PLOS | ONE

RESEARCH ARTICLE

A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition

Yu Hu[1], Yongkang Wong[2], Wentao Wei[1], Yu Du[1], Mohan Kankanhalli[3], Weidong Geng[1]*

1 State Key Lab of CAD&CG, College of Computer Science and Technology, Zhejiang University, Hangzhou, China, **2** Smart Systems Institute, National University of Singapore, Singapore, Singapore, **3** School of Computing, National University of Singapore, Singapore, Singapore

* gengwd@zju.edu.cn

Abstract

The surface electromyography (sEMG)-based gesture recognition with deep learning approach plays an increasingly important role in human-computer interaction. Existing deep learning architectures are mainly based on Convolutional Neural Network (CNN) architecture which captures spatial information of electromyogram signal. Motivated by the sequential nature of electromyogram signal, we propose an attention-based hybrid CNN and RNN (CNN-RNN) architecture to better capture temporal properties of electromyogram signal for gesture recognition problem. Moreover, we present a new sEMG image representation method based on a traditional feature vector which enables deep learning architectures to extract implicit correlations between different channels for sparse multi-channel electromyogram signal. Extensive experiments on five sEMG benchmark databases show that the proposed method outperforms all reported state-of-the-art methods on both sparse multichannel and high-density sEMG databases. To compare with the existing works, we set the window length to 200ms for NinaProDB1 and NinaProDB2, and 150ms for BioPatRec subdatabase, CapgMyo sub-database, and csl-hdemg databases. The recognition accuracies of the aforementioned benchmark databases are 87.0%, 82.2%, 94.1%, 99.7% and 94.5%, which are 9.2%, 3.5%, 1.2%, 0.2% and 5.2% higher than the state-of-the-art performance, respectively.

Citation: Hu Y, Wong Y, Wei W, Du Y, Kankanhalli M, Geng W (2018) A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition. PLoS ONE 13(10): e0206049. https://doi.org/10.1371/journal.pone.0206049

Editor: Huiguang He, Chinese Academy of Sciences, CHINA

Received: April 11, 2018

Accepted: October 7, 2018

Published: October 30, 2018

Data Availability Statement: Five benchmark databases used in this manuscript can be downloaded as follows: 1) NinaProDB1 & NinaProDB2, Official Website: http://ninaweb.hevs.ch/, Dataset Download URL: https://datadryad.org/resource/doi:10.5061/dryad.1k84r, Related Paper DOI: https://doi.org/10.1038/sdata.2014.53 2) BioPatRec26MOV, Official Website: https://github.com/biopatrec/biopatrec, Dataset Download URL: https://github.com/biopatrec/biopatrec/tree/Data_Repository/6mov8chFUS, Related Paper DOI: https://doi.org/10.1186/1751-0473-8-11 3) CapgMyo-DBa, Official Website: http://zju-capg.

Introduction

The surface electromyogram signal [1] records muscle's information by putting non-invasive surface sEMG electrodes on the skin. The electrical activity recorded by sEMG electrodes allows us to develop human-computer interface (HCI) system which has been employed in four major areas [2]: **(1)** Assistive technology (e.g., myoelectric controlled prosthesis [3], wheelchair [4] and assistive robots [5]), **(2)** Rehabilitative technology (e.g., sEMG-driven Exoskeletons [6] and serious games [7, 8]), **(3)** Input technology (e.g., armbands and MCI [9]), and **(4)** Silent speech recognition [10]. Among these applications, sEMG-based hand gesture

org/myo/, Dataset Download URL: https://figshare.com/articles/Data_from_Gesture_Recognition_by_Instantaneous_Surface_EMG_Images_CapgMyo-DBa/7210397, Related Paper DOI: https://doi.org/10.1038/srep36571 4)csl-hdemg, Official Website: https://www.uni-bremen.de/en/csl/research/motion-recognition.html, Dataset Download URL: http://www.csl.uni-bremen.de/CorpusData/download.php?crps=cslhdemg, Related Paper DOI: https://doi.org/10.1145/2702123.2702501.

Funding: This work was supported by National Key Research and Development Program of China (No. 2016YFB1001302), the National Natural Science Foundation of China (No. 61379067), and the National Research Foundation, Prime Ministers office, Singapore under its International Research Centre in Singapore Funding Initiative.

Competing interests: The authors have declared that no competing interests exist.

recognition plays an important and fundamental role for computers or assistive devices to understand human body language.

The traditional sEMG-based gesture recognition framework consists of data preprocessing, feature extraction, feature selection and gesture classification. Among these stages, feature extraction and gesture classification are two important stages in sEMG-based gesture recognition framework. Therefore, researchers have focused on presenting discriminative feature sets with domain knowledge [11–15], as well as employing conventional machine learning algorithms to classify hand gestures [16–19]. These works often require excessive parameter tuning and rich domain knowledge.

In recent years, deep learning techniques achieve promising performance in various fields [20–23] and provide a new perspective to analyze sEMG for hand gestures recognition. Inspired by the excellent performance of deep learning techniques, the Convolutional Neural Network (CNN) has been exploited for sEMG-based gesture recognition [24–29]. Park and Lee [24] proposed a CNN model with adaptive feature learning to improve the inter-subject accuracy. The CNN-based sEMG gesture recognition was studied in [25] which achieved comparable performance with traditional methods on the NinaPro database. Geng *et al.* [26, 27] presented a new CNN architecture for instantaneous sEMG images and the recognition accuracies are 77.8% for 52 gestures, 99.5% for 8 gestures and 89.3% for 27 gestures on three sEMG benchmark databases. Du *et al.* [28] designed a semi-supervised deep CNN framework which employed data glove to provide auxiliary information.

Overall speaking, existing deep learning methods for sEMG-based gesture recognition are mainly based on CNN architecture. However, the sEMG is a form of sequential data by its nature. In the field of video classification and human activity analysis, the hybrid CNN-RNN architecture has obtained good performance when compared with pure CNN-based approaches [30–33]. Aiming at modeling the temporal information better than conventional CNN-based architectures, we investigate a hybrid CNN-RNN architecture for sEMG-based gesture recognition to capture both spatial and temporal information. Moreover, attention mechanism has been applied to the proposed hybrid CNN-RNN architecture for it has proven successful in sequential data modeling (e.g., machine translation [34], image caption generation [35] and speech recognition [36]).

The main contributions of this work are twofold:

1. We propose an attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition, which models both the spatial and temporal information of sEMG and focuses on the subsegments which contain more discriminative information for gesture recognition. Compared with the CNN module of proposed architecture which only models the spatial information, it improves the recognition accuracies from 83.5% to 84.8% on the first sub-database of NinaPro (denoted as NinaProDB1), from 73.4% to 74.8% on the second sub-database of NinaPro (denoted as NinaProDB2), from 83.9% to 92.5% on a sub-database of BioPatRec (denoted as BioPatRec26MOV), from 97.7% to 99.7% on a sub-database of CapgMyo (denoted as CapgMyo-DBa) and from 92.1% to 94.9% on the csl-hdemg, respectively.

2. Motivated by the idea of signal image [37], we present a new sEMG image representation method based on a feature vector for sparse multi-channel electromyogram signals. It rearranges feature vectors based on the classical feature set Phinyomark [13], and is evaluated on three sparse multi-channel sEMG databases using the CNN, hybrid CNN-RNN and attention-based hybrid CNN-RNN frameworks, respectively. The results show that the proposed sEMG image representation method is superior to the existing sEMG image

representation method. The improvements are at least 2.0% for NinaProDB1, 7.4% for NinaProDB2 and 1.6% for BioPatRec26MOV.

The remainder of the paper is organized as follows. Firstly, we review related works on sEMG-based gesture recognition methods, hybrid CNN and RNN architectures and the attention mechanism. Secondly, we introduce the proposed attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition, and describe the details of the new feature vector based sEMG image representation methods. Thirdly, we show the experimental results on five benchmark sEMG databases. Finally, we draw the conclusion and discuss our future work.

Related work

In this section, we present related works on sEMG-based gesture recognition methods, the hybrid CNN-RNN architectures and attention mechanism in the literature.

The handcrafted features and traditional machine learning classifiers have been extensively researched in early sEMG-based gesture recognition frameworks. Existing sEMG-based handcrafted features can be divided into three categories [38]: time domain, frequency domain, and time-frequency domain features. Many researchers focused on presenting new sEMG features based on their domain knowledge [14, 15] or analyzing existing features to propose new feature sets [13, 39]. Traditional machine learning classifiers have been employed to recognize sEMG-based gestures, such as k-Nearest Neighbor (kNN) [16], Linear Discriminate Analysis (LDA) [17, 40, 41], Hidden Markov Model (HMM) [18, 42], and Support Vector Machine (SVM) [14, 19]. The SVM is the most popular classifier in early sEMG-based gesture recognition frameworks. Patricia *et al.* [19] utilized Geodesic Flow Kernel with SVM classifier to classify 10 gestures. Doswald *et al.* [14] applied SVM classifier with Pearson VII Universal Kernel to recognize 5 gestures. As electromyogram signal is sequential data, HMM is suitable for modeling electromyogram signal with hidden information. Yun *et al.* [42] used HMM classifier to create a sign language recognition system based on sEMG.

