Worker Activity Recognition in Smart Manufacturing Using IMU and sEMG Signals with Convolutional Neural Networks

Wenjin Tao, Ze-Hao Lai, Ming C. Leu and Zhaozheng Yin

August 1, 2018

Worker Activity Recognition in Smart Manufacturing Using IMU and sEMG Signals with Convolutional Neural Networks

Wenjin Tao[a,*], Ze-Hao Lai[a], Ming C. Leu[a], Zhaozheng Yin[b]

[a] *Department of Mechanical and Aerospace Engineering, Missouri University of Science and Technology, Rolla, MO 65409, USA*
[b] *Department of Computer Science, Missouri University of Science and Technology, Rolla, MO 65409, USA*

Abstract

In a smart manufacturing system involving workers, recognition of the worker's activity can be used for quantification and evaluation of the worker's performance, as well as to provide onsite instructions with augmented reality. In this paper, we propose a method for activity recognition using Inertial Measurement Unit (IMU) and surface electromyography (sEMG) signals obtained from a Myo armband. The raw 10-channel IMU signals are stacked to form a signal image. This image is transformed into an activity image by applying Discrete Fourier Transformation (DFT) and then fed into a Convolutional Neural Network (CNN) for feature extraction, resulting in a high-level feature vector. Another feature vector representing the level of muscle activation is evaluated with the raw 8-channel sEMG signals. Then these two vectors are concatenated and used for work activity classification. A worker activity dataset is established, which at present contains 6 common activities in assembly tasks, i.e., grab tool/part, hammer nail, use power-screwdriver, rest arm, turn screwdriver, and use wrench. The developed CNN model is evaluated on this dataset

[*] Corresponding author
E-mail address: w.tao@mst.edu (Wenjin Tao).

and achieves 98% and 87% recognition accuracy in the half-half and leave-one-out experiments, respectively.

Keywords: Smart Manufacturing; Activity Recognition; Deep Learning; Convolutional Neural Networks (CNN); IMU; sEMG.

1. Introduction

The availability of low-cost sensors and the development of Internet-of-Things (IoT) technologies enable access to big data for the manufacturing industry [1], which builds up the data foundation for smart manufacturing. A variety of methods and algorithms have been developed to learn valuable information from the data, and to make the manufacturing smarter [2]. The fast-growing artificial intelligence technologies, particularly deep learning [3], are promising to further boost this industry. In a smart manufacturing system involving workers, recognition of the worker's activity can be used for quantification and evaluation of the worker's performance, as well as to provide onsite instructions with augmented reality. Wearable devices, such as an armband embedded with an Inertial Measurement Unit (IMU) or surface electromyography (sEMG) sensors, directly sense the movement of human body or the level of muscle activation, which can provide information of the body status. In addition, there are a lot of inexpensive wearable devices in the market, such as Myo armbands [4] and smartphones, which are widely used in activity recognition tasks.

For activity recognition in the manufacturing area, Stiefmeire et al. [5] utilized ultrasonic and IMU sensors for worker activity recognition in a bicycle maintenance scenario using a Hidden Markov Model classifier. Later they proposed a string-matching based segmentation and classification method using multiple IMU sensors for recognizing worker activity in car manufacturing tasks [6, 7]. Koskimaki et al. [8] used a wrist-worn IMU sensor to capture the arm movement and a K-Nearest Neighbors model to classify five activities for industrial assembly lines. Maekawa et al. [9] proposed an unsupervised measurement method for lead time estimation of factory work using signals from a smartwatch with an IMU sensor.

In general, the activity recognition task can be broken down into two subtasks: feature extraction and subsequent multiclass classification. To extract more discriminative features, various methods have been applied to the raw signals in the time or frequency domain, e.g., mean, correlation, and Principal Component Analysis [10, 11, 12, 13]. Different classifiers have been explored on the features for activity recognition, such as the Support Vector Machine [10, 12], Random Forest, K-Nearest Neighbors, Linear Discriminant Analysis [11], and Hidden Markov Model [13]. To effectively learn the most discriminative features, Jiang et al. [14] proposed a method based on Convolutional Neural Networks (CNN). They assembled the raw IMU signals into an activity image, which enabled the CNN model to automatically learn the discriminative features from the activity image for classification.

In the present research, we choose a Myo armband to capture the worker's activity because it can provide both IMU and sEMG signals. Motivated by the study of Jiang et al. [14], we stack the raw IMU signals to form a signal image. This image is transformed into an activity image by applying Discrete Fourier Transformation (DFT) and then fed into a CNN for feature extraction, resulting in a high-level feature vector. Another feature vector representing the level of muscle activation is calculated from the raw sEMG signals. Then these two vectors are concatenated and used for worker activity classification. An overview of our method is illustrated in Figure 1. To evaluate the method, a worker activity dataset containing 6 common activities in assembly tasks is established.

