

Intersected EMG Heatmaps and Deep Learning Based Gesture Recognition

Weijie Ke

CeSIP, Royal College Building
University of Strathclyde, Glasgow
+44 7597080995

weijie.ke@strath.ac.uk

Lykourgos Petropoulakis

CeSIP, Royal College Building
University of Strathclyde, Glasgow

Yannan Xing

CeSIP, Royal College Building
University of Strathclyde, Glasgow
+44 7741004094

yannan.xing@strath.ac.uk

John Soraghan

CeSIP, Royal College Building
University of Strathclyde, Glasgow

Gaetano Di Caterina

CeSIP, Royal College Building
University of Strathclyde, Glasgow

ABSTRACT

Hand gesture recognition in myoelectric based prosthetic devices is a key challenge to offering effective solutions to hand/lower arm amputees. A novel hand gesture recognition methodology that employs the difference of EMG energy heatmaps as the input of a specific designed deep learning neural network is presented. Experimental results using data from real amputees indicate that the proposed design achieves 94.31% as average accuracy with best accuracy rate of 98.96%. A comparison of experimental results between the proposed novel hand gesture recognition methodology and other similar approaches indicates the superior effectiveness of the new design.

CCS Concepts

• **Computing methodologies** → **Machine learning** → **Machine learning approaches** → **Neural Networks**

Keywords

Convolutional Neural Network, Gesture Recognition, EMG, Signal Processing.

1. INTRODUCTION

Electromyography (EMG) involves the measurement of electrical activity from functional skeletal muscles [1]. The arm movements usually are triggered by the central nervous system, which is controlled by the brain sending signals to the forearm. The points where muscle fibres are triggered by the nerves are called motor units (MU), and these transmit electrical signals which cause muscles to contract and relax. The voltage generated by the motor units is called motor unit action potential (MUAP).

Nowadays EMG is used worldwide in medical examinations for evaluating the health condition of muscles and the nerve cells that control them. Diseases such as disorders affecting the connection between nerves and muscles are easily identified through EMG [2]. Furthermore, EMG signals are also commonly utilized as control signals for prosthetic devices [3] (prosthetic hands, arms, and lower limbs).

SAMPLE: Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Conference'10, Month 1–2, 2010, City, State, Country.
Copyright 2010 ACM 1-58113-000-0/00/0010 ...\$15.00.

DOI: <http://dx.doi.org/10.1145/12345.67890>

Currently, there are two types of frequently used sensors for EMG signal recording, surface and intramuscular [3]. Surface sensors, also known as sEMG sensors, record muscle activity from above the skin, providing data from the motor units that are near the skin. Often, sEMG sensors comprise two electrodes, one of which is used as a reference and one to measure the potential difference with respect to the first sensor. Other configurations, such as a bipolar sensor, also exist. Intramuscular EMG is generally performed by inserting a fine wire directly into a muscle, measuring the activity of a single or various motor units. In this configuration a surface EMG sensor is used as a reference point.

The interest in EMG signal has increased over the years among research and industrial communities because of their efficiency in representing the muscle activities [4]. In the application of hand gesture recognition, various studies have demonstrated the applicability of gesture prediction utilizing EMG signals [5, 6, 7, 8, 9, 10]. However, currently one of the main issues in the field is the limitations in recognition accuracy, due to the difficulty of extracting features from multi-sensory EMG signals. Most of the studies emphasize the selection and noise reduction of raw EMG signals, which leads to a lack of universality.

In this work, we introduce a novel approach for gesture recognition by transferring raw EMG signals to heatmaps and feeding these into a convolutional deep neural network. We evaluated the performance of this approach using an EMG dataset [11] combined with the convolutional neural network.

The remainder of this paper is organized as follows. Section 2 introduces the sEMG dataset used for this work. Section 3 presents the experimental methodology including the generation method of differed heatmaps and the structure of the convolutional network. Section 4 presents the results and compares them with other hand gesture recognition methodologies and ours. Conclusions are presented in section 5.

2. DATASET

The dataset used in this work was obtained using sEMG sensors. The sEMG signal consists of the summation of the different MUAP within its reach plus some additional external noise. When humans perform a muscle contraction, different MUs are activated in turn, thus preventing the individual motor units from becoming fatigued [11]. MUAP's recorded from different MUs show different waveforms, whereas MUAP's recorded by the same electrode from the same MU show an almost identical waveform. Our dataset was recorded using two HD EMG arrays, each comprising 64 sensors. One EMG array was placed on the surface of extensor forearm muscles (upper band?), while the other was

placed on the flexor muscles (Lower band?). Of the 22 participants involved in the study, 9 were able bodied and the rest were amputees with different levels of hand/lower arm amputation. Sensor locations remained the same for both amputees and able handed participants when performing the same test. Figure 1 shows array attachments of test subjects' forearms.

