An Extended Spatial Transformer Convolutional Neural Network for Gesture Recognition and Self-calibration Based on Sparse sEMG Electrodes

Wei Chen, Lihui Feng, Jihua Lu, and Bian Wu

***Abstract*—sEMG-based gesture recognition is widely applied in human-machine interaction system by its unique advantages. However, the accuracy of recognition drops significantly as electrodes shift. Besides, in applications such as VR, virtual hands should be shown in reasonable posture by self-calibration. We propose an armband fusing sEMG and IMU with autonomously adjustable gain, and an extended spatial transformer convolutional neural network (EST-CNN) with feature enhanced pretreatment (FEP) to accomplish both gesture recognition and self-calibration via a one-shot processing. Different from anthropogenic calibration methods, spatial transformer layers (STL) in EST-CNN automatically learn the transformation relation, and explicitly express the rotational angle for coarse correction. Due to the shape change of feature pattern as rotational shift, we design the fine tuning layer (FTL) which is able to regulate rotational angle within 45°. By combining STL, FTL and IMU-based posture, EST-CNN is able to calculate non-discretized angle, and achieves high resolution of posture estimation based on sparse sEMG electrodes. Experiments collect frequently-used 3 gestures of 4 subjects in equidistant angles to evaluate EST-CNN. The results under electrodes shift show that the accuracy of gesture recognition is 97.06%, which is 5.81% higher than CNN, the fitness between estimated and true rotational angle is 99.44%.**

***Index Terms*—Self-calibration, spatial transformer, surface electromyography, gesture recognition, robustness.**

This work was supported by the National Natural Science Foundation of China under Grant 61675025. *(Corresponding authors: Lihui Feng and Jihua Lu)*.

Wei Chen and Lihui Feng are with the Laboratory of Photonics Information Technology, Ministry of Industry and InformationTechnology, Beijing Institute of Technology, Beijing 100081, China (e-mail: 3120195296@bit.edu.cn; lihui.feng@bit.edu.cn).

Jihua Lu and Bian Wu are with the School of Integrated Circuits and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: lujihua@bit.edu.cn; 3220210670@bit.edu.cn).

I. Introduction

Gesture recognition, as a human-machine interface (HMI) technique, has been adopted and utilized in numerous fields of biomedical science, such as upper limb and hand rehabilitation, minimally invasive robotic surgery, prosthetic technology, telesurgery navigation, virtual reality (VR) assisted therapy, psychotherapy, and neural interface. Currently, the major technical routes for gesture recognition include computer vision based, inertial sensor based and strain sensor based methods. Every approach of these acquisition and recognition with its own strengths plays a role in different scenarios. However, these approaches can only trace movements of hands and fingers through outward manifestations, and are not suitable for amputees. Gesture recognition based on surface electromyography (sEMG) is capable to perceive part of human ideation by electrodes against the skin around muscles that significantly distinct from aforementioned methods [1]. Consequently, by processing of multi-channel biological signals, sEMG-based method is able to not only recognize gestures, but also provide a theoretical basis for psychological diagnosis [2]. Besides, based on its characteristic of low power consumption, sEMG detection is deployed on wearable devices such as smart watches and armbands, which enable biological monitoring whenever and wherever [3], [4].

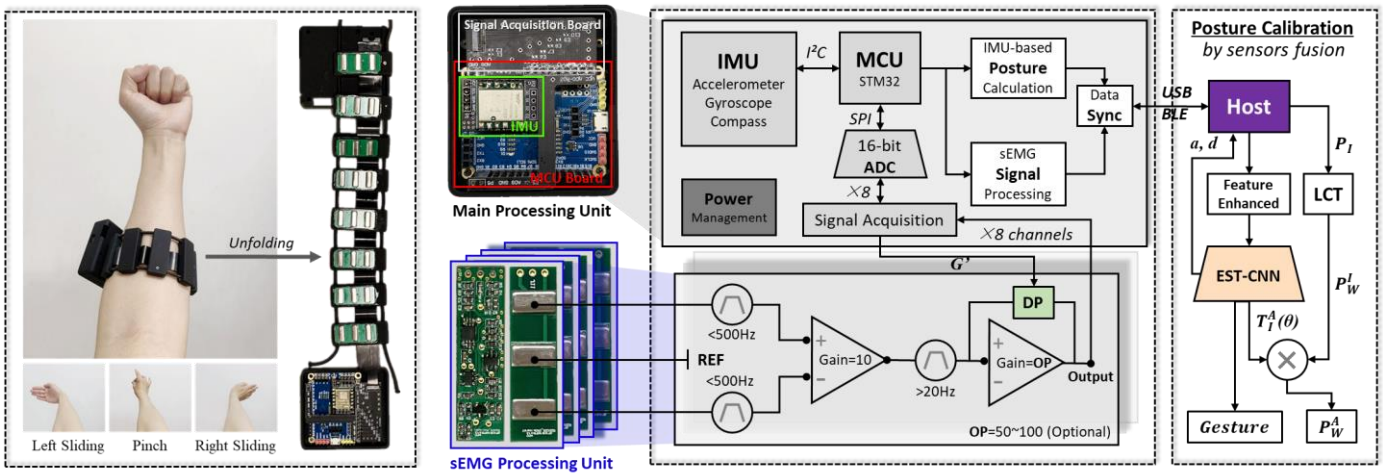
Nevertheless, due to the rotational shift of sEMG armband in practice, matching algorithms of feature patterns show descending accuracy and robustness of gesture recognition. Besides the deviation of rotational angle from wearing factors, the sliding between skin and muscle when rotating forearm brings rotational angle as well. In some interaction scenes such as remote manipulator controlling and VR application [5] [6], the deviation of rotational angles would lead to opposite results. Meanwhile, frequently manual calibrations may bring inconvenience and misoperation which is not beneficial to long-term wearing. Therefore, gesture recognition algorithm endowed with self-calibration is more friendly and reliable to users.

In response to gesture recognition robustness concern and self-calibration issue, investigators employed various machine learning (ML) approaches in recent years. Among them, convolutional neuronal network (CNN) have been employed most extensively [7], [1]. Wei *et al.* proposed a multi-view

CNN framework based on sparse sEMG electrode array. The framework aggregates sEMG feature maps at early and late phases of neural networks, then select the most reliable feature for improving classification accuracy [8]. Further, Wei *et al*. replaced the sparse sEMG electrodes with high density (HD) sEMG electrode array, and presented a divide-and-conquer strategy to increase recognition accuracy by tandem fusing features extracted from multi-stream CNN [9]. Zhang *et al*. considered the dimension of the temporal information and proposed STF-GR that decomposing primitive signals into a series of stationary signals and utilizing RCNN to establish the gesture recognition model [10]. To improve the training efficiency of neural networks, Chen *et al*. applied transfer learning and demonstrated the superiority of combining CNN and LSTM for recognition [11]. Tsinganos *et al*. exploited Hilber curve to characterize the 2-D bioelectrical signals image [12], [13]. Advantages of this approach include making time domain data serialized and increasing the processing efficiency. These approaches optimizing gesture recognition by derived CNN are based on simple sEMG data. Mao *et al.* fused data from accelerometers and sEMG electrodes by 12 detection unit. Though GRNN trained by the fusion dataset, Mao's approach allows continuously motion tracking of fingers [14]. Besides CNN and its variations, many ML-based methods had been adopted to fulfill classification tasks. Cheng *et al*. took rapid spiking neural network (SNN) learning approach, in which the machine is able to save power consumption and support high computing capability for classification [15]. Cote-Allard *et al*. proposed an adversarial neural network and a new dataset to improve the online accuracy of EMG-based gesture recognition [16]. Jaber *et al*. proposed three types of spatial feature sets, and combined histogram oriented gradient (HOG) algorithm and support vector machines (SVM) to achieve advantageous performance [17].

However, only gesture recognition is not enough in practical application. For example, in interactive virtual scenarios of VR therapy, patients need to know the correct posture of gesture and avoid holding props in inconsistent direction. Thus, researchers paid more attentions to the calibration of electrode shift. Li *et al.* proposed shift estimation and adaptive correction (SEAR) method based on activation polar angle (APA). By calculating the mean absolute value (MAV) of bioelectricity electrical data from 8 electrodes, the approach is able to get the APA in polar coordinate system and calibrate the bias of armband rotation, then classify 8 types of gestures by a pre-trained SVM [18]. Hu *et al.* presented an approach which is able to self-calibrate based on the novel conception of muscle core activation regions. Simultaneously the approach establishes hybrid model consisting of CNN and LSTM to classify, and has risen the gesture recognition accuracy by 5.72% [19]. Wu *et al*. proposed a strategy of electromyography enhancement against electrode shift by median filters and interpolation. After data augmentation, the method enables gesture recognition by utilizing dilated convolutional neural network. This method has achieved promising results with accuracy over 95.34% as electrodes shift [20]. Depending on arm's internal conductivity profile and the anatomic principle, Kim *et al.* presented an approach addressing muscle activation source in forearm for recognizing sEMG interface rotation [21], [22]. He *et al.* proposed a novel framework called position identification (PI). As the anchor gesture performed by user, selected optimal classifier achieved position and gesture recognition [23], [24], [25].

