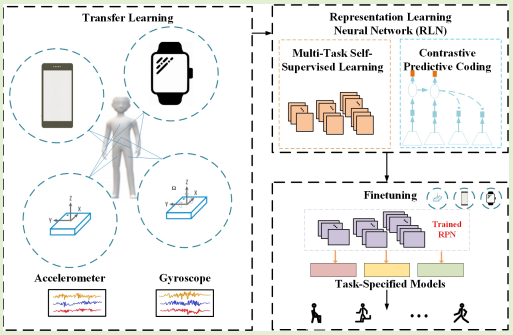
Transfer Learning Model Knowledge Across Multi-Sensors Locations Over Body Sensor Network

Xiaoye Qian, Huan Chen, Yi Cai, Kuo-Chung Chu, Wenyao Xu, and Ming-Chun Huang

***Abstract*—With the growth of sensing technologies, the sensors are applied to diversified fields including health care, elderly protection, human activity abnormal detection, and surveillance. The advanced sensors embedded in mobile devices generate a large amount of valuable data. In recent years, to deal with the massive volumes of data, representation learning emerged as an alternative approach to extract the features without manual feature extraction. In this paper, we develop an unsupervised representation learning system for mining features across multiple sensors placed on different parts of the human body for recognizing human daily activities. The unsupervised representation learning approach allows models to learn the feature representation among a large number of unlabeled data samples collected from different parts of the human body. In order to demonstrate the feasibility of our system, extensive experiments on human daily activities recognition are carried out to evaluate the effectiveness of the learned representations.**



***Index Terms*—Representation learning, transfer learning, deep learning, body sensor network.**

I. Introduction

THE growing availability of the data collected from smart mobile devices is changing the way of data analysis [1]. The sensors are widely used in body sensor networks (BSNs) technology. As sensing technologies develop, the emerging advanced sensors are embedded and extensively applied to the mobile devices for a wide range of applications such as health care [2], [3], elderly protection [4], [5], human activity abnormal detection [6], surveillance [7], and eating detection [8]. The advanced sensors embedded in mobile devices generate a large amount of valuable data. It provides information on human physical activities, evaluation on individual's independence, and health status [9], [10].

Convolutional neural networks (CNN) based systems have been widely recognized as the efficient way to detect human activities with human-annotated labels [5]. Millions of labeled training data are the foundations of great success. However, collecting the massive volumes of labeled data requires manual annotation, which is time-consuming and expensive. It is crucial to train the model from a large number of unlabeled data for human activities recognition. Besides, the unlabeled data can be complicated due to the data are acquired from different mobile devices and from various parts of the human body. The signal waveform generated from different mobile device positions can be dissimilar [11]. A person while jogging holding the smartphone in hand is expected to wave the device more than placing it in the pocket. Besides, the orientation of mobile devices can be different. Those differences in data increase the difficulty of using the data collected from different positioning directly.

Learning reusable discriminative representations across different sensors has become an active research domain [12]. There is an emerging paradigm named representation learning based on the deep learning approach, which converts the

Manuscript received January 13, 2022; accepted March 30, 2022. Date of publication April 11, 2022; date of current version May 31, 2022. This work was supported by the 2021-2022 Kunshan Government Research Fund under Grant R97030024S. The associate editor coordinating the review of this article and approving it for publication was Prof. Pierluigi Salvo Rossi. *(Corresponding authors: Ming-Chun Huang; Wenyao Xu.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Duke Kunshan University Institutional Review Board (DKU IRB) under IRB No. FWA00021580.

Xiaoye Qian, Huan Chen, and Yi Cai are with the Department of Electrical and Computer Science, Case Western Reserve University, Cleveland, OH 44106 USA (e-mail: xxq82@case.edu; hxc556@case.edu; yxc757@case.edu).

Kuo-Chung Chu is with the Department of Information Management, National Taipei University of Nursing & Health Sciences, Taipei 112, Taiwan (e-mail: kcchu@ntunhs.edu.tw).

Wenyao Xu is with the Department of Computer Science and Engineering, State University of New York at Buffalo, Buffalo, NY 14260 USA (e-mail: wenyaoxu@buffalo.edu).

Ming-Chun Huang is with the Department of Data and Computational Science, Duke Kunshan University, Suzhou, Jiangsu 215316, China (e-mail: mh596@duke.edu).

Digital Object Identifier 10.1109/JSEN.2022.3166187

data into relevant critical features and can be further utilized for activity recognition or classification [13]. While the deep learning-based approaches have led to significant performance improvement, the model's depth and complexity are limited by the quantity and quality of well-annotated data [14]. It is typically straightforward to get large quantities of unlabeled data directly from mobile devices. The deep learning-based unsupervised representation learning algorithms relieve the burden of massive manual annotation [15]. Specifically, most of the self-supervised methods are proposed for learning representations by solving the "pretext" tasks from the data itself, which significantly reduces the need of domain expertise and the substantial effort to manually annotate the training data.

In this paper, we aim to extract the signal representations of the sensor data collected from the mobile devices placed on different human body parts. The learned representations can be transferred across the mobile devices in different positions. To further demonstrate our system, we apply the framework for learning representations on both self-collected public benchmark datasets including fall detection (FD) and Activities of Daily Living (ADLs). Extensive experiments are implemented and validated to demonstrate the feasibility and effectiveness of the systems that are aided by learned representations. We summarize the following contributions. First, we build an unsupervised representation learning system to extract the generalizable features through multiple sensor positions and analyze two different representation learning frameworks, the multi-task self-learning and Contrastive Predictive Coding (CPC). Second, we analyze the effectiveness of applied learned representations on human daily activities recognition (HAR) including fall detection and activities of daily living. Third, we demonstrate that fine-tuning the pre-trained model with limited available well-annotated data improves the model accuracy significantly compared to the direct training from those annotated data.

II. Related Work

As the revolution of intelligence devices with sophisticated hardware progresses, a large amount of valuable data are generated from mobile devices. Mobile-based machine learning has been one important area of research in recent years [16]. The mobile sensing enables applications to quantify the user's exercise patterns [17], [18], monitor the elderly people's falls [11], and Human daily activities recognition [19]. With the increasing computation and storage capabilities of mobile devices, the utilization of mobile devices with sensors for those applications has become an effective approach. Recently, CNNs have been successfully applied in sensor-based human activities detection systems [5]. However, it is expensive and time-consuming to get a large amount of well-curated or human-annotated data for learning. An alternative approach to model learning is self-supervised representation learning, where the models extract the deep feature without the requirement of a large number of well-curated or human-annotated data.

A. Representation Learning

One of the fundamental data mining methods is to develop models to extract a meaningful set of features from the data. The extracted features are utilized within machine learning algorithms to learn the mapping with those features [20]. Representation learning has been widely applied for the computer vision and natural language processing problems, such as learning the representations from colorization [21], predicting image rotations [22], and context-based self-supervised learning [23]. The role of the representations in activity recognition is investigated by applying divergent features to train the model [24].

B. Self-Supervised Representation Learning of HAR

Due to the ease of collection and set up [25], as well as the protection of privacy [26], the sensor-based HAR approaches have become more prevalent and are widely used all over the world [27]. The common sensor modalities include accelerometers, gyroscopes, and magnetometers [28], [29]. Traditional approaches have made tremendous progress on HAR by applying data-driven machine learning algorithms including decision tree [30], support vector machine [31], naive Bayes [32], and hidden Markov models [33]. It is necessary to apply the feature extraction algorithms before building machine learning models to extract proper and important features for subsequent modeling procedure [34]. However, the process to extract the features from the data of complex correlations and non-linearity is complicated [35]. The feature extraction of the HAR system highly relies on domain-specific knowledge, which is time-consuming and expensive. In most daily HAR tasks, the methods highly rely on hand-crafted extracted features, which requires proficient knowledge in that domain.

Representation learning is typically extracting features from the input data [35]. Most of the methods find compact representations for the sensor data. The representations can be extracted through time and frequency domain [36]. The representation learning approaches such as principal component analysis (PCA) have been developed to extract the features from the sensor signal [37]. Other representation learning models based on deep learning are formed by the composition of multiple nonlinear transformations. The models include deep belief network (DBN) [38] and stacked auto-encoder (SAE) [39]. Recently, self-supervised learning approaches have been implemented by defining a "pretext" task from the data itself. Those approaches have been widely applied in computer vision and natural language processing problems. Multi-task self-supervised representation learning can learn the shared representations across various tasks by collaboratively optimizing multiple objectives. Biologically, multitask learning can be regarded as exploiting relations among tasks and applying the knowledge that has been acquired by the surrogate or auxiliary tasks to the new tasks. Peng *et al.* [10] applied the multitask learning on ADLs for exploring the relations between simple activity and complex activity. Saeed *et al.* [12] applied multi-task learning to find

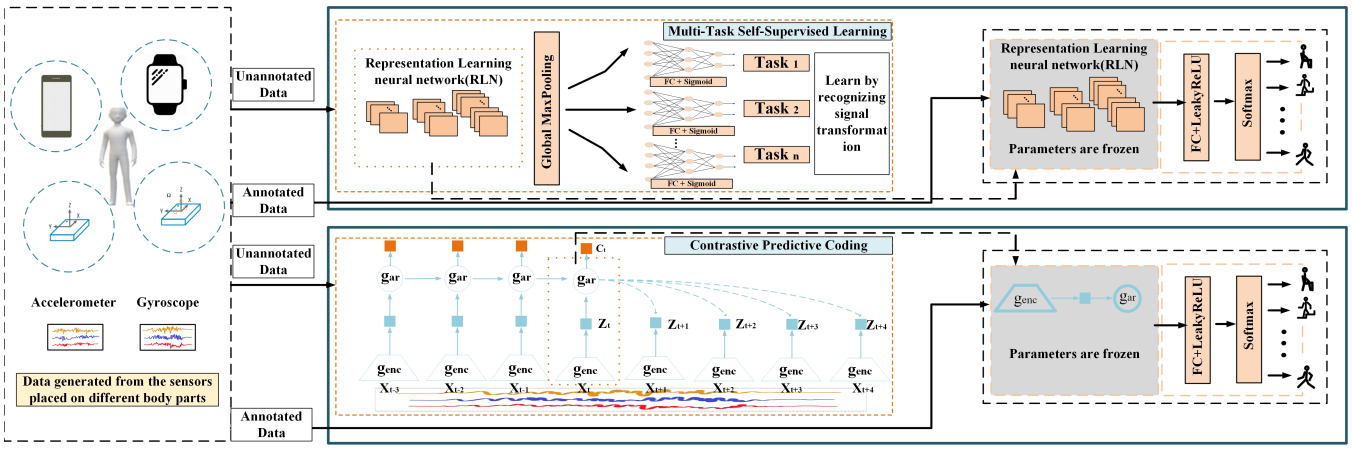


Fig. 1. The figure illustrates the work flow of the system. The representation are learning from the unlabeled data collected from sensors placed on different body parts. The learned representations are applied to human daily activities recognition and fine-tuning the model with annotated data.

the representative relations of accelerometer signals. Contrastive predictive coding (CPC) [40] is a self-supervised approach and encodes high-level temporal information from the time-series data. The data are mapped to a latent space followed by the auto-regressive network for the subsequently encoded context representations.

III. System

In this paper, we apply the representation learning, the multi-task self-supervised learning and the CPC framework, to extract the representations across the sensor data collected from different body parts. The system framework is illustrated in Fig. 1.

A. Multi-Task Self-Supervised Learning

To learn the information from a large amount of the unlabeled data from different sensors that are placed on different human body parts, the "pretext" tasks are defined in order to perform unsupervised representation learning. The tasks are designed to recognize the signal transformations motivated by references [12], [41]. All the data from different sensor positions are collected together and applied to the signal transformations before feeding into the representation learning network (RLN). The RLN contains three 5*1 convolutional layers with 32, 64, 96 feature maps. The dropout layer is used after each of the convolutional layers. The distinct task-specific layers for each task are implemented by applying multilayer perceptron with 2-hidden layers including 1728 and 512 neural nodes and are followed by the LeakyReLu activation functions. Additionally, we apply the early-stopping if the network fully converges to avoid over-fitting. The L2 regularization is applied with a rate of 0.0003. The model consists of shared layers transferred from the RLN and the distinct last two layers with 1728 and 512 units. The RLN learns the representations of sensor data and the feature mappings through the signal transformations classification. The multi-task learning (MTL) framework is applied to distinguish the signal transformations. The common trunk (shallow layers) neural network of the MTL model is able to learn the generic representation and extract the representations of multiple sensor data. We define multiple T tasks according to detecting whether the sensor data has been transformed and classifying which transformations have taken place. The whole network is divided into two parts by commonly shared trunk and task-specify distinct layers. The design enables the learning process to push extracted task-specific features to the last distinct layer and let the shared layers extract generic representations through recognizing the signal transformations for all sensor data collected from different human body parts. The learned representations can be further transferred to perform HAR with different body-part sensor models.

The signal transformations including noising, scaling, rotating, cropping, permuting, time-warping, magnitude-warping, and the combination of permuting and time-warping are implemented. More specifically, the random noise transformation (or jitter) adds noise to the original signal. The model can capture the features of minor signal changes through noise transformation. The scaling transformation enables the model to learn the features of amplitude and offset in-variances from changing the magnitude of the signals. Besides, the rotation transformation enables the model to learn the orientation in-variances through implementing the inversion to the signals. The time-warping transformation makes the model learn the features of temporal location by distorting the time intervals between samples in the windows. The magnitude-warping transformation involves the data window with a smooth curve around 1. What is more, the magnitude transformation facilitates the model to learn the features of convolution changing around the samples. In addition, the random sampling and the permutation transformation perturb the signal waveforms by removing and permuting some data points in the signal waveforms respectively, which allows the model to learn the features of the relative sampling rate between the raw signal and transformed signal.

Each data signal collected from the different positioning mobile devices $e \in E$ are transformed $J(x_i)$ before feeding into the representation learning neural network, where the

signal transformations are defined as $J(.)$. The model is feed-forwarding to process the different tasks simultaneously. where the loss function for signal transformation recognition tasks is defined in equation 1. We define the y_t to be the output of the model with a set of n training instances.

$$L(w) = \sum_{t \in T} \phi_t (-\frac{1}{n_t} \sum_{i=1}^{n_t} (y_i (log(y_t)) + (1 - y_i) log(1 - y_t)) + \beta ||w||^2) \quad (1)$$

The weights of RLN are updated according to equation 2.

$$w_i^{new} = w_i^{old} - \alpha \left[\frac{1}{m} \nabla_{w_i^{(old)}} L(w) \right] \quad (2)$$

In equation 2, m is the mini-batch data size, α is the learning rate.

When the training converges, the shallow common trunk of the RLN model can be transferred to any HAR models trained with different sensor data collected from different human body parts. The last task-specific layers can be fine-tuned according to the needs of the task.

Algorithm 1 Representation Learning Across Sensors Placed on Different Body Parts

Input: Unlabeled dataset and labeled dataset: N_u, N_l
Output: Human activity recognition model

```
/*Initialization for representation learning*/
Establish the representation learning model by defining the "pretext" tasks.
/*Training*/
for each traning instance x_i in N_u do
    for each task t ∈ T do
        Learning the representations through a CPC-based model or through a multi-task-based model.
/*Transferring the learned representations to the HAR*/
The representation learning network is transferred to the HAR model by freezing the parameters.
while labeled sensor data e ∈ E do
    for each mini-batch input n ∈ N_l do
        Fine-tune the last layers of human activity recognition.
```

B. Contrastive Predictive Coding

The key of contrastive predictive coding (CPC) is to predict future time-steps in the latent space using auto-regressive modeling [40]. The assumption behind doing such a task is to extract high-level features by predicting the future time-steps, which is an established approach in signal processing [42]. The framework of CPC is illustrated in Fig. 1. The encoder model is defined as g_{enc} and transforms the signal to a sequence of latent representations. Besides, an auto-regressive model g_{ar} is designed to extract the context of latent representations of the output of encoder model $z_t = g_{enc}(x_t)$. The noise contrastive estimation (NCE) loss is utilized to update the model parameters according to Equation 4.

$$L_{NCE} = -\mathbb{E}_x \left[log \frac{f(x_{t+k}, c_t)}{\sum_{x_j \in X} f(x_j, c_t)} \right] \quad (3)$$

where $f(x_{t+k}, c_t)$ is the density ratio to preserves the mutual information between the x_{t+k} (the signal at time $t + k$), and the c_t, the context latent representations learned by the auto-regressive model $c_t = g_{ar}(z_{\leq t})$.

$$f(x_{t+k}, c_t) \propto \frac{p(x_{t+k}|c_t)}{p(x_{t+k})} \quad (4)$$

After training the representation learning model, the parameters of the g_{ar} and g_{enc} are frozen and to be transferred to the classification model to extract the signal representations on HAR.

C. Representations Transferring

In the above section, the representation learning frameworks are introduced. The subsequent procedure is to do the classification based on those learned representations. When the training process of the representation learning becomes convergence, the parameters of the model are transferred to the classification model with the frozen parameters. The classification backend includes batch normalization layer, LeakyReLU activation layer, and dropout layer with $p = 0.1$. Generally, the representation learning model is to learn the generic representations from the abundant unlabeled sensor data. The classification model is transferred from the representation learning model, and then fine-tuning the last task-specific layers with labeled sensor data. The summary of system workflow is concluded in Algorithm 1.

IV. Experiment and Results

A. Datasets and Evaluation Metrics

The experiments are conducted to validate the representations learned from different body-part sensor datasets. The system is evaluated with a self-collected dataset and a publicly available dataset. The HHAR dataset [43] collects 36 smartphones and smartwatches, consisting of 13 different device models. Two embedded sensors including the accelerometer and gyroscope signals are used to recognize the activities of daily living (ADLs) including biking, sitting, standing, walking, stair up, and stair down. To further evaluate the system, more data are self-collected. 10 ADLs on 10 subjects are collected including sitting down, sitting (sitting on the chair and sitting on the sofa), walking (walking upstairs, walking downstairs, and walking on the ground), bending, bending to pick up items, standing, lying, squatting and squatting to pick up items and a group of falls containing backward falls, forward hard falls, forward soft falls, left falls and right falls as described in [44]. All the data is collected from the accelerometer and gyroscope placed in the trouser's pocket (thigh), on the arm, and on the chest. All signals are re-sampled to 50 Hz before training according to the literature [45] and adopted the segmentation technique around founded peak [46] into fixed windows with 200 samples and 50% overlap for human activity recognition and fall detection.

TABLE I
COMPARING THE PERFORMANCE OF MODELS TRANSFERRED FROM REPRESENTATION LEARNING AND SUPERVISED LEARNING WITH TWO DIFFERENT ACCELEROMETER SENSOR POSITIONS ON PUBLIC HHAR DATASETS

Algorithms	Sensor Positions	Labeled Data Per Class	Accuracy	Precision	Recall	F1-Score
Supervised Learning	Thigh	50	0.4918	0.4487	0.4686	0.4464
		100	0.5843	0.5630	0.5647	0.5486
		500	0.7735	0.7775	0.7593	0.7614
	Arm	50	0.6003	0.5878	0.6047	0.5890
		100	0.6752	0.6682	0.6851	0.6682
		500	0.7620	0.7615	0.7639	0.7559
Multi-task Self-supervised Learning	Thigh	50	**0.7404**	**0.7206**	**0.7239**	**0.7187**
		100	**0.7926**	**0.7813**	**0.7797**	**0.7791**
		500	**0.8799**	**0.8743**	**0.8709**	**0.8716**
	Arm	50	0.6772	0.6745	0.6811	0.6755
		100	0.7098	0.7103	0.7152	0.7106
		500	0.8032	0.8013	0.8058	0.8032
CPC Representation Learning	Thigh	50	0.7328	0.7132	0.7167	0.7115
		100	0.7861	0.7745	0.7732	0.7745
		500	0.8744	0.8688	0.8655	0.8661
	Arm	50	0.6475	0.6439	0.6512	0.6458
		100	0.7002	0.7005	0.7056	0.7009
		500	0.7820	0.7828	0.7873	0.7847

The same sampling rate and the fixed-length sliding window enable the model to learn the features related to the location and make it more feasible to transfer between different devices. Other than the public dataset, the number of self-collected data samples is 3380 and a total of 30420 samples after applying the signal transformations.

To better evaluate and assess the performance of the multi-classification in the experiments, the metric of overall accuracy, Macro-precision, Macro-recall, and Macro-F1 [47], [48] is used. Assume there is a C classes:

$$Accuracy = \sum \frac{TP_c + TN_c}{TP_c + FN_c + FP_c + TN_c} \tag{5}$$

$$Macro - precision = \frac{1}{C} \sum_{c \in C} \frac{TP_c}{TP_c + FP_c} \tag{6}$$

$$Macro - recall = \frac{1}{C} \sum_{c \in C} \frac{TP_c}{TP_c + FN_c} \tag{7}$$

$$Macro - F1 = \frac{1}{C} \sum_{c \in C} \frac{2TP_c}{2TP_c + FN_c + FP_c} \tag{8}$$

B. Results

1) Demonstrate the Effectiveness of the Learned Sensor Representations: To evaluate the effectiveness of the learned representations across the sensors placed on different body parts, the experiments are conducted on two HAR datasets including the public benchmark dataset HHAR and self-collected dataset.

As shown in Table I, representation learning increases the performance of activity recognition, especially under the scenario that the available labeled data are limited per class compared to using supervised learning. A 5-folder cross-validation is implemented for each dataset. Both representation learning architectures have a better performance than the system trained from scratch on the HHAR dataset with the number of available labeled data per class equal to 50, 100, and 500. Transferring the representation model to the HAR model can improve the performance of human daily activities with the smartphone (placed on the thigh), even the available label is limited. The accuracy of the multi-task self-supervised learning on 50 available labeled data per class is 74.04%, the 100 labeled data per class is 79.26%, and 500 labeled data per class, which is increased 24.86%, 20.83% and 10.64% to the accuracy than the supervised learning, respectively. The representation learning achieves attractive performance on the HHAR dataset. In addition, the model transferred from learned representation has a better convergence rate.

As shown in Table II, the learned representations are applied to HAR. The representations learning can improve the system performance on human daily activities recognition. We get the impressive results that learning the representations from sensors increases the system accuracy, especially when the available data are limited. The representation learning based on CPC can increase 22% of accuracy compared to train model from scratch. The consistently improved performance over training from scratch makes a compelling indicator that the representation learning algorithm extracts important features across sensor data collected from different body parts and makes a significant improvement in performance over random weights of the neural network.

2) Validating Representation Learning on Gyroscope: In this section, we apply the representation learning on gyroscope data collected from different human body parts to validate the effectiveness of the representation learning on the gyroscope sensor. The experiments on HHAR dataset demonstrate that the learned representations can still improve the performance of gyroscope data. As shown in Table III, where the ACC, P, R, and F1 refer to Accuracy, Macro-Precision, Macro-Recall, and Macro F1-Score. Despite the improvement of gyroscope performance, accuracy, precision, recall, and F1-score are not as significant as the accelerometer, the results demonstrate that the beneficial representations can be extracted from both gyroscope and accelerometer sensor data.

TABLE II
COMPARING THE PERFORMANCE OF MODELS TRANSFERRED FROM REPRESENTATION LEARNING AND SUPERVISED LEARNING WITH TWO DIFFERENT ACCELEROMETER SENSOR POSITIONS ON SELF-COLLECTED DATASETS

Algorithms	Sensor Positions	Labeled Data Per Class	Accuracy	Precision	Recall	F1-Score
Supervised Learning	Thigh	10	0.6263	0.5986	0.6776	0.5715
		30	0.7858	0.7256	0.8381	0.7421
		100	0.9167	0.8437	0.8475	0.8417
	Arm	10	0.4272	0.5268	0.5518	0.4298
		30	0.7598	0.7015	0.7974	0.7203
		100	0.8504	0.7997	0.7710	0.7745
	Waist	10	0.5571	0.5068	0.6604	0.5299
		30	0.6949	0.6333	0.7839	0.6689
		100	0.8740	0.7835	0.8074	0.7645
Multi-task Self-supervised Learning	Thigh	10	0.7146	0.6548	0.7324	0.6560
		30	0.8366	0.7927	0.8547	0.8090
		100	**0.9198**	**0.8713**	**0.8482**	**0.8409**
	Arm	10	0.6378	0.5844	0.6566	0.5760
		30	0.7677	0.7050	0.7936	0.7254
		100	0.8740	0.8129	0.7720	0.7751
	Waist	10	0.6634	0.6031	0.6966	0.6169
		30	0.7527	0.6714	0.7998	0.6976
		100	0.8858	0.8123	0.8137	0.7902
CPC Representation Learning	Thigh	10	**0.8473**	**0.7765**	**0.7841**	**0.7696**
		30	**0.8675**	**0.7811**	**0.8194**	**0.7763**
		100	0.9148	0.8689	0.8614	0.8429
	Arm	50	0.6475	0.6439	0.6512	0.6458
		30	0.7002	0.7005	0.7056	0.7009
		100	0.8669	0.8336	0.8738	0.8409
	Waist	10	0.7284	0.8087	0.6718	0.6936
		30	0.8355	0.7979	0.8361	0.8093
		100	0.8669	0.8336	0.8738	0.8409

TABLE III
REPRESENTATION LEARNING FOR HHAR GYROSCOPE DATA

Arm	Acc	P	R	F1
Supervised: 50 labeled	0.4700	0.4565	0.4712	0.4304
Representations: 50 labeled	0.4980	0.5004	0.4963	0.4905
Supervised: 100 labeled	0.5035	0.4998	0.4981	0.4726
Representations: 100 labeled	0.5447	0.5398	0.5401	0.5376
Thigh	Acc	P	R	F1
Supervised: 50 labeled	0.5311	0.5267	0.5179	0.5159
Representations: 50 labeled	0.5251	0.5189	0.5143	0.5055
Supervised: 100 labeled	0.5611	0.5609	0.5404	0.4883
Representations: 100 labeled	0.5971	0.5850	0.5874	0.5496

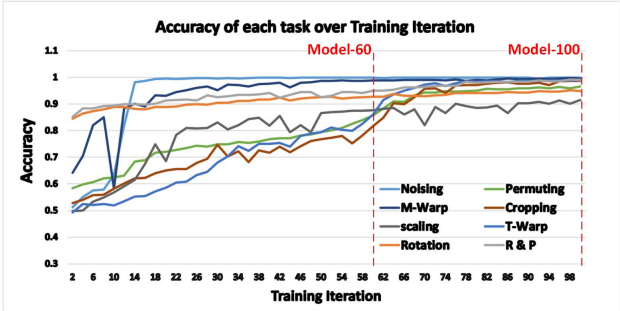


Fig. 2. This figure shows the accuracy of each signal transformation tasks during the representation learning based on multi-task learning with our dataset.

V. DISCUSSION

A. The Influence of Accurate Learned Representation on Classification Model

We have demonstrated the effectiveness of the representation learning on the classification task under realistic, challenging requirements. In this section, the influence of accurate learned representation on the classification model is discussed. Weather more accurate representations lead to a better performance of the classification task? Will there be an over-fitting problem? As shown in Figure 2, which illustrates the tendency of the accuracy of each signal transformation task along with the training iterations on the self-collected dataset. As shown in the figure, model-60 refers to the RLN network trained by 60 iterations and model-100 refers to the RLN network trained by 100 iterations. Apparently, model-100 has better performance on recognizing the signal transformations than model-60. We further evaluate whether a better representation learning model can result in a better performance.

As shown in Fig. 3, as expected, the performance of model-100 is better than both model-60 and supervised learning models. Better performance of representation learning leads to higher accuracy of the classification task, especially with the limited available labeled data. At the same time, we found that the more accurate the representation learned, the better performance with a higher convergence rate on the classification model.

B. Could the Learned Representation Applied to Different Users?

To further evaluate the effectiveness of the representations learned from sensors, we evaluate the model for the new incoming client (leave-one-out cross-validation). As shown in the Fig. 4, the learned representation significantly improves the performance of the HAR model even for a new incoming user. In particular, the representation learning model has better performance, which increases the accuracy of 12%, the precision of 20%, improves the recall of 16%, and the F1 score of 18% than the supervised learning with limited available private user sensor data.

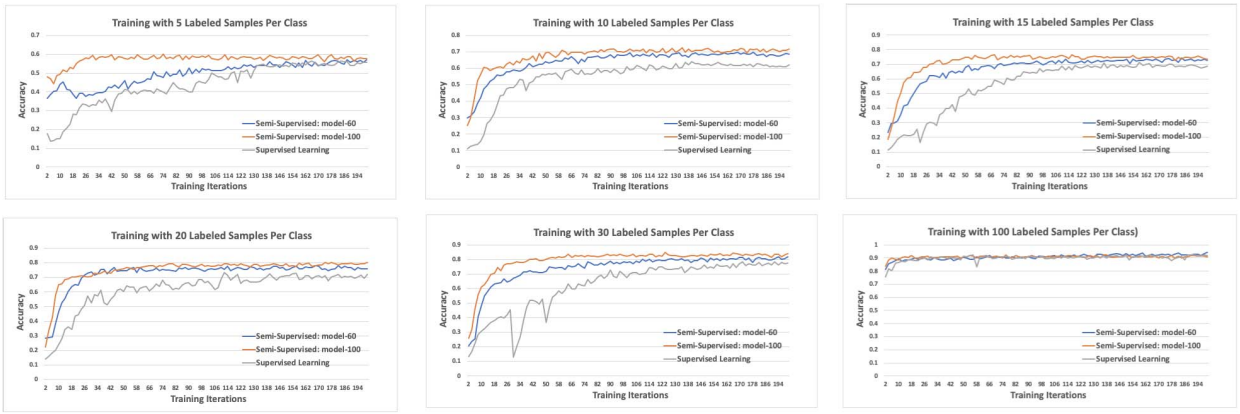


Fig. 3. The results of the representation learning. The representation learning model is pre-trained on an entire dataset from different location sensor data with 5, 10, 15, 20, 30, and 100 labeled instances per class.

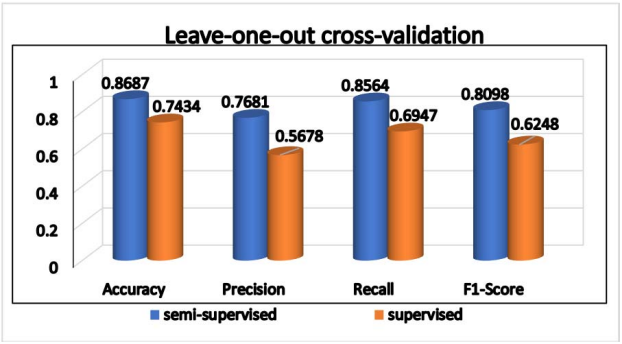


Fig. 4. Evaluate the performance of representation learning across different user.

TABLE IV
COMPUTATION COMPLEXITY FOR REPRESENTATION LEARNING MODELS

Model	FLOPs	Parameters
CPC	214.982 MB	3.7478 MB
Multi-task	29.649 MB	0.605 MB

C. Computation Complexity

As shown in the Table IV, the computation complexity is summarized. The float point operations (FLOPs) of CPC representation learning and multi-task learning models are calculated. Besides, the task-specific layers (last layers) take about 3.9 MB FLOPs.

VI. CONCLUSION

In this paper, we applied representation learning for different body-part sensors. The representation learning system is proposed to learn the representations from a large amount of unlabeled data collected from sensors placed on different body parts. Collecting the data from all different body parts make it easier to get a large quantity of sensor data. The collective data across sensors has a potential value not only can improve the system, but also has business value. With representation learning, the model learned from the unlabeled data can improve the system performance especially when the available well-curated data are limited to a specific body part. We found that learning the representations from the data can facilitate different body-part sensor models and significantly improve the learning efficiency and effectiveness. The system is validated through both public and self-collected datasets. As expected, the performance of the proposed system has a great advantage over using traditional supervised learning.

ACKNOWLEDGMENT

The funders had no role in the study design; data collection, analysis, or interpretation; in the writing of the report; or in the decision to submit the article for publication.

REFERENCES

[1] M. M. Dhanvijay and S. C. Patil, "Internet of Things: A survey of enabling technologies in healthcare and its applications," *Comput. Netw.*, vol. 153, pp. 113–131, Apr. 2019.
[2] P. Gope and T. Hwang, "BSN-Care: A secure IoT-based modern healthcare system using body sensor network," *IEEE Sensors J.*, vol. 16, no. 5, pp. 1368–1376, Mar. 2016.
[3] A. Subasi, M. Radhwan, R. Kurdi, and K. Khateeb, "IoT based mobile healthcare system for human activity recognition," in *Proc. 15th Learn. Technol. Conf. (LT)*, Feb. 2018, pp. 29–34.
[4] G. Sebestyen, I. Stoica, and A. Hangan, "Human activity recognition and monitoring for elderly people," in *Proc. IEEE 12th Int. Conf. Intell. Comput. Commun. Process. (ICCP)*, Sep. 2016, pp. 341–347.
[5] X. Qian, H. Chen, H. Jiang, J. Green, H. Cheng, and M.-C. Huang, "Wearable computing with distributed deep learning hierarchy: A study of fall detection," *IEEE Sensors J.*, vol. 20, no. 16, pp. 9408–9416, Aug. 2020.
[6] N. Irvine, C. Nugent, S. Zhang, H. Wang, and W. W. Y. Ng, "Neural network ensembles for sensor-based human activity recognition within smart environments," *Sensors*, vol. 20, no. 1, p. 216, Dec. 2019.
[7] K.-E. Ko and K.-B. Sim, "Deep convolutional framework for abnormal behavior detection in a smart surveillance system," *Eng. Appl. Artif. Intell.*, vol. 67, pp. 226–234, Jan. 2018.
[8] A. Bedri *et al.*, "EarBit: Using wearable sensors to detect eating episodes in unconstrained environments," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 3, pp. 1–20, Sep. 2017.
[9] D. Castro, W. Coral, C. Rodriguez, J. Cabra, and J. Colorado, "Wearable-based human activity recognition using an IoT approach," *J. Sensor Actuator Netw.*, vol. 6, no. 4, p. 28, Nov. 2017.
[10] L. Peng, L. Chen, Z. Ye, and Y. Zhang, "AROMA: A deep multi-task learning based simple and complex human activity recognition method using wearable sensors," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 2, pp. 1–16, 2018.
[11] A. T. Özdemir, "An analysis on sensor locations of the human body for wearable fall detection devices: Principles and practice," *Sensors*, vol. 16, no. 8, p. 1161, 2016.

[12] A. Saeed, T. Ozcelebi, and J. Lukkien, "Multi-task self-supervised learning for human activity detection," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 3, no. 2, pp. 1–30, Jun. 2019.

[13] P. A. Colpas, E. Vicario, E. De-La-Hoz-Franco, M. Pineres-Melo, A. Oviedo-Carrascal, and F. Patara, "Unsupervised human activity recognition using the clustering approach: A review," *Sensors*, vol. 20, no. 9, p. 2702, May 2020.

[14] F. Zhuang, X. Cheng, P. Luo, S. J. Pan, and Q. He, "Supervised representation learning: Transfer learning with deep autoencoders," in *Proc. 24th Int. Joint Conf. Artif. Intell.*, 2015, pp. 1–7.

[15] Z. Feng, C. Xu, and D. Tao, "Self-supervised representation learning from multi-domain data," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3245–3255.

[16] E. Bulbul, A. Cetin, and I. A. Dogru, "Human activity recognition using smartphones," in *Proc. 2nd Int. Symp. Multidisciplinary Stud. Innov. Technol. (ISMSIT)*, Oct. 2018, pp. 1–6.

[17] S. Gleadhill, J. B. Lee, and D. James, "The development and validation of using inertial sensors to monitor postural change in resistance exercise," *J. Biomechanics*, vol. 49, no. 7, pp. 1259–1263, May 2016.

[18] M. O'Reilly, B. Caulfield, T. Ward, W. Johnston, and C. Doherty, "Wearable inertial sensor systems for lower limb exercise detection and evaluation: A systematic review," *Sports Med.*, vol. 48, no. 5, pp. 1221–1246, May 2018.

[19] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognit. Lett.*, vol. 119, pp. 3–11, Mar. 2019.

[20] S. Sprager and D. Zazula, "A cumulant-based method for gait identification using accelerometer data with principal component analysis and support vector machine," *WSEAS Trans. Signal Process.*, vol. 5, no. 11, pp. 369–378, 2009.

[21] R. Zhang, P. Isola, and A. A. Efros, "Colorful image colorization," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, Oct. 2016, pp. 649–666.

[22] S. Gidaris, P. Singh, and N. Komodakis, "Unsupervised representation learning by predicting image rotations," 2018, *arXiv:1803.07728*.

[23] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 3111–3119.

[24] H. Haresamudram, D. V. Anderson, and T. Plötz, "On the role of features in human activity recognition," in *Proc. 23rd Int. Symp. Wearable Comput.*, Sep. 2019, pp. 78–88.

[25] M. Quwaider and Y. Jararweh, "An efficient big data collection in body area networks," in *Proc. 5th Int. Conf. Inf. Commun. Syst. (ICICS)*, Apr. 2014, pp. 1–6.

[26] A. Jain and V. Kanhangad, "Human activity classification in smartphones using accelerometer and gyroscope sensors," *IEEE Sensors J.*, vol. 18, no. 3, pp. 1169–1177, Feb. 2018.

[27] K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu, and Y. Liu, "Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities," *ACM Comput. Surv.*, vol. 54, no. 4, pp. 1–40, May 2022.

[28] Y. Wang, S. Cang, and H. Yu, "A data fusion-based hybrid sensory system for older people's daily activity and daily routine recognition," *IEEE Sensors J.*, vol. 18, no. 16, pp. 6874–6888, Aug. 2018.

[29] M. Munoz-Organero, "Human activity recognition based on single sensor square HV acceleration images and convolutional neural networks," *IEEE Sensors J.*, vol. 19, no. 4, pp. 1487–1498, Feb. 2019.

[30] L. Fan, Z. Wang, and H. Wang, "Human activity recognition model based on decision tree," in *Proc. Int. Conf. Adv. Cloud Big Data*, Dec. 2013, pp. 64–68.

[31] D. N. Tran and D. D. Phan, "Human activities recognition in Android smartphone using support vector machine," in *Proc. 7th Int. Conf. Intell. Syst., Modelling Simulation (ISMS)*, Jan. 2016, pp. 64–68.

[32] X. Yang and Y. L. Tian, "EigenJoints-based action recognition using Naïve-Bayes-nearest-neighbor," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2012, pp. 14–19.

[33] C. A. Ronao and S.-B. Cho, "Recognizing human activities from smartphone sensors using hierarchical continuous hidden Markov models," *Int. J. Distrib. Sensor Netw.*, vol. 13, no. 1, 2017, Art. no. 1550147716683687.

[34] X. Yuan, L. Ye, L. Bao, Z. Ge, and Z. Song, "Nonlinear feature extraction for soft sensor modeling based on weighted probabilistic PCA," *Chemometrics Intell. Lab. Syst.*, vol. 147, pp. 167–175, Oct. 2015.

[35] S. R. Ramamurthy and N. Roy, "Recent trends in machine learning for human activity recognition—A survey," *WIREs, Data Mining Knowl. Discovery*, vol. 8, no. 4, p. e1254, Jul. 2018.

[36] D. Figo, P. C. Diniz, D. R. Ferreira, and J. M. P. Cardoso, "Preprocessing techniques for context recognition from accelerometer data," *Pers. Ubiquitous Comput.*, vol. 14, no. 7, pp. 645–662, 2010.

[37] T. Plötz, N. Y. Hammerla, and P. L. Olivier, "Feature learning for activity recognition in ubiquitous computing," in *Proc. 22nd Int. Joint Conf. Artif. Intell.*, 2011, pp. 1–6.

[38] L. M. Dang, K. Min, H. Wang, M. J. Piran, C. H. Lee, and H. Moon, "Sensor-based and vision-based human activity recognition: A comprehensive survey," *Pattern Recognit.*, vol. 108, Dec. 2020, Art. no. 107561.

[39] B. Almaslukh, J. Almuhtadi, and A. Artoli, "An effective deep autoencoder approach for online smartphone-based human activity recognition," *Int. J. Comput. Sci. Netw. Secur.*, vol. 17, no. 4, pp. 160–165, 2017.

[40] A. van den Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," 2018, *arXiv:1807.03748*.

[41] T. T. Um *et al.*, "Data augmentation of wearable sensor data for Parkinson's disease monitoring using convolutional neural networks," in *Proc. 19th ACM Int. Conf. Multimodal Interact.*, Nov. 2017, pp. 216–220.

[42] B. S. Atal and M. R. Schroeder, "Adaptive predictive coding of speech signals," *Bell Syst. Tech. J.*, vol. 49, no. 8, pp. 1973–1986, Oct. 1970.

[43] A. Stisen *et al.*, "Smart devices are different: Assessing and mitigating mobile sensing heterogeneities for activity recognition," in *Proc. 13th ACM Conf. Embedded Networked Sensor Syst.*, Nov. 2015, pp. 127–140.

[44] X. Qian *et al.*, "The smart insole: A pilot study of fall detection," in *Proc. EAI Int. Conf. Body Area Netw.* Cham, Switzerland: Springer, Oct. 2019, pp. 37–49.

[45] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. M. Havinga, "Fusion of smartphone motion sensors for physical activity recognition," *Sensors*, vol. 14, no. 6, pp. 10146–10176, 2014.

[46] D. Micucci, M. Mobilio, and P. Napoletano, "UniMiB SHAR: A dataset for human activity recognition using acceleration data from smartphones," *Appl. Sci.*, vol. 7, no. 10, p. 1101, 2017.

[47] G. Tsoumakas and I. Vlahavas, "Random k-labelsets: An ensemble method for multilabel classification," in *Proc. Eur. Conf. Mach. Learn.* Berlin, Germany: Springer, Sep. 2007, pp. 406–417.

[48] A. Murad and J.-Y. Pyun, "Deep recurrent neural networks for human activity recognition," *Sensors*, vol. 17, no. 11, p. 2556, 2017.