

Contrastive Domain Adaptation: A Self-Supervised Learning Framework for sEMG-Based Gesture Recognition

Anonymous IJCB 2022 submission

Abstract

Gesture recognition using surface electromyography (sEMG) shows its great potential in the field of human-computer interaction (HCI). Previous works achieve relatively good performance based on the assumption of invariant statistic distribution. However, the practical application effect is unsatisfactory due to the problem of domain shift. Existing approaches need plenty of labeled sEMG samples from target scenarios for calibration, which is burdensome for experimenters and users. In this work, we present a contrastive self-supervised learning framework (ConSSL) for sEMG-based gesture recognition to realize domain adaptation in target domains. After pretraining on a bunch of unlabeled samples, only a small number of labeled samples are needed for calibration and domain adaptation. Experimental results indicate that the proposed framework outperforms other approaches even if $\leq 50\%$ labeled samples in target scenarios are available and achieves the state-of-the-art.

1. Introduction

Hand gesture recognition is one of the most promising research fields to realize natural human-computer interaction (HCI), which can be achieved by Wi-Fi [1], inertial measurement unit [8], camera [16] and surface electromyography (sEMG). sEMG is a bio-signal derived from the muscle fibers' action potential [15]. sEMG-based gesture recognition approaches directly sense and decode human muscular activities [19, 18]. sEMG-based gesture recognition is capable of distinguishing subtle finger movements and wrist movements and is insensitive to environmental light and sound, which arise the interest of many researchers [15, 19, 18].

The sEMG-based gesture recognition has achieved great success in academia while the application in real-world scenarios is limited. Since the sEMG signals are bio-signals sensed from the skin and easily disturbed by several factors, such as electrode shifts, muscular fatigue and skin sweating,

statistical distributions of sEMG data are variable on different days or different subjects, which cause the problem of domain shift [14, 11] in inter-session and inter-subject scenarios. The domain shift problem leads to the poor performance of real-time applications. Furthermore, participants may be bored with long-time signal collection while the acquisition equipment varies greatly, it is expensive to collect a large number of sEMG signals and label them. For the reason above, establishing a framework for domain adaptation using a few labeled samples shows extremely significance.

The domain adaptation of sEMG-based gesture recognition approaches usually follows a two-stage paradigm composed of pretraining and fine-tuning that are pretrained on existing sessions or subjects (source domain) and fine-tuned on target sessions or subjects (target domain) [11, 7, 6, 4]. Following the two-stage paradigm composed of pretraining and fine-tuning, we develop a contrastive self-supervised learning framework for domain adaptation called *ConSSL*, which is inspired by contrastive learning models in computer vision [2, 3, 17]. Considering sEMG signals as time-series data, we notice that sEMG samples from the same trials, especially from the adjacent sampled windows, have similar data distribution. During pretraining, *ConSSL* needs just sEMG signals with no gesture labels and works in a task-agnostic self-supervised way, which learns critical knowledge by comparing samples in pairs. Instead of a bunch of labeled data, just a small amount of labeled signals (one or two trials for each gesture) are needed for calibration during task-specific fine-tuning. Fine-tuning achieves adaptation to the target domain using labeled sEMG samples as few as possible, which reduces the labor of experimenters and improves the participant experience degrees.

The main contributions of our work can be summarized as:

- We propose contrastive self-supervised learning framework *ConSSL* for sEMG-based gesture recognition to tackle the issue of domain shift. The proposed framework needs only unlabeled samples during pretraining and a small number of labeled samples in target domain, which is of vital importance for real-world

applications.

- We propose construction approaches of sEMG pairs and a contrastive pairwise loss for the pretraining task of gesture recognition task which also can be applied in the other time-series tasks.
- We conduct experiments on two datasets which comprise inter-session and inter-subject scenarios with limited labeled sEMG samples and examine the performance of ConSSL with the labels increasing for two positive pair construction approaches. Results show that ConSSL achieves the state-of-the-art.

The remainder of this paper is organized as follows. Section 2 presents related work. Section 3 depicts the detail of the proposed framework. Section 4 describes the experimental setting and results followed by a conclusion in Section 5.