There were 12 distinct movements performed by all subjects, which were all formed from rest and maintained for 3 to 5 seconds. Each gesture was repeated 10 times for able-bodied subjects and 5 times for amputees in a random order. However, some of the gestures performed by the amputees were different compared to the able bodies. For example, some of the amputees were unable to perform a grip. Hence, only the 8 hand activities, which both sets of subjects could achieve, were used for analysis and classification. Example images of the gestures used from the dataset are shown in figures 2 and 3. All 8 gestures shown were performed under the same conditions.

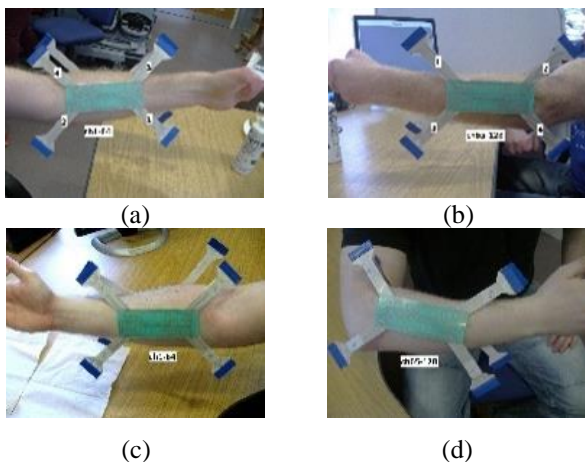


Figure 1. (a) and (b) show 1 to 128 sensors attached on amputee's forearm. (c) and (d) show 1 to 128 sensors attached on able-handed participant's forearm.

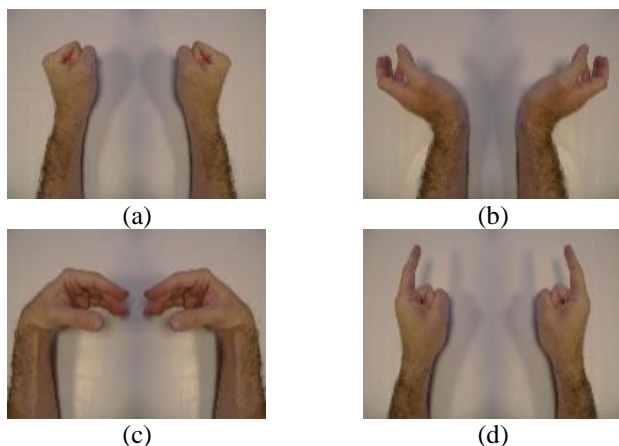


Figure 2. Classified gestures (a) close; (b) extension; (c) flexion; (d) point.

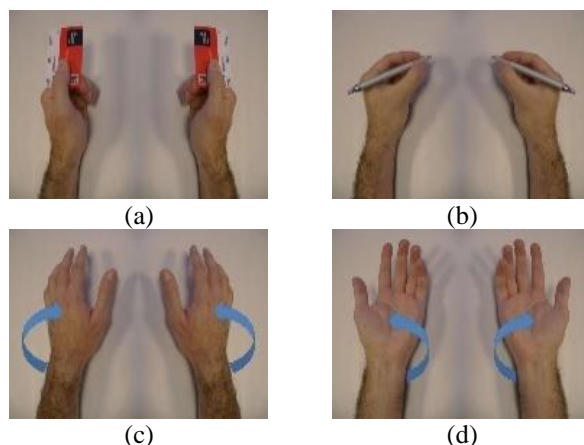


Figure 3. Classified gestures (a) lateral; (b) tripod closed; (c) palm down; (d) palm up.

3. Heatmap Difference Technique

The recorded signals were amplified and sampled at 2048 Hz. The sampled data is subsequently passed through built-in hardware filters comprising a high-pass filter with cut off frequency at 3Hz, and a low-pass filter with cut off frequency at 900 Hz, as significant EMG activity happens between 5 Hz and 450 Hz [12]. A 3rd order Butterworth digital filter was employed to remove electronic equipment noise and motion artefacts. After filtering, signals are modified by Hamming windows with length 250ms, with 50% overlap as described in [11].

