Wearable Supernumerary Robotic Limb System Using a Hybrid Control Approach Based on Motor Imagery and Object Detection

Zhichuan Tang, Lingtao Zhang, Xin Chen, Jichen Ying, Xinyang Wang, and Hang Wang

Abstract—**Motor disorder of upper limbs has seriously affected the daily life of the patients with hemiplegia after stroke. We developed a wearable supernumerary robotic limb (SRL) system using a hybrid control approach based on motor imagery (MI) and object detection for upper-limb motion assistance. SRL system included an SRL hardware subsystem and a hybrid control software subsystem. The system obtained the patient's motion intention through MI electroencephalogram (EEG) recognition method based on graph convolutional network (GCN) and gated recurrent unit network (GRU) to control the left and right movements of SRL, and the object detection technology was used together for a quick grasp of target objects to compensate for the disadvantages when using MI EEG alone like fewer control instructions and lower control efficiency. Offline training experiment was designed to obtain subjects' MI recognition models and evaluate the feasibility of the MI EEG recognition method; online control experiment was designed to verify the effectiveness of our wearable SRL system. The results showed that the proposed MI EEG recognition method (GCN+GRU) could effectively improve the MI classification accuracy (90.04% ± 2.36%) compared with traditional methods; all subjects were able to complete the target object grasping tasks within 23 seconds by controlling the SRL, and the highest average grasping success rate achieved 90.67% in bag grasping task. The SRL system can effectively assist people with upper-limb motor disorder to perform upper-limb tasks in daily life by natural human-robot interaction, and improve their ability of self-help and enhance their confidence of life.**

Zhichuan Tang is with the Industrial Design Institute, Zhejiang University of Technology, Hangzhou 310023, China, and also with the Faculty of Science and Technology, Bournemouth University, Poole BH12 5BB, U.K. (e-mail: ttzzcc@zju.edu.cn).

Lingtao Zhang, Jichen Ying, Xinyang Wang, and Hang Wang are with the Industrial Design Institute, Zhejiang University of Technology, Hangzhou 310023, China.

Xin Chen is with the College of Information Engineering, Zhejiang University of Technology, Hangzhou, 310023 China.

I. Introduction

WITH the growing number of aging population in the worldwide, stroke incidence has continuously and sharply increased in recent years. More than 80% of stroke patients with hemiplegia suffered upper-limb motor disorder in different degree due to the nervous system injury [1]. Upper-limb motor disorder severely affects the daily life of patients, resulting in serious financial and life burden to families and great mental pressure to patients. How to help these people improve their ability of self-help and living independently is especially important for social development [2].

Supernumerary Robotic Limb (SRL) is a wearable robotic device to enhance wearer's capability in different tasks. Parietti and Asada [3] designed a wearable supernumerary robotic upper limb that could provide extra support for workers; Ciullo *et al.* [2] developed a supernumerary robotic hand that could assist patients with severe stroke to perform a variety of hand motions. Unlike the traditional exoskeletons and prostheses, SRL can attach to different parts of a patient's body to assist him to execute one-hand tasks or to augment the functionality of the remaining healthy hand to perform the two-hand tasks [1]. SRL can supplement and enhance wearer's capability additional robotic structures, which has clear advantages over exoskeletons and prostheses [4]. While the physical hardware is developed, how to control SRL naturally and accurately remains an important challenge. There are two strategies to control SRLs: direct control and indirect control [5]. Direct control is a way that users input control commands directly to SRLs by buttons, visual interfaces or physiological signals. Some previous studies used different parts of body, not just limbs, to control SRLs. Guggenh *et al.* [6] used different levels of force on the fingertips to control different grasp motions of an SRL in order not to affect operator's normal actions in the work; Salvietti *et al.* [7] obtained the wearer's motion intentions by analyzing surface electromyography (sEMG) signals of muscles around the eyebrows to control the "six" finger. Indirect control is a way that users do not need to input specific control commands but the SRLs can move automatically through analyzing users' body movements. Peternel *et al.* [8] constructed a human motion recognition model based on sEMG signals of muscles on limbs, and this model could predict body motion trajectories automatically to control an SRL

indirectly to help users complete target tasks. Some other studies have focused on using hybrid control methods (direct control + indirect control) to control SRLs. Muhammad *et al.* [9] applied a hybrid control strategy based on sEMG signals and inertial sensors to control a supernumerary robotic arm; Song and Asada [10] preset the automatic motion trajectory of SRLs based on the users' posture data when using a knife and fork, and controlled an SRL by different motions of left foot; Cio *et al.* [11] proposed a hybrid control system based on stereo vision and eye-tracking technologies to control an SRL in real time. The above hybrid control strategies could effectively transfer multiple control information to let users to interact with SRLs. However, most of these control methods used the parts of body which were not the limbs to be assisted to control SRLs, e.g., use feet to control an SRL for upper-limb assistance, resulting in spending a lot of pre-training time on adapting to this control manner. Besides, the unnatural interaction experience between user and SRL led to greater cognitive efforts and higher learning costs for users.

Brain-computer interface (BCI) provides communication and control capabilities to people with severe motor disabilities. BCI systems support communication through direct measures of electroencephalogram (EEG), functional magnetic resonance imaging (FMRI) and electrocorticogram (ECOG) to operate and control devices like cursor, wheelchair, and exoskeleton [12]. Patients with upper-limb motor disorder are unable to move their limbs, but their brains are normal. Therefore, patient's EEG signals can be decoded and converted into control signals to control SRLs in two-dimensional or three-dimensional space [13]. Motor imagery (MI) EEG classification is a widely applied paradigm in BCI systems. Patients can imagine limb movements to control external devices [14]. Pinheiro *et al.* [15] designed an intelligent wheelchair system based on MI EEG control for patients with lower-limb motor disorder to improve their walking ability; Lee *et al.* [16] developed a control approach by classifying users' MI intentions to control an exoskeleton to move in three different directions. The control strategy based on MI EEG is more natural and can reduce cognitive efforts for patients during SRL operation.

One problem in current MI-based BCI applications is that the low signal-to-noise ratio of EEG signals lead to a low EEG decoding accuracy. To tackle this problem, researchers have focused on looking for reliable features to improve the classification performance [17], [18]. Some latest studies aimed to select optimal EEG features which do not contain redundant information and noise through feature selection methods like the common spatial pattern (CSP) based on L1-Norm and Dempster–Shafer theory [19], the correlation-based channel selection (CCS) method [20], etc. The others aimed to obtain the deep features which can describe different MI classes through deep learning methods like the deep belief network (DBN) [21], the long short-term memory (LSTM) network [22], the convolutional neural network (CNN) [23], etc. In these methods, CNN can directly be used for automatic feature extraction of the raw input signals, and can obtain deeper-level and more differentiated feature information for EEG signal recognition [24]–[26]. However, the traditional CNN approaches do not consider the topological relationship

and structure information of EEG electrodes. The distribution structure of EEG channel is irregular, and EEG signals do not belong to the standard non-Euclidean structure data [27]. Graph convolutional network (GCN) provides an effective way to describe the intrinsic relationship between different nodes of the graph, which is suitable for the topological feature extraction of EEG signals in the discrete spatial domain. Some previous studies have focused on using GCN for emotion feature extraction and recognition based on EEG signals, and they found that GCN achieved the higher recognition accuracies than CNN [28]. In addition, when human imagines the movement of unilateral limb, the power of mu and beta rhythms will decrease or increase in the sensorimotor area of the contralateral hemisphere and the ipsilateral hemisphere in different time periods [29]. Therefore, MI EEG signals also contain temporal feature information. Long short-term memory (LSTM) [30] and gated recurrent unit network (GRU) [31] can solve some problems such as gradient vanishing and gradient exploding. Both LSTM and GRU use gating mechanisms to memorize as much long-term information as possible. GRU has the fewer parameters, simpler structure, shorter model training time and faster convergence than LSTM [32]. In our study, we used GCN and GRU to extract spatial topological features and temporal features of EEG, respectively.

The other problem is that when only MI is used in the control strategy, no enough control dimensions/commands can be output to control different motions of SRLs, which affects the control accuracy and efficiency during SRL operation. For example, common MI can output three control commands by imagining left hand, right hand and feet movements, but the control of SRLs needs four control commands (up, down, left and right) at least. Some previous studies have used hybrid control approaches by combing MI and other assistant methods (electrooculography (EOG) [33], P300 [34], object detection [35], etc.) to improve the control efficiency of BCI applications. In these assistant methods, object detection can identify semantic objects of a certain class (such as humans, buildings, or cars) in digital images and videos [36]. This technology has been applied in the control of external devices like robotic arm and prosthesis. Shi *et al.* [37] used computer vision-based object detection technology and sEMG to control a multi-fingered prosthetic hand, which helped users rapidly complete grasping tasks on various daily objects; Chen *et al.* [38] developed a robotic arm control system by combing augmented reality (AR), object detection, and steady-state visual evoked potential (SSVEP) to autonomously picked up the target objects. Object detection technology can be used together with MI for a quick grasp of target objects to improve the control accuracy and efficiency of SRLs.