2. Related work

sEMG-based gesture recognition. Researchers devoting considerable energy to establishing powerful algorithms for sEMG-based gesture recognition. Recently, deep learning approaches have prospered in many fields including sEMG-based recognition. Geng et al. [9] view sEMG signals as instantaneous sEMG images and use deep convolutional networks to classify them followed by majority voting. Zhang et al. [20] extract spatial-temporal features by convolutional neural networks (CNN) and long short-term memory (LSTM) to recognize hand gestures. Lai et al. [13] propose the concept of sEMG graph and use graph neural network (GNN) and temporal neural network (TCN) to realize gesture recognition. An important assumption of these methods is that the statistic distributions of offline training and online application are invariant, so these works usually validate their algorithms in intra-session scenarios, which cannot guarantee the effectiveness in inter-session or inter-subject scenarios.

Domain adaptation. Domain shift problem creates a barrier between offline training and practical application, and many researchers have been dedicated to solving it. Du et al. [7] present a standard benchmark database for inter-session gesture recognition and adopt AdaBN for data calibration. Kanoga et al. [10] apply a source selection and semi-supervised transfer mapping algorithm to overcome inter-subject variabilities. Kim et al. [12] propose a subject-independent method based on a muscle source activation model, in which a reference source model facilitates parameter transfer from a specific subject. Ketykó et al. [11] use a deep learning model to approximate the domain shift that exists in inter-session and inter-subject scenarios. These works need a large amount of labeled samples and rich

mathematical knowledge, which increases extra labor and limits the flexibility.

3. Methods

3.1. Problem statement

Given by a large pooled unlabeled sEMG dataset $D_u = \{x_i^u\}_{i=0}^{N_u-1}$ obtained from different sessions or subjects (source domain) and a small number of labeled signals $D_l = \{(x_i^l, y_i^l)\}_{i=0}^{N_l-1}$ from target sessions or subjects (target domain), where $x_i \in \mathbb{R}^{T \times C}$ denotes a sEMG sample composed of T C -channels frames and y_i is the corresponding gesture label, N_u and N_l are the number of unlabeled and labeled samples, respectively. Domain adaptation of sEMG-based gesture recognition hopes to learn abstract knowledge from unlabeled sEMG signals and transfer it to target scenarios for faster learning or better performance. The key issue that domain adaptation must tackle is the domain shift that mainly comes from inter-session and inter-subject scenarios. sEMG signals coming from human muscles are disturbed by many factors, e.g., power-line interference, sensor shift, muscle fatigue, which can be formulated as:

$$X \sim \mathcal{D}_{emg} \circ \mathcal{D}_{noise}, \quad (1)$$

where signal variable X follows a mixture distribution composed of $\mathcal{D}_{emg}$ and $\mathcal{D}_{noise}$. $\mathcal{D}_{emg}$ is the real component derived from motor units [9] and $\mathcal{D}_{noise}$ means the noise jamming signal that comes from sources mentioned above. $\circ$ denotes the mixed mode.

By constructing a series of tasks $\mathcal{T} = \cup_{k=1}^K \mathcal{T}_k$ based on D_u and D_l , in which K is the number of tasks, finally we can get a model with strong adaptability that is easy to adapt to new scenarios. Significantly task $\mathcal{T}_k$ can be supervised or unsupervised, which needs a careful design to make the best use of sEMG signals.

3.2. Contrastive domain adaptation framework

From expression 1 we can know that the statistic distribution of real sEMG signal $\mathcal{D}_{emg}$ is an invariant distribution, noise distribution $\mathcal{D}_{noise}$ is inconsistent in the source domain and the target domain. Instead of approximating or learning the noise distribution using sEMG signals coupled with corresponding gesture labels during pre-training [11], a more efficient approach is to directly learn the general representation of sEMG signals and filter out the noises that we do not care about. What's more, labeling for a large number of sEMG data seems to be extra labor, because some useful information (from same or not same trial/session/subject) has been included when collecting sEMG data. For these reasons, we propose this contrastive self-supervised learning framework *ConSSL* that comprises tasks $\mathcal{T} = \{\mathcal{T}_{pretrain}, \mathcal{T}_{fine-tune}\}$, where

$\mathcal{T}_{pretrain}$ is a contrastive self-supervised task-agnostic pre-training task for general knowledge learning, i.e., $\mathcal{D}_{emg}$, and $\mathcal{T}_{fine-tune}$ denotes task-specific fine-tuning task for domain adaptation and gesture recognition in target scenarios downstream. More details can be seen in Algorithm 1.