The CNN architecture is the most widely used deep learning technique for sEMG-based gesture recognition, which can be divided into two categories based on different evaluation methods. The first study focuses on improving recognition accuracy of intra-session evaluation [25, 26]. Atzori *et al.* [25] constructed sEMG images which contain both spatial and temporal information and trained a CNN model to extract high-level features. Geng *et al.* [26] provided a novel CNN model to extract spatial information from the instantaneous sEMG images and achieved state-of-the-art performance. The second study is devoted to the difference between sessions or subjects [24, 27]. Park and Lee [24] draw adaptation method into CNN model to learn better features for inter-subject evaluation. Du *et al.* [27] applied domain adaptation based on the GengNet [26] to improve the inter-session accuracy. Zhai *et al.* [29] extracted useful information from the sEMG spectrogram to form sEMG images and a CNN-based architecture was employed to model the relationship between sEMG images and gesture labels.

The RNN architecture has been applied for sEMG-based hand problems, such as pose estimation [43, 44] and sEMG feature extraction [45, 46]. Hioki and Kawasaki [43] presented a neural network with recurrent structure to estimate finger joint angles using sEMG signal. Quivira *et al.* [44] proposed a sEMG-based hand pose estimation method using RNN with LSTM cells. It constructed a model for predicting the hand joint kinematics through sEMG signals and captured hand pose kinematics accurately. Amor *et al.* [45] applied Myo armband to collect sEMG signals for sign language recognition and employed the RNN architecture to extract features from sequential data for analyzing sign language gestures. Shin *et al.* [46]

exploited an RNN architecture with three LSTM layers to extract features from sEMG and Inertial Measurement Unit (IMU) signals for Korean sign recognition.

The hybrid CNN-RNN architecture has obtained good performance in recognition of video and wearable sensors. Ebrahimi Kahou *et al.* [30] presented a hybrid CNN-RNN architecture for facial expression analysis. The CNN and RNN module of the architecture were trained separately. Wu *et al.* [31] designed a hybrid deep learning framework to extract spatial, short-term and long-term features which consist of two hybrid CNN-RNN architecture and a regularized fusion layer. Ordóñez and Roggen [32] proposed a deep hybrid CNN-RNN for activity recognition with multimodal wearable sensors. This architecture provided a natural sensor fusion and modeled temporal information of the activity. Wang *et al.* [33] recommended a novel CNN-LSTM model to solve both gesture recognition and pose estimation problem with only RGB videos. The CNN block was employed to extract spatial features from each frame, and the proposed sequentially supervised LSTM (SS-LSTM) used auxiliary knowledge instead of the class label to supervise learning process.

The attention mechanism has been injected into RNN architectures for performance enhancement in many application scenarios. Dzmitry *et al.* [34] proposed an RNN encoder-decoder model with attention for machine translation. It jointly learned for alignment and translation and achieved significant performance improvement over the basic encoder-decoder method. Kelvin *et al.* [35] introduced two attention-based image caption generators and the results on three benchmark databases showed the effectiveness of attention. Chorowski *et al.* [36] presented a novel attention-based neural speech recognition architecture. The performance was comparable to that of the traditional methods on the TIMIT dataset. Song *et al.* [47] provided an end-to-end LSTM network with spatial and temporal attention modules for skeleton-based human action recognition. The recognition accuracies on two benchmark databases outperformed other state-of-the-art methods.

As mentioned above, the hybrid CNN-RNN architecture has been successfully applied to activity recognition based on video and wearable sensors. The attention mechanism is also an effective way to enhance the performance of RNN architecture.

Since the electromyogram signal is noisier than other wearable sensor signals [32], we extract the Phinyomark feature set [13] of each channel to generate new sEMG images and employ deep neural network to extract useful information between each channel. However, the generated sEMG image is a monochrome image and has much smaller pixel resolutions than normal images or video frames. We carefully fine-tune parameters of each layer and add locally-connected layer [26] to our proposed attention-based hybrid CNN-RNN architecture which has been applied in sEMG-based gesture recognition for the first time.

Methods

Attention-based hybrid CNN-RNN architecture

The attention mechanism has been proposed in deep learning to learn from the way a human perceives the real-world that paying attention to different regions [48]. The motor unit action potential (MUAP) generates and propagates along the muscle fibers [26, 49] and muscles have varying importance in contributing to different hand movements [50]. If the learned classification model can effectively capture these important factors of the involved muscles, it may bring performance improvements on sEMG-based gesture recognition. The attention framework in deep learning usually models the importance inside the training data through weights and has been successfully applied to various tasks, such as image caption generation [35], speech recognition [36], sentiment analysis [51] and *etc*. Therefore, we focus on how to embed the attention mechanism into the classification model and

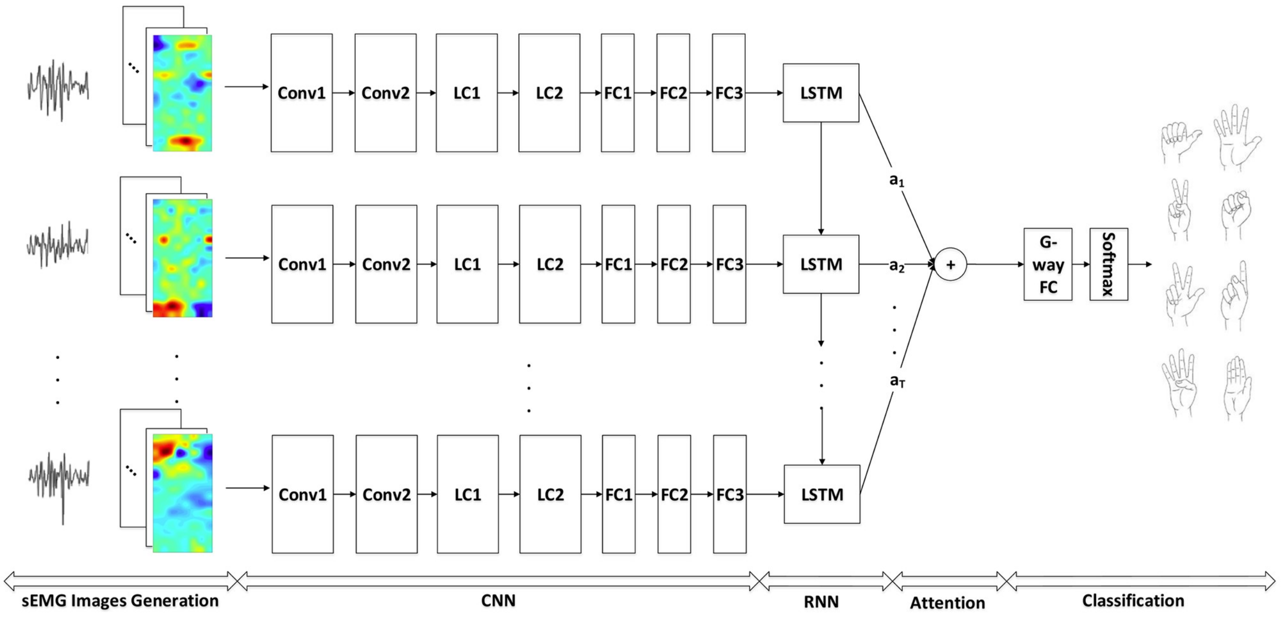

Fig 1. Proposed attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition.

https://doi.org/10.1371/journal.pone.0206049.g001

accordingly propose a novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition (see Fig 1).

The proposed architecture models both the spatial-temporal information and importance of the input electromyogram signals and the layers configuration of the proposed network are shown in Table 1.

Each sample recorded from C electrodes with L frames is denoted as $\boldsymbol{X}$. We first use sliding window method to split $\boldsymbol{X}$ into subsegments which are denoted as $\{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_T\}$, where T ($T \leq L$) is the number of subsegments and also time steps of RNN. Each subsegment $\boldsymbol{X}_t$, $\forall t = 1, 2, \ldots, T$ has N frames with the size of $1 \times C$. Then, the $\boldsymbol{X}_t$ is converted into an image $\boldsymbol{I}_t$ with size $(N \times W \times H)$, where $W \times H = C$, W and H are width and height of the image, respectively. The detailed image representation method can be seen in section. Given the converted images $\{\boldsymbol{I}_1, \boldsymbol{I}_2, \ldots, \boldsymbol{I}_T\}$, CNN is applied as feature extractor to transform them into feature

Table 1. The layers configuration of proposed attention-based hybrid CNN-RNN.

Layers	Name	Configurations	Modules
1	**Conv1**	64 kernels, kernel size (3 × 3)	CNN
2	**Conv2**	64 kernels, kernel size (3 × 3)	
3	**LC1**	64 kernels	
4	**LC2**	64 kernels	
5	**FC1**	512 outputs	
6	**FC2**	512 outputs	
7	**FC3**	128 outputs	
8	**LSTM1**	LSTM, 512 hidden unit outputs	RNN
9	**Attention**		Attention
10	**G-way FC**		Classification
11	**Softmax**		

https://doi.org/10.1371/journal.pone.0206049.t001

vectors $\{\boldsymbol{F_1}, \boldsymbol{F_2}, \ldots, \boldsymbol{F_T}\}$. The CNN model consists of seven layers. The first two layers are convolutional layers with 64 3 × 3 kernels, followed by two locally-connected layers. The locally-connected layer with 64 1 × 1 kernels is employed to extract local features of the sEMG image. Batch normalization [52] is used for each of the layers mentioned before reducing the internal covariate shift. The last three layers are all fully-connected layers with batch normalization, and a dropout with the probability of 0.5 is applied to the first two fully-connected layers. For the sequence modeling stage, each RNN unit has a dropout with the probability of 0.5 and 512 hidden units, followed by a G-way fully-connected layer and a softmax classifier. G is the number of gestures to be recognized. The final label is decided by average-pooling of the softmax outputs.

The recurrent neural network contains feedback loops and encodes contextual information of a temporal sequence. Given the input sequence $\{\boldsymbol{F_1}, \boldsymbol{F_2}, \ldots, \boldsymbol{F_T}\}$ (feature vectors extracted from CNN model), the hidden states $\boldsymbol{h_t}$ and outputs $\boldsymbol{y_t}$ can be calculated as follows:

$$\begin{aligned} \boldsymbol{h}_t &= H(\boldsymbol{W}_{ih}\boldsymbol{F}_t + \boldsymbol{W}_{hh}\boldsymbol{h}_{t-1} + \boldsymbol{b}_h) \\ \boldsymbol{y}_t &= \boldsymbol{W}_{ho}\boldsymbol{h}_t + \boldsymbol{b}_o \end{aligned} \tag{1}$$

where $\boldsymbol{W_{ih}}, \boldsymbol{W_{hh}}, \boldsymbol{W_{ho}}$ are weight matrices between input, hidden and output layers. As standard RNN suffers from gradient vanishing or exploding problem, long short-term memory (LSTM) [53] has been proposed to alleviate this issue. Each LSTM unit consists of input gate, output gate, forget gate and cell, and the calculating relations between them are as follows:

$$\begin{aligned} \boldsymbol{i}_t &= \delta(\boldsymbol{W}_i[\boldsymbol{h}_{t-1}, \boldsymbol{F}_t] + \boldsymbol{b}_i) \\ \boldsymbol{f}_t &= \delta(\boldsymbol{W}_f[\boldsymbol{h}_{t-1}, \boldsymbol{F}_t] + \boldsymbol{b}_f) \\ \boldsymbol{o}_t &= \delta(\boldsymbol{W}_o[\boldsymbol{h}_{t-1}, \boldsymbol{F}_t] + \boldsymbol{b}_o) \\ \hat{\boldsymbol{c}}_t &= tanh(\boldsymbol{W}_c[\boldsymbol{h}_{t-1}, \boldsymbol{F}_t] + \boldsymbol{b}_c) \\ \boldsymbol{c}_t &= \boldsymbol{f}_t \odot \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \odot \hat{\boldsymbol{c}}_t \\ \boldsymbol{h}_t &= \boldsymbol{o}_t \odot tanh(\boldsymbol{c}_t) \end{aligned} \tag{2}$$

where δ is the logistic sigmoid function and $\boldsymbol{i}, \boldsymbol{f}, \boldsymbol{o}$ and $\boldsymbol{c}$ are input, forget, ouput gate and cell activation.

An attention layer [51] is employed to enhance the performance of hybrid CNN-RNN architecture. Its calculation formula is as follows.

$$\begin{aligned} \boldsymbol{M}_t &= tanh(\boldsymbol{W}_h\boldsymbol{h}_t) \\ \alpha_t &= softmax(\boldsymbol{w}^T\boldsymbol{M}_t) \\ \boldsymbol{r} &= \textstyle\sum_{t=1}^{T} \alpha_t\boldsymbol{h}_t \end{aligned} \tag{3}$$

where h_t is the output of the t-th hidden unit of RNN module, α_t is the t-th attention weight, $\boldsymbol{W_h}$ and $\boldsymbol{w^T}$ are weighted matrices and r is the output of attention module. The output $\boldsymbol{r}$ is followed by a G-way fully-connected layer and a softmax classifier.

Loss function. The loss function of attention-based hybrid CNN-RNN architecture is:

$$loss = \alpha \cdot loss_{\text{attention}} + \beta \cdot loss_{\text{target}} + \lambda \cdot ||\boldsymbol{w}||^2 \tag{4}$$

where the first term is the attention loss, the second term is the target replication loss [54] and the last term is the regularization term. The α, β and λ are three weight parameters.

$$loss_{\text{attention}} = \frac{1}{T} \cdot l(g_1(\boldsymbol{X}), y) \tag{5}$$

$$l(g_1(\boldsymbol{X}), y) = -\sum_{i=1}^{G} \mathbf{1}_i(y) log\ g_1(\boldsymbol{X})^i \tag{6}$$

$$g_1(\boldsymbol{X}) = f_s(f_a(f_h(\boldsymbol{X_1}), f_h(\boldsymbol{X_2}), \ldots, f_h(\boldsymbol{X_T}))) \tag{7}$$

where $\boldsymbol{X}$ is the electromyogram signal to be recognized, y is the ground-truth label, and T is the number of time steps of RNN. G is the number of gestures to recognize, $g_1(\boldsymbol{X})^i$ is the i-th dimension of $g_1(\boldsymbol{X})$ and $\mathbf{1}_i()$ is the indicator function. f_h, f_a and f_s stand for the hybrid CNN-RNN architecture, attention module and the last softmax layer, respectively.

$$loss_{\text{target}} = \frac{1}{T} \sum_{t=1}^{T} l(g_2(\boldsymbol{X}_t), y) \tag{8}$$

$$l(g_2(\boldsymbol{X}_t), y) = -\sum_{i=1}^{G} \mathbf{1}_i(y) log\ g_2(\boldsymbol{X}_t)^i \tag{9}$$

$$g_2(\boldsymbol{X}_t) = f_s(f_h(\boldsymbol{X_t})) \tag{10}$$

where $\boldsymbol{X}_t$ is the t-th subsegment of $\boldsymbol{X}$ and $g_2(\boldsymbol{X}_t)^i$ is the i-th dimension of $g_2(\boldsymbol{X}_t)$. f_h and f_s stand for the hybrid CNN-RNN architecture and the last softmax layer.

Image representation from temporal electromyogram signals

Existing sEMG databases can be divided into two categories: sparse multi-channel sEMG database [55] and high-density sEMG database [26, 56]. We generate sEMG images for both sparse multi-channel and high-density sEMG databases. As mentioned in [26], we convert a segment of electromyogram signal into a sEMG image which has the same dimensions (i.e., color channel, width, and height) as the RGB image.

An intuitive sEMG image representation method is to use the placement of electrodes and each electrode can be regarded as a pixel of sEMG images. It is a feasible sEMG image representation method for high-density sEMG databases csl-hdemg and CapgMyo-DBa, as electromyogram signals are collected by a grid of sEMG electrodes. The detailed image representation procedure is described as follows. The input is a segment of electromyogram signal of the high-density electromyogram signal with size $L \times W \times H$, where L is the number of frames, W is rows of the array electrode and H is columns of the array electrode. The raw signal is converted into a sEMG image with size $L \times W \times H$, where L is the number of color channels of the sEMG image, W is the width of sEMG image and H is the height of sEMG image. The sEMG image size of csl-hdemg and CapgMyo-DBa are $L \times 24 \times 7$ and $L \times 16 \times 8$, respectively.

However, for sparse multi-channel sEMG database, the number of electrodes is limited and the placement is sparse. Inspired by the image representation method used in [37] for human activity recognition with accelerometer and gyroscope, there are six image representation methods for raw electromyogram signal, namely **raw-image1**, **raw-image2**, **signal-image1**, **signal-image2**, **activity-image1** and **activity-image2**. The input of the sEMG image representation methods is a segment of electromyogram signal of NinaProDB1 with size $L \times C$, where L

is the number of frames, C is the number of signal channels and $C = 10$ for NinaProDB1. The detailed sEMG image representation methods are described as follows.