In this paper, we developed a wearable SRL system using a hybrid control approach based on motor imagery and object detection for upper-limb motion assistance. SRL system included an SRL hardware subsystem and a hybrid control software subsystem. The SRL hardware subsystem consisted of a self-developed robotic arm and a camera. The hybrid control software subsystem consisted of an MI recognition module, an object detection module and a hybrid control module. MI recognition module was used to recognize users'
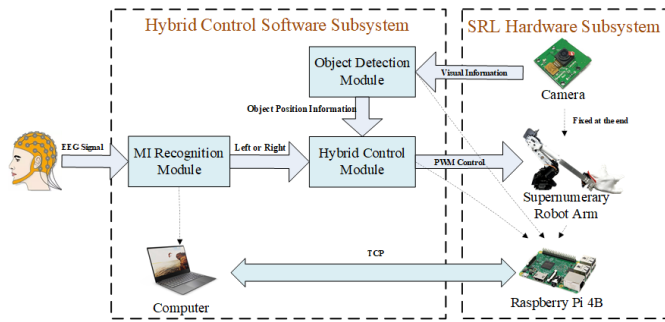
Fig. 1. The architecture diagram of the wearable SRL system.

motion intentions through an MI recognition model based on GCN that extracted spatial topological features of EEG and GRU that extracted temporal features of EEG; the object detection module was used to recognize different target objects in the target area and provide position information of this object; the hybrid control module was used to estimate the motion trajectory of SRL and control the SRL in real time. During the online control, users' MI EEG data were processed by MI recognition module to output the control commands of left/right movement, and the YOLO object detection system in the object detection module was used together for a quick grasp of target objects to compensate for the disadvantage of MI control dimension and improve the control accuracy of the SRL. An offline training experiment and an online control experiment were designed to evaluate and verify our proposed method and SRL system.

II. WEARABLE SRL SYSTEM

A. System Architecture

The wearable SRL system included an SRL hardware subsystem and a hybrid control software subsystem, as shown in Fig. 1. The SRL hardware subsystem including a robotic arm, a bionic hand and some other hardware accessories was used for upper-limb assistance; the hybrid control software subsystem including an MI recognition module, an object detection module and a hybrid control module was used to recognize users' motion intention, detect target objects precisely and output SRL's control signals. In hybrid control software subsystem, the MI recognition module recognized users' motion intention based on GCN and GRU; the object detection module acquires the visual information of target area from the camera on SRL hardware subsystem to analyze the category of target object and output accurate position information; the hybrid control module controlled the SRL to move to left or right according to the recognized MI intentions firstly, and then realized the precise grasping of the target object based on accurate position information which was output from the object detection module when the camera captured the target object (i.e., the target object appeared in the target area).

B. SRL Hardware Subsystem

We fabricated a lightweight wearable SRL hardware subsystem, as shown in Fig. 2. SRL hardware subsystem consisted of a robotic arm, a three-fingered bionic hand, a camera, a fixed
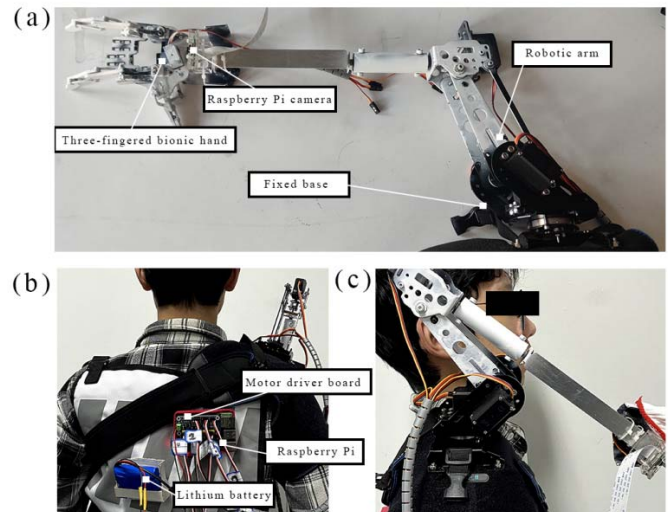

Fig. 2. SRL hardware subsystem. (a) robotic arm and bionic hand; (b) motor driver board, lithium battery and Raspberry Pi; (c) fixing method of SRL.

base, a motor driver board, a lithium battery and a Raspberry Pi microcomputer. The robotic arm with 5 degrees of freedom could simulate the normal arm motions and postures to perform complex upper-limb tasks. The five joints of the robotic arm were actuated by steering gears. A camera was fixed at the end of the robot arm to capture the target object for object detection. The three-fingered bionic hand actuated by steering gears was located at the end of the robotic arm. The fixed base was mounted on the users' right shoulder with an elastic strap to support the SRL and make it steady, as shown in Fig. 2c. All steering gears of The SRL hardware subsystem was controlled by a motor driver board and powered by a lithium battery (12V, 2400mA), as shown in Fig. 2b. The motor driver board (PCA9685, WeixueElectronics Co. Ltd, China) and the micro camera were connected to a Raspberry Pi microcomputer. The microcomputer with a 64-M storage space, a linux system and a Tensorflow Lite framework [39] could run deep learning models independently for MI recognition and object detection, and transmit data to the computer through the TCP port.

C. Hybrid Control Software Subsystem

The hybrid control software subsystem including an MI recognition module, an object detection module and a hybrid control module was used for the recognition of patient's motion intent, the precise detection of target objects and the output of hardware control signals.

1) MI Recognition Module: We constructed an MI recognition model based on GCN and GRU, including the feature extraction part and the classification part, as shown in Fig. 3. GCN and GRU were used to extract spatial topological features and temporal features of EEG, respectively. The input of the model was MI EEG signals and the output was motion intentions (left or right). EEG signals of different EEG frequency bands in one input sample were divided into several time segmentations through overlapping window method, and these time segmentations were input into the feature extraction part separately.
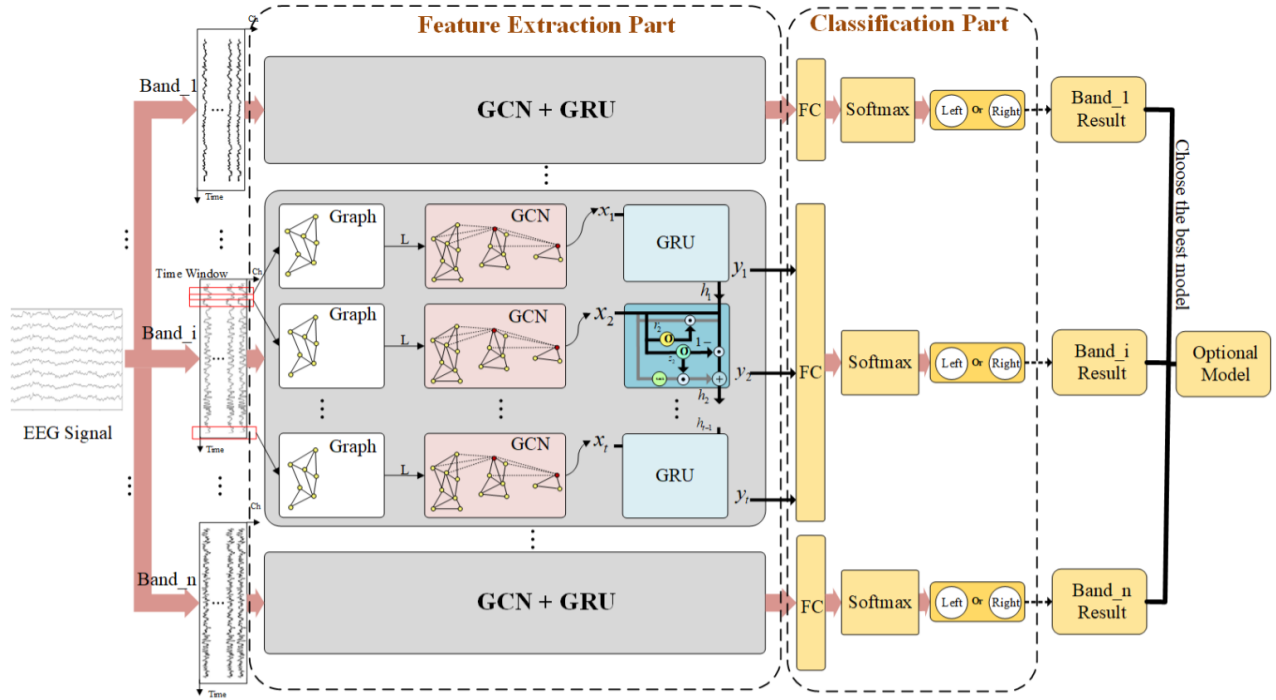
Fig. 3. Framework diagram of MI recognition model. Different EEG frequency bands of one input sample were segmented into several time windows, and input into the feature extraction part based on GCN and GRU models and the classification part based on a softmax classifier to obtain the MI motion intention.

