Upper Limb Movement Decoding Scheme Based on Surface Electromyography Using Attention-Based Kalman Filter Scheme

Anyuan Zhang, Qi Li, Zhenlan Li, and Jiming Li

Abstract—**Convolutional neural network (CNN)-based models are widely used in human movement decoding based on surface electromyography. However, they capture only the spatial information of the surface electromyography and lack prior knowledge of the system, resulting in unsatisfactory decoding accuracy. To address these issues, we propose an attention-based Kalman filter scheme (AKFS), which uses an attention-based CNN model to better extract temporal information and a KF to add prior knowledge of the system. We further solve the problem of insufficient data due to the short training time of new subjects by using a transfer learning method based on a fine-tuning strategy. The proposed scheme was tested in four scenarios: intra-session, intra-session long-time use, inter-subject, and inter-subject with a fine-tuning strategy. The proposed attention-based CNN model outperformed the vanilla CNN model and a hybrid CNN-long short-term memory (LSTM) model in intra-session and intra-session long-time use. After fine-tuning with a small amount of data on a new subject, the attention-based CNN model achieved higher decoding accuracy than the vanilla CNN model and lower response time than CNN-LSTM model. Furthermore, the schemes with KF outperformed the schemes without KF in all scenarios. Our proposed scheme improves the decoding accuracy of the traditional CNN model for a single subject by better capturing the temporal information of the surface electromyography signal and increasing the prior knowledge of the system. Additionally, the proposed scheme can be easily transferred to a new subject using only a small amount of data.**

Index Terms—**Attention mechanism, convolutional neural network, human movement decoding, Kalman filter, long short-term memory.**

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of Changchun University of Science and Technology (CUST) under Application No. 20210013.

Anyuan Zhang and Jiming Li are with the School of Computer Science and Technology, Changchun University of Science and Technology, Changchun 130022, China.

Qi Li is with the School of Computer Science and Technology, Changchun University of Science and Technology, Changchun 130022, China, and also with the Zhongshan Institute of Changchun University of Science and Technology, Zhongshan 528437, China (e-mail: liqi@cust.edu.cn).

Zhenlan Li is with the Department of Physical Medicine and Rehabilitation, The First Hospital of Jilin University, Changchun 130012, China.

Digital Object Identifier 10.1109/TNSRE.2023.3262269

I. INTRODUCTION

HUMAN movement decoding based on surface electromyography (sEMG) has been widely used in prosthesis control and human–machine interaction (HMI) [1], [2], [3], [4]. Many researchers have adopted convolutional neural network (CNN) models as practical approaches to decode human movement from sEMG [5], [6], [7], [8]. In contrast with machine learning approaches [9], [10], these CNN models are not as time-consuming in feature engineering and can realize multiple degrees of freedom (DoFs) angle prediction in human movement decoding [11]. Moreover, these CNN models can use small amounts of data after electrode shifting and user change to retrain the model through a fine-tuning strategy to improve the robustness of the decoding method in the case of electrode movement and user change [12], [13], [14], [15]. However, due to neglect of the temporal information of sEMG and lack of prior knowledge of the system, the accuracy of human movement decoding based on CNN models is unsatisfactory, which limits its clinical application [16].

Many researchers have tried to improve the accuracy of human movement decoding schemes based on CNN models by proposing mixed structure models [17], [18], [19], [20]. Although these CNN models are popular solutions for human movement decoding based on sEMG, single CNN methods focus only on the spatial information of sEMG and ignore the temporal information of sEMG. Thus, a novel structure model to better capture spatial information and temporal information of sEMG needs to be explored. Because the long short-term memory (LSTM) structure is suitable for processing time series and has achieved considerable improvements in human movement decoding based on sEMG [21], [22], [23], [24], many researchers have proposed hybrid CNN-LSTM structures to improve the accuracy of human movement decoding based on CNN models by better capturing the temporal information of sEMG. Xia et al. [25] proposed a CNN-LSTM model to decode the upper limb movement in three dimensions (3D) space. This method performed better than the support vector regression (SVR) and CNN models. Bao et al. [26] also proposed a CNN-LSTM method; they adopted a CNN model to extract features from Fourier-transformed sEMG and adopted an LSTM model to estimate the degrees of angles from human joints. The method is superior to a single CNN model in decoding accuracy.

The attention mechanism has recently been proposed to improve the performance of LSTM-based models in natural language processing [27], [28], [29], [30] by focusing on more important information in time series through an attention weight matrix. Hu et al. [31] proposed an attention-based CNN architecture for sEMG-based gesture classification that outperformed CNN, LSTM, and CNN-LSTM models. However, their method focused solely on human discrete movement decoding. Consequently, the effectiveness of the attention mechanism in human continuous movement decoding needs investigation.

Inspired by the application of the attention mechanism in human discrete movement decoding, we hypothesized that it can further improve the performance of CNN-LSTM models in human continuous movement decoding using sEMG by better capturing the temporal information of sEMG.

Lack of prior knowledge of the system also leads to low accuracy of human movement decoding based on CNN models. To address this issue, many researchers have adopted post-processing strategies to increase the prior knowledge of human movement decoding systems. Englehart and Hudgins [32] proposed a majority vote approach to determine the current output of the decoding model through voting on the previous three or more outputs of the decoding model to reduce classification error outputs. Amsuss et al. [33] proposed a post-processing strategy based on an artificial neural network to reduce classification errors. The movement was determined by the maximum likelihood estimation of the artificial neural network and the forearm muscle activity. Their method significantly reduced the classification errors of discrete movement decoding. Zhang et al. [34] proposed a post-processing strategy based on a threshold-based motion onset detection method to reduce classification errors. The method significantly reduced the classification errors compared to the original classification. These post-processing strategies add prior knowledge to human movement decoding schemes and significantly reduce the classification errors of the schemes. However, as they focus solely on human discrete movement decoding, a post-processing strategy for human continuous movement decoding needs to be explored.

The Kalman filter (KF) has been applied to CNN-LSTM models to improve the performance of human continuous movement decoding. KF uses Kalman gain to determine the weight of the internal transition and observation models and uses noise measurement to calculate the optimal output of the model [35]. Bao et al. [36] proposed a CNN-LSTM with a KF scheme to estimate the single DoF of the finger and wrist. The CNN-LSTM with KF scheme achieved significantly higher performance than the CNN scheme and CNN-LSTM schemes. However, their method focused solely on the prediction performance of a single DoF joint angle under the same subject. The performance of KF in predicting multiple DoF joint angles between different subjects requires further investigation.

To solve the low decoding accuracy of CNN due to the neglect of temporal information of sEMG and the lack of prior knowledge of the system, we propose a human movement decoding scheme named the attention-based KF scheme (AKFS), which combines an attention-based CNN model with a KF. We adopt the attention-based CNN model to improve the decoding accuracy of the CNN model by better capturing the temporal information of sEMG and adopt the KF as a post-processing strategy to improve the decoding accuracy of the CNN model by adding prior knowledge of the system. In contrast to previous studies, we use the attention mechanism to solve the problem of continuous motion decoding based on sEMG signals, and explore the problem of multi-degree of freedom angle prediction of KF in different subjects. The proposed scheme was evaluated in four scenarios: Intra-session, intra-session long-time use, inter-subject, and inter-subject with a fine-tuning strategy. To evaluate the effectiveness of each part of the proposed scheme, we examined the performance of six different schemes (CNN, CNN-LSTM, attention-based CNN, CNN with KF, CNN-LSTM with KF, and AKFS) by combining the three decoding models with the two post-processing strategies under the four scenarios.

II. METHODS

A. Subjects

Ten non-disabled subjects (all right-handed males, age: 26–30 years) participated in this experiment. None of the subjects had a history of neuromuscular diseases. Before the experiment, we informed the subjects of the process and purpose of the experiment, and they provided informed consent. The study was approved by the ethics committee of Changchun University of Science and Technology (CUST), 20210013, July 3, 2021.