In this paper, we proposed a self-developed armband composed of 8 sEMG electrodes and a 9-axis IMU. The armband is able to process sEMG signals on board, and adjusts the signal gain of hardware autonomously for different users. Furthermore, by combining IMU and sEMG data in a loosely coupled way, the armband achieves gesture posture calibration. Besides, an extended spatial transformer convolutional neural network (EST-CNN) is proposed to improve robustness of sEMG-based gesture recognition, and self-calibrate rotational bias simultaneously when armband is worn freely. Due to the rotational, scaling, shearing, and inversional invariance of spatial transformer network (STN) [26], adaptability and reliability of gesture recognition can be optimized. Another characteristic of STN is able to denote affine parameters



(a) (b)

Fig. 1. (a) The diagram of armband and gestures. (b) The structure diagram of self-developed circuit and system.

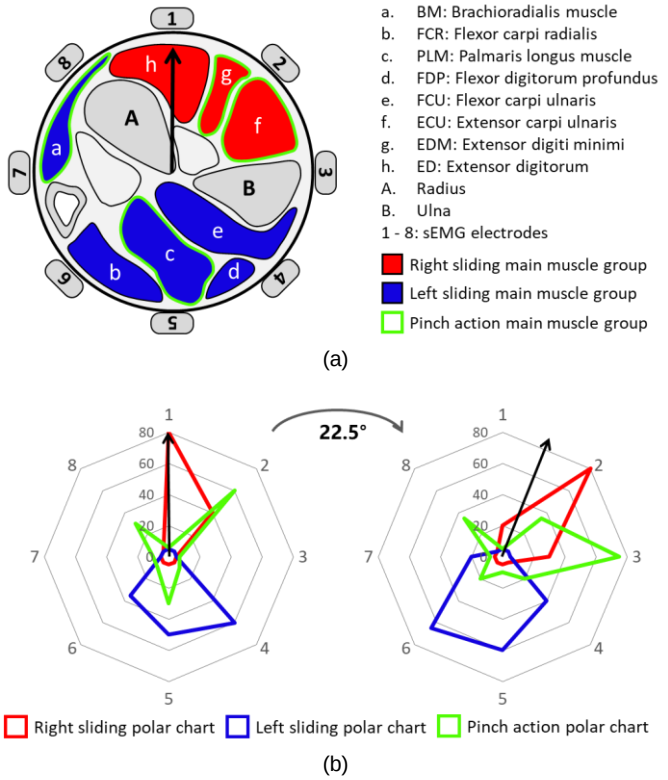


Fig. 2. (a) Electrode and arm cross section. (b) Radar mapping of different gesture in different angles.

explicitly, which can be used for calibrating. However, as shown in Fig. 2, due to the muscle distribution of forearm is separate, the sEMG feature patterns would be different as rotating armband. Therefore, STN cannot be adopted in this case directly. Instead, the EST-CNN combines STN and fine-tuning layer (FTL) together for estimating continuous correcting angle. Additionally, a dataset, including 4 subjects' 3 gestures at 32 rotational positions, is produced for supervised learning.

In conclusion, results under electrodes shift show that the accuracy of gesture recognition is 97.06%, which is 5.81% higher than CNN, the fitness between estimated rotational angle and true value is 99.44%.

The remainder of this paper is structured as follows. First, the process of EST-CNN is elucidated considerable detail in Section II. Section III presents and analyzes experimental results and comparisons. The associated discussion has been taken in Section IV. Finally, the paper is summed up in Section V.

II. Method

To self-calibrate the posture of the armband in the world coordinate system while recognizing gestures, data from the IMU and bioelectric array are required to be fused for processing. The posture of the armband relative to the world coordinate system is obtained through the posture calculation of the 9-axis IMU, and the sEMG electrode array can explicitly obtain the offset angle of the armband through EST-CNN network processing. By combining the two posture conversion relations, the final calibration result can be obtained. For different gestures, muscle groups of forearm generate different intensity of sEMG signals. After signals are acquired by sensing modules, system will pre-process the data firstly, and extract features as the type of radar image in polar coordinates. In the present study, we choose three gestures (as shown in Fig. 1), which are often used in interactive scene (i.e. leftward sliding, right sliding and click). To estimate the continuously rotational angle in this algorithm, we recorded these gestures at specific wearing rotational angles (i.e. 0°, 15° and 30° respect to initial position) as the dataset for EST-CNN training and validation. Among the network, spatial transformer layers (STL) are set to learn the affine relation between rotating feature patterns. The results from ST explicitly express the transformations such as translation, shearing, scaling and rotation by affine matrix with 6 parameters. But the affine matrix of ST layer only shows the rotational relation in one type, the angles between "15°-Click" and "30°-Click" (as shown in Fig. 8) should be estimated by FT layer. By integrating the outputs from ST and FT layers, the algorithm enables gesture recognition and rotational correction simultaneously.

A. Circuit and System

As shown in Fig. 1, the circuit system of the armband mainly consists of 8 sEMG processing units, MCU (Microcontroller Unit) main board, IMU and signal acquisition board.

sEMG processing unit: The sEMG processing unit acquires differential signals through a pair of electrodes based on copper coated with AgCl. According to the characteristics of sEMG signals, such as frequency from 20Hz to 500Hz and voltage ranging from 0.35mV to 1mV, we designed the circuit with adjustable capability. As shown in Fig. 1, sEMG signals conducted through the electrodes are firstly processed by low-pass filters to eliminate part of noise, and then their differential signals are amplified by the instrumentation amplifier at a gain of 10, as well as by a high-pass filter to further extract the effective components. We set the digital potentiometer MAX5439 into the operational amplifier circuit for automatic adjustment of the amplification gain, which helps the consistency and adaptability of the product.

However, the specific value of the gain G' depends on current gain G and the scaling a, d of Eq. (11) fed by the STL in the EST-CNN, as can be seen in section D. (3) of Section II.

$$G' = \frac{a+d}{2} G \tag{1}$$

Main processing unit: After the signal processing in hardware, the sEMG signals are firstly received by the signal acquisition board and processed by the 16-bit high-precision ADC conversion unit. Then, the high speed communication with MCU (STM32 selected in this case) through SPI protocol ensures the processing in real time. At the same time, the STM32 can also adjust the digital potentiometer in reverse to suit different users. In addition, STM32 configures the IMU via I2C bus to set the sampling rate, accuracy and other parameters of each sensor (accelerometer, gyroscope and compass) to solve the IMU postures. Since the refresh rate of the IMU posture

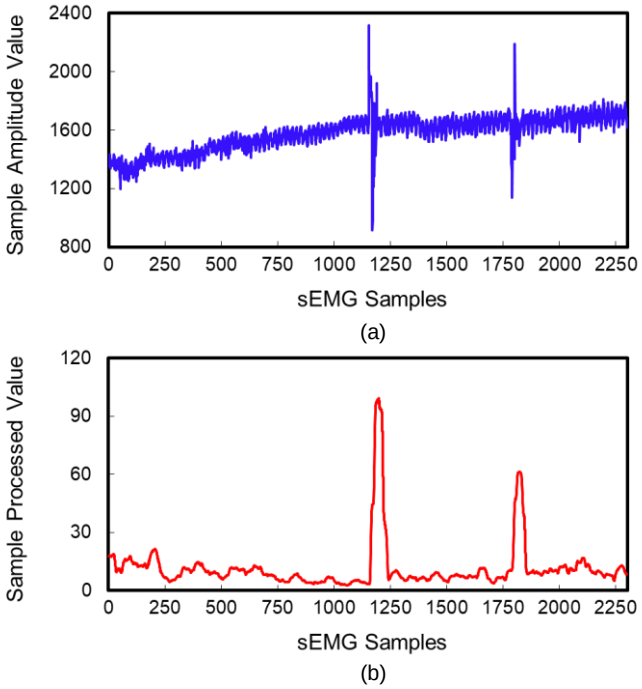


Fig. 3. (a) The raw signal from one of electrodes of sEMG-based armband. (b) The processed signal after pretreatment.

depends on the sampling rate of the magnetometer and the calculation of the processor, and is lower than the sampling rate of the sEMG. In the data synchronization module, we set the sampling rate as 100Hz, and package sEMG and IMU data together for sending to the host.

