Convolutional Neural Network in Brain-computer Interfaces-exoskeleton System

Yanbo Wang*

School of Mechanical Engineering and Automation, Northeastern University, Shenyang 200240, China

* Corresponding Author Email: 20215066@stu.neu.edu.cn

Abstract. Brain-computer interfaces (BCIs) have emerged as a groundbreaking technology that has the potential to revolutionize the field of stroke rehabilitation. These innovative systems allow individuals who have suffered from strokes to regain lost motor function by directly connecting their brains with external devices, such as exoskeletons. One of the most commonly used paradigms in BCIs is motor imagery (MI), which involves generating electroencephalograms (EEGs) through imagined movements. This means that stroke patients can perform motor tasks simply with the help of exoskeletons by only thinking about them, without any physical movement required. The ability to decode these EEG signals is crucial for enabling effective communication between the brain and external devices. Deep learning (DL), particularly convolutional neural networks (CNNs) which are able to extract meaningful features from raw EEG data and accurately classify different types of imagined movements, has proven to be highly effective in decoding EEG signals generated during motor imagery tasks and has already made numerous contributions in this area. This article provides an initial overview of BCIs and CNNs, followed by an explanation of how CNNs decode EEG signals derived from motor imagery, as well as the utilization of BCIs for controlling exoskeleton devices. Finally, the current limitations and future directions in this field are discussed.

Keywords: Brain-computer interfaces, motor imagery, convolutional neural networks, exoskeleton.

1. Introduction

Brain-computer interfaces (BCIs) have the remarkable ability to capture a vast array of electroencephalogram (EEG) signals. Patients with stroke impairments can use motor imagery (MI) to generate these brain signals, which are then captured and processed by BCIs to drive exoskeletons, enabling them to perform actions they have lost. BCIs are considered to have the potential to revolutionize patients rehabilitation, opening up new possibilities for restoring the independence of those with spinal cord injuries or neurological diseases. Currently, EEG-based motor imagery has been applied in some medical applications, such as stroke patients using exoskeletons to control elbow or lower limb movements to regain arm motion and walking abilities [1]. However, EEG signals have a highly complex structure, and there are many challenges in classifying EEG signals, such as susceptibility to interference from physiological responses like heartbeats and eye blinks, as well as electronic device radiation, and environmental factors like noise and lighting. As a result, most BCI systems are still in the clinical trial stage and have not been widely adopted.

The EEG signal processing involves several key components: (1) noise processing and filtering, (2) feature extraction, and (3) classification. Currently, methods using deep learning to process EEG signals are widely applied in BCI-related experiments. Deep learning is believed to have powerful computational capabilities. The application of deep learning in BCIs has been researched for more than a decade with many achievements. Among them, convolutional neural networks (CNN) are the most frequently used in decoding EEG signals, with statistics showing that 78% of models in the past decade have employed convolutional neural networks (including CNNs and hybrid CNNs)[1]. In some of the most cutting-edge research, Mwata-Velu used a convolutional neural network (CNN) with a K-fold method to achieve an average classification accuracy of 94.8% [2], successfully classifying visual EEG signals into 40 categories; Yu and colleagues constructed a multi-branch convolutional network (MBCNN) and a temporal convolutional network (TCN) for decoding, with an average accuracy reaching 75.08% [3]. As of now, there are over 160 articles published in 2024 on the Web of Science, indicating the rapid development of research in this field.

Convolutional Neural Networks (CNNs), as the most commonly used deep learning models, have numerous applications in various steps of electroencephalogram (EEG) signal processing. This review will summarize the whole possess of the use of CNNs for filtering, feature extraction, and classification control of EEG signals. It will also discuss the advantages and disadvantages of existing methods and propose some possibilities for future development.