B. Collection of sEMG Data

Seven dual-electrode wireless sEMG sensors (Noraxon, USA) were used to collect sEMG signal from seven positions corresponding to the following muscles: biceps, brachioradialis, flex carpi rad, flex carpi ulnaris, triceps, ext carpi ulnaris and ext dig communis (as shown in Fig 1.). Sensors were placed on the belly of muscles, according to the Noraxon instructions. The locations of the electrodes on the skin were cleaned with 75% alcohol to reduce the impedance between the skin and the electrodes. The sEMG were sampled at 1500 Hz and were band-pass filtered by a seventh-order Butterworth filter in the range of 20–500 Hz.

Each subject performed nine upper limb movements, which included elbow flexion (EF), wrist flexion (WF), wrist extension (WE), wrist pronation (WP), wrist supination (WS), wrist flexion and pronation (WFP), wrist flexion and supination (WFS), wrist extension and pronation (WEP), and wrist extension and supination (WES).

Data from two sessions were collected for each subject. Each session contained 9 upper limb movements and each movement contained 5 repetitions. Each repetition consisted of four time periods: 3 s of relaxing in the original position, 3 s of reaching the target position, 3 s of keeping the target position, and 3 s of returning to the original position; a total of 12 s. Thus, each movement contains 60 s continuous data. Between every two movements, all subjects had enough rest to avoid fatigue. Each session lasted approximately 30 min.

C. Kinematic Data Collection

To ensure that all the subjects' motions had similar movement trajectories and to label the sEMG data to obtain sufficient training data, we developed a HoloLens virtual system (HVS), as shown in Fig. 2.

A red cross was generated in front of the subject according to the limb position by a unity program run on HoloLens. The trajectory of the upper limb motion was simulated by the trajectory of the red cross. A laser pen was fixed to the hand of the subject. A green cross was also projected in front of the subject by the laser pen. The subject was instructed to control the green cross to track the trajectory of the red cross. An MPU6050 sensor was fixed on the laser pen to collect the continuous kinematic data of upper limb movement (the Euler angles of three axes were calculated from the data of the accelerometer and gyroscope). To synchronize the kinematic data with the sEMG of upper limb movements, the MPU6050 sensor communicated with an STM32F103 board at 115200 baud through a serial port and synchronized with Noraxon sEMG acquisition software. The kinematic data of the human upper limb were sampled by MPU6050 at 100 Hz. We normalized the collected sEMG signal and kinematic data, respectively, as shown in (1).

$$\hat{x}_i = \frac{x_i - u}{\sigma} \tag{1}$$

where x_i is the ith sample of data, u is the mean value of the data, σ is the variance of the data, and $\hat{x}_i$ is the ith sample of normalized data.

During the experiment, the hand of each subject was placed on a cross-shaped metal rod with the elbow supported by a bracket. When the subjects felt uncomfortable, they were allowed to quit at any time. All subjects conducted nine upper limb movements each by tracking the trajectory of the red cross movement in a unity program on a HoloLens.

The trajectories of the nine movements are shown in Fig. 3. All subjects were encouraged to control the green cross projected by a laser pen to follow the trajectories of a red cross projected by HoloLens without fatigue. Before the experiment, the subjects familiarized themselves with the nine upper limb movements and the HVS.

D. Data Segmentation

The sEMG signals were segmented using a 250 ms sliding window with 80% overlapping (Fig. 4). To label the sEMG signal, we downsampled the kinematic data from 100 Hz to 20 Hz (taking the last point every five points). The number of samples in the first second was 16, and thereafter, 20 samples per second. In this way, the kinematic data collected by MPU6050 can be labeled with the sEMG signals to interpret the movement of the human upper limb.

E. Attention-Based CNN

An attention-based CNN model was introduced in this study (Fig. 5). Raw sEMG signals from seven channels in a sliding window were used as the input of the model to directly decode the human movement intention. The model was mainly composed of one input layer, three convolutional modules, two LSTM modules, one attention mechanism module, one fully connected layer, and one output layer. The structure of the CNN module was adopted from [14], and the structure of the LSTM module was adopted from [25]. Because the decoding test in our study was significantly simpler, the structure of the proposed model was significantly simpler than those of the original models. Each sliding window had 375 samples of 7 channels as the input of the model. The filter numbers of the 3 convolution layers were 64, 96, and 128, and the kernel sizes of the 3 layers were 23, 13, and 11. All convolution layers adopted the same padding in the stride of 1.

There were five layers in the first and second CNN modules (convolution layer, batch normalization layer, rectified linear unit (ReLU) layer, average pooling layer, and a dropout layer). The epsilon of the batch normalization layers was set to 0.001. The pooling size and the stride of the first pooling layer were set to 15. The pooling size and the stride of the second pooling layer were set to five. The drop rate of the two dropout layers was set to 0.15. The third CNN module had a special design: There was no pooling or dropout layer in the third CNN module. The number of hidden units of the two LSTM modules was 128. The number of time steps of the two LSTM modules was set to five according to the increment of the sliding window. An attention mechanism module was followed by the last LSTM module to enhance the relationship between the sequences. The fully connected layer was followed by the attention mechanism module, which comprised 128 hidden units. Finally, the three Euler angles of the three axes were output by a regression layer.

The attention mechanism adopted in this study is shown in Fig. 6. After the second LSTM module, the shape of the sequence was 128×5. Q, K, and V matrices were defined. Q, K, and V were obtained from the dot product of the output of the second LSTM and three 128×128 matrices. The dimensions of the Q, K, and V matrices were 5×128. The Q matrix was multiplied by the transpose of the K matrix and then passed through a softmax layer to obtain a 5×5 relationship matrix between sequences. Then, the dot product of the relationship matrix with the V matrix produced the attention weight matrix. The output attention weight matrix can be described as in (2).

$$Attention(Q, K, V) = softmax(QK^T)V \tag{2}$$

where Q, K, and V are three matrices. A softmax function was adopted in the attention mechanism.

For human decoding schemes, the model was trained using an Nvidia Quadro P620 GPU. The stochastic gradient descent optimizer with momentum was applied to the model. The momentum and learning rate were set to 0.001 and 0.8 based on the results of multiple experiments.

F. Kalman Filter (KF)

A KF was adopted in this study to add prior knowledge of the system. The kinematics of the transition model and

observation model are described by (3) and (4):

$$x_i = Fx_{i-1} + w \tag{3}$$
$$z_i = Hx_i + v \tag{4}$$

where x_i is the output of the transition model at time i, F is the state transition matrix, z_i is the output of the observation model at time i, H is the measurement transformation matrix, and w and v are the state and measurement uncertainties drawn from Gaussian distributions N(0, Q) and N(0, R).

After these parameters are determined, the state transition equation and the prior state covariance matrix of the system can be expressed using (5) and (6):

$$\hat{x}_i = Fx_{i-1} \tag{5}$$
$$\hat{P}_i = FP_{i-1}F^T + Q \tag{6}$$

where is the prior state estimate for the time i and is the prior state covariance matrix.

The Kalman gain can be calculated using (7):

$$K_i = \frac{\hat{P}_i H^T}{H \hat{P}_i H^T + R} \tag{7}$$

The output and covariance matrix P_i optimized by the KF can be expressed using (8) and (9):

$$x_i = \hat{x}_i + K_i(z_i - H\hat{x}_i) \tag{8}$$
$$P_i = (1 - \hat{K}_i H)\hat{P}_i \tag{9}$$

In KF, the parameters H, Q, and R need to be determined manually. H was set to 1, Q was set to 10^{-4}, R was set to 10^{-3}, and z_i was the real output of the scheme. The designed KF was followed by a decoding model to add prior knowledge of the system.