Posture calibration: Since the posture calculated by IMU, P_I, is based on its own initial coordinate system, it needs to be transferred by linear coordinate transformation (LCT), as T_W^I, to the reference coordinate system as P_W^I. Besides, the sEMG signal needs further data enhancement to form a radar map, which is processed by EST-CNN in host (as shown in Fig. 1) to obtain a rotation matrix $T_I^A(\theta)$ relative to the IMU coordinate system. The final calibrated posture P_w^A is obtained by the product of P_W^I and T_I^A, as in Eq. (3).

$$P_w^I = T_W^I P_I \tag{2}$$
$$P_w^A = T_I^A(\theta) P_W^I \tag{3}$$

The θ in Eq. (3) is the calibration angle between forearm and IMU that acquired from Eq. (15).

B. Biological Basis and Signal Acquisition

As the terminal execution unit of neuromuscular system (NMS), skeletal muscles are precisely regulated by cerebral cortex through descending pathways. Hence machine is capable to understand a part of human intention according to myoelectric signals collection and processing. In NMS, motor neuron, axons, skeletal muscle fibers and neuromuscular junctions make up the motor unit (MU) which is the most fundamental unit of NMS. Among them, the neuromuscular junction (NMJ), as the chemical synapse between muscle fibers and motor neurons, plays a role to convert bioelectrical energy from chemical energy by acetylcholine. Then bioelectrical signals are transferred between volume conductor such as body fluid and fat, finally sensed by electrode on skin. Based on the above theoretical underpinnings of biology, we carried out this study of sEMG-based gesture recognition and self-calibration.

To easily collect the data of sEMG, a scalable armband including 8 sEMG signal processing modules is designed. As shown in Fig. 2, these modules are uniformly appressed to different regions of forearm at initial moment. Every signal processing module sets 3 filters and 2 amplifiers which are regulated to match the amplitude frequency characteristic of sEMG signals (i.e. spectral range is 20-500Hz, voltage amplitude is 0.35-1mV). The customized processing module (shown in Fig. 1) is helpful to save computational power of processor and enhance the readability of sEMG signals.

C. Feature Enhanced Pretreatment

Different from most commonly employed feature extraction approaches which are based on analysis of signal characteristics in the time domain, the frequency domain and the time–frequency domain [27], we acquire 8 channels of sEMG signal to organize radar image on polar coordinates as the feature pattern (Fig. 2). Besides, to obtain a reliable radar pattern, it is essential to process raw sEMG signals by some approaches, such as finite impulse response (FIR) filters, normalization, and envelop calculation. Fig. 3 shows the signal process performance in this case.

Not only processing signals for each sEMG channel independently, but to diminish experimental artefacts, this paper presents the FEP (feature enhancement pretreatment) method by considering 8 channels together. As shown in Fig. 4, the prick at C-7 in the radar map of (a) is diminished, meanwhile the triangle becomes sharper.

Every sample c is represented as an array with 8 sEMG signal measurements as:

$$\mathrm{c} = [c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8] \tag{4}$$

The root mean square (RMS) of c is calculated to evaluate the average level as:

$$c_{RMS} = \sqrt{\sum_{n=1}^{N} |c_n|^2 \Big/ N} \tag{5}$$

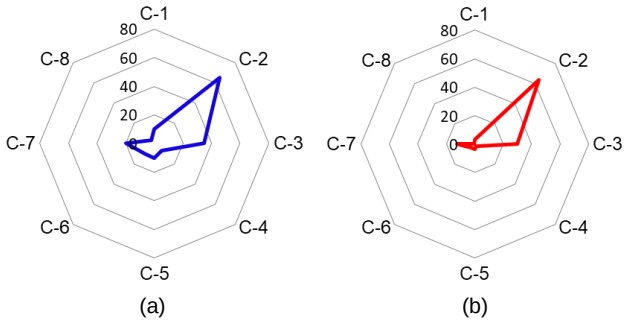


Fig. 4. (a) The raw feature pattern. (b) The augmented feature pattern.

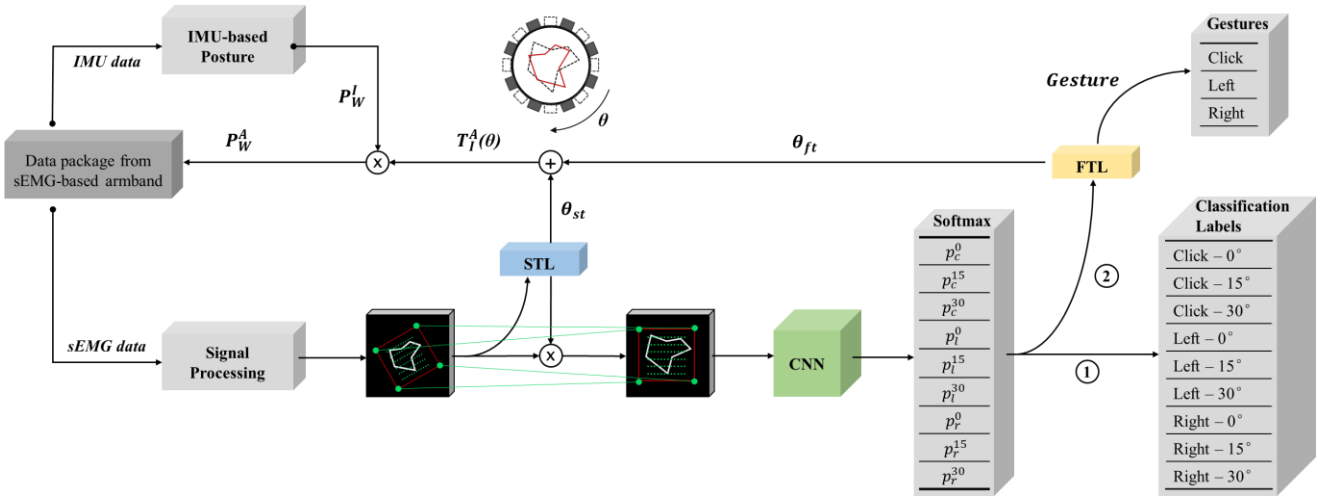


Fig. 5. The standard processing flow of EST-CNN which includes STL, FTL and CNN modules. 1) Training path. 2) Estimation path.

w_n is the weight for regulating the value of channel n.

$$w_n = c_n / c_{RMS} \tag{6}$$

To keep distinct values and diminish the noise, we present a specified sigmoid function to process w_n as:

$$w_n' = 2\left(\frac{1}{1 + e^{-2w_n}} - 1\right) \tag{7}$$

By the array of w' (whose element is w_n'), we get the enhanced sample c_e as:

$$c_e = c \cdot w'^T \tag{8}$$

D. EST-CNN Construction

CNN only has rotational invariance of feature maps in a small range of pattern shift. But in the application of sEMG-based gesture recognition by armband, feature maps always rotate in a large scale and weaken the reliability of recognition. To overcome these challenges, we set ST layers in neural networks to learn the affine relation and correct input feature patterns. After processed by ST layers, Feature maps are much easier classified by CNN, and improve the robustness of gesture recognition. In turn, due to the affine relation is able to be expressed explicitly by STL, we can obtain the rotational relation between input feature patterns and correct feature patterns on polar coordinates. In this approach, the armband coordinate is able to be reset automatically to reference coordinate when gesture is recognized.

Skeletal muscles of forearm act independently. To each of them, the strongest sEMG signal intensity is from the region with the most abundant skeletal muscle fibers, and weakens nonlinearly with fibers lessening. According to this principle, as rotating armband on forearm, the feature pattern cannot remain unchanged, just like patterns at different rotation angles in range from 0 to 45 degrees shown in Fig. 8. Therefore, to solve the limitation of STL which is only able to estimate the rotation between same type feature patterns but the patterns in 45°, we proposed the FTL which is able to analyze the probability of every category for estimating the rotational angle within 45 degrees. To estimate the small rotation, we need a protractor with 3 ticks (at 0°, 15° and 30°). By comparing the similarity between input pattern and the general patterns at 3 ticks, FTL is able to estimate the rotational angle in the protractor.

1) Standard Neural Networks Building

Fig. 5 shows the system flow of the EST-CNN typical structure. However, considering about the performance in different cases, many other possible network structures exist, such as inserting more STLs and CNNs, or adapting the order of layers in a sequence. Before processed in EST-CNN, the signals sampled by sEMG-based armband should be preprocessed to generate feature patterns as described in Section II. C. Next, the EST-CNN would operate in training and estimating modes.