Heatmaps indicate the electrical activity (energy level) of different sEMG sensors as a result of different gestures, which result in different levels of target muscle activity.

The original signal dataset contains 128 channels. Each channel represents a specific muscle area. The values shown in the heatmaps for each channel were obtained from the mean absolute value of the original EMG signal and plotted in various colors.

As an example, graphs of the heatmaps of a gesture performed by participant 5 are shown in figure 4, where the left heatmap shows the array placed on the flexor muscles and the right heatmap shows the array placed in the extensor muscles. White color areas indicate high muscle activation. The numbers indicate sensor locations in the arrays.

As seen from figure 4, most of the peak muscle activation occurred in small areas of the heatmap. This implies that when the arrays are being divided into smaller rectangles and sensors are

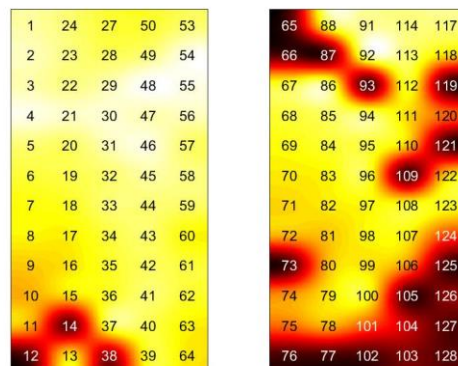


Figure 4. Heatmap for close gesture of participant 5

averaged, vital information on the specific area where the muscle activation has occurred can be lost. Noting that different gestures show different areas where muscle activation is at its highest, it is intuitively obvious that selecting sensors from an area with high activation levels will significantly enhance the performance of a classifier, as it would be able to better capture the muscle activation patterns and omit readings from sensors which provide little or no useful information.

To investigate this, we transformed RGB heatmaps into Grey heatmaps so as to enhance visibility of sensors with highest energy level (brightest color). After transformation, only the most active sensors were represented by white color. Less active or non-active sensors were turned into black. Based on previous experiments, an optimal number of sensors which provide useful information in each case is necessary to minimize overlapping common features. However, with no clear knowledge of what this optimal sensor number is in each case, an empirical strategy was employed to determine optimal sensor numbers based on sensor signal strength levels.

The procedure for evaluating these optimal numbers is as follows. Starting with just one sensor with the highest signal (brightest color) for a given gesture, a partial heatmap is formed. The second partial heatmap uses the two highest signal sensors (this includes the first sensor). This process of selecting partial heatmaps continues, each time including the sensor with the next highest signal. The process stops when 50 (out of 64) sensors are selected for each band – the last 14 sensors with very low signal strengths were ignored. Examples of the heatmaps generated by the 4 topmost sensors forming the “close” gesture for the lower band are shown in figures 5 and 6.

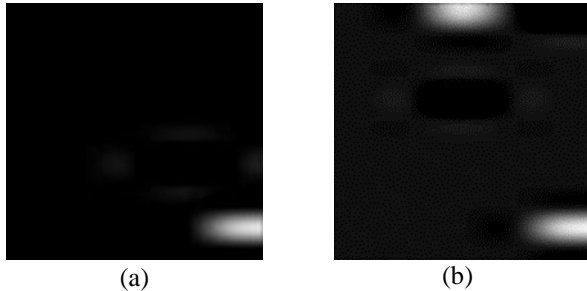


Figure 5. (a) for 1 sensor; (b) for 2 sensors

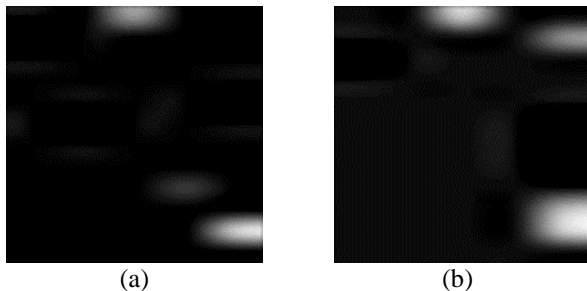


Figure 6. (a) for 3 sensors; (b) for 4 sensors (2 of the sensors are close to each other).