The remainder of this paper is organized as follows. Section 2 discusses how we build up the worker activity dataset. Our proposed method is detailed in Sections 3 and 4. The experimental setups and results are described in Sections 5 and 6, respectively. Finally, Section 7 provides the conclusions.

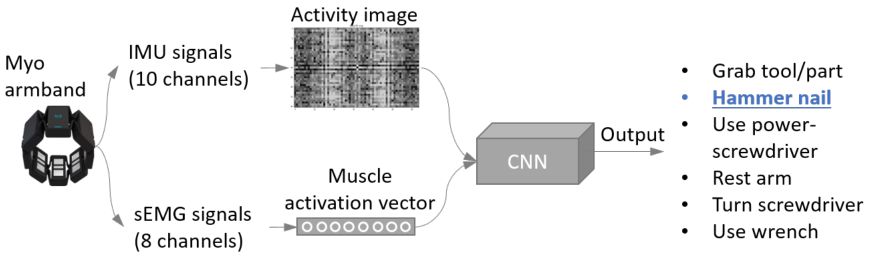

Figure 1: Overview of our worker activity recognition method.

2. Data Collection

To establish our dataset of worker activity, six activities commonly performed in assembly tasks are chosen, which are: grab tool/part (GT), hammer nail (HN), use power-screwdriver (UP), rest arm (RA), turn screwdriver (TS), and use wrench (UW).

A Myo armband equipped with IMU and sEMG sensors from Thalmic Labs is used for data acquisition. The IMU returns three types of signals (3-channel acceleration, 3-channel angular velocity, and 4-channel orientation) at the sample rate of 50Hz. A set of 8 sEMG pods attached to the skin return 8 channels of unitless signals in the range of [-128, 127] at the sample rate of 200Hz, which represent the corresponding muscle activations. These 18-channel signals are transmitted via Bluetooth to the computer.

There are 8 subjects recruited to conduct a set of tasks (listed in Table 1) containing the 6 activities. As demonstrated in Figure 2(a), the subject is asked to stand in front of the workbench, wear a Myo armband on his/her right forearm with a fixed orientation (Figure 2(b)), and perform the tasks on assembly dummies in a natural way. The Myo data are collected during the tasks and an overhung camera is used to record the assembly activities simultaneously for monitoring the process. Examples of the 6 activities are shown in Figure 3, which are taken from the overhung camera.

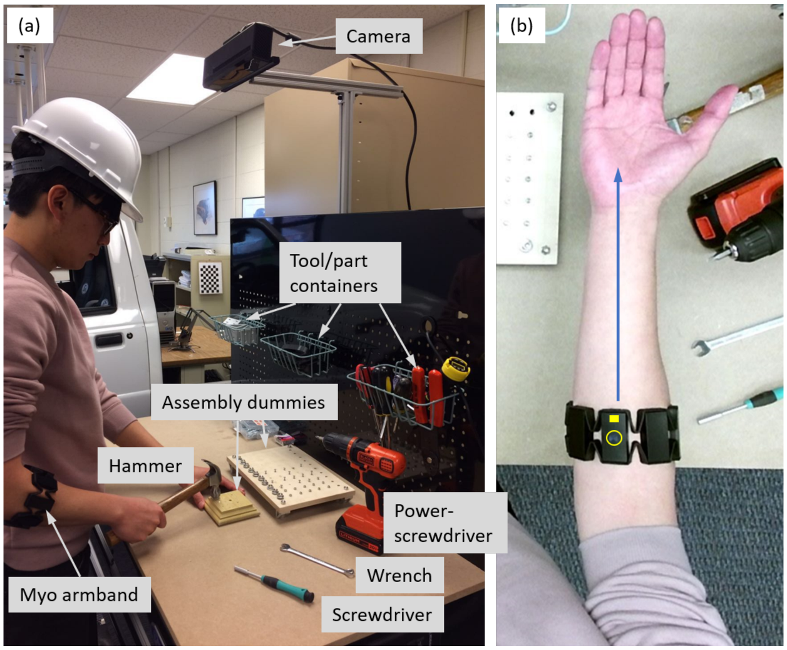

Figure 2: (a) Data collection setup; (b) Wearing orientation of a right-hand.