In training mode, STL forward propagates incoming feature maps and outputs parameters of affine matrix for correcting. Additionally, STLs can be putted in different links of networks to transform the posture of feature maps. Then, the corrected feature map would be transmit to CNN to classification. Based on the dataset described in Section II. E, we defined nine labels for NLL (Negative Log-Likelihood) loss-function calculation which is used to regulate the weights of network by back propagation.

In estimation mode, feature maps are imported to EST-CNN and propagated forward directly. Different from the training mode, FTL plays the role of accomplishing gesture classification, and combines STL together to support rotational angle calibration.

2) sEMG-based Gesture Recognition

To recognize the gesture more precisely, FTL sums up the probability value of same gesture from nine classifications. By comparing each sum value of gesture, FTL selects the category with the maximum value as the final gesture. Denote the probability array p_{cat} with nine categories from softmax as:

$$p_{cat} = (p_c^0, p_c^{15}, p_c^{30}, p_l^0, p_l^{15}, p_l^{30}, p_r^0, p_r^{15}, p_r^{30}) \tag{9}$$

where p_c^0, p_l^{15}, and p_r^{30} represents the category probability of gesture "Click", "Left" and "Right" at rotational angle of 0°,

15° and 30°, respectively. The similar expression for others. Then, the probability of each gesture is summed up by Eq. (10), and select the maximum one as gesture by Eq. (11).

$$\begin{cases} p_c = p_c^0 + p_c^{15} + p_c^{30} \\ p_l = p_l^0 + p_l^{15} + p_l^{30} \\ p_r = p_r^0 + p_r^{15} + p_r^{30} \end{cases} \tag{10}$$

$$Gesture = max\{p_c, p_l, p_r\} \tag{11}$$

3) Rotational Angle Correction

Spatial Transform Layer: STL is a kind of neural networks that specifically utilized to process the affine transformation. As shown in Fig. 6, the typical structure of STL is consisted of localization net, grid generator and sampler. It also supports the end-to-end method to train the transformer networks without indicating detail rotational angles. In this case, feature patterns are set up in planar polar coordinates with the center of origin, thus it could barely be translated or scaled in large amplitude. According to this particular situation, we specified the ST layer, and made its output parameters associated with rotation, shearing and scaling. Then, as described in Eq. (12) and (13), the affine matrix would be established by these parameters. What's more, the sampler in STL applies bilinear interpolation for making coordinates differentiable.

$$\begin{bmatrix} x_l \\ y_l \end{bmatrix} = \begin{bmatrix} cos\,\theta_{l-1}^l & sin\,\theta_{l-1}^l & 0 \\ -sin\,\theta_{l-1}^l & cos\,\theta_{l-1}^l & 0 \end{bmatrix} \begin{bmatrix} x_{l-1} \\ y_{l-1} \\ 1 \end{bmatrix} \tag{12}$$

In Eq. (12) (13), (x_l, y_l) is the index number of pixel in feature maps of l layer, (x_{l-1}, y_{l-1}) represents the index in $l-1$ layer. Eq. (12) presents rotational matrix, and θ_{l-1}^l is the rotational angle from layer L-1 to layer L. In Eq. (13), a, d are magnifications of scale, and c, b stand for shearing.

$$\begin{bmatrix} x_l \\ y_l \end{bmatrix} = \begin{bmatrix} a & b & 0 \\ c & d & 0 \end{bmatrix} \begin{bmatrix} x_{l-1} \\ y_{l-1} \\ 1 \end{bmatrix} \tag{13}$$

In EST-CNN, different transformations are adopted for improving performance. Rotational STL is mainly used to estimate rotational angle, and scaling and shearing STLs are applied for correcting the shape of feature pattern.

On the other hand, the scaling size of the pattern also represents the signal amplitude of the sEMG. The system proposed in this paper has the ability of gain adaption of hardware, and the gain value to be adjusted can be obtained according to Eq. (1) in Section II.

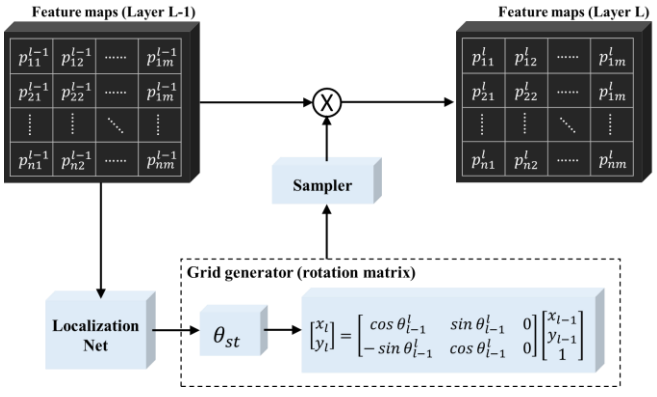


Fig. 6. The flow-process diagram of STL.

Fine Tuning Layer: FTL is utilized to estimate the rotational degree of input feature maps in 45 degrees. As the 8 electrodes of armband divide 360 degrees into 8 equal parts, feature patterns are repetitive at every 45 degrees. However, during rotating in 45 degrees, patterns cannot keep invariant as shown in Fig. 8. In FTL, 45 degrees are divided into 3 parts equally as protractor ticks (i.e. 0°, 15° and 30°). After getting the output of Softmax, we would obtain the probability of each category which is helpful to judge the similarity between input patterns and the patterns on ticks. The higher similarity stands for the closer to degree ticks. Besides, due to the patterns at 0° and 45° are similar, the p^0 represents 0° or 45° in options as shown in Fig. 7. If the estimated angle is in the S1 or S2 area, the probability p^{30} or p^0 would be the minimum, at this moment, p^0 stands for the similarity with the pattern at 0°. However, in the S3 area which is away from the 15° tick, p^0 would represent the similarity with the pattern at 45°. The process of FTL is illustrated by equations below.

$$\theta_{ft} = T(p^0, p^{15}, p^{30}) = \begin{cases} \frac{\pi}{12}p^{15} + \frac{\pi}{6}p^{30} & ,min(p^0, p^{15}, p^{30}) = p^0 \text{ or } p^{30} \\ \frac{\pi}{12}p^{15} + \frac{\pi}{6}p^{30} + \frac{\pi}{4}p^0 & ,min(p^0, p^{15}, p^{30}) = p^{15} \end{cases} \tag{14}$$

The coarse tuning angle by STL is θ_{st}, and the fine tuning angle by FTL is θ_{ft}, the calibration angle θ is:

$$\theta = \theta_{st} + \theta_{ft} \tag{15}$$

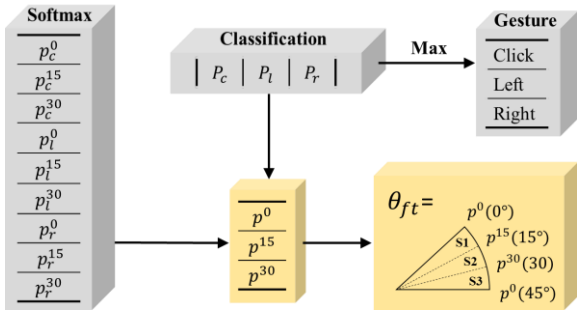


Fig. 7. The flow-process diagram of FTL.

E. Experimental Design

In the experiments designed, we adopt the armband as described previously in Fig.1, to provide biological data to upper host for data processing and ML. The experiments enlisted 4 subjects (2 males and 2 females, maximum girth of forearm is 23.7~27.5cm), who finished a sequence of gestures. By wearing the armband at the region with the most abundant muscle fibers (i.e. maximum girth) of forearm, subjects made a gesture separated by 5 seconds of silence, and hold every gesture for 1 second. We collected about 5 to 10 times of action

This article has been accepted for publication in IEEE Transactions on Biomedical Circuits and Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TBCAS.2022.3222196