In the pre-training paradigm based on the contrastive self-supervised framework, the construction of sEMG pairs is a crucial problem. We construct positive pairs and negative pairs based on segmented samples from trials. When a subject has performed a gesture, sEMG signals collected from him/her are treated as a trial. The sEMG samples generated from trials using a sliding window compose unlabeled dataset D_u . Significantly, the two adjacent samples x_t, x_{t+1} in a trail compose of a positive pair. The positive dataset D_{pos} is made up of positive sEMG pairs. On contrary, nonadjacent samples, which include samples from nonadjacent windows, are treated as negative pairs. Considering the balance of positive and negative pairs and the convenience of calculation, we construct negative pairs based on the sampled positive pair set B_{pos} instead of positive dataset D_{pos} . The construction of positive pairs considers the stability and continuity of gestures, especially gesture transition. What's more, it is difficult to ensure that external factors remain unchanged even in the same trial, e.g., force level, which are mixed in sEMG signals as noises. Nonadjacent samples as negative pairs can filter out these noises to a certain extent during pretraining. Furthermore, noises may be introduced in the different trials which cause domain shifts including inter-session and inter-subject domain shifts. Negative pairs formed by samples from different trials can eliminate these noises. By constructing positive pairs and negative pairs, essential information of sEMG can be decoded while interference information will be filtered out.

Whether pretraining or fine-tuning, the role of encoder can not be ignored. In our framework, we choose STCN-GR [13] as encoder $f(\cdot|\theta_f)$. STCN-GR is a novel model for sEMG-based gesture recognition, which is the first introduction of sEMG graph to represent the sEMG signals and use graph neural networks and temporal neural networks to extract spatial-temporal information. STCN-GR has achieved state-of-the-art in the field of sEMG-based gesture recognition. In proposed ConSSL, we discard the dropout layer and use the rest as the encoder. What follows is a similarity head $g(\cdot|\theta_g)$ that maps representations to the space where a contrastive pairwise series loss is applied when pretraining. The encoder are followed by a classification head $h(\cdot|\theta_h)$ that projects sEMG representations to gesture space and recognizes corresponding gestures when fine-tuning.

The pretraining framework is illustrated in Figure 1. During pretraining, each trials of collected sEMG signals will be segmented as sEMG samples using sliding window

Algorithm 1 Contrastive Domain Adaptation Algorithm

Input: Unlabeled dataset D_u , labeled dataset D_l , numbers of epochs E_1 and E_2 , batch size S_1 and S_2 , encoder $f(\cdot|\theta_f)$, similarity head $g(\cdot|\theta_g)$, classification head $h(\cdot|\theta_h)$.