G. Scheme Adaptability Analysis Experiment

In human movement decoding based on sEMG, the model is usually challenged by long-time use and user changes. Thus, a robust model that can cope with long-time use and user changes is essential. To solve this issue, we proposed AKFS that combines the attention-based CNN model with KF. To evaluate its effectiveness, we examined the performance of six different schemes (CNN, CNN-LSTM, attention-based CNN, CNN with KF, CNN-LSTM with KF, and AKFS) under four scenarios. The four scenarios were as follows.

1) Scenario 1. Intra-Session: The first session from each subject was chosen as the data, and five-fold cross-validation was performed on the data. The first session contained 9 movements and each movement contained 5 repetitions. Each repetition was segmented by a 250 ms sliding window with 80% overlapping. Thus, there were 236 samples in each movement trial. Therefore, the training set contained 8496 samples (9 movements × 4 trials × 236 samples) and used the remaining trials of all movements as the testing set (9 movements × 1 trial × 236 samples). Five-fold cross-validation was conducted. One repetition of one subject was chosen as the training set and the remaining repetitions were chosen as the testing set.

2) Scenario 2. Intra-Session Long-Time Use: Two sessions from each subject were used as the data. Each session contained 9 movements. To simulate the long-time use of movements, the five consecutive repetitions of each movement (60 s) were segmented. Each movement was segmented by a 250 ms sliding window, with 80% overlapping. Thus, there were 10764 samples in a session (9 movements × 1196 samples) after segmentation. Two-fold cross-validation was conducted. One session of one subject was chosen as the training set and the other session was chosen as the testing set.

3) Scenario 3. Inter-Subject: The first session of each subject was chosen as the data. The continuous 60s data of each movement of the first session was segmented by a 250 ms sliding window with 80% overlapping. The first session of each subject contained 10764 samples (9 movements × 1196 samples). Two-fold cross-validation was conducted on each pair of subjects. We chose data from one subject as the training set and the trained model was tested on the first session of the remaining subjects. To investigate whether the proposed scheme can learn the human muscle activities of multiple subjects, we also investigated whether a model trained on multiple subjects can accurately predict data from a new subject. Ten-fold cross-validation was conducted. We chose data from the first session of nine subjects as the training set (9 movements × 9 subjects × 1196 samples) and data from the first session of the remaining subject as the testing set (9 movements × 1196 samples).

4) Scenario 4. Inter-Subject With a Fine-Tuning Strategy: After changing the subjects, the performance of the model with the data from new subjects will be significantly reduced. To solve this problem, a fine-tuning strategy was applied, in which only a small amount of sEMG data was used to retrain the model to obtain a satisfactory performance of the model for the specific subject. Two-fold cross-validation was conducted on each pair of subjects. The data from the first session of one subject was chosen as the training set (9 movements × 1196 samples). The data from the first session of the remaining subjects were used to construct testing and calibration data. Five-fold cross-validation was conducted on the test subject. We used one repetition of each movement as a calibration set (9 movements × 236 samples) to retain the trained model. The remaining four repetitions were used as the testing set (9 movements × 4 trials × 236 samples) to evaluate the fine-tuned model. The efficiency of the model trained on multiple subjects was also investigated. Ten-fold cross-validation was conducted. We used the first session of nine subjects as the training set (9 subjects × 9 movements × 1196 samples). Five-fold cross-validation was conducted on the test subject. We used one repetition from each movement as a calibration set (9 movements × 236 samples) to fine-tune the trained model. The remaining four repetitions were used as the testing set (9 movements × 4 trials × 236 samples) to evaluate the fine-tuned model.

H. Performance Index and Statistical Analysis

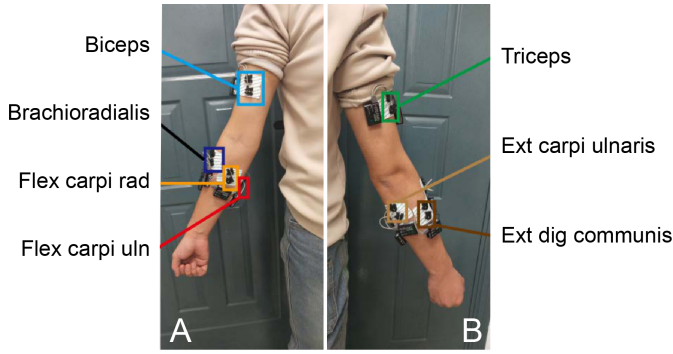The coefficient of determination (R^2) was used as the performance index to evaluate the efficiency of the decoding

Fig. 1. Experimental setup of sEMG sensors on seven muscles of subjects. (A) Positions of sensors on biceps, brachioradialis, flex carpi rad, and fulnaris. (B) Positions of sensors on triceps, ext carpi ulnaris, and ext dig communis.
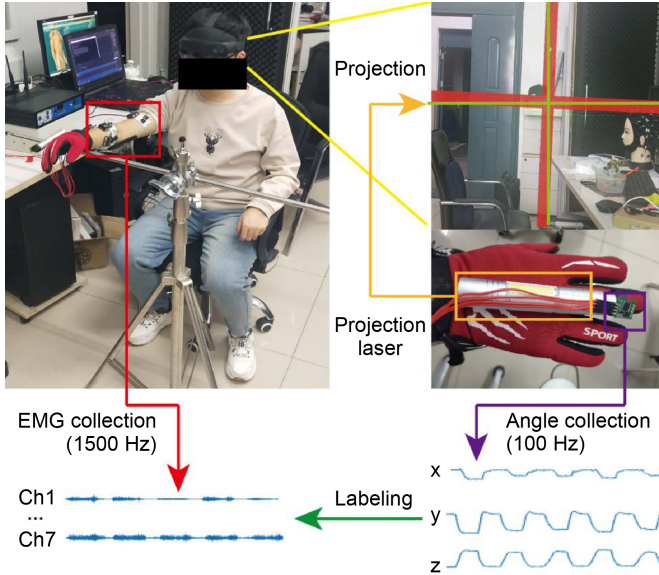

Fig. 2. Flow of operation of the HoloLens virtual system (HVS). A HoloLens projects a red cross in a real environment and the subject controls a green cross projected by a laser pen to track the red cross. The upper limb movement is interpreted by the kinematic data of MPU6050 (Euler angles of x, y, and, z axes). A computer records the Euler angles of MPU6050 while recording the sEMG signals of the subject.

schemes. R^2 can be described as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{D} \sum_{t=0}^{N} (P_i(t) - F_i(t))^2}{\sum_{i=1}^{D} \sum_{t=0}^{N} (F_i(t) - \overline{F_i(t)})^2} \quad (10)$$

where $F_i(t)$ is the real angle of the ith axis, $\overline{F_i(t)}$ is the average angle of the ith axis, $P_i(t)$ is the angle predicted by the scheme, N is the number of samples in a sliding window, and D is the number of axes.