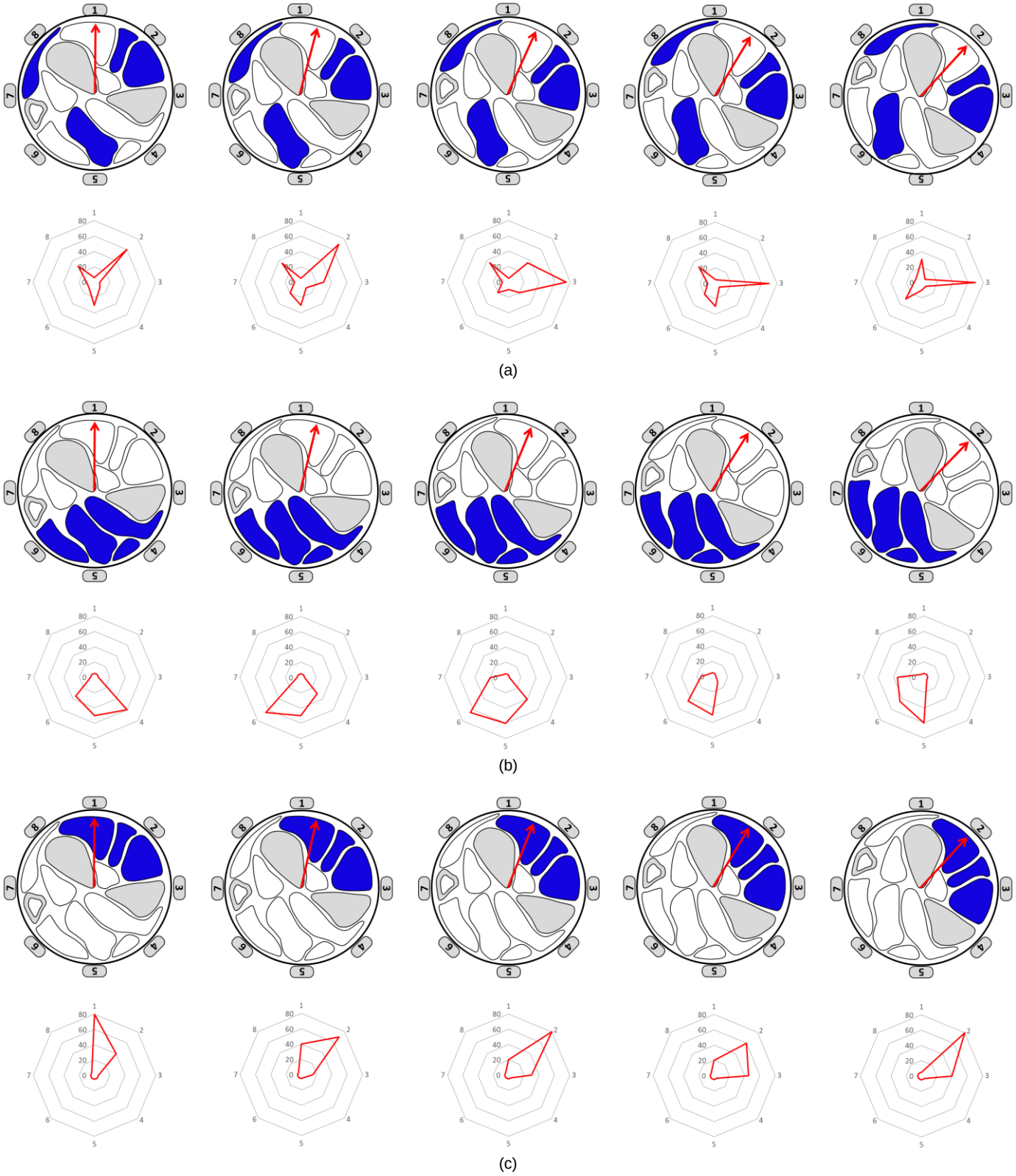


Fig. 8. Feature patterns of 3 gestures (from top to bottom: click, left, right) at 5 rotational angles(from left to right: 0°, 15°, 22.5°, 30°, 45°).

as a group for every classification.

To valid the performance of gesture classification as electrodes shift, we set the armband at rotational angle of 0 degree, 15 degree and 30 degree related to initial position to record 3 gestures as shown in Fig. 8. In other words, there are 9 classifications (i.e. 3 gestures at 3 postures) in the dataset. Subsequently, we selected good data and ignored dirty data from the raw dataset. Based on the selected dataset from 8 sEMG channels, the feature pattern is established on an 8-dimensional polar coordinate system as image file type. After

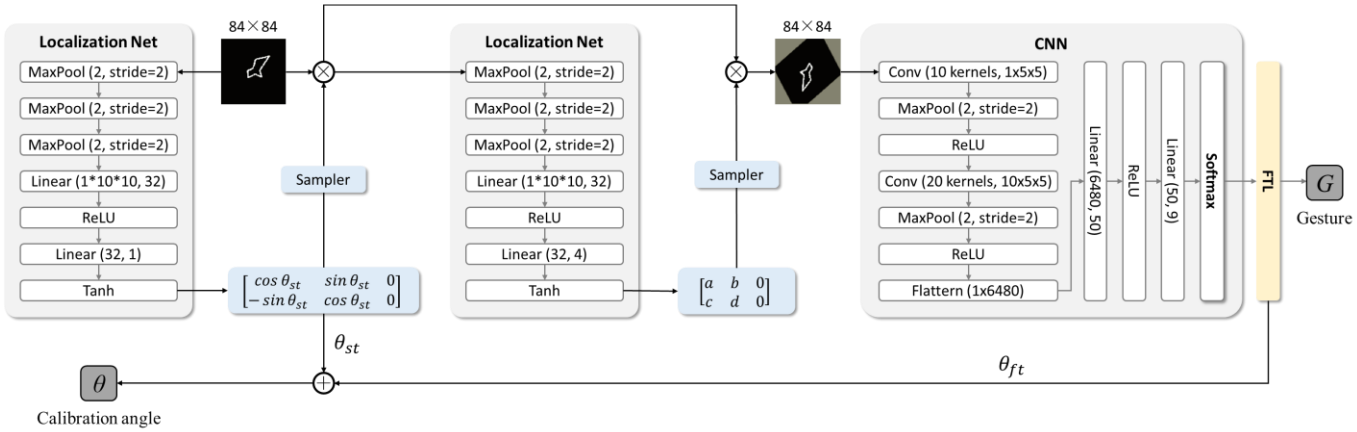


Fig. 9. The network structure of EST-CNN in detail.

that, the upper computer specifically enhanced the selected dataset by rotating patterns in 45 degree for 7 times, and saved the augmented patterns in each time. For machine learning, the augmented dataset was divided into 90% and 10% for training and validation respectively (i.e. 42904 training data and 5303 validation data).

To compare the classification accuracy of different structures of deep neural network, especially including or not STL, we designed an experiment based on our dataset. What's more, we designed experiments to valid the calibration performance by using STL and FTL in estimating rotational angles.

The network of experiments consists of two STL modules as shown in Fig. 9, one FTL module and CNN module with input data size of 1×84×84. The STL and FTL are described in D. (3) of this section, and their standard structures are used in the experiment. In the CNN module, the first convolutional layer uses 10 kernels with kernel size of 1×5×5 and downsampling with maxpool; the second convolutional layer uses 20 kernels with kernel size of 10×5×5 and downsampling with maxpool. Then the flattened data (1×6840) is imported into the fully connected layer (6840, 50) with ReLU activation function. After dropping out some neurons, the data are finally imported into the fully connected layer (50, 9) and classified by softmax.

III. Results

The primary purpose of proposing EST-CNN is to enhance the robustness of gesture recognition and enable rotational self-calibration of sEMG-based armband. We comprehensively evaluate the robustness by testing accuracies under different scenarios, such as armband worn at different postures by different subjects. The accuracy of estimated rotational angle is an essential indicator to evaluate the performance of self-calibration, which is demonstrated in the following experiments.

A. Robustness of Gesture Recognition

To evaluate the robustness of EST-CNN in gesture recognition, we designed experiments to test the performance of armband over two aspects of rotation and worn by different people. In the first experiment, we compared the accuracy for gesture recognition with the datasets including or not rotation data. Besides, to verify the robustness difference between persons, we enrolled 4 subjects' gesture data and compared the accuracies of the three networks. In addition, to intuitively demonstrate the effects of transforming, feature maps before and after STL are shown in Fig. 10. From this figure, although 4 subjects' feature patterns are in different shapes and scales, STLs transform them to similar shape, scale and direction.

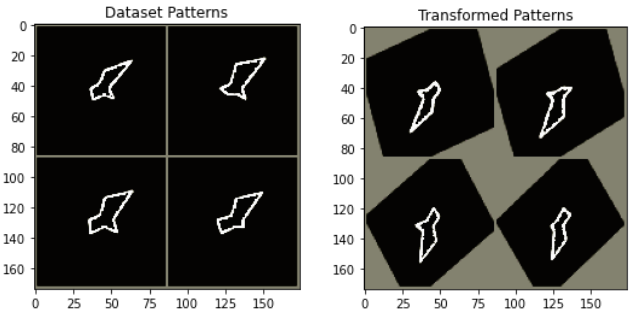


Fig. 10. (a) The raw feature patterns of 4 subjects. (b) The transformed feature patterns.

As Table I shows, three different networks (CNN, EST-CNN with rotation ST, EST-CNN with rotation/scaling/shearing ST) are trained over the same condition. The table shows, in the dataset without rotation, accuracies for recognition are quite close, which means the performance improvement of gesture recognition by EST-CNN is not magnificent. Whereas, as evaluating the dataset with augmented data (data are rotated), EST-CNN exhibits positive effect than CNN, and increasing the success rate of gesture recognition by 5.81% in average. Therefore, we conclude that EST-CNN has better robustness of

TABLE I
ACCURACIES OF NETWORKS FOR THREE GESTURES

Model		**Accuracy (%)** (Original dataset)			**Accuracy (%)** (Augmented dataset)		
		Click	Left	Right	Click	Left	Right
CNN		97.96	98.40	97.72	93.38	89.90	94.85
EST-CNN	Rot ~~Scale~~ ~~Shear~~	98.34	98.03	98.64	97.75	98.12	97.48
EST-CNN	Rot Scale Shear	97.67	98.64	98.79	97.82	97.23	98.13

Three kinds of networks, i.e. CNN, EST-CNN (with rotational matrix), EST-CNN (with rotational matrix, scaling matrix, shearing matrix).

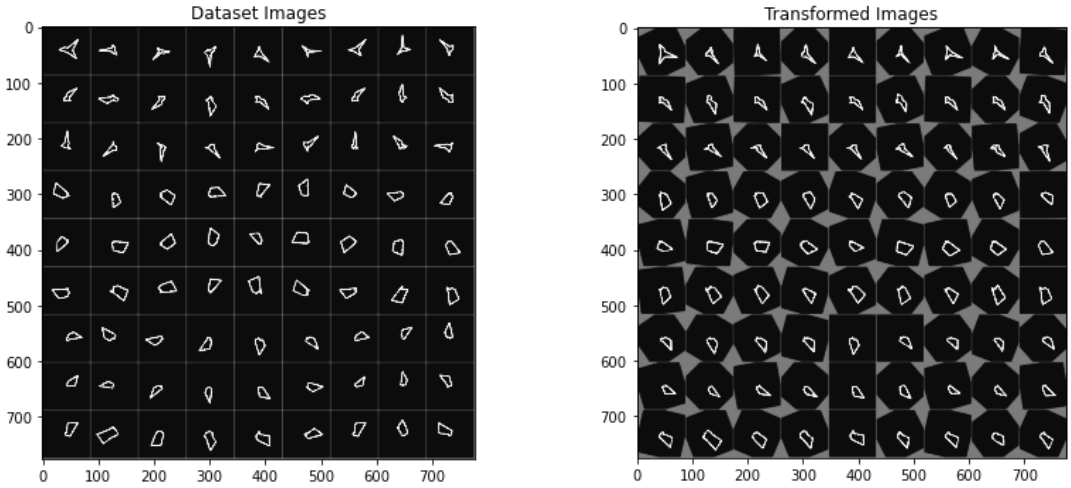


Fig. 11. (a) The patterns of 9 types (9 rows: click-0°, click-15°, click-30°, left-0°, left-15°, left-30°, right-0°, right-15°, right-30°), at 9 rotational angles (9 columns: 0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°, 360°). (b) The transformed patterns by EST-CNN (axis unit: pixel).

TABLE II
ACCURACIES OF NETWORKS FOR FOUR SUBJECTS

Model		**Accuracy (%)** (Augmented dataset)			
		Subject-A	Subject-B	Subject-C	Subject-D
CNN		91.99	91.07	93.83	93.95
EST-CNN	Rot ~~Scale~~ ~~Shear~~	98.11	97.98	96.82	98.22
EST-CNN	Rot Scale Shear	97.82	97.47	98.19	97.43

Three kinds of networks, i.e. CNN, EST-CNN (with rotational matrix), EST-CNN (with rotational matrix, scaling matrix, shearing matrix).

gesture recognition under the scenes of rotational shift between electrodes and muscles.

Table II shows the gesture recognition results of 4 subjects (2 males and 2 females, whose maximum girth of forearm is 23.7~27.5cm). Here, the EST-CNN with scaling and shearing STL performs better than other networks and improves the reliability of classification between persons. Fig. 10 intuitively shows gesture patterns of 4 subjects before and after processing.

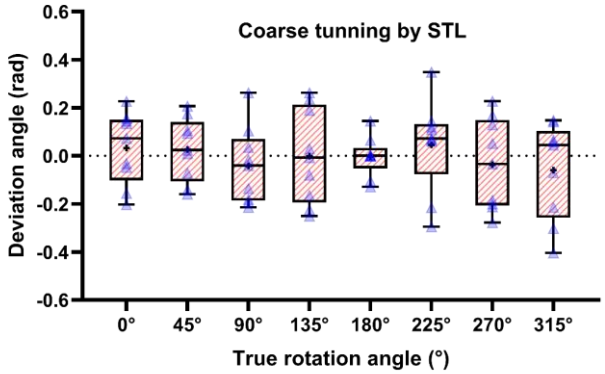


Fig. 12. The deviation angle between estimated and true values by coarse tuning of STL.

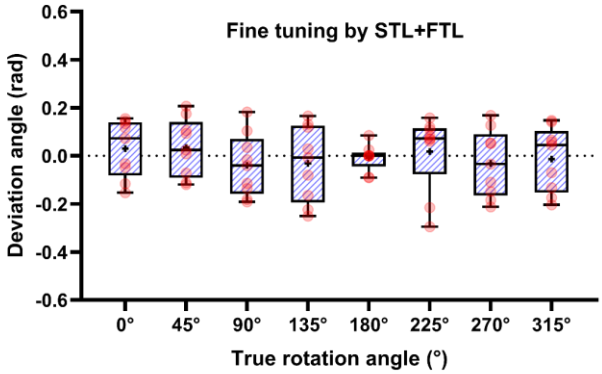


Fig. 13. The deviation angle between estimated and true values by fine tuning of STL+FTL.

We observe that all patterns in the same row are transformed to the same direction.

B. *Accuracy of Angle Calibration*

After training EST-CNN, the feature maps of 9 categories at 8 angles are processed. Fig. 12 shows the deviations of estimated angles by STL and true values, whose intuitive representation of transformation is shown in Fig. 11. Observing from Fig. 11, all the feature patterns are corrected to the same

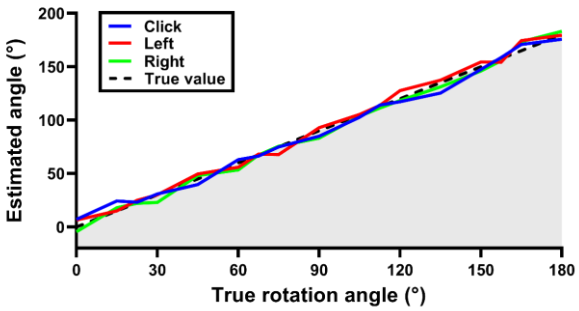


Fig. 14. The estimated rotational angle curves of different gestures.

TABLE III
PERFORMANCES OF DIFFERENT METHODS FOR ELECTRODES SHIFT

Reference	Algorithm	Dataset	Device	Result
Li, Z. Y. *et al.* 2021 [17]	Calibration: APA Classification: SEAR	Subject: 10 Gesture: 8 Shift Position: 9 **(Their self-built dataset)**	Sparse: 8 channels	Accuracy: 79.32% Calibration: -0.017±0.13 rad (Non-discrete)
Hu, R. C, *et al.* 2021 [18]	Calibration: CAR Classification: CNN+LSTM	Subject: 11 Gesture: 9 Shift Position: 3(H) × 5(V) **(Their self-built dataset)**	HD-sEMG: 2×6×8	Accuracy: 94.51±4.56% Calibration: Pixel-based stride (Discrete)
Wu, L. *et al.* 2020 [19]	Calibration: N/A Classification: AUG+DCNN	Subject: 10 Gesture: 6 Shift Position: in 10×10 mm^2 **(Their self-built dataset)**	HD-sEMG: 10×10	Accuracy: 95.34% Calibration: N/A
Kim, M. *et al.* 2018 [27]	Calibration: ASM Classification: NN	Subject: (Not mentioned) Gesture: 6 Shift Position: 8 **(Their self-built dataset)**	Sparse: 8 channels	Accuracy: 95.63% Calibration: 0.22±0.62 rad (Non-discrete)
Chen, W. *et al.* (This method)	Calibration: EST-CNN Classification: EST-CNN	Subject: 4 Gesture: 3 Shift Position: 32 **(Our self-built dataset & CSL-HDEMG)**	Sparse: 8 channels **with Self-adjustment**	Accuracy: 97.06% Calibration:-0.0052±0.063 rad (Non-discrete)

direction, which means the regulation by STL achieves the expected effect of coarse calibration.

