Enhancing EEG and sEMG Fusion Decoding Using a Multi-Scale Parallel Convolutional Network With Attention Mechanism

Xianlun Tang, *Senior Member, IEEE*, Yidan Qi, Jing Zhang, Ke Liu, Yin Tian, and Xinbo Gao, *Fellow, IEEE*

***Abstract*— Electroencephalography (EEG) and surface electromyography (sEMG) have been widely used in the rehabilitation training of motor function. However, EEG signals have poor user adaptability and low classification accuracy in practical applications, and sEMG signals are susceptible to abnormalities such as muscle fatigue and weakness, resulting in reduced stability. To improve the accuracy and stability of interactive training recognition systems, we propose a novel approach called the Attention Mechanism-based Multi-Scale Parallel Convolutional Network (AM-PCNet) for recognizing and decoding fused EEG and sEMG signals. Firstly, we design an experimental scheme for the synchronous collection of EEG and sEMG signals and propose an ERP-WTC analysis method for channel screening of EEG signals. Then, the AM-PCNet network is designed to extract the time-domain, frequency-domain, and mixed-domain information of the EEG and sEMG fusion spectrogram images, and the attention mechanism is introduced to extract more fine-grained multi-scale feature information of the EEG and sEMG signals. Experiments on datasets obtained in the laboratory have shown that the average accuracy of EEG and sEMG fusion decoding is 96.62%. The accuracy is significantly improved compared with the classification performance of single-mode signals. When the muscle fatigue level reaches 50% and 90%, the accuracy is 92.84% and 85.29%, respectively. This study indicates that using this model to fuse EEG and sEMG signals can improve the accuracy and stability of hand rehabilitation training for patients.**

***Index Terms*— Rehabilitation training, signal fusion, multi-scale parallel network, fusion decoding, muscular fatigue.**

Manuscript received 21 July 2023; revised 9 December 2023; accepted 23 December 2023. Date of publication 26 December 2023; date of current version 16 January 2024. This work was supported in part by the National Nature Science Foundation of China under Grant 61673079, in part by the Natural Science Foundation Project of Chongqing under Grant CSTB2022NSCQ-MSX0380, and in part by the Science and Technology Research Program of Chongqing Municipal Education Commission under Grant KJZD-M202200603 and Grant KJQN202200626. *(Corresponding author: Xianlun Tang.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Research Ethics Committee of Chongqing University of Posts and Telecommunications.

Xianlun Tang is with the Chongqing Key Laboratory of Complex Systems and Bionic Control, Chongqing University of Posts and Telecommunications, Chongqing 400065, China, and also with the Guangyang Bay Laboratory, Chongqing Institute for Brain and Intelligence, Chongqing 400064, China (e-mail: tangxl@cqupt.edu.cn).

Yidan Qi, Jing Zhang, and Ke Liu are with the Chongqing Key Laboratory of Complex Systems and Bionic Control, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: S210301041@stu.cqupt.edu.cn).

Yin Tian and Xinbo Gao are with the Guangyang Bay Laboratory, Chongqing Institute for Brain and Intelligence, Chongqing 400064, China (e-mail: tianyin@cqupt.edu.cn; gaoxb@cqupt.edu.cn).

Digital Object Identifier 10.1109/TNSRE.2023.3347579

I. Introduction

THE prevalence of cerebrovascular disease and the occurrence of frequent accidents have contributed to a rising population of individuals with paralysis [1]. For paralyzed patients, postoperative rehabilitation is an effective treatment that can help improve and restore movement motor function. Hand exercise rehabilitation training is essential when patients relearn daily movements [2]. The widely used single mode Electroencephalography (EEG) or Surface Electromyography (sEMG) signals cannot fully meet the requirements of effective patient control. Integrating surface electromyography (sEMG) and electroencephalography (EEG) for fusion recognition and decoding holds theoretical promise in enhancing the classification accuracy of single-mode actions, thereby offering a novel approach to hand sports injury rehabilitation. This method aims to facilitate hand rehabilitation training and improve the overall effectiveness of rehabilitation interventions for patients.

The EEG signal is the reflection on the surface of the cerebral cortex when the brain neurons are active, which can be used to decode action information [3]. Given that the paralysis and stroke patients with EEG signal performances are not compromised, the Brain-computer interface (BCI) technology can be used to control objects such as wheelchairs and prosthetic hands to improve Patient's lost body function [4], [5]. Recent research has shown that BCI technology can help paralyzed patients with motor rehabilitation, such as patients after stroke) [6]. However, due to the limited number of classifiable modes for a single EEG mode, precise control of output devices is restricted. EEG signals are susceptible to external noise interference, and the low classification accuracy and poor user adaptability often limit the application of EEG signals in clinical patient rehabilitation.

Multifunctional prosthetics are often used for patients with muscle function loss to restore their lost motor function. sEMG is a bioelectrical signal generated by neuromuscular activity recorded from the surface of skeletal muscle through

electrodes when the human body moves autonomously [7]. The sEMG signal contains rich neural information, which can extract features from fewer channels that can control multiple action modes. The sEMG signal plays a vital role in the control of modern mobile prosthetics [8] and rehabilitation robots [9]. It is important to acknowledge that certain limitations may arise due to variations among subjects and the specific application environment. These limitations may include issues such as muscle fatigue arising from prolonged usage and the inability of subjects to generate consistent and sufficient sEMG power due to muscle weakness or disability.

Fusing multimodal signals is a feasible method to improve the accuracy and stability of classification. By combining EEG signals with sEMG signals [10], sufficient information is provided for motion decoding. According to the level of information fusion, the fusion methods can be categorized into three groups: data layer fusion, feature layer fusion, and decision layer fusion. The first group is data layer fusion, which involves directly fusing the EEG signal and sEMG signal data obtained from different acquisition sensors, and then feature extraction and classification are performed on the fused signal data. The obtained information has a certain degree of redundancy. The information loss of data layer fusion is the smallest, but the fault tolerance is the smallest and the anti-interference ability is the worst. The second group is feature layer fusion, where feature vectors are extracted from the obtained EEG and sEMG signals, followed by feature data fusion processing. Finally, the fused features are used for the decision-making of system classification. The feature layer fusion extracts effective features from various channels of EEG and sEMG data, preserving useful information while compressing it, resulting in high accuracy. Yang et al. [11] proposed a method based on graph theory as a multimodal fusion strategy for EEG and sEMG. Functional connectivity, often considered as the weight of edges, enhances the robustness and accuracy of hand motion recognition. The third group is decision layer fusion, which involves separately processing and classifying EEG and sEMG signals for decision-making. Both EEG and sEMG signals are required to have independent decision-making capabilities, with the greatest information loss, ignoring the synergistic complementarity between EEG and sEMG. Tryon et al. [12] proposed a decision layer fusion strategy based on EEG and sEMG sources, which achieved flexion–extension motion recognition and improved the accuracy and stability of the system.

At present, feature extraction plays a very important role in the research of EEG and sEMG signals. Ji et al. [13] proposed a feature extraction method based on discrete wavelet transform (DWT) and empirical mode decomposition (EMD) to improve the effectiveness of EEG signals. Li et al. [14] proposed an effective feature fusion method TS-SEFFNet to enhance the temporal and spectral dependencies in MI-EEG. Zhu et al. [15] designed a network framework that combines CWT with AlexNet, and collected sEMG signal data from various gesture actions to extract rich time-frequency domain features of sEMG signals.

With the continuous development of brain science technology, multiple researchers have proposed the hybrid BCI system that combines EEG signals and sEMG signals. The proposal compensates for the shortcomings of existing brain-computer interfaces. Li et al. [16] merged EEG and sEMG into parallel control inputs, extracted four time-domain features, and inputted them into the linear discriminant analysis (LDA). Combined with the sequential forward selection (SFS) algorithm to optimize performance, the highest recognition accuracy was 87.0%. Chowdhury et al. [17] used the correlation between power-limited time processes as the fusion feature of EEG and sEMG to classify hand movements. For the disabled patient group, the accuracy was 84.53 Â± 4.58%. Shi et al. [18] proposed a multimodal enhanced fusion network based on a dense non-attention mechanism and introduced the Joint attention structure with an accuracy of 88.44%. However, existing methods all have some shortcomings. First, the coherence and functional coupling relationship between EEG signals and sEMG signals are ignored [19]. Secondly, feature extraction is mostly a machine learning method, which can result in the loss of some features [20]. In this work, we propose the Attention Mechanism-based Multi-Scale Parallel Convolutional Network (AM-PCNet) for recognizing and decoding fusion signals to improve the effective control of patient rehabilitation training.

The four main contributions of this paper can be summarized as follows:

1) We propose an EEG channel selection method based on ERP-WTC analysis. Three channels of EEG signals are selected and fused with sEMG signals in the feature layer to obtain time-frequency domain fusion features of EEG and sEMG signals.
2) We propose a novel AM-PCNet for EEG and sEMG fusion decoding, which uses a parallel structure to extract features in the time, frequency, and mixed domains. In this way, the loss of multi-domain representations in a single-branch network is preserved.
3) Pyramid Split Attention (PSA) and Squeeze Excitation (SE) attention mechanisms, which enable the network to extract fine-grained multi-scale features of EEG and sEMG signals effectively, are introduced to establish long-distance channel dependencies and enhance information exchange between EEG and sEMG signals.
4) We design a synchronous collection scheme for the EEG and sEMG signals. Experiments are conducted on the dataset collected in the laboratory, and the rationality and effectiveness of the multimodal fusion decoding method are analyzed and verified.

II. Methods

To integrate the advantages of human-machine interaction between a single EEG and sEMG modality and improve system performance, we researched the fusion and recognition of EEG and sEMG signals based on attention mechanism multi-scale parallel convolutional neural networks.

A. The Design of Synchronous Collection Scheme for EEG and sEMG Signals

We independently designed a synchronous acquisition scheme for EEG signals and sEMG signals, achieving

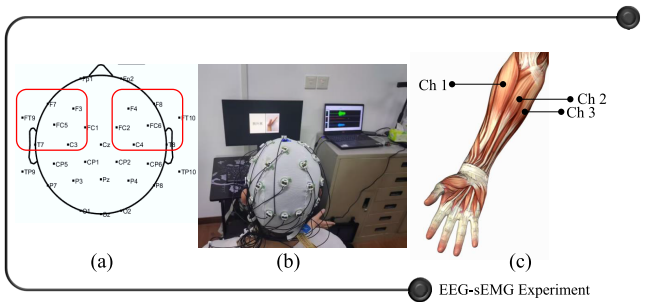


Fig. 1. Synchronous collection of EEG and sEMG signals (a) Distribution map of electroencephalogram electrodes, (b) The EEG and sEMG collection system, (c) Muscle position distribution map.

synchronous triggering and collection of EEG and sEMG signals and storing real-time signals. This scheme is implemented by a set of sensor devices for collecting sEMG and EEG signals, a device for signal visualization and storage, and a computer for displaying experimental paradigms and sending trigger signals. During the synchronous collection process, the E-prime software synchronously sends a serial signal at the beginning of the action. Then it sends it to the sEMG signal collection device through a trigger. When the trigger command is received, the EEG and sEMG signals start recording and storing synchronously. The synchronous collection process of electroencephalogram signals is shown in Fig. 1(b). In the experiment, Brain Products equipment was used to collect EEG signals, and 31 channels of EEG signals were recorded from the scalp electrode of the EEG. The channel position information is shown in Fig. 1(a). We select the Delsys device to collect sEMG signals and record three channels of sEMG signals from the Delsys wireless acquisition sensor. Fig. 1(c) shows the muscle selection positions, where Ch1 represents the flexor carpi brevis, Ch2 the palmaris longus, and Ch3 the flexor carpi ulnaris.

B. Screening of EEG Channels Based on ERP-WTC

Although BCI systems with a large number of electrodes can record more working features of brain regions, it is necessary to consider the response of EEG electrodes to different types of stimuli and the coherence between EEG and sEMG signals [19], in fusion motion recognition systems. We screen EEG channels to achieve the optimal combination of EEG and sEMG signals channels and prepare for subsequent fusion at the feature layer.

1) ERP Analysis: ERP (Event Related Potential) refers to the potential changes in the brain area caused by applying a specific stimulus to the sensory system or a certain part of the brain when the stimulus is given or withdrawn [21]. For the hand movements in the experiment, We use the ERP topographic map analysis to select activated brain regions. The activation status at the spatial scale corresponding to the peak of the ERP waveform can be observed in the ERP2D or 3D brain topographic map shown in Fig. 2. ERP2D and 3D brain topographic maps can select activated brain regions and detect faulty or noisy electrodes, so we preliminarily select the two activated regions of interest in the red box in Fig. 1(a). Based on the specific analysis of the ERP waveform curve in Fig. 2,

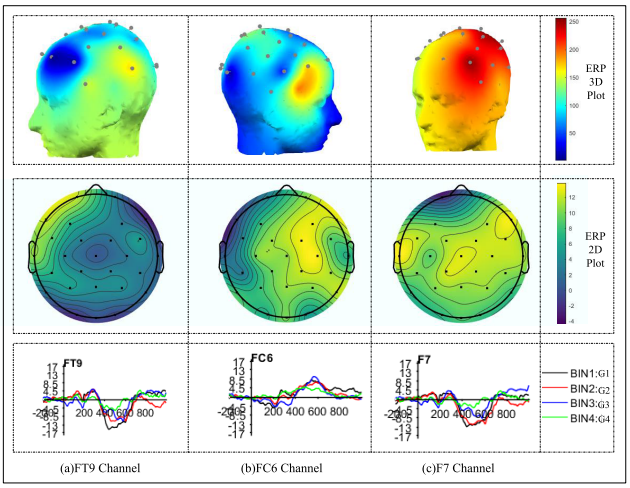


Fig. 2. The activation of EEG channels (a) FT9 channel, (b) FC6 channel, (c) F7 channel.

six channels, FT9, F7, C3, FC6, F8, and F4, were preliminarily selected from the two activated regions.

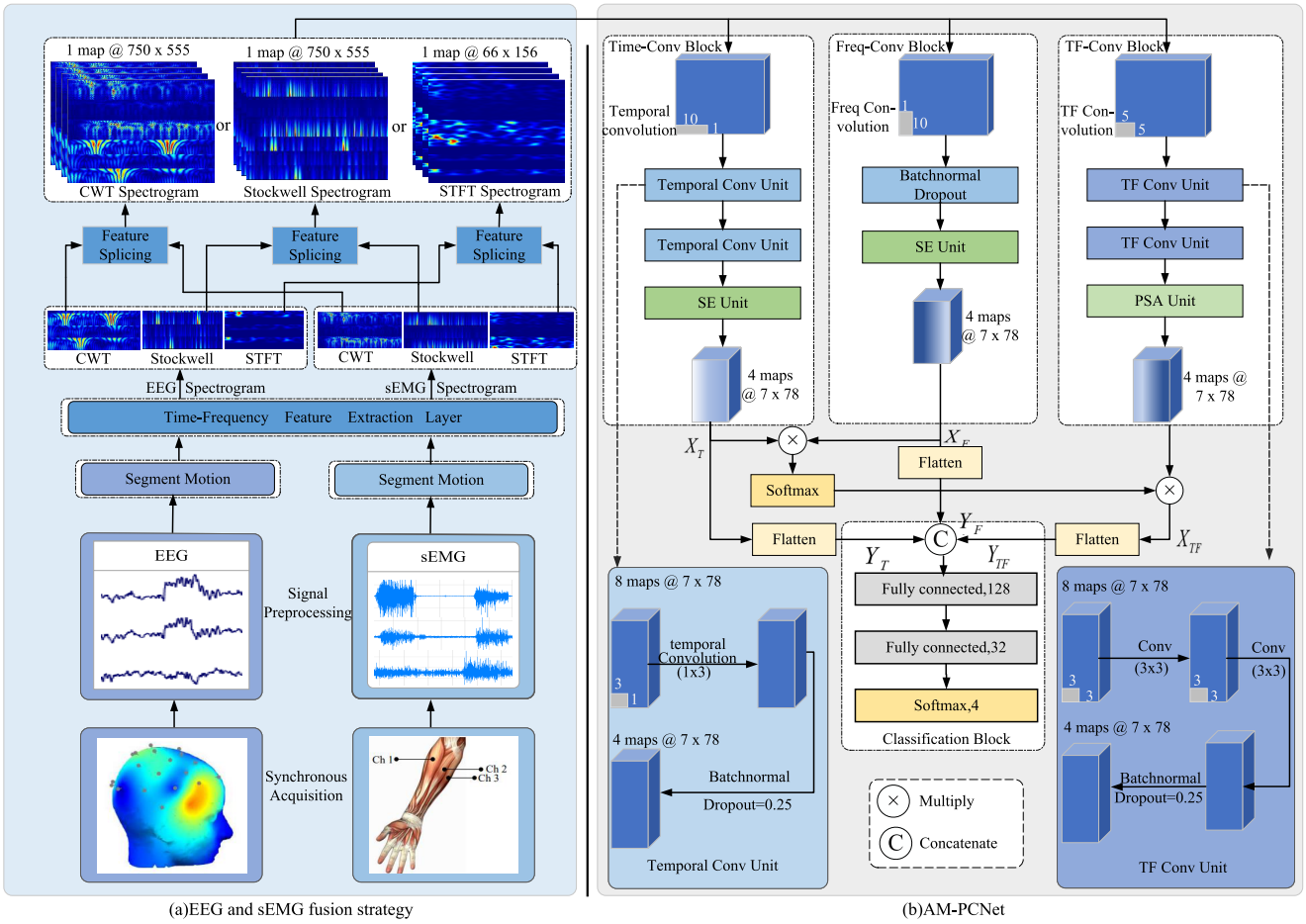
2) WTC Analysis: In the EEG and sEMG signals fusion motion recognition system, the brain's motor cortex sends commands to control the limbs and completes the muscles through the brain stem and the spinal cord along the motor nerve pathway exercise [19]. Simultaneously, the limb signals are sent back to the cerebral cortex along the sensory nerve pathway for fusion analysis. This interaction between cerebral cortex activity and muscle movement can be evaluated through the consistency of EEG and sEMG. Thus, it is possible to screen the EEG channels most relevant to sEMG signals of different muscle movements. Wavelet coherence analysis: spectral coherence [19] measures the correlation degree of two signals in the frequency domain. The formula is as follows:

$$Coh_{xy} = \frac{|Pxy(f)|^2}{Pxx(f) \bullet Pyy(f)} \tag{1}$$

Pxy refers to the cross-spectral density of EEG and sEMG signals, *Pxx* and *Pyy* refer to the self spectral density of *x* and *y* respectively, that is, the power spectral density. The range of *Cohxy* values is [0,1], and the larger the value, the greater the degree of correlation between EEG and sEMG signals at frequency f. The coherence analysis reflects the consistency of the EEG and sEMG signals. Informed by coherence analysis, FT9, F7, and FC6 EEG channels are finally selected.

C. Extracting Synchronous Features of the EEG and sEMG Signals

We comprehensively consider the common features information of EEG and sEMG signals and use time-frequency spectrogram images to demonstrate the signal oscillation behavior of EEG and sEMG signals. By maximizing the preservation of multi-domain information of EEG and sEMG signals, the fusion of EEG and sEMG feature layers is achieved. We use short-time Fourier transform [22], wavelet transform [15], and Stockwell transform [23] to perform time-frequency transformation on the EEG and sEMG signals in each channel.



(a)EEG and sEMG fusion strategy (b)AM-PCNet

Fig. 3. The EEG and sEMG fusion strategy based on AM-PCNet (a)The Overview of EEG and sEMG fusion strategy, (b) The structure of AM-PCNet.

1) Short Time Fourier Transform: As non-stationary signals, the internal characteristics of sEMG and EEG can be fully characterized by STFT. STFT divides the signal into smaller segments, computes the Fourier transform of each segment, and visualizes the frequency characteristics over time. Fourier transform can be applied to the sliding window to obtain the following local spectra:

$$STFT_x(\tau, f) = \int_{-\infty}^{+\infty} [x(t)\, w(t-\tau)]\, e^{-j2\pi f t} dt \quad (2)$$

where $x(t) \in L^2(R)$ is the original signal, $t, \tau \in R$ refer to time, $f \in R$ is the frequency, and $w(t) \in L^2(R)$ is the applied window function. The selected window in STFT is the Hanning window.

2) Continuous Wavelet Transform: The ψ in the continuous wavelet transform is called the fundamental wavelet or mother wavelet. The continuous wavelet transform can be expressed as follows:

$$(W_\psi f)(a,b) = \langle f, \psi_{a,b} \rangle = \sqrt{a} \int_{-\infty}^{+\infty} \bar{\psi}\left(\frac{t-b}{a}\right) dt \quad (3)$$

In formula 3, a represents frequency, while b represents time or spatial position. The reciprocal of scale *1/a* corresponds to frequency ω.

3) Stockwell Transform: We attempt to use Stockwell transform for time-frequency feature extraction of EEG and EMG. The Stockwell transform formula for continuous signal x(t) is:

$$S_x(\tau, f) = \int_{-\infty}^{+\infty} x(t) \frac{|f|}{\sqrt{2\pi}} e^{\frac{-f^2(\tau-t)^2}{2}} e^{-i2\pi f t} dt \quad (4)$$

We preserve the frequency domain information of EEG signals ranging from 8-30Hz and the effective frequency domain features of sEMG signals ranging from 10-200Hz. Due to the synchronous collection of EEG and sEMG, the EEG and sEMG are synchronized in the time dimension after time-frequency transformation. We concatenate the EEG time-frequency image and the sEMG time-frequency image in the frequency domain dimension, with the three channels of sEMG on top of the three channels of EEG. The fused features of EEG and sEMG are saved as concatenated time-frequency images, as shown in Fig. 3.

D. Proposed AM-PCNet Structure

As shown in Fig. 3, we used the spectrogram images of the spliced EEG and sEMG signals as the input of AM-PCNet. We selected three different scales of convolutional kernels, namely 1×10, 10×1, and 5×5, to obtain the time-domain features, frequency-domain features, and time-frequency mixed

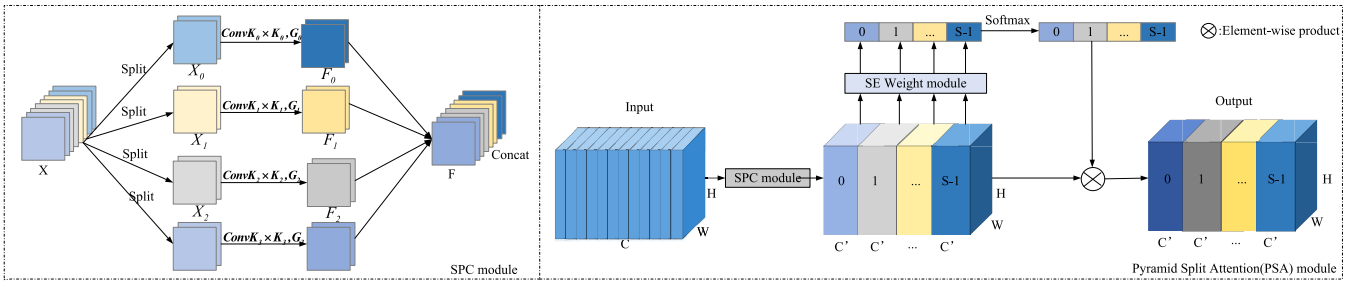


Fig. 4. The specific composition structure of PSA attention module.

domain features of the input feature map to better obtain rich network features [24].

1) The Design of the Time-Conv Block: The EEG and sEMG fusion spectrogram images contain a large amount of time-domain feature information, so a 1 × 10 convolutional kernel is first selected in this block to extract time-domain features. To further explore deeper time-frequency features, continuous convolutional units on the time dimension are used. The Time-Conv block in Fig. 3 includes two successive Temporal Conv Units, and the structure of the units is shown in Fig. 3. Each unit starts from a 1 × 3 convolution kernel, with a convolution stride size of (2,15). The Temporal Conv Unit is only convolved at the time scale, which can extract the internal time-domain features of the EEG and sEMG fusion spectrogram images without damaging the EEG-sEMG signals in each channel. In addition, dropout and batch normalization operations are used in each unit to minimize overfitting.

Time-Conv block arranges all units in order and extracts deeper feature representations from basic shallow time domain features. The number of units and the size of the kernel will be verified by subsequent related experiments.

2) The Design of the Freq-Conv Block: We concatenate the spectrogram images of EEG and sEMG signals in the frequency domain dimension. We design the Freq-Conv block to extract frequency domain features of EEG and sEMG signals while establishing long-range EEG channel dependencies. We select a 10 × 1 convolution kernel with a convolution step size of (2,15) and then perform dropout and batch normalization operations.

3) The Design of the TF-Conv Block: In the TF-Conv block of Fig. 3, a 5×5 convolutional kernel is first selected to extract the neighborhood features of each pixel in the spectrogram image. The time and frequency channels are mixed to extract the contextual features of the EEG and sEMG spectrogram images. In the TF-Conv block layout, there are two successive TF Conv Units, and the structure of each unit is shown in Fig. 3. Each unit starts from two successive convolution kernels with the size of 3 × 3, which perform convolution simultaneously in both time and frequency domains. It can extract correlation time-frequency features between each channel. The convolution stride is (2,15). The activation function is ReLu. In addition, dropout and batch normalization operations are used in each unit to reduce overfitting. The TF-Conv block arranges all units to extract time-frequency deep feature information further.

4) Attention Module: In the study of decoding sEMG-EEG signals, we aim to extract features with high discriminability and robustness. We introduce two attention models: SE (Squeeze Excitation) [25] and PSA (Pyramid Split Attention) [26].

We add the SE attention model to the Time-Conv block and Freq-Conv block, enabling the network to selectively amplify valuable time-domain and frequency-domain characteristic channels based on global information.

We added the PSA attention module to the TF Conv block. First, each group is convolved with different convolution kernel sizes through group convolution to obtain the Receptive field of different scales and extract information about different scales. The SPC module structure is demonstrated in Fig. 4.

$$F_i = Conv(k_i \times k_i, G_i)(X_i) \tag{5}$$

where the number of groups is set at 4, so i=0,1,2,3.

Next, through the SE attention module, we extract the weighted values of each group of channels and obtain spatial information of multi-scale input spectrogram images.

$$Z_i = SEWeight(F_i) \tag{6}$$

$$att_i = Soft\max(Z_i) \tag{7}$$

We multiply the feature map of the corresponding scale with the attention vector at the channel-wise level. We concatenate the weighted feature maps into dimensions, and the PSA module Output is shown in Fig. 4:

$$Out = Cat([Y_0, Y_1, \cdots, Y_{S-1}]) \tag{8}$$

where the number of S is set at 4, and Y is the output of the PSA module corresponding to the convolutional weighting of different scale groups.

The Time-Conv block and Freq-Conv block focus on some information in the spectrogram image, while the information obtained in the TF-Conv block is a mixture of time and frequency channel features. We use the similarity between time-domain and frequency-domain features to weight the attention of the time-frequency convolution module [27] so that the time-frequency convolution module focuses on time-frequency features, improves the interaction between the three parallel network branches, and enhances the accuracy and stability of the network.

$$att_i = Soft\max(X_{T_i} X_{F_i}{}^T) = \frac{\exp(X_{T_i} X_{F_i}{}^T)}{\sum_{i=0}^{K-1} \exp(X_{T_i} X_{F_i}{}^T)} \tag{9}$$

$$X_{TF_i} = x_{TF_i} \odot att_i \tag{10}$$

where the number of K is set at 4, so i=0,1,2,3. Where X_{T_i} is the output of Time-Conv Block, X_{F_i} is the output of Freq-Conv Block, and X_{TF_i} is the output of TF-Conv Block.

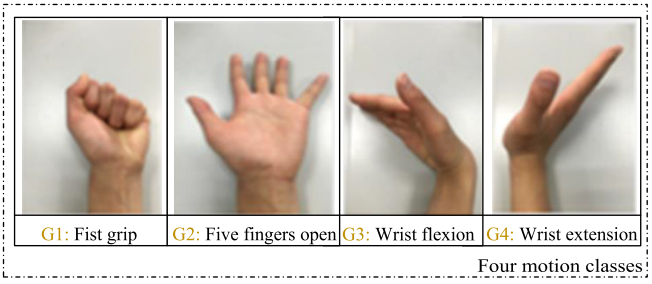


Fig. 5. Four hand movements in the EEG and sEMG acquisition experiments.

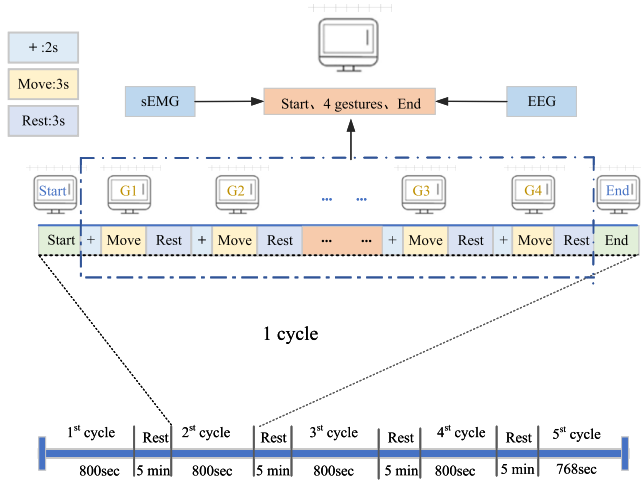


Fig. 6. Experimental scheme for synchronous acquisition of the EEG and sEMG signals.

5) The Design of Classification Block: We flatten the feature maps of three blocks into one-dimensional feature vectors and then perform feature stitching on them.

$$\{Y_T, Y_F, Y_{TF}\} = flatten\{X_T, X_F, X_{TF}\} \quad (11)$$

$$Out = Cat([Y_T, Y_F, Y_{TF}]) \quad (12)$$

We send the concatenated feature vectors *Out* to two fully connected layers connected in series. Finally, the Softmax function converts the output into classification probability.

III. Experiments

A. Acquisition Experiments of EEG and sEMG Signals

Eight subjects were recruited to collect their EEG and sEMG signals. They were all healthy students with an average age of 23.5 years old. They are right-handed. Before the experiment, all the subjects signed a written informed consent form. This experiment was performed in a quiet room. The subject sat on a chair and waited for the start of the experiment. They perform the four common hand movements, as shown in Fig. 5. The specific process of the experimental paradigm in Fig. 6 is as follows:

- 0-2s: When the computer is heard, a "+" picture appears in the computer screen, reminding the subject to prepare to perform the corresponding actions.
- 2-5s: The subject can see the corresponding action picture in the middle of the computer screen, and then the screen is in a black screen state. The subjects perform corresponding unilateral upper limb hand movements and EEG motor imagination. The entire action is completed within 3s.
- 5-8s: The subjects relax and return to their position prior to the experiment, getting ready for the following experiment.

The EEG and sEMG signal preprocessing steps are as follows:

- EEG: Firstly, the EEG signal needs to be band-pass filtered and 50Hz power frequency interference removed. Then, the original collected EEG signal needs to be channel selected and baseline interference removed. Next, the EEG signal needs to be segmented to obtain the corresponding action response signal.
- sEMG: Firstly, the original sEMG signal is filtered to retain the signal information in the 10-200Hz frequency band. Then, the action segment signal is extracted, and finally, the sEMG signal is downsampled to facilitate synchronous feature extraction with the EEG signal.

In the experiments, the first four rounds of 400 trials were used as a training set, and the 96 trials of the last round were used as the test set data. Tensflow1.15.0 is used to build the AM-PCNet network. The loss function uses cross-entropy. The dropout probability is 0.25. The Adam method is used to optimize our network. The learning rate is 0.0001. The batch size is 32, and 300 epochs are trained. we used classification accuracy (ACC) [28], kappa coefficient (K) [29], F1 score (F1) [30], and Recall [28] to evaluate the proposed AM-PCNet.