In the feature extraction part, firstly, the EEG data of each segmentation were converted to graph data to input into the GCN model. Each EEG channel corresponded to a node of the graph data, and the connections between two different nodes corresponded to the edges of the graph. A undirected graph to describe the topological structure of the EEG data is defined as $G = \{V, E\}$, where V represents the set of nodes with the number of $|V| = N$ and E represents the set of edges connecting these nodes. Graph convolutional operation in GCN model can be expressed as

$$H^{l+1} = f\left(L^{sym} H^l W^l\right), \qquad (1)$$

where H^l and H^{l+1} are the lth graph convolutional layer and $l + 1$th graph convolutional layer, $f(\cdot)$ is ReLU activation function, W^l is weight matrix of the lth graph convolutional layer, and L^{sym} is symmetric normalized Laplacian matrix. L^{sym} can be calculated by

$$L^{sym} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = D^{-\frac{1}{2}} (D - A) D^{-\frac{1}{2}} = I_N - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}, \qquad (2)$$

where $L = D - A$ is the combinatorial Laplacian matrix, D is the degree matrix with diagonal entries, I_N is the unit matrix, and A is the adjacency matrix describing the connection relationships between any two nodes. The entry of adjacency matrix is denoted by w_{ij} to measure the importance of the connection between the i-th node and the j-th node. The distance function method [40] is used to determine the entries w_{ij} of the adjacency matrix. The Gaussian kernel function was selected as the distance function in our study, and it can be expressed as

$$w_{ij} = \begin{cases} \exp\left(-\dfrac{[dist\,(i,\,j)]^2}{2\theta^2}\right), & if\,dist\,(i,\,j) \leq \tau \\ 0, & otherwise, \end{cases} \qquad (3)$$

where τ and θ are two fixed parameters, $dist\,(i,\,j)$ represents the distance between the i-th node and the j-th node. The output graph features were extracted from the input graph data through graph convolutional operations in graph convolutional layers, and were transformed to a one-dimension vector x by linear affine transformation and flattening operation. The flattening vectors of all time segmentations formed the time series data $[x_1, x_2 \ldots x_t]$, where t is the number of time segmentations in one EEG frequency band. Secondly, the time series data were input into the GRU model to extract temporal features. There are two kinds of gate in the GRU model, i.e. update gate z_t and reset gate r_t. The update gate helps the model to determine how much of the past information needs to be passed along to the future, and the reset gate is used from the model to decide how much of the past information to forget. At each timestamp t, it takes an input x_t and the hidden state h_{t-1} from the previous timestamp $t - 1$. Later it outputs y_t and a new hidden state h_t which again passed to the next timestamp. The formula for each gate and outputs at time step t are

$$z_t = \sigma\,(W_{xz} x_t + W_{hz} h_{t-1} + b_z), \qquad (4)$$
$$r_t = \sigma\,(W_{xr} x_t + W_{hr} h_{t-1} + b_r), \qquad (5)$$
$$\tilde{h}_t = \tanh(W_{x\tilde{h}} x_t + r_t \odot W_{h\tilde{h}} h_{t-1} + b_{\tilde{h}}), \qquad (6)$$
$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t, \qquad (7)$$
$$y_t = \sigma\,(W_{hy} h_t), \qquad (8)$$

Fig. 4.   Visual information from the camera and the recognition results of object category and coordinate.

where σ denotes the sigmoid function, tanh denotes the hyperbolic tangent function, W is the weight matrix, b is the bias, $\tilde{h}$ is the candidate hidden state, and $\odot$ denotes Hadamard product.

In the classification part, the outputs of the GRU model were poured into a fully connected layer, followed by a softmax layer and a classification output layer to generate predicted class labels (left or right) and recognize the MI motion intents. The model with best classification performance in different EEG frequency bands was selected as the optimal MI recognition model for this subject.

The hyperparameters in the model training process were mainly empirically chosen. We set the learning rate was 0.001, the batch size was 1024, the dropout rate was 50% for the fully connected layer, and the maximum number of iterations was 10000. The Stochastic Gradient Descent (SGD) algorithm implemented by the Adam optimizer was used to update the model parameters [41]. Batch normalization (BN) was employed for graph convolutions. The early stopping method was used to determine the optimal parameters of model, i.e., when the accuracy of the validation set data did not increase within 20 epochs, the parameters at this epoch was saved as the optimal parameters. The loss curves during the model training process was used to evaluate the convergence of the models.

2) Object Detection Module: The object detection module was used to recognize different target objects in the target area and provide position information of this object to further control the SRL precisely.

We used YOLO object detection system [42] which was running on the Raspberry Pi with the Tensorflow Lite [39] framework to real-time detect target objects. The visual information from the camera fixed at the end of the robot arm was transferred to YOLO to output the probability of the category of the object and the coordinate values of the object's border, as shown in Fig. 4.

3) Hybrid Control Module: The hybrid control module was used to estimate the motion trajectory of SRL and control the SRL in real time based on the motion intention information from MI recognition module and the target position information from object detection module. The SRL could move following the patient's motion intention and could perform the grasping task quickly and precisely through the hybrid control. The flow chart of the hybrid control module is shown in Fig. 5.

Firstly, the SRL needed to ensure whether the target objects was in the target area or not. The robotic arm of the SRL could move to left or right based on patient's motion intention from MI recognition module. At the same time, the camera at the end of the SRL captured the visual information of the target
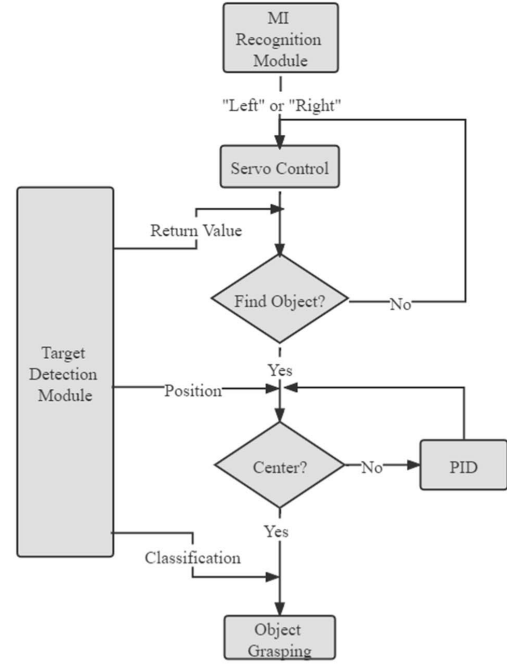

Fig. 5.   The flow chart of the hybrid control module.

area and recognized the target object by the object detection module. Then, the hybrid control module determined whether the target object was located in the target area or not according to the output information returned by the object detection module. If the target object was not in the target area, patient could control the SRL to move by MI until the target object was detected.

Secondly, when the target object was detected in the target area, the hybrid control module would adjust the SRL automatically to face to the target object, i.e., let the target object locate in the center of the camera view. We set the coordinate origin at the top left of the camera view and establish a plane coordinate system, as shown in Fig. 6a. The size of the camera view was 640px * 480px, and the coordinate values of the bounding box's four points of the target object $(X_{min}, Y_{min}; X_{max}, Y_{min}; X_{min}, Y_{max}; X_{max}, Y_{max})$ were obtained from the object detection module. The coordinate values of the center point of the camera was (320, 240), and the coordinate values of the center point of the target object (X_{center}, Y_{center}) could be calculated by

$$\begin{cases} X_{center} = X_{max} - (X_{max} - X_{min})/2 \\ Y_{center} = Y_{max} - (Y_{max} - Y_{min})/2 \end{cases} \quad (9)$$

If the target object was not in the center of the camera view, i.e., $X_{center} \neq 320$ and $Y_{center} \neq 240$, the incremental PID algorithm was used for SRL's adaptive adjustment to face to the target object. This algorithm could calculate the rotation angle values $\Delta \mu$ that different steering gears needed to adjust in each time point based on the input deviation values e (the distance between the center of the camera view and the center of the object, as shown in Fig. 6b):

$$\Delta \mu (k) = K_p (e(k) - e(k-1)) + K_i e(k) + K_d(e(k) \\ - 2e(k-1) + e(k-2)) \quad (10)$$
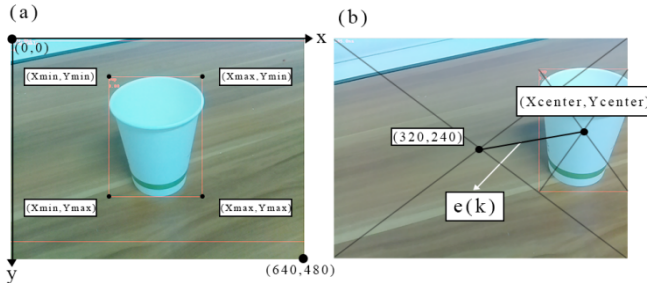
Fig. 6. Coordinate system (a) for coordinate values calculation and deviation value (b) between the center of the camera view and the center of the object.

where $\Delta\mu(k)$ is the rotation angle value at the kth control cycle, the K_p denotes the proportional coefficient, K_i denotes the integral coefficient, K_d denotes the differential coefficient, and $e(k)$ is the deviation value at the kth control cycle.

Finally, when the target object was detected in the center of the camera view, the hybrid control module automatically executed the corresponding grasp command based on the object category information from the object detection module to grasp the object; the SRL would hold the object for 5 seconds, and then the hybrid control module automatically executed the corresponding release command to release the object on the desk. This grasp-release process represents one successful grasping trial. We preset three grasp commands and the corresponding release commands for different target objects (cup, bag and bowl), and the time-sequence control commands of different steering gears in the SRL were set in each grasp command and each release command.

