S-CONVNET: A SHALLOW CONVOLUTIONAL NEURAL NETWORK ARCHITECTURE FOR NEUROMUSCULAR ACTIVITY RECOGNITION USING INSTANTANEOUS HIGH-DENSITY SURFACE EMG IMAGES

Md. Rabiul Islam, Daniel Massicotte, Francois Nougarou, Philippe Massicotte, and Wei-Ping Zhu[*]

Dept. of Electrical and Computer Engineering, Université du Québec à Trois-Rivières, QC, Canada
[*] Dept. of Electrical and Computer Engineering, Concordia University, Montreal, QC, Canada

ABSTRACT– The recent progress in recognizing low-resolution instantaneous high-density surface electromyography (HD-sEMG) images opens up new avenues for the development of more fluid and natural muscle-computer interfaces. However, the existing approaches employed a very large deep convolutional neural network (ConvNet) architecture and complex training schemes for HD-sEMG image recognition, which requires learning of >5.63 million(M) training parameters only during *fine-tuning* and *pre-trained* on a very large-scale labeled HD-sEMG training dataset, as a result, it makes high-end resource-bounded and computationally expensive. To overcome this problem, we propose S-ConvNet models, a simple yet efficient framework for learning instantaneous HD-sEMG images from scratch using *random-initialization*. Without using any pre-trained models, our proposed S-ConvNet demonstrate very competitive recognition accuracy to the more complex state of the art, while reducing learning parameters to only $\approx$ 2M and using $\approx 12\times$ smaller dataset. The experimental results proved that the proposed S-ConvNet is highly effective for learning discriminative features for instantaneous HD-sEMG image recognition, especially in the data and high-end resource-constrained scenarios.

Keywords: Neuromuscular activity recognition, Shallow convolutional neural networks, Feature learning, HD-sEMG, Gesture recognition, Muscle-computer interface, Deep neural networks.

I. INTRODUCTION

Neuromuscular activity recognition has a growing motivation for research because of its respective novel applications in real life. The major application domains are non-invasive control of active prosthesis [1], wheelchairs [2], exoskeletons [3] or providing interaction methods for video games [4] and neuromuscular diagnosis [5]. The conventional approaches for neuromuscular activity recognition immensely rely on sparse multi-channel surface electromyography (sEMG) sensors and windowed descriptive and discriminatory sEMG features [6-10]. However, the sparse multi-channel sEMG based methods are not suitable for real-world applications due to their lack of robustness to electrode shift and positioning and therefore malfunctioning in any one of the channels requires retraining the entire system [11], [12]. In recent years, the high-density sEMG (HD-sEMG) based methods have been proposed to address this problem [11-13], [29]. The HD-sEMG consists of two-dimensional (2D) arrays of closely spaced electrodes that used to record the myoelectric activity over the skin surface [13], [14]. The recorded HD-sEMG data are spatially correlated enabled both spatial and temporal changes and robust to electrode shift and positioning [12]. The windowed sEMG and descriptive and discriminative features are used by the existing HD-sEMG based methods for neuromuscular activity recognition. However, finding an optimal window size would still require that reflects the compromise between classification accuracy and controller delay (both increase with the window increase) especially in the application of assistive technology, physical rehabilitation, and human-computer interfaces [13].

To address this problem, the distinctive patterns inside the instantaneous sEMG images has been explored for developing more fluid and natural muscle-computer interfaces (MCI's) in recent years by Geng *et al.*, [13] and M. R. Islam *et al.,* [15], [33]. This scheme enables neuromuscular activity recognition solely with the sEMG images spatially composed from HD-sEMG signals recorded at a specific instant. The instantaneous values of HD-sEMG signals at each sampling instant were arranged in a 2D grid following the electrode positioning. Afterwards, this 2D grid was converted to a grayscale sEMG image. Using Histogram of Oriented Gradients (HOG) as discriminative features and pairwise SVM's classification method in [15], a competitive neuromuscular activity recognition accuracy of an 8-hand gesture has been achieved as par with the state-of-the-art method for an intra-subject test.

However, a DeepFace [17] like very large deep convolutional neural network (CNN or ConvNet) architecture is employed by the state-of-the-art methods [13], [16] for sEMG image classification, which requires to be pre-trained on a very large-scale training dataset ($\approx$ 0.76 million), as a result, it makes computationally expensive to be practical for real-world MCIs applications. Following are the other critical limitations of using *pre-trained* networks for instantaneous HD-sEMG image recognition:

(i) Constrained structure design space – pre-trained networks are very deep and large and trained on a large-scale HD-sEMG dataset, therefore, containing a massive number of parameters. Hence, there is a little flexibility to control/adjust the network structures (even for small changes) by directly adopting the *pre-trained* network to the *target task*. The requirement of computing resources and large-scale pre-trained datasets are also bounded by large network structures.

(ii) Domain mismatch – the existing sEMG based neuromuscular activity recognition methods are usually trained and evaluated on the data acquired from the able-bodied subjects. However, in real time

sEMG-based MCIs applications (e.g., assistive technology, physical rehabilitation, etc.) are most of the time designed for elderly people, amputees and patients. These differences impose a serious problem due to the varied sEMG distributions in the *source* and *target task*. Though the fine-tuning of the pre-trained model can reduce the gap, however, it is still a serious problem, when there is a huge mismatch between the *source* and the *target task* [18]. Also, this conventional wisdom of pre-training is recently challenged by He *et al.* [31], where *pre-training* does not necessarily improve the *target task* accuracy is proved to be claimed.

(iii) Learning bias – the distributions and the loss functions between the *source task* and the *target task* may vary significantly, which may lead to different searching/optimization spaces. Therefore, the learning may be biased towards a local minimum which is not optimal for the *target task* [19].

To overcome these above-mentioned problems, our work is motivated by the following research question- *is it possible to learn neuromuscular activities from scratch utilizing HD-sEMG datasets available only for the target task without any pre-training?* To achieve this goal, we propose shallow and lightweight convolutional neural network (S-ConvNet) architectures, a simple yet effective framework, which could learn neuromuscular activity from scratch using $\approx 12 \times \mathit{smaller\ dataset}$ than its pre-trained counterparts for HD-sEMG image recognition.

For instantaneous sEMG image-based neuromuscular activity recognition, the challenge remains open because very limited research has been done on it.

The rest of the paper is organized as follows. Sections II and III present the proposed framework and S-ConvNet models respectively. Section IV describes the testing database and experimental validation. Section V offers some conclusive remarks.

II. The Framework

The proposed S-ConvNet framework for neuromuscular activity recognition using instantaneous HD-sEMG images has three phases: (i) pre-processing and HD-sEMG image generation (ii) architectural design of the S-ConvNet model and (iii) classification. Fig.1 describes the proposed S-ConvNet framework of muscular activity recognition by instantaneous sEMG images. The All-ConvNet [33] is a fully convolutional neural network, where the depth of the All-ConvNet network is much higher than the propoed S-ConvNet. First, the power-line interferences were removed from the acquired HD-sEMG signals with a band-stop filtered between 45 and 55 Hz using a 2nd order Butterworth filter. Then, the HD-sEMG signals at each sampling instant were arranged in a 2-D grid according to their

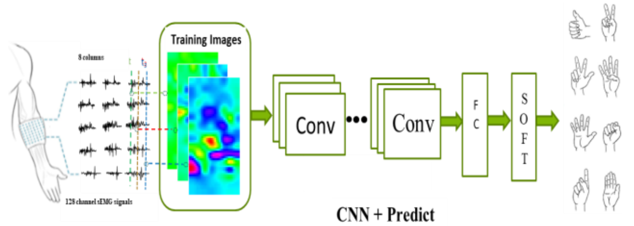

Fig. 1. Schematic diagram of the proposed framework of muscular activity recognition by instantaneous sEMG images.