1. The **raw-image1** is obtained by transforming the input into a sEMG image with size $L \times 1 \times 10$, where L is the number of color channels of the sEMG image, 1 is the width of sEMG image, and 10 is the height of sEMG image.
2. The **raw-image2** [25] is obtained by transforming the input into a sEMG image with size $1 \times L \times 10$, where 1 is the number of color channels of the sEMG image, L is the width of sEMG image, 10 is the height of sEMG image, and `width × height = signal channels`.
3. The **signal-image1** [37] is formed by rearranging the data of each signal channel in [37] with size $L \times 1 \times 51$.
4. The **signal-image2** [37] is formed by the same procedure as **signal-image1** with size $1 \times L \times 51$.
5. The **activity-image1** [37] is generated by FFT transformation of signal-image1 with size $L \times 1 \times 51$.
6. The **activity-image2** [37] is generated by FFT transformation of signal-image2 with size $1 \times L \times 51$.

We evaluate the six sEMG image representation methods using the existing CNN architecture [26] on NinaProDB1. Firstly, the electromyogram signal is segmented by the sliding window with 200ms length and converted into six sEMG images which can be found in Fig 2. Then, the training set and test set are the same as those described in the experimental setup. Finally, GengNet [26] is employed to respectively extract useful information of the six sEMG images. The classification accuracy in Table 2 shows that the signal-image method is a feasible sEMG image representation method to improve recognition accuracy for sparse multi-channel electromyogram signal. The signal-image method achieves higher classification accuracy than the raw-image method for signal-image method contains more information between different channels. The activity-image methods perform the worst in three image representation methods because of the FFT transform which may cause time-domain information loss.

The feature extraction plays a significant role in traditional sEMG-based gesture recognition methods and many classical feature sets have achieved good performance [11, 13, 15]. Therefore, we want to generate a new sEMG image based on the traditional feature vector. The most obvious idea is to flatten all the feature vectors of different channels into one vector with size

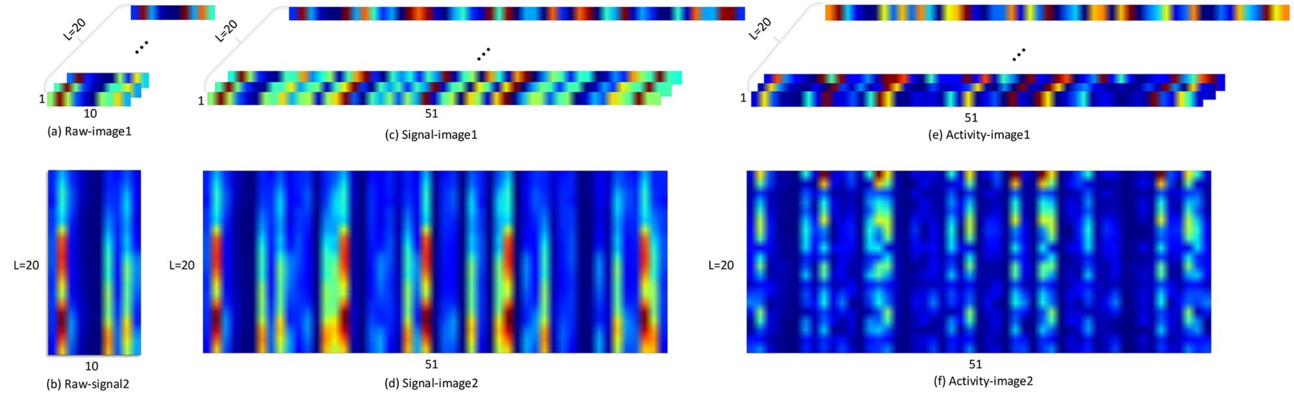

Fig 2. Six raw signal based sEMG image representation methods.

https://doi.org/10.1371/journal.pone.0206049.g002

Table 2. Comparison of gesture recognition accuracy with various image representation methods on NinaProDB1. Here, we employ the GengNet [26], and the sliding window length is fixed at 200ms for all experiments.

sEMG Image	Classification Accuracy
raw-image1	83.5%
raw-image2	82.9%
signal-image1	84.9%
signal-image2	79.8%
activity-image1	78.1%
activity-image2	72.8%
feature-signal-image1	**86.3%**

https://doi.org/10.1371/journal.pone.0206049.t002

`feature dimension = signal channels × feature vector dimension` and conduct a sEMG image with size 1 × 1 × `feature dimension` which is denoted as feature-flatten-image. We have evaluated this sEMG image representation method on the NinaProDB1, and the recognition accuracy is 81.9% which is lower than that of raw signal based sEMG images (i.e., "raw-image1" and "signal-image1"). Inspired by the good performance of signal-images, we propose a new sEMG image representation method "feature-signal-image1" which makes full use of the traditional feature vector and achieves better performance among existing sEMG image representation methods.

Experiments and results

In this section, we first delineate the experimental setup, followed by the performance comparisons between proposed architecture and state-of-the-art methods on five sEMG benchmark databases. Secondly, we discuss the effectiveness of attention mechanism. Then, the results of different image representation methods are presented. Finally, we evaluate and discuss the impacts of various parameters of the architecture on recognition accuracy.

Experimental setup

In this work, we follow the experimental process which consists of data acquisition, preprocessing, segmentation and gesture recognition. The proposed architecture is implemented using MxNet [57], and the evaluations are carried out on five sEMG benchmark databases, namely NinaProDB1, NinaProDB2, BioPatRec26MOV, CapgMyo-DBa and csl-hdemg. The details summary of all database are shown in Table 3.

The first and second sub-database of NinaPro database [55] are denoted as NinaProDB1 and NinaProDB2, respectively. NinaProDB1 contains a total of 52 gestures from 27 subjects, including 9 wrist movements, 8 hand postures, 12 finger movements and 23 grasping and functional movements. The electromyogram signal is filtered by a low-pass Butterworth filter [26, 55]. NinaProDB2 collects 50 gestures from 40 subjects, including 23 grasping and

Table 3. Details of five sEMG benchmark databases.

Database	Subjects	Gestures	Sessions	Trials	Number of electrodes	Sampling rate (Hz)
NinaproDB1	27	52	1	10	10	100
NinaProDB2	40	50	1	6	12	2000
BioPatRec26MOV	17	26	1	3	8	2000
CapgMyo-DBa	18	8	1	10	128	1000
csl-hdemg	5	27	5	10	192	2048

https://doi.org/10.1371/journal.pone.0206049.t003

functional movements, 9 wrist movements, 8 hand postures, 9 finger force patterns and the rest position. The electromyogram signal is filtered by a low-pass Butterworth filter [26, 55] and is downsampled to 100HZ which is NinaProDB1's sampling rate.

A subset of BioPatRec toolbox is available online [58] (denoted as BioPatRec26MOV), which collects 26 hand movements from 17 subjects using 8 sEMG electrodes. The duration of the contraction is based on a contraction time percentage, which is set to the default value 0.7 [15].

The first sub-database of CapgMyo database [26] (denoted as CapgMyo-DBa), which contains 8 hand gestures from 18 subjects and each gesture performed 10 trials. The electromyogram signal is band-pass filtered [26] in the data collection.

The csl-hdemg database [56] contains 27 finger gestures from 5 subjects, where each subject was recorded 5 sessions and performed each gesture 10 trials in each session. The electromyogram signal is rectified [59] and filtered by a low-pass Butterworth filter in pre-processing.

Given the preprocessed electromyogram signal, we decompose it into small segments using the sliding window strategy with overlapped windowing scheme to fully utilize the computing capacity of the system [60]. The window length must be shorter than 300ms [11] to satisfy real-time usage constraints. To compare our proposed method with previous works, we follow the segmentation strategy in previous studies. The window length is fixed to 150ms and 200ms for NinaProDB1, 200ms for NinaProDB2, 50ms and 150ms for BioPatRec26MOV, 40ms and 150ms for CapgMyo-DBa and 150ms and 170ms for csl-hdemg. For NinaProDB2 and BioPatRec26MOV, the sliding window steps of test sets are the same as those in existing works [15, 29] which are 100ms and 50ms, respectively.

In previous works on NinaProDB1 and NinaProDB2 [25, 29], the training set consists of approximately 2/3 of the gesture trials of each subject and the remaining trials constitute the test set. For BioPatRec26MOV, we conduct the intra-session cross-validation scheme mentioned in [15]. As each gesture has 3 repetitions in BioPatRec26MOV, the first repetition is applied as the training set and the other two repetitions are applied as the test set [15]. According to previous works on csl-hdemg [26, 27], the intra-session cross-validation scheme was adopted. For each session, a leave-one-out cross-validation is performed, in which each of the 10 trials is used as the test set and the remaining 9 trials are used as training set. For CapgMyo-DBa, the training set consisted of half of the trials, and the other half constitute the test set [26].