3. Signal Preprocessing

Although the Myo data are collected separately for different tasks and each task consists of only one activity, there still might be some noise inside the data, such as preparing activities between hammering nails. To address it, the recorded videos are reviewed to locate the time durations, each of which contains only one of the six activities. These durations are used to segment the raw Myo data.

Usually, the duration of a segmented instance ranges from a few seconds to more than one minute, which consists of repeated activity patterns. Thus, sampling is needed to prepare the data samples for recognition. As depicted in Figure 4, the 50Hz IMU signals are sampled using a sliding window with the width of 64 timestamps and 75% overlap between two steps. Thus each IMU sample lasts for about 1.3 seconds, which covers at least one activity pattern.

Table 1: Tasks for collecting worker activity.

No	Tasks	Activities
1	Grab 30 tools/parts from the 3 containers	GT
2	Hammer 15 nails into the wooden dummy	HN
3	Tighten 20 screws using a power-screwdriver	UP
4	Rest arms for about 60 seconds	RA
5	Tighten 10 nuts using a screwdriver	TS
6	Tighten 10 nuts using a wrench	UW

After sampling the IMU signals, the 200Hz sEMG signals are sampled according to the time durations of the IMU samples. Then each sEMG sample has an approximate width of 256 timestamps.

After sampling, suppose we have N IMU samples and N sEMG samples, by using the method proposed by Jiang et al. [14], the 10-channel signals in an IMU sample are stacked and shuffled, forming a signal image with the size of 42×64. Then this signal image is transformed into an activity image by applying two-dimensional (2D) Discrete Fourier Transform (DFT) and taking its logarithmic magnitude. Figure 5 shows activity image examples for each activity. For a sEMG sample, the 8-channel signals are averaged along each channel, forming an 8-dimensional vector, which represents the level of muscle activation.

4. CNN Architecture

The architecture of our CNN model is illustrated in Figure 6. It accepts two inputs, the IMU activity image and the sEMG vector, and outputs a probability distribution of the 6 activities.

After the preprocessing steps described in Section 3, there are N activity images X_i^{IMU} and N sEMG vectors X_i^{sEMG}, where $i \in [1, N]$. X_i^{IMU} has the size of $42 \times 64 \times 1$ (height, width, depth, respectively) and is normalized to the interval $[0, 1]$ before being fed into three 5×5 convolutional layers for feature extraction. Each convolutional layer is down-sampled to a half by implementing a 2×2 max pooling layer.

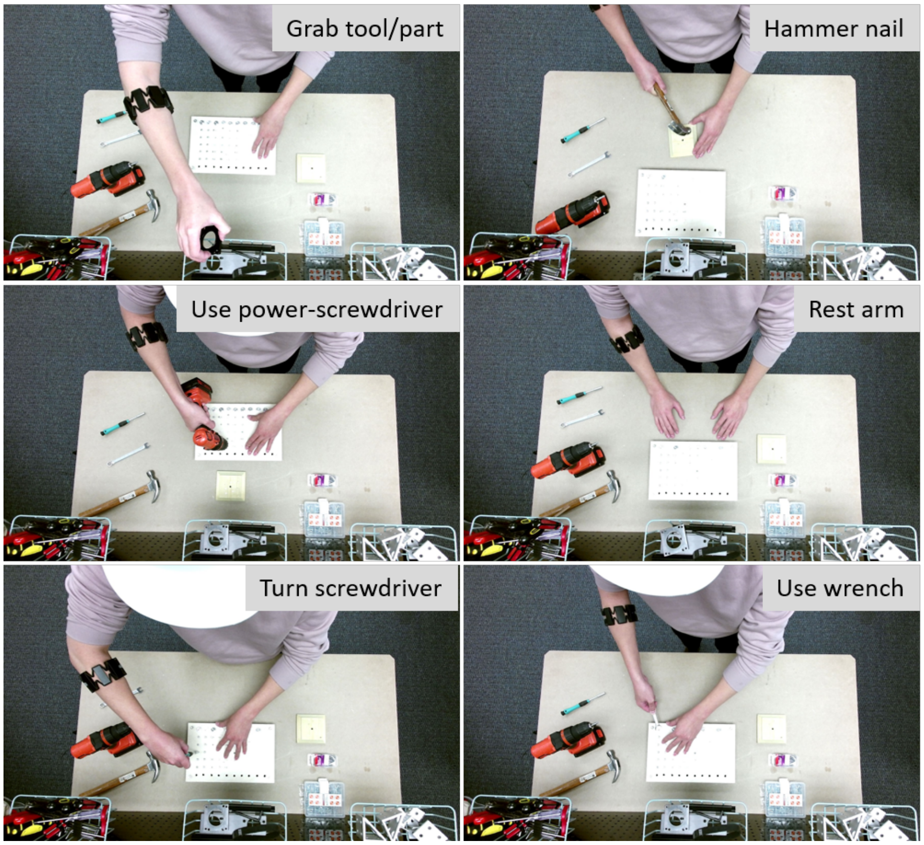

Figure 3: Examples of the 6 activities.