2. Computer Interfaces and EEG Signals

2.1. Computer Interfaces

Brain-computer interface (BCI) systems collect brain signals under specific task conditions, capture neural activity, and then process and decode the collected signals into specific instructions, which are then transmitted to external devices to achieve functional substitution or neural feedback. Based on the method of EEG signal acquisition, BCIs can be divided into invasive and non-invasive categories. Invasive BCIs were initially used to improve the motor function or communication abilities of patients with high-level spinal cord injuries and amyotrophic lateral sclerosis (ALS), reducing their dependence on the outside world, with the aim of functional substitution. The advantage of invasive BCIs lies in the precision of signal extraction and analysis, but due to the need for surgical implantation of electrodes, there are certain risks involved, and their application is relatively limited. Currently, with the further development of signal analysis and decoding technologies, an increasing number of non-invasive BCIs are being used for the rehabilitation intervention of stroke patients.

2.2. Electroencephalogram Signal

Electroencephalogram (EEG) signals are the sum of the postsynaptic potentials of neuronal populations in the cerebral cortex recorded from the surface of the cerebral cortex or the scalp.

Paradigms in the research process of brain-computer interfaces (BCIs):

P300: It is a type of Event-Related Potentials (ERPs) that typically appears about 300 milliseconds after a specific cognitive event. P300 is often generated through visual evoked potentials and is observed in the parietal and occipital regions of the brain, being one of the strongest neural characteristics observed in EEG.

Error-Related Negativity (ERN): It appears within about 50 milliseconds after making a mistake and is usually observed in the prefrontal cortex. It is the amplitude disturbance of the electroencephalogram following the perception of error feedback in a sensory BCI.

Steady-State Visual Evoked Potentials (SSVEPs): When a visual stimulus flickers at a certain frequency, the brain generates a steady-state visual evoked potential that is synchronized with the flickering frequency. The frequency of SSVEPs is consistent with the flicker frequency of the stimulus and can be used for high-speed BCI communication.

Movement-Related Cortical Potentials (MRCPs): Potential changes related to actual or imagined limb movements, including preparation potentials and movement execution potentials.

Sensory Motor Rhythm (SMR): Electroencephalographic rhythms related to sensory and motor functions, typically in the range of 12-30 Hz, reflecting the activity of sensory and motor areas.

3. Convolutional Neural Networks

It is currently the most commonly used method for processing brain-computer interface (BCI) systems. Convolutional Neural Networks (CNNs) learn different features from input signals through nonlinear transformations, mainly used for processing data with a clear grid-like topological structure, such as images and audio waveforms. The tight connections between layers and spatial information in Convolutional Neural Networks make them particularly suitable for image processing and understanding. Currently, image recognition technologies such as face recognition and license plate recognition all use Convolutional Neural Networks.

Convolutional Neural Networks are mainly composed of three structural blocks: convolutional layers (for feature extraction), pooling layers (for feature dimension reduction), and fully connected layers (FC) (for classification).

The main job of the convolutional layer is to train a smaller number of parameters to extract feature information from data images. Convolutional layers use convolutional kernels to convolve each local feature of the input image and extract feature information, which is essentially the process of the neural network extracting the external features of the image by learning local features.

The main function of the pooling layer is to reduce the number of parameters and shrink the size of the feature map, which helps to prevent overfitting. The max pooling operation reduces the input 5×5 matrix to a 3×3 output matrix. After one max pooling operation, it effectively reduces the model parameters and can avoid model overfitting [4].

Convolutional and pooling layers form the basic building blocks of Convolutional Neural Networks (CNNs), and networks of different depths are established by stacking multiple convolutional and pooling layers. For deeper CNNs, activation functions are required to enhance the network's nonlinear expressive power and its ability to fit the complexity of the data.

The main role of the fully connected (FC) layer is to transform the image features output by the convolutional neural network after pooling into a one-dimensional vector for image recognition. Typically, the fully connected operation comes after the convolution and pooling operations, and it plays a role in classification within the entire neural network.

4. Convolutional Neural Network and Motor Imagery

EEG signals are two-dimensional images that combine temporal and spatial features, and they have a low signal-to-noise ratio. Traditional CNN structures are not very suitable for processing EEG signals. To prevent the features after convolution from mixing both spatial and temporal information, the convolution kernels in the convolutional layers need to be specifically set as vectors rather than matrices used in general image recognition, so as to extract only spatial or temporal features.

This article will use a case to explain to readers how CNN processes EEG signals for motor imagery [5].