The R^2 in the four scenarios was analyzed using a repeated measure analysis of variance (ANOVA) with factors including the three decoding models (CNN, CNN-LSTM, and attention-based CNN) and two post-processing strategies (with KF and without KF) with a significance level of 0.05. A Bonferroni corrected post-hoc test was also conducted. All results were analyzed on a personal computer running IBM ®SPSS Statistics 22 software.
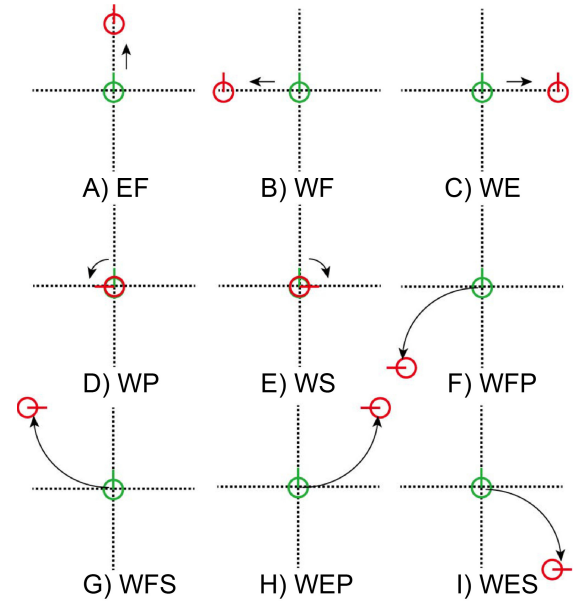

Fig. 3. Trajectories of nine upper limb movements projected by HoloLens in a real environment. The green and red marks represent the crosses projected by the laser pen and HoloLens, respectively. A)–D) are the 1-DoF movements, whereas E)–I) are the 2-DoF movements.
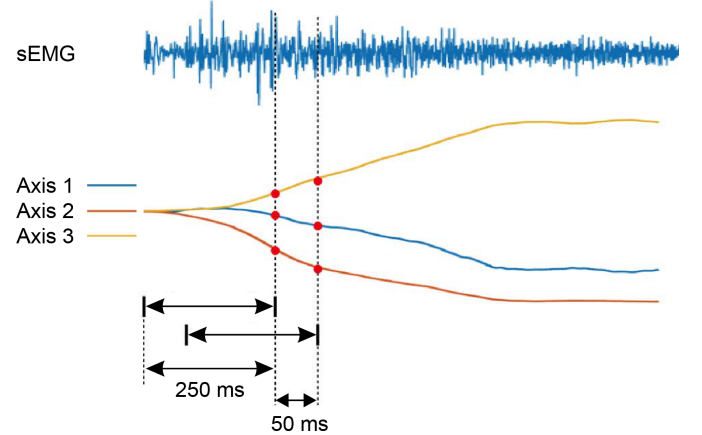

Fig. 4. Relationship between kinematic data and sEMG signals. The red dots represent the labels of the sEMG samples.
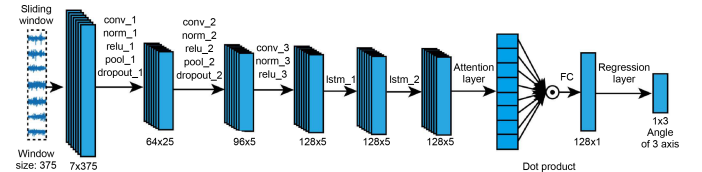

Fig. 5. Structure of the attention-based CNN model used to decode the angle of the three axes.

III. RESULTS

A. Results on Intra-Session Decoding Accuracy

Table I summarizes the decoding accuracy (R^2) of the six decoding schemes of the ten subjects in the intra-session scenario. An ANOVA analysis showed that both the decoding model ($F_{2,8} = 13.297$, $p = 0.003$) and post-processing strategy ($F_{1,9} = 76.278$, $p < 10^{-3}$) had a significant influence on the decoding accuracy of the schemes and there was no significant interaction between the decoding model and post-processing strategy ($F_{2,8} = 2.145$, $p = 0.179$). The schemes
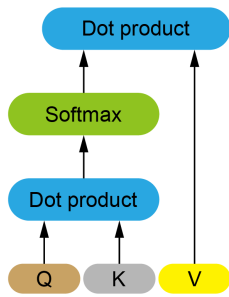
Fig. 6. Attention mechanism adopted in this study. Q, K, and V are 5×128 matrices.

with attention-based CNN (attention-based CNN and AKFS) had significantly higher accuracy than the schemes with CNN (CNN and CNN with KF) (p = 0.005) and the schemes CNN-LSTM models (CNN-LSTM and CNN-LSTM with KF) (p = 0.005). There was no significant difference in the decoding accuracy between the schemes with CNN and CNN-LSTM (p = 0.443). The schemes (CNN with KF, CNN-LSTM with KF, and AKFS) with KF had a significantly higher accuracy than the schemes (CNN, CNN-LSTM, and attention-based CNN) without KF (p < 10^{-3}).

B. Results on Intra-Session Long-Time Use

Table II summarizes the decoding accuracy (R^2) of the six decoding schemes obtained after combining them with three decoding models and two post-processing strategies of the ten subjects in the intra-session long-time use scenario. An ANOVA analysis showed that both the decoding model ($F_{2,8} = 25.988$, p < 10^{-3}) and post-processing strategy ($F_{1,9} = 138.798$, p < 10^{-3}) had a significant influence on the accuracy of the decoding schemes. A significant interaction was found between the decoding model and post-processing strategy ($F_{2,8} = 6.651$, p = 0.02).

The analysis of the simple effect between the decoding model and post-processing strategy showed that the decoding model had a significant effect on the accuracy of the schemes without KF (CNN, CNN-LSTM, and attention-based CNN) ($F_{2,8} = 27.200$, p < 10^{-3}) or the schemes with KF (CNN with KF, CNN-LSTM with KF, and AKFS) ($F_{2,8} = 27.731$, p < 10^{-3}). Without KF, the attention-based CNN had higher accuracy than the CNN (p = 0.001) and the CNN-LSTM (p = 0.033). With KF, the AKFS model had significantly higher accuracy than the CNN with KF (p = 0.001) and the CNN-LSTM with KF (p = 0.049). However, there was no significant difference in accuracy between the schemes with CNN and CNN-LSTM with or without KF (CNN, CNN with KF, CNN-LSTM, and CNN-LSTM with KF) (p > 0.05).

A simple effect analysis showed that post-processing strategy had a significant effect on the accuracy of the schemes with CNN (CNN and CNN with KF) ($F_{1,9} = 138.857$, p < 10^{-3}), the schemes with CNN-LSTM (CNN-LSTM and CNN-LSTM with KF) ($F_{1,9} = 110.152$, p < 10^{-3}) and the schemes with attention-based CNN (attention-based CNN and AKFS) ($F_{1,9} = 113.778$, p < 10^{-3}). The schemes with KF (CNN with KF, CNN-LSTM with KF, and AKFS) had significantly higher accuracy than the schemes without KF (CNN, CNN-LSTM, and attention-based CNN) (p < 10^{-3}).

TABLE I
R^2 OF CNN, CNN-LSTM, ATTENTION-BASED CNN, CNN WITH KF, CNN-LSTM WITHKF, AND AKFS IN THE OF INTRA-SESSION LONG-TIME USE SCENARIO

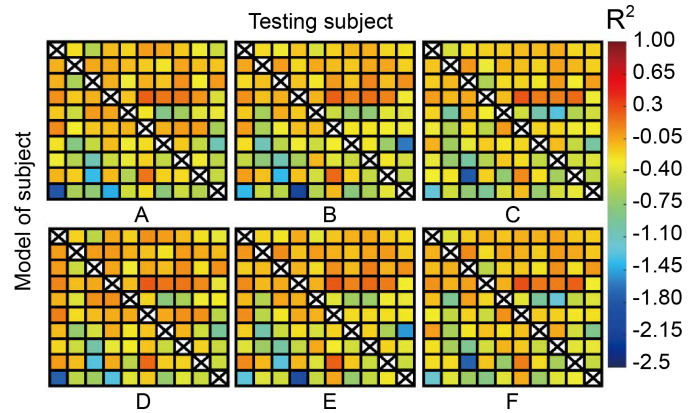| Subject	CNN	CNN-LSTM	Attention-based CNN	CNN with KF	CNN-LSTM with KF	AKFS
1	0.76	0.76	0.77	0.78	0.79	0.80
2	0.80	0.80	0.80	0.82	0.81	0.82
3	0.83	0.83	0.84	0.85	0.84	0.85
4	0.68	0.69	0.71	0.71	0.72	0.73
5	0.62	0.61	0.65	0.66	0.66	0.69
6	0.69	0.70	0.71	0.73	0.73	0.75
7	0.78	0.79	0.79	0.80	0.82	0.82
8	0.66	0.64	0.66	0.70	0.69	0.71
9	0.76	0.79	0.79	0.79	0.81	0.81
10	0.64	0.68	0.69	0.70	0.71	0.73
Average	0.72	0.73	0.74	0.75	0.76	0.77
Std	0.07	0.07	0.07	0.06	0.06	0.05