III. EXPERIMENTS

We designed the offline training experiment and online control experiment to verify the effectiveness of the SRL system. In the offline training experiment, we built the EEG-based MI recognition model for each subject to recognize his motion intention. In the online control experiment, subjects control the SRL to grasp the target objects based on MI and object detection, and the results of the grasping tasks were evaluated.

A. Subjects

Ten healthy graduate students (5 males and 5 females, aged 22-29 years) participated in our experiment. Each subject signed informed consent forms before the experiment. The experimental procedure was reviewed and approved by the human ethical clearance committee of Zhejiang University of Technology.

B. Offline Training Experiment

The aim of the offline training experiment was to obtain the MI recognition model for each subject. In the offline training experiment, subjects wore an EEG cap connecting to the Biosemi acquisition device (ActiveTwo, BioSemi instrumentation, the Netherlands) to collect 64-channel EEG data. The reference electrode was placed at the left ear mastoid and the ground electrode was replaced by two separate electrodes
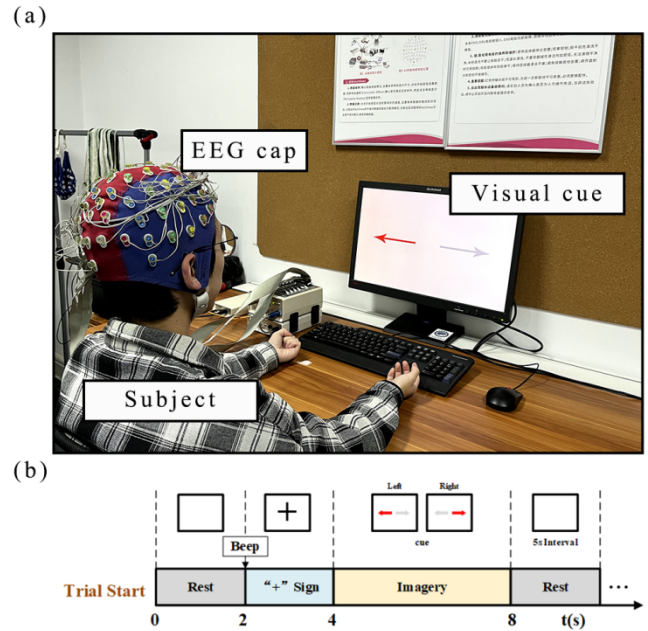

Fig. 7. Offline training experimental scenario (a) and timing diagram of one trial (b).

(CMS and DRL). Before placing the electrodes, the conductive gel was used to reduce the impedance between the electrodes and the scalp. The impedance of each electrode was kept below 5 kΩ. The sample rate was 1024 Hz.

After all electrodes were placed, subjects sat in front of the computer screen and put their hands naturally on the table, as shown in Fig. 7a. They were asked to avoid blinking their eyes and unnecessary head/body movements as possible. Each subject was required to complete 280 cue-based trials in the offline experiment, including 140 trials of left hand MI and 140 trials of right hand MI. The timing diagram of one trial is shown in Fig. 7b. Each trial lasted for 8 seconds, and started with an acoustical warning tone at second 2 and a 2-s "+" sign presentation. Then a cue ("←" or "→") randomly appeared on the screen (second 4-8) to indicate the left or right hand movement that the subject needed to imagine, corresponding to the left or right movement of the SRL. To avoid fatigue, there was a 5-s interval between each trial and a 2-min rest period between every 35 trials.

The 4-s MI EEG (second 4-8) of each trial was selected as one input sample of MI recognition model. Due to the individual differences of EEG, event-related desynchronization (ERD) and event-related synchronization (ERS) patterns which are used as important features in the discrimination between right/left MI presented in different EEG frequency bands for different subjects [43], [44]. To select the optimal frequency bands to build the MI recognition model, we divided the 8-30 Hz MI EEG signal into 11 frequency bands (8-10 Hz, 10-12 Hz, 12-14 Hz, 14-16 Hz, 16-18 Hz, 18-20 Hz, 20-22 Hz, 22-24 Hz, 24-26 Hz, 26-28 Hz, 28-30 Hz). Then, MI EEG of each frequency band was divided into 7 time segmentations through 1-s time window with 0.5-s overlapping, and the time segmentations were converted to the graph data for MI recognition model training.
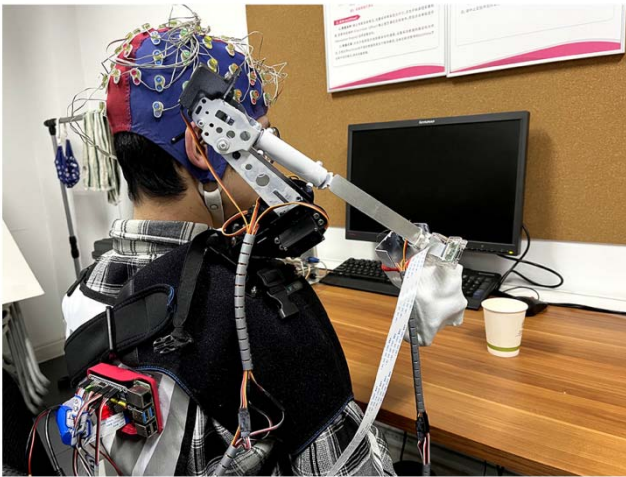
Fig. 8. Online control experimental scenario.

Fig. 9. The three grasping tasks in the online control experiment.

In the MI recognition model training process, each subject's data were used to train his own model. All data were divided into five parts randomly: three parts were used as the training set (60% of the data) for model construction, one part was used as the validation set (20% of the data) for optimal parameter selection, and one part was as the testing set (20% of the data) for model evaluation. Besides, to compare with the other methods, we selected three methods, i.e., CNN [23], LSTM [22] and CSP+SVM [45], to train MI recognition models respectively on the same training set, and then these models were tested using the same testing set.

C. Online Control Experiment

The aim of the online control experiment was to evaluate the control performance of the SRL system. The experimental scenario is shown in Fig. 8. An experimenter fixed the SRL hardware subsystem on the subject's right shoulder using an elastic strap, and adjusted the robotic arm in the sagittal plane and kept the camera and the robotic arm in a straight line. The subject was seated in a chair naturally and wore the 64-channel EEG cap, and different target objects (cup, bag or bowl) were placed on the table in front of the subject.

The task of the online control experiment was to control the SRL system to grasp the target object. Subjects were asked to perform MI according to the position of the target object to control the SRL system to move to the target area. When the target object appeared in the camera view, the SRL system automatically adjusted the robotic arm to let the target object locate in the center of the camera view, and then performed the grasping tasks for different target objects based on the corresponding preset grasp commands. Three target objects were put in different positions on the table, i.e., the cup placed in front on subject's right side, the handbag placed in front on subject's left side, and the bowl placed in front of the subject. The distance between the coronal plane of the three target objects and the coronal plane of the camera was same (20cm). Subjects were asked to control the SRL to grasp the target object through MI after hearing the "beep". The three grasping tasks are shown in Fig. 9. Each subject was required
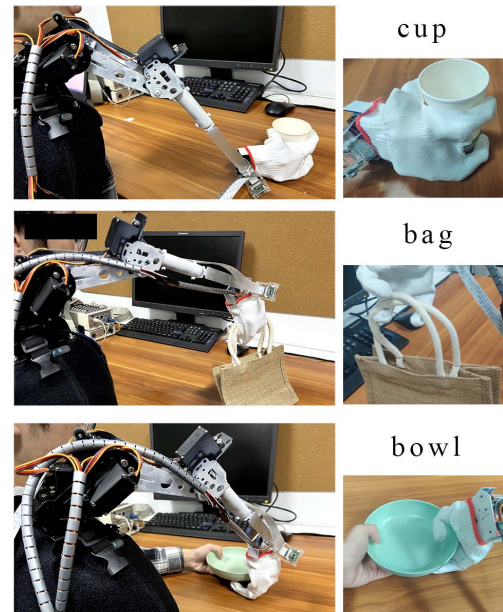
to perform 15 grasping trials for each target object, resulting in a total of 45 grasping trials.

IV. RESULTS

A. Public Dataset Results

To verify the effectiveness of the MI recognition model based on GCN and GRU, we firstly applied it to the public dataset and compared it with the other three methods (CNN, LSTM, and CSP+SVM). The public dataset we used is dataset IIIa from BCI Competition III [46]. In this dataset, three subjects labeled k3, k6 and l1were asked to imagine the left hand, right hand, and tongue/foot movements according to different visual cues. The EEG signals were recorded from 64 channels with a sample rate of 250 Hz. The number of trials per class is 60 for each subject. We only selected the trials of left and right hand MI tasks in this study. More details of experimental paradigm see the following website: http://www.bbci.de/competition/iii/. The classification accuracies of four methods on the dataset IIIa for each subject are shown in Table I. For all three subjects, the average classification accuracy of GCN+GRU (91.20%±4.68%) is higher than those of CNN (88.06%±4.28%), LSTM (83.44%±3.17%), and CSP+SVM (81.39%±3.76%). The real subjects had to be participated in the online control experiment to control the SRL in real time, since our proposed MI recognition method was further applied to the subjects' EEG data collected in the offline training experiment to train their own models.