Output: Classifier $F_2(\cdot|\Theta_2)$.

```

1: function PRETRAIN( $D_u, E_1, S_1, f(\cdot|\theta_f), g(\cdot|\theta_g)$ )
2:   Define  $F_1(\cdot|\Theta_1) = g(f(\cdot|\theta_f)|\theta_g)$ 
3:   Initialize parameters  $\Theta_1$ 
4:   Construct  $D_{pos} = \{(x_t^u, x_{t+1}^u) | x_t^u, x_{t+1}^u \in D_u\}$ 
5:   for  $e = 0 \rightarrow E_1 - 1$  do
6:     while True do
7:       Sample  $B_{pos} = \{p_i^{pos}\}_{i=0}^{S_1-1}$  from  $D_{pos}$ 
8:       if  $|B_{pos}| < S_1$  then
9:         break
10:      end if
11:      for  $i = 0 \rightarrow S_1 - 1$  do
12:         $z_{i,0} = F_1(p_{i,0}^{pos}|\Theta_1)$ 
13:         $z_{i,1} = F_1(p_{i,1}^{pos}|\Theta_1)$ 
14:      end for
15:      for  $i, j \in \{0, \dots, S_1 - 1\}, m, n \in \{0, 1\}$  do
16:        Compute pair similarity  $sim(i, m, j, n)$ 
17:      end for
18:      Compute contrastive loss  $\mathcal{L}_1$ 
19:      Update  $\Theta_1$  to minimize  $\mathcal{L}_1$ 
20:    end while
21:  end for
22:  return  $f(\cdot|\theta_f)$ 
23: end function

24: function FINE_TUNE( $D_l, E_2, S_2, f(\cdot|\theta_f), h(\cdot|\theta_h)$ )
25:   Define  $F_2(\cdot|\Theta_2) = h(f(\cdot|\theta_f)|\theta_h)$ 
26:   Initialize parameters  $\theta_h$ 
27:   for  $e = 0 \rightarrow E_2 - 1$  do
28:     while True do
29:       Sample  $B_l = \{(x_i^l, y_i^l)\}_{i=0}^{S_2-1}$  from  $D_l$ 
30:       if  $|B_l| < S_2$  then
31:         break
32:       end if
33:       for  $i = 0 \rightarrow S_2 - 1$  do
34:         $\hat{y}_i = F_2(x_i^l|\Theta_2)$ 
35:      end for
36:      Compute classification loss  $\mathcal{L}_2$ 
37:      Update  $\Theta_2$  to minimize  $\mathcal{L}_2$ 
38:    end while
39:  end for
40:  return  $F_2(\cdot|\Theta_2)$ 
41: end function

```

with or without overlap. Each positive samples pair or negative samples pair composed of sEMG samples are fed into the encoder $f(\cdot|\theta_f)$ followed by similarity head $g(\cdot|\theta_g)$ to