electrode positioning. This grid was further transformed into an instantaneous sEMG image by linearly transforming the values of sEMG signals from mV to color intensity as $[-2.5mV,\ 2.5mV]$ to $[0\ 255]$. Thus, an instantaneous grayscale sEMG image was formed with a size of 16×8. Secondly, we devised different S-ConvNet models which describe in Section III. Finally, providing instantaneous HD-sEMG images and their corresponding labels, our devised S-ConvNet model is trained offline to predict to which muscular activity an instantaneous HD-sEMG image belongs. Then, this trained S-ConvNet model is used to recognize different neuromuscular activities at test time from the unseen instantaneous HD-sEMG images.

III. Model Description- The Shallow Convolutional Neural Network (S-ConvNet)

We train our S-ConvNet on a multi-class neuromuscular activity recognition task, namely, to recognize an activity class through an instantaneous HD-sEMG image. The overall architecture of S-ConvNet models are described in Table I. Starting from the simplest Model A, the depth and number of parameters in the network gradually increase to Model C. The instantaneous HD-sEMG image is passed through a convolutional (conv.) layers, where a small receptive field with a 3×3 filters are used. The smallest receptive field with 3×3 filters is the minimum filter size to allow overlapping convolutions and spatial pooling with a stride of 2, which also captures the notion of left, right and center amicably. It can be observed that the Model B from the Table I is a variant of the Network in Network architecture [24], where only 1×1 convolution is performed after each normal 3×3 convolutions layers. The 1×1 convolution act as a linear transformation of the input channels followed by a non-linearity [25]. We also highlight that the model C is a variant of the simple ConvNet models introduced by J. T. Springenberg *et. al.,* [20] for object recognition in which the spatial pooling is performed by using a stridden CNN. The output of a convolution map f produced by a convolution layer c is computed as follows:

$$c_{i,j,o}(f) = \emptyset\left(\sum_{h=1}^{k}\sum_{w=1}^{k}\sum_{u=1}^{n}\theta_{h,w,u,o}\cdot f_{g(h,w,i,j,u)}\right) \qquad (1)$$

where θ are the convolutional weights or filters; $g(h,w,i,j,u) = (r.i+h,\ r.j+w,\ u)$ is the function mapping from a position in c to a position in f respecting the stride r; w and h are respectively the width and height of the filters; n is the number of channels (in case f is the output of a convolutional layer, n is the number of filters); $o \in [1, M]$ is the number of output feature or channels of the convolutional layer and $\emptyset(\cdot)$ is the activation function, an exponential linear unit ELU defined as:

$$\emptyset(x) = \begin{cases}\alpha(\exp(x)-1), & if\ x < 0\\ x, & if\ x \geq 0\end{cases} \qquad (2)$$

Afterwards, the convolution maps produced by the final convolutional layer of each of the model networks, illustrated in Table I, are flattened out to form a multi-dimensional feature vector. Then, the flattened feature vector is inputted to a fully connected layer where each of the feature elements is connected to all its input neurons. This fully connected layer can capture correlations between features extracted in the distant part of the instantaneous sEMG images.

TABLE I THE THREE S-CONVNET NETWORKS MODELS FOR NEUROMUSCULAR ACTIVITY RECOGNITION

A	B	C
Input 16×8 Gray-level Image		
3 × 3 Conv. 32 ELU	3 × 3 Conv. 32 ELU	3 × 3 Conv. 32 ELU
3 × 3 Conv. 64 ELU	1 × 1 Conv. 32 ELU	3 × 3 Conv. 32 ELU
3 × 3 Conv. 64 ELU	3 × 3 Conv. 64 ELU	3 × 3 Conv. 32 ELU, with stride $r = 2$
FC1 256 ELU	1× 1 Conv. 64 ELU	3 × 3 Conv. 64 ELU
FC2 G-way softmax	FC1 256 ELU	3 × 3 Conv. 64 ELU
-	FC2 G-way softmax	3 × 3 Conv. 64 ELU, with stride $r = 2$
-	-	FC1 256 ELU
-	-	FC2 G-way softmax

Finally, the output of the fully connected layer is fed to a G-way softmax layer (where G is the number of neuromuscular activity classes) which produces a distribution over the class labels. If we denote $\hat{y}^{(j)}$ as the jth element of the G dimensional output vector of the layer preceding the softmax layer, the class probabilities are estimated using the softmax function $\sigma(.)$ defined as below:

$$\sigma\left(\hat{y}^{(j)}\right) = \frac{\exp\left(\hat{y}^{(j)}\right)}{\sum_G \exp\left(\hat{y}^{(G)}\right)} \tag{3}$$

The goal of this training is to maximize the probability of the correct neuromuscular activity class. We achieve this by minimizing the cross-entropy loss [26] for each training sample. If y is the true label for a given input, the loss is

$$L = -\sum_j y^{(j)} \ln\left(\sigma(\hat{y}^{(j)}\right) \tag{4}$$

The loss is minimized over the parameters by computing the gradient of L with respect to the parameters and by updating the parameters using the state-of-the-art Adam (adaptive moment estimation) gradient descent-based optimization algorithm [27].

Having trained the network, an instantaneous HD-sEMG image is recognized as in the neuromuscular activity class C by simply propagating the input image forward and computing: $C = argmax_j(\hat{y}^{(j)})$.

The major advantage of the proposed S-ConvNet models are easily scalable and does not increase the learning parameters with the enhancement of input HD-sEMG image size. Whereas, the ConvNet employed by the state of the art [13] is unscalable. For example, the learning parameters of [13] increase to ≈ 5.63M to ≈ 11M with a little augmentation of input HD-sEMG image size from 16×8 to 16×16 due to the use of an unshared weight strategy [32].

IV. THE PERFORMANCE EVALUATION OF THE PROPOSED S-CONVNET MODELS

In order to quantify the effect of simplifying the proposed S-ConvNet model architecture, we perform experiments on CapgMyo data sets [16] (These data sets are made publicly available from the following website: http://zju-capg.org/myo/data/index.html). This dataset was developed for providing a standard benchmark database (DB) to explore new possibilities for studying next-generation muscle-computer interfaces (MCIs). The CapgMyo database comprises 3 sub-databases (referred as DB-a, DB-b and DB-c). However, DB-a has been used in our experiments to evaluate the performance of our proposed methods for *intra-session* neuromuscular activity recognition because the maximum number of subjects (18) have participated in DB-a. In DB-a, 8 isotonic and isometric hand gestures were obtained from 18 of the 23 subjects and each gesture was also recorded 10 times. For each subject, the recorded HD-sEMG data is filtered, sampled and instantaneous sEMG image is generated using the method described in Section II. More explicitly, 8 different hand gestures are performed by every subject and each hand gestures are recorded 10 times with a 1000 Hz sampling rate, which in total generates (8×10×1 000) = 80 000 or 80k instantaneous sEMG images individually. Then, our S-ConvNet models are learned from *scratch* through *random initialization*. We performed training, validation and testing using only 80 000 images produced by 18 subjects individually through a leave one trial out cross-validation. We kept one trial out from each of the 8 different hand gestures i.e 8 000 images for validation and testing. The remaining 9 trials for 8 different hand gestures i.e 72k images have been used for training. The cross-validation accuracy A is computed for each class i, as the number of totals correctly recognized hand gestures, divided by the total number of tests sEMG images

$$\text{Accuracy, } A = \frac{C}{N} = \frac{\sum C_i}{N} \tag{5}$$

where $i = \{1, 2, \ldots, G\}$ and G is the number of gesture classes.

TABLE II THE AVERAGE RECOGNITION ACCURACY (%) OF 8 HAND GESTURES WITH INSTANTANEOUS HD-SEMG IMAGES FOR 18 DIFFERENT SUBJECTS AND RECOGNITION APPROACHES. MAJORITY VOTING (ON 40 SEMG IMAGE) RESULTS ARE SHOWN IN PARENTHESES

Model	Average Recognition Accuracy (%)	# Learning Parameters
S-ConvNet-A	**87.95 (98.87)**	**≈ 2.09M**
S-ConvNet-B	86.94	≈ 2.12M
S-ConvNet-C	87.02	≈ 2.10M
W.Geng *et.al.,* [13]	89.3 (99.00)	≈ 5.63M + Pre-training

In contrast, existing approaches (e.g., [13] and [16]) for instantaneous HD-sEMG image recognition used a total of (18×40 000) = 720 000 or 720k training images for pre-training, while 40 000 images from each of the subject are used separately for *fine-tuning*. Therefore, the existing approaches involve a total of (720 000+40 000) = 760 000 or 760k images only in the training process.

In our experiments, we compared all the proposed S-ConvNet models described in Section III on the CapgMyo DB-a datasets without any *pre-training* or data augmentations. The connection weights for all S-ConvNet networks were *randomly initialized* using Xavier and He initialization schemes [21], [28]. However, we found that the models with He initialization scheme perform on average 1-1.5% worse than the Xavier initialization. We also propose to use a computationally efficient stochastic optimization algorithm, Adam [27], which provides fast and reliable learning convergence than the stochastic gradient descent (SGD) optimization algorithm used in the literature for instantaneous HD-sEMG image recognition. Our proposed all S-ConvNet models were trained using Adam optimization algorithms with a momentum decay and scaling decay are initialized to 0.9 and 0.999 respectively. In contrast to SGD, Adam is an adaptive learning rate algorithm, therefore, it requires less tuning of the learning rate hyperparameter. The learning rate of 0.001 is initialized to all our experiments. The smaller batches of 256 randomly chosen samples from the training dataset are fed to the network during consecutive learning iterations for all our experiments. We set a maximum of 100 epochs for training our S-ConvNet models. However, to avoid overfitting we have also applied early stopping in which the training process is interrupted if no improvements in validation loss are noticed for 5 consecutive epochs. The Batch normalization [22] is applied after the input and before each non-linearity. The Dropout [23] was applied on all layers with probabilities 35% for all S-ConvNet models. The S-ConvNet models were trained on a workstation with an Intel Core, 3.60 (i7-4790) processor, 16GB RAM and an NVIDIA RTX Ti GPU. Each epoch was completed in approximately 4s while training with S-ConvNet-A. The test results for all the S-ConvNet models are presented in Table II and compared with state-of-the-art methods.

As can be seen in the Table II, the simple S-ConvNet models (on the order of ≈ 2M learning parameters) trained from *random-initialization* with 3×3 convolutions and a dense layer with only a smaller number of neuron performs comparably to the state of the art for CapgMyo DB-a dataset even though the state of the art methods use more complicated network architectures and training schemes which requires to learn over ≈ 5.63M parameters during fine-tuning only and also pre-trained with over 720k instantaneous HD-sEMG images.

Fig. 2 presents the recognition accuracy obtained by our proposed different S-ConvNet models for 18 different subjects and their statistical significance. We achieve 87.95%, 86.94%, and 87.02% average recognition accuracy for the proposed S-ConvNet-A, B, and C models respectively, which is very competitive to the more complex, highly resource-based and fine-tuned pre-trained models proposed by the existing approaches while also reducing the learning parameters to a large extent. These high recognition accuracies for neuromuscular activity recognition based on instantaneous HD-sEMG images indicate the stability and potentiality of the proposed S-ConvNet models.