Based on the recognition results of test sets, the classification accuracy is calculated for each database as given below:

$$\text{Classification Accuracy} = \frac{\text{Number of correct classifications}}{\text{Total number of test samples}} * 100\% \tag{11}$$

Comparison with existing deep learning approaches

We compare proposed attention-based hybrid CNN-RNN architecture with the state-of-the-art deep learning approaches on five sEMG benchmark databases and the results can be found in Table 4.

In this work, the compared approaches are AtzoriNet [25], GengNet [26] and ZhaiNet [29]. We also compare the proposed method with the state-of-the-art traditional machine learning method using a random forest classifier (namely Traditional-RF [55]) and new feature set with LDA classifier (namely Feature-LDA [15]). The proposed architecture on NinaProDB1 using raw-image1 achieves 84.7% classification accuracy of 52 hand gestures which is 6.9% higher than the GengNet. The feature-signal-image1 improves the accuracy from 84.7% to 86.7%

Table 4. Classification accuracy of the proposed method and previous works.

	NinaProDB1			NinaProDB2		BioPatRec26MOV			CapgMyo-DBa			csl-hdemg		
	150ms	200ms	Trial	200ms	Trial	50ms	150ms	Trial	40ms	150ms	Trial	150ms	170ms	Trial
Feature-LDA [15]	-	-	-	-	-	86.3%	92.9%	-	-	99.0%	-	-	-	-
Traditional-RF [55]	-	75.3%	-	-	-	-	-		-	-	-	-	-	-
AtzoriNet [25]	-	66.6%	-	-	-	-		-	-	-	-	-	-	-
GengNet [26]	-	77.8%	96.7%	-	-	-	-	-	99.0%	99.5%	-	89.3%	90.4%	96.8%
ZhaiNet [29]	-	-	-	78.71%	-	-	-	-	-	-	-	-	-	-
RNN Module with raw-signal	78.1%	79.8%	95.0%	Did not converge		76.4%	82.3%	92.6%	71.8%	80.4%	90.4%	65.3%	71.1%	75.8%
CNN Module with raw-image1	82.6%	83.5%	96.5%	73.4%	97.6%	82.1%	83.9%	92.2%	98.0%	97.7%	98.9%	92.0%	92.1%	95.2%
CNN Module with feature-signal-image1	85.4%	86.3%	97.2%	81.4%	97.5%	85.2%	90.0%	95.8%	-	-	-	-	-	-
Hybrid CNN-RNN with raw-image1	83.5%	84.7%	96.5%	74.6%	97.7%	88.5%	92.2%	96.8%	99.1%	99.6%	99.9%	94.3%	94.8%	96.1%
Hybrid CNN-RNN with feature-signal-image1	86.4%	86.7%	97.1%	82.0%	97.5%	89.9%	93.9%	97.5%	-	-	-	-	-	-
Attention-based hybrid CNN-RNN with raw-image1	83.7%	84.8%	96.5%	74.8%	97.6%	88.7%	92.5%	96.8%	**99.3%**	**99.7%**	**99.9%**	**94.5%**	**94.9%**	**96.1%**
Attention-based hybrid CNN-RNN with feature-signal-image1	**86.8%**	**87.0%**	**97.3%**	**82.2%**	**97.6%**	**90.0%**	**94.1%**	**97.7%**	-	-	-	-	-	-

https://doi.org/10.1371/journal.pone.0206049.t004

which is 8.9% higher than state-of-the-art deep learning approach and 11.4% higher than state-of-the-art traditional machine learning method. The attention mechanism has improved the accuracy from 86.7% to 87.0%. For NinaProDB2, the proposed attention-based hybrid CNN-RNN architecture using feature-signal-image1 achieves 82.2% classification accuracy of 50 hand gestures from 40 subjects which is 78.71% in previous work [29]. The classification accuracy of 26 gestures from BioPatRec26MOV database is 94.1% which is 92.9% in the existing work [15].

The classification accuracy of CapgMyo database is close to saturation, and attention-based hybrid CNN-RNN achieves 99.7% classification accuracy which is 0.2% higher than the GengNet. For csl-hdemg database, the accuracy is improved from 89.3% to 94.5% by the proposed attention-based hybrid CNN-RNN architecture.

After training the attention-based hybrid CNN-RNN model on GPUs, we achieved the trained model which can be applied for sEMG-based gesture recognition on any machine that contains GPU or CPU. In order to discuss the recognition time of each sample for five benchmark databases, we test the trained model on a workstation with one NVIDIA TITAN Xp GPU and one Intel 6850k CPU. The results are shown in Table 5. The recognition time of each sample on GPU is less than 10ms and the model can be applied for prosthetic control and human-computer interaction [61]. The recognition time of each sample on CPU is less than 350ms which satisfies response time constraints for human-computer interaction [61].

Ablation studies on the proposed architecture

To prove the advantage of hybrid CNN-RNN, we evaluate the CNN and RNN module which are constructed using model parameters mentioned before. In the CNN module evaluation,

Table 5. Recognition time of each sample on five benchmark databases with attention-based hybrid CNN-RNN architecture. The recognition window length is 200ms for NinaProDB1 and NinaProDB2, 150ms for BioPatRec26MOV, CapgMyo-DBa and csl-hdemg.

	NinaProDB1	NinaProDB2	BioPatRec26MOV	CapgMyo-DBa	csl-hdemg
GPU	3.0ms	3.6ms	4.1ms	7.8ms	6.0ms
CPU	106ms	140ms	107ms	258ms	327ms

https://doi.org/10.1371/journal.pone.0206049.t005

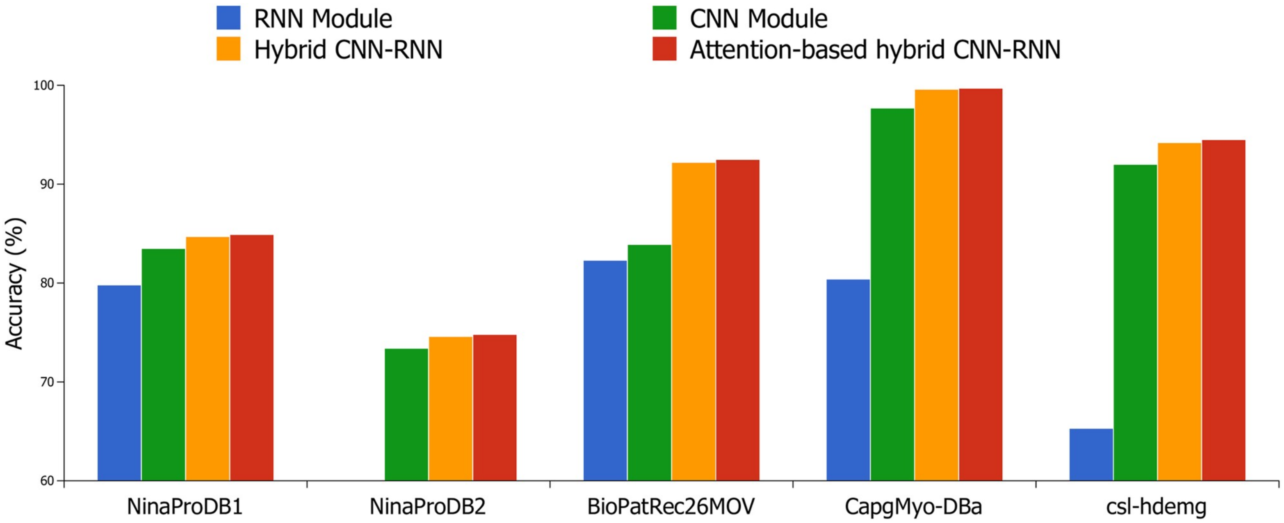

Fig 3. Classification accuracy of RNN module with raw-signal, CNN module hybrid CNN-RNN and attention-based hybrid CNN-RNN architectures with raw-image1 on five benchmark databases.

https://doi.org/10.1371/journal.pone.0206049.g003

the input sEMG data are the same as those in hybrid CNN-RNN, and we employ a softmax layer instead of RNN unit. In the RNN module evaluation, the input is the same as that in hybrid CNN-RNN, but there is no need to convert the electromyogram signal into a sEMG image. Each frame of the input electromyogram signal is directly followed by RNN without extracting high-level features by CNN. Moreover, we inject the attention mechanism into RNN module of proposed hybrid CNN-RNN architecture, and it allows the model to pay attention to the subsegments which contain more discriminative information for gesture recognition. As can be seen in Table 4 and Fig 3, the attention-based hybrid CNN-RNN architecture outperforms the other three frameworks on five sEMG benchmark databases. The improvements of recognition accuracy for attention-based hybrid CNN-RNN architecture are 0.3% on NinaProDB1, 0.2% on NinaProDB2, 0.3% on BioPatRec26MOV, 0.1% on CapgMyo-DBa and 0.2% on csl-hdemg, respectively.