However, classifiers are more efficient when using common features or, in this case, common heatmap features for each gesture for all subjects. Given the mixture of healthy subjects and amputees (with varying degrees of injuries), the partial heatmaps obtained unstable information, such as outliers and signals not directly generated by the corresponding gestures and they could be slightly different for each subject. For this reason, an

intersection methodology (Algorithm 1) was used on these partial heatmaps so as to determine the common heatmap features for each gesture for all users which could then be used for classification.

Algorithm 1: Intersection Algorithm

Input: partial heatmaps for all subjects’ gestures for all numbers of sensors considered

Parameter: image type: jpeg; image size: 8×8

Output: intersected heatmaps

- 1: collect all partial heatmaps for same sensor number for all subjects’ gestures
 - 2: select one partial heatmap for one gesture from one subject
 - 3: subtract (pixelwise) the heatmap of the selected subject’s gesture from all remaining heatmaps (from all other users) for the same gesture.
 - 4: inverse pixel value of obtained subtracted heatmaps to generate intersected heatmaps between the chosen subject and each of the other subjects
 - 5: repeat steps 2 to 4 for all other subjects
 - 6: repeat steps 2 to 5 for all remaining gestures
 - 7: repeat steps 1 to 6 for all remaining sensor number selections.
-

The algorithm obtains the common heatmap features for the selected number of sensors considered in each case for all users, for each gesture in turn. In this case we had a maximum of 50 sensors and 8 gestures. These were: 1 palm up; 2 palm down; 3 tripod closed; 4 lateral; 5 point; 6 flexion; 7 extension; 8 close.

For each gesture, 340 partial heatmaps (30 partial heatmaps from 9 able handed subjects and 5 from 14 amputees) were obtain for each set of sensors selected (from 1 to 50). In total, 136000 partial heatmaps are obtained (340 per-gesture × 8 gestures × 50 sensors selection conditions). After applying the intersection algorithm for each set of sensors selected, 57630 intersected heatmaps were obtained per-gesture. In total, 23052000 intersected heatmaps were applied (57630 per-gesture × 8 gestures × 50 sensors selection conditions). For the combined two bands shown in figure 8 the best results were obtained when a maximum of 35 sensors were used.

4. NEURAL NETWORK

Due to the characteristics of the generated heatmaps, it was decided to use a VGG[13] based CNN[14] as the deep learning architecture for the classification model because of its simple and stable architecture, which enables easier modification of the network structure with low risk. The main structure of this network is shown in figure 7. The network contains three convolution blocks with max-pooling layer [15] in between, which substantially reduces the computing free-parameters and thus accelerates the training speed. Each convolution block contains 3 convolutional layers, and each convolutional layer is implemented with the fixed size 3x3 filter. Two dropout layers [16] with 512 neurons are connected to the last convolutional layer of the 3rd convolution block, which prevent the network from overfitting. The output layers consist of 8 neurons which correspond to the predicting gestures. The network is trained using the Adam gradient descent algorithm [17] combining a Softmax Classifier [18] for efficient recognition.

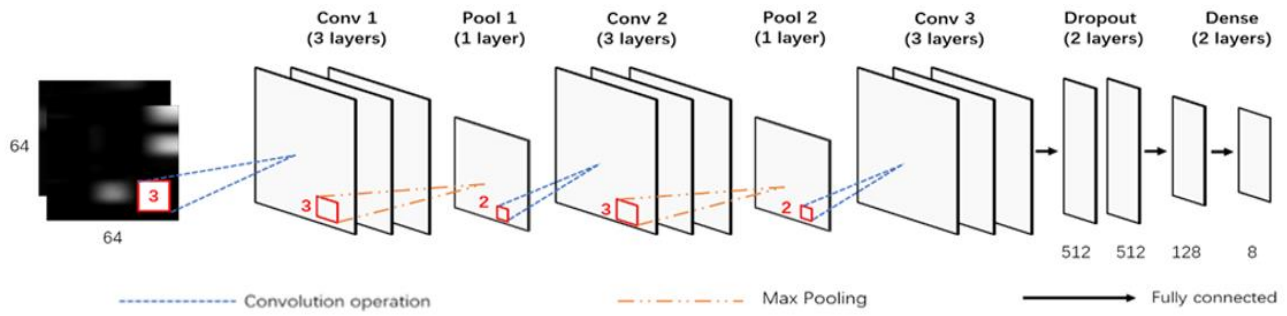


Figure 8. Network Structure

5. EXPERIMENTAL RESULTS

In order to fit the designed CNN to the heatmap, numerous experiments on network tuning were tested. We trained the network with 10 epochs and different hyper parameters, using different number of sensor inputs ranging from 1 to 50 sensors.

The 8 hand gesture classification results for the intersected heatmaps for the combined upper and lower bands are shown in figure 8. Figure 8 clearly indicates 35 sensors as being the set of gestures considered, the accuracy of the classification increases with increasing number of sensors, until 35 sensors are used. Beyond this point adding more sensors appears to have a slightly negative effect in the classification rates. Hence, for this set of gestures it is considered that 35 is the optimal number of sensors which can produce the best classification results. It should be noted that, even though figure 8 indicates 35 sensors as being the optimal number, the actual number of sensors which are common to all subjects is significantly less than 35 for each gesture. This is clearly shown in Table 1 where the common contributing sensors for each gesture are shown. Table 1 indicates that, for most cases, only around 20 sensors are contributing significantly for all users and many of them are contributing to more than one gesture. There are more contributing sensors from the upper band than the lower one. The best classification rate obtained was 98.96% when 35 sensors were considered.

The proposed method demonstrates nearly perfect results for the given dataset. Table 2 is a comparison of several methods

showing the number of gestures considered, and the number of amputees involved in the study. The last 3 entries in the table indicate the results using the proposed approach when 6, 7 and 8 gestures were considered with corresponding classification rates of 98.24%, 98.77% and 98.96% respectively. As shown in Table 2, Xiaolong Zhai et al. [19] used SVM classification methodology and achieved 77.44% accuracy on NinaPro dataset. Rezwanaul et al. [20] fed time frequency features of raw EMG signals to an artificial neural network and achieved 88.40% accuracy. Ugur Sahin et al. [21] applied wavelet transform into raw EMG dataset and obtained 96.60% with 6 hand movements. Xun Chen et al. [22] used Hudgins' Time Domain features, autocorrelation and cross-correlation coefficients and spectral power magnitudes of EMG signals and achieved 90% real time recognition accuracy with six subjects. Côté-Allard et al. [23] transfer learning algorithm on EMG signals collected from 17 subjects and obtained 97.8 % recognition accuracy with 7 gestures.

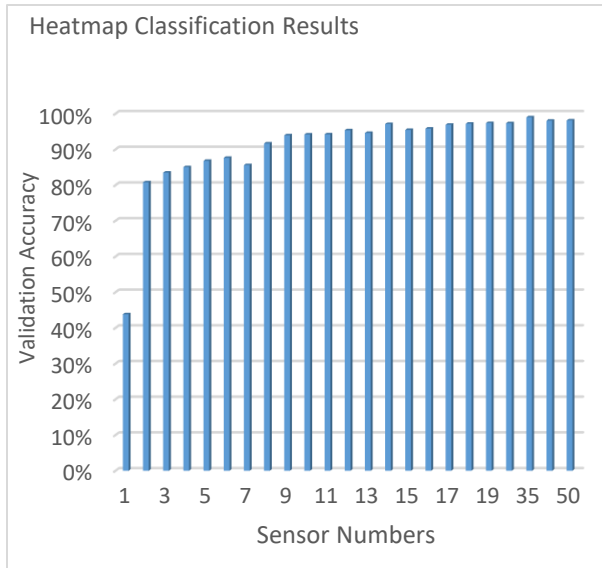


Figure 7. Classification results for combined two bands for 8 gestures.

Table 1. Common active 35 sensors for two bands

Gestures	Lower Band Common Sensor Locations	Upper Band Common Sensor Locations
palm up	48,49,50,53,54,55	103,104,105,117,118,119,125,126,127,128
palm down	24, 27, 50	107, 108, 109, 117,118,119,121,122,123,124,125,126,127,128
tripod closed	3, 4, 5, 6, 54, 55	95,96,97,105,106,107,108,117,118,122,123,126,127,128
lateral	1, 2, 3, 4, 5, 6	67,105,106,113,114,117,118,120,121,122,123,124,126,127
point	3, 4, 5, 6, 20, 21, 22	117,118,120,121,122,123,124,126,127,128
flexion	2, 3, 4, 20, 21, 22	91,92,93,94,95,96,97,117,118,126,127,128
extension	46,47,48,49,53,54,55,56,57	95,96,97,111,112,113,117,118,119,125,126,127
close	3,4,5,6,7,8,27,28,29,49,50	110,111,112,113,117,118,119,125,126,127