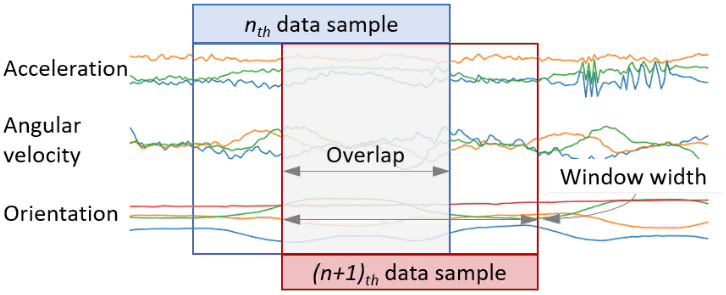

Figure 4: Sampling method.

Then the feature map from the third pooling layer having the size of 2 × 5 × 32

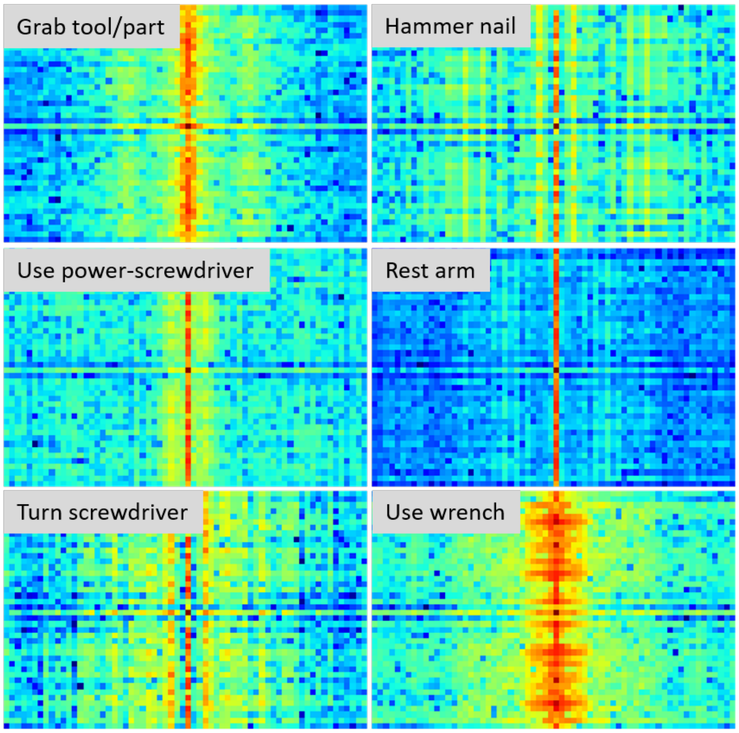

Figure 5: Examples of activity image.

is flattened into a 320-dimension feature vector, which is subsequently densified by a fully connected layer to a 16 dimensional feature vector.

On the other side, X_i^{sEMG} representing muscle activation levels at 8 positions is injected directly as a high-level feature. It is concatenated to the previous 16-dimension feature vector from IMU signals, resulting in the total dimension of 24. Then another fully connected layer is used to densify the feature vector to the dimension of 6, which is the number of activities. Then this 6 dimensional score vector $\{S_j, j = 1, 2, ..., 6\}$ is transformed to output the predicted probabilities using a softmax function [15] as follows:

$$P(y = j|[X_i^{IMU}, X_i^{sEMG}]) = \frac{\exp(S_j)}{\sum_{k=1}^{6} \exp(S_k)} \tag{1}$$

where $P(y = j|[X_i^{IMU}, X_i^{sEMG}])$ is the predicted probability of being class j based on the inputs X_i^{IMU} and X_i^{sEMG}.

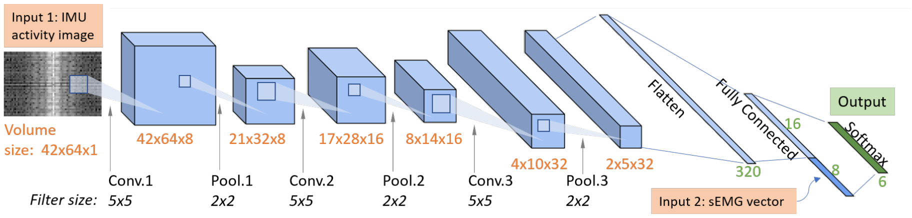

Figure 6: The architecture of our CNN model. The volume size is represented in *height* × *width* × *depth*. ‘Conv.’ and ‘Pool.’ denote the operations of convolution and pooling, respectively.

Training a CNN model involves optimizing the network’s weights w to minimize a chosen cost function. We select the cross entropy [15] as the cost function:

$$\mathcal{L}(w) = \sum_{i=1}^{N} \sum_{j=1}^{6} y_{ij} \log[P(y = j|[X_i^{IMU}, X_i^{sEMG}])] + \lambda l_2(w) \qquad (2)$$

where y_{ij} is 0 if the ground truth label of the ith data $[X_i^{IMU}, X_i^{sEMG}]$ is the jth label, and is 1 otherwise. The L2 regularization term [16] is added to the cost function to penalize large weights, and λ is its coefficient. The Adam optimization method [17] is used in the training.

The dropout regularization [18] randomly drops units from the neural network during training, which is commonly used to avoid the overfitting. It is implemented after the flatten layer in the CNN model.

5. Experiment

We evaluate our method on an established worker activity dataset, which has 6 activities performed by 8 subjects. The quantitative information of the dataset is listed in Table 2. There are 11,211 data samples in total. These subjects use different amounts of time to finish each task, therefore they have different numbers of data samples for each activity.

Two evaluation policies are conducted, i.e., half-half and leave-one-out policies. In the half-half evaluation, after randomly shuffling, one half of the dataset

Table 2: Number of data samples for each activity of different subjects.

Subject No.	GT	HN	UP	RA	TS	UW
1	193	140	364	266	222	442
2	302	408	195	56	274	751
3	198	183	171	251	214	567
4	204	172	188	29	82	344
5	187	204	142	43	213	372
6	216	77	179	47	129	301
7	213	196	203	254	231	576
8	200	184	262	145	148	273

is prepared for training and the other half is kept for testing. In the leave-one-out evaluation, the samples from 7 out of 8 subjects are used for training, and the samples of the left one subject are reserved for testing. We employ several commonly used metrics [16] to evaluate the classification performance, which are listed as follows:

- Accuracy

$$Accuracy = \frac{\sum_i^N 1(\hat{y}_i = y_i)}{N} \tag{3}$$

- Precision and Recall

$$\begin{aligned} Precision &= \frac{TP}{TP + FP} \\ Recall &= \frac{TP}{TP + FN} \end{aligned} \tag{4}$$

- F_1 score

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{5}$$

where 1 is an indicator function in Equation 3. For a certain class y_i, True Positive (TP) is defined as a sample of class y_i that is correctly classified as y_i; True Negative (TN) means a sample from a class other than y_i is correctly classified as 'not y_i'; False Positive (FP) means a sample from a class other than y_i is misclassified as y_i; False Negative (FN) means a sample from the class y_i is misclassified as another 'not y_i' class. F_1 score is the harmonic mean of Precision and Recall, which ranges in the interval [0,1].

The CNN model described in Section 4 is created using the Google TensorFlow library. For training hyperparameters, we choose the batch size as 512, the learning rate as 0.001, the dropout rate as 0.5, and the regularizer coefficient as 1e-5. The number of epochs is 1000. We use a workstation with one 12 core Intel Xeon processor, 64GB of RAM and one Nvidia Geforce 1080 Ti graphic card for the CNN training.

6. Results

To explore the optimal combination of inputs for the CNN model, we first compare the performance of three cases using different inputs: 1). activity images from the IMU signals (IMU-AI); 2). activity images from the sEMG signals (sEMG-AI); and 3). vectors representing the muscle activation levels from the sEMG signals (sEMG-V). The model described in Section 4 is adapted accordingly to fit these 3 cases. For case 1, the lower stream of the CNN model for the second input shown in Figure 6 is abandoned. Case 2 uses a CNN model similar to the one in case 1. Since the size of an activity image from the sEMG signals is 25×64, this model only has two sets of convolutional layers with the maximum depth of 16, instead of 32. For case 3, only the lower stream of the CNN model shown in Figure 6 is reserved, which is a fully connected neural network from 8 nodes to 6 nodes.

The performance of these three cases in terms of accuracy, precision, recall and F_1 score with two evaluation strategies (half-half and leave-one-out) is summarized in Table 3. Case 1 has the highest performance among the three, which is about 30% higher than the other two. It demonstrates that the activity images from the IMU signals provide more discriminative features for activity recognition. Compared to case 2, case 3 has higher performance as well as lower computational cost due to the simplicity of its model.