Fig. 7. Confusion matrices of different human decoding schemes in the inter-subject scenario: (A) CNN, (B) CNN-LSTM, (C) attention-based CNN, (D) CNN with KF, (E) CNN-LSTM with KF, and (F) AKFS.
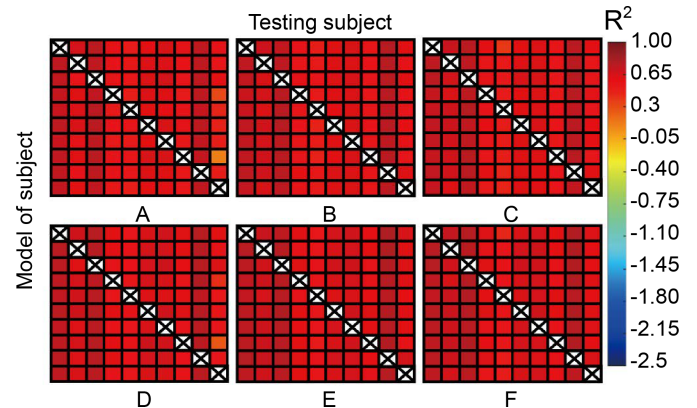

Fig. 8. Confusion matrices of different human decoding schemes in the inter-subject with fine-tuning strategy scenario. (A) CNN, (B) CNN-LSTM, (C) attention-based CNN, (D) CNN with KF, (E) CNN-LSTM with KF, and (F) AKFS.

C. Results on Inter-Subject

The inter-subject confusion matrices of the six schemes are depicted in Fig. 7.

Each small square represents the decoding accuracy (R^2 value) obtained by training (row) and testing (column) based on the first session data from different subjects. The ANOVA analysis showed that the decoding model had an insignificant influence on the accuracy of the decoding schemes ($F_{2,43} = 1.461$, p = 0.243). However, the post-processing strategy had a
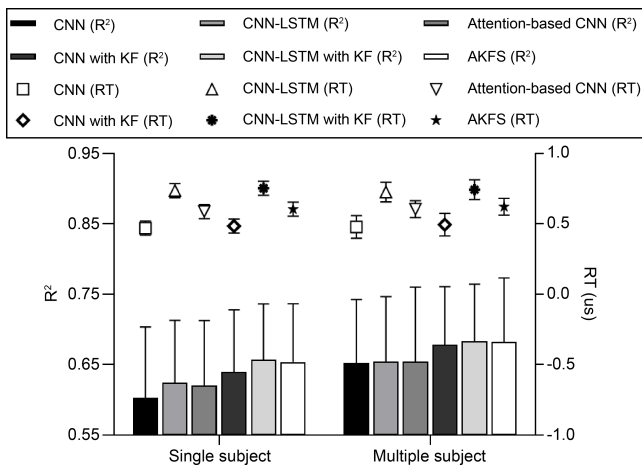
Fig. 9. Average decoding accuracy (R^2) and response time (RT) of different human decoding schemes with fine-tuning strategy in the two different training strategies.

significant influence on the accuracy of the decoding schemes ($F_{1,44} = 1042.700$, $p < 10^{-3}$). A significant interaction between the decoding model and post-processing strategy was observed ($F_{2,43} = 5.309$, $p = 0.009$). A simple effect analysis indicated that the decoding model had an insignificant influence on the accuracy of the schemes with KF (CNN with KF, CNN-LSTM with KF, and AKFS) ($F_{2,43} = 1.865$, $p = 0.167$) or the schemes without KF (CNN, CNN-LSTM, and attention-based CNN) ($F_{2,43} = 1.131$, $p = 0.332$) and post-processing strategy had a significant influence on the accuracy of the CNN models (CNN, and CNN with KF) ($F_{1,44} = 531.145$, $p < 10^{-3}$), the CNN-LSTM models (CNN-LSTM, and CNN-LSTM with KF) ($F_{1,44} = 702.746$, $p < 10^{-3}$) and the attention-based CNN models (attention-based CNN, and AKFS) ($F_{1,44} = 963.357$, $p < 10^{-3}$). The CNN with KF model, the CNN-LSTM with KF model, and AKFS had significantly higher accuracy than the schemes (CNN, CNN-LSTM, and attention-base CNN) without KF ($p < 10^{-3}$).

As previous results show (R^2 lower than 0.11), the accuracy of decoding new subject data using the model trained on the data of a single subject is very low. Therefore, we analyzed the results of the schemes trained with multiple-subject data to determine whether the proposed scheme can learn common features from multiple-subject data and thereby enhance the decoding accuracy of the model for new-subject data.

The results of the schemes trained on multiple subjects are presented in Table III.

An ANOVA analysis showed that the post-processing strategy had a significant influence on the accuracy ($F_{1,9} = 102.265$, $p < 10^{-3}$). However, the decoding model had no significant influence on the accuracy ($F_{2,8} = 1.405$, $p = 0.3$). There was no significant interaction between the decoding model and post-processing strategy ($F_{2,8} = 0.576$, $p = 0.584$). The schemes with KF (CNN with KF, CNN-LSTM with KF, and AKFS) had significantly higher accuracy than the schemes without KF (CNN, CNN-LSTM, and attention-based CNN) ($p < 10^{-3}$).

TABLE II
R^2 OF CNN, CNN-LSTM, ATTENTION-BASED CNN, CNN WITH KF, CNN-LSTM WITH KF, AND AKFS TRAINED ON MULTIPLE SUBJECTS UNDER THE INTRA-SESSION SCENARIO

Subject	CNN	CNN-LSTM	Attention-based CNN	CNN with KF	CNN-LSTM with KF	AKFS
1	0.65	0.67	0.69	0.68	0.7	0.72
2	0.48	0.52	0.54	0.51	0.54	0.57
3	0.65	0.68	0.68	0.67	0.7	0.69
4	0.65	0.66	0.67	0.67	0.69	0.69
5	0.47	0.47	0.5	0.53	0.52	0.54
6	0.64	0.59	0.65	0.68	0.63	0.68
7	0.52	0.56	0.58	0.56	0.59	0.61
8	0.4	0.36	0.41	0.43	0.41	0.45
9	0.67	0.71	0.72	0.71	0.73	0.75
10	0.54	0.57	0.57	0.58	0.61	0.61
Average	0.57	0.58	0.60	0.60	0.61	0.63
Std	0.10	0.11	0.10	0.09	0.10	0.09

TABLE III
R^2 OF CNN, CNN-LSTM, ATTENTION-BASED CNN, CNN WITH KF, 28 CNN-LSTM WITH KF, AND AKFS TRAINED ON MULTIPLE SUBJECTS UNDER 29 THE INTER-SUBJECT SCENARIO

Subject	CNN	CNN-LSTM	Attention-based CNN	CNN with KF	CNN-LSTM with KF	AKFS
1	0.02	0.05	0.04	0.11	0.10	0.09
2	0.05	-0.10	0.02	0.10	-0.06	0.06
3	0.07	0.01	0.08	0.10	0.05	0.13
4	0.04	-0.02	-0.03	0.10	0.06	0.04
5	0.04	0.06	-0.05	0.06	0.10	-0.01
6	0.22	0.16	0.24	0.25	0.22	0.29
7	0.07	0.01	-0.02	0.08	0.04	0
8	0.18	0.01	0.09	0.21	0.06	0.13
9	0.16	0.23	0.18	0.22	0.28	0.23
10	-0.08	-0.04	-0.05	-0.04	0.01	0.01
Average	0.08	0.04	0.05	0.12	0.09	0.10
Std	0.09	0.10	0.10	0.09	0.10	0.10

D. Results on Inter-Subject With a Fine-Tuning Strategy

The confusion matrices of the six inter-subject schemes with a fine-tuning strategy are depicted in Fig. 8. Each small square represents the accuracy (R^2 value) obtained by training (row) and testing (column). The results showed that both the decoding model ($F_{2,43} = 6.157$, $p = 0.004$) and post-processing strategy ($F_{2,43} = 808.327$, $p < 10^{-3}$) had a significant influence on the accuracy. A significant interaction between the decoding model and post-processing strategy was observed ($F_{2,43} = 3.637$, $p = 0.035$).