Since the accuracy enhancement capability of attention mechanism is influenced by the length of input sequence (i.e., the number of subsegments), we present the results of different numbers of subsegments on NinaProDB1. The number of subsegments is set as {2,5,10,20} and the results are shown in Fig 4. If we set the number of subsegments as 2, the recognition accuracy of the attention-based hybrid CNN-RNN is 0.3% higher than that of the hybrid CNN-RNN. However, if we increase the number of subsegment to 10, accuracy of the attention-based model is 0.7% higher than that of the model without attention.

Evaluation of different image representation methods

We compare the "feature-signal-image1" with two raw signal based methods "raw-image1" and "signal-image1" for CNN, hybrid CNN-RNN and attention-based hybrid CNN-RNN frameworks. The results in Fig 5 show that the "feature-signal-image1" achieves the best performance for all the three frameworks on three sparse multi-channel databases NinaProDB1, NinaProDB2 and BioPatRec26MOV.

We also evaluate the eight image representation methods on the sparse multi-channel database NinaProDB1 and the results can be seen in Table 6. The raw-image1, raw-image2, signal-image1, signal-image2, activity-image1, activity-image2 and feature-signal-image1 are

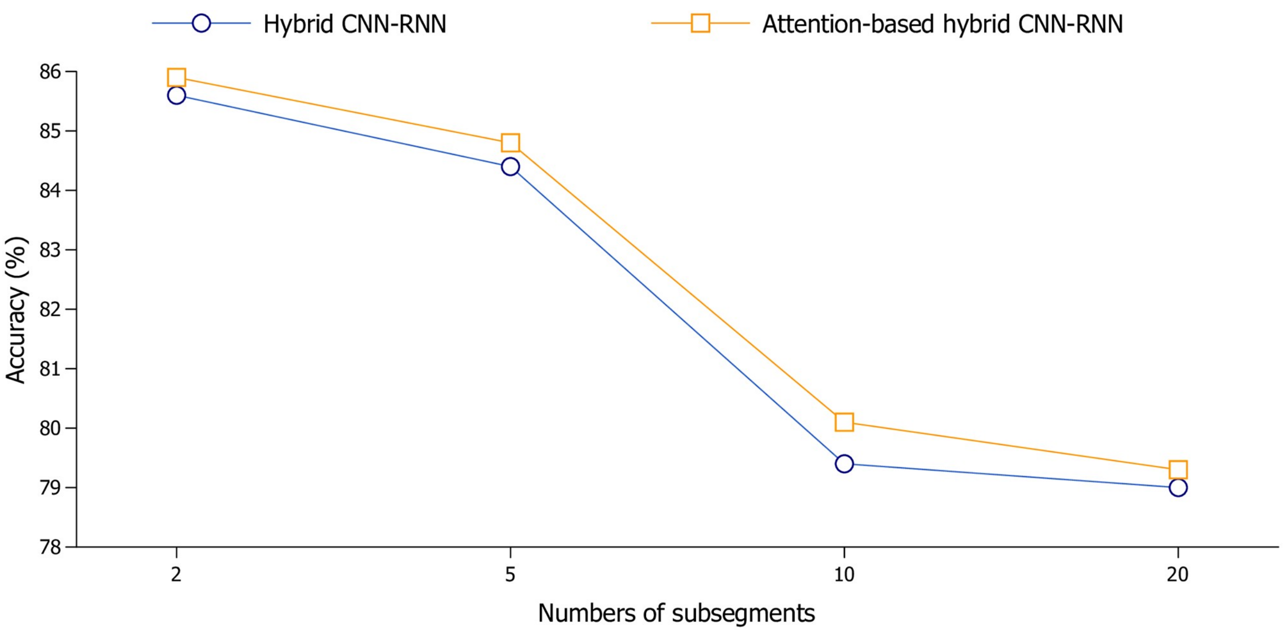

Fig 4. Classification accuracy of attention-based hybrid CNN-RNN architecture with with different numbers of subsegments on NinaProDB1.

https://doi.org/10.1371/journal.pone.0206049.g004

mentioned in section Methods. The feature-signal-image2 uses the same generation procedure as that of signal-image2 and the raw-image2 has been used in existing work [25]. In Table 6, we find that feature-signal-images formed by feature vectors achieve higher accuracy than sEMG images formed by raw signals. The raw-image1, signal-image1, activity-image1 and feature-signal-image1 obtain higher accuracy than the general image representation methods raw-image2, signal-image2, activity-image2 and feature-signal-image2, respectively. We draw the same conclusion for CNN module, hybrid CNN-RNN and attention-based hybrid CNN-RNN architectures that the **feature-signal-image1** achieves the highest accuracy in the eight evaluated sEMG image representation methods. For the input of RNN module is a vector instead of an image, we also compare the raw signal with feature vector for the RNN module and the accuracies are 79.8% and 74.5%, respectively.

Conclusion

In this work, we propose an attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition, which consists of feature extraction stage and attention-based sequential modeling stage. It makes full use of spatial and temporal information of electromyogram signals and the attention mechanism makes the network more intelligent to pay attention to different parts of the electromyogram signal. The evaluations are performed on five sEMG benchmark databases, namely NinaProDB1, NinaProDB2, BioPatRec26MOV, CapgMyo-DBa and csl-hdemg databases. The results show that 1) the hybrid CNN-RNN architecture outperforms both CNN and RNN modules; 2) the attention mechanism enhances the performance of the hybrid CNN-RNN architecture. Moreover, we present a new feature vector based sEMG image representation method "feature-signal-image1" for sparse multi-channel databases. Compared with the sEMG image representation method "raw-image1", it improves the recognition accuracy from 84.8% to 87.0% on NinaProDB1, from 74.8% to 82.2% on NinaProDB2, from 92.5% to 94.1% on BioPatRec26MOV. Overall, the recognition accuracies of proposed sEMG-based gesture recognition method are 87.0% for NinaProDB1, 82.2% for NinaProDB2,

