



Research article

Motion intention recognition of the affected hand based on the sEMG and improved DenseNet network

Qunfeng Niu¹, Lei Shi¹, Yang Niu, Kunming Jia, Guangxiao Fan, Ranran Gui, Li Wang^{*}

College of Electrical Engineering, Henan University of Technology, Zhengzhou, China

ARTICLE INFO

Keywords:

sEMG
Motion intention of the affected hand
Improved DenseNet
Channel optimization
PCMIRR

ABSTRACT

The key to sEMG (surface electromyography)-based control of robotic hands is the utilization of sEMG signals from the affected hand of amputees to infer their motion intentions. With the advancements in deep learning, researchers have successfully developed viable solutions for CNN (Convolutional Neural Network)-based gesture recognition. However, most studies have primarily concentrated on utilizing sEMG data from the hands of healthy subjects, often relying on high-dimensional feature vectors obtained from a substantial number of electrodes. This approach has yielded high-performing sEMG recognition systems but has failed to consider the considerable inconvenience that the abundance of electrodes poses to the daily lives and work of patients. In this paper, we focused on transradial amputees and used sEMG data from the Ninapro DB3 database as our dataset. Firstly, we introduce a STFT (Short-Time Fourier Transform)-based time-frequency feature fusion map for sEMG. This map includes both time-frequency features and the time-frequency localization of sEMG signals. Secondly, we propose an Improved DenseNet (Dense Convolutional Network) model for recognizing motion intentions in the affected hand of amputees based on their sEMG signals. Finally, addressing the issue of optimizing the number of electrodes carried by amputees, we introduce the PCMIRR (Pearson Correlation and Motion Intention Recognition Rate) algorithm. This algorithm optimizes the number of channels by considering the Pearson correlation between the sEMG channels of amputees and the recognition rate of motion intentions in the affected hand based on single-channel sEMG data. The experimental results reveal that the recognition accuracy, recall, and F1 score achieved by the Improved DenseNet model were 93.82%, 93.61%, and 93.65%, respectively. When the number of electrodes was optimized to 8, the recognition accuracy reached 94.50%. In summary, this paper ultimately attained precise recognition of motion intentions in amputees' affected hands while utilizing the minimum number of sEMG channels. This method offers a novel approach to sEMG-based control of bionic robotic hands.

1. Introduction

According to the results of the National Sample Survey of Persons with Disabilities published in 2010, the total number of

* Corresponding author.

E-mail address: hautwangli@163.com (L. Wang).

¹ These authors contributed equally to this work.

individuals with different types of disabilities in China was 85.02 million by the end of 2010. This number represented 6.2% of the country's total population. Among these individuals, those with physical disabilities comprised the largest portion, accounting for 29.07% of the total number of people with disabilities, which amounted to 24.72 million [1]. The hand, being the most versatile limb of the human body, presents a significant challenge for individuals striving to engage in daily activities and regain functional independence in the presence of upper extremity disability [2]. The bionic robot hand is a specialized robotic system that complements or even substitutes for medical professionals in conducting patients' rehabilitation training and supporting their independent living. By incorporating robotics into the realm of clinical rehabilitation medicine, it addresses the limitations of traditional rehabilitation exercise therapy, thus paving the way for innovative approaches to upper limb rehabilitation. Consequently, bionic robotic hands offer considerable potential and advantages for various applications. However, one significant challenge in robotic hand control lies in the ability and precision of recognizing the intention behind the motion of the affected hand, as derived from the sEMG signals of amputees [3].

The sEMG (surface electromyography) is an electrical signal generated by the activity of muscle fibers. It offers valuable insights into muscle activity patterns, strength, timing, and coordination [4]. Several recent studies, including Ortiz-Catalan, M [5] and Tan, D. W. et al. [6], have demonstrated the presence of sEMG in residual muscle tissue even in cases of physical disability or amputation, unless the amputation is accompanied by nerve damage or other motor deficits. The non-invasive acquisition of this signal through surface electrodes placed on the skin over residual muscles offers valuable insights into the electrical activity generated by the remaining muscles of the upper limb. sEMG serves as a reflection of the underlying muscle activity linked to voluntary movements, serving as a window into the intricate mechanisms of neurological function that can extract information related to motion control [7].

At present, researchers combine sEMG with machine learning models to explore the field of human hand motion recognition. Classification methods for gesture recognition based on sEMG can be broadly categorized into two groups: traditional machine learning classification methods and deep learning classification methods.

Traditional machine learning classification methods refers to a class of machine learning techniques characterized by training models with the knowledge and experience of domain experts and manually selected, extracted features to assign data points to pre-defined categories or labels. These methods are usually based on statistical and mathematical models that involve feature engineering to build models for classification tasks, such as Support Vector Machine [8–11], Random Forest [9,12], Bayesian techniques [13,14] and Linear Discriminant Analysis [8,10], etc. Common feature extraction can rely on various signal processing methods, including those in the time domain, frequency domain, and time-frequency domain. However, when dealing with complex classification tasks, characterized by a larger number of gesture actions and high similarity among these actions, manually selected features prove inadequate in capturing all the pertinent information present in the sEMG signals. This limitation can result in the oversight of subtle yet valuable signals, ultimately exerting a substantial influence on the classification outcomes.

With the rapid advancement of deep learning, researchers have begun to employ deep learning techniques to accomplish human motion recognition [15,16]. The deep learning model can directly extract features from the sEMG signal, enabling hierarchical feature representation and learning within the model thanks to its end-to-end learning approach. This allows the model to effectively capture dynamic features and temporal information in the original sEMG without the need for manually selecting features. Substantial progress has been achieved in the field of gesture recognition based on HD-sEMG (High-Density surface electromyography) [17]. HD-sEMG is an electromyography acquisition technique that employs electrode arrays containing hundreds of electrodes to achieve high-density EMG data collection. Du et al. [18] used 2D-CNN (Convolutional Neural Network) to recognize sEMG from 22 gestures collected using an 8×16 electrode array, and the classification accuracy was 86.5%. However, using sparsely placed electrodes to collect multi-channel sEMG for gesture recognition remains a challenge. This could be attributed to the relatively low spatial sampling density of sparse multi-channel sEMG when compared to HD-sEMG, which may fail to capture subtle muscle activity and local features. Sparse multi-channel sEMG typically consists of a few to a dozen electrodes. Atzori et al. [19] used 1D-CNN to classify sEMG signals from 52 types of hand gestures collected using 10 electrodes, achieving a recognition accuracy of 66.59%. CNN has been widely acknowledged for their exceptional performance in analyzing visual data and extracting meaningful features for classification tasks [20,21]. The input to the model can be either the original data signal itself or an image generated from the data signal. Utilizing images derived from data signals as input to a CNN enables more effective utilization of spatial information and the application of image processing techniques to extract valuable features. More researchers convert sEMG to images as input to CNN. Cheng et al. [22] generated a multi-sEMG feature map based on 10-channel sEMG from intact subjects performing 52 gesture motions and proposed a CNN model for this feature map classification with an accuracy of 82.54% for recognition. Qureshi, M. F. et al. [23] generated Mel spectrograms based on a 6-channel sEMG dataset from 8 intact subjects performing 11 gesture motions, repeated daily for 7 days, and proposed a CNN model for the classification of Mel Spectrograms, achieving a recognition accuracy of 99.42%. Qureshi, M. F. et al. [24] generated Log-Mel Spectrograms based on a 10-channel sEMG dataset from intact subjects performing 10 gesture motions and proposed an E2CNN model for the classification of Log-Mel Spectrograms, achieving a recognition accuracy of 91.27%. Zhang et al. [25] generated Hilbert graphs from 10-channel sEMG of 52 gesture actions by the Hilbert transform and proposed a dual-view multi-scale convolutional neural network for Hilbert graph classification with recognition accuracy of 86.72%.

Currently, despite the research on gesture recognition using deep learning methods being well-established with promising recognition performance, the majority of researchers have focused on complete subjects' hand sEMG data. There is relatively less research on recognizing hand motion intentions based on the weak and incomplete surface electromyographic signals from amputees. The quality of life for amputees could experience significant improvement if the sEMG signals from the affected hand can be accurately identified and applied to control a bionic robot. Moreover, in the field of designing high-performance sEMG-based gesture recognition systems, researchers often rely on high-dimensional feature vectors obtained from multiple electrodes to classify more complex hand motion commands with greater accuracy. However, they tend to overlook the inconvenience that an excessive number of electrodes

can pose to amputees during rehabilitation and daily life. To address these challenges, this paper introduces an Improved DenseNet model for accurately identifying the motion intentions of the affected hand based on the sEMG signals from amputees. Additionally, it proposes a PCMIRR-based channel number optimization algorithm. The combined approach achieves precise motion intention identification for the affected hand of amputees while using the minimum number of sEMG electrodes. In summary, the innovations presented in this paper can be summarized as follows.

- A sEMG time-frequency feature fusion map was introduced, which is based on STFT (Short-Time Fourier Transform) and consolidates information from both the time and frequency domains of sEMG. Subsequently, an extensive dataset was created, comprising 20,524 sEMG time-frequency feature fusion maps, utilizing the sEMG data from 17 motions performed by 8 transradial amputees sourced from the NinaproDB3 library. To ensure its applicability in practical scenarios, effective measures were taken to prevent overfitting.

Table 1
17 basic finger and wrist motions [27].

Basic movements of the figure								
1	Thumb up		2	Extension of index and middle, flexion of the others		3	Flexion of ring and little finger, extension of the others	
4	Thumb opposing base of little Finger		5	Abduction of all fingers		6	Fingers flexed together in fist	
7	Pointing index		8	Adduction of extended fingers				
Basic movements of the wrist								
9	Wrist supination (axis: middle finger)		10	Wrist pronation (axis: middle finger)		11	Wrist supination (axis: little finger)	
12	Wrist pronation (axis: little finger)		13	Wrist flexion		14	Wrist extension	
15	Wrist radial deviation		16	Wrist ulnar deviation		17	Wrist extension with closed hand	

- An Improved DenseNet (Dense Convolutional Network) model was proposed for the recognition of motion intention in the affected hand based on the sEMG of amputees. The model utilizes DenseNet121 as the baseline architecture and introduces the Multi-Scale structure and DANet (Dual Attention Network) module to enrich feature information in the image. This augmentation enhances the significance of effective features within the input image by incorporating network branches, resulting in improved network performance.
- A channel number optimization algorithm based on PCMIRR (Pearson Correlation and Motion Intention Recognition Rate) was introduced by correlating the Pearson correlation between sEMG channels of amputees with the motion intention recognition rate of the affected hand using single-channel sEMG data. This algorithm simplifies the number of sEMG channels while enhancing the accuracy of the Improved DenseNet model for the sEMG time-frequency feature fusion map dataset based on the amputee's affected hand sEMG. To some extent, it mitigates the reduction in the quality of life for amputees due to the excessive number of sensors they need to wear.

This paper is structured as follows. Section 2 presents the dataset, network architecture, and research methodology. Section 3 discusses the experimental platform, network evaluation metrics, experiments, and experimental results. Finally, Section 4 offers a discussion, and Section 5 presents the conclusion.

2. Materials and methods

2.1. Dataset

Ninapro is a publicly accessible multimodal database specifically created to support research on human, robotic, and prosthetic hands, as well as machine learning-based control systems [26]. The database comprises over 300 data collections divided into 10 datasets. The goal of this paper is to capture the motion intention of the affected hand in individuals with one-arm disabilities using sEMG. Hence, this paper focuses on 17 fundamental finger and wrist movements (Table 1) executed by eight transradial amputees (Table 2) sourced from the Ninapro DB3 dataset [27].

The dataset uses 12 active double differential wireless electrodes from a Delsys Trigno Wireless EMG system. Electrode placement: Eight electrodes were evenly distributed at the level of the radial-humeral joint on the forearm. Additionally, two electrodes were positioned at the key activity points of the superficial finger flexors and superficial finger extensors, while another two electrodes were situated at the primary activity points of the biceps and triceps. During data collection, participants were instructed to perform these motions with their affected hand. Each repetition of a motion lasted for 5 s, followed by a 3-s rest period, with a total of six repetitions.

Table 1 provides details of the 17 fundamental finger and wrist movements. In Table 2, clinical characteristics of the eight subjects are summarized, including handedness, amputated hand, remaining forearm, years since amputation, Phantom Limb Sensation (ranging from 0 to 5, where 0 signifies no sensation, and 5 denotes strong sensation), and DASH score [28] (ranging from 0 to 100, where 0 indicates no dysfunction or pain, and 100 represents the most severe dysfunction and pain).

2.2. Dataset processing

2.2.1. Filtering and normalization

Because of the relatively low intensity of sEMG signals in the human forearm, the sEMG recorded from amputees tends to be even weaker than that of individuals without amputation. Additionally, the sEMG acquisition process is susceptible to environmental influences, resulting in the inclusion of various types of noise in the final sEMG data, such as power frequency noise and baseline drift. These noise factors significantly hinder the extraction and identification of essential information within the original signal [29–31]. The DB3 dataset has been preliminarily preprocessed for sEMG, including signal full-wave rectification, synchronization and retagging. Thus, only high-frequency noise from the acquisition device needs to be considered. The electrical activity of the muscles was smoothed using a first-order low-pass Butterworth filter with a cutoff frequency of 1 Hz to eliminate the underlying high-frequency noise from the signal [19,26]. Subsequently, Minmax normalization was applied to the filtered sEMG signal to enhance the extraction of signal features.

Table 2

Clinical characteristics of the 8 transradial amputees [27].

Number	Handedness	Amputated Hand	Remaining Forearm (%)	Years since Amputation	Phantom Limb Sensation	DASH Score
1	R	R	50	13	2	1.67
2	R	L	70	6	5	15.18
3	R	R	30	5	2	22.50
4	L	L	90	1	2	11.67
5	R	R	50	5	2	33.33
6	R	R	90	14	5	3.33
7	R	R	50	2	5	11.67
8	R	R	90	5	4	12.50

2.2.2. STFT-based fusion map of time-frequency features of sEMG

The sEMG data from eight transradial amputees in the Ninapro DB3 database used in this paper consists of sparse multi-channel sEMG. This type of sEMG, due to its limited number of electrodes and sparse arrangement compared to HD-sEMG signals, cannot comprehensively cover the muscle surface to provide extensive information. Additionally, amputees' sEMG signals are characterized by lower signal intensity and a vestigial nature when compared to those of healthy individuals. Hence, this paper augments the time domain EMG of the residual limb EMG with frequency domain information to incorporate additional frequency characteristics and temporal variations of the EMG. This enhancement allows for a richer representation of the EMG, which reflects the intention of motion in the affected hand.

Given the non-smooth, faint, and vestigial nature of amputees' sEMG signals, it becomes challenging to determine when frequency components emerge using the FFT (Fast Fourier Transform). In contrast, the STFT offers the ability to capture the frequency domain information of the signal within specific time segments by employing a sliding window to divide the signal into smaller segments and calculating the Fourier transform for each segment individually. Hence, this paper introduces a STFT-based time-frequency feature fusion map of sEMG. This fusion map encompasses both time-frequency characteristics and the time-frequency localization of the amputee sEMG. The steps involved in creating and implementing this STFT-based fusion map of time-frequency features from sEMG signals are detailed as follows.

Firstly, the preprocessed sEMG data (following filtering and normalization) is segmented using a sliding window of 256 ms in size. In order to enhance real-time performance and increase the number of images for model training, this paper shifts the sliding window with a time interval of 184 ms. This process is iteratively applied throughout the entire length of the signal. For illustration, Fig. 1 depicts the division of the sEMG time domain signal using a sliding window. It is demonstrated using a segment of sEMG data from channel 1 of subject 1 performing action 1 (thumb up) as an example.

Secondly, the DFT (Discrete Fourier Transform), which is one of the forms of Fourier Transform, is applied to each processing window, and the formula is shown in (1).

$$Y(k) = \sum_{n=1}^N \left[X(n) e^{\left(\frac{-j2\pi kn}{N} \right)} \right] \quad (1)$$

where n denotes the index of the time series, which N represents the number of sampling points, k denotes the index of the frequency domain series ($k \in (0, N-1)$), j is an imaginary unit and e is the base of the natural logarithm.

Finally, for each sliding window, the DFT operation generates sEMG and its corresponding frequency domain signals, creating a LTFU (Local Time-Frequency Unit). This unit provides comprehensive information regarding both the time-frequency characteristics and the localization of these characteristics within the signal. Following this, all the LTFUs are gathered, and the time and frequency signals of sEMG are separately transformed into grayscale intensity. The transformation is then linearly applied to create grayscale images, resulting in the sEMG time domain feature map, sEMG frequency domain real part feature map, and sEMG frequency domain imaginary part feature map. These three feature maps are subsequently combined to generate the sEMG time-frequency feature fusion map, where each feature map corresponds to one channel of the RGB image.

Fig. 2 illustrates the sEMG time domain feature map (a), sEMG frequency domain real part feature map (b), sEMG frequency domain imaginary part feature map (c), and sEMG time-frequency feature fusion map (d) for subject 1's action 1 (thumb up) using a single sliding window. These four plots share the same dimensions, and the resolution of this plot is indicated as $C \times W$, where C represents the number of EMG sensors (channels), and W denotes the current sliding window length. In this case, with 12 electrodes, $C = 12$, and a window size of 256 ms results in $W = 512$, given a signal frequency of 2 kHz.

The sEMG data for 17 actions performed by the affected hands of 8 transradial amputees from the Ninapro DB3 database were used

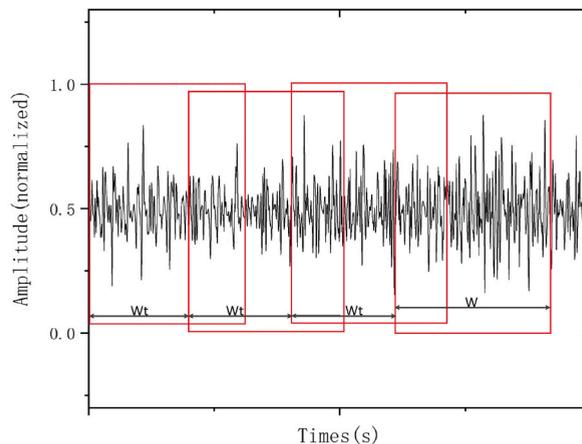


Fig. 1. Segmentation of single channel sEMG.

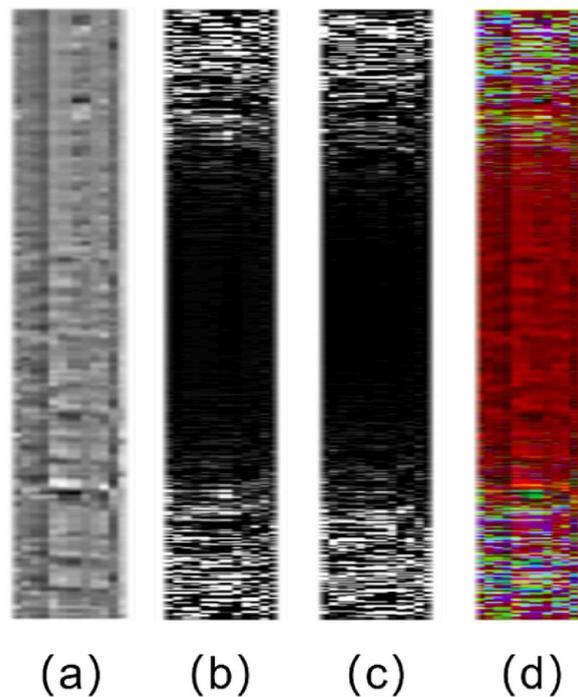


Fig. 2. 12-channel sEMG feature map of an action under 1 sliding window: (a) sEMG time domain feature map; (b) sEMG frequency domain real part feature map; (c) sEMG frequency domain imaginary part feature map; (d) sEMG time-frequency feature fusion map.

to generate a dataset of 20,524 sEMG time-frequency feature fusion maps, following the steps outlined above. This dataset was utilized in this paper to study the recognition of motion intentions in the affected hands of amputees. For the dataset, 16,428 images were randomly selected for the training set, while the remaining 4096 images were allocated to the test set, maintaining an 8:2 ratio. Detailed information about this dataset is provided in [Table 3](#).

2.3. Model building

2.3.1. Improved DenseNet modeling framework

The DenseNet was introduced by Huang et al., in 2017 [32]. This network model has proven effective due to its small number of parameters and feature reuse, particularly on datasets with weak features and complex categories. The sEMG time-frequency feature fusion dataset of amputees studied in this paper presents weak features and complex categories due to the low sEMG signal strength, numerous gesture types, and residual phenomena. In response, this paper introduces an Improved DenseNet model, which enhances DenseNet121 by incorporating a Multi-Scale structure. This modification improves the model's ability to capture shallow edge information and deep semantic information from the sEMG of amputees' affected hands. Additionally, a DANet module is introduced to optimize the model's attention mechanism, both between channels in the multilayer feature maps and within each layer of the feature maps. The overall framework of the Improved DenseNet model is depicted in [Fig. 3](#).

As seen in [Fig. 3](#), the sEMG time-frequency feature fusion map generated from the 12-channel sEMG of the amputees is used as input of the model. Layer 1 of the model is a convolutional layer with a convolutional kernel size of 7×7 for extracting features of the input image; Layer 2 is a DANet module for augmenting the model's weights for each channel of the sEMG time-frequency feature fusion map and each location of each channel that benefits from the identification of the resulting features; Layer 3 is a maximum pool layer with a size of 3×3 for reducing the spatial size of the feature map. This is followed by 4 Dense Blocks of the Improved DenseNet model, each

Table 3

The sEMG time-frequency feature fusion map dataset.

Gesture number	1	2	3	4	5	6	7	8	9
Training set	1039	1067	987	1018	954	898	1152	1079	1114
Test set	259	266	246	254	238	224	288	269	278
Total	1298	1333	1233	1272	1192	1122	1440	1348	1392
Gesture number	10	11	12	13	14	15	16	17	
training set	811	1084	860	872	859	956	922	756	
test set	202	271	214	217	214	238	230	188	
Total	1013	1355	1074	1089	1073	1194	1152	944	

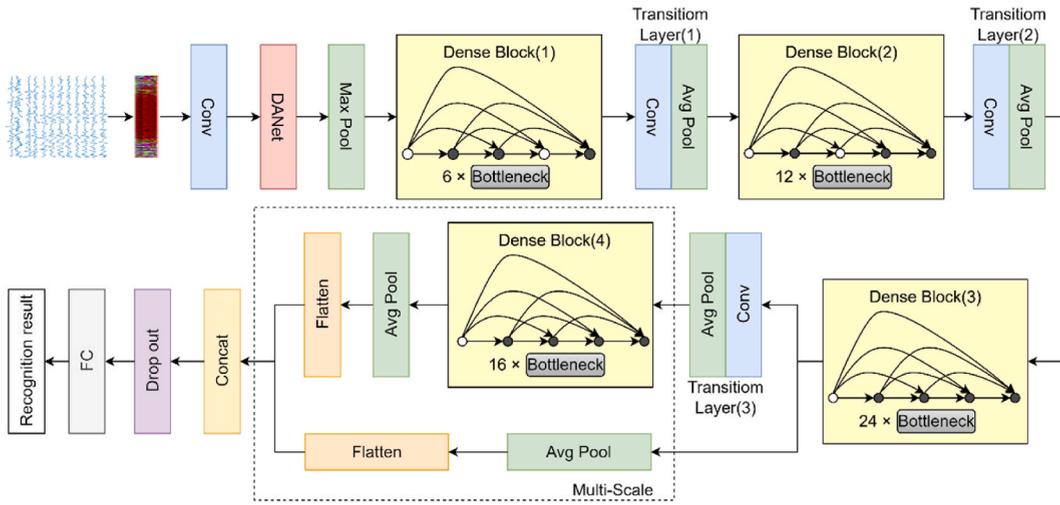


Fig. 3. Improved DenseNet overall model framework.

Dense Block (a) consists of multiple bottlenecks (the bottleneck contains 1 convolutional layer with 1×1 convolutional kernel size and 1 convolutional layer with 3×3 convolutional kernel size, see (b), which are densely connected in such a way that each convolutional layer can receive the feature maps of all previous layers as input, increasing feature reuse and information flow (Fig. 4). Between each Dense Block, the number of channels and the spatial size of the feature map are reduced by using the transition layer, which consists of a convolutional layer and an average pool layer for reducing the dimensionality and compressing the size of the feature map. Another branch is added after the third Dense Block, and its output features are passed through an average pool layer and a flatten layer, respectively, with the output features of the fourth Dense Block, and then the concat layer is used for feature stitching for output. Finally, the final result of the gesture action classification is obtained after the fully connected layer.

2.3.2. Multi-scale of improved DenseNet

The sEMG time-frequency feature fusion map utilized as input for the network model in this paper is a pixel image derived from the sEMG data of amputees’ affected hands. Unlike conventional images with readily interpretable features, pixel images pose challenges in direct interpretation. In this case, the convolutional neural network model must account for both shallow features, such as color, texture, and local edges of the sEMG time-frequency feature fusion map, as well as deeper features encompassing semantic information and local details. This holistic approach enhances the utilization of feature information from the sEMG time-frequency feature fusion map.

To tackle this challenge, an additional branch was incorporated into Dense Block3, building upon the DenseNet121 model’s architecture (Fig. 3). The output from Dense Block3 and Dense Block 4 was routed through an average pooling layer and a flatten layer, respectively. Subsequently, the two sets of features were integrated through the concat layer. Finally, the combined feature information traversed the fully connected layer to achieve motion intention recognition in amputees’ affected hands based on their sEMG data, as illustrated in Fig. 5 (a) and (b). This design preserves the benefits of feature reuse and information flow inherent in Dense Blocks, while simultaneously strengthening low-level features, retaining finer details, and enhancing the model’s overall representation and performance.

2.3.3. DANet of improved DenseNet

In the case of multichannel images, the convolutional neural networks typically assign uniform weights to the feature maps of each

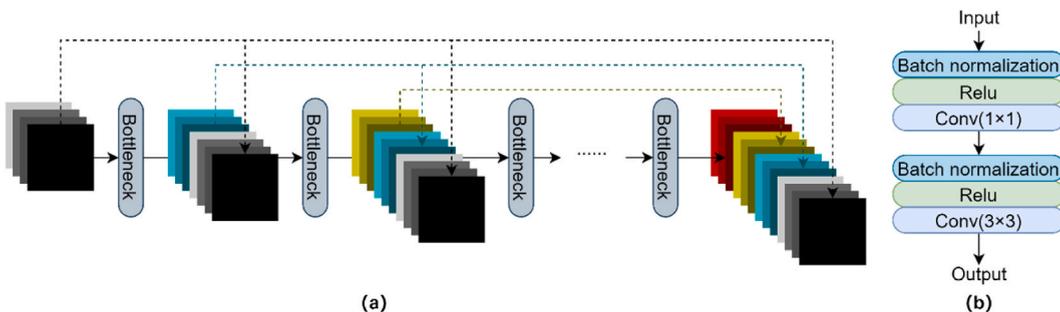


Fig. 4. (a) Structure of Dense Block; (b) Structure of the bottleneck.

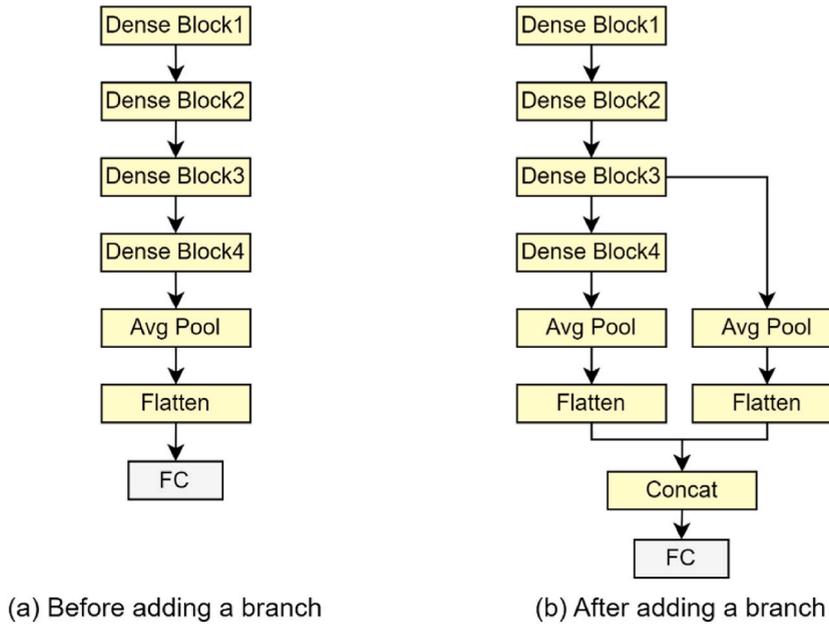


Fig. 5. Add additional branches to Dense Block 3 based on DenseNet 121: (a) Before adding a branch; (b) After adding a branch.

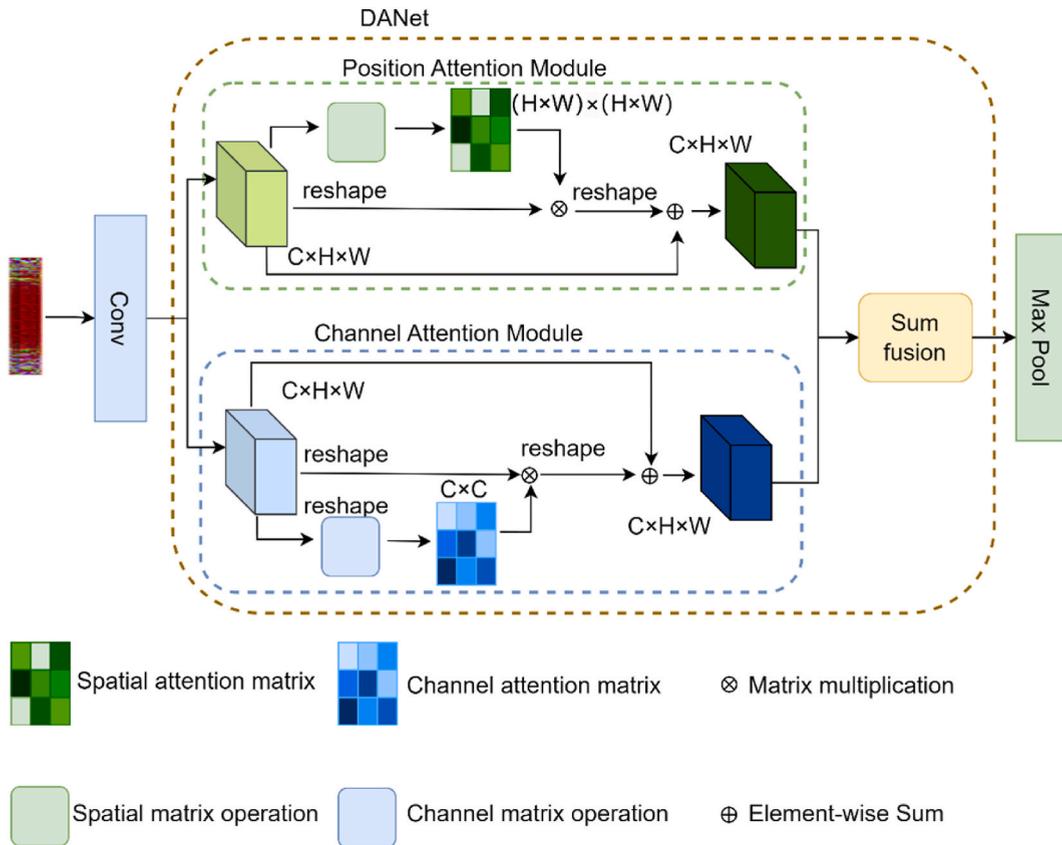


Fig. 6. The location of DANet in Improved DenseNet and the internal structure of DANet.

channel within the image. However, the sEMG time-frequency feature fusion map employed as input for Densenet121 in this paper is a 3-channel RGB map. It comprises one layer of time domain information and two layers of frequency domain information, generated by segmenting the sEMG data of a complete gesture action using a sliding window. Importantly, the time domain and frequency domain data within the sEMG time-frequency feature fusion map exhibit varying degrees of dynamism across different sliding windows. If the model network were to assign identical weights to the feature maps of all three channels, it would not adequately capture these dynamic time-frequency features. Furthermore, the sEMG data used for the sEMG time-frequency feature fusion map originates from amputees and exhibits a residual phenomenon. This phenomenon can lead to discrepancies in the feature information extracted by the model network for corresponding positions in each layer of the image's channels, potentially impacting the ultimate recognition accuracy. To address these two issues, this paper introduces the DANet module [33], which is built upon the Densenet121 model. The DANet module comprises two vital components:

CAM (Channel Attention Module): The CAM assigns optimized weights to the feature maps of each channel within the sEMG time-frequency feature fusion map. This ensures that both time domain and frequency domain information is more effectively harnessed, accommodating the dynamic nature of the time-frequency feature fusion map.

PAM (Position Attention Module): The PAM assigns weights to the feature information at each location across the feature maps within each channel layer of the sEMG time-frequency feature fusion map. This strategic allocation of weights ensures that the model network focuses its attention on valuable feature information within each channel layer, mitigating the impact of the residual phenomenon.

These modifications collectively enhance the network's ability to extract and utilize dynamic time-frequency feature information while maintaining focus on critical features for improved accuracy.

Fig. 6 illustrates the placement of the DANet module within the Improved DenseNet model, providing insight into its internal structure. This figure reveals that the DANet module incorporates two parallel structures, CAM and PAM, allowing it to simultaneously optimize features for a given image. Subsequently, it aggregates and outputs the two sets of optimized features. Additionally, Fig. 6 shows that the DANet module is integrated after the first convolutional layer of the Improved DenseNet model and before the maximum pool layer. This strategic positioning empowers the DANet module to adjust the weights of each channel within each layer of the input feature map, as well as the feature weights of each position within each channel of each layer, directly from the source. Importantly, the input and output feature maps of the DANet module maintain the same size, ensuring that they do not interfere with feature extraction in subsequent layers.

2.4. PCMIRR-based channel number optimization algorithm

In the study, it was observed that in the Ninapro DB3 dataset, subjects 6 and 7 did not use the two electrodes placed at the primary active sites of finger flexors and finger extensors due to arm disability. This finding underscores the notion that excessive electrodes can inconvenience amputees in their daily lives. As a solution, this paper introduces a channel number optimization algorithm based on PCMIRR. This algorithm combines the Pearson correlation between amputee sEMG channels with the motion intention recognition rate of the affected hand, utilizing single-channel sEMG data from amputees.

The PC (Pearson Correlation) was proposed by Karl Pearson in 1896. PC is a method that leans towards the 'Hard' approach. It is employed to quantify the degree of correlation between two variables and is commonly used in the field of machine learning to compute the similarity between features. The specific formula for calculation is shown in (2).

$$r_{A,B} = \frac{\sum_{i=1}^n (a_i - \bar{A})(b_i - \bar{B})}{n\sigma_A\sigma_B} = \frac{\sum_{i=1}^n (a_i b_i) - n\bar{A}\bar{B}}{n\sigma_A\sigma_B} \quad (2)$$

where n denotes the number of samples, a_i denotes the i -th element in the feature vector A , and b_i denotes the i -th element in the feature vector B . \bar{A} , \bar{B} are the mean values of vectors A and B , respectively. σ_A and σ_B denote the standard deviation of vectors A and B , respectively, while $\sum_{i=1}^n a_i b_i$ is the sum of the fork products of A and B . $r_{A,B}$ is between -1 and $+1$. When $r_{A,B} \geq 0$, there is a positive correlation between A and B . The larger the correlation coefficient, the stronger the correlation. A larger $r_{A,B}$ indicates that one of A or B can be removed as redundancy. When $r_{A,B} = 0$, there is no linear correlation between A and B . When $r_{A,B} \leq 0$, there is a negative correlation between A and B , and B will decrease as A increases. In statistics and analysis, it is necessary not only to know whether the variables are correlated with each other, but also the degree of correlation. Usually, the absolute value of the correlation coefficient and the strength of correlation between variables exist as shown in Table 4.

The channel optimization algorithm based on Pearson correlation proposed in this paper proceeds as follows.

Table 4
Correlation intensity table.

Correlation coefficient	Correlation Strength
0.8–1	Extremely strong correlation
0.6–0.8	Strong correlation
0.4–0.6	Medium correlation
0.2–0.4	Weak correlation
0.0–0.2	Extremely weak or no correlation

The EMG signals obtained from the 1st experiment of the 1st movement of subject 1 were expressed as X . The specific formula for calculation is shown in (3) and (4).

$$X = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,j} & \cdots & x_{1,C} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,j} & \cdots & x_{2,C} \\ \vdots & \vdots & \ddots & x_{i,j} & \ddots & \vdots \\ x_{T,1} & x_{T,2} & \cdots & x_{T,j} & \cdots & x_{T,C} \end{pmatrix}^{T \times C} \quad (3)$$

$$X = [X_1, X_2, \dots, X_K, \dots, X_C] \quad (4)$$

where T denotes the temporal sample point of the original sEMG, C denotes the number of electrodes (channels) used to acquire EMG ($C = 12$, $T \gg C$), and $x_{i,j}$ denotes the sample element of the sEMG signal acquired at the i -th time point, j -th channel. The EMG signal X_K in each channel is considered a feature, which is a one-dimensional vector, and the correlation coefficient between two channels m and n is $r_{m,n}$ according to the definition of Pearson's correlation coefficient. The specific formula for calculation is shown in (5).

$$r_{m,n} = \frac{\sum_{m,n=1}^C (X_m - \bar{X}_m)(X_n - \bar{X}_n)}{\sqrt{\sum_{m=1}^C (X_m - \bar{X}_m)^2} \sqrt{\sum_{n=1}^C (X_n - \bar{X}_n)^2}} \quad (5)$$

where \bar{X}_m , \bar{X}_n are the mean values of the m -th and n -th channels in the time dimension, respectively, and the correlation matrix between channels can be defined as R . The specific formula for calculation is shown in (6).

$$R = \begin{pmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,j} & \cdots & r_{1,C} \\ r_{2,1} & r_{2,2} & \cdots & r_{2,j} & \cdots & r_{2,C} \\ \vdots & \vdots & \ddots & r_{i,j} & \ddots & \vdots \\ r_{C,1} & r_{C,2} & \cdots & r_{C,j} & \cdots & r_{C,C} \end{pmatrix}^{C \times C} \quad (6)$$

where $r_{m,n}$ are the correlation coefficients between the channels m -th and n -th. $r_{m,n} = r_{n,m}$. The correlation coefficient matrices corresponding to the 6 repetitions of the experiment of 17 movements of 8 subjects, totaling 816, were derived separately. These 816 correlation matrices were averaged to obtain the mean Pearson correlation matrix \bar{R} . The specific formula for calculation is shown in (7).

$$\bar{R} = \frac{1}{L \times N \times M} \sum_{a=1}^L \sum_{b=1}^M \sum_{c=1}^N R_{a,b,c} \quad (7)$$

where L denotes the number of subjects, N denotes the number of actions, M denotes the number of repeated experiments, and $R_{a,b,c}$ denotes the Pearson correlation coefficient matrix for the c -th experiment of the b -th action of the a -th subject. The Pearson correlation heat map is generated on the basis of the final Pearson correlation matrix \bar{R} obtained.

According to the correlation strength table presented in Table 4, when $r_{m,n} \geq 0.6$, there is a relatively strong correlation between the m -th and n -th channels. The final optimized number of channels in this study was selected from those corresponding to the correlation coefficients $r_{m,n}$ in the range of 0.6–1. However, it's important to note that $r_{m,n}$ represents a pair of channels, and once $r_{m,n}$ is determined, the specific channel numbers to be optimized need to be established. This paper determined the channel numbers to be removed based on the motion intention recognition rate of the affected hand using single-channel sEMG data from amputees. The 12 sEMG channels were sequentially retained, while other channel data were set to 0. This resulted in the creation of 12 individual single-channel time-frequency feature fusion datasets, each representing a specific channel. These 12 single-channel sEMG time-frequency feature fusion datasets were individually input into the Improved DenseNet model to obtain 12 single-channel motion intention recognition rates for the affected hand based on single-channel sEMG data from amputees.

A higher single-channel motion intention recognition rate indicates that the channel contributes more significantly to the final recognition of the motion intention of the affected hand in amputees, while a lower rate signifies a less significant contribution. To determine the channel for optimization, the individual recognition rates for the two channels of interest are compared. The channel with the lower single-channel motion intention recognition rate is the one selected for optimization. Subsequently, the time-frequency feature fusion map is generated by progressively reducing one channel from the 12 channels sEMG. The sEMG time-frequency feature fusion maps, with varying numbers of optimized channels, are sequentially input into the Improved DenseNet network model. This process enables the calculation of recognition rates for sEMG time-frequency feature fusion maps with different channel optimization configurations. By comparing these recognition rates, the optimal number of channels is selected as the final configuration.

3. Experiment and result

3.1. Experimental setup

3.1.1. Experimental platform

The experiments in this paper were based on the Windows 10 operating system. The GPU used was the GeForce GTX 3080 GPU (10 GB graphics memory), the processor used was the Intel Core i7-12700K CPU@3.61GHz, and the running memory was 64 GB. The development environment for creating sEMG time-frequency feature fusion maps used Matlab2020b. The development environment for model building and training testing used Pycharm, the deep learning framework used PyTorch, the parallel computing framework used CUDA version 11.3, and the Python version used is 3.8.

3.1.2. Network evaluation

The evaluation metrics used in this paper are the calculated accuracy ACC, precision P, recall R, and F1 score [34]. Specific formulas are given in (8) to (12).

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

$$P = \frac{TP}{TP + FP} \quad (9)$$

$$R = \frac{TP}{TP + FN} \quad (10)$$

$$F1 = 2 \times \frac{P \times R}{P + R} \quad (11)$$

$$Avg\ index = \frac{\sum_{j=1}^{17} i_j}{\sum_{j=1}^{17} N_j} \quad (12)$$

in which.

ACC	Accuracy rate	The ratio of the number of samples correctly predicted by the model to the total number of samples.
P	Precision rate	The ratio of the number of samples correctly predicted by the model as positive cases to the total number of samples predicted as positive cases among all samples classified as positive cases.
R	Recall rate	The ratio of the number of samples in which the model correctly predicts a positive case to the total number of actual positive cases in all actual positive cases samples.
TP	True Positive	The number of samples correctly predicted by the model as positive cases.
TN	True Negative	The number of samples correctly predicted by the model as counterexamples.
FP	False Positive	The number of samples incorrectly predicted by the model as positive cases.
FN	False Negative	The number of samples incorrectly predicted by the model as counterexamples.
F1	F1 score	The F1 score is a combined Precision and Recall metric, and it is the summed average of Precision and Recall.
i_j		An evaluation indicator.
N_j		Total number of samples for a given action type.

In addition to the aforementioned performance metrics, time complexity is also a crucial factor. Therefore, this paper includes T-time (training time) and P-time (prediction time) in the evaluation criteria in the model improvement experiments. T-time represents the total time required for training over all epochs, while P-time indicates the time taken to predict a single image. Furthermore, the Kruskal-Wallis test is employed to evaluate comparisons with previous studies on the Ninapro DB3 dataset, taking into account variations in the number of classes. Any p-values below 0.05 are considered statistically significant.

3.1.3. Experiment details and parameter settings

For the experiments, the order of the images in the training set was randomly arranged, and the resolution of the input images was adjusted to 224×244 . Subsequently, each channel of the input images was normalized according to the mean value of 0.5 and the standard deviation of 0.5.

SGD (Stochastic Gradient Descent) was chosen as the optimizer for the experimental model with an initial lr (learning rate) of 0.002, an initial lrf (learning rate factor) of 0.1, a momentum setting of 0.9, and a weight decay of 10^{-4} . The NAG (Nesterov accelerated gradient) algorithm is used to accelerate the update of model parameters. Subsequently, a cosine function-based learning rate

scheduler is designed, which will update lrf between 0.1 and 1 according to the maximum number of iterations epoch of the current training and increase linearly to 1 with the training process to implement the cosine annealing learning rate scheduling strategy. Set the batch size to 32 and the maximum number of iterations epochs to 60. After each iteration, the accuracy of the model is evaluated using a test set, and the results of each evaluation and the model parameters are kept. In the subsequent model improvement experiments, the parameters remain unchanged.

3.2. Result

3.2.1. Performance test of Densenet121

This section evaluated the model performance of DenseNet121 for classifying sEMG time-frequency feature fusion map datasets. This paper chose the baseline CNN: ResNet50 [35], GoogLeNet [36], MobileNetV2 [37], ShuffleNetV2 [38], RegNet [39] and DenseNet121 [32].

Table 5 and Fig. 7 present a performance comparison between DenseNet121 and other baseline convolutional neural network models using the sEMG time-frequency feature fusion map dataset. It is evident that RegNet demonstrates the lowest accuracy at 63.50%. The baseline model chosen for this study, DenseNet121, exhibits the highest accuracy at 89.62%. Following that, the performance ranking is as follows: ResNet50, ShuffleNetV2, MobileNetV2, and GoogleNet with accuracy rates of 87.26%, 80.76%, 80.18%, and 68.77%, respectively. In summary, DenseNet121 stands out as the top-performing baseline model in this research.

3.2.2. Network evaluation of improved DenseNet

In this section, the DenseNet121 model equipped with the Multi-Scale structure and DANet module, as described in Sections 2.3.2 and 2.3.3, is assessed using a dataset comprising sEMG time-frequency feature fusion maps. The performance metrics, as well as the confusion matrix for the Improved DenseNet, are presented in Table 6 and Fig. 9, respectively.

Among these options, there are three choices for adding the Multi-Scale structure to DenseNet121 (a): (b) add branches to Dense Block1~Dense Block3 respectively; (c) add branches to Dense Block2~Dense Block3 respectively; (d) add branche to Dense Block3 (Fig. 8). As indicated in Table 6, the addition of a branch in Dense Block 3 and the concatenation of the output features from Dense Block 3 and Dense Block 4 for the final output yielded the optimal experimental results. This modification increased the network recognition accuracy by 2.93%, while the T time increased by only 0.17%.

After selecting the optimal Multi-Scale structure and incorporating the DANet module, the resulting MS-DA-DenseNet model improved recognition accuracy by 1.27% compared to the MS-DenseNet model. However, there was an 11.61% increase in T-time. This demonstrates that adjusting the weights of different channels for multi-layer feature maps and different position features within each layer channel of the model can be beneficial in enhancing network performance. However, it also incurs an increase in time cost. The MS-DA-DenseNet model represents the Improved DenseNet model proposed in this paper. In the end, Improved DenseNet achieved a 4.2% increase in accuracy compared to DenseNet121, reaching 93.82%, a 4.31% increase in recall, reaching 93.61%, and an F1 score of 93.65%. Simultaneously, there was an 11.81% increase in T-time, and the P-time experienced a 6.54% increment.

3.2.3. Comparison of results with previous methods

To further demonstrate the performance of the DA-MS-DenseNet model proposed in this paper, a comparison with state-of-the-art gesture recognition techniques was performed. Table 7 summarizes the experimental results from earlier studies as well as those obtained using our method on Ninapro DB3.

The table indicates that, compared to the average accuracy of recognizing 10 actions by CNN [40] and CNN [42], DA-MS-DenseNet achieved an improvement of 20.51% and 23.82%, respectively, in average accuracy when recognizing 17 actions.

3.2.4. Evaluation of channel number optimization

In accordance with section 2.4, a heatmap of the Pearson correlation coefficients based on the sEMG of amputees was generated. Fig. 10 provides an overview, with lighter areas indicating weak correlations between channels and darker areas representing strong correlations. The horizontal and vertical axes represent the 12 sEMG channels. Based on Fig. 10, the correlation channels with correlation coefficients ($r_{m,n}$) greater than or equal to 0.6 were identified in descending order: 5–6, 1–3, 8–9, 4–12, 6–7, 11–12, 9–10, 7–10, 2–4, 4–6, 4–5, 5–7, and 1–8, totaling 12 pairs.

Based on section 2.4, single-channel sEMG time-frequency feature fusion maps were generated, and they were individually fed into

Table 5
Comparison of performance metrics of CNN baseline models.

Model	Accuracy	Avg Precision	Avg Recall	Avg F1 Score
ResNet50	87.26%	87.48%	86.94%	87.02%
GoogleNet	68.77%	69.46%	69.39%	68.59%
MobileNetV2	80.18%	80.74%	79.92%	79.95%
ShuffleNetV2	80.76%	80.59%	80.54%	80.53%
RegNet	63.50%	63.16%	63.05%	63.00%
DenseNet121	89.62%	89.53%	89.30%	89.36%

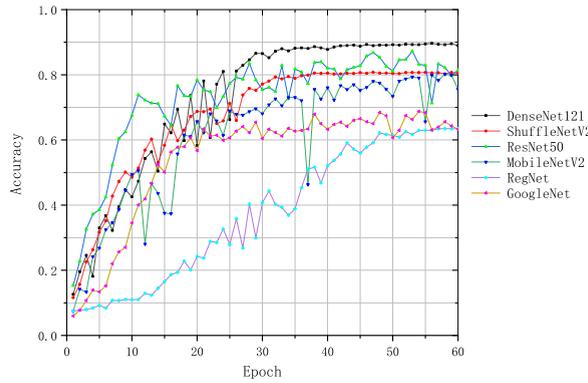


Fig. 7. Accuracy of different CNN baseline models trained of 60 rounds.

Table 6
Comparison of performance metrics of different DenseNet121 improvements.

Model	Accuracy	Avg Precision	Avg Recall	Avg F1 Score	T-time	P-time
DenseNet121 (a)	89.62%	89.53%	89.30%	89.36%	4615s	3.21 ms
MS(a)-DenseNet (b)	89.89%	90.03%	89.75%	89.78%	4701s	3.29 ms
MS(b)-DenseNet (c)	90.92%	90.91%	90.71%	90.71%	4650s	3.26 ms
MS(c)-DenseNet (d)	92.55%	92.61%	92.32%	92.41%	4623s	3.23 ms
MS(c)-DA-DenseNet (e)	93.82%	93.74%	93.61%	93.65%	5160s	3.42 ms

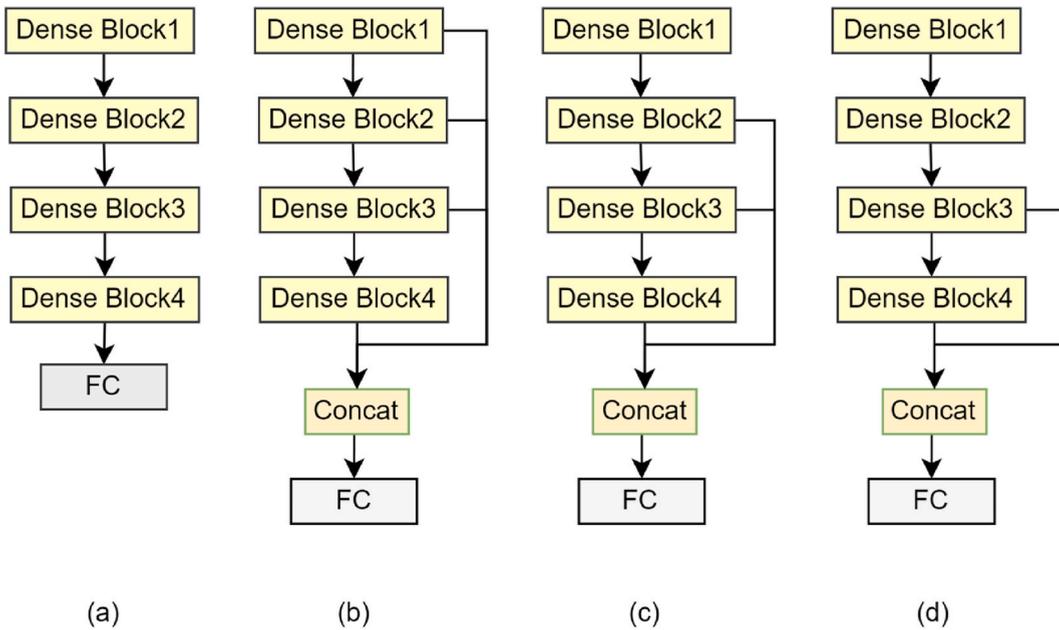
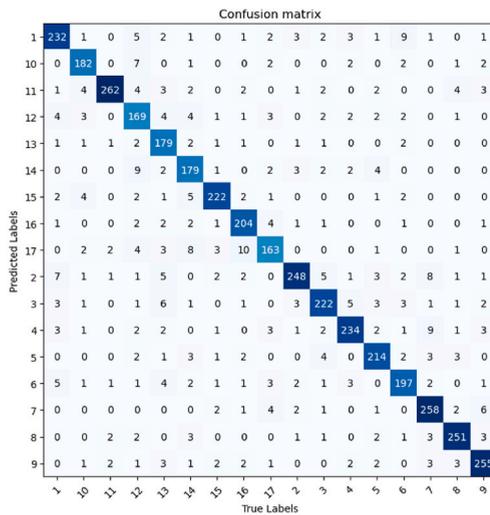
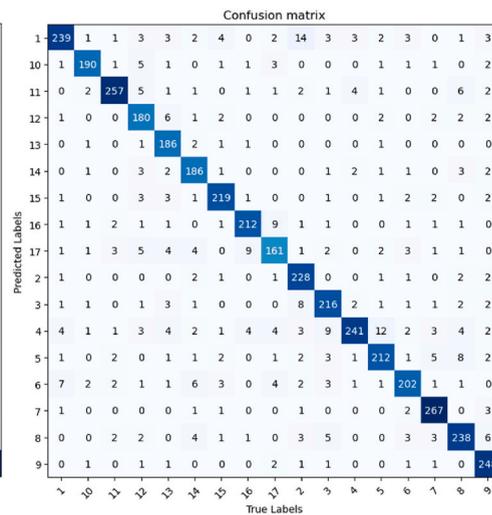


Fig. 8. (a) Multi-Scale structure is not added; (b) Dense Block1~Dense Block3 are added to branches separately; (c) Dense Block2~Dense Block3 are added to branches separately; (d) Dense Block3 is added to branch.

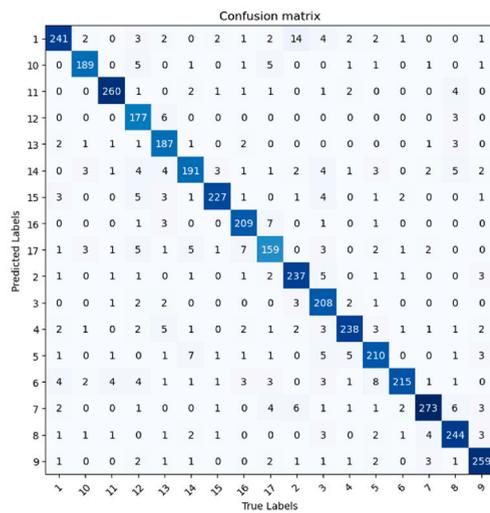
the Improved DenseNet network model to obtain recognition rates for single-channel sEMG (Table 8). From Tables 8 and it can be observed that the recognition rate for Channel 5 is higher than that for Channel 6, indicating that the first channel to be optimized is Channel 6. Optimized sEMG time-frequency feature fusion maps were generated and fed into the Improved DenseNet network model to obtain recognition rates for the affected hand’s motion intention based on 11 channels of sEMG. Subsequently, with each optimization iteration, the number of optimized channels increased by one, resulting in recognition rates for the affected hand’s motion intention by amputees with different numbers of optimized channels (Table 9).



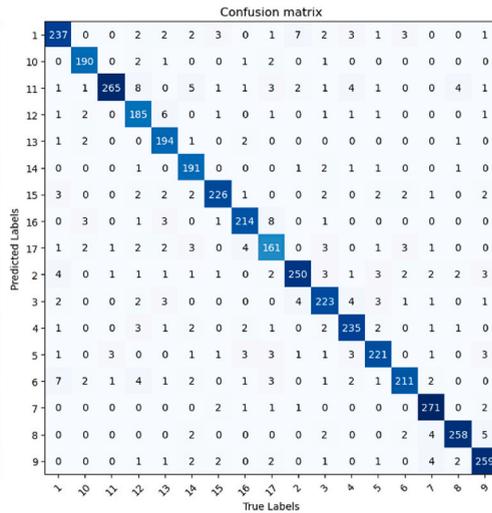
(a)



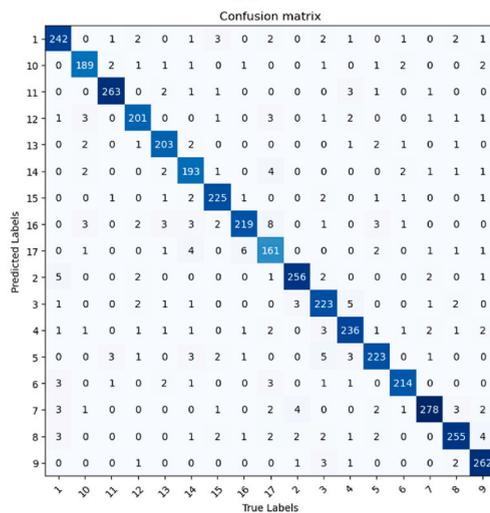
(b)



(c)



(d)



(e)

(caption on next page)

Fig. 9. Confusion matrix of DenseNet121 and improved networks: (a) DenseNet121; (b) MS-DenseNet (DenseBlock1~DenseBlock3); (c) MS-DenseNet (DenseBlock2~DenseBlock3); (d) MS-DenseNet (DenseBlock3); (e) MS-DA-DenseNet (DenseBlock3).

The results presented in Table 9 demonstrate that as the number of optimized channels increases from 1 to 4, the recognition rate of the Improved DenseNet network model for the time-frequency feature fusion map corresponding to the number of channels gradually improves. The best experimental result was achieved when the number of channels was optimized to 8, with a recognition rate of 94.50%, representing a 0.68% increase compared to the pre-optimization rate. In this case, channels 3, 6, 9, and 12 were removed.

Table 7
Comparison of DA-MS-DenseNet with previous studies implemented on Ninapro DB3.

Reference	Year	Classes	Classifier	Avg Accuracy
Atzori, M. et al. [26]	2014	50	SVM	46.27%
Atzori, M. et al. [19]	2016	50	CNN	38.09%
Zhai et al. [40]	2017	10	CNN	73.31%
Wei et al. [41]	2019	50	Multi-View CNN	>63%
Gopal. et al. [42]	2022	10	CNN	>70%
Fatayer. et al. [43]	2022	41	ALR-CNN	85.58%
This Work	2023	17	DA-MS-DenseNet	93.82%



Fig. 10. Pearson correlation coefficient heat map.

Table 8
The motion intention recognition rate of the affected hand based on the single-channel sEMG of the affected hand of amputees.

Channel number	Accuracy
1	21.19%
2	15.23%
3	13.26%
4	21.72%
5	14.96%
6	14.24%
7	16.06%
8	20.80%
9	14.75%
10	14.99%
11	19.72%
12	20.51%

Table 9

The motion intention recognition rate of the affected hands based on sEMG with different optimized number of channels.

Number of remaining channels	Accuracy
12	93.82%
11 (del 6)	94.23%
10 (del3.6)	94.47%
9 (del3.6.9)	94.45%
8(del3.6.9.12)	94.50%
7 (del3.6.9.10.12)	93.67%
6 (del 2.3.6.9.10.12)	92.23%

However, when the number of channels was optimized to 7, the recognition rate drops to 93.67%. This concludes the optimization process.

4. Discussion

In this section, the paper will discuss the experimental results based on the results in Section 3.2.

For Section 3.2.1, this section evaluates the model performance of recognition classification using six baseline convolutional neural networks on the sEMG time-frequency feature fusion map dataset. The experimental results show that only the Densenet121 and Resnet50 models perform well, achieving recognition accuracy above 85%. This can be attributed to the unique characteristics of the sEMG time-frequency feature fusion map datasets used in this paper. These datasets are pixel images generated from the sEMG of the amputee's affected hand and possess traits such as weak signal intensity and noise, making them challenging to interpret and understand compared to conventional images with easily recognizable features. The Resnet50 model's success can be attributed to its use of residual blocks, which establish connections between preceding and following layers. This residual connectivity allows ResNet50 to train deeper networks more effectively and retain more detailed information during feature extraction, resulting in superior performance. On the other hand, the DenseNet121 model borrows concepts from the ResNet model but takes the dense connection mechanism further. In this model, all layers are connected to each other, and each layer's additional input includes features from all previous layers. This extensive connectivity enhances feature reuse and contributes to the model's strong performance. However, the remaining four CNN models, which are designed to classify images under resource constraints, are not suitable for the dataset used in this paper.

For Section 3.2.2, this section is an ablation study of the Improved DenseNet. First, when adding the multi-scale structure to the DenseNet121 base, the results in Table 6 show that network performance improves in all three scenarios mentioned. However, when examining the confusion matrix (b) and (c) in Fig. 9, it becomes evident that for the first two scenarios, the recognition rate for some actions actually decreases, and the F1 score for the first scenario also decreases. This indicates that while increasing image feature information by adding a multi-scale structure can enhance network performance, it must be done in moderation. Adding too much feature information can result in information redundancy and adversely affect network performance. After determining the Multi-Scale structure, based on the results of the model after adding the DANet module, it is confirmed that adjusting the model weights for different channels in the multilayer feature map and the varying location features within each layer channel is beneficial for improving network performance. Despite the increase in time cost after adding the DANet, it is noticeable, as indicated in Table 6, that the substantial increase is primarily observed in T-time. In practical testing scenarios, the training process can be completed in advance. The increase in time for a single prediction (P-time) is only 0.21 ms, which is deemed entirely acceptable.

For Section 3.2.3, this section highlighted the effectiveness of our proposed model through a comparative analysis with state-of-the-art gesture recognition methods using sEMG data from Ninapro DB3. It is essential to note that variations in methodologies, such as pre-processing techniques and differences in the number of movement classes, may impact a fair comparison between studies. Statistical analysis on the Ninapro DB3 dataset was performed using the Kruskal-Willis test. The results indicated a p-value of 0.0032 (<0.05) and an h-value of 19.63, signifying a statistically significant difference. Subsequent Multcompare analysis revealed that DAMS-DenseNet achieved the highest mean rank, outperforming other classifiers and demonstrating its superior performance.

For Section 3.2.4, the channel optimization of multi-channel sEMG is performed based on the Pearson correlation coefficient heat map and the motion intention recognition rate based on single-channel affected hand sEMG. According to Table 9, the motion intention recognition rate of the affected hand sEMG based on a different optimized number of channels increases and then decreases with the increase in the optimized number of channels. The best result is obtained when the number of channels was optimized to 8. This demonstrates the feasibility of optimizing the number of channels, which not only provides convenience for amputees and improves their quality of life but also eliminates unnecessary redundant information in the signal. This optimization improves the accuracy of the Improved DenseNet model in identifying the motion intention of the affected hand based on the sEMG of the amputee's affected hand. However, it's important to consider the optimization scale. When the number of removed channels is too large, essential feature information may be lost, leading to a reduction in the model's recognition accuracy.

5. Conclusions

Recognizing the motion intention of the affected hand in amputees through sEMG control is a crucial challenge in the field of sEMG-controlled bionic manipulators. This paper introduces a novel approach by employing an STFT-based sEMG time-frequency feature fusion map. It transforms the time-domain EMG signals from amputees' affected hands into images and incorporates the frequency domain information of the sEMG to capture more distinctive details from the subtle residual sEMG of the affected hand. To address this challenge, the study utilizes the DenseNet121 model as a baseline and introduces the Improved DenseNet model. This specialized model enhances the accuracy of recognizing motion intentions based on sEMG data collected from amputees' affected hands. Recognizing the practical constraints faced by amputees, such as limited space and health considerations, the paper presents an innovative algorithm for optimizing the number of sEMG channels based on PCMIRR. This optimization process results in highly accurate motion intention recognition while using a minimal number of sEMG channels, improving the convenience and usability for amputees. Based on the aforementioned statements, the following innovations were achieved:

- An sEMG time-frequency feature fusion map was developed based on the STFT algorithm, providing time-frequency characteristics and localization information of the EMG signal, thereby enhancing the representation of the motor intention of the affected hand. Using surface EMG data from transradial amputees within the NinaproDB3 dataset, a dataset containing 20,524 fusion maps of the time-frequency characteristics of surface EMG signals during transradial amputee movements was constructed. This approach effectively mitigated overfitting, ensuring its applicability to practical field use.
- The DenseNet121 model is chosen as the baseline model, and a branch is added to DenseBlcok3 of the DenseNet121 model so that the output features of DenseBlcok3 and DenseBlcok4 are fused and sent to the next layer, reinforcing the low-level features while retaining more detailed information, and adding DANet after the first convolutional layer for optimizing the weights between channels of the multichannel feature map and the weights between the internal feature information of each channel. The proposed Improved DenseNet enhances the recognition accuracy of the motion intention of the affected hand in amputees by 4.2%, achieving an accuracy rate of 93.82% compared to the DenseNet121 model.
- The PCMIRR channel number optimization algorithm is introduced, which combines the Pearson correlation between the sEMG channels of amputees and the motion intention recognition rate based on the single-channel sEMG of the affected hand. This optimization method enhances the quality of life for amputees and effectively reduces redundant signal information. Ultimately, when the number of electrodes was optimized to 8, the Improved DenseNet model achieved the highest accuracy of 94.50% in recognizing the motion intention of the affected hand based on the sEMG signals from the amputees.

Accordingly, the proposed method in this study achieves precise recognition of the motion intention of amputees' affected hands using the minimum number of sEMG electrodes. However, there are limitations in terms of dataset sources and the number of categories in the dataset. Future work should consider the following aspects:

- As mentioned earlier, the data source in this study is the Ninapro DB3 database, and more reference to the original author's methods is needed for data preprocessing. Further filtering of subjects and action types is also required. In the future, we plan to create a dataset following the Ninapro official sEMG collection protocol.
- It is crucial to consider that an increase in the number of categories in the dataset directly impacts the model's accuracy. This issue will be a focus of future research because our goal is to improve the accuracy of the DA-MS-DenseNet model as the number of classes increases. These future studies will provide a deeper understanding of the DA-MS-DenseNet model's capabilities in electromyographic gesture classification tasks and offer valuable insights for the future development of this field.

Funding

This study has received funding from the Science and Technology Research Project of Henan Province (Grant No. 222102220080) and the Science Foundation of Henan University of Technology (Grant No. 2019BS055).

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Additional information

No additional information is available for this paper.

CRedit authorship contribution statement

Qunfeng Niu: Writing – review & editing, Writing – original draft, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Lei Shi:** Writing – original draft, Visualization, Validation, Software, Methodology, Data curation. **Yang Niu:** Supervision, Software, Investigation, Formal analysis, Data curation. **Kunming Jia:** Writing – original draft, Software, Methodology, Data curation. **Guangxiao Fan:** Software, Methodology, Data curation. **Ranran Gui:** Investigation, Data

curation. **Li Wang:** Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] S.J. Chen, G. Chen, Zheng, J. China disability population Survey and data, *International Journal of Reproductive Health/Family Planning*. 30 (3) (2011) 2, <https://doi.org/10.3969/j.issn.1674-1889.2011.03.016>.
- [2] G.J. Garcia, A. Alepuz, G. Balastegui, L. Bernat, J. Mortes, S. Sanchez, E. Vera, C.A. Jara, V. Morell, J. Pomares, J.L. Ramon, A. Ubeda, ARMA: a sensorized arm wearable for motor rehabilitation, *Biosens. Bioelectron.* 12 (7) (2022), <https://doi.org/10.3390/bios12070469>.
- [3] S. Wang, J.J. Zheng, B. Zheng, X.T. Jiang, Phase-based grasp classification for prosthetic hand control using sEMG, *Biosens. Bioelectron.* 12 (2) (2022), <https://doi.org/10.3390/bios12020057>.
- [4] H. Wang, Y. Zhang, C. Liu, H.H. Liu, sEMG based hand gesture recognition with deformable convolutional network, *International Journal of Machine Learning and Cybernetics* 13 (6) (2022) 1729–1738, <https://doi.org/10.1007/s13042-021-01482-7>.
- [5] M. Ortiz-Catalan, Engineering and surgical advancements enable more cognitively integrated bionic arms, *Sci. Robot.* 6 (58) (2021), <https://doi.org/10.1126/scirobotics.abk3123>.
- [6] D.W. Tan, M.A. Schiefer, M.W. Keith, J.R. Anderson, J. Tyler, D.J. Tyler, A neural interface provides long-term stable natural touch perception, *Sci. Transl. Med.* 6 (257) (2014), <https://doi.org/10.1126/scitranslmed.3008669>.
- [7] J. Duchateau, R.M. Enoka, Human motion unit recordings: origins and insight into the integrated motion system, *Brain Res.* 1409 (2011) 42–61, <https://doi.org/10.1016/j.brainres.2011.06.011>.
- [8] J.C. Gonzalez-Ibarra, C. Soubervielle-Montalvo, O. Vital-Ochoa, H.G. Perez-Gonzalez, EMG pattern recognition system based on neural networks, *Mexican International Conference on Artificial Intelligence* (2012), <https://doi.org/10.1109/MICAI.2012.23>.
- [9] A. Truong, H. Boujot, T. Zaharia, Laban descriptors for gesture recognition and emotional analysis, *Vis. Comput.* 32 (1) (2016) 83–98, <https://doi.org/10.1007/s00371-014-1057-8>.
- [10] S. Shen, K. Gu, X.R. Chen, C.X. Lv, R.C. Wang, Gesture recognition through sEMG with wearable device based on deep learning, *Mobile Network. Appl.* 25 (6) (2020) 2447–2458, <https://doi.org/10.1007/s11036-020-01590-8>.
- [11] S.A. Abboud, S. Al-Wais, S.H. Abdullah, F. Alnajjar, A. Al-Jumaily, Label self-advised support vector machine (LSA-SVM)-Automated classification of foot drop rehabilitation case study, *Biosens. Bioelectron.* 9 (4) (2019), <https://doi.org/10.3390/bios9040114>.
- [12] H. Wang, B. Ru, X. Miao, Q. Gao, M. Habib, L. Liu, S. Qiu, MEMS devices-based hand gesture recognition via wearable computing, *Micromachines* 14 (5) (2023), <https://doi.org/10.3390/mi14050947>.
- [13] A. Cenedese, G.A. Susto, G. Belgioioso, G.I. Cirillo, F. Fraccaroli, Home automation oriented gesture classification from inertial measurements, *IEEE Trans. Autom. Sci. Eng.* 12 (4) (2015) 1200–1210, <https://doi.org/10.1109/TASE.2015.2473659>.
- [14] K. Bergmann, S. Kopp, Modeling the production of coverbal iconic gestures by learning bayesian decision networks, *Appl. Artif. Intell.* 24 (6) (2010) 530–551, <https://doi.org/10.1080/08839514.2010.492162>.
- [15] Z.W. Yang, D. Jiang, Y. Sun, B. Tao, X.L. Tong, G.Z. Jiang, M.M. Xu, J.T. Yun, Y. Liu, B.J. Chen, J.Y. Kong, Dynamic gesture recognition using surface EMG signals based on multi-stream residual network, *Front. Bioeng. Biotechnol.* 9 (2021), <https://doi.org/10.3389/fbioe.2021.779353>.
- [16] M. Akmal, S. Khalid, M. Moiz, M.J. Abbass, M.F. Qureshi, Z. Mushtaq, Leveraging training strategies of artificial neural network for classification of multiday electromyography signals, in: *2022 International Conference on Emerging Trends in Electrical, Control, and Telecommunication Engineering, ETECTE*, Lahore, Pakistan, 2022, pp. 1–5, <https://doi.org/10.1109/ETECTE55893.2022.10007103>.
- [17] X.Y. Jiang, X.Y. Liu, J.H. Fan, X.M. Ye, C.Y. Dai, E.A. Clancy, D. Farina, W. Chen, Enhancing IoT security via cancelable HD-sEMG-based biometric authentication password, encoded by gesture, *IEEE Internet Things J.* 8 (22) (2021) 16535–16547, <https://doi.org/10.1109/JIOT.2021.3074952>.
- [18] Y. Du, W. Jin, W. Wei, Y. Hu, W. Geng, Surface EMG-based inter-session gesture recognition enhanced by deep domain adaptation, *Sensors* 17 (3) (2017), <https://doi.org/10.3390/s17030458>.
- [19] M. Atzori, M. Cognolato, H. Muller, Deep learning with convolutional neural networks applied to electromyography data: a resource for the classification of motions for prosthetic hands, *Front. Neurobot.* 10 (2016) 9, <https://doi.org/10.3389/fnbot.2016.00009>.
- [20] L.X. Yang, X.H. Xie, J.H. Lai, Learning discriminative visual elements using part-based convolutional neural network, *Neurocomputing* 316 (2018) 135–143, <https://doi.org/10.1016/j.neucom.2018.07.059>.
- [21] Y.C. Wei, Y. Zhao, C.Y. Lu, S.K. Wei, L.Q. Liu, Z.F. Zhu, S.C. Yan, Cross-modal retrieval with CNN visual features: a new baseline, *IEEE Trans. Cybern.* 47 (2) (2017) 449–460, <https://doi.org/10.1109/TCYB.2016.2519449>.
- [22] Y. Cheng, G. Li, M. Yu, D. Jiang, J. Yun, Y. Liu, D. Chen, Gesture recognition based on surface electromyography-feature image, *Concurrency Comput. Pract. Ex.* 33 (6) (2020), <https://doi.org/10.1002/cpe.6051>.
- [23] M.F. Qureshi, Z. Mushtaq, M.Z.U. Rehman, E.N. Kamavuako, Spectral image-based multiday surface electromyography classification of hand motions using CNN for human-computer interaction, *IEEE Sensor. J.* 22 (21) (2022) 20676–20683, <https://doi.org/10.1109/JSEN.2022.3204121>.
- [24] M.F. Qureshi, Z. Mushtaq, M.Z.U. Rehman, E.N. Kamavuako, E2CNN: an efficient concatenated CNN for classification of surface EMG extracted from upper limb, *IEEE Sensor. J.* 23 (8) (2023) 8989–8996, <https://doi.org/10.1109/JSEN.2023.3255408>.
- [25] Y. Zhang, F. Yang, Q. Fan, A. Yang, X. Li, Research on sEMG-based gesture recognition by dual-view deep learning, *IEEE Access* 10 (2022) 32928–32937, <https://doi.org/10.1109/ACCESS.2022.3158667>.
- [26] M. Atzori, A. Gijbets, C. Castellini, B. Caputo, A.G. Hager, S. Elsig, G. Giatsidis, F. Bassetto, H. Muller, Electromyography data for non-invasive naturally-controlled robotic hand prostheses, *Sci. Data* 1 (2014) 140053, <https://doi.org/10.1038/sdata.2014.53>.
- [27] M. Atzori, A. Gijbets, C. Castellini, B. Caputo, A.G. Hager, S. Elsig, G. Giatsidis, F. Bassetto, H. Muller, Effect of clinical parameters on the control of myoelectric robotic prosthetic hands, *J. Rehabil. Res. Dev.* 53 (3) (2016) 345–358, <https://doi.org/10.1682/JRRD.2014.09.0218>.
- [28] P.L. Hudak, P.C. Amadio, C. Bombardier, D. Beaton, D. Cole, A. Davis, G. Hawker, J.N. Katz, M. Makela, R.G. Marx, L. Punnett, J. Wright, Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder, and head), *Am. J. Ind. Med.* 29 (6) (1996) 602–608, [10.1002/\(SICI\)1097-0274\(199606\)29:6<602::AID-AJIM4>3.0.CO;2-L](https://doi.org/10.1002/(SICI)1097-0274(199606)29:6<602::AID-AJIM4>3.0.CO;2-L).
- [29] Y. Sun, C. Xu, G. Li, W. Xu, D. Chen, Intelligent human computer interaction based on non redundant EMG signal, *AEJ - Alexandria Engineering Journal.* 59 (3) (2020), <https://doi.org/10.1016/j.aej.2020.01.015>.
- [30] J. Qi, G. Jiang, G. Li, Y. Sun, B.J. Tao, Intelligent human-computer interaction based on surface EMG gesture recognition, *IEEE Access* 7 (2019) 61378–61387, <https://doi.org/10.1109/ACCESS.2019.2914728>.
- [31] S. Kyeong, W. Shin, M. Yang, U. Heo, J.R. Feng, J.J. Kim, Recognition of walking environments and gait period by surface electromyography, *Frontiers of Information Technology & Electronic Engineering* 20 (3) (2019) 11, <https://doi.org/10.1631/FITEE.1800601>.
- [32] G. Huang, Z. Liu, L.V.D. Maaten, K.Q. Weinberger, Densely Connected Convolutional Networks, *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, <https://doi.org/10.1109/CVPR.2017.243>.
- [33] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, H. Lu, Dual Attention Network for Scene Segmentation, 2018, <https://doi.org/10.1109/CVPR.2019.00326>.
- [34] G.N. Lance, W.T. Williams, A general theory of classificatory sorting strategies, *Comput. J.* (1967), <https://doi.org/10.1093/comjnl/10.3.271>.

- [35] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, IEEE Conference on Computer Vision and Pattern Recognition (2016), <https://doi.org/10.1109/CVPR.2016.90>.
- [36] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, A. Rabinovich, Going deeper with convolutions, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition, 2015, <https://doi.org/10.1109/CVPR.2015.7298594>.
- [37] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.C. Chen, MobileNetV2: inverted residuals and linear bottlenecks, IEEE/CVF Conference on Computer Vision and Pattern Recognition 2018 (2018), <https://doi.org/10.1109/CVPR.2018.00474>.
- [38] N. Ma, X. Zhang, H.T. Zheng, J. Sun, ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design, European Conference on Computer Vision, 2018, https://doi.org/10.1007/978-3-030-01264-9_8.
- [39] I. Radosavovic, R.P. Kosaraju, R. Girshick, K. He, P. Dollár, Designing network design spaces, 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (2020), <https://doi.org/10.1109/CVPR42600.2020.01044>.
- [40] X.L. Zhai, B. Jelfs, R.H.M. Chan, C. Tin, Self-recalibrating surface EMG pattern recognition for neuroprosthesis control based on convolutional neural network, Front. Neurosci. 11 (2017), <https://doi.org/10.3389/fnins.2017.00379>.
- [41] W.T. Wei, Q.F. Dai, Y.K. Wong, Y. Hu, M. Kankanhalli, W.D. Geng, Surface-electromyography-based gesture recognition by multi-view deep learning, IEEE Trans. Biomed. Eng. 66 (10) (2019) 2964–2973, <https://doi.org/10.1109/TBME.2019.2899222>.
- [42] P. Gopal, A. Gesta, A. Mohebbi, A systematic study on electromyography-based hand gesture recognition for assistive robots using deep learning and machine learning models, Sensors 22 (10) (2022), <https://doi.org/10.3390/s22103650>.
- [43] A. Fatayer, W.P. Gao, Y.L. Fu, sEMG-based gesture recognition using deep learning from noisy labels, Ieee Journal of Biomedical and Health Informatics 26 (9) (2022) 4462–4473, <https://doi.org/10.1109/JBHI.2022.3179630>.