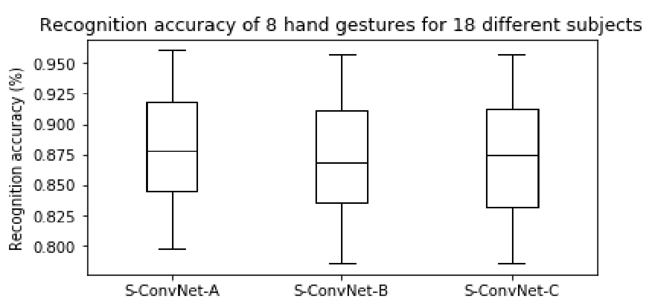

Fig. 2. The recognition accuracy of 8 hand gestures for 18 different subjects with our proposed S-ConvNet recognition approaches.

The recognition accuracy of 8 hand gestures of all 18 subjects in CapgMyo DB-a which obtained through leave one trial out cross-validation for 10 different trials using S-ConvNet-A and their statistical significance are presented in Fig. 3. It is observed that the average recognition accuracy >93% and >88%. have been achieved at least for 6 and 5 different subjects respectively. Moreover, the high average recognition accuracies 94.29%, 96.55% and 98.87% are achieved by a simple majority voting with 3, 20 and 40 instantaneous images respectively (Table II, S-ConvNet-A). All these highly promising and competitive results proved that the proposed S-

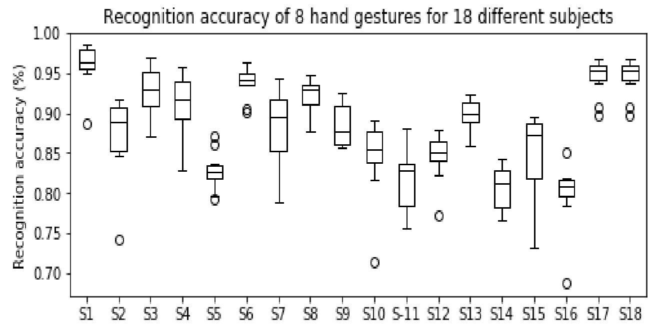

Fig. 3. The recognition accuracy of 8 hand gestures for 18 different subjects with our proposed S-ConvNet-A.

ConvNet models trained from *random-initialization* can learn all the necessary invariances that require to build a discriminant representation using only the available target dataset for neuromuscular activity recognition based on instantaneous HD-sEMG images. Therefore, the performance of the proposed S-ConvNet models is no worse than its more complex, highly resource-based, pre-trained and fine-tuned state of the art models. It is also worth mentioning that why we did compare our results only with [13] and not with the [16] and [30]. Because the same pre-trained and fine-tuned model employed in [13] was used in these successive studies, however, to address the same problem with a different view. Now, for the fair comparison with the state of the art, the following points are required to be highlighted.

We introduce a leave one trial out cross-validation in which our proposed S-ConvNet models are tested with 80k different samples for every subject. Existing instantaneous HD-sEMG image recognition approaches are tested with 40k samples for each of the subjects. Whereas we have used 80k samples (twice the number of testing samples) for recognition and achieved comparable performance on par with the state of the art. It is also noteworthy that the recognition results of all S-ConvNet models are obtained without any hyper-parameter tuning. Therefore, we also want to stress out that the results of all models evaluated in this paper could potentially be improved or even surpass the state of the art by a thorough hyperparameter tuning.

Finally, we argue that as aforementioned briefly, training from scratch is of critical importance at least for the following reasons. First, *Domain mismatch*– the distributions of the sEMG signals vary considerably even between recording sessions of the same subject. This problem becomes even more challenging, where the learned model is used to recognize muscular activities in a different recording session. Though the fine-tuning of the pre-trained model can reduce the gap due to the deformations in a new recording session. But what if we have a technique that can learn HD-sEMG images from scratch for recognizing neuromuscular activities. Second, the fine-tuned pre-trained model restricts the structure design space for neuromuscular activity recognition. This is very critical for the deployment of deep neural network models to the resource limited scenarios.

