comparative models such as FWC, ResNet, and VGG. In addition, ablation experiments showed that by introducing attention mechanisms and occlusion independent neural responses, the final classification accuracy was improved to 93.94%. From the above, the research on the fusion intelligent computing model based on DCNN and brain visual cognition effectively improved the accuracy of computer vision image classification.

Although research focused on EEG image classification and achieved good classification results in the relevant areas of ventral flow and visual regions, the current scope of research has not yet covered other brain tissue and neural mechanisms. Therefore, future research can be extended to explore the functions of other brain regions, such as their contributions to tasks such as cognitive control and emotion recognition in different brain regions. In addition, combining different neural mechanisms and multi-modal data will help improve the comprehensiveness and accuracy of cognitive image classification, thereby promoting further development in the field of BCIs. Future work will strive to further enhance the analytical ability of EEG information for complex visual stimuli through the integration of broader neural regions and mechanisms, in order to promote the widespread application of intelligent computing models in practical applications.

Funding

This study was supported by the Key Science and Technology Program of Henan Province (Project Name: Research on NoC Routing Algorithm and Fault-Tolerant Technology Based on Spanning Tree Sub-Domains; Grant No. 252102210225).

References

[1] Wilson H, Chen X, Golbabaee M, Proulx, M. J., & O'Neill, E. Feasibility of decoding visual information from electroencephalogram. Brain-Computer Interfaces, 2024, 11(1-2): 33-60. DOI: 10.1080/2326263X.2023.2287719

[2] Finlayson S G, Subbaswamy A, Singh K, Bowers, J, Kupke, A., Zittrain, J & Saria, S. The clinician and dataset shift in artificial intelligence. New England Journal of Medicine, 2021, 385(3): 283-286. DOI: 10.1056/NEJMc2104626

[3] Masana M, Liu X, Twardowski B, Menta, M., Bagdanov, A. D & Van De Weijer, J. Class-incremental learning: survey and performance evaluation on image classification. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(5): 5513-5533. DOI: 10.1109/TPAMI.2022.3213473

[4] Zhu Y, Zhuang F, Wang J, Ke, G., Chen, J., Bian, J & He, Q. Deep subdomain adaptation network for image classification. IEEE Transactions on Neural Networks and Learning Systems, 2020, 32(4): 1713-1722. DOI: 10.1109/TNNLS.2020.2988928

[5] Hong D, Gao L, Yao J, Zhang, B., Plaza, A & Chanussot, J. Graph convolutional networks for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(7): 5966-5978. DOI: 10.1109/TGRS.2020.3015157

[6] Gao Z, Sun X, Liu M, Dang, W., Ma, C., & Chen, G. Attention-based parallel multiscale convolutional neural network for visual evoked potentials electroencephalogram classification. IEEE Journal of Biomedical and Health Informatics, 2021, 25(8): 2887-2894. DOI: 10.1109/JBHI.2021.3059686

[7] Ahirwal M K, Kose M R. Audio-visual stimulation based emotion classification by correlated electroencephalogram channels. Health and Technology, 2020, 10(1): 7-23. DOI: 10.1007/s12553-019-00394-5

[8] Komolovaitė D, Maskeliūnas R, Damaševičius R. Deep convolutional neural network-based visual stimuli classification using electroencephalography signals of healthy and alzheimer's disease subjects. Life, 2022, 12(3): 374-379. DOI: 10.3390/life12030374

[9] Kumari N, Anwar S, Bhattacharjee V. Time series-dependent feature of electroencephalogram signals for improved visually evoked emotion classification using EmotionCapsNet. Neural Computing and Applications, 2022, 34(16): 13291-13303. DOI: 10.1007/s00521-022-06942-x

[10] Santamaria-Vazquez E, Martinez-Cagigal V, Vaquerizo-Villar F, & Hornero, R. electroencephalogram-inception: a novel deep convolutional neural network for assistive ERP-based brain-computer interfaces. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2020, 28(12): 2773-2782. DOI: 10.1109/TNSRE.2020.3048106

[11] Yıldırım Ö, Baloglu U B, Acharya U R. A deep convolutional neural network model for automated identification of abnormal electroencephalogram signals. Neural Computing and Applications, 2020, 32(20): 15857-15868. DOI: 10.1007/s00521-018-3889-z

[12] Miao M, Hu W, Yin H, & Zhang, K. Spatial-Frequency feature learning and classification of motor imagery electroencephalogram based on deep convolution neural network. Computational and Mathematical Methods in Medicine, 2020, 2020(1): 1981728-1981752. DOI: 10.1155/2020/1981728

[13] Li F, He F, Wang F, Zhang, D & Li, X. A novel simplified convolutional neural network classification algorithm of motor imagery electroencephalogram signals based on deep learning. Applied Sciences, 2020, 10(5): 1605-1624. DOI: 10.3390/app10051605

[14] Tang X, Shen H, Zhao S, Li, N., & Liu, J. Flexible brain – computer interfaces. Nature Electronics, 2023, 6(2): 109-118. DOI: 10.1038/s41928-022-00913-9

[15] Kawala-Sterniuk A, Browarska N, Al-Bakri A, Pelc, M., Zygarlicki, J., Sidikova, M., et al. Summary of over fifty years with brain-computer interfaces—a review. Brain Sciences, 2021, 11(1): 43-45. DOI: 10.3390/brainsci11010043

[16] Cohn N. Your brain on comics: a cognitive model of visual narrative comprehension. Topics in Cognitive Science, 2020, 12(1): 352-386. DOI: 10.1111/tops.12421

[17] Finlayson S G, Subbaswamy A, Singh K, Bowers, J, Kupke, A., Zittrain, J & Saria, S. The clinician and dataset shift in artificial intelligence. New England Journal of Medicine, 2021, 385(3): 283-286. DOI: 10.1056/NEJMc2104626

[18] Bicanski A, Burgess N. Neuronal vector coding in spatial cognition. Nature Reviews Neuroscience, 2020, 21(9): 453-470. DOI: 10.1038/s41583-020-0336-9

[19] Franzen L, Stark Z, Johnson A P. Individuals with dyslexia use a different visual sampling strategy to read text. Scientific Reports, 2021, 11(1): 6449-6455. DOI: 10.1038/s41598-021-84945-9

[20] Zhou S K, Greenspan H, Davatzikos C, Duncan, J. S, Van Ginneken, B, Madabhushi, A & Summers, R. M. A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. Proceedings of the IEEE, 2021, 109(5): 820-838. DOI: 10.1109/JPROC.2021.3054390

[21] Basso M A, Bickford M E, Cang J. Unraveling circuits of visual perception and cognition through the superior colliculus. Neuron, 2021, 109(6): 918-937. DOI: 10.1016/j.neuron.2021.01.013

[22] Jeong J J, Tariq A, Adejumo T, Trivedi, H., Gichoya, J. W., & Banerjee, I. Systematic review of generative adversarial networks (GANs) for medical image classification and segmentation, Journal of Digital Imaging, 2022, 35(2): 137-152. DOI: 10.1007/s10278-021-00556-w

[23] Wang F. Automatic ink painting rendering technique based on deep convolutional neural networks. Informatica, 2025, 49(5): 95-108. DOI: 10.31449/inf.v49i5.7112