On the Use of Deeper CNNs in Hand Gesture Recognition Based on sEMG Signals

Nikolaos Tsagkas, Panagiotis Tsinganos, Athanassios Skodras
Department of Electrical and Computer Engineering, University of Patras, 26504 Patras, Greece
n.tsagkas@upnet.gr, {panagiotis.tsinganos, skodras}@ece.upatras.gr

Abstract—**In the past few years, a great interest for the classification of hand gestures with Deep Learning methods based on surface electromyography (sEMG) signals has been developed in the scientific community. In line with latest works in the field, the objective of our work is the construction of a novel Convolutional Neural Network architecture, for the classification of hand-gestures. Our model, while avoiding overfitting, did not perform significantly better compared to a much shallower network. The results suggest that the lack of diversity in the sEMG recordings between certain hand-gestures limits the performance of the model. In addition, the classification accuracy on a database we developed using a commercial device (Myo Armband) was substantially higher (approximately 24%) than a similar benchmark dataset recorded with the same device.**

Keywords—Deep Learning, surface electromyography, sEMG, Convolutional Neural Networks, CNN, hand gesture recognition, classification, database, data acquisition, signal processing

I. Introduction

In the field of Human-Machine Interaction, a great interest has been developed over the past decades for the creation of user interfaces that take advantage of hand-gesture recognition. The use of surface electromyography signals (sEMG), has been widely adopted by the scientific community for the classification of hand-gestures, especially in applications like artificial limbs. However, challenges such as the lack of robustness in prosthetics control, the difficulty of adjusting the sEMG sensors without professional assistance and the limited amount of data for applying Machine Learning algorithms, lead to unnatural movement of the prosthetic.

In the past few years extensive research has been conducted to tackle the issues concerning the efficiency of the algorithms that back sEMG prosthetics. Neural networks are being used for the classification of hand-gestures from sEMG recordings since 1993. In [1], an accuracy of 91.2% was achieved for the classification of 4 hand movements. Recent approaches, on the other hand, focus on using Deep Learning (DL) algorithms, where Convolutional Neural Networks (CNN) have played a vital role in achieving high accuracy in sEMG classification. The first end-to-end DL architecture was proposed in [2]. This approach managed to surpass in accuracy the SVM classification of 6 hand-gestures, using a CNN. The authors of [3] adapted the well-known LeNet [4] architecture, for the classification of 53 gestures. Even with such a simple architecture, results were comparable to traditional classification techniques. Additionally, the works of [5] and [6] prove that the use of dropout [7] and batch normalization [8] can improve the accuracy of an sEMG – CNN classifier.

Even though the birth of the theory of DL took place in the 1940s and CNNs were first successful in 1995 [4], DL did not revolutionize Machine Learning until the construction of large databases, since DL requires a huge amount of data to achieve high accuracy. Unfortunately, this is still the case with sEMG databases, where collecting sEMG signals is time consuming and requires a great number of volunteers. Recently however, steps have been taken with the development of databases such as the capgMyo [9] and the Ninapro [10]. Nevertheless, both databases have been developed with custom sEMG devices that require professional assistance to calibrate.

Our aim is to further continue the research in hand-gesture recognition from sEMG recordings, by proposing a novel CNN architecture and developing a new sEMG database, with the Myo Armband. The contributions of this work are the development of:

- a novel CNN architecture for the accurate classification of hand-gestures from sEMG signals,
- the Myo University of Patras (MyoUP) database.

II. The MyoUP Database

In order to contribute to the acquisition of sEMG data, particularly from devices that do not require professional calibration, we developed a sizeable sEMG database. Our database, MyoUP, was inspired by the Ninapro database and all of the recorded hand-gestures, presented in Fig. 1, are identical to some of the Ninapro [10]. The recording device we used was the Myo Armband, by Thalmic labs. The Myo Armband is a relatively cheap and easy-to-wear device, with a sampling frequency of 200Hz and 8 dry sEMG channels that has been widely adopted in scientific research [11, 12, 13].

The MyoUP database contains recordings from 8 intact subjects (3 females, 5 males; 1 left handed, 7 right handed; age 22.38 ± 1.06 years). The acquisition process was divided into three parts: 5 basic finger movements, 12 isotonic and isometric hand configurations and 5 grasping hand-gestures. Volunteers became accustomed with the procedure before performing each set of exercises. Subjects were instructed to repeat each gesture 5 times, for a 5sec period, interleaved with 5sec interruptions to avoid muscle fatigue. A supervisor assisted the subjects in wearing the Myo Armband to their dominant hand so that the device would be placed in a comfortable position for the subject and the device would detect the sEMG signals accurately. The sEMG was visible to the subject on a screen along with a picture of the hand-gesture that had to be performed.

III. Data Analysis and Network Architecture

A. Data Analysis

The datasets used for the classification tasks are:

- Ninapro – DB1 [10]: Recorded with 10 Otto Bock MyoBock 13E200 electrodes,

978-1-7281-4959-2/19/$31.00 ©2019 IEEE

- Ninapro – DB5 [14]: Recorded with the Myo Armband,
- MyoUP: Recorded with Myo Armband.

Recordings from the DB1 consist of 10 sEMG signals, one from each sensor, sampled at 100Hz. Following the same methodology with [3], the signals were low pass filtered at 1Hz and a sliding window method of length 150ms and 60% overlap [15] was applied to construct the CNN input images. This way, images were constructed with a resolution of 10 x 15. The exact same methodology was used with the Myo Armband recordings, however, the images had only 8 rows (thus a size of 8 x 15), since the Myo Armband has 8 sEMG sensors.

In order to avoid overfitting and increase the model's accuracy, data augmentation was applied, thus doubling the number of images. More specifically, images were used twice in training, once in their original form and once with Additive Gaussian White Noise, with SNR equal to 30dB.

Finally, the partition of the images into training, validation and testing sets was based on the repetition of each hand-gesture during the recording. For the Ninapro database (DB1 and DB5), the testing set consisted of repetitions 2 and 7, