Sensors & Actuators: A. Physical 363 (2023) 114693

Contents lists available at ScienceDirect

Sensors and Actuators: A. Physical

journal homepage: www.journals.elsevier.com/sensors-and-actuators-a-physical

Deep transfer learning-based adaptive gesture recognition of a soft e-skin patch with reduced training data and time

Yu Rong [a,b], Guoying Gu [a,b,c,*]

[a] Robotics Institute, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China
[b] State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai, 200240, China
[c] Meta Robotics Institute, Shanghai Jiao Tong University, Shanghai, 200240, China

ARTICLE INFO

Keywords:
Soft e-skin patch
Gesture recognition
Deep transfer learning
Adaptive strategy

ABSTRACT

Deep learning-based classification algorithms are promising in gesture recognition with soft e-skin patches. However, the reported algorithms usually require large amount of training data, resulting in the time-consuming data collection process. In this paper, we present a deep transfer learning-based adaptive strategy for accurate gesture recognition of a soft e-skin patch with reduced training data and time. To this end, we first train a base neural network as the general feature extraction network. Next, we transfer the front layers of the pre-trained base network to target networks of new gesture recognition tasks. Further, we apply the fine-tune technique to refine the copied parameters. Finally, with our custom-built soft e-skin patch, we experimentally verify the developed strategy on two typical transfer cases, termed as the user transfer case (Case I) and the gesture transfer case (Case II). The experimental results show that, to ensure the stable accuracy of 95 %, the training data with and without the adaptive strategy are 1,312 vs 10,912 for Case I, and 8,192 vs 12,032 for Case II, respectively. In this sense, the training time of target networks can be reduced by 62.96 % for Case I and 34.20 % for Case II, respectively. This work shows the potential to promote the widespread application of e-skins in human computer interaction.

1. Introduction

Hand gestures are of great significance to human-to-human communication [1] and human computer interaction [2]. For example, the communication gap between the hearing impaired and healthy people can be bridged by sign language [3], and the disabled can accurately control robotic wheelchairs assisted by a gesture recognition system [4]. E-skin with embedded soft strain sensors is a promising device for gesture recognition due to its intrinsic stretchability and biocompatibility [5]. To efficiently achieve fast and robust hand gesture recognition, data-driven classification algorithms are crucial [6].

In the literature, many data-driven classification algorithms, such as K-Nearest Neighbor (KNN), Decision Tree (DT), and Linear Discriminant Analysis (LDA), have been widely used [7–9]. However, when dealing with multidimensional data or large datasets, these algorithms generally suffer from relatively low classification accuracy and high computational complexity [10]. On the other hand, features of the input data need to be manually selected, which dramatically increases users' burden [11].

Due to the remarkable expressive power and the ability to automatically extract latent features, the deep learning algorithms are investigated and applied for gesture recognition [12]. Among reported deep learning algorithms, artificial neural networks (ANNs) are mostly used in static gesture recognition due to their relatively simple structures [13]. However, ANNs generally neglect the spatial and temporal features in the input data. This limitation impedes the wide applications for the e-skins consisting of sensor arrays or dynamic gesture recognition tasks. Alternatively, convolutional neural networks (CNNs) are employed to identity the spatial features, since filters in each layer are able to extract local information revealing strain patterns of the skin or bending angles of fingers [14]. For dynamic gesture recognition, recurrent neural networks (RNNs) are introduced to analyze sequential signals. For example, by dividing the collected data into several sequences, Zhang et al. [15] presented a long short-term memory (LSTM) neural network to extract the temporal features, which improved the classification accuracy.

Despite the promising achievements on gesture recognition with deep learning approaches, the practical application in gesture recognition with an e-skin system is still limited. This gap may be mainly caused by the huge training data demand of neural networks to distinguish diverse sensor signal distributions among different tasks. The

* Corresponding author at: Robotics Institute, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China.
E-mail address: guguoying@sjtu.edu.cn (G. Gu).

https://doi.org/10.1016/j.sna.2023.114693
Received 3 April 2023; Received in revised form 12 September 2023; Accepted 25 September 2023
Available online 27 September 2023
0924-4247/

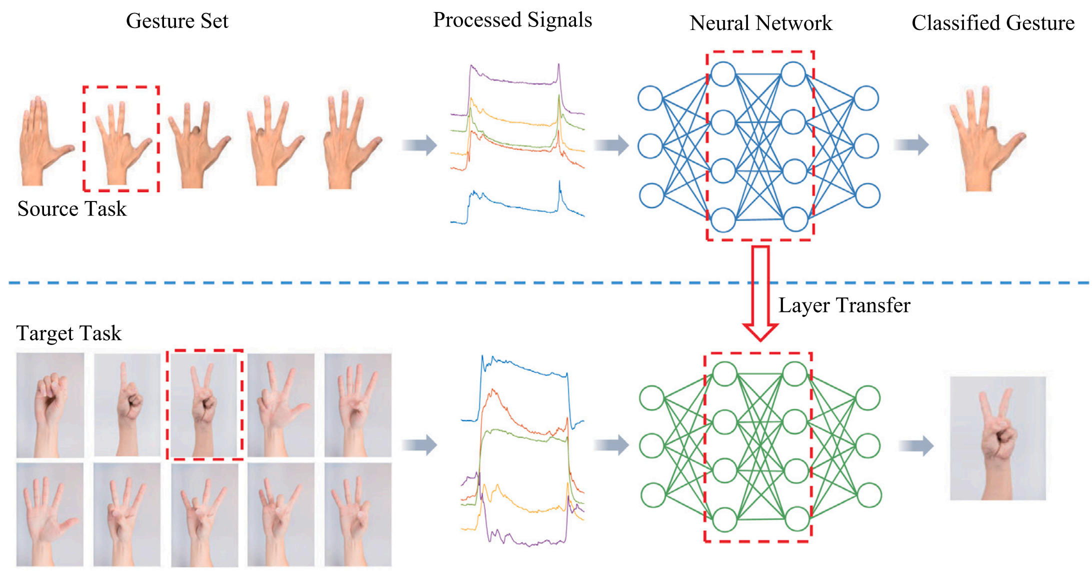

Fig. 1. Overview of realizing the rapid adaption of an e-skin patch to a new gesture recognition task with the proposed adaptive strategy for the gesture transfer case.

large amount of high-quality data with manual annotation are time-consuming to collect, which is unrealistic for daily use. Alternatively, deep transfer learning is an effective approach to release the requirement of heavy training data. By leveraging learned knowledge from source tasks (i.e., original tasks) to target tasks (i.e., new tasks), deep transfer learning can ensure high accuracy of target neural networks with relatively small training datasets, which have been well demonstrated in the field of gesture recognition based on Surface Electromyography (sEMG) [16], computer-aided diagnosis [17], and simulations of chemical, biological systems [18], image colorization [19], machinery fault diagnosis [20], and face mask detection [21].

Recently, some interesting works have been reported to apply deep transfer learning to soft sensor-based devices, but they mainly focus on the utilization of a single sensor [22]. Although Thuruthel et al. [23] introduced deep transfer learning algorithms into a multi-sensor system, they achieved the reduced training time of networks at a cost of decreased accuracy. Moreover, the effectiveness of transfer learning is researched when the device is reattached by the same user or reused by different users [24]. Besides, the used approaches are usually limited to the same gesture set with a fixed transfer method. Currently, it is still elusive on how to adaptively adjust the transfer strategy for different gesture recognition tasks in order to release the enormous data burden.

In this paper, we present an adaptive deep transfer learning strategy for gesture recognition with a soft e-skin patch with five embedded strain sensors (as shown in Fig. 1), which achieves high-accuracy classifications with reduced training data and time. To this end, we firstly train a base neural network as the general feature extraction network. Next, we transfer the front layers of the pre-trained base network to target networks of new gesture recognition tasks. Further, we apply the fine-tune technique to further refine the copied parameters. Finally, with our custom-built soft e-skin patch, we experimentally verify the developed algorithms on two typical transfer cases, termed as the user transfer case (Case I) and the gesture transfer case (Case II). The experimental results show that, to ensure the stable accuracy of 95%, the training data with and without the adaptive strategy are 1312 vs. 10,912 for Case I, and 8192 vs. 12,032 for Case II, respectively. In this sense, the training time of target networks can be reduced by 62.96% for Case I and 34.20% for Case II, respectively. These results demonstrate that the proposed adaptive strategy can decrease the data demand and training time of target networks, which is particularly marked for the user transfer case. Therefore, it is promising to enhance the application of the soft e-skin patch in various fields. Overall, the main novelties and contributions of this work can be summarized as follows.

(1) A deep transfer learning-based adaptive strategy is proposed for a soft e-skin patch. It can enhance the rapid adjustment of the soft e-skin patch to diverse gesture recognition tasks by circumventing lengthy data acquisitions and training time.

(2) A parameter determination scheme of the number of transferred layers is developed based on the similarity between the source task and the target task. For the user transfer case, all layers of the source network are transferred; for the gesture transfer case, only the first layer of the source networks is transferred, which rarely investigated by previous researches.

(3) Experimental results are presented and compared to validate the effectiveness and feasibility of the proposed strategy.

The reminder of this paper is organized as follows. Section 2 proposes the basic gesture recognition algorithm. Section 3 introduces the detailed adaptive strategy. Section 4 analyzes experimental results and verifies the development, and Section 5 concludes this paper.

2. Basic gesture recognition strategy

In this section, we firstly introduce the basic gesture recognition strategy based on the deep learning classification algorithm, including its structure and its training scheme. Then, we explicitly explain how to determine the optimal strategy for a single gesture recognition task.

2.1. Network structure and training scheme

We adopt the Multilayer Feedforward Neural Networks (MLFNN) due to its remarkable expressive power and simple structure [25]. The developed network in our work contains an input layer, hidden layers, and an output layer. Each neuron in a layer is connected with all neurons in adjacent layers. Rectified Linear Units (ReLU) are selected as the activation functions of hidden layers to accelerate computation and avoid the vanishing gradient problem [26], and the output layer utilizes Sigmoid activation function.

To tackle the typical training problem, a customized training scheme for the applied MLFNNs is utilized. Firstly, the adaptive moment estimation (Adam) optimizer is adopted to update the parameters of networks because of its ability to deal with sparse gradients [27]. Meanwhile, mini-batch training strategy is employed to address the converge problem [28], and the batch size in this research is set to 32.

Besides, early stopping technique is employed to avoid overfitting [29]. It means that the training process would be terminated ahead when the classification performance of the MLFNN stagnates for 10 training iterations. The gesture dataset collected from one task is split into 3 datasets: a training dataset to update parameters, a validation dataset to prevent overfitting, and a test dataset to evaluate the performance. The proportions of 3 datasets are 70%, 15% and 15%, respectively. The standard cross-entropy loss function is employed as a performance metric during training process:

$$L(\pi; y) = -\sum_{k=1}^{K} y_k \log \pi_k \tag{1}$$

where y_k is the correct labels, and π_k is the predicted probability.

2.2. Determination of the optimal basic strategy

A basic gesture recognition strategy which attains outstanding classification performance can fundamentally promote the effectiveness of the proposed adaptive strategy. The determination of the basic gesture recognition strategy can be achieved by two steps. In the first step, the number of hidden layers and neurons of each layer should be properly set. In the second step, a training setup is required.

For step 1, the most important factor is the number of hidden layers, because excess hidden layers can cause overfitting problem and long training time. MLFNNs containing 1 hidden layer or 2 hidden layers are powerful for most classification problems [25], so MLFNNs with 1 or 2 hidden layers are compared in this step. The numbers of neurons contained in each layer can be determined by grid search method. For step 2, the most important factor is the data preprocess method, since improper preprocess methods tend to result in the difficulty of convergence [30]. In this work, we compare two techniques for data preprocess for comparisons: (i), extracting the relative change rates of sensor readings, which is widely used in recalibration of sensors [31–33] (ii), normalizing sensor readings, which is commonly employed in the field of deep learning. Moreover, unprocessed data are adopted as a benchmark. Detailed experiment results are presented in Section 4.2.

3. Adaptive strategy for rapid adjustment to new gesture recognition tasks

In this section, we present an adaptive deep transfer learning strategy for a soft e-skin patch to facilitate its rapid adaption in new gesture recognition tasks. We firstly give an overview of the deep transfer learning. Then, we introduce 2 typical transfer cases and their corresponding datasets. Finally, the detailed adaptive strategy is developed based on the basic gesture recognition strategy mentioned in Section 2.

3.1. Overview of deep transfer learning

Deep transfer learning is a significant tool to alleviate the enormous data burden. The key idea of deep transfer learning is to reapply knowledge learned from one task to another similar task [11]. In general, the task that is used to acquire prior knowledge is named the source task, and the task that receives transferred knowledge is named the target task. Datasets of source tasks and target tasks are called source datasets and target datasets, respectively.

Deep transfer learning can be explicitly defined using mathematic notations. First of all, a source domain can be represented by $S = \{X, P(X)\}$. The source dataset $X = \{x_1, \dots, x_n\}$ consists of numerous instances, and the source probability distribution $P(X)$ is approximately represented by a neural network. Then, given the target dataset $Y = \{y_1, \dots, y_m\}$, deep transfer learning aims to acquire the target probability distribution $P(Y)$. Specifically, the target probability distribution is a conditional probability distribution $P(Y|S)$, since it leverages prior knowledge learned in the source domain. In addition, the size of the target dataset tends to be much smaller than the size the source dataset, i.e., $m \ll n$.

Fig. 2. Gesture sets: (a) American Sign Language Digits (ASLD) ; (b) Single Finger Movements (SFM) [9].

3.2. Two typical transfer cases

In this work, 2 typical transfer cases with a soft e-skin patch are investigated: 1) the user transfer case (Case I): the soft e-skin patch is worn by another user to classify the same gesture sets; 2) the gesture transfer case (Case II): the e-skin patch is worn by the same user to classify another gesture set.

For Case I, we employ the widely used American Sign Language Digits (ASLD) [6,34] as the gesture sets of the source task and the target task as shown in Fig. 2(a). The source dataset and the target dataset are collected from 2 users. For Case II, the Single Finger Movement (SFM) set (as shown in Fig. 2(b)) is selected as the source gesture set, and ASLD set is used as the target gesture set. The SFM set is the simplest hand gestures, and may be regarded as basic components of complicated gestures of the ASLD set. Therefore, neural networks for the SFM recognition tasks may imply general knowledge about extracting latent representations, which can be reapplied to the ASLD recognition tasks. The source dataset and the target dataset of Case II are SFM gesture dataset and ASLD dataset collected from the same user.

3.3. Adaptive strategy

In the proposed adaptive strategy, the network-based deep transfer learning is applied due to its efficient implementation and interpretability [11]. In this sense, a source MLFNN is firstly trained. As shown in Fig. 3(a), a well-trained source MLFNN can be divided into two parts: general feature extractors and specific feature extractors. The front layers of the source MLFNN tend to extract general latent features implied in both source datasets and target datasets, and succeeding layers are prone to extract specific features which differ remarkably among diverse tasks [11]. By transferring front layers, parameters of general feature extractors are reused as the initialization of corresponding layers in the target network as shown in Fig. 3(b). Leveraging prior knowledge acquired from the source task can reduce the data to train the general feature extractors of the target MLFNN. Therefore, the data-efficiency of target networks is increased, and the enormous data burden is released without the loss of classification accuracy.

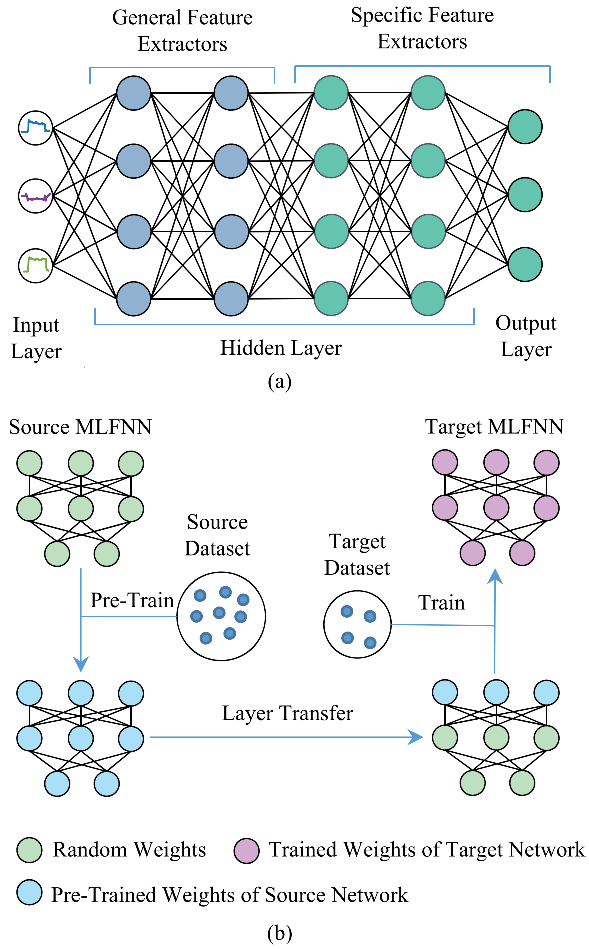

Fig. 3. (a) MLFNNs for classification can be typically divided into two part: general feature extractors and specific feature extractors. (b) The structure of adaptive network-based deep transfer learning strategy. A source MLFNN is firstly pre-trained. Then, general feature extractors of the source MLFNN are transferred, which can promote the data efficiency of the target MLFNN. Finally, the target MLFNN is trained with the target dataset.

Parameters of specific feature extractors of the target MLFNN are randomly initialized, because misuse of parameters in these layers can cause serious optimization difficulties [35]. Finally, the target MLFNN is trained with the target dataset.

4. Experimental results

In this section, we firstly illustrate the setup of the experimental platform and data collection process. We then determine the optimal basic gesture recognition strategy. Finally, we show the experimental results of the 2 transfer cases with the proposed adaptive strategy.

4.1. Experimental platform and data collection

Fig. 4 shows the experimental platform with our customized soft e-skin patch. The soft e-skin patch [9] contains 5 W-shaped soft strain sensors. When the hand moves to make a gesture, the stretching skin on the back of the hand shall elongate the strain sensors of the soft e-skin patch and thus increase their resistances. Soft sensors are set along 1st, 2nd, 3rd, 4th, and 5th metacarpal. Their locations are optimized by a sequence of feature reduction methods (refer to [9] for details). The middle sensor is orthogonal to the metacarpal, and others are along corresponding metacarpals. Each soft sensor is composed of a sandwich structure: two polyethylene terephthalate (PET) substrates

Table 1
Detailed setups of networks.

Network	Number of hidden layers	Data preprocess method
Network 1	2	Normalized
Network 2	2	Relative change
Network 3	2	Unprocessed
Network 4	1	Normalized
Network 5	1	Relative change
Network 6	1	Unprocessed

Table 2
Experiment results of different strategies for Case I.

Strategy	Accuracy/%	Training time/s
S1T	96.31 ± 1.85	275.83 ± 41.32
S2T	78.75 ± 4.59	227.92 ± 21.81
S1T+	95.61 ± 2.18	249.51 ± 28.33
S2T+	85.65 ± 4.72	175.87 ± 28.98
S3T+	98.69 ± 0.21	203.39 ± 21.43

protecting the middle soft electrode layer made from carbon grease (847, MG Chemical, Canada). The sensors feature low hysteresis and high stability across days.

A microcontroller unit (MCU, STM32L475, STMicroelectronics) with embedded analog-to-digital converter (ADC) records the changing resistances of strain sensors at a sampling frequency of 100 Hz. The data are gathered and transmitted online from the MCU to the computer. Once the data are obtained, a digital low-pass filter in MATLAB (2019a, Mathworks) is used to filter the raw data. Finally, a Python deep learning framework (PyTorch 1.6.0, with Graphics Processing Unit NVIDIA GTX 2070s) is applied to classify gestures using the processed data.

Twelve healthy subjects (7 males and 5 females aged 21–33 years old) have participated in data acquisition. The soft e-skin patch was attached to the subjects' left hands with the help of atoxic, adhesive gel. A screen was used to display the gestures to make. Each gesture was required to hold for 5 s following a 5 seconds' relaxation. Subjects could ask for a rest at any time if necessary. A computer stored the collected data online and processed data offline. The middle 4 s of collected data during making gestures were retained, because the first and last 0.5 s were treated as transition time. For simplicity, we use the notations of SFMn and ASLDn ($n = 1, 2, \ldots, 11, 12$) to represent SFM datasets and ASLD datasets collected from the subject n.

4.2. Experiments for determining the optimal basic strategy

6 MLFNNs with different setups (as shown in Table 1) are trained and compared to identify the optimal gesture recognition strategy. To eliminate the contingency caused by random initialization, each network is repeatedly trained with 20 different sets of initialized parameters. The evaluation criteria are average classification accuracy and training time.

The results are shown in Fig. 5. The results show that preprocessed data can dramatically promote the classification accuracies and reduce the training time. Comparing to the two data preprocess methods, the relative change rate of input data occasionally causes relatively low average prediction accuracies (about 80%), which is obvious in network 2. By contrast, networks trained on normalized input data steadily maintain their prediction accuracies at almost 100%. Networks with single hidden layer require more training time to ensure stable classification accuracies (>95%) compared to networks containing 2 hidden layers. Based on above experimental results, the network with 2 hidden layers trained on normalized input data is the optimal choice for the basic gesture recognition strategy.

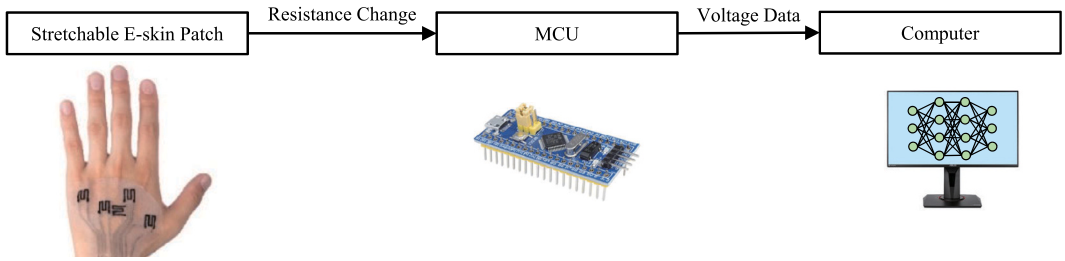

Fig. 4. Diagram of the experimental setup. When making gestures, the stretching skin can elongate the soft strain sensors and thus increase their resistance. Then, an MCU (STM32L475) with embedded ADCs is installed between sensors and external resistors as a voltage divider to record the changing voltage signals. All data are collected and saved online, passing through the MCU to the computer.

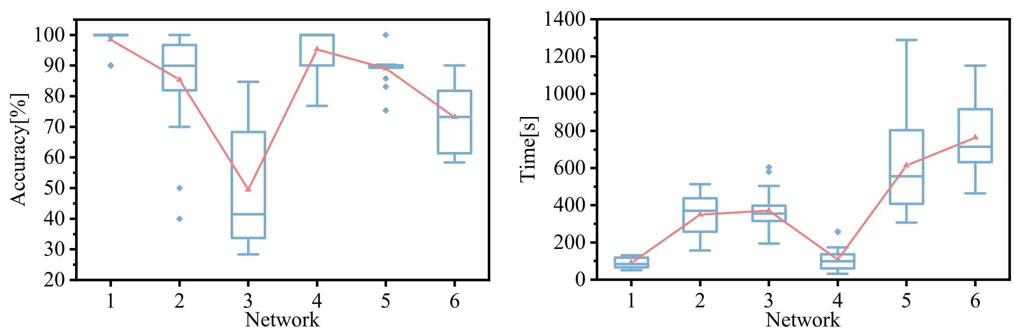

Fig. 5. Experimental comparison among different networks in terms of average classification accuracy (left) and training time (right). The network with 2 hidden layers trained on normalized input data is the optimal choice for the basic gesture recognition strategy.

Table 3
Prediction accuracies for Case I trained with different amounts of training data.

Data amount	787	918	1049	1181	1312	1443
Accuracy/%	88.01	90.21	91.45	93.26	95.50	95.92

Table 4
Experiment results of different strategies for Case II.

Strategy	Accuracy/%	Training time/s
S1T	94.07 ± 4.74	133.01 ± 24.20
S2T	69.26 ± 10.36	321.06 ± 29.71
S1T+	95.24 ± 3.81	69.43 ± 11.73
S2T+	80.85 ± 8.86	52.83 ± 8.86

4.3. Experiments of two transfer cases

Implementation of deep transfer learning varies with different pairs of source tasks and target tasks. Therefore, the adaptive strategy for each transfer case should be investigated separately. There are two main factors for the adaptive strategy. The first factor is the usage of fine-tune technique, because the rate of source training data and target training data determines whether to utilize fine-tune technique [35]. The second factor is the number of copied layers from pre-trained source networks which mainly depends on similarity between source tasks and target tasks. After determining the optimal adaptive strategy for each case, its effectiveness and feasibility would be further confirmed.

The entire experiment process is performed as follows. Firstly, several alternative adaptive strategies are designed, tested and compared in terms of classification accuracy and training time. In this process, a pair of the source task and the target is utilized. It should be noted that the target dataset is only half the size of the source dataset, because an effective adaptive strategy should ensure high recognition accuracy with small target datasets. For simplicity, notations SnT and SnT+ ($n = 1, 2, 3$) are used to represent networks trained according to different strategies. Specifically, the parameter n represents the number of layers copied to target networks. The front n hidden layers are transferred for $n = 1, 2$, and 2 hidden layers and the output layer are transferred for $n = 3$. The symbol + signifies the utilization of the fine-tune technique. Secondly, networks with and without the determined adaptive strategies are trained and compared to verify the effectiveness. Moreover, the prediction accuracies of the adaptive strategy trained with different amounts of training data are analyzed. Besides, the minimum data demanded by target networks to ensure robust and high recognition accuracy (>95%) are confirmed. Finally, additional pairs of source tasks and target tasks are employed to validate the generalization of the proposed adaptive strategies. For Case I, the pair of the source dataset ASLD1 and the target dataset ASLD2 are employed to determine the optimal transfer strategy. 10 pairs of the source dataset ASLD1 and the target datasets ASLDn ($n = 3, 4, \ldots, 10, 12$) are used to verify the generalization. For Case II, the pair of the source dataset SFM1 and the target dataset ASLD1 are employed to determine the optimal transfer strategy. 11 pairs of the source dataset SFMn and the target datasets ASLDn ($n = 2, 3, \ldots, 10, 12$) are used to verify the generalization.

(1) Experimental Results of Case I: Table 2 demonstrates classification accuracies and training time of target networks under different adaptive strategies. Fine-tuned networks attain an average classification accuracy of 90.63% compared to 87.48% without the fine-tune technique. Besides, S1T+ and S2T+ achieve 9.54% and 22.84% reduction in average training time compared to S1T and S2T. These results demonstrate that the fine-tune technique can effectively promote classification accuracy and reduce training time. As more layers of source networks are copied to target networks and then fine-tuned, average classification accuracy firstly decreases from 95.61% of S1T+ to 85.65% of S2T+, and then increase to 98.69% of S3T+. In addition, average training time has a similar fluctuation trend. Among all fine-tuned networks attaining 95% classification accuracy, S3T+ requires the shortest training time. Therefore, copying and fine-tuning all parameters of source networks is the optimal adaptive strategy for Case I.

Experimental results of target networks with and without the proposed adaptive strategy are shown in Fig. 6. When the same training data are utilized, target networks with the proposed adaptive strategy

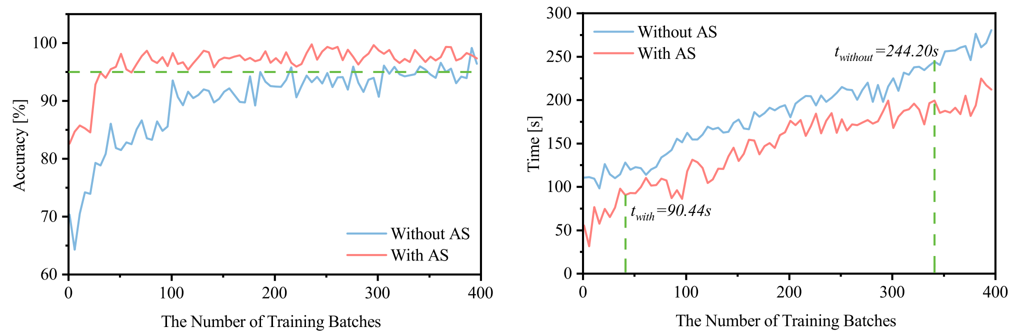

Fig. 6. Experimental results of networks with and without the adaptive strategy (AS) for Case I in terms of average classification accuracy (left) and training time (right).

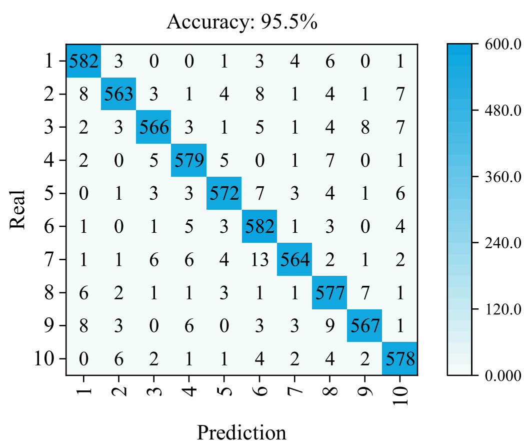

Fig. 7. Confusion matrix of the adaptive strategy for Case I.

Table 5
Prediction accuracies for Case II trained with different amounts of training data.

Data amount	4915	5734	6553	7372	8192	9011
Accuracy/%	89.31	92.87	93.15	94.31	96.10	95.62

obviously achieve higher classification accuracy. Prediction accuracies of the adaptive strategy trained with different amounts of training data are summarized in Table 3. The results show that a decrease in training data results in a reduced prediction accuracy. To achieve an acceptable classification accuracy of 95%, the required minimal number of training data in target networks with the adaptive strategy is 1312. For the neural network trained without the adaptive strategy, the demanded amount of training data is 10,912. Owing to the reduced input data, the training time of the target network is reduced from 244.20 s to 90.44 s. Moreover, the inference time, including data acquisition, data process, and forward calculation, of the neural network trained with the adaptive strategy is 12.46 ms. This result shows that the trained neural network can be used in real-time gesture recognition (about 65 Hz). The confusion matrix of the adaptive strategy tested on the test dataset of ASLD2 is shown in Fig. 7. The results demonstrate that the adaptive strategy can dramatically release the heavy data demand and accelerate the training process of networks.

With the optimal adaptive strategy, the average classification accuracy of networks trained on 10 other target datasets are 95.15%. Therefore, the result demonstrates the generalization of the proposed adaptive strategy.

(2) Experimental Results of Case II: Table 4 lists the classification accuracies and training time of target networks with different adaptive strategies. With fine-tune method, S1T+ and S2T+ achieve higher classification accuracies of 95.24% and 80.85%, respectively, compared to 94.07% and 69.26% of S1T and S2T. Besides, S1T+ and S2T+ both demand less training time (69.43 s and 52.83 s, respectively) than S1T and S2T (133.01 s and 321.06 s, respectively). Overall, the fine-tune technique can promote the prediction accuracy and dramatically reduce training time. For two fine-tuned networks, more copied layers of source networks can evidently decrease classification accuracies from 95.24% of S1T+ to 80.85% of S2T+. Therefore, copying and fine-tuning the first layer of source networks is the optimal adaptive strategy for Case II.

Experimental results of target networks with and without the proposed adaptive strategy are shown in Fig. 8. When the same training data are utilized, average classification accuracies of target networks with the proposed adaptive strategy are slightly higher. Moreover, average training time is nearly identical for two kinds of target networks given the same dataset. The prediction accuracy changes of the adaptive strategy trained on different amounts of training data are summarized in Table 5. The results show that a higher prediction accuracy necessitates a larger amount of training data, which is similar to Case I. Besides, the determined amount of training data for Case II is 8192. For the neural network trained without the adaptive strategy, the demanded amount of training data is 12,032. Owing to the reduced input data, the training time of the target network is reduced from 246.64 s to 162.30 s. Besides, the inference time of the trained network for Case II (13.05 ms) is nearly the same as Case I (12.46 ms), which also verifies the capability of the real-time gesture recognition (about 65 Hz). The confusion matrix of the adaptive strategy tested on the test dataset of ASLD1 is shown in Fig. 9. These results demonstrate that the proposed adaptive strategy can ensure robust and outstanding performance with small target datasets.

The average prediction accuracy of target networks trained on 11 additional target datasets is 95.42%. The result verifies the adaptability of proposed adaptive strategy to different tasks.

4.4. Discussion

The experimental results demonstrate that the proposed adaptive gesture recognition strategy realizes efficient reutilization of the trained network for two cases. This accomplishment primarily stems from the transfer of knowledge about various strain patterns. Specifically, when different users present the same gesture, the strain patterns on the back of their hands exhibit similarities due to the resemblances in their hand structures. Furthermore, even when a single user performs different gesture sets, similarities in strain patterns can emerge due to consistent hand motion habits. These similarities can be inherited by transferring trained layers, thus leading to training data reduction and training time reduction.

Moreover, it is worth noting that the adaptive strategy employed in Case I features more transferred layers and more remarkable time reduction. This means that more knowledge is inherited by transferred layers in Case I. This phenomenon may primarily arise from the increased similarity in hand strain patterns when different users make the same gesture. For example, making gesture nine of ASLD set involves

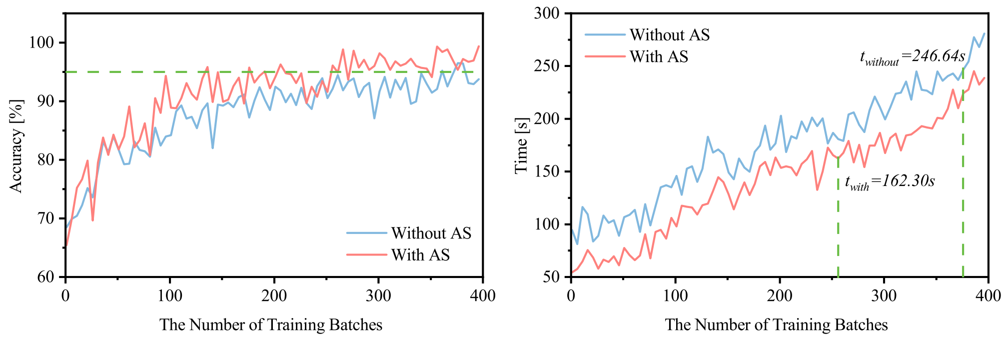

Fig. 8. Experimental results of networks with and without the adaptive strategy (AS) for Case II in terms of average classification accuracy (left) and training time (right).

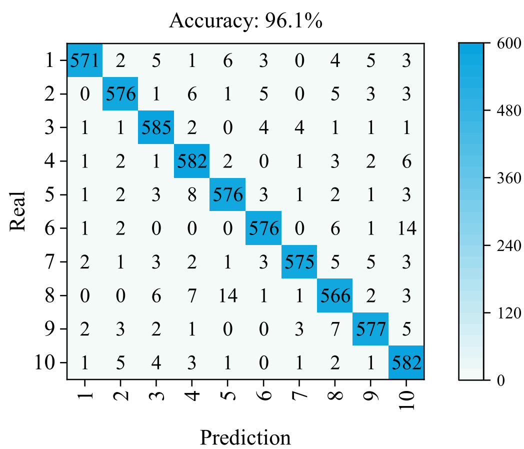

Fig. 9. Confusion matrix of the adaptive strategy for Case II.

bending thumbs and forefingers, leading to larger readings of sensors placed near thumbs and forefingers. This pattern holds true for most users, and thus there are more similarities between two tasks in Case I. On the other hand, there is no such similarity between ASLD set and SFM set. Therefore, less trained layer can be transferred in Case II.

5. Conclusions

In this paper, we present an efficient adaptive deep transfer learning strategy for a soft e-skin patch. The adaptive strategy combines the model-based deep transfer learning algorithm and MLFNNs. With the proposed strategy, gesture recognition systems achieve rapid adaption to inexperienced tasks by releasing heavy training data demand. Owing to the reduced training data, the collecting time of data and training time of target networks are decreased. The proposed adaptive strategy varies based on the similarity between source tasks and target tasks. For the user transfer case, all layers of the source network are transferred; for the gesture transfer case, only the first hidden layer of the source networks is transferred. In the future work, we will further investigate the influence of the e-skin patch reattachment with the adaptive transfer strategy. Overall, the proposed adaptive strategy holds potential to enhance the application of an e-skin in human–computer interaction.

CRediT authorship contribution statement

Yu Rong: Conceptualization, Methodology, Data curation, Software, Writing – original draft. **Guoying Gu:** Supervision, Funding acquisition, Writing – review & editing.

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.

Data availability

Data will be made available on request.

Acknowledgments

This study was supported in part by the National Natural Science Foundation of China (Grant Nos. 52025057 and T2293725), the Science and Technology Commission of Shanghai Municipality, China (Grant No. 22511101700).

References

[1] David McNeill, Language and Gesture, Vol. 2, Cambridge University Press, 2000.
[2] Jess McIntosh, Asier Marzo, Mike Fraser, Carol Phillips, Echoflex: Hand gesture recognition using ultrasound imaging, in: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 2017, pp. 1923–1934.
[3] Himadri Nath Saha, Sayan Tapadar, Shinjini Ray, Suhrid Krishna Chatterjee, Sudipta Saha, A machine learning based approach for hand gesture recognition using distinctive feature extraction, in: 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), 2018, pp. 91–98.
[4] Xiang Gao, Lei Shi, Qiang Wang, The design of robotic wheelchair control system based on hand gesture control for the disabled, in: 2017 International Conference on Robotics and Automation Sciences (ICRAS), 2017, pp. 30–34.
[5] Jun Chang Yang, Jaewan Mun, Se Young Kwon, Seongjun Park, Zhenan Bao, Steve Park, Electronic skin: recent progress and future prospects for skin-attachable devices for health monitoring, robotics, and prosthetics, Adv. Mater. 31 (48) (2019) 1904765.
[6] Ming Jin Cheok, Zaid Omar, Mohamed Hisham Jaward, A review of hand gesture and sign language recognition techniques, Int. J. Mach. Learn. Cybern. 10 (2019) 131–153.
[7] Chieh-Ju Wu, Kai-Hsiang Lin, Meng-Lin Hsieh, Jen-Yuan Chang, Realization of natural user interface for computer control with KNN classifier enhanced smart glove, in: Information Storage and Processing Systems, Vol. 59124, 2019, V001T13A004.
[8] Kuang-Yow Lian, Chun-Chieh Chiu, Yong-Jie Hong, Wen-Tsai Sung, Wearable armband for real time hand gesture recognition, in: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2017, pp. 2992–2995.
[9] Shuo Jiang, Ling Li, Haipeng Xu, Junkai Xu, Guoying Gu, Peter B. Shull, Stretchable e-skin patch for gesture recognition on the back of the hand, IEEE Trans. Ind. Electron. 67 (1) (2019) 647–657.
[10] Sotiris B. Kotsiantis, Ioannis Zaharakis, P. Pintelas, et al., Supervised machine learning: A review of classification techniques, Emerg. Artif. Intell. Appl. Comput. Eng. 160 (1) (2007) 3–24.
[11] Chuanqi Tan, Fuchun Sun, Tao Kong, Wenchang Zhang, Chao Yang, Chunfang Liu, A survey on deep transfer learning, in: Artificial Neural Networks and Machine Learning - ICANN 2018, 2018, pp. 270–279.
[12] Ajay Shrestha, Ausif Mahmood, Review of deep learning algorithms and architectures, IEEE Access 7 (2019) 53040–53065.
[13] Noor Adnan Ibraheem, RafiqulZaman Khan, Survey on various gesture recognition technologies and techniques, Int. J. Comput. Appl. 50 (7) (2012) 38–44.

[14] Oliver Glauser, Shihao Wu, Daniele Panozzo, Otmar Hilliges, Olga Sorkine-Hornung, Interactive hand pose estimation using a stretch-sensing soft glove, ACM Trans. Graph. (ToG) 38 (4) (2019) 1–15.

[15] Xiaoliang Zhang, Ziqi Yang, Taiyu Chen, Diliang Chen, Ming-Chun Huang, Cooperative sensing and wearable computing for sequential hand gesture recognition, IEEE Sens. J. 19 (14) (2019) 5775–5783.

[16] Zhipeng Yu, Jianghai Zhao, Yucheng Wang, Linglong He, Shaonan Wang, Surface EMG-based instantaneous hand gesture recognition using convolutional neural network with the transfer learning method, Sensors 21 (7) (2021) 2540.

[17] Hoo-Chang Shin, Holger R. Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, Ronald M. Summers, Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning, IEEE Trans. Med. Imaging 35 (5) (2016) 1285–1298.

[18] Justin S. Smith, Benjamin T. Nebgen, Roman Zubatyuk, Nicholas Lubbers, Christian Devereux, Kipton Barros, Sergei Tretiak, Olexandr Isayev, Adrian E. Roitberg, Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning, Nat. Commun. 10 (1) (2019) 2903.

[19] Shanshan Huang, Xin Jin, Qian Jiang, Li Liu, Deep learning for image colorization: Current and future prospects, Eng. Appl. Artif. Intell. 114 (2022) 105006.

[20] Chuan Li, Shaohui Zhang, Yi Qin, Edgar Estupinan, A systematic review of deep transfer learning for machinery fault diagnosis, Neurocomputing 407 (2020) 121–135.

[21] Mohamed Loey, Gunasekaran Manogaran, Mohamed Hamed N. Taha, Nour Eldeen M. Khalifa, A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic, Measurement 167 (2021) 108288.

[22] DongWook Kim, Junghan Kwon, Byungjun Jeon, Yong-Lae Park, Adaptive calibration of soft sensors using optimal transportation transfer learning for mass production and long-term usage, Adv. Intell. Syst. 2 (6) (2020) 1900178.

[23] Thomas George Thuruthel, Kieran Gilday, Fumiya Iida, Drift-free latent space representation for soft strain sensors, in: 2020 3rd IEEE International Conference on Soft Robotics (RoboSoft), 2020, pp. 138–143.

[24] Kosuke Kikui, Yuta Itoh, Makoto Yamada, Yuta Sugiura, Maki Sugimoto, Intra-/inter-user adaptation framework for wearable gesture sensing device, in: Proceedings of the 2018 ACM International Symposium on Wearable Computers, 2018, pp. 21–24.

[25] Ronen Eldan, Ohad Shamir, The power of depth for feedforward neural networks, in: Conference on Learning Theory, 2016, pp. 907–940.

[26] Xavier Glorot, Antoine Bordes, Yoshua Bengio, Deep sparse rectifier neural networks, in: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, 2011, pp. 315–323.

[27] Diederik P. Kingma andJimmy Ba, Adam: A method for stochastic optimization, in: 3rd International Conference on Learning Representations (ICLR), 2015.

[28] Dominic Masters, Carlo Luschi, Revisiting small batch training for deep neural networks, 2018, arXiv preprint arXiv:1804.07612.

[29] Rich Caruana, Steve Lawrence, C. Giles, Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping, in: Advances in Neural Information Processing Systems, Vol. 13, 2000.

[30] Kun Lan, Dan-tong Wang, Simon Fong, Lian-sheng Liu, Kelvin K.L. Wong, Nilanjan Dey, A survey of data mining and deep learning in bioinformatics, J. Med. Syst. 42 (2018) 1–20.

[31] HaiPeng Xu, ZeQun Shen, GuoYing Gu, Performance characterization of ionic-hydrogel based strain sensors, Sci. China Technol. Sci. 63 (6) (2020) 923–930.

[32] Zequn Shen, Xiangyang Zhu, Carmel Majidi, Guoying Gu, Cutaneous ionogel mechanoreceptors for soft machines, physiological sensing, and amputee prostheses, Adv. Mater. 33 (38) (2021) 2102069.

[33] Guoying Gu, Haipeng Xu, Sai Peng, Ling Li, Sujie Chen, Tongqing Lu, Xiaojun Guo, Integrated soft ionotronic skin with stretchable and transparent hydrogel–elastomer ionic sensors for hand-motion monitoring, Soft Robot. 6 (3) (2019) 368–376.

[34] Kshitij Bantupalli, Ying Xie, American sign language recognition using deep learning and computer vision, in: 2018 IEEE International Conference on Big Data (Big Data), 2018, pp. 4896–4899.

[35] Jason Yosinski, Jeff Clune, Yoshua Bengio, Hod Lipson, How transferable are features in deep neural networks? in: Advances in Neural Information Processing Systems, Vol. 27, 2014.

Yu Rong Ph.D. candidate; School of Mechanical Engineering; Shanghai Jiao Tong University; His research interests include data-driven modeling and control of soft robotics, and machine intelligence.

Guoying Gu received the B.E. degree (with honors) in electronic science and technology, and the Ph.D. degree (with honors) in mechatronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2006 and 2012, respectively.

Since October 2012, Dr. Gu has worked at Shanghai Jiao Tong University, where he is currently appointed as a Professor of School of Mechanical Engineering. He was a Humboldt Fellow with University of Oldenburg, Germany. He was a Visiting Scholar at Massachusetts Institute of Technology, National University of Singapore and Concordia University. His research interests include soft robotics, bioinspired and wearable robots, smart materials sensing, actuation and motion control. He is the author or co-author of over 100 publications, which have appeared in Nature Biomedical Engineering, Science Robotics, Science Advances, Advanced Materials, National Science Review, IEEE/ASME Trans., Advanced Functional Materials, Soft Robotics, etc., as book chapters and in conference proceedings.

Dr. Gu received the National Science Fund for Distinguished Young Scholars. Now he serves as Associate Editor of IEEE Transactions on Robotics and IEEE Robotics and Automation Letters. He has also served for several journals as Editorial Board Member, Topic Editor, or Guest Editor, and several international conferences/symposiums as Chair, Co-Chair, Associate Editor or Program Committee Member.