Therefore, we choose the two inputs, i.e., IMU-AI and sEMG-V, for our CNN

Table 3: Overall performance (%) of the half-half (hh) and leave-one-out (loo) experiments.

Inputs[*]		Accuracy	Precision	Recall	F_1 Score
IMU-AI	hh	97.5	97.5	97.5	97.5
	loo	85.0	87.2	87.3	85.3
sEMG-AI	hh	60.4	64.0	60.3	61.8
	loo	49.2	52.8	49.1	48.4
sEMG-V	hh	66.4	66.8	67.4	67.0
	loo	50.7	52.5	53.1	47.9
IMU-AI,	hh	97.6	97.8	97.5	97.7
sEMG-V	loo	87.4	89.0	89.5	87.6

[*] 'AI' denotes activity images from either IMU or sEMG signals, and 'V' denotes vectors that represent the muscle activation levels from sEMG signals.

model. As shown in Table 3, its performance of the leave-one-out experiment is about 2% higher than that in case 1 with only one IMU-AI input. For the half-half experiment, 97.6% of the testing samples are correctly recognized. Also, as shown in Figure 7, only a small number of samples are misclassified and not along the diagonal. It is about 10% higher than 87.4% of the leave-one-out experiment. This is because all the testing subjects are seen in the half-half experiment, while the testing subject in the leave-one-out experiment is unseen.

The leave-one-out results on each testing subject are detailed in Table 4. The 4th subject has the highest performance, which reaches 98.2%, 97.1%, 98.8% and 97.9% in accuracy, precision, recall and F_1 score, respectively. The lowest performance is from the 7th subject, which has about 37% of the testing samples misclassified. The UW activity of the 7th subject has the largest recognition errors, which is shown in Figure 8(7). The majority of UW are misclassified as TS, GT and UP. By reviewing the recorded videos, as illustrated in Figure 9 where the arrows show the approximate trajectories of the arm movements, we find the reason for the low performance of the UW activity is that the 7th subject performed the UW task significantly differently from other subjects and thus it is difficult for the CNN model to learn using the leave-one-out strategy.

Table 4: Results (%) of the leave-one-out experiments evaluated on each test subject.

Subject No.	Accuracy	Precision	Recall	F1 Score
1	92.3	93.2	92.7	92.7
2	92.0	88.0	92.6	89.8
3	90.0	88.8	88.9	88.0
4	98.2	97.1	98.8	97.9
5	93.6	93.0	94.7	93.6
6	84.4	91.8	85.2	87.0
7	63.3	74.7	76.7	66.7
8	85.1	85.4	86.5	84.9

To address the confusing issues and further improve the model performance, some directions for future work are considered, such as recruiting more subjects to learn more working styles, using data augmentation techniques to add more variations to the collected data, and exploring different methods of signal preprocessing and feature extraction to fully exploit the sEMG signals. In addition, the recording videos can also be utilized to create an image-based activity recognition module.

7. Conclusion

In this paper, we develop a Convolutional Neural Network (CNN) model for worker activity recognition in smart manufacturing using the Inertial Measurement Unit (IMU) and surface electromyography (sEMG) signals obtained from a Myo armband. A worker activity dataset is established, which involves 8 subjects and contains 6 common activities in assembly tasks (i.e., grab tool/part, hammer nail, use power-screwdriver, rest arm, turn screwdriver and use wrench). The developed CNN model is evaluated on this dataset and achieves 98% and 87% recognition accuracy in the half-half and leave-one-out experiments, respectively.

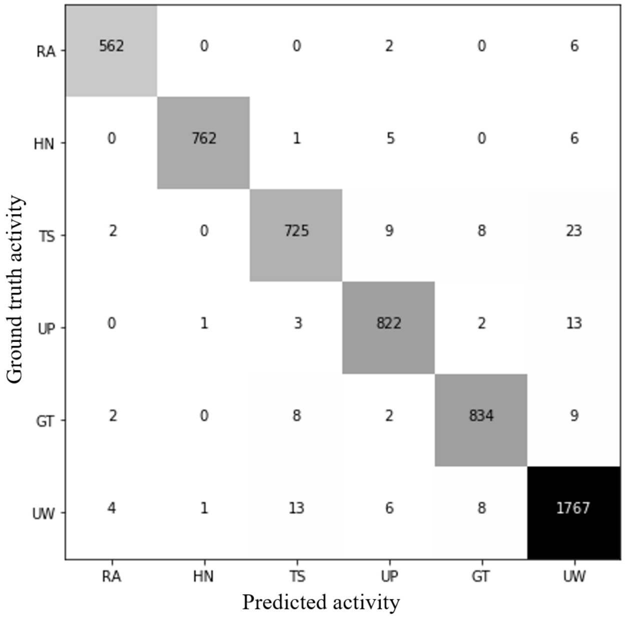

Figure 7: Confusion matrix of the half-half experiment. The values represent the number of samples, e.g., the '562' on the upper-left corner means there are 562 samples of actual 'rest arm' (RA) correctly predicted as RA, and the '6' on the upper-right corner means there are 6 samples of actual RA incorrectly predicted as 'use wrench' (UW).