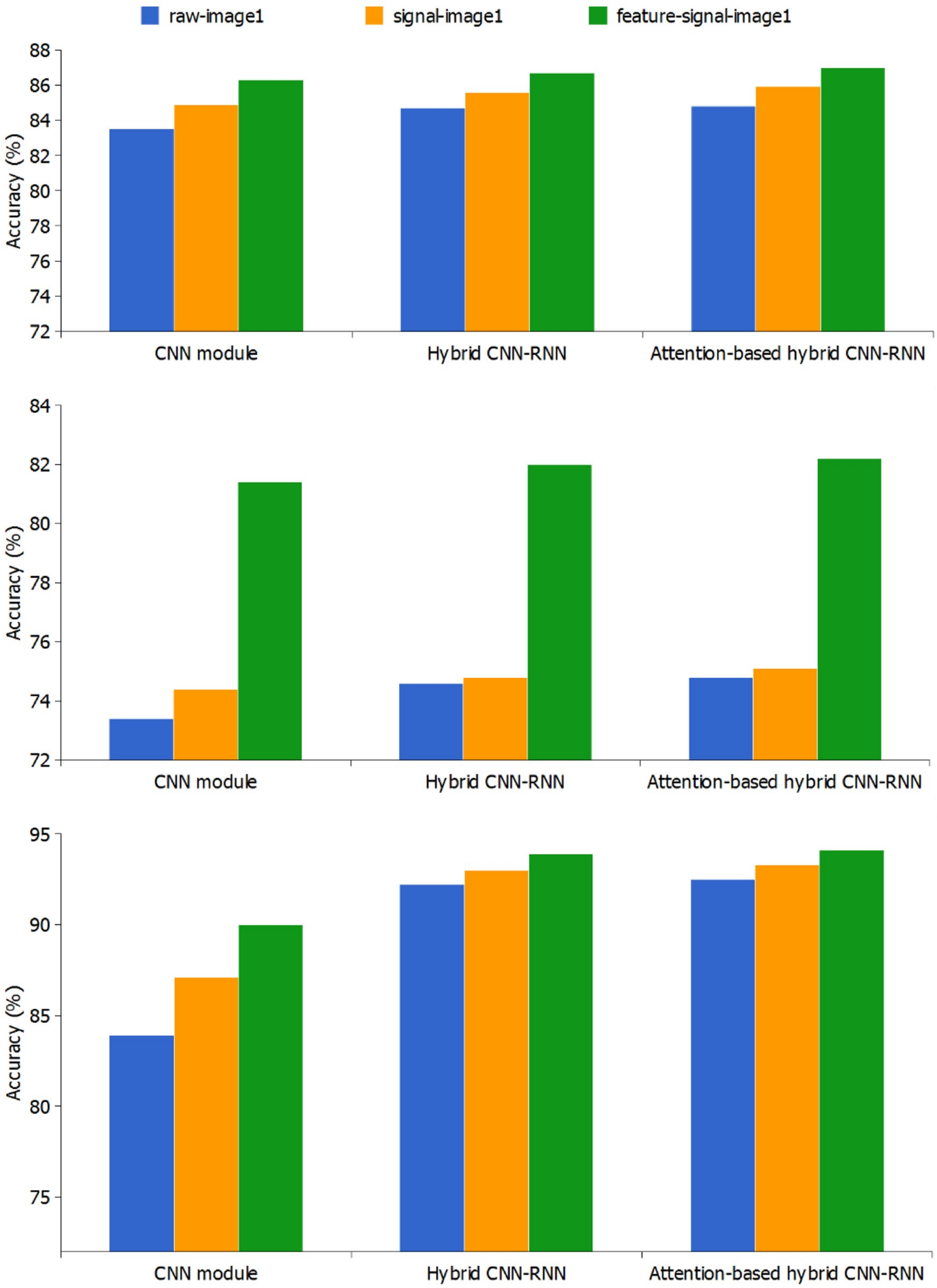

Fig 5. Classification accuracy of CNN module, hybrid CNN-RNN and attention-based hybrid CNN-RNN architectures with three sEMG image representation methods on three sparse multi-channel benchmark databases.

https://doi.org/10.1371/journal.pone.0206049.g005

Table 6. Classification accuracy of different image representation methods on NinaProDB1. We use the same sliding window length (200ms) for all experiments mentioned bellow.

	CNN Module	hybrid CNN-RNN	Attention-based hybrid CNN-RNN
raw-image1	83.5%	84.7%	84.8%
raw-image2	82.9%	80.8%	80.9%
signal-image1	84.9%	85.6%	85.9%
signal-image2	79.8%	81.6%	82.0%
activity-image1	78.1%	78.8%	79.1%
activity-image2	72.8%	74.1%	74.5%
feature-signal-image1	**86.3%**	**86.7%**	**87.0%**
feature-signal-image2	83.1%	84.5%	84.7%

https://doi.org/10.1371/journal.pone.0206049.t006

94.1% for BioPatRec26MOV, 99.7% for CapgMyo-DBa and 94.5% for csl-hdemg. The improvements are 9.2% (NinaProDB1), 3.5% (NinaProDB2), 1.2% (BioPatRec26MOV), 0.2% (CapgMyo-DBa) and 5.2% (csl-hdemg) higher than the state-of-the-art performances [15, 26, 29, 55].

The electromyogram signal is a kind of biological signal which is severely affected by the difference between subjects. It makes the accuracy of Leave-One-Subject-Out cross-validation (LOSOCV) much lower than that of Within-Subject cross-validation (WSCV) in previous works [14, 27]. Future research will be to improve the accuracy of LOSOCV which is significant for a new user to interact with computers. We will first extend our framework to fuse the sEMG data with IMU data and extract common features of different subjects to improve the LOSOCV accuracy. Then, we will propose a framework to integrate information from various sensors in the HCI system to allow both intact-limbed and amputees to communicate with different kinds of machines efficiently.

Acknowledgments

This work was supported by grants from the National Key Research and Development Program of China (No. 2016YFB1001302), the National Natural Science Foundation of China (No. 61379067), and the National Research Foundation, Prime Ministers office, Singapore under its International Research Centre in Singapore Funding Initiative.

Author Contributions

Data curation: Yu Hu, Wentao Wei, Yu Du.

Formal analysis: Wentao Wei, Yu Du.

Methodology: Yu Hu, Yongkang Wong, Weidong Geng.

Software: Yu Hu.

Supervision: Yongkang Wong, Mohan Kankanhalli, Weidong Geng.

Validation: Yongkang Wong.

Writing – original draft: Yu Hu.

Writing – review & editing: Yu Hu, Yongkang Wong, Wentao Wei, Mohan Kankanhalli, Weidong Geng.

References

1. Reaz MB, Hussain M, Mohd-Yasin F. Techniques of EMG signal analysis: detection, processing, classification and applications. Biological Procedures Online. 2006; 8(1):11–35. https://doi.org/10.1251/bpo124
2. Hakonen M, Piitulainen H, Visala A. Current state of digital signal processing in myoelectric interfaces and related applications. Biomed Signal Process Control. 2015; 18:334–359. https://doi.org/10.1016/j.bspc.2015.02.009
3. Karlik B, Tokhi MO, Alci M. A fuzzy clustering neural network architecture for multifunction upper-limb prosthesis. IEEE Transactions on Biomedical Engineering. 2003; 50:1255–1261. https://doi.org/10.1109/TBME.2003.818469 PMID: 14619995
4. Moon I, Lee M, Ryu J, Mun M. Intelligent robotic wheelchair with EMG-, gesture-, and voice-based interfaces. In: IEEE/RSJ International Conference on Intelligent Robots and Systems; 2003. p. 3453–3458.
5. Zhang X, Wang X, Wang B, Sugi T, Nakamura M. Meal assistance system operated by electromyogram (EMG) signals: Movement onset detection with adaptive threshold. International Journal of Control, Automation and Systems. 2010; 8:392–397.
6. Rosen J, Fuchs MB, Arcan M. Performances of Hill-Type and Neural Network Muscle models-Toward a Myosignal-Based Exoskeleton. Computers and Biomedical Research. 1999; 32:415–439. https://doi.org/10.1006/cbmr.1999.1524 PMID: 10529300
7. Lyons GM, Sharma P, Baker M, O'Malley S, Shanahan A. A computer game-based EMG biofeedback system for muscle rehabilitation. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2003. p. 1625–1628.
8. van Dijk L, van der Sluis CK, van Dijk HW, Bongers RM. Learning an EMG Controlled Game: Task-Specific Adaptations and Transfer. PLOS ONE. 2016; 11(8):1–14. https://doi.org/10.1371/journal.pone.0160817
9. Asai Y, Tateyama S, Nomura T. Learning an Intermittent Control Strategy for Postural Balancing Using an EMG-Based Human-Computer Interface. PLOS ONE. 2013; 8(5):1–19. https://doi.org/10.1371/journal.pone.0062956
10. Jorgensen C, Dusan S. Speech interfaces based upon surface electromyography. Speech Communication. 2010; 52(4):354–366. https://doi.org/10.1016/j.specom.2009.11.003
11. Hudgins B, Parker P, Scott RN. A new strategy for multifunction myoelectric control. IEEE Transactions on Biomedical Engineering. 1993; 40(1):82–94. https://doi.org/10.1109/10.204774 PMID: 8468080
12. Du YC, Lin CH, Shyu LY, Chen T. Portable hand motion classifier for multi-channel surface electromyography recognition using grey relational analysis. Expert Syst Appl. 2010; 37(6):4283–4291. https://doi.org/10.1016/j.eswa.2009.11.072
13. Phinyomark A, Phukpattaranont P, Limsakul C. Feature reduction and selection for EMG signal classification. Expert Systems with Applications. 2012; 39(8):7420–7431. https://doi.org/10.1016/j.eswa.2012.01.102
14. Doswald A, Carrino F, Ringeval F. Advanced Processing of sEMG Signals for User Independent Gesture Recognition. In: Mediterranean Conference on Medical and Biological Engineering and Computing; 2014. p. 758–761.
15. Khushaba RN, Al-Timemy AH, Al-Ani A, Al-Jumaily A. A Framework of Temporal-Spatial Descriptors-Based Feature Extraction for Improved Myoelectric Pattern Recognition. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2017; 25(10):1821–1831. https://doi.org/10.1109/TNSRE.2017.2687520 PMID: 28358690
16. Kim J, Mastnik S, André E. EMG-based hand gesture recognition for realtime biosignal interfacing. In: International Conference on Intelligent User Interfaces; 2008. p. 30–39.
17. Naik GR, Acharyya A, Nguyen HT. Classification of finger extension and flexion of EMG and Cyberglove data with modified ICA weight matrix. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2014. p. 3829–3832.
18. Samadani AA, Kulic D. Hand gesture recognition based on surface electromyography. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2014. p. 4196–4199.
19. Patricia N, Tommasit T, Caputo B. Multi-Source Adaptive Learning for Fast Control of Prosthetics Hand. In: International Conference on Pattern Recognition; 2014. p. 2769–2774.
20. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition; 2016. p. 770–778.
21. Farabet C, Couprie C, Najman L, LeCun Y. Learning hierarchical features for scene labeling. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2013; 35(8):1915–1929. https://doi.org/10.1109/TPAMI.2012.231 PMID: 23787344

22. Kiros R, Zhu Y, Salakhutdinov RR, Zemel R, Urtasun R, Torralba A, et al. Skip-thought vectors. In: Advances in neural information processing systems; 2015. p. 3294–3302.

