論 文 内 容 の 要 旨

Electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) are neuroimaging techniques used to study brain function. fMRI provides high spatial resolution, allowing for detailed visualization of brain structures, while EEG offers high temporal resolution, allowing for the study of rapid brain processes. These techniques are crucial for understanding various cognitive processes, such as perception, attention, memory, and language. Deep learning (DL) algorithms as high-performance tools for pattern recognition and regression problem have been increasingly applied to fMRI and EEG data. DL-based techniques can automatically learn complex and meaningful representations from the raw data, which can extract high-level features from the voxel-level fMRI volumes and be applied to tasks such as brain state classification or decoding cognitive processes. For EEG analysis, DL-based techniques can learn temporal patterns and relationships in the electrical signals, facilitating tasks like emotion recognition or seizure detection. Therefore, this thesis conducted research based on deep learning algorithms from three perspectives, including P300 detection (a type of EEG signal), EEG denoising, and fMRI classification, and proposed a spatial-temporal neural network (STNN), a multi-module neural network (MMNN), and a multi-pooling 3D convolutional neural network (MP3DCNN) to resolve these challenges. Specifically, the first chapter provided a brief overview of EEG, fMRI, and deep learning algorithms, where the definition, characteristics, and application of EEG and fMRI signals are given. Then the principle, key concepts, and real applications of deep learning algorithms are introduced. Finally, the potential and risks of EEG and fMRI signal processing based on deep learning are listed. From the second to the fourth chapter, the background, materials, methods, experiments, and discussion about the proposed algorithms will be described, respectively. The fifth chapter summarizes this thesis. The main works are brief described as follows:

1) Spatial-temporal neural networks (STNN) for P300 detection: The P300 signal, also known as P3 or P300 component, is an electrophysiological response observed in the event-related potential (ERP) waveform recorded from the scalp using EEG tools. It typically occurs around 250-500 milliseconds after the presentation of a rare or unexpected stimulus and is associated with attention, cognitive processes, and decision-making. P300 spellers are common brain-computer interface (BCI) systems designed to transfer information between human brains and computers. In most P300 detections, the P300 signals are collected by averaging multiple electroencephalographic (EEG) changes to the same target stimuli, so the participants are obliged to endure multiple repeated stimuli. Therefore, an STNN is proposed to detect P300 signals, which can detect signals by combining the outputs from a temporal unit and a spatial unit. The temporal unit is a flexible framework consisting of several temporal modules designed for analyzing brain potential changes in the time domain. The spatial unit combines one-dimensional convolutions (Conv1Ds) and linear layers to generalize P300 features from the space domain, and it can decode EEG signals recorded using different numbers of electrodes. We demonstrate the effectiveness of our model using three public databases: P300 speller with ALS patients, covert and overt ERP-based BCI, and BCI Competition III-dataset II. In the dataset of P300 speller with ALS patients, EEG signals from eight ALS patients were recorded, and every participant in the study went through 35 trials, with 10 rounds of repeated stimuli in each trial. Every round of stimuli contained two target stimuli and 10 nontarget stimuli. In the dataset of covert and overt ERP-based BCI, 10 healthy subjects took part in the experiment. The EEG data are recorded on Farwell and Donchin's paradigm and the Geometric Speller. Each experiment included three sessions, with six trials in each session. Each trial contained eight rounds of repeated stimuli with 12 stimuli within each round of stimuli. In the dataset of BCI Competition III-dataset II, both EEG signals from two subjects (A and B) were divided into a training set (85 trials) and a testing set (100 trials). Every trial contained 15 rounds of repeated stimuli. We implemented a within-subject P300 detection and a cross-subject P300 detection, respectively using the dataset of P300 speller with ALS patients as well as covert and overt ERP-based BCI.

The results showed that both amyotrophic lateral sclerosis (ALS) patients and healthy subjects can benefit from this study. In the within-subject P300 detection and the cross-subject P300 detection, the proposed STNN gained higher performance with fewer repeated stimuli than other comparative approaches. Furthermore, we applied the proposed STNN in the P300 detection challenge of BCI Competition III. The accuracy score was 89% in the fifth round of repeated stimuli, outperforming the best result in the literature (accuracy = 80%) to the best of our knowledge. The results demonstrate that the proposed STNN performs well with limited stimuli and is robust enough for various P300 detections. The main reasons are as follows: 1) the temporal unit, as a flexible DL-based network dedicated to time-domain modeling, can capture the temporal dependencies from brain potential changes by constructing an end-to-end multi-level sequential mapping, so it is more sensitive than the previously mentioned approaches when detecting P300 signals; 2) the spatial unit can constantly generalize and compress P300 features in the space domain, which hedges complex noise interference to a certain extent; 3) a joint decision-making mechanism is built into the network by connecting the temporal unit and the spatial unit concurrently, which can utilize the above advantages of the two units, thus achieving both better performance and stronger robustness. In the future, the proposed STNN is predicted to reach a high information transfer rate (ITR) when implementing online P300 detection. Moreover, we consider that this network has potential for applications in EEG-BCI systems and some other areas of signal processing, such as Electrocardiogram (ECG) classification, seeing that it is designed with a flexible structure and can be fast training and testing with limited data.