Further, FTL plays a role as protractor with angle ticks for the fine tuning of calibrated angle within 45 degrees. Fig. 13 shows the deviations of estimated angles by STL+FTL and true values. In experiments, we sample 4 values (0°, 15°, 22.5° and 30°) in every 45° of 180°, and compare the angle deviation of three gestures between estimated and true values. In Fig. 14, we find that, by FTL tuning, the fitted curves of estimated rotational angle are quite close to the curve of true value whose fitness is 99.44% in average.

IV. DISCUSSION

The proposed system solution includes a signal gain adaptive armband and a self-calibrating recognition algorithm, and achieve mutual adjustment between hardware and software to improve the accuracy of gesture recognition under non-ideal conditions. Spurred by the rise of metaverse, extensive and intensive studies have been carried out to improve the performance of myo-based gesture recognition in recent years. As far as we know, the accuracy of recognition algorithm proposed by these studies can be 96% under ideal conditions. Nevertheless, as rotational shift of armband, the robustness would greatly be worsened. What's more, in VR system, users would have an avatar, and their real hands would be mapped to the virtual world. Thus, when users wear armband upside down, the system should be calibrated to ensure the virtual hand is in right posture. The EST-CNN is able to complete two tasks (including gesture recognition and self-calibration) via one-shot processing by automatically learning the transformation of feature patterns.

Hardware automatic adjustment: Compared to other solutions that mostly use Thalmic Myo armband, digital potentiometers controlled by MCU are added to our hardware system for realizing the gain adjustment of the sEMG processing unit by scale transformation of EST-CNN. In addition, a 9-axis IMU and a high-precision ADC converter are set to enhance the calibration accuracy. To summarize, the software and hardware coupling to adjust each other is the signature of this system.

Two tasks in one-shot: Several methods and models [28] [19] [18] [29] estimate rotational shift by anthropogenic calibration, then recognize gestures by neural networks. However, due to users are multifarious, the calibration parameters set by people are hardly self-adaptive for extensive compatibility. To improve the compatibility for individuals, we prefer to machine learning the transformation between feature patterns by itself, rather than by human intervention. Furthermore, different from state-of-art methods which process correction and classification in two phases, we combine the two phases in a network and process in one-shot to improve the coupling.

Li et al. presented SEAR (Shift Estimation and Adaptive) for electrode shift estimation and adaptive correction which is based on the APA (activation polar angle) [18]. By calculating MAV of every sEMG electrode, APA would be obtained to measure the shift between armband's current and initial angles. Then, process adaptive correction based on Sigmoid and classify by pre-trained SVM. The error of angle estimation is about -0.017±0.13 radians, and the accuracy of 8 gestures recognition is 79.32% in average.

The proposed EST-CNN is an end-to-end model for gesture recognition and feature correction by deploying ST layers which include locational neural networks. The weights of STL, which are used to generate affine matrix, would be regulated while being trained. The trained ST layers not only feedback affine matrixes to virtual system for calibrating virtual hand posture, but also coupling with CNN to improve the feasibility and superiority of gesture recognition. By the presented experiments in Section III, the average accuracy of gesture recognition is about 97.06% as electrodes shift, and the goodness-of-fit between estimated and true values is about 99.44% in average.

Overcoming resolution limit of sparse electrodes: to improve sampling resolutions for high accuracy of shift

calibration, researchers adopted HD-sEMG electrodes to get more detail sEMG signals. However, considering about hardware cost and computing consumption, if sparse electrodes enable to estimate small shifts, the sEMG-based armband would have better market competitiveness.

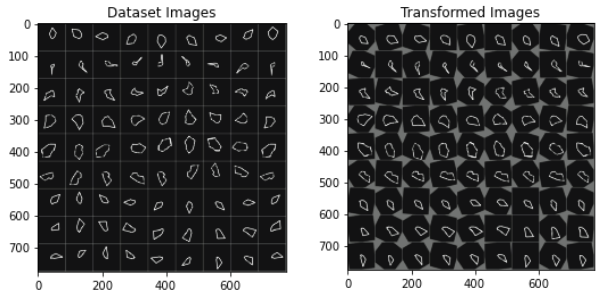
Based on HD-sEMG electrode array, Hu et al. extracted useful signals by FastICA and built signal matrix and parameter matrix. Through analyzing the source signal which has the largest energy to locate the core activation region of muscles. Further, these regions are aligned for correcting electrode array shift. Overall, this approach shows better performance that increasing classification accuracy about 5.72~7.69% after using calibration algorithm.

To provide a non-discretized estimation of rotational shift, the FT layers, as post-treatment of EST-CNN, set 3 ticks to evenly divide 45°. By calculating the similarity between feature maps based on softmax layer, the fine tuning angle would be settled down in the 45° protractor. The experiment results show the error between estimated and true values is further decreased after applying FTL.

Comparison of related methods: As shown in Table III, five gesture recognition methods of anti-electrode shift are compared by several aspects. First, EST-CNN is the only algorithm in the table that can perform both recognition and calibration tasks at one-shot. Second, the self-built dataset collects more data from different positions for better analysis. In addition, the self-developed hardware device has the function of gain self-adjustment, which is more convenient for calibration. Finally, it has better performance in both gesture recognition accuracy and rotational calibration.

Verification of open access dataset: We tried to test EST-CNN with open dataset, limited by the number of electrode columns, it is difficult to simulate 0°, 15°, 30°, and test intermediate states of FTL such as 22.5°, 36°, etc. at the same time. Most of the existing public datasets increase the resolution in the longitudinal direction (rows), such as Ninapro, CapeMyo, and Hyser, while in the transverse direction (columns) most of them increase to 16, which is unable to simulate our rotation. We selected the HD-sEMG dataset with the largest number of electrode columns, CSL-HDEMG. By rotating its 7x24 data array for simulation, we selected patterns of three types of angles at intervals (0°, 15° and 30°, without 22.5°) and rotated them at multiple angles to simulate the rotation.

The CSL-HDEMG will differ from the patterns captured by our hardware due to hardware consistency issues. However, since the features of the three gestures are distinctly different, it does not affect the rotation calibration of EST-CNN for each type of data. Considering the coupling of software and hardware, it is recommended to use the dataset from our self-researched hardware.



(a) (b)

Fig. 15. (a) The patterns of 9 types (9 rows: G14-0°, G14-15°, G14-30°, G21-0°, G21-15°, G21-30°, G23-0°, G23-15°, G23-30°), at 9 rotational angles (9 columns: 0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°, 360°). (b) The transformed patterns by EST-CNN (axis unit: pixel).

Fig. 15 shows the radar plots of the data from CSL-HDEMG and the results of its processing by EST-CNN. The recognition of the three gestures (G14, G21 and G23, which are same gesture as 'click', 'left' and 'right') has an accuracy of 97.68%, higher than self-built dataset. The reason may be that the CSL-HDEMG dataset cannot simulate the baseline drift and gain change from putting on and taking off armband. In addition, the trained EST-CNN has a calibration accuracy of 0.052+0.34 rad for the CSL-HDEMG. However, the public dataset is unable to simulate the intermediate state values, such as 22.5° and 36°, therefore only a limited discussion can be provided here as a reference.

V. Conclusion and Future Work

For upper limb and hand rehabilitation, minimally invasive robotic surgery, prosthetic technology, VR-assisted therapy and neural interface, both good robustness of gesture recognition and self-calibration of gesture posture are required in human-machine interaction system. Based on these abilities, users would not hold virtual props upside down or execute the contrary command. The purpose of this approach is to recognize gestures as wearing with bias, as well as automatically calibrating the gesture posture in virtual interaction system.

First, this paper introduces some novel research in sEMG-based gesture recognition and electrodes shift correction, such as algorithms of CNN based on sparse and HD electrodes, fusion algorithms of sEMG and IMU, and anti-electrodes-shift algorithm. By analyzing these approaches, we propose an armband fusing sEMG and IMU with autonomously adjustable gain and EST-CNN. Then, describe the system process including data flow of hardware, FEP, neuron network building, key transformers (STL and FTL). In the end, we designed experiments to evaluate the robustness of gesture recognition and the accuracy of self-calibration.

However, in the future, we still need to improving algorithm by recognizing more gestures, even though enabling the input by virtual keyboard. Meanwhile, the work of expanding datasets including upper extremity amputations and the nondisabled people is necessary.

Reference

[1] K. Li, J. Zhang, L. Wang, M. Zhang, J. Li, and S. Bao, "A review of the key technologies for sEMG-based human-robot interaction systems," *Biomed. Signal Proces.*, vol. 62, p. 102074, 2020-01-01. 2020.

[2] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Troster, and U. Ehlert, "Discriminating Stress From Cognitive Load Using a Wearable EDA Device," *IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE*, vol. 14, no. 2, pp. 410-417, 2010-01-01. 2010.