A simple effect analysis showed that the decoding model had a significant influence on the accuracy of the schemes with KF (CNN with KF, CNN-LSTM with KF, and AKFS) ($F_{2,43} = 6.320$, $p = 0.004$) or the schemes without KF (CNN, CNN-LSTM, and attention-based CNN) ($F_{2,43} = 6.009$, $p = 0.005$). Without KF, both the CNN-LSTM model ($p = 0.005$) and the attention-based CNN ($p = 0.03$) had significantly higher accuracy than the CNN. However, there was no significant difference in accuracy between the CNN-LSTM and the attention-based CNN ($p = 0.216$). With KF, both the CNN-LSTM with KF ($p = 0.004$) and the AKFS ($p = 0.029$) achieved higher accuracy than the CNN. There was no significant difference in accuracy between the CNN-LSTM with KF and the AKFS model ($p = 0.278$).

TABLE IV
R^2 OF CNN, CNN-LSTM, ATTENTION-BASED CNN, CNN WITH KF, CNN-LSTM WITH KF, AND AKFS TRAINED ON MULTIPLE SUBJECTS UNDER THE INTER-SUBJECT WITH A FINE-TUNING STRATEGY SCENARIO

Subject	CNN	CNN-LSTM	Attention-based CNN	CNN with KF	CNN-LSTM with KF	AKFS
1	0.74	0.75	0.75	0.76	0.77	0.77
2	0.69	0.70	0.70	0.71	0.72	0.71
3	0.77	0.76	0.77	0.78	0.77	0.78
4	0.60	0.62	0.62	0.63	0.65	0.65
5	0.53	0.54	0.52	0.56	0.58	0.56
6	0.65	0.63	0.64	0.68	0.66	0.67
7	0.65	0.65	0.66	0.67	0.68	0.69
8	0.62	0.6	0.61	0.66	0.64	0.65
9	0.76	0.78	0.80	0.78	0.80	0.81
10	0.51	0.51	0.47	0.55	0.56	0.53
Average	0.65	0.65	0.65	0.68	0.68	0.68
Std	0.09	0.09	0.11	0.08	0.08	0.09

A simple effect analysis showed that the post-processing strategy had a significant effect on the accuracy of the CNN models (CNN, and CNN with KF) ($F_{1,44} = 362.192$, $p < 10^{-3}$), CNN-LSTM models (CNN-LSTM, and CNN-LSTM with KF) ($F_{1,44} = 853.719$, $p < 10^{-3}$), and attention-based CNN models (attention-based CNN, and AKFS) ($F_{1,44} = 943.348$, $p < 10^{-3}$). The CNN with KF, CNN-LSTM with KF, and AKFS had significantly higher accuracy than the schemes without KF (CNN, CNN-LSTM, and attention-based CNN) ($p < 10^{-3}$).

The results of the schemes trained on multiple subjects with a fine-tuning strategy are presented in Table IV. An ANOVA analysis showed that the post-processing strategy had a significant influence on the accuracy of the decoding schemes ($F_{1,9} = 59.220$, $p < 10^{-3}$). However, the decoding model had no significant influence on the accuracy of the decoding schemes ($F_{2,8} = 0.232$, $p = 0.798$). There was a significant interaction between the decoding model and post-processing strategy ($F_{2,8} = 7.178$, $p = 0.016$).

A simple effect analysis showed that the decoding model had an insignificant influence on the accuracy of the schemes with KF ($F_{2,8} = 0.008$, $p = 0.993$) or without KF ($F_{2,8} = 0.841$, $p = 0.466$) and the post-processing strategy had a significant influence on the accuracy of the CNN models (CNN, and CNN with KF) ($F_{1,9} = 63.334$, $p < 10^{-3}$), the CNN-LSTM models (CNN-LSTM, and CNN-LSTM with KF) ($F_{1,9} = 65.010$, $p < 10^{-3}$) and the attention-based CNN models (attention-based CNN and AKFS) ($F_{1,9} = 44.511$, $p < 10^{-3}$). The CNN with KF, CNN-LSTM with KF, and AKFS had a significant accuracy compared with the schemes without KF (CNN, CNN-LSTM, and attention-based CNN) ($p < 10^{-3}$).

E. Comparison of the Accuracy and Response Time Between the Model Trained by A Single Subject and the Model Trained by Multiple Subjects

We also analyzed the differences between the two training strategies (the schemes trained on a single subject with a fine-tuning strategy and the schemes trained on multiple subjects with a fine-tuning strategy) in the inter-subject scenarios to determine whether the model trained on multiple subjects can learn the generalization features of sEMG signals of different subjects. The average decoding accuracy (R^2) and average response time (RT) of the different schemes in the two different training strategies are shown in Fig. 9.

A repeated measure analysis of variance (ANOVA) with factors including the three decoding models (CNN, CNN-LSTM, and attention-based CNN), two post-processing strategies (with KF and without KF), and two training strategies (the schemes trained on a single subject with a fine-tuning strategy and the schemes trained on multiple subjects with a fine-tuning strategy). An ANOVA analysis showed that the training strategy had a significant influence on the accuracy of the decoding schemes ($F_{1,9} = 16.025$, $p = 0.003$). There was a significant interaction between training strategies and the post-processing strategy ($F_{1,9} = 42.09$, $p < 10^{-3}$).

A simple effect analysis indicated that the training strategy had a significant influence on the accuracy of the schemes with KF (CNN with KF, CNN-LSTM with KF and AKFS) ($F_{1,9} = 17.702$, $p = 0.002$) or without KF (CNN, CNN-LSTM, and attention-based CNN) ($F_{1,9} = 14.236$, $p = 0.004$). Under the schemes with KF (CNN with KF, CNN-LSTM with KF and AKFS) and the schemes without KF (CNN, CNN-LSTM, and attention-based CNN), the schemes trained on a single subject with a fine-tuning strategy had a significantly lower accuracy compared with the schemes trained on multiple subjects with a fine-tuning strategy ($p < 0.05$).

By comparing the response time results, we found that there was no significant difference between the response time in schemes using single-subject data for training and that in schemes using multiple-subject data for training ($p = 0.260$). The response time of the schemes with KF (CNN with KF, CNN-LSTM with KF, and AKFS) is significantly higher than that of the schemes without KF (CNN, CNN-LSTM, and attention-based CNN) ($p < 0.001$). The response time of the schemes with CNN-LSTM (CNN-LSTM and CNN-LSTM with KF) ($p < 0.001$) and the schemes with attention-based CNN (attention-based CNN and AKFS) ($p < 0.001$) is significantly higher than that of the schemes with CNN (CNN and CNN with KF). Increasing the attention mechanism to CNN-LSTM significantly reduces the response time of the model ($p < 0.001$). The response time of all frames should be less than 1 ms.