B. Results of Offline Training Experiment

1) ERD/ERS Analysis: The time course of ERD/ERS and EEG topographies are shown in Fig. 10. For left and right hand MI, the EEG power of C3, Cz and C4 at 8-12 Hz frequency band was averaged across all trials and all subjects, and was displayed as the relative percentage to the EEG power during the reference period. During MI (second 4-8),
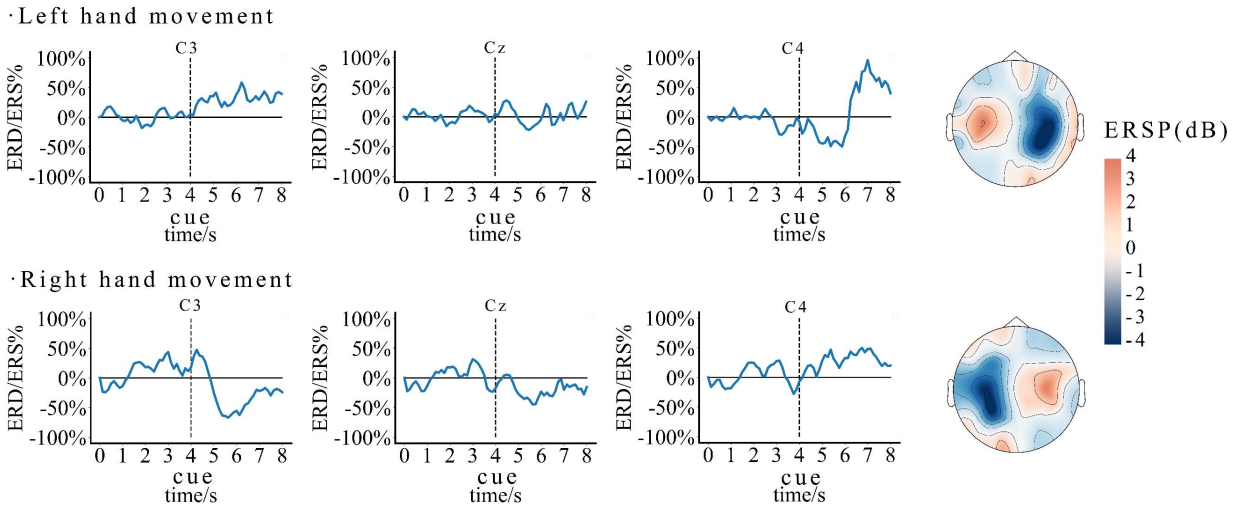
Fig. 10. The time course of ERD/ERS from second 0 to 8 and EEG topographies from second 5 to 7 for left and right hand MI across all trials and all subjects.

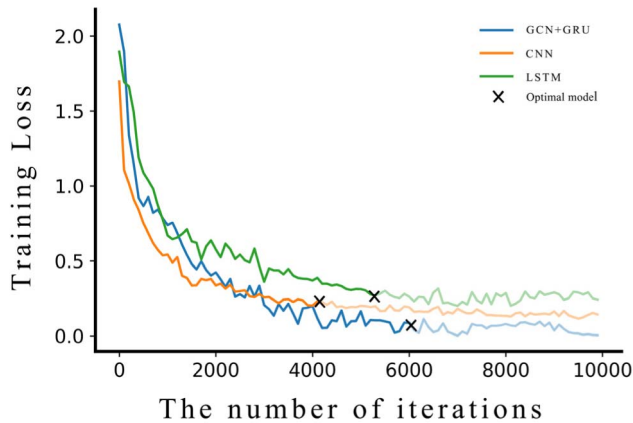| Subject	GCN+GRU	CNN	LSTM	CSP+SVM
1	94.44%	91.67%	86.67%	85.00%
2	85.83%	83.33%	80.33%	77.50%
3	93.33%	89.17%	83.33%	81.67%
Average± std	91.20%± 4.68%	88.06%± 4.28%	83.44% ± 3.17%	81.39%± 3.76%

Fig. 11. A representative example of training loss curves of three deep neural network models (GCN+GRU, CNN and LSTM) from subject 2.

EEG data revealed a significant ERD (second 5–7) and a post-movement ERS (second 7–8) over the contralateral side, and showed a weak ERS over the ipsilateral side and at Cz. The EEG topographies of left and right hand MI from second 5 to 7 across all trials and all subjects were calculated to further analyze ERD/ERS patterns. Blue color stands for ERD (power decrease), and pink color stands for ERS (power increase).

2) Model Training Analysis: Each subject's data were used to train his own models using four methods. All subjects obtained the models with good convergence results. A representative example of training loss curves of three deep neural network models (GCN+GRU, CNN and LSTM) from subject 2 is shown in Fig. 11. The horizontal x-axis represents the number of iterations, the vertical Y-axis represents the value of loss, and the black cross represents the model with optimal parameters selected by the early stopping method. The GCN+GRU model has a lower convergence speed (i.e., the model with optimal parameters appears later) than the other two models. The reason was that the model structure became more complicated when combining GCN and GRU algorisms. However, low training speed during the model training process did not affect the real-time control of the SRL in the online control experiment. In addition, the GCN+GRU model has a smaller loss value at the epoch with optimal parameters than the other two models, i.e., the trained GCN+GRU model has a better classification performance.

3) Model Classification Results: Table II shows the MI classification accuracies of GCN+GRU models at 11 EEG frequency bands for all subjects. Different EEG frequency bands presented different classification accuracies for one subject, e.g., the highest MI classification accuracy was in 12-14 Hz frequency band (91.43%) and the lowest MI classification accuracy was in 28-30 Hz frequency band (76.43%) for subject 1. We selected the GCN+GRU model in the frequency band that achieved the highest MI classification accuracy as the optimal model for each subject.

Three other methods, i.e., CNN, LSTM, and CSP+SVM, were also used to train the recognition models with the same training set and the same optimal frequency band for each subject to compare with the proposed GCN+GRU model. The average confusion matrices of four methods for testing set across all subject are shown in Fig. 12. The main diagonal entries represent the average percentages of correct classification samples and standard deviations, and the off-diagonal entries represent the average percentages of misclassification samples and standard deviations. To evaluate the different recognition models, the classification accuracies of four methods on the same testing set for each subject are calculated and shown in Table III. According to Fig. 12 and Table III, the

TABLE II
MI CLASSIFICATION ACCURACIES OF GCN+GRU MODELS AT 11 EEG FREQUENCY BANDS FOR ALL SUBJECTS

| Subject	8-10 Hz band	10-12 Hz band	12-14 Hz band	14-16 Hz band	16-18 Hz band	18-20 Hz band	20-22 Hz band	22-24 Hz band	24-26 Hz band	26-28 Hz band	28-30 Hz band
1	82.14%	88.57%	91.43%	86.07%	82.50%	81.43%	81.07%	79.64%	79.29%	77.50%	76.43%
2	85.36%	89.64%	93.21%	90.71%	89.29%	86.79%	84.29%	82.14%	83.21%	84.64%	80.36%
3	75.36%	79.64%	81.07%	80.36%	84.64%	82.50%	78.93%	82.14%	84.64%	87.86%	86.79%
4	80.00%	84.29%	91.07%	82.86%	82.14%	81.79%	79.64%	83.57%	89.29%	86.07%	78.93%
5	89.64%	91.07%	88.21%	82.86%	80.36%	77.50%	76.79%	76.07%	77.86%	79.64%	80.00%
6	82.14%	89.64%	82.14%	78.21%	77.50%	76.43%	75.71%	78.21%	78.57%	81.07%	83.93%
7	87.14%	92.14%	88.93%	82.50%	78.93%	75.00%	77.14%	78.93%	77.86%	80.36%	82.86%
8	75.36%	76.79%	80.00%	82.50%	81.43%	78.93%	81.07%	82.14%	85.36%	81.79%	78.21%
9	81.43%	87.14%	85.71%	80.71%	76.43%	77.50%	76.79%	78.57%	78.21%	76.07%	81.07%
10	85.36%	91.43%	91.07%	82.14%	83.93%	80.36%	78.93%	81.43%	77.14%	76.07%	80.36%
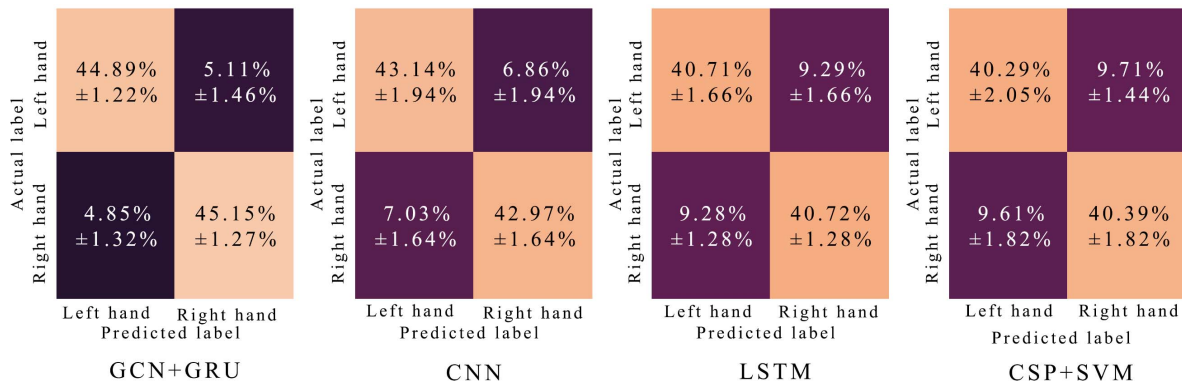
Fig. 12. The average confusion matrices of four methods for testing set across all subjects.

TABLE III
THE OPTIMAL FREQUENCY BAND AND THE CLASSIFICATION
ACCURACIES OF FOUR METHODS ON THE SAME TESTING
SET FOR EACH SUBJECT

Subject	Optimal frequency band	GCN+GRU	CNN	LSTM	CSP+SVM
1	12-14Hz	91.43%	86.07%	82.86%	79.29%
2	12-14Hz	93.21%	90.36%	75.00%	78.93%
3	26-28Hz	87.86%	85.71%	77.86%	79.64%
4	12-14Hz	91.07%	80.00%	84.64%	86.07%
5	10-12 Hz	91.07%	91.79%	82.86%	79.64%
6	8-10Hz	89.64%	85.36%	83.21%	85.36%
7	10-12Hz	92.14%	86.79%	78.93%	84.29%
8	24-26Hz	85.36%	81.07%	83.21%	72.14%
9	10-12Hz	87.14%	87.86%	82.50%	82.14%
10	10-12Hz	91.43%	86.07%	83.21%	79.29%
Average± std		90.04%± 2.36%	86.11%± 3.42%	81.43% ± 2.92%	80.68%± 3.85%

average classification accuracy using GCN+GRU (90.04% ± 2.36%) is 3.93%, 8.61% and 9.36% higher than those using CNN (86.11% ± 3.42%), LSTM (81.43% ± 2.92%) and CSP+SVM (80.68% ± 3.85%) respectively, which demonstrate that GCN+GRU model has a better MI classification performance. A one-way analysis of variance (ANOVA) was applied to evaluate the main effect of the MI recognition method on the classification accuracy. A confidence level of 95% was selected. ANOVA results demonstrated that there was a significant main effect of the recognition method on the classification accuracy (F = 16.044, p < 0.05). Tukey post-hoc test indicated significant differences between GCN+GRU and the other three recognition methods (all p < 0.05).

To further evaluate the classification performance of each class (left hand movement or right hand movement) using different recognition methods, the precision, recall and F-score of the two classes using four methods for each subject are calculated and shown in Table IV. The higher precision, recall and F-score values indicate the better classification performance (GCN+GRU). A 4 (GCN+GRU, CNN, LSTM, and CSP+SVM) ×2 (left hand movement and right hand movement) two-way ANOVA was applied to evaluate the interaction of recognition method × MI class and the main effect of recognition method and MI class on the classification performance. A confidence level of 95% was selected. ANOVA results demonstrated that no significant interaction between the recognition method and MI class (F = 0.002, p > 0.05); there was a main effect of recognition method on the classification performance (F = 98.551, p < 0.05), and there was no main effect of MI class on the classification performance (F = 0.003, p > 0.05).

C. Results of Online Control Experiment

The results of average grasping success rate, task execution time and target object recognition accuracy for each subject in the online control experiment are shown in Table V. The grasping success rate was defined as the percentage of the number of times that the subject successfully grasped the target objects; the task execution time was defined as the time duration between the start of MI and the end of successful grasping; the target object recognition accuracy was defined as the percentage of the number of times that the target objects were correctly recognized. As shown in Table V, for

THE PRECISION, RECALL AND F-SCORE OF THE TWO CLASSES USING FOUR METHODS FOR EACH SUBJECT

Subject		GCN+GRU		CNN		LSTM		CSP + SVM	
		Left hand	Right hand	Left hand	Right hand	Left hand	Right hand	Left hand	Right hand
1	Precision	0.9143	0.9143	0.8592	0.8623	0.8239	0.8333	0.7971	0.7887
	Recall	0.9143	0.9143	0.8652	0.8561	0.8357	0.8214	0.7857	0.8000
	F-Score	0.9143	0.9143	0.8622	0.8592	0.8298	0.8273	0.7914	0.7943
2	Precision	0.9353	0.9291	0.9007	0.9065	0.7536	0.7465	0.7872	0.7914
	Recall	0.9286	0.9357	0.9071	0.9000	0.7429	0.7571	0.7929	0.7857
	F-Score	0.9319	0.9324	0.9039	0.9032	0.7482	0.7518	0.7900	0.7885
3	Precision	0.8681	0.8897	0.8623	0.8521	0.7826	0.7746	0.7943	0.7986
	Recall	0.8929	0.8643	0.8500	0.8643	0.7714	0.7857	0.8000	0.7929
	F-Score	0.8803	0.8768	0.8561	0.8582	0.7770	0.7801	0.7972	0.7957
4	Precision	0.9014	0.9203	0.7917	0.8088	0.8440	0.8489	0.8582	0.8633
	Recall	0.9209	0.9007	0.8143	0.7857	0.8500	0.8429	0.8643	0.8571
	F-Score	0.9110	0.9104	0.8028	0.7971	0.8470	0.8459	0.8612	0.8602
5	Precision	0.9137	0.9078	0.9149	0.9209	0.8286	0.8286	0.7986	0.7943
	Recall	0.9071	0.9143	0.9214	0.9143	0.8286	0.8286	0.7929	0.8000
	F-Score	0.9104	0.9110	0.9181	0.9176	0.8286	0.8286	0.7957	0.7972
6	Precision	0.9044	0.8889	0.8561	0.8511	0.8298	0.8345	0.8511	0.8561
	Recall	0.8849	0.9078	0.8500	0.8571	0.8357	0.8286	0.8571	0.8500
	F-Score	0.8945	0.8982	0.8530	0.8541	0.8327	0.8315	0.8541	0.8530
7	Precision	0.9143	0.9286	0.8652	0.8705	0.7914	0.7872	0.8429	0.8429
	Recall	0.9275	0.9155	0.8714	0.8643	0.7857	0.7929	0.8429	0.8429
	F-Score	0.9209	0.9219	0.8683	0.8674	0.7885	0.7900	0.8429	0.8429
8	Precision	0.8561	0.8511	0.8271	0.7959	0.8345	0.8298	0.7246	0.7727
	Recall	0.8500	0.8571	0.7857	0.8357	0.8286	0.8357	0.7692	0.7286
	F-Score	0.8530	0.8541	0.8059	0.8153	0.8315	0.8327	0.7463	0.7500
9	Precision	0.8768	0.8662	0.8681	0.8897	0.8227	0.8273	0.8261	0.8169
	Recall	0.8643	0.8786	0.8929	0.8643	0.8286	0.8214	0.8143	0.8286
	F-Score	0.8705	0.8723	0.8803	0.8768	0.8256	0.8244	0.8201	0.8227
10	Precision	0.8649	0.8591	0.8582	0.8633	0.8298	0.8345	0.7929	0.7929
	Recall	0.8591	0.8649	0.8643	0.8571	0.8357	0.8286	0.7929	0.7929
	F-Score	0.8620	0.8620	0.8612	0.8602	0.8327	0.8315	0.7929	0.7929
Average ±std	Precision	0.8949±0.0265	0.8955±0.0290	0.8603±0.0342	0.8621±0.0390	0.8141±0.0286	0.8145±0.0332	0.8072±0.0393	0.8118±0.0314
	Recall	0.8949±0.0293	0.8953±0.0270	0.8622±0.0408	0.8599±0.0345	0.8143±0.0350	0.8143±0.0269	0.8112±0.0324	0.8078±0.0383
	F-Score	0.8948±0.0270	0.8953±0.0271	0.8612±0.0365	0.8609±0.0357	0.8142±0.0317	0.8144±0.0299	0.8092±0.0352	0.8097±0.0342

all subjects, the average grasping success rate, the average task execution time and the average target object recognition accuracy are 86.44 ± 3.44%, 18.85 ± 2.52 seconds and 96 ± 5.6%, respectively. In successful grasping trials, the SRL system could be controlled by subjects to complete the target object grasping tasks within 23 seconds, and achieved a high target object recognition accuracy. Fig. 13 shows the average grasping success rate of three grasping tasks across all subjects. The results showed that the bag grasping task achieved the highest average grasping success rate of 90.67% ± 6.44%, and the cup grasping task and the bowl grasping task achieved the average grasping success rate of 86.67% ± 7.70% and 82.00% ± 9.45%, respectively.

V. DISCUSSION

The offline training experiment (MI recognition model training) and online control experiment (target object grasping tasks) were designed to verify the effectiveness of our proposed SRL system, and both of them showed the suitable results.