V. Conclusion

We present S-ConvNet models, a simple yet efficient framework for learning instantaneous HD-sEMG images from scratch for neuromuscular activity recognition. Without using any pre-trained models, our proposed S-ConvNet demonstrates very competitive accuracy to the more complex state of the art for neuromuscular activity recognition based on instantaneous HD-sEMG images, while using ≈ 12×smaller dataset and reducing learning parameters to ≈ 2M. The proposed S-ConvNet has great potential for learning and recognizing neuromuscular activities on resource-bounded devices. Our future work will consider improving inter-session neuromuscular activity recognition performances, as well as learning S-ConvNet models to support resource-bounded devices.

Acknowledgment

This work was supported in part by the Regroupement stratégique en microsystèmes du Québec (ReSMiQ) and the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

[1] D. Farina *et al.,* "The extraction of neural information from the surface EMG for the control of upper-limb prostheses: merging avenues and challenges," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* Vol 22**,** no. 2, pp. 797–809, 2014.

[2] G. Jang, J. Kim, J, S. Lee, and Y. Choi, "EMG-based continuous control scheme with simple classifier for electric-powered wheelchair," *IEEE Transactions on Industrial Electronics,* Vol. 63, no. 6, pp. 3695–3705, 2016.

[3] R. Jimenez-Fabian, and O. Verlinden, "Review of control algorithms for robotic ankle systems in lower-limb orthoses, prostheses, and exoskeletons," *Medical Engineering & Physics,*Vol 34, no. 4, pp. 397–408, 2012.

[4] D.-H. Kim, *et al.,* "Epidermal electronics," *Science* 333**,** 838–843, 2011.

[5] Y. Hu, J. N. Mak, & K. Luk, "Application of surface EMG topography in low back pain rehabilitation assessment," *International IEEE/EMBS Conference on Neural Engineering,* pp. 557–560, 2007.

[6] E. Costanza., S. A. Inverso, R. Allen, R. and P Maes, "Intimate interfaces in action: assessing the usability and subtlety of EMG-based motionless gestures," *Conference on Human Factors in Computing Systems*, ACM, pp. 819–828, 2007.

[7] T. S. Saponas, D. S. Tan, D. Morris, and R. Balakrishnan, "Demonstrating the feasibility of using forearm electromyography for muscle-computer interfaces," *Conference on Human Factors in Computing Systems*, ACM, pp. 515–524, 2008.

[8] T. S. Saponas, D. S. Tan, D. Morris, D, J. Turner and J. A. Landay, "Making muscle-computer interfaces more practical," *Conference on Human Factors in Computing Systems,* pp. 851–854, ACM, 2010.

[9] M. Atzori *et al.,* "Electromyography data for non-invasive naturally-controlled robotic hand prostheses," *Scientific Data* 1**,** 2014.

[10] N. Patricia, T. Tommasi. & B. Caputo, "Multi-source adaptive learning for fast control of prosthetics hand," *International Conference on Pattern Recognition,* pp. 2769–2774, 2014.

[11] C. Amma, T. Krings, J. Ber, J. and T. Schultz, "Advancing muscle-computer interfaces with high-density electromyography," *Conference on Human Factors in Computing Systems,* pp. 929–938, ACM, 2015.

[12] A. Stango, F. Negro and D. Farina, "Spatial correlation of high density EMG signals provides features robust to electrode number and shift in pattern recognition for myocontrol," *IEEE Transactions on Neural Systems and Rehabilitation Engineering* Vol 23, no.2, pp. 189–198, 2015.

[13] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu and J. Li, "Gesture recognition by instantaneous surface EMG images," *Sci. Rep.,* Vol 15, no.*6*, 36571, Nov 2016.

[14] R. Casale and A. Rainoldi, "Fatigue and fibromyalgia syndrome: clinical and neurophysiologic pattern," *Best Practice & Research Clinical Rheumatology* Vol 25**,** no. 2, 241–247, 2011.

[15] M. R. Islam, D. Massicotte, F. Nougarou and W. Zhu, "HOG and Pairwise SVMs for Neuromuscular Activity Recognition Using Instantaneous HD-sEMG Images," *2018 16th IEEE International New Circuits and Systems Conference (NEWCAS)*, Montreal, QC, 2018, pp. 335-339.

[16] Y. Du., W. Jin, W. Wei, Y. Hu and W Geng, "Surface EMG-based inter-session gesture recognition enhanced by deep domain adaptation," *Sensors*, Vol. 17, no. 3, 458, 2017.

[17] Y. Taigman, M. Yang, M. Ranzato and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, pp. 1701-1708, 23-28 June, 2014.

[18] S. Gupta, J. Hoffman, and J. Malik., "Cross modal distillation for supervision transfer," *In CVPR*, 2016.

[19] Z. Shen, Z. Liu, J. Li, Y.-G. Jiang, Y. Chen, and X. Xue., "DSOD: Learning deeply supervised object detectors from scratch," *In ICCV*, 2017.

[20] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. A. Riedmiller., "Striving for Simplicity: The All Convolutional Net," CoRR, abs/1412.6806, 2014.

[21] X. Glorot and Y. Bengio., "Understanding the difficulty of training deep feedforward neural networks," *In AISTATS*, 2010.

[22] S. Ioffe and C. Szegedy., "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *In ICML*, 2015.

[23] Srivastava, Nitish, Hinton, Geoffrey E, Krizhevsky, Alex, Sutskever, Ilya, and Salakhutdinov, Ruslan., "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, Vol 15, no. 1, pp.1929–1958, 2014.

[24] Lin, Min, Chen, Qiang, and Yan, Shuicheng., "Network in network," *In ICLR*: Conference Track, 2014.

[25] Simonyan, Karen and Zisserman, Andrew., "Very deep convolutional networks for large-scale image recognition," In arxiv:cs/arXiv:1409.1556, 2014.

[26] Janocha, Katarzyna, and Wojciech Marian Czarnecki. "On loss functions for deep neural networks in classification." *arXiv preprint arXiv:1702.05659,* 2017.

[27] Kingma, Diederik and Ba, Jimmy., "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

[28] K. He, X. Zhang, S. Ren, and J. Sun., "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," *In ICCV*, 2015.

[29] F. Nougarou, A. Campeau-Lecours, R. Islam, D. Massicotte and B. Gosselin, "Muscle Activity Distribution Features Extracted from HD sEMG to Perform Forearm PatternRecognition," *2018 IEEE Life Sciences Conference (LSC)*, Montreal, QC, pp. 275-278, Oct. 2018.

[30] W. Wei, Q. Dai, Y. Wong, Y. Hu, M. Kankanhalli and W. Geng, "Surface-Electromyography-Based Gesture Recognition by Multi-View Deep Learning," in *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 10, pp. 2964-2973, Oct. 2019.

[31] K. He, R. Girshick, and P. Dollar, "Rethinking imagenet pretraining", *In ICCV*, 2019.

[32] M. R. Islam, D. Massicotte, F. Nougarou, P. Massicotte and W. Zhu., "S-ConvNet: A shallow convolutional neural network architecture for neuromuscular activity recognition using instantaneous high-density surface EMG images," *arXiv preprint arXiv:1906.03381*, 2019.

[33] M. R. Islam, D. Massicotte and W. Zhu., "All-ConvNet: A Lightweight All CNN for Neuromuscular Activity Recognition Using Instantameous High-Density sEMG Images," IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Dubrovnik, Croatia, 2020, pp. 1-5.