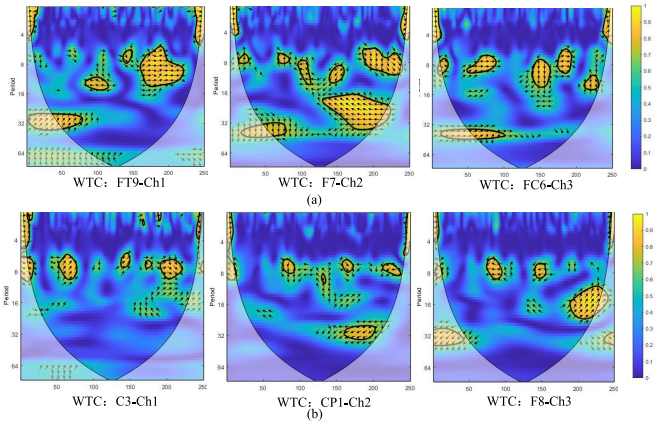


Fig. 7. Analysis of EEG and sEMG wavelet coherence (a) The EEG and sEMG strong coherence, (b) The EEG and sEMG weak coherence.

B. Comparison of EEG Channels Selection

Fig. 7 is a time-frequency map of wavelet coherence coefficients for different channel combinations of EEG and sEMG, reflecting the differences in coherence coefficients between different EEG and sEMG channels. Fig. 7(a) shows strong coherence in combination, and Fig. 7(b) shows weak coherence in combination. The comparison of strong and weak coherence provides theoretical support for improving the accuracy and stability of EEG and sEMG fusion. Eventually, we chose the FT9, FC6, and F7 channels.

TABLE I
COMPARISON OF CLASSIFICATION ACCURACY(%) BETWEEN DIFFERENT TIME-FREQUENCY TRANSFORMS AND BASELINE NETWORKS

Subject		S1	S2	S3	S4	S5	S6	S7	S8	Aver	Std
CWT	TFCNN	96.88	97.92	93.75	94.79	91.67	95.83	95.83	79.17	93.23	6.00
	TFCNN-LSTM	94.79	97.92	95.83	96.88	**94.79**	91.67	96.88	79.17	93.49	6.09
	AM-PCNet	97.92	**98.96**	96.88	**96.88**	91.67	96.88	**98.96**	88.54	95.84	3.74
Stockwell	TFCNN	94.79	96.88	93.75	90.62	88.54	91.67	94.79	88.54	92.45	**3.09**
	TFCNN-LSTM	94.79	97.92	94.79	91.67	88.54	92.71	96.88	80.21	92.19	5.68
	AM-PCNet	95.83	97.92	94.79	95.83	90.62	96.88	**98.96**	78.12	93.62	6.74
STFT	TFCNN	96.88	96.88	94.79	95.83	86.46	95.83	96.88	81.25	93.10	5.92
	TFCNN-LSTM	97.92	97.92	93.75	95.83	86.46	96.88	97.92	83.33	93.75	5.71
	AM-PCNet	**98.96**	**98.96**	**97.92**	**96.88**	90.62	**97.92**	**98.96**	**92.71**	**96.62**	3.19

TABLE II
COMPARISON OF CLASSIFICATION PERFORMANCE OF DIFFERENT COMBINATIONS OF THE EEG AND SEMG CHANNELS

Subject	C3/CP1/F8				FT9/F7/FC6			
	ACC(%)	F1	Recall	K	ACC(%)	F1	Recall	K
S1	98.96	0.990	0.989	0.986	98.96	0.990	0.989	0.986
S2	97.92	0.978	0.976	0.972	98.96	0.989	0.989	0.986
S3	94.79	0.915	0.916	0.880	97.92	0.978	0.978	0.972
S4	93.75	0.934	0.931	0.916	96.88	0.968	0.966	0.958
S5	87.50	0.871	0.872	0.833	90.62	0.905	0.906	0.875
S6	95.83	0.947	0.950	0.931	97.92	0.978	0.980	0.972
S7	97.92	0.978	0.980	0.972	98.96	0.989	0.988	0.986
S8	88.54	0.888	0.886	0.847	92.71	0.930	0.928	0.903
Aver	94.40	0.938	0.938	0.917	**96.62**	**0.966**	**0.966**	**0.955**

We fuse the EEG channels (FT9 / F7 / FC6) and EEG channels (C3 / CP1 / F8) with sEMG signals at the feature layer, and the classification results are shown in Table II. The average classification accuracy of a strong coherent combination is 96.62%, while the average classification accuracy of a weak coherent channel combination is 94.4%. After screening, the accuracy has increased by 2.22%, and the Kappa value has increased by 0.038. Screening of EEG channels suggests that each subject showed improvements in accuracy, Recall, F1, and Kappa. The result shows that the stronger the coherence between EEG and sEMG channels, the better the fusion classification performance. The conclusion presented in Table II also proves that reasonable screening of EEG channels can effectively enhance the coupling correlation between EEG and sEMG signals and improve the accuracy and stability of hand motion decoding.

C. Comparison of Feature Extraction for Different EEG and sEMG Signals

Table I summarizes the accuracy obtained by eight subjects in different time-frequency methods and classification methods. The three network structures used are:

- AM-PCNet: Multi-scale parallel convolutional network based on attention mechanism.
- TFCNN: TF-Conv Block branch module in AM-PCNet.
- TFCNN-LSTM: TFCNN connects LSTM network.

According to the comparative results, the best combination, which employs STFT to extract time-frequency features and the AM-PCNet network as the classifier, achieves an average accuracy of 96.62%. The Table I indicates that the performance of STFT is slightly better than that of Stockwell and CWT, boosting accuracy by 3% and 0.78%, respectively, in the AM-PCNet network. The result shows that by selecting the window width, STFT adapts to signal frequency content changes and improves the extraction feature's quality. In Table I, subject 5 performs the best with the combination of CWT and TFCNN-LSTM, with an accuracy rate of 94.79%. The other subjects obtain the best classification accuracy with the combination of STFT and AM-PCNet. The subject standard deviation of the STFT+AM-PCNet combination in Table I is 3.19. The results indicate that the STFT+AM-PCNet method we selected is the most robust.

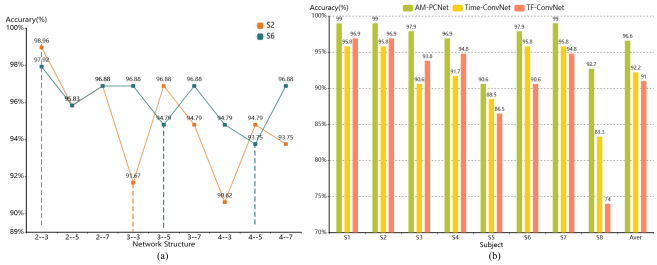


Fig. 8. (a) Accuracy comparison of different network structure in AM-PCNet. The number before "-" represents the unit numbers and the number after "-" indicates the convolution kernel sizes, (b) Accuracy comparison of AM-PCNet, Time-ConvNet, and TF-ConvNet.

D. Network Parameter and Structure Comparison Experiment

Section II-D mentions that the size of convolutional kernels and the number of units connected in series significantly impact the performance of network structure. We consider selecting two values to explore the optimal network structure [31]. Temporal Conv Unit convolution size is $1 \times K$. TF Conv Unit Convolutional Size is $K \times K$. K takes three values, i.e., 3, 5, and 7. We set the number of units in series to M, and select M from 2, 3, and 4. We have considered a total of 9 network structures. We randomly choose two subjects for testing.

In Fig. 8(a), it can be observed that when M is 2 and K is 3 (with a horizontal coordinate of "2-3"), the accuracy is the highest, reaching 98.96% and 97.92%, respectively. The accuracy rate gradually decreases with the increase in the number of series units, and overfitting occurs. Therefore,

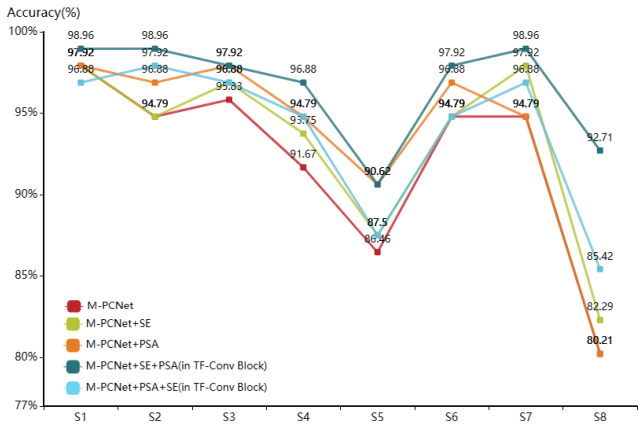


Fig. 9. Accuracy comparison of different attention mechanisms.

we choose the "2-3" parameter combination in the subsequent experiments.

To evaluate the classification effectiveness of each network branch in the model, we remove the other two network branches from the proposed model to compare decoding accuracy. The input images are EEG and sEMG Spectrogram images (time–frequency domain). A comparison classification method uses the Time-Conv branch network, where the convolutional kernel is only convoluted through the time of the input images. The other uses the TF-Conv branch network to simultaneously perform convolution operations on both time and frequency. The comparative results are shown in Fig. 8(b). Our proposed AM-PCNet network model exhibits the best performance, with an average accuracy of 96.6%. When using the Time-Conv branch network for classification, the average accuracy is 92.2%, while using the TF-Conv branch network for classification has an average accuracy of 91%. Compared with Time-Conv and TF-Conv independent networks, the AM-PCNet designed in this paper combines the advantages of each branch network and incorporates attention-based branch module interaction. The results verify the effectiveness of the AM-PCNet network and the importance of module interaction.

E. Comparison of Attention Mechanisms

To evaluate the effectiveness of introducing attention mechanisms in the model, we compare decoding accuracy by removing one or two attention mechanisms from the proposed model. We conducted the following ablation experiments with M-PCNet as the baseline model for four experiments. The EEG and sEMG feature fusion time-frequency images are used as input for the network. Experiment 1: M-PCNet (Multi-scale Parallel Convolution). Experiment 2: M-PCNet+ Squeeze Exception. Experiment 3: M-PCNet+ Pyramid Split Attention. Experiment 4: M-PCNet+ Squeeze Exception + Pyramid Split Attention(in TF-Conv Block). Experiment 5: M-PCNet + Pyramid Split Attention + Squeeze Exception(in TF-Conv Block).

The results of the average accuracy are presented in Table III. Fig. 9 displays the classification accuracy results of the ablation experiments for each subject. In the results, the average gesture recognition accuracy of 8 subjects using the baseline method M-PCNet was 92.06%; Adding SE to M-PCNet can improve the average recognition accuracy by 1.17%; Introducing PSA in M-PCNet improves accuracy by 1.6%; Simultaneously introducing both SE and PSA improves accuracy by 4.56%. In Table III and Fig. 9, we can see that the classification accuracy decreased by 2.74% after swapping the positions of the SE module and PSA module. This is mainly because the SE attention mechanism is not effective in extracting effective features in the two-dimensional time-frequency domain. By using the grouping convolution principle of PSA, the most suitable convolution kernel size can be obtained, effectively extracting finer-grained multi-scale two-dimensional spatial information of EEG and sEMG fusion. The above results indicate that the introduction of attention modules enables the network to effectively extract fine-grained multi-scale spatial information from EEG and sEMG fusion, amplify valuable EEG and sEMG fusion feature channels, and establish longer distance EEG and sEMG channel dependency. For these reasons, the accuracy and stability of EEG and sEMG fusion recognition classification are improved.

TABLE III
COMPARISON OF AVERAGE PERCENTAGE CLASSIFICATION ACCURACY WITH DIFFERENT ATTENTION MECHANISMS

Methods	Average ACC(%)
Multi-Scale Parallel CNN	92.06
+SE(Squeeze Excitation)	93.23
+PSA(Pyramid Split Attention)	93.75
+SE+PSA(in TF-Conv Block)	**96.62**
+PSA+SE(in TF-Conv Block)	93.88

IV. DISCUSSIONS

A. Comparison of Different Decoding Methods

To demonstrate the advantages of our model, we conducted a comparative analysis between the AM-PCNet network recognition model and other models investigated in recent years. Here is a concise introduction to these methods:

- DeepNet [32]: A deep-level model that uses more than three convolutional layers between time dimensions.
- EEGNet [33]: A compact EEG analysis Convolutional neural network uses deep convolution and separable convolution to construct the network model.
- ShallowNet [32]: A deep learning model with two simple convolutional layers and one mean pooling layer.
- EEG-TCNet [34]: A new time convolutional network (TCN) that achieves excellent accuracy with minimal trainable parameters.
- TCNet-Fusion [35]: A network structure that utilizes time convolutional networks, separable convolutions, and deep convolutional layer fusion to improve the interactivity of classification systems.
- ATCNet [36]: A network structure that uses a multi-head self-attention method and time convolution to extract advanced temporal features.

Table IV compares the AM-PCNet model with several state-of-the-art models on the laboratory dataset. Compared with the ShallowNet method, the AM-PCNet method has improved classification accuracy by 6.38% on average. Table IV shows that our proposed method outperforms other models among

TABLE IV
COMPARISON OF THE AVERAGE PERCENTAGE CLASSIFICATION ACCURACY(%) OF DIFFERENT MODELS

Subject	DeepNet 2017	ShallowNet 2017	EEGNet 2018	EEG-TCNet 2020	TCNetFusion 2021	ATCNet 2022	AM-PCNet
S1	95.83	91.67	95.83	94.79	94.79	93.75	**98.96**
S2	97.92	96.88	97.92	95.83	96.88	96.88	**98.96**
S3	88.54	88.54	94.79	90.62	91.67	92.71	**97.92**
S4	90.62	87.50	92.71	95.83	95.83	92.71	**96.88**
S5	89.58	87.50	**92.71**	89.58	87.50	92.71	90.62
S6	92.71	91.67	95.83	89.58	94.79	94.79	**97.92**
S7	96.88	96.88	96.88	95.83	94.79	97.92	**98.96**
S8	83.33	81.25	87.50	82.29	84.38	84.38	**92.71**
Aver	91.93	90.24	94.27	91.79	92.58	93.23	**96.62**
Std	4.91	5.22	3.29	4.77	4.43	4.09	**3.19**

TABLE V
COMPARISON OF THE AVERAGE PERCENTAGE CLASSIFICATION ACCURACY(%) OF DIFFERENT FUSION MODELS

Subject	DCA Fusion 2021	Decision-Level Fusion 2019	GFSEs 2022	AM-PCNet
S1	87.50	94.79	96.88	98.96
S2	92.71	95.83	97.92	98.96
S3	92.71	92.71	94.79	97.92
S4	88.54	92.71	95.83	96.88
S5	94.79	95.83	87.50	90.62
S6	89.58	82.29	94.79	97.92
S7	87.50	94.79	98.96	98.96
S8	91.67	90.62	86.46	92.71
Aver	90.63	92.45	94.14	**96.62**
Std	2.73	4.48	4.66	**3.19**

all subjects except for subject 5. Compared with ATCNet, our method considers both time-domain, frequency-domain, and neighborhood features, and introduces multiple attention mechanisms, improving 3.39% of accuracy. The results indicate that AM-PCNet can decode multi-task EEG and sEMG fusion signals more accurately and effectively. The standard deviation of the model between subjects reached 3.19, indicating that the model performed stably among all participants and had good decoding effects on hand movements.

B. Comparison of Different EEG and sEMG Fusion Methods

To demonstrate the advantages of our EEG and sEMG signal fusion method, we conducted a comparative analysis with other fusion methods investigated in recent years. Here is a concise introduction to these methods:

- DCA Fusion [37]: A feature layer post-fusion method, which first extracts the time-domain features of EEG signals and sEMG signals separately, and then fuses and classifies them in the later stage of the feature layer.
- Decision-Level Fusion [12]: A decision-level fusion method for EEG signals and sEMG signals, which decodes and classifies the EEG and EMG signals separately and then fuses the classification results using the decision-level fusion.
- GFSEs [11]: A method based on graph theory as a multimodal fusion strategy for EEG and sEMG. Functional connectivity, often considered as the weight of edges, enhances the robustness and accuracy of hand motion recognition.

We compared the proposed EEG signals and sEMG signals fusion decoding method with the three aforementioned fusion methods, and the results are shown in Table V.

In Table V, we can see that using the dataset collected in the laboratory and comparing it with the other three EEG signals and sEMG signals fusion methods, our proposed fusion classification result is the highest. Compared with DCA Fusion in the later stage of feature layer fusion, our accuracy has improved by 5.99%. Compared with the Decision-Level Fusion method, our EEG and sEMG signal feature layer spectrum concatenation method improved the accuracy by 4.17%. Compared with GFSEs using functional connectivity node features between EEG and sEMG signals, our method focuses more on the common frequency domain features of EEG and sEMG signals, with an accuracy improvement of 2.48%. In summary, our proposed ERP-WTC EEG channel screening method and early feature layer fusion using common time-domain and frequency-domain features of EEG and sEMG signals achieved maximum expression of common features of signals. The final results also indicate that our proposed method has stronger capabilities in feature fusion and extraction.

C. Single-Mode Signal Classification and Decoding

We compare the decoding performance of single-mode EEG signals with sEMG signals for action classification. Table VII shows the classification performance of single-mode sEMG signals for each subject. Table VIII displays the classification performance of single-mode EEG signals for each subject in different combinations of EEG signals. Table VIII indicates that the classification accuracy of the strong coherent combination of EEG is 5.34% higher than that of the weak coherent combination, which once again verifies the effectiveness of the EEG channel screening method.

In Table VII and Table VIII, the action classification methods utilizing sEMG signals demonstrate superior accuracy compared to those relying on EEG signals. The result indicates that sEMG signals contain more information related to unilateral upper limb movements, while EEG signals poorly classify fine hand movements.

In Table VII, Table VIII, and Table IV, the classification results of single-mode signals were compared with those of the

TABLE VI
COMPARISON OF CLASSIFICATION PERFORMANCE OF THE EEG AND sEMG FUSION UNDER MUSCLE FATIGUE

Subject	Fatigue level 0%				Fatigue level 20%				Fatigue level 50%				Fatigue level 90%			
	ACC(%)	F1	Recall	K	ACC(%)	F1	Recall	K	ACC(%)	F1	Recall	K	ACC(%)	F1	Recall	K
S1	98.96	0.990	0.990	0.986	97.92	0.979	0.980	0.972	96.88	0.969	0.907	0.958	87.50	0.875	0.875	0.833
S2	98.96	0.989	0.989	0.986	96.88	0.967	0.968	0.958	95.83	0.945	0.944	0.930	91.67	0.902	0.904	0.875
S3	97.92	0.978	0.978	0.972	95.83	0.955	0.954	0.944	93.75	0.935	0.934	0.917	78.12	0.787	0.788	0.708
S4	96.88	0.968	0.966	0.958	91.67	0.914	0.912	0.889	92.71	0.924	0.924	0.903	85.42	0.844	0.845	0.805
S5	90.62	0.905	0.906	0.875	86.46	0.862	0.864	0.820	85.42	0.850	0.852	0.806	83.33	0.817	0.819	0.764
S6	97.92	0.978	0.98	0.972	95.83	0.956	0.956	0.944	95.83	0.957	0.958	0.944	89.58	0.872	0.873	0.846
S7	98.96	0.989	0.988	0.986	97.92	0.978	0.98	0.972	96.88	0.968	0.968	0.958	91.67	0.911	0.911	0.889
S8	92.71	0.930	0.928	0.903	90.62	0.907	0.904	0.875	85.42	0.847	0.844	0.791	75.00	0.756	0.751	0.665
Aver	96.62	0.966	0.966	0.955	94.14	0.940	0.940	0.922	92.84	0.924	0.924	0.901	85.29	0.846	0.846	0.798

TABLE VII
CLASSIFICATION PERFORMANCE OF DIFFERENT SUBJECTS UNDER sEMG SIGNAL MODALITY

Subject	sEMG: Ch1/ Ch2/ Ch3			
	ACC(%)	F1	Recall	K
S1	98.96	0.990	0.990	0.986
S2	97.92	0.978	0.976	0.972
S3	95.83	0.958	0.960	0.944
S4	96.88	0.968	0.966	0.958
S5	87.50	0.866	0.880	0.834
S6	96.88	0.968	0.970	0.958
S7	98.96	0.989	0.988	0.986
S8	91.67	0.917	0.914	0.889
Aver	95.58	0.954	0.956	0.941

TABLE VIII
CLASSIFICATION PERFORMANCE OF DIFFERENT EEG CHANNEL COMBINATIONS UNDER EEG SIGNAL MODALITY

Subject	C3 /CP1 /F8				FT9/ F7/ FC6			
	ACC(%)	F1	Recall	K	ACC(%)	F1	Recall	K
S1	36.46	0.361	0.365	0.154	39.58	0.383	0.397	0.197
S2	35.54	0.342	0.351	0.136	37.50	0.381	0.375	0.171
S3	34.38	0.332	0.336	0.124	44.79	0.424	0.436	0.261
S4	30.21	0.304	0.307	0.080	36.46	0.364	0.367	0.156
S5	30.21	0.300	0.303	0.068	34.38	0.320	0.328	0.106
S6	37.50	0.373	0.373	0.164	42.71	0.417	0.417	0.222
S7	31.25	0.286	0.304	0.076	37.50	0.370	0.371	0.164
S8	28.12	0.191	0.291	0.054	35.54	0.357	0.353	0.139
Aver	32.94	0.311	0.329	0.107	38.28	0.377	0.381	0.177

fusion of EEG and sEMG signals. The average classification accuracy of electroencephalogram fusion was 96.62%, and the average classification accuracy of single-mode sEMG signals and EEG signals was 95.58% and 38.28%, respectively. After the fusion of EEG and sEMG signal features, the synergy and complementarity between signals were reflected, and the recognition rate was higher than that of single sEMG features and EEG features, with strong generalization. The results indicate that the accuracy, stability, and model robustness of action intention recognition can be improved through the complementary fusion of multiple features.

D. EEG and sEMG Fusion Decoding Under Muscle Fatigue

When sEMG signals are used alone, muscle fatigue issues may occur due to differences in subjects and application environments. Due to muscle fatigue, signal quality deteriorates, and classification accuracy decreases. We conduct multimodal fusion analysis using EEG and sEMG signals to alleviate the adverse effects of muscle fatigue.

In this study, we use simulation methods to represent the degree of muscle fatigue. We study fatigue sEMG signals of varying degrees from 0% to 90%. As the fatigue level increases, the sEMG amplitude decreases, the variance of the amplitude increases, and its root mean square increases [38]. We simulated fatigue signals by reducing the amplitude of surface sEMG signals and adding Gaussian noise, where reducing the amplitude simulated the exhaustion of the subjects.

As the degree of muscle fatigue increases, the changes in various indicators after EEG and sEMG fusion are shown in Table VI. In Table VI, the classification accuracy decreases as the degree of muscle fatigue increases. with 0% of muscle fatigue, the accuracy of EEG and sEMG signals fusion stands at 96.62%. When the degree of muscle fatigue reaches 20%, the classification accuracy after fusion is 94.14%, and the K-score is 0.922, indicating good network recognition performance. When the simulated fatigue sEMG signal amplitude is 50%, the recognition accuracy is 92.84%. Still, when the degree of muscle fatigue reaches 90%, the recognition rate of EEG and sEMG signals fusion significantly decreases, with accuracy and K-value being 85.29% and 0.798, respectively. The results indicate that the approach is robust even in weak sEMG signals, reducing the impact of partial loss of motor function and exercise fatigue in subjects. The fusion classification network we have constructed effectively utilizes the synergistic complementarity between EEG and sEMG signals to improve system stability, and recognition accuracy.

V. CONCLUSION

We propose a new attention mechanism-based multi-scale parallel convolutional network (AM-PCNet) for identifying and decoding EEG and sEMG fusion signals. We save the selected EEG and EMG signals in the form of time-frequency map concatenation and use AM-PCNet to extract their time-domain, frequency-domain, and mixed-domain information. We introduce attention mechanisms to effectively extract finer-grained multi-scale EEG feature information, improve the expression ability of the network model, and thus improve the accuracy and stability of signal fusion recognition.

We conduct experiments on the dataset collected in the laboratory to evaluate the effectiveness and generalization of the proposed method. Our proposed AM-PCNet network outperforms other state-of-the-art methods in accuracy, Kappa value,

Recall, and F1 value. The experiment and discussion also illustrate that our proposed method effectively improves the accuracy of recognition classification compared to single-mode decoding approaches, ensures decoding stability during muscle fatigue, and improves the accuracy and stability of the interactive recognition system.