Acknowledgments

This research work is supported by the National Science Foundation grant CMMI-1646162 and also by the Intelligent Systems Center at Missouri University of Science and Technology. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] S. Jeschke, C. Brecher, T. Meisen, D. Özdemir, T. Eschert, Industrial internet of things and cyber manufacturing systems, in: Industrial Internet of Things, Springer, 2017, pp. 3–19.

[2] K. Nagorny, P. Lima-Monteiro, J. Barata, A. W. Colombo, Big data analysis in smart manufacturing: A review, International Journal of Communications, Network and System Sciences 10 (03) (2017) 31.

[3] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.

[4] Thalmic Labs Inc., Myo armband, [Online; accessed 15-November-2017] (2017).
URL `https://www.myo.com/`

[5] T. Stiefmeier, G. Ogris, H. Junker, P. Lukowicz, G. Troster, Combining motion sensors and ultrasonic hands tracking for continuous activity recognition in a maintenance scenario, in: Wearable Computers, 2006 10th IEEE International Symposium on, IEEE, 2006, pp. 97–104.

[6] T. Stiefmeier, D. Roggen, G. Troster, Fusion of string-matched templates for continuous activity recognition, in: Wearable Computers, 2007 11th IEEE International Symposium on, IEEE, 2007, pp. 41–44.

[7] T. Stiefmeier, D. Roggen, G. Ogris, P. Lukowicz, G. Tröster, Wearable activity tracking in car manufacturing, IEEE Pervasive Computing 7 (2).

[8] H. Koskimaki, V. Huikari, P. Siirtola, P. Laurinen, J. Roning, Activity recognition using a wrist-worn inertial measurement unit: A case study for industrial assembly lines, in: Control and Automation, 2009. MED'09. 17th Mediterranean Conference on, IEEE, 2009, pp. 401–405.

[9] T. Maekawa, D. Nakai, K. Ohara, Y. Namioka, Toward practical factory activity recognition: unsupervised understanding of repetitive assembly work in a factory, in: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ACM, 2016, pp. 1088–1099.

[10] D. Anguita, A. Ghio, L. Oneto, X. Parra, J. L. Reyes-Ortiz, A public domain dataset for human activity recognition using smartphones., in: ESANN, 2013.

[11] T. Peterek, M. Penhaker, P. Gajdoš, P. Dohnálek, Comparison of classification algorithms for physical activity recognition, in: Innovations in Bio-inspired Computing and Applications, Springer, 2014, pp. 123–131.

[12] W. Chang, L. Dai, S. Sheng, J. T. C. Tan, C. Zhu, F. Duan, A hierarchical hand motions recognition method based on imu and semg sensors, in: Robotics and Biomimetics (ROBIO), 2015 IEEE International Conference on, IEEE, 2015, pp. 1024–1029.

[13] C. A. Ronao, S.-B. Cho, Human activity recognition using smartphone sensors with two-stage continuous hidden markov models, in: Natural Computation (ICNC), 2014 10th International Conference on, IEEE, 2014, pp. 681–686.

[14] W. Jiang, Z. Yin, Human activity recognition using wearable sensors by deep convolutional neural networks, in: Proceedings of the 23rd ACM international conference on Multimedia, ACM, 2015, pp. 1307–1310.

[15] C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[16] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016, `http://www.deeplearningbook.org`.

[17] D. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980.

[18] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting., Journal of machine learning research 15 (1) (2014) 1929–1958.