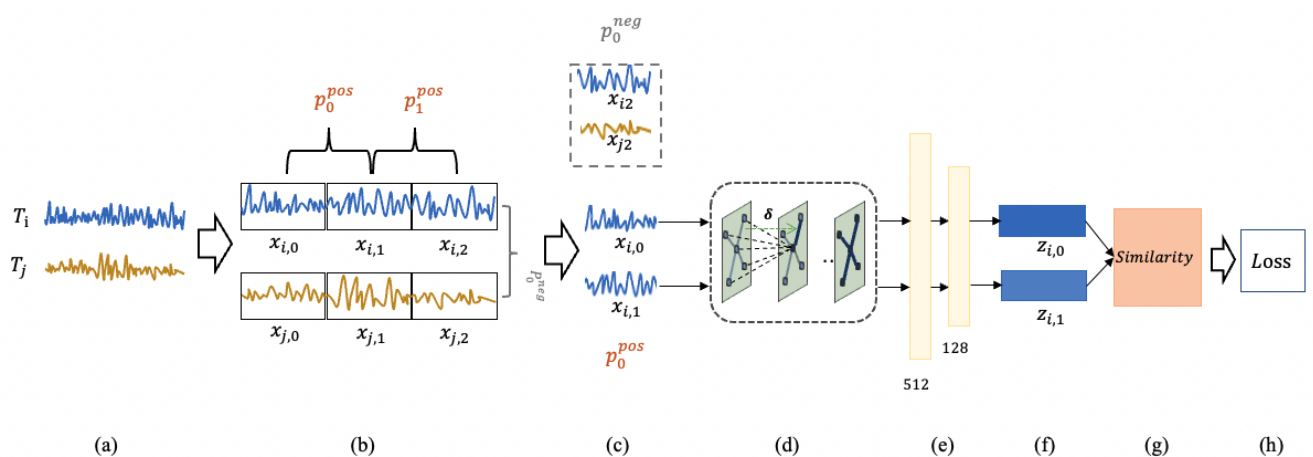


Figure 1. Contrastive pretraining in the source domain. Take two trials for example. (a) Different sEMG trials T_i and T_j . (b) For each trial, a sliding window will be performed to generate samples, $x_{i,t}$ and $x_{j,t}$. Adjacent samples compose positive pairs (orange font), e.g., $p_0^{pos} = (x_{i,0}, x_{i,1})$, and nonadjacent samples compose negative pairs (grey font), e.g., $p_0^{neg} = (x_{i,2}, x_{j,2})$. (c) Surface EMG samples are fed into networks in pairs. (d) Spatial-temporal convolutional networks for feature extraction. (e) The similarity head comprises two full connection layers that are 512 dimensions and 128 dimensions, respectively. (f) Surface EMG representations of sEMG samples in pair. (g) Similarity for sEMG samples in pair. (h) Loss calculation according to similarity scores of all pairs in a batch.

get semantic feature vectors. Pairwise similarity scores of sEMG pairs are calculated for final contrastive loss. For pairs p_i^{pos} and p_j^{pos} , the similarity score can be formulated as:

$$\text{sim}(i, m, j, n) = z_{i,m}^T z_{j,n} / (\|z_{i,m}\| \cdot \|z_{j,n}\|), \quad (2)$$

where $m, n \in \{0, 1\}$, and $z_{i,0}, z_{i,1}$ represent the feature vectors of the first sample and the second sample in pair p_i^{pos} calculated by $F_1(\cdot|\Theta)$ in Algorithm 1. The pairwise series loss is defined as:

$$\ell(i, m, j, n) = -\log(\exp(\text{sim}(i, m, j, n)/\tau) / \sum_{s=0}^{S_1-1} (1 - \mathbb{1}_{[s=i, m=n]}) \exp(\text{sim}(i, m, s, n)/\tau)), \quad (3)$$

where $\mathbb{1}_{[\cdot]}$ indicates whether they are the same sample and τ is a temperature parameter. For given a batch size S_1 , the final contrastive loss can be represented as:

$$\mathcal{L}_1 = \sum_{i=0}^{S_1-1} (\ell(i, 0, i, 1) + \ell(i, 1, i, 0)). \quad (4)$$

From the expression of 3 and 4, we can see that the similarities of sEMG samples in positive pairs are considered in numerator and denominator while the others are only considered in the denominator, i.e., the similarities of negative sEMG pairs. For the convenience of calculation, overlapping samples in two adjacent pairs are treated as negative pairs, e.g., $(x_{i,1}, x_{i,1})$ formed by samples of positive pairs p_0^{pos} and p_1^{pos} in Figure 1. By minimizing $\mathcal{L}_1$, the extracted

features of sEMG samples from the same trials will be similar while that from different trials will be dissimilar, which realizes the learning of general knowledge and complete the preparation of domain adaptation. Significantly, during pretraining, sEMG data are unlabeled, which means that no extra labor is consumed when collecting them. Pretraining completely works in a self-supervised and task-agnostic way. The dataset can be continuously expanded and trained at all hours if necessary in actual scenarios. What's more, the contrastive loss $\mathcal{L}_1$ can be applied in the other time series tasks based on contrastive learning to learn the essential knowledge contained in signals.

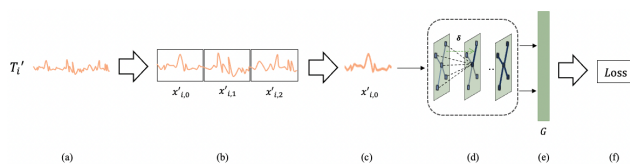


Figure 2. Fine-tuning in the target domain. For trial T'_i , it will be windowed into sEMG samples and fed into the pretrained feature extractor followed by a classification header to calculate classification loss.