2) Multi-module neural networks (MMNN) for EEG denoising: This study proposed an MMNN to remove ocular artifacts (OAs) and myogenic artifacts (MAs) from noisy single-channel electroencephalogram (EEG) signals. This network consists of multiple denoising modules connected in parallel. Each denoising module is built using one-dimensional convolutions (Conv1Ds) and fully connected (FC) layers, and it estimates not only clean EEG signals but also artifacts. The proposed MMNN has two main advantages. First, the multiple denoising modules can purify noisy input EEG signals by continuously removing artifacts in the forward propagation. Second, the parallel architecture allows the parameters of each denoising module to be updated concurrently in the backpropagation, thereby improving the learning capacity of neural networks. We tested the network denoising performance using a recent public database, namely, EEGdenoiseNet. This database provides large-scale clean EEG and artifact epochs, involving 4514 clean EEG epochs, 3400 EOG epochs, and 5598 EMG epochs. These epochs were used to synthesize the training and testing data. Among them, we implemented the model evaluation using 3000 pairs and 400 pairs of training and testing epochs for the OA removal, as well as 5000 pairs and 598 pairs of training and testing epochs for the MA removal. These epochs are synthesized at ten different noise levels, aiming to simulate the real applications. And we performed a 10-fold cross-validation on the training samples for hyperparameter tuning.

The testing results revealed that the proposed network reduced the temporal relative root mean square error (T-RRMSE) and spectral relative root mean square error (S-RRMSE) by at least 6% and enhanced the correlation coefficient (CC) by at least 3% over the state-of-the-art approaches. Observing the deviation distribution between the denoised and clean signals confirmed these significant performance improvements. Furthermore, the proposed network achieved a similar performance efficiency with only 60% of the training data compared to the existing DL models. Finally, the proposed model was compared with the non-deep learning techniques. According to the ANOVA results with Holm-Bonferroni correction, the performance improvement is significant (all p-values < 0.001) in both the OA and MA removals. In the future, there are some challenges worth exploring using the proposed model. For example, OAs and MAs are entangled with motion artifacts in a real EEG epoch, however, there is no available public database to evaluate the model performance for the mixed signals. Given that the proposed model offers significant advantages over the conventional and DL models in this study, the related research is within the scope of further work.

3) Multi-pooling 3D Convolutional Neural Networks (MP3DCNN) for fMRI Classification: Neural decoding of visual object classification via functional magnetic resonance imaging (fMRI) data is challenging and is vital to understand underlying brain mechanisms. The previous study performed categorical (face vs. object), face sub-categorical (male face vs. female face), and object sub-categorical (natural object vs. artificial object) classifications via a classic three-layer 3DCNN, revealing that the human visual system recognizes objects following the principle of going from categories into sub-categories. However, the previous classification model did not significantly present a high accuracy even with 9-fold fMRI data averaging, especially for sub-categorical classification tasks. Therefore, a novel multi-pooling 3D convolutional neural network (MP3DCNN) is proposed, which is expected to reach a higher accuracy than the previous model and play a valuable role in decoding brain mechanisms. The proposed MP3DCNN included a feature extraction, a feature combination, and a classifier, where the feature extraction has a mainchain and two branches. The mainchain is a three-layer 3DCNN, where each 3D convolution is combined with a batch normalization, an average 3D pooling layer, and a rectified linear unit (ReLU). The first and second 3D convolutions each have a branch connection, where an average 3D pooling layer and a linear layer are used to generalize further and connect the extracted features. In the end, through the feature combination and classifier, the model can provide a sophisticated decision using multi-level features in fMRI classification.

In the work, the fMRI dataset is from the previous study, where 53 healthy subjects participated in the visual stimulus task, and each one experienced an average of 9.96 ± 2.88 rounds of visual stimuli, where the subject clicked the button corresponding to a random visual stimulus (an image of a male face, female face, natural object, or artificial object) within 0.5s. During this period, a Siemens 3T MRI scanner recorded the subject's brain states as T1-weighted 3D fMRI volumes. Through SPM12, the fMRI volumes were realigned, co-registered, normalized to the standard Montreal Neurological Institute (MNI) template, and resampled to 2-mm isotropic voxels. After data cleaning, there are 17306 fMRI volumes from 50 subjects available, including 4453, 4399, 4214, and 4240 volumes corresponding to the visual stimuli of male face, female face, natural object, and artificial object, respectively. To suppress the background noise and the irrelevant neural activities, the fMRI dataset of each subject was multi-fold averaged as an option to improve the data quality (for example 9-fold fMRI data averaging). We performed 9-fold cross-validation with 25 training iterations, where the batch size was 64, the learning rate was 0.00001, and the loss function was binary cross entropy (BCE). Within the 25 iterations, the model parameters with the highest accuracy score on each validation dataset were used for model testing. Finally, we used a majority voting scheme to ensemble the nine results in determining the classification accuracy.

The results showed that this model can improve the classification accuracy for categorical (Face vs. Object), face sub-categorical (Male face vs. Female face), and object sub-categorical (Natural object vs. Artificial object) classifications from 1.684% to 14.918% over the previous study in decoding brain mechanisms. The main reasons are speculated as follows: 1) we considered that the multiple 3D average pooling layers can against the local feature redundancy in the feature extraction process and pass the global information to the classifier as much as possible; 2) through the branch connections, the model decision can depend on the merged features of the three 3D convolutions, thereby improving the model's robustness. In future research, we look forward to using grid-search to optimize the model hyperparameters and exploring the visual explanations based on the reached classification results.

Overall, this thesis proposed high-performance deep learning algorithms for P300 detection, EEG denoising, and fMRI classification, respectively. And they are expected to improve the efficiency in various brain decoding tasks in the future.