Researchers, on one hand, use public EEG datasets such as the Dataset 1 from BCI Competition IV, and on the other hand, recruit participants. Subjects wear corresponding EEG signal collection devices and perform motor imagery based on computer prompts, collecting the corresponding EEG signals. Subsequently, the features of the EEG signals are learned through deep learning methods.

Tang et al. proposed a network composed of five layers: The first layer is the input layer; the second (convolutional layer) and third (convolutional layer) constitute the feature extraction part; the output of the third layer (feature values) and the fourth and fifth layers (fully connected layers) constitute the classification part.

(1) First layer (L1): This layer is the input layer, with each input sample being a 28×60 input matrix, where 28 represents the number of channels, and 60 indicates the time sampling points in each channel.

(2) Second layer (C2): This layer is the convolutional layer (the first hidden layer), which mainly performs spatial filtering on the original input samples, so the connections between this layer and the input layer are local connections. Eight filters are used in this layer, and each filter convolves the input matrix to obtain different feature mappings, resulting in 8 feature maps. The convolution kernel is set to [28×1], and the size of each feature map is (1×60). The convolution kernel is set as a vector rather than a matrix used in general image recognition to prevent the features after convolution from mixing two types of information, containing only spatial features.

(3) Third layer (C3): This layer is the convolutional layer (the second hidden layer), which mainly extracts temporal features of the EEG signals, so the concept of local connections and weight sharing

is also applied. Five filters are used for each feature map in the C2 layer, so after mapping through this part, C3 layer has a total of 40 feature maps. The convolution kernel is set to [1×10], and the size of each feature map is (1×6). The reason for setting the convolution stride to be the same as the convolution kernel length is to reduce parameters to prevent overfitting, while achieving convolution and downsampling at the same time.

(4) Fourth layer (F4): This layer is the fully connected layer (the third hidden layer), which, in conjunction with the previous layer and the output layer, forms the classification part, so this layer is fully connected both before and after. The number of neurons is set to 100.

(5) Fifth layer (O5): This layer is the output layer, containing 2 neurons, representing a binary classification problem (left hand motor imagery or foot motor imagery).

During the CNN training phase, each subject's data is used individually to train their respective classification models. The dataset is divided into 5 parts, with 3 parts as the training set (60% of the data), 1 part as the validation set (20% of the data), and 1 part as the test set (20% of the data). The training set is used for model construction, the validation set for model optimal parameter selection, and the test set for model recognition rate assessment.

Researchers compared the proposed CNN model with traditional classification methods such as power value + SVM, CSP + SVM, and MRA + LDA and found that the average recognition rate of the CNN method (90.75% ±2.47%) is higher than the latter three methods (82.25% ± 1.06%, 85.75%±1.77%, and 87.75%±2.47%).

5. Brain-Computer Interface and Exoskeletons Control

Subjects wear devices capable of collecting EEG signals, such as EEG caps, and begin to perform certain motor imagery tasks. The collected EEG signals are processed through a convolutional neural network to extract features. When the wearer performs the same motor imagery again, the collected EEG signals are matched with the learned signals, and then the derived signal categories are transmitted to the exoskeleton device via Bluetooth. The exoskeleton device executes the corresponding actions and sends the execution status back to the host part. The content of the motor imagery must match the actions the exoskeleton is to perform. For example, if the lower limb exoskeleton is to perform "standing," the subject imagines the action of "standing." At this time, the device collects EEG signals and learns the features, ultimately determining a "standing" EEG signal. Thus, when the subject imagines "standing" again, by comparing the EEG signals, the computer derives the "standing" instruction and sends this command to the exoskeleton, which also has a program set to execute the standing movement and carries it out, thereby driving the exoskeleton.

Case: Composition of the Assistive Exoskeleton Brain-Computer Interface System [6]:

The lower limb motor imagery brain-computer interface system consists mainly of an upper computer part responsible for decoding EEG motor intentions and a state machine part for controlling the exoskeleton to complete the transition from motor intention to executing actions. The system is composed as shown in the Figure 1 below. The state machine selects executable commands from a preset control command set for the current state of the exoskeleton and finally sends them to the exoskeleton to perform actions. During online experiments, the subject wears an EEG cap, and the upper computer part of the brain-computer interface system collects EEG data, while the state machine detects the exoskeleton's status and establishes a connection with it. When EEG signal collection begins, the subject starts performing the corresponding motor imagery task.

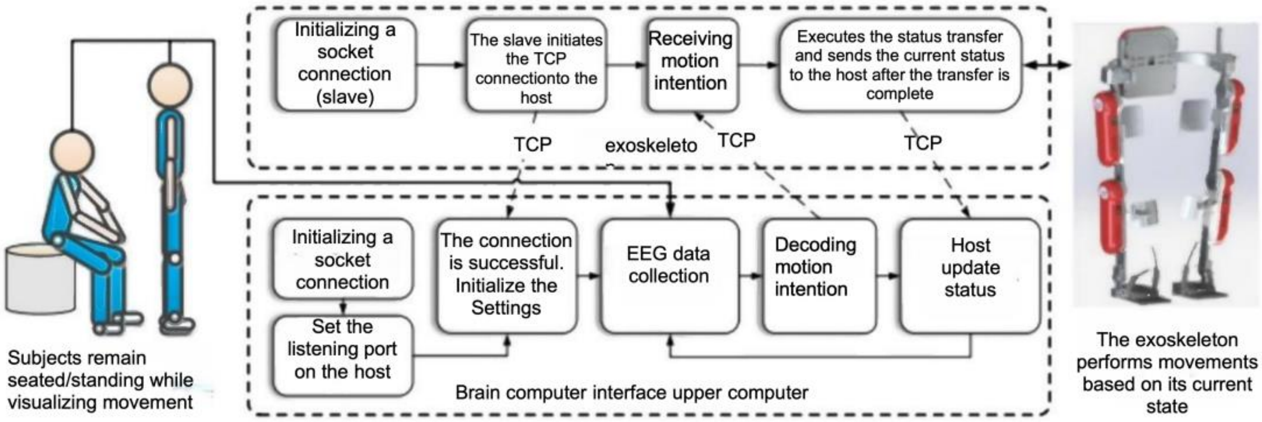

Fig. 1 Communication flow of lower limb movement imagination brain computer interface system [6]

The communication process between the two components of the assistive exoskeleton brain-computer interface is shown in the figure below. When the upper computer obtains the motion intention decoded by the motor imagery classification algorithm, it encodes the motion intention and transmits it to the state machine through a TCP connection. In this transmission process, the upper computer, which decodes the motor imagery EEG signals, is defined as the server, and the state machine is the client. It is required that the upper computer and the state machine be in the same local area network and communicate through a pre-defined port. The state machine communicates with the exoskeleton via Bluetooth.

The exoskeleton sends data packets on a broadcast channel, and after receiving the broadcast, the state machine sends a scan request containing the device address to the exoskeleton, which then returns a scan response and opens the connection. Subsequently, whenever the state machine (host) receives a motion intention, it transmits control commands to the exoskeleton robot via Bluetooth based on the current status of the exoskeleton. For online testing, the Aider lower limb assistive exoskeleton is used.

Considering safety factors, the exoskeleton robot is designed as a finite state machine, including three states: sitting/resting state, standing/reset state, and walking/stepping state.

The state transitions between the three states are controlled by the motion intentions (stand up, walk, sit down) obtained, but these state transitions are restricted by the current state of the exoskeleton robot. Under the constraints of this state machine, the exoskeleton robot will not directly transition to the sitting state while in the stepping state. Suppose the exoskeleton robot receives a command to sit down while in the stepping state.

6. Discussion and Prospect

6.1. Advantages of Convolutional Neural Networks

Convolutional Neural Networks (CNNs) leverage their powerful computational capabilities and strengths in recognizing and processing two-dimensional images, making them widely applied in learning electroencephalogram (EEG) signals. CNNs possess a multi-layered structure that allows for integration with other machine learning methods. For instance, in popular research, researchers often employ integrated neural networks such as Compact CNN [7], CNN+RNN (Recurrent Neural Networks) and CNN+GAN (Generative Adversarial Networks) to enhance learning capabilities and classification accuracy.