Fine-tuning illustrated in Figure 2 transfers critical knowledge to target scenarios and realizes domain adaptation. A classifier $F_2(\cdot|\Theta)$ comprising pretrained encoder $f(\cdot|\theta_f)$ and randomly initialized classification header $h(\cdot|\theta_h)$ will be trained using a few labeled sEMG samples from D_t in target domain. Good performance can be achieved even if the samples are not abundant. The loss

function $\mathcal{L}_2$ used at this stage is a cross-entropy loss:

$$\mathcal{L}_2 = - \sum_{i=0}^{S_2-1} \sum_{j=0}^{G-1} (y_{ij} \log \hat{y}_{ij}). \quad (5)$$

where S_2 and G are the batch size and the number of gestures respectively. y_{ij} and $\hat{y}_{ij}$ are the ground truth and the predicted gesture labels, respectively. By minimizing $\mathcal{L}_2$, conSSL is able to achieve domain adaptation in target scenarios and learn the knowledge of gesture recognition.

4. Experiments

4.1. Datasets and settings

For evaluating the performance of our work in inter-session and inter-subject scenarios, we conducted experiments on high-density sEMG dataset *CapgMyo* [7] and sparse sEMG dataset *MyoArmbandDataset* [5]. Due to the difficulty of multi-session and multi-subject data collection, very few public datasets are available and these two datasets are the most suitable datasets that we can find to examine the proposed framework.

CapgMyo, which is collected by 8×16 electrode array at $1000Hz$, has 3 sub-datasets (DB-a, DB-b and DB-c). *DB-a* is for fine-tuning hyperparameters. *DB-b* that has 8 gestures are obtained from 10 subjects in two sessions on different days. *DB-c* contains 12 gestures of 10 subjects. Particularly, DB-b is used for inter-session and inter-subject evaluation while DB-c is just used for inter-subject evaluation. *MyoArmbandDataset* (abbreviated as *armband*) is recorded by 8-channels Myo Armband (Thalmic Labs) at $200Hz$, in which participants are comprised of 19 able-bodied participants for pretraining and 17 able-bodied participants for evaluation. The data of 7 gestures, each for 5s, form a circle and 4 circles are referred to a round. For a fair comparison, the evaluation followed previous work [16, 11, 5]. Before training and fine-tuning, sEMG signals were normalized on the time dimension [13].

All the details in our work are implemented by PyTorch and trained on a Linux server with two NVIDIA RTX 3090 GPUs. We chose temperature parameter $\tau = 0.5$ and number of epochs $E_1 = E_2 = 100$. Adam optimizer with weight decay of $1e - 6$ is used for both pretraining and fine-tuning. Batch sizes for fine-tuning are set to 16 and 32 for *CapgMyo* and *armband*, respectively. The code of our work are made publicly to facilitate communication and reproducibility in <https://github.com/laizhiping/ConSSL>.

4.2. Experimental results

In the two scenarios where domain shift problems exist, we conducted comprehensive experiments to examine the performance of the proposed framework with the number

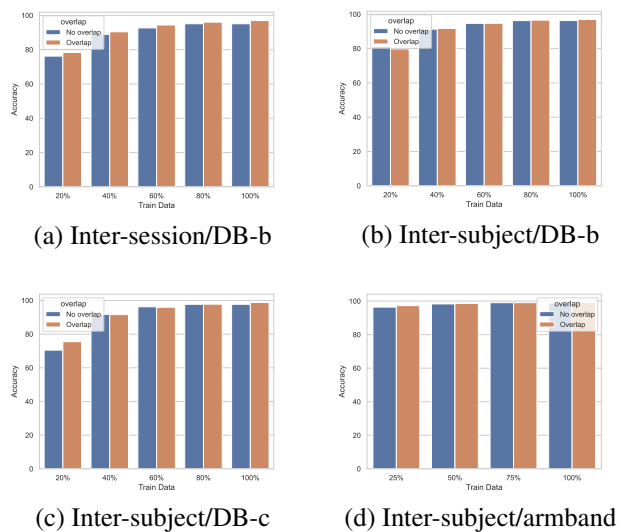


Figure 3. Performance with the increase of labeled sEMG samples in different scenarios. The recognition accuracy continuously improves with the increase of labeled samples, especially when a small number of labeled samples are available.