[3] S. Tam, M. Boukadoum, A. Campeau-Lecours, and B. Gosselin, "A Fully Embedded Adaptive Real-Time Hand Gesture Classifier Leveraging HD-sEMG and Deep Learning," *IEEE T. Biomed. Circ. S.*, vol. 14, no. 2, pp. 232-243, 2020-01-01. 2020.

[4] J. Y. He, H. Luo, J. Jia, J. Yeow, and N. Jiang, "Wrist and Finger Gesture Recognition With Single-Element Ultrasound Signals: A Comparison With Single-Channel Surface Electromyogram," *IEEE T. Bio.-Med. Eng.*,

vol. 66, no. 5, pp. 1277-1284, 2019-01-01. 2019.

[5] F. Nougarou, A. Campeau-Lecours, D. Massicotte, M. Boukadoum, C. Gosselin, and B. Gosselin, "Pattern recognition based on HD-sEMG spatial features extraction for an efficient proportional control of a robotic arm," *Biomed. Signal Proces.*, vol. 53, p. 101550, 2019-01-01. 2019.

[6] Y. Hu, Y. K. Wong, Q. F. Dai, M. Kankanhalli, W. D. Geng, and X. D. Li, "sEMG-Based Gesture Recognition With Embedded Virtual Hand Poses Adversarial Learning," *IEEE ACCESS*, vol. 7, pp. 104108-104120, 2019-01-01. 2019.

[7] T. Luo *et al.*, "Convolutional Neural Network with Data Augmentation for Robust Myoelectric Control," *2019 IEEE INTERNATIONAL CONFERENCE ON COMPUTATIONAL INTELLIGENCE AND VIRTUAL ENVIRONMENTS FOR MEASUREMENT SYSTEMS AND APPLICATIONS (CIVEMSA 2019)*, 2019, pp. 129-133.

[8] W. T. Wei, Q. F. Dai, Y. K. Wong, Y. Hu, M. Kankanhalli, and W. D. Geng, "Surface-Electromyography-Based Gesture Recognition by Multi-View Deep Learning," *IEEE T. Bio.-Med. Eng.*, vol. 66, no. 10, pp. 2964-2973, 2019-01-01. 2019.

[9] W. Wei, Y. Wong, Y. Du, Y. Hu, M. Kankanhalli, and W. Geng, "A multi-stream convolutional neural network for sEMG-based gesture recognition in muscle-computer interface," *Pattern Recogn. Lett.*, vol. 119, pp. 131-138, 2019-01-01. 2019.

[10] Y. W. Zhang, Y. Q. Chen, H. C. Yu, X. D. Yang, and W. Lu, "Learning Effective Spatial-Temporal Features for sEMG Armband-Based Gesture Recognition," *IEEE INTERNET OF THINGS JOURNAL*, vol. 7, no. 8, pp. 6979-6992, 2020-01-01. 2020.

[11] X. Chen, Y. Li, R. C. Hu, X. Zhang, and X. Chen, "Hand Gesture Recognition based on Surface Electromyography using Convolutional Neural Network with Transfer Learning Method," *IEEE J. Biomed. Health*, vol. 25, no. 4, pp. 1292-1304, 2021-01-01. 2021.

[12] P. Tsinganos, B. Cornelis, J. Cornelis, B. Jansen, and A. Skodras, "A Hilbert Curve Based Representation of sEMG Signals for Gesture Recognition," *PROCEEDINGS OF 2019 INTERNATIONAL CONFERENCE ON SYSTEMS, SIGNALS AND IMAGE PROCESSING (IWSSIP 2019)*, S. Rimacdrlje, D. Zagar, I. Galic, G. Martinovic, D. Vranjes, and M. Habijan, eds., 2019, pp. 201-206.

[13] P. Tsinganos, B. Cornelis, J. Cornelis, B. Jansen, and A. Skodras, "Hilbert sEMG data scanning for hand gesture recognition based on deep learning," *Neural Comput. Appl.*, vol. 33, no. 7, pp. 2645-2666, 2021-01-01. 2021.

[14] H. Mao, P. Fang, and G. L. Li, "Simultaneous estimation of multi-finger forces by surface electromyography and accelerometry signals," *Biomed. Signal Proces.*, vol. 70, 2021-01-01. 2021.

[15] L. Cheng, Y. Liu, Z. G. Hou, M. Tan, D. J. Du, and M. R. Fei, "A Rapid Spiking Neural Network Approach With an Application on Hand Gesture Recognition," *IEEE TRANSACTIONS ON COGNITIVE AND DEVELOPMENTAL SYSTEMS*, vol. 13, no. 1, pp. 151-161, 2021-01-01. 2021.

[16] U. Cote-Allard *et al.*, "A Transferable Adaptive Domain Adversarial Neural Network for Virtual Reality Augmented EMG-Based Gesture Recognition," *IEEE T. Neur. Sys. Reh.*, vol. 29, pp. 546-555, 2021-01-01. 2021.

[17] H. A. Jaber, M. T. Rashid, and L. Fortuna, "Online myoelectric pattern recognition based on hybrid spatial features," *Biomed. Signal Proces.*, vol. 66, 2021-01-01. 2021.

[18] Z. Y. Li, X. G. Zhao, G. J. Liu, B. Zhang, D. H. Zhang, and J. D. Han, "Electrode Shifts Estimation and Adaptive Correction for Improving Robustness of sEMG-Based Recognition," *IEEE J. Biomed. Health*, vol. 25, no. 4, pp. 1101-1110, 2021-01-01. 2021.

[19] R. C. Hu, X. Chen, X. Zhang, and X. Chen, "Adaptive Electrode Calibration Method Based on Muscle Core Activation Regions and Its Application in Myoelectric Pattern Recognition," *IEEE T. Neur. Sys. Reh.*, vol. 29, pp. 11-20, 2021-01-01. 2021.

[20] L. Wu, X. Zhang, K. Wang, X. Chen, and X. Chen, "Improved High-Density Myoelectric Pattern Recognition Control Against Electrode Shift Using Data Augmentation and Dilated Convolutional Neural Network," *IEEE T. Neur. Sys. Reh.*, vol. 28, no. 12, pp. 2637-2646, 2020-01-01. 2020.

[21] M. Kim, W. K. Chung, and J. Kosecka, "Muscle Activation Source Model-based sEMG Signal Decomposition and Recognition of Interface Rotation," *2018 IEEE/RSJ INTERNATIONAL CONFERENCE ON INTELLIGENT ROBOTS AND SYSTEMS (IROS)*, A. A. Maciejewski *et al.*, eds., 2018, pp. 2780-2786.

[22] M. Kim, K. Kim, and W. K. Chung, "Simple and Fast Compensation of sEMG Interface Rotation for Robust Hand Motion Recognition," *IEEE T. Neur. Sys. Reh.*, vol. 26, no. 12, pp. 2397-2406, 2018-01-01. 2018.

[23] J. Y. He, X. J. Sheng, X. Y. Zhu, and N. Jiang, "A Novel Framework Based on Position Verification for Robust Myoelectric Control Against Sensor Shift," *IEEE Sens. J.*, vol. 19, no. 21, pp. 9859-9868, 2019-01-01. 2019.

[24] J. Y. He, M. V. Joshi, J. Chang, and N. Jiang, "Efficient correction of armband rotation for myoelectric-based gesture control interface," *J. Neural Eng.*, vol. 17, no. 3, 2020-01-01. 2020.

[25] J. Y. He, X. J. Sheng, X. Y. Zhu, and N. Jiang, "Position Identification for Robust Myoelectric Control Against Electrode Shift," *IEEE T. Neur. Sys. Reh.*, vol. 28, no. 12, pp. 3121-3128, 2020-01-01. 2020.

[26] J. M, S. K, and Z. A, "Spatial Transformer Networks," *Advances in Neural Information Processing Systems*, pp. 2017-2025. 2015.

[27] J. X. Qi, G. Z. Jiang, G. F. Li, Y. Sun, and B. Tao, "Surface EMG hand gesture recognition system based on PCA and GRNN," *Neural Comput. Appl.*, vol. 32, no. 10, pp. 6343-6351, 2020-01-01. 2020.

[28] J. Kim, B. Koo, Y. Nam, and Y. Kim, "sEMG-Based Hand Posture Recognition Considering Electrode Shift, Feature Vectors, and Posture Groups," *Sensors-Basel*, vol. 21, no. 22, 2021-01-01. 2021.

[29] G. Huang, Z. E. Xian, F. Tang, L. L. Li, L. Zhang, and Z. G. Zhang, "Low-density surface electromyographic patterns under electrode shift: Characterization and NMF-based classification," *Biomed. Signal Proces.*, vol. 59, 2020-01-01. 2020.