6.2. Limitation

Traditional Convolutional Neural Networks (CNNs) have made significant strides in the field of EEG signal processing, but they still face challenges. EEG signals are influenced by various factors during collection, such as errors inherent in the collection equipment, individual differences among subjects, and even interference from the same subject's varying mental states at different times [1]. These factors can directly affect the quality of the collected EEG signals, leading to a scarcity of high-quality EEG data and a bias towards a small sample size. Moreover, traditional CNNs are limited by the size of the training set; they require a large amount of labeled data for training, otherwise their performance may decline. Furthermore, as mentioned in the case in Chapter 3, current neural networks can only distinguish between significantly different human activities such as lifting the left hand and standing. There is still much room for research in terms of detailed classification.

6.3. Development Direction

6.3.1. Optimization of neural network structure

The performance of a single Convolutional Neural Network (CNN) is limited. Currently, hybrid neural networks are more commonly used, which have stronger computational capabilities and classification abilities, promising to improve the accuracy rate under various classification conditions. At the same time, they can also integrate cutting-edge technologies such as signal twins.

6.3.2. Data sources

The data currently used for experimental training comes from a special database, and there may be a certain difference between the EEG signals of the patients themselves and the results trained by the theoretical model. There are also differences between different patients. The direction of making the same brain-computer interface widely applicable to the population will be one of the future research directions.

6.3.3. Training time and reaction time

Currently, the training of brain-computer interface (BCI) systems requires a long time and also necessitates high-performance processors. There is still significant room for improvement in the latency time it takes for the neural network to collect EEG signals and relay them to the exoskeleton to execute movements.

7. Conclusion

In this study, we review the fundamental concept of brain-computer interface (BCI), electroencephalograms (EEGs) and convolutional neural networks (CCN). There are also analysises of the process of convolutional neural networks (CNNs) in processing motor imagery EEG signals and the entire brain-computer interface (BCI) system that transmits the output signals to the exoskeleton for actuation. Raw signal data were used extensively with deep learning methods, with or without minimal preprocessing. Currently, exoskeleton systems based on BCIs have made significant strides in achieving simple activities through motor imagery under the learning classification of CNNs. This remarkable advancement has opened up new possibilities for individuals with limited mobility or physical disabilities, as they can now control external devices using their thoughts alone. Furthermore, exoskeleton systems based on BCIs can already achieve some simple activities through motor imagery under the learning classification of CNNs. In future research, the BCI-exoskeleton system can be further improved by optimizing the neural network structure, improving data sources, and optimizing training time.

References

[1] Altaheri, H., Muhammad, G., Alsulaiman, M., Amin, S.U., Altuwaijri, G.A., Ab- dul, W., Bencherif, M.A., Faisal, M.: Deep learning techniques for classification of electroencephalogram (eeg) motor imagery (mi) signals: A review. Neural Computing and Applications, 2023, 35(20): 14681-14722.

[2] Mwata-Velu, T., Avina-Cervantes, J.G., Cruz-Duarte, J.M., Rostro-Gonzalez, H., Ruiz-Pinales, J.: Multiclass Classification of Visual Electroencephalogram Based on Channel Selection, Minimum Norm Estimation Algorithm, and Deep Network Architectures. Sensors, 2024, 24(12): 3968.

[3] Yu SQ, Wang ZD, Wang F, Chen K, Yao DZ, Xu P, Zhang Y, Wang HS, Zhang T.: Multiclass classification of motor imagery tasks based on multi-branch convolutional neural network and temporal convolutional network model, Cerebral Cortex, 2024, 34(2).

[4] Luo FG, Song Q, Qin YC, Shi YY.: Research on the application of convolutional neural network in image recognition. Computer and information technology, 2024, 32(03): 51-54.

[5] Tang ZC, Zhang KJ, Li C, Sun SQ, Huang Q, Zhang SY.: Motion imagination classification based on deep convolutional neural networks and its application in brain-controlled exoskeletons. Journal of Computer science, 2017, 40(6): 1367-1378.

[6] Qin YG: Research on lower limb motion imagination classification algorithm for walking exoskeleton brain-computer interface. UESTC (University of Electronic Science and Technology of China), 2023.

[7] Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ.: EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces. J Neural Eng, 2018, 15(5): 56013.