REFERENCES

[1] K. Nagamune, M. Kubota, H. Matsuo, and Y. Mifuku, "A development of wiping rehabilitation system using leap motion for patients with upper limb paralysis," in *Proc. World Autom. Congr. (WAC)*, Oct. 2022, pp. 7–11.

[2] D.-Y. Lee, M. Lee, and S.-W. Lee, "Decoding imagined speech based on deep metric learning for intuitive BCI communication," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1363–1374, 2021.

[3] C. Guger, H. Ramoser, and G. Pfurtscheller, "Real-time EEG analysis with subject-specific spatial patterns for a brain-computer interface (BCI)," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 447–456, Dec. 2000.

[4] H. Wang, X. Dong, Z. Chen, and B. E. Shi, "Hybrid gaze/EEG brain computer interface for robot arm control on a pick and place task," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 1476–1479.

[5] R. Zhang et al., "Control of a wheelchair in an indoor environment based on a brain-computer interface and automated navigation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 1, pp. 128–139, Jan. 2016.

[6] K. K. Ang and C. Guan, "EEG-based strategies to detect motor imagery for control and rehabilitation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 4, pp. 392–401, Apr. 2017.

[7] W. Wei, Q. Dai, Y. Wong, Y. Hu, M. Kankanhalli, and W. Geng, "Surface-electromyography-based gesture recognition by multi-view deep learning," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 10, pp. 2964–2973, Oct. 2019.

[8] S. Wang et al., "Improved multi-stream convolutional block attention module for sEMG-based gesture recognition," *Frontiers Bioeng. Biotechnol.*, vol. 10, Jun. 2022, Art. no. 909023.

[9] Y. Fan and Y. Yin, "Active and progressive exoskeleton rehabilitation using multisource information fusion from EMG and force-position EPP," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 12, pp. 3314–3321, Dec. 2013.

[10] S. Kim, D. Y. Shin, T. Kim, S. Lee, J. K. Hyun, and S.-M. Park, "Enhanced recognition of amputated wrist and hand movements by deep learning method using multimodal fusion of electromyography and electroencephalography," *Sensors*, vol. 22, no. 2, p. 680, Jan. 2022.

[11] S. Yang, M. Li, and J. Wang, "Fusing sEMG and EEG to increase the robustness of hand motion recognition using functional connectivity and GCN," *IEEE Sensors J.*, vol. 22, no. 24, pp. 24309–24319, Dec. 2022.

[12] J. Tryon, E. Friedman, and A. L. Trejos, "Performance evaluation of EEG/EMG fusion methods for motion classification," in *Proc. IEEE 16th Int. Conf. Rehabil. Robot. (ICORR)*, Jun. 2019, pp. 971–976.

[13] N. Ji, L. Ma, H. Dong, and X. Zhang, "EEG signals feature extraction based on DWT and EMD combined with approximate entropy," *Brain Sci.*, vol. 9, no. 8, p. 201, Aug. 2019.

[14] Y. Li, L. Guo, Y. Liu, J. Liu, and F. Meng, "A temporal–spectral-based squeeze-and- excitation feature fusion network for motor imagery EEG decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1534–1545, 2021.

[15] K. Zhu, X. Zhang, H. Liu, Y. Xiong, Y. Zhang, and C. He, "An approach for sEMG-based gesture recognition using continuous wavelet transform and AlexNet convolutional neural network," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2021, pp. 762–767.

[16] X. Li, O. W. Samuel, X. Zhang, H. Wang, P. Fang, and G. Li, "A motion-classification strategy based on sEMG-EEG signal combination for upper-limb amputees," *J. NeuroEng. Rehabil.*, vol. 14, no. 1, pp. 1–13, Dec. 2017.

[17] A. Chowdhury, H. Raza, Y. K. Meena, A. Dutta, and G. Prasad, "An EEG-EMG correlation-based brain-computer interface for hand orthosis supported neuro-rehabilitation," *J. Neurosci. Methods*, vol. 312, pp. 1–11, Jan. 2019.

[18] K. Shi et al., "Multimodal human-exoskeleton interface for lower limb movement prediction through a dense co-attention symmetric mechanism," *Frontiers Neurosci.*, vol. 16, Apr. 2022, Art. no. 796290.

[19] X. Xi et al., "Enhanced EEG–EMG coherence analysis based on hand movements," *Biomed. Signal Process. Control*, vol. 56, Feb. 2020, Art. no. 101727.

[20] P. Autthasan et al., "MIN2Net: End-to-end multi-task learning for subject-independent motor imagery EEG classification," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 6, pp. 2105–2118, Jun. 2022.

[21] G. L. Cerone, A. Giangrande, M. Ghislieri, M. Gazzoni, H. Piitulainen, and A. Botter, "Design and validation of a wireless body sensor network for integrated EEG and HD-sEMG acquisitions," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 61–71, 2022.

[22] U. Cote-Allard et al., "Deep learning for electromyographic hand gesture signal classification using transfer learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 4, pp. 760–771, Apr. 2019.

[23] R. G. Stockwell, L. Mansinha, and R. P. Lowe, "Localization of the complex spectrum: The S transform," *IEEE Trans. Signal Process.*, vol. 44, no. 4, pp. 998–1001, Apr. 1996.

[24] Y. Han, B. Wang, J. Luo, L. Li, and X. Li, "A classification method for EEG motor imagery signals based on parallel convolutional neural network," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103190.

[25] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[26] H. Zhang, K. Zu, J. Lu, Y. Zou, and D. Meng, "EPSANet: An efficient pyramid squeeze attention block on convolutional neural network," in *Proc. Asian Conf. Comput. Vis.*, 2021, pp. 1161–1177.

[27] E. Su, S. Cai, L. Xie, H. Li, and T. Schultz, "STAnet: A spatiotemporal attention network for decoding auditory spatial attention from EEG," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 7, pp. 2233–2242, Jul. 2022.

[28] Y. Li, J. Liu, Z. Tang, and B. Lei, "Deep spatial–temporal feature fusion from adaptive dynamic functional connectivity for MCI identification," *IEEE Trans. Med. Imag.*, vol. 39, no. 9, pp. 2818–2830, Sep. 2020.

[29] J. Chen, Z. Yu, Z. Gu, and Y. Li, "Deep temporal-spatial feature learning for motor imagery-based brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 11, pp. 2356–2366, Nov. 2020.

[30] S. Ren, W. Wang, Z.-G. Hou, X. Liang, J. Wang, and W. Shi, "Enhanced motor imagery based brain-computer interface via FES and VR for lower limbs," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 8, pp. 1846–1855, Aug. 2020.

[31] X. Tang, C. Yang, X. Sun, M. Zou, and H. Wang, "Motor imagery EEG decoding based on multi-scale hybrid networks and feature enhancement," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1208–1218, 2023.

[32] R. T. Schirrmeister et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, Nov. 2017.

[33] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056013.

[34] T. M. Ingolfsson, M. Hersche, X. Wang, N. Kobayashi, L. Cavigelli, and L. Benini, "EEG-TCNet: An accurate temporal convolutional network for embedded motor-imagery brain–machine interfaces," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2020, pp. 2958–2965.

[35] Y. K. Musallam et al., "Electroencephalography-based motor imagery classification using temporal convolutional network fusion," *Biomed. Signal Process. Control*, vol. 69, Aug. 2021, Art. no. 102826.

[36] H. Altaheri, G. Muhammad, and M. Alsulaiman, "Physics-informed attention temporal convolutional network for EEG-based motor imagery classification," *IEEE Trans. Ind. Informat.*, vol. 19, no. 2, pp. 2249–2258, Feb. 2023.

[37] M. S. Al-Quraishi, I. Elamvazuthi, T. B. Tang, M. Al-Qurishi, S. Parasuraman, and A. Borboni, "Multimodal fusion approach based on EEG and EMG signals for lower limb movement recognition," *IEEE Sensors J.*, vol. 21, no. 24, pp. 27640–27650, Dec. 2021.

[38] V. de A. Rocha, J. C. do Carmo, and F. D. O. Nascimento, "Weighted-cumulated S-EMG muscle fatigue estimator," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 6, pp. 1854–1862, Nov. 2018.