IV. DISCUSSION

In this study, we proposed a human movement decoding scheme using an attention-based CNN model and a KF to improve the decoding accuracy of a traditional CNN model. An attention-based CNN model was adopted to further improve the spatial information and the temporal information extraction ability of CNN models by capturing the important information between various time series of sEMG signals. A KF was adopted to improve the decoding accuracy of the CNN model by adding prior knowledge of the system. Furthermore, a fine-tuning strategy was adopted to solve the problem of insufficient data due to the short training time for a new subject. To evaluate the effectiveness of the proposed

scheme, we compared the decoding accuracy of the proposed scheme with that of five combined schemes of decoding models and post-processing strategies (CNN, CNN-LSTM, attention-based CNN, CNN with KF, and CNN-LSTM with KF schemes) in four scenarios (intra-session, intra-session long-time use, inter-subject and inter-subject with a fine-tuning strategy).

The attention-based CNN significantly improved the decoding accuracy of CNN models in the intra-session, inter-session long-time use, and inter-subject with fine-tuning strategy scenarios. As reported in previous studies, the hybrid CNN-LSTM model has been extensively used in human movement decoding and achieved satisfactory performance [25], [26] by capturing the temporal and spatial information of sEMG signals. To further improve the ability of the model to use temporal information and improve the decoding accuracy of the model in human upper limb movements, we applied the attention mechanism to CNN-LSTM. Attention-based CNN can extract effective information in the time domain to improve the utilization of the CNN-LSTM time domain information. The attention mechanism can also reduce information redundancy and reduce model calculation time. Thus, the proposed attention-based CNN achieved a higher decoding accuracy than CNN and a lower response time than CNN-LSTM.

Consistent with a previous study on human movement classification [31], the attention mechanism can also improve the performance of the CNN model in human continuous movement decoding. However, previous studies ignored the factors influencing sEMG pattern recognition in real life, such as user changes. In this study, we further extended the attention-based CNN model to the inter-subject scenario. However, its performance was unsatisfactory. This may be a result of the different data distributions of different subjects in the same movement because the data of different subjects have different mean values and variances. Even when we developed an HVS to unify the movement range and normalize the sEMG signals of different subjects, the exercise habits of different subjects still differed.

The models trained with multiple subjects achieved significantly higher performance than those trained with a single subject under the inter-subject scenario. This is because the deep learning method can extract generalization features from multiple subjects and adapt the difference between the distribution of the training and test sets under the inter-subject scenario. However, although the model trained with multiple subjects' data showed improved performance in the inter-subject scenario, the performance of its implementation is still unsatisfactory. There is still a gap between the data distribution of multiple subjects and that of a new subject. Therefore, it was necessary to propose a fine-tuning strategy to improve the decoding accuracy of a specific subject.

The fine-tuning strategy can reduce the gap between the data distribution of different datasets. After fine-tuning with a small amount of data on a new subject, the CNN, CNN-LSTM, and attention-based models achieved higher performance than the models without fine-tuning. Furthermore, both the CNN-LSTM and attention-based CNN models outperformed the CNN model under the inter-subject with fine-tuning strategy scenario. After fine-tuning, all deep learning models can adjust parameters to adapt to the data distribution of the new subject through a small amount of data. Because CNN-LSTM and attention-based CNN can better extract the temporal information and spatial information of the sEMG signal, these two models can achieve a smoother prediction trajectory after fine-tuning compared with the CNN model.

We also evaluated the model trained with multiple-subject data in the context of inter-subject fine-tuning to determine whether the model can learn multiple generalization features to improve the fine-tuning performance. With the increase in training subjects, the gap between the three models became smaller after fine-tuning. Previous studies reported that mixture models can achieve better performance than single-structure models [25], [26]. However, we arrived at a different conclusion in the condition of the model trained on multiple subjects.

Because the CNN model uses data from multiple subjects to obtain generalized initialization parameters, using a small amount of data from the target subjects to fine-tune the model can result in a similar performance to that of the fine-tuned CNN-LSTM and fine-tuned attention-based CNN models. Furthermore, because the model gets more generalized initial parameters from multiple subjects' data, the decoding performance of the model trained with multiple subjects' data is higher than that of the model trained with data from a single subject. The initialization strategy of model parameters has more influence than the difference of model structure.

The proposed KF significantly improved the decoding accuracy of the schemes. Inspired by the post-processing strategies of discrete movement classification based on sEMG, we proposed a post-processing strategy based on KF for continuous movement decoding. In contrast to Bao et al. [36], who focused only on a single subject or a single DoF, we tested not only the effectiveness of KF decoding multiple DoFs for a single subject, but also the effectiveness of KF in decoding multiple DoFs between different subjects. The results indicated that KF significantly improves the decoding accuracy of the schemes in intra-session, intra-session long-time use, and inter-subject with a fine-tuning strategy. Thus, adding prior knowledge to the decoding scheme improves the decoding accuracy of the decoding scheme for inter-subject with multiple DoFs.

V. CONCLUSION

In this study, we proposed a decoding scheme that employs an attention-based CNN model to improve efficiency by using the temporal and spatial information of sEMG signals. In addition, it uses a KF to increase its prior knowledge and thereby improve its decoding accuracy in human movement decoding. The proposed scheme achieved good performance for a single subject under the scenarios of intra-session and intra-session long-time use, and the model could be easily transferred to a new subject through a fine-tuning strategy. Nevertheless, in this study, there was no significant difference in decoding accuracy under the scenario of inter-subject with a fine-tuning strategy

between models trained on multiple subjects due to different data distributions. Therefore, in future work, we will use the domain adaption method to further reduce the data distribution difference between different subjects and further improve the decoding accuracy of the AKFS using the fine-tuning strategy. We will also apply this scheme to an upper limb exoskeleton with multiple DoFs to achieve simultaneous and proportional control.

REFERENCES

[1] F. Lunardini, C. Casellato, A. D'Avella, T. D. Sanger, and A. Pedrocchi, "Robustness and reliability of synergy-based myocontrol of a multiple degree of freedom robotic arm," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 9, pp. 940–950, Sep. 2016, doi: 10.1109/TNSRE.2015.2483375.

[2] J. Chae, Y. Jin, Y. Sung, and K. Cho, "Genetic algorithm-based motion estimation method using orientations and EMGs for robot controls," *Sensors*, vol. 18, no. 1, p. 183, Jan. 2018, doi: 10.3390/s18010183.

[3] Z. Lu, K.-Y. Tong, X. Zhang, S. Li, and P. Zhou, "Myoelectric pattern recognition for controlling a robotic hand: A feasibility study in stroke," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 2, pp. 365–372, Feb. 2019, doi: 10.1109/TBME.2018.2840848.

[4] Z. Feng et al., "iLeg-A lower limb rehabilitation robot: A proof of concept," *IEEE Trans. Hum. Mach. Syst.*, vol. 46, pp. 761–768, 2016.

[5] M. Atzori, M. Cognolato, and H. Müller, "Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands," *Frontiers Neurorobotics*, vol. 10, p. 9, Sep. 2016, doi: 10.3389/fnbot.2016.00009.

[6] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu, and J. Li, "Gesture recognition by instantaneous surface EMG images," *Sci. Rep.*, vol. 6, p. 36571, Nov. 2016, doi: 10.1038/srep36571.

[7] M. Z. Ur Rehman et al., "Multiday EMG-based classification of hand motions with deep learning techniques," *Sensors*, vol. 18, no. 8, p. 2497, 2018, doi: 10.3390/s18082497.

[8] P. Gulati, Q. Hu, and S. F. Atashzar, "Toward deep generalization of peripheral EMG-based human–robot interfacing: A hybrid explainable solution for neurorobotic systems," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 2650–2657, Apr. 2021, doi: 10.1109/LRA.2021.3062320.