23. Luong M, Pham H, Manning CD. Effective Approaches to Attention-based Neural Machine Translation. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing; 2015. p. 1412–1421.

24. Park KH, Lee SW. Movement intention decoding based on deep learning for multiuser myoelectric interfaces. In: International Winter Conference on Brain-Computer Interface; 2016. p. 1–2.

25. Atzori M, Cognolato M, Müller H. Deep Learning with Convolutional Neural Networks Applied to Electromyography Data: A Resource for the Classification of Movements for Prosthetic Hands. Frontiers in Neurorobotics. 2016; 10. https://doi.org/10.3389/fnbot.2016.00009 PMID: 27656140

26. Geng W, Du Y, Jin W, Wei W, Hu Y, Li J. Gesture recognition by instantaneous surface EMG images. Scientific Reports. 2016; 6:36571. https://doi.org/10.1038/srep36571 PMID: 27845347

27. Du Y, Jin W, Wei W, Hu Y, Geng W. Surface EMG-Based Inter-Session Gesture Recognition Enhanced by Deep Domain Adaptation. Sensors. 2017; 17(3). https://doi.org/10.3390/s17030458

28. Du Y, Wong Y, Jin W, Wei W, Hu Y, Kankanhalli M, et al. Semi-supervised Learning for Surface EMG-based Gesture Recognition. In: International Joint Conference on Artificial Intelligence; 2017. p. 1624–1630.

29. Zhai X, Jelfs B, Chan RH, Tin C. Self-recalibrating surface EMG pattern recognition for neuroprosthesis control based on convolutional neural network. Frontiers in neuroscience. 2017; 11:379. https://doi.org/10.3389/fnins.2017.00379 PMID: 28744189

30. Ebrahimi Kahou S, Michalski V, Konda K, Memisevic R, Pal C. Recurrent neural networks for emotion recognition in video. In: ACM International Conference on Multimodal Interaction; 2015. p. 467–474.

31. Wu Z, Wang X, Jiang YG, Ye H, Xue X. Modeling spatial-temporal clues in a hybrid deep learning framework for video classification. In: ACM international conference on Multimedia; 2015. p. 461–470.

32. Ordóñez FJ, Roggen D. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. Sensors. 2016; 16(1):115. https://doi.org/10.3390/s16010115

33. Wang P, Song Q, Han H, Cheng J. Sequentially supervised long short-term memory for gesture recognition. Cognitive Computation. 2016; 8(5):982–991. https://doi.org/10.1007/s12559-016-9388-6

34. Bahdanau D, Cho K, Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate. In: International Conference on Learning Representations; 2015.

35. Xu K, Ba J, Kiros R, Cho K, Courville AC, Salakhutdinov R, et al. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. In: Proceedings of the International Conference on Machine Learning; 2015. p. 2048–2057.

36. Chorowski JK, Bahdanau D, Serdyuk D, Cho K, Bengio Y. Attention-Based Models for Speech Recognition. In: Advances in Neural Information Processing Systems; 2015. p. 577–585.

37. Jiang W, Yin Z. Human activity recognition using wearable sensors by deep convolutional neural networks. In: ACM International Conference on Multimedia; 2015. p. 1307–1310.

38. Oskoei MA, Hu H. Myoelectric control systems-A survey. Biomed Signal Process Control. 2007; 2:275–294.

39. Phinyomark A, Limsakul C, Phukpattaranont P. A Novel Feature Extraction for Robust EMG Pattern Recognition. Journal of Computing. 2009; 1:71–80.

40. Menon R, Caterina GD, Lakany H, Petropoulakis L, Conway B, Soraghan J. Study on interaction between temporal and spatial information in classification of EMG signals in myoelectric prostheses. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2017; PP(99):1–1.

41. Jarrassé N, Nicol C, Touillet A, Richer F, Martinet N, Paysant J, et al. Classification of Phantom Finger, Hand, Wrist, and Elbow Voluntary Gestures in Transhumeral Amputees With sEMG. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2017; 25(1):71–80. https://doi.org/10.1109/TNSRE.2016.2563222

42. Yun LK, Swee TT, Anuar R, Yahya Z, Yahya A, Kadir MRA. Sign Language Recognition System using SEMG and Hidden Markov Model. In: International Conference on Mathematical Methods, Computational Techniques and Intelligent Systems; 2013. p. 50–53.

43. Hioki M, Kawasaki H. Estimation of finger joint angles from sEMG using a neural network including time delay factor and recurrent structure. ISRN Rehabilitation. 2012; 2012. https://doi.org/10.5402/2012/604314

44. Quivira F, Koike-Akino T, Wang Y, Erdogmus D. Translating sEMG signals to continuous hand poses using recurrent neural networks. In: Biomedical & Health Informatics (BHI), 2018 IEEE EMBS International Conference on. IEEE; 2018. p. 166–169.

45. Amor ABH, Ghoul O, Jemni M. Toward sign language handshapes recognition using Myo armband. In: Information and Communication Technology and Accessibility (ICTA), 2017 6th International Conference on. IEEE; 2017. p. 1–6.

46. Shin S, Baek Y, Lee J, Eun Y, Son SH. Korean sign language recognition using EMG and IMU sensors based on group-dependent NN models. In: Computational Intelligence (SSCI), 2017 IEEE Symposium Series on. IEEE; 2017. p. 1–7.

47. Song S, Lan C, Xing J, Zeng W, Liu J. An End-to-End Spatio-Temporal Attention Model for Human Action Recognition from Skeleton Data. In: AAAI; 2017. p. 4263–4270.

48. Goferman S, Zelnik-Manor L, Tal A. Context-aware saliency detection. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2012; 34(10):1915–1926. https://doi.org/10.1109/TPAMI.2011.272 PMID: 22201056

49. Merletti R, Farina D. Surface electromyography: physiology, engineering and applications. John Wiley & Sons; 2016.

50. Huang YY, Low KH, Lim HB. Objective and quantitative assessment methodology of hand functions for rehabilitation. In: 2008 IEEE International Conference on Robotics and Biomimetics; 2009. p. 846–851.

51. Baziotis C, Pelekis N, Doulkeridis C. DataStories at SemEval-2017 Task 4: Deep LSTM with Attention for Message-level and Topic-based Sentiment Analysis. In: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017); 2017. p. 747–754.

52. Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In: International Conference on Machine Learning; 2015. p. 448–456.

53. Sundermeyer M, Schlüter R, Ney H. LSTM Neural Networks for Language Modeling. In: Interspeech; 2012. p. 194–197.

54. Lipton ZC, Kale DC, Elkan C, Wetzell R. Learning to diagnose with LSTM recurrent neural networks. In: International Conference on Learning Representations; 2016.

55. Atzori M, Gijsberts A, Castellini C, Caputo B, Hager AGM, Elsig S, et al. Electromyography data for non-invasive naturally-controlled robotic hand prostheses. Scientific data. 2014; 1. https://doi.org/10.1038/sdata.2014.53 PMID: 25977804

56. Amma C, Krings T, Böer J, Schultz T. Advancing Muscle-Computer Interfaces with High-Density Electromyography. In: ACM Conference on Human Factors in Computing Systems; 2015. p. 929–938.

57. Chen T, Li M, Li Y, Lin M, Wang N, Wang M, et al. MXNet: A Flexible and Efficient Machine Learning Library for Heterogeneous Distributed Systems. Statistics. 2015;.

58. Ortiz-Catalan M, Brånemark R, Håkansson B. BioPatRec: A modular research platform for the control of artificial limbs based on pattern recognition algorithms. Source code for biology and medicine. 2013; 8(1):11. https://doi.org/10.1186/1751-0473-8-11 PMID: 23597283

59. Konrad P. The ABC of EMG. A practical introduction to kinesiological electromyography. 2005; 1.

60. Englehart K, Hudgins B. A robust, real-time control scheme for multifunction myoelectric control. IEEE Transactions on Biomedical Engineering. 2003; 50(7):848–854. https://doi.org/10.1109/TBME.2003.813539 PMID: 12848352

61. Milosevic B, Benatti S, Farella E. Design challenges for wearable EMG applications. In: Design, Automation Test in Europe Conference Exhibition (DATE), 2017; 2017. p. 1432–1437.