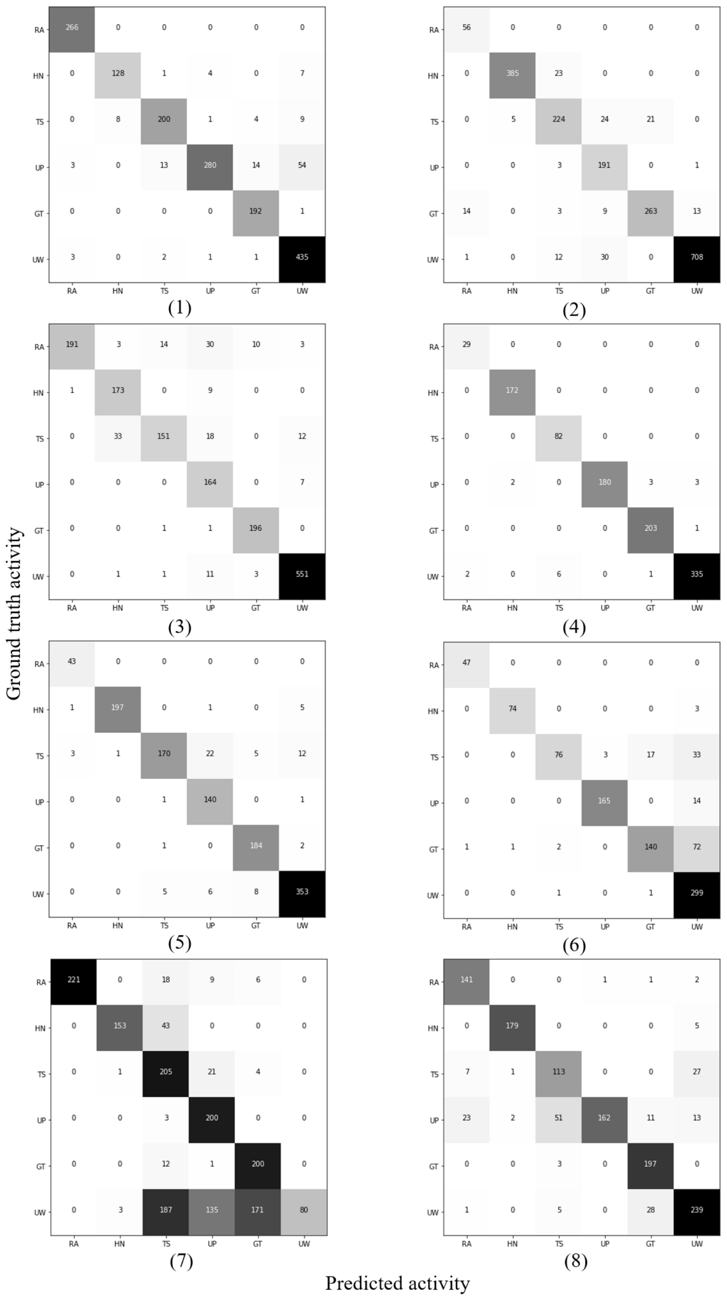

Figure 8: Confusion matrix of the leave-one-out experiment on each of the eight testing subjects.

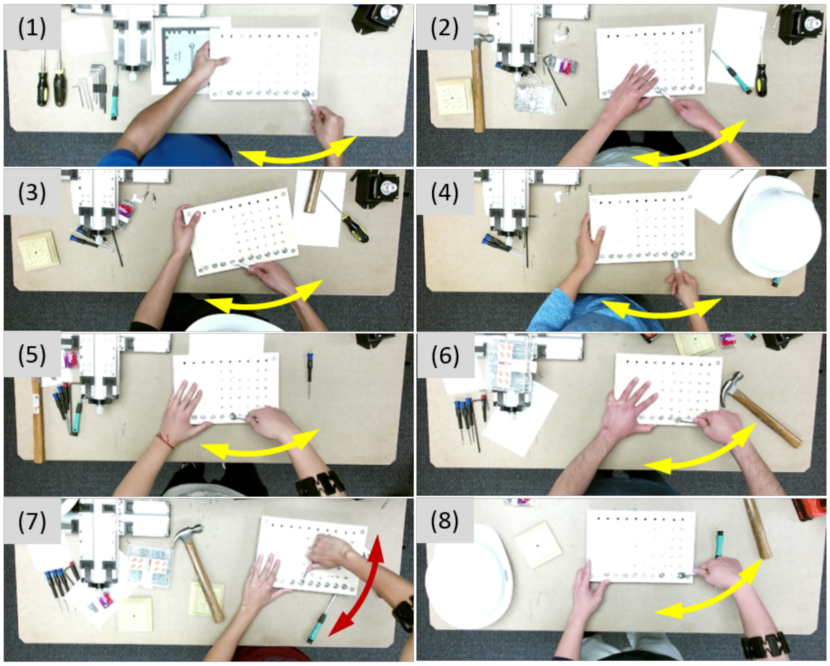

Figure 9: Use of wrench (UW) activities of each of the eight subjects. The arrows show the approximate trajectories of the arm movements.