of labeled samples increasing. The experimental results are illustrated in Figure 3. We can see the trend that the performance of ConSSL improved when the labeled samples increasing, which means that knowledge of gesture and target domain can be learned gradually through fine-tuning using more labeled samples. The performance improvement is particularly obvious when few labeled samples are available ($\leq 50\%$). When the number of labeled samples increases, the improvement begins to slow down, which indicates that trade-off between performance and labor cost should be considered in practice. Furthermore, in two scenarios, ConSSL achieves the recognition accuracy of more than 90% even if only $\leq 50\%$ data are available in the target domain, i.e., only a small amount of samples in target domain need to be collected for domain adaptation, and good performance can be achieved, which shows excellent adaptability and the potential of reducing labor. In another way, we compare the performance of two different construction methods of positive pairs, overlapping way (e.g., $(x_{i,0}, x_{i,1})$ and $(x_{i,1}, x_{i,2})$) and non-overlapping way (e.g., $(x_{i,0}, x_{i,1})$ and $(x_{i,2}, x_{i,3})$). From Figure 3 we can see that there was no significant difference in performance, but the overlapping way slightly outperforms the non-overlapping way. The reason is that the overlapping way considers more positive signals and realizes information transmission due to the overlap of samples.

To verify the superiority of ConSSL, we compare it with some comparable methods that trained with 100% labeled samples in the target domain. From Table 1, we can see that ConSSL is always superior to methods and shows a great

Table 1. Comparison between different methods in two scenarios. The results in bold shows that our method outperforms other works and achieves state-of-the-art in two scenarios.

Scenario	Dataset	Method	Accuracy (%)
Inter-session	CapgMyo DB-b	DMA [7]	63.3
		2sRNN [11]	83.8
		Ours (40%)	90.4
		Ours	97.1
Inter-subject	CapgMyo DB-b	DMA [7]	55.3
		2sRNN [11]	89.9
		Ours (40%)	91.7
		Ours	97.3
	CapgMyo DB-c	DMA [7]	35.1
		2sRNN [11]	85.4
		Ours (40%)	91.6
		Ours	98.4
	Armband	LDA [5]	98.1
		DLTL [5]	98.3
		Ours (50%)	98.4
		Ours	99.0

advantage in tackling both two domain shift issues. Specifically, the improvements of **13.3%** and **7.4%**, **13.0%**, **0.7%** are realized in inter-session and inter-subject scenarios, respectively. If abundant labeled signals can be obtained for fine-tuning when calibrating, the recognition accuracy can be as high as that of offline training. Significantly, we show that our algorithm achieves better performance in two scenarios even if only $\leq 50\%$ of labeled data are available while counterparts need full labeled data. Our proposed ConSSL is superior to other works by **6.6%** with only 40% labeled samples in inter-session scenario, and in inter-subject scenario outperforms others by **1.8%**, **6.2%** and **0.1%** with 40% and 50% labeled sample. The comparison results shows that ConSSL is an excellent algorithm which achieves the state-of-the-art.

5. Conclusion

In this paper, we present a contrastive self-supervised learning framework for domain adaptation. To enhance the performance of our framework, we propose construction approaches of sEMG pairs during pretraining. A contrastive series loss is proposed by us to learn the general knowledge contained in sEMG signals. From experimental results, we can know that our work shows great advantage in tackling domain shift issues existing in inter-session scenario and inter-subject scenario, which is very helpful for practical application. Although a relatively small number of labeled samples ($\leq 50\%$) are available, the proposed framework still achieves $> 90\%$ recognition accuracy and outperforms the other works. The essential knowledge learned

during pretraining is the task-agnostic information, in our feature work, we will consider more sEMG-based tasks downstream besides gesture recognition. Furthermore, we will extend this work to other time-series tasks, e.g., electroencephalogram classification tasks.

References

- [1] J. Chen, F. Li, H. Chen, S. Yang, and Y. Wang. Dynamic gesture recognition using wireless signals with less disturbance. *Personal and Ubiquitous Computing*, 23(1):17–27, 2019.
- [2] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.
- [3] T. Chen, S. Kornblith, K. Swersky, M. Norouzi, and G. Hinton. Big self-supervised models are strong semi-supervised learners. *arXiv preprint arXiv:2006.10029*, 2020.
- [4] U. Côté-Allard, C. L. Fall, A. Campeau-Lecours, C. Gosselin, F. Laviolette, and B. Gosselin. Transfer learning for semg hand gestures recognition using convolutional neural networks. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1663–1668. IEEE, 2017.
- [5] U. Côté-Allard, C. L. Fall, A. Drouin, A. Campeau-Lecours, C. Gosselin, K. Glette, F. Laviolette, and B. Gosselin. Deep learning for electromyographic hand gesture signal classification using transfer learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(4):760–771, 2019.
- [6] U. Côté-Allard, G. Gagnon-Turcotte, A. Phinyomark, K. Glette, E. Scheme, F. Laviolette, and B. Gosselin. A transferable adaptive domain adversarial neural network for virtual reality augmented emg-based gesture recognition. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29:546–555, 2021.
- [7] Y. Du, W. Jin, W. Wei, Y. Hu, and W. Geng. Surface emg-based inter-session gesture recognition enhanced by deep domain adaptation. *Sensors*, 17(3):458, 2017.
- [8] J. Ducloux, P. Colla, P. Petrashin, W. Lancioni, and L. Toledo. Accelerometer-based hand gesture recognition system for interaction in digital tv. In *2014 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, pages 1537–1542. IEEE, 2014.
- [9] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu, and J. Li. Gesture recognition by instantaneous surface emg images. *Scientific reports*, 6(1):1–8, 2016.
- [10] S. Kanoga, T. Hoshino, and H. Asoh. Subject transfer framework based on source selection and semi-supervised style transfer mapping for semg pattern recognition. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1349–1353. IEEE, 2020.
- [11] I. Ketykó, F. Kovács, and K. Z. Varga. Domain adaptation for semg-based gesture recognition with recurrent neural networks. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7. IEEE, 2019.

- [12] M. Kim, W. K. Chung, and K. Kim. Subject-independent semg pattern recognition by using a muscle source activation model. *IEEE Robotics and Automation Letters*, 5(4):5175–5180, 2020.
- [13] Z. Lai, X. Kang, H. Wang, W. Zhang, X. Zhang, P. Gong, L. Niu, and H. Huang. Stcn-gr: Spatial-temporal convolutional networks for surface-electromyography-based gesture recognition. In *International Conference on Neural Information Processing*, pages 27–39. Springer, 2021.
- [14] E. Rahimian, S. Zabihi, A. Asif, S. F. Atashzar, and A. Mohammadi. Few-shot learning for decoding surface electromyography for hand gesture recognition. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1300–1304. IEEE, 2021.
- [15] E. Rahimian, S. Zabihi, S. F. Atashzar, A. Asif, and A. Mohammadi. Xceptiontime: independent time-window xceptiontime architecture for hand gesture classification. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1304–1308. IEEE, 2020.
- [16] J. Song, G. Sörös, F. Pece, and O. Hilliges. Real-time hand gesture recognition on unmodified wearable devices. In *IEEE Conference on computer vision and pattern recognition*. Citeseer, 2015.
- [17] W. Zhang, H. Wang, Z. Lai, and C. Hou. Constrained contrastive representation: Classification on chest x-rays with limited data. In *2021 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2021.
- [18] X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang, and J. Yang. A framework for hand gesture recognition based on accelerometer and emg sensors. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 41(6):1064–1076, 2011.
- [19] X. Zhang, X. Chen, W.-h. Wang, J.-h. Yang, V. Lantz, and K.-q. Wang. Hand gesture recognition and virtual game control based on 3d accelerometer and emg sensors. In *Proceedings of the 14th international conference on Intelligent user interfaces*, pages 401–406, 2009.
- [20] Y. Zhang, Y. Chen, H. Yu, X. Yang, and W. Lu. Learning effective spatial-temporal features for semg armband-based gesture recognition. *IEEE Internet of Things Journal*, 7(8):6979–6992, 2020.