Table 2. Classification results comparison

Method	Gesture numbers	Amputee numbers	Input data type	Results
SVM ([19])	10	11	Spectrogram	77.44 %
ANN ([20])	4	0	Time Frequency Features	88.40 %
CNN ([21])	6	0	Wavelet Transformation	96.60 %
JL ([22])	4	0	TD and other features	90.00 %
CNN ([23])	7	0	Transfer Learning	97.81 %
IA-CNN	6	14	Proposed Heatmap for both	98.24 %
IA-CNN	7	14	Proposed Heatmap for both	98.77 %
IA-CNN	8	14	Proposed Heatmap for both	98.96 %

6. CONCLUSION

In this paper, we have presented a novel heatmap difference method for personalizing EMG-based models with deep learning techniques and we have evaluated it with our high-density sEMG dataset. We have compared this performance with some proposed prior work, which used different techniques such as wavelet transforms, spectrograms and Hudgins time Domain features with various machine learning methodologies. Even though these experiment results were achieved using different categories and forms of sEMG datasets, the collection instrument of all sEMG datasets remains the same i.e. the dataset used by others contain sEMG signals just like the sEMG dataset we have used. Our experimental results show that the heatmap difference approach significantly outperforms other methodologies in terms of accuracy though some of the methodologies perform better (The latter has slightly higher accuracy than our method because of the number of gestures and subjects used, which in our case is 8 gesture classification from 23 participants, including 14 amputees), which was as high as 98.96% under and 98.96% under 35 sensors conditions for 8 hand gesture recognition based on the combined two bands. The heatmap difference approach minimizes the input data dimensions by reducing required number of attaching sensors without recognition accuracy deterioration. It also has the added advantage of minimizing the possibility of erroneous readings. This study has shown that the active sensors are not spread equally in the two bands. Thus, one would expect that the individual recognition accuracy for each band may differ. Furthermore, given that there are several sensors which contribute to more than one gestures, it is reasonable to assume that judicious selection of sensor inputs can be used to minimize the number of sensors required to obtain very high classification accuracy. For future work, we plan to implement more sEMG datasets to our neural network since only 8 hand gesture recognitions have been achieved so far.

7. ACKNOWLEDGMENTS

This work is supported by the University of Strathclyde. We gratefully acknowledge the assistance of all members in the deep learning group, Centre for Signal and Image Processing (CeSIP), EEE department.

8. REFERENCES

- [1] M. B. I. Reaz, M. S. Hussain, and F. Mohd-Yasin, "Techniques of EMG signal analysis: detection, processing, classification and applications," *Biological procedures online*, vol. 8, no. 1, p. 11, 2006.
- [2] H. J. Hermens, B. Freriks, R. Merletti, D. Stegeman, J. Blok, G. Rau, C. Disselhorst-Klug, G. Hägg, "European recommendations for surface electromyography," *Roessingh research and development*, vol. 8, no. 2, pp. 13-54, 1999.
- [3] P. McCool, "Surface myoelectric signal analysis and enhancement for improved prosthesis control," PhD Thesis, University of Strathclyde, 2014.
- [4] C. Disselhorst-Klug, T. Schmitz-Rode, and G. Rau, "Surface electromyography and muscle force: limits in sEMG-force relationship and new approaches for applications," *Clinical biomechanics*, vol. 24, no. 3, pp. 225-235, 2009.
- [5] D. Farina et al., "The extraction of neural information from the surface EMG for the control of upper-limb prostheses: emerging avenues and challenges," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 4, pp. 797-809, 2014.
- [6] 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.
- [7] D. Farina and R. Merletti, "Comparison of algorithms for estimation of EMG variables during voluntary isometric contractions," *Journal of Electromyography and Kinesiology*, vol. 10, no. 5, pp. 337-349, 2000.
- [8] E. A. Clancy, E. L. Morin, and R. Merletti, "Sampling, noise-reduction and amplitude estimation issues in surface electromyography," *Journal of electromyography and kinesiology*, vol. 12, no. 1, pp. 1-16, 2002.
- [9] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," *Expert Systems with Applications*, vol. 39, no. 8, pp. 7420-7431, 2012.
- [10] L. Hargrove, K. Englehart, and B. Hudgins, "A training strategy to reduce classification degradation due to electrode displacements in pattern recognition based myoelectric control," *Biomed. Signal Process. Control*, vol. 3, no. 2, pp. 175-180, 2008.
- [11] R. Menon, G. Di Caterina, H. Lakany, L. Petropoulakis, B. A. Conway, and J. J. Soraghan, "Study on interaction between temporal and spatial information in classification of EMG signals for myoelectric prostheses," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 10, pp. 1832-1842, 2017.
- [12] A. Van Boxtel, "Optimal signal bandwidth for the recording of surface EMG activity of facial, jaw, oral, and neck muscles," *Psychophysiology*, vol. 38, no. 1, pp. 22-34, 2001.
- [13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition." Sep.2014.

- [14] [Online]. Available: <https://arxiv.org/pdf/1409.1556.pdf>. [Accessed: 03- Mar- 2019].
- [15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," NIPS, pp. 1106-1114, 2012.
- [16] B. Graham, "Fractional max-pooling," Dec.2014. [Online]. Available: <https://arxiv.org/pdf/1412.6071.pdf>. [Accessed: 03- Mar- 2019].
- [17] J. Ba and B. Frey, "Adaptive dropout for training deep neural networks," Advances in Neural Information Processing Systems, pp. 3084-3092, 2013.
- [18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Dec.2014.[Online].Available:<https://arxiv.org/pdf/1412.6980.pdf>. [Accessed: 02- Mar- 2019].
- [19] D. Heckerman and C. Meek, "Models and selection criteria for regression and classification," Proceedings of the thirteenth annual conference on uncertainty in artificial intelligence, pp. 198-207, 1997.
- [20] X. Zhai, B. Jelfs, R. H. M. Chan, and C. Tin, "Self-recalibrating surface EMG pattern recognition for neuroprosthesis control based on convolutional neural network," Frontiers in neuroscience, vol. 11, p. 379, 2017.
- [21] M. R. Ahsan, M. I. Ibrahimy, and O. O. Khalifa, "Electromyography (EMG) signal based hand gesture recognition using artificial neural network (ANN)," Mechatronics (ICOM), pp.1-6, 2011.
- [22] U. Sahin and F. Sahin, "Pattern recognition with surface EMG signal based wavelet transformation," Systems, Man, and Cybernetics (SMC), pp. 295-300, 2012.
- [23] X. Chen and Z. J. Wang, "Pattern recognition of number gestures based on a wireless surface EMG system," Biomedical Signal Processing and Control, vol. 8, no. 2, pp. 184-192, 2013.
- [24] U. Côté-Allard, C. L. Fall, A. Campeau-Lecours, C. Gosselin, F. Laviolette, and B. Gosselin, "Transfer learning for sEMG hand gestures recognition using convolutional neural networks," in Systems, Man, and Cybernetics (SMC), pp. 1663-1668, 2017.
- [25] Brown, L. D., Hua, H., and Gao, C. 2003. A widget framework for augmented interaction in SCAPE. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology* (Vancouver, Canada, November 02 - 05, 2003). UIST '03. ACM, New York, NY, 1-10. DOI= <http://doi.acm.org/10.1145/964696.964697>.
- [26] Yu, Y. T. and Lau, M. F. 2006. A comparison of MC/DC, MUMCUT and several other coverage criteria for logical decisions. *J. Syst. Softw.* 79, 5 (May. 2006), 577-590. DOI= <http://dx.doi.org/10.1016/j.jss.2005.05.030>.
- [27] Spector, A. Z. 1989. Achieving application requirements. In *Distributed Systems*, S. Mullender, Ed. ACM Press Frontier Series. ACM, New York, NY, 19-33. DOI= <http://doi.acm.org/10.1145/90417.90738>.

Columns on Last Page Should Be Made As Close As Possible to Equal Length

Authors' background

Your Name	Title*	Research Field	Personal website
Weijie ke	Phd candidate	Deep learning	N/A
Yannan xing	Phd candidate	Deep learning	N/A
Gaetano Di Caterina	Lecturer	Video and Image processing	https://www.strath.ac.uk/staff/dicaterinagaetanodr/
Lykourgos Petropoulakis	Senior Lecturer	Robotics - Prosthetics	https://www.strath.ac.uk/staff/petropoulakislykourgosdr/
John Soraghan	Professor	DSP, deep learning	https://pureportal.strath.ac.uk/en/persons/john-soraghan

*This form helps us to understand your paper better, **the form itself will not be published.**

*Title can be chosen from: master student, Phd candidate, assistant professor, lecture, senior lecture, associate professor, full professor