[9] A. Waris, I. K. Niazi, M. Jamil, K. Englehart, W. Jensen, and E. N. Kamavuako, "Multiday evaluation of techniques for EMG-based classification of hand motions," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 4, pp. 1526–1534, Jul. 2018, doi: 10.1109/JBHI.2018.2864335.

[10] J. Luo, C. Liu, and C. Yang, "Estimation of EMG-based force using a neural-network-based approach," *IEEE Access*, vol. 7, pp. 64856–64865, 2019, doi: 10.1109/ACCESS.2019.2917300.

[11] M. Houston, A. Wu, and Y. Zhang, "Optimizing input for gesture recognition using convolutional networks on HD-sEMG instantaneous images," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Nov. 2021, pp. 6539–6542.

[12] S. Tam, M. Boukadoum, A. Campeau-Lecours, and B. Gosselin, "Intuitive real-time control strategy for high-density myoelectric hand prosthesis using deep and transfer learning," *Sci. Rep.*, vol. 11, no. 1, p. 11275, May 2021.

[13] A. Ameri, M. A. Akhaee, E. Scheme, and K. Englehart, "A deep transfer learning approach to reducing the effect of electrode shift in EMG pattern recognition-based control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 2, pp. 370–379, Feb. 2020, doi: 10.1109/TNSRE.2019.2962189.

[14] D. Yang and H. Liu, "An EMG-based deep learning approach for multi-DOF wrist movement decoding," *IEEE Trans. Ind. Electron.*, vol. 69, no. 7, pp. 7099–7108, Jul. 2022, doi: 10.1109/TIE.2021.3097666.

[15] E. Rahimian, S. Zabihi, A. Asif, D. Farina, S. F. Atashzar, and A. Mohammadi, "FS-HGR: Few-shot learning for hand gesture recognition via electromyography," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 6, no. 2, pp. 2650–2657, Apr. 2021, doi: 10.1109/TNSRE.2021.3077413.

[16] K.-T. Kim, C. Guan, and S.-W. Lee, "A subject-transfer framework based on single-trial EMG analysis using convolutional neural networks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 1, pp. 94–103, Jan. 2020, doi: 10.1109/TNSRE.2019.2946625.

[17] W.-K. Tam, T. Wu, Q. Zhao, E. Keefer, and Z. Yang, "Human motor decoding from neural signals: A review," *BMC Biomed. Eng.*, vol. 1, no. 1, p. 22, Dec. 2019, doi: 10.1186/s42490-019-0022-z.

[18] J.-L. Ren, Y.-H. Chien, E.-Y. Chia, L.-C. Fu, and J.-S. Lai, "Deep learning based motion prediction for exoskeleton robot control in upper limb rehabilitation," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 5076–5082, doi: 10.1109/ICRA.2019.8794187.

[19] F. Demir, V. Bajaj, M. C. Ince, S. Taran, and A. Şengür, "Surface EMG signals and deep transfer learning-based physical action classification," *Neural Comput. Appl.*, vol. 31, no. 12, pp. 8455–8462, Dec. 2019, doi: 10.1007/s00521-019-04553-7.

[20] A. Gautam, M. Panwar, D. Biswas, and A. Acharyya, "MyoNet: A transfer-learning-based LRCN for lower limb movement recognition and knee joint angle prediction for remote monitoring of rehabilitation progress from sEMG," *IEEE J. Transl. Eng. Health Med.*, vol. 8, 2020, Art. no. 2100310, doi: 10.1109/JTEHM.2020.2972523.

[21] D. Xiong, D. Zhang, X. Zhao, and Y. Zhao, "Deep learning for EMG-based human-machine interaction: A review," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 3, pp. 512–533, Mar. 2021, doi: 10.1109/JAS.2021.1003865.

[22] D. Kim et al., "Simultaneous estimations of joint angle and torque in interactions with environments using EMG," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2020, pp. 3818–3824, doi: 10.1109/ICRA40945.2020.9197441.

[23] Y. Chen et al., "A continuous estimation model of upper limb joint angles by using surface electromyography and deep learning method," *IEEE Access*, vol. 7, pp. 174940–174950, 2019, doi: 10.1109/ACCESS.2019.2956951.

[24] C. Grech, T. Camilleri, and M. Bugeja, "Using neural networks for simultaneous and proportional estimation of upper arm kinematics," in *Proc. 25th Medit. Conf. Control Autom. (MED)*, Jul. 2017, pp. 247–252, doi: 10.1109/MED.2017.7984126.

[25] P. Xia, J. Hu, and Y. Peng, "EMG-based estimation of limb movement using deep learning with recurrent convolutional neural networks," *Artif. Organs*, vol. 42, no. 5, pp. E67–E77, May 2018, doi: 10.1111/aor.13004.

[26] T. Bao, S. A. R. Zaidi, S. Xie, P. Yang, and Z.-Q. Zhang, "A CNN-LSTM hybrid model for wrist kinematics estimation using surface electromyography," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021, doi: 10.1109/TIM.2020.3036654.

[27] V. Ashish et al., "Attention is all you need," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.* San Jose, CA, USA: Curran Associates, 2017, pp. 6000–6010.

[28] X. Cao, M. Iwase, J. Inoue, and E. Maeda, "Gesture recognition based on ConvLSTM-attention implementation of small data sEMG signals," in *Proc. Adjunct Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput. Proc. ACM Int. Symp. Wearable Comput.*, Sep. 2019, pp. 21–24, doi: 10.1145/3341162.3343835.

[29] H. Ye et al., "Attention bidirectional LSTM networks based mime speech recognition using sEMG data," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2020, pp. 3162–3167, doi: 10.1109/SMC42975.2020.9282863.

[30] L. Mou et al., "Driver stress detection via multimodal fusion using attention-based CNN-LSTM," *Exp. Syst. Appl.*, vol. 173, Jul. 2021, Art. no. 114693, doi: 10.1016/j.eswa.2021.114693.

[31] Y. Hu, Y. Wong, W. Wei, Y. Du, M. Kankanhalli, and W. Geng, "A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition," *PLoS ONE*, vol. 13, no. 10, Oct. 2018, Art. no. e0206049, doi: 10.1371/journal.pone.0206049.

[32] K. Englehart and B. Hudgins, "A robust, real-time control scheme for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 7, pp. 848–854, Jul. 2003, doi: 10.1109/TBME.2003.813539.

[33] S. Amsuss, P. M. Goebel, N. Jiang, B. Graimann, L. Paredes, and D. Farina, "Self-correcting pattern recognition system of surface EMG signals for upper limb prosthesis control," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 4, pp. 1167–1176, Apr. 2014, doi: 10.1109/TBME.2013.2296274.

[34] X. Zhang, X. Li, O. W. Samuel, Z. Huang, P. Fang, and G. Li, "Improving the robustness of electromyogram-pattern recognition for prosthetic control by a postprocessing strategy," *Frontiers Neurorobotics*, vol. 11, p. 51, Sep. 2017, doi: 10.3389/fnbot.2017.00051.

[35] J. A. George, T. S. Davis, M. R. Brinton, and G. A. Clark, "Intuitive neuromyoelectric control of a dexterous bionic arm using a modified Kalman filter," *J. Neurosci. Methods*, vol. 330, Jan. 2020, Art. no. 108462, doi: 10.1016/j.jneumeth.2019.108462.

[36] T. Bao, Y. Zhao, S. A. R. Zaidi, S. Xie, P. Yang, and Z. Zhang, "A deep Kalman filter network for hand kinematics estimation using sEMG," *Pattern Recognit. Lett.*, vol. 143, pp. 88–94, Mar. 2021, doi: 10.1016/j.patrec.2021.01.001.