The results of offline training experiment showed that MI recognition model trained by EEG data of different frequency bands achieved different classification accuracies for one subject, i.e., the optimal EEG frequency band for different subjects were different. Due to the individual differences, one subject' MI EEG presented the different levels of ERD/ERS patterns in different frequency bands, and different subjects' MI EEG presented the different levels of ERD/ERS patterns in one frequency band [47], [48]. The results also showed
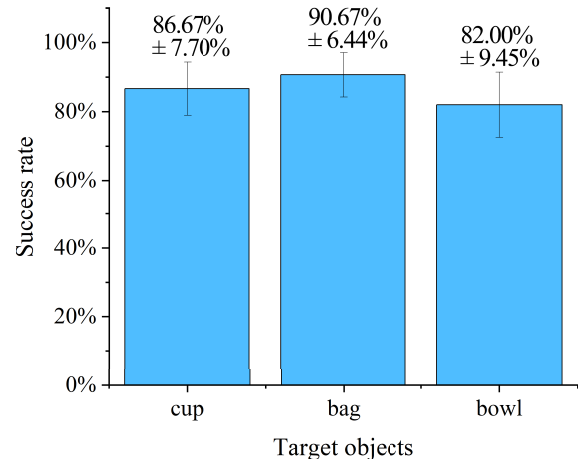

Fig. 13. The average grasping success rate of three grasping tasks across all subjects.

that GCN+GRU model had the higher average classification accuracy (90.04%) than CNN, LSTM and CSP+SVM models. One reason is that GCN is suitable for processing non-Euclidean structure data (EEG signals) by extracting the deep abstract features and topological features of graph data. Lun *et al.* [41] obtained topological features of EEG signals by GCN to improve the MI classification accuracy based on the initial CNN model. The other reason is that GRU can extract the temporal features of EEG. The ERD/ERS patterns change over time during MI. For example, the EEG data reveals a significant ERD at the start and a post-movement

Subject	Success rate(%)	Execution time(s)	Recognition accuracy (%)
1	97.78±3.14	17.57±2.21	97.78±3.14
2	84.44±3.14	17.37±2.79	97.78±3.14
3	91.11±3.14	18.97±4.56	88.89±3.14
4	86.67±5.44	18.67±1.63	100±0
5	86.67±10.89	16.67±3.02	95.56±6.29
6	88.89±3.14	17.54±1.64	97.78±3.14
7	82.22±11.33	19.33±1.44	97.78±3.14
8	82.22±8.31	20.33±1.35	93.33±9.43
9	80±5.44	19.67±3.83	93.33±5.44
10	84.44±6.29	22.57±4.51	97.78±3.14
Average	86.44±8.34	18.85±2.52	96±5.6

ERS over the contralateral side during left/right hand MI [49]. Qiao *et al.* [50] used GRU to extract EEG temporal features from the output data of CNN model for MI classification, and the method with GRU achieved the higher classification accuracy than the traditional CNN model. To compare with the traditional deep learning models, GCN+GRU model could obtain the spatial topological features and temporal features of EEG at the same time, and could achieve the better MI classification performance.

The results of the online control experiment showed that the average grasping success rate was 86.44%, the task execution time of successful grasping trials were within 16-23s, and the average target object recognition accuracy was 96%. All subjects could control the SRL by MI to grasp the target objects in a short time. The reason is that we added object detection function together with MI control, and the target object's position information provided by the object detection module could help SRL move to the target area faster and grasp the target objects more precisely. Cio *et al.* [11] also used the object detection technology to help the robotic arm move faster and to reduce time the robotic arm moved to the target area. The average grasping success rate of bowl grasping task was lower than those of the other two grasping tasks. The possible reason is that the bionic hand could not control the strength and speed during the grasping process so that the target object might move when the bionic hand touched it. During the online control experiment, we tried to test the reliability of the YOLO object detection system under the complex environment. We found that the object recognition also had a high level of accuracy under multi-target environments (e.g., there were more than one object on the desk), but the accuracy was affected adversely by a low-light situation or an occlusion situation. Therefore, some other algorithms like YOLO-in-the-Dark model [51] and CompositionalNets [52], can be further discussed in our next research work to improve object detection accuracy under complex environmental conditions. In addition, one of ten subjects (Subject 10) spent more time in grasping tasks than the others. The possible reason is that the pre-training time before the experiment was not enough and the SRL system was new to him, resulting in the lower MI classification accuracy and more times of MI to control the SRL.

One limitation in our study was that our proposed SRL system were capable of being applied for heathy subjects to perform grasping tasks. However, the SRL system is always useful to augment upper limb performance for elderly or patients suffered upper-limb motor disorder. So it is necessary to test our proposed methods on these users in our future study.

VI. CONCLUSION

In this paper, we developed a wearable SRL system using a hybrid control approach based on MI and object detection, including an SRL hardware subsystem and a hybrid control software subsystem. The system obtained the patient's motion intention through MI EEG signal recognition method based on GCN and GRU to control the left and right movements of SRL, and the object detection technology was used together for a quick grasp of target objects to compensate for the disadvantages of MI control dimension and improve the control accuracy of the SRL. The results showed that the proposed MI EEG recognition method (GCN+GRU) could effectively improve the MI classification accuracy compared with traditional methods; all subjects were able to complete the target object grasping tasks by controlling the SRL in a short time, and they all had a high grasping success rate. The proposed SRL system can effectively assist people with upper-limb motor disorder to perform upper-limb tasks in daily life by natural human–robot interaction, and improve their ability of self-help and enhance their confidence of life. In addition, our study can provide the new thoughts and novel approaches for the application of BCI technology in SRL control and upper-limb motion assistance.

REFERENCES

[1] F. Y. Wu and H. H. Asada, "Implicit and intuitive grasp posture control for wearable robotic fingers: A data-driven method using partial least squares," *IEEE Trans. Robot.*, vol. 32, no. 1, pp. 176–186, Feb. 2016, doi: 10.1109/TRO.2015.2506731.

[2] A. S. Ciullo *et al.*, "A novel soft robotic supernumerary hand for severely affected stroke patients," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 5, pp. 1168–1177, May 2020, doi: 10.1109/TNSRE.2020.2984717.

[3] F. Parietti and H. Asada, "Supernumerary robotic limbs for human body support," *IEEE Trans. Robot.*, vol. 32, no. 2, pp. 301–311, Apr. 2016, doi: 10.1109/TRO.2016.2520486.

[4] C. Khazoom, P. Caillouette, A. Girard, and J.-S. Plante, "A supernumerary robotic leg powered by magnetorheological actuators to assist human locomotion," *IEEE Robot. Autom. Lett.*, vol. 5, no. 4, pp. 5143–5150, Oct. 2020, doi: 10.1109/LRA.2020.3005629.

[5] B. Yang, J. Huang, X. Chen, C. Xiong, and Y. Hasegawa, "Supernumerary robotic limbs: A review and future outlook," *IEEE Trans. Med. Robot. Bionics*, vol. 3, no. 3, pp. 623–639, Aug. 2021, doi: 10.1109/TMRB.2021.3086016.

[6] J. Guggenheim, R. Hoffman, H. Song, and H. H. Asada, "Leveraging the human operator in the design and control of supernumerary robotic limbs," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 2177–2184, Apr. 2020, doi: 10.1109/LRA.2020.2970948.

[7] G. Salvietti *et al.*, "Integration of a passive exoskeleton and a robotic supernumerary finger for grasping compensation in chronic stroke patients: The SoftPro wearable system," *Frontiers Robot. AI*, vol. 8, Jun. 2021, Art. no. 661354, doi: 10.3389/FROBT.2021.661354.

[8] L. Peternel, N. Tsagarakis, and A. Ajoudani, "A human-robot co-manipulation approach based on human sensorimotor information," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 7, pp. 811–822, Jul. 2017, doi: 10.1109/TNSRE.2017.2694553.

[9] U. Muhammad, K. A. Sipra, M. Waqas, and S. Tu, "Applications of myo armband using EMG and IMU signals," in *Proc. 3rd Int. Conf. Mechatronics, Robot. Autom. (ICMRA)*, Oct. 2020, pp. 6–11, doi: 10.1109/ICMRA51221.2020.9398375.

[10] H. Song and H. H. Asada, "Integrated voluntary-reactive control of a human-SuperLimb hybrid system for hemiplegic patient support," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 1646–1653, Apr. 2021, doi: 10.1109/LRA.2021.3058926.

[11] Y.-S. L.-K. Cio, M. Raison, C. L. Menard, and S. Achiche, "Proof of concept of an assistive robotic arm control using artificial stereovision and eye-tracking," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 12, pp. 2344–2352, Dec. 2019.

[12] C.-R. Phang and L.-W. Ko, "Global cortical network distinguishes motor imagination of the left and right foot," *IEEE Access*, vol. 8, pp. 103734–103745, 2020, doi: 10.1109/ACCESS.2020.2999133.

[13] N. Waytowich, A. Henderson, D. Krusienski, and D. Cox, "Robot application of a brain computer interface to staubli tx40 robots–early stages," in *Proc. World Automation Congr.*, Sep. 2010, pp. 1–6.

[14] M. A. Khan, R. Das, H. K. Iversen, and S. Puthusserypady, "Review on motor imagery based BCI systems for upper limb post-stroke neurorehabilitation: From designing to application," *Comput. Biol. Med.*, vol. 123, pp. 1–33, Aug. 2020, doi: 10.1016/j.compbiomed.2020.103843.

[15] O. R. Pinheiro, L. R. G. Alves, and J. R. D. Souza, "EEG signals classification: Motor imagery for driving an intelligent wheelchair," *IEEE Latin Amer. Trans.*, vol. 16, no. 1, pp. 254–259, Jan. 2018, doi: 10.1109/TLA.2018.8291481.

[16] K. Lee, D. Liu, L. Perroud, R. Chavarriaga, and J. D. R. Millán, "A brain-controlled exoskeleton with cascaded event-related desynchronization classifiers," *Robot. Auton. Syst.*, vol. 90, pp. 15–23, Apr. 2017, doi: 10.1016/j.robot.2016.10.005.

[17] J. Jin, Z. Wang, R. Xu, C. Liu, X. Wang, and A. Cichocki, "Robust similarity measurement based on a novel time filter for SSVEPs detection," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Oct. 14, 2021, doi: 10.1109/TNNLS.2021.3118468.

[18] Z. Tang, C. Li, and S. Sun, "Single-trial EEG classification of motor imagery using deep convolutional neural networks," *Optik*, vol. 130, pp. 11–18, Feb. 2017, doi: 10.1016/j.ijleo.2016.10.117.

[19] J. Jin, R. Xiao, I. Daly, Y. Miao, X. Wang, and A. Cichocki, "Internal feature selection method of CSP based on L1-norm and Dempster–Shafer theory," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 4814–4825, Nov. 2021, doi: 10.1109/TNNLS.2020.3015505.

[20] J. Jin, Y. Miao, I. Daly, C. Zuo, D. Hu, and A. Cichocki, "Correlation-based channel selection and regularized feature optimization for MI-based BCI," *Neural Netw.*, vol. 118, pp. 262–270, Oct. 2019, doi: 10.1016/j.neunet.2019.07.008.

[21] Y. Chu, X. Zhao, Y. Zou, W. Xu, J. Han, and Y. Zhao, "A decoding scheme for incomplete motor imagery EEG with deep belief network," *Frontiers Neurosci.*, vol. 12, pp. 1–17, Sep. 2018, doi: 10.3389/FNINS.2018.00680.

[22] J. Zhou, M. Meng, Y. Gao, Y. Ma, and Q. Zhang, "Classification of motor imagery EEG using wavelet envelope analysis and LSTM networks," in *Proc. Chin. Control Decis. Conf. (CCDC)*, Jun. 2018, pp. 5600–5605, doi: 10.1109/CCDC.2018.8408108.

[23] N. Shajil, S. Mohan, P. Srinivasan, J. Arivudaiyanambi, and A. Arasappan Murrugesan, "Multiclass classification of spatially filtered motor imagery EEG signals using convolutional neural network for BCI based applications," *J. Med. Biol. Eng.*, vol. 40, no. 5, pp. 663–672, Oct. 2020, doi: 10.1007/s40846-020-00538-3.

[24] B. Xu *et al.*, "Wavelet transform time-frequency image and convolutional network-based motor imagery EEG classification," *IEEE Access*, vol. 7, pp. 6084–6093, 2018, doi: 10.1109/ACCESS.2018.2889093.

[25] G. Dai, J. Zhou, J. Huang, and N. Wang, "HS-CNN: A CNN with hybrid convolution scale for EEG motor imagery classification," *J. Neural Eng.*, vol. 17, no. 1, pp. 1–11, Jan. 2020, doi: 10.1088/1741-2552/ab405f.

[26] Y. Hou *et al.*, "Deep feature mining via attention-based BiLSTM-GCN for human motor imagery recognition," 2020, *arXiv:2005.00777*.

[27] H. Wang, L. Xu, A. Bezerianos, C. Chen, and Z. Zhang, "Linking attention-based multiscale CNN with dynamical GCN for driving fatigue detection," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021, doi: 10.1109/TIM.2020.3047502.

[28] T. Zhang, X. Wang, X. Xu, and C. L. P. Chen, "GCB-Net: Graph convolutional broad network and its application in emotion recognition," *IEEE Trans. Affect. Comput.*, vol. 13, no. 1, pp. 379–388, Jan./Mar. 2019, doi: 10.1109/TAFFC.2019.2937768.

[29] G. Pfurtscheller and F. H. L. da Silva, "Event-related EEG/MEG synchronization and desynchronization: Basic principles," *Clin. Neurophysiol.*, vol. 110, pp. 1842–1857, Nov. 1999, doi: 10.1016/S1388-2457(99)00141-8.

[30] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[31] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*.

[32] L. Zhao, Y. Song, C. Zhang, and Y. Liu, "T-GCN: A temporal graph convolutional network for traffic prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 9, pp. 3848–3858, Sep. 2020, doi: 10.1109/TITS.2019.2935152.

[33] Y. Yu, Y. Liu, E. Yin, J. Jiang, Z. Zhou, and D. Hu, "An asynchronous hybrid spelling approach based on EEG–EOG signals for Chinese character input," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 6, pp. 1292–1302, Jun. 2019, doi: 10.1109/TNSRE.2019.2914916.

[34] Y. Yu *et al.*, "Toward a hybrid BCI: Self-paced operation of a P300-based speller by merging a motor imagery-based 'brain switch' into a P300 spelling approach," *Int. J. Hum. Comput. Interact.*, vol. 33, no. 8, pp. 623–632, Aug. 2017, doi: 10.1080/10447318.2016.1267450.

[35] C. Penaloza, D. Hernandez-Carmona, and S. Nishio, "Towards intelligent brain-controlled body augmentation robotic limbs," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2018, pp. 1011–1015, doi: 10.1109/SMC.2018.00180.

[36] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019, doi: 10.1109/TNNLS.2018.2876865.

[37] C. Shi, D. Yang, J. Zhao, and H. Liu, "Computer vision-based grasp pattern recognition with application to myoelectric control of dexterous hand prosthesis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 9, pp. 2090–2099, Sep. 2020, doi: 10.1109/TNSRE.2020.3007625.

[38] X. Chen, X. Huang, Y. Wang, and X. Gao, "Combination of augmented reality based brain-computer interface and computer vision for high-level control of a robotic arm," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 12, pp. 3140–3147, Dec. 2020, doi: 10.1109/TNSRE.2020.3038209.

[39] R. David *et al.*, "TensorFlow lite micro: Embedded machine learning on TinyML systems," 2020, *arXiv:2010.08678*.

[40] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," *IEEE Trans. Affect. Comput.*, vol. 11, no. 3, pp. 532–541, Jul./Sep. 2020.

[41] Y. Hou *et al.*, "GCNs-Net: A graph convolutional neural network approach for decoding time-resolved EEG motor imagery signals," 2020, *arXiv:2006.08924*.

[42] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[43] A. Al-Saegh, S. A. Dawwd, and J. M. Abdul-Jabbar, "Deep learning for motor imagery EEG-based classification: A review," *Biomed. Signal Process. Control*, vol. 63, Jan. 2021, Art. no. 102172, doi: 10.1016/j.bspc.2020.102172.

[44] Y. Pei *et al.*, "A tensor-based frequency features combination method for brain-computer interfaces," in *Proc. Int. Conf. Cognit. Syst. Signal Process.*, 2021, pp. 511–526.

[45] H. Sun, Y. Xiang, Y. Sun, H. Zhu, and J. Zeng, "On-line EEG classification for brain-computer interface based on CSP and SVM," in *Proc. 3rd Int. Congr. Image Signal Process.*, Oct. 2010, pp. 4105–4108, doi: 10.1109/CISP.2010.5648081.

[46] B. Blankertz *et al.*, "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 153–159, Jun. 2006.

[47] Z. Wang, Y. Yu, M. Xu, Y. Liu, E. Yin, and Z. Zhou, "Towards a hybrid BCI gaming paradigm based on motor imagery and SSVEP," *Int. J. Hum.-Comput. Interact.*, vol. 35, no. 3, pp. 197–205, 2019.

[48] Y. Yu *et al.*, "Toward brain-actuated car applications: Self-paced control with a motor imagery-based brain-computer interface," *Comput. Biol. Med.*, vol. 77, pp. 148–155, Oct. 2016.

[49] Y. Jeon, C. S. Nam, Y.-J. Kim, and M. C. Whang, "Event-related (De)synchronization (ERD/ERS) during motor imagery tasks: Implications for brain–computer interfaces," *Int. J. Ind. Ergonom.*, vol. 41, no. 5, pp. 428–436, Sep. 2011, doi: 10.1016/j.ergon.2011.03.005.

[50] W. Qiao and X. Bi, "Deep spatial-temporal neural network for classification of EEG-based motor imagery," in *Proc. Int. Conf. Artif. Intell. Comput. Sci.*, Jul. 2019, pp. 265–272.

[51] Y. Sasagawa and H. Nagahara, "YOLO in the dark–domain adaptation method for merging multiple models," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland, 2020, pp. 345–359, doi: 10.1007/978-3-030-58589-1_21.

[52] A. Kortylewski, Q. Liu, A. Wang, Y. Sun, and A. Yuille, "Compositional convolutional neural networks: A robust and interpretable model for object recognition under occlusion," *Int. J. Comput. Vis.*, vol. 129, no. 3, pp. 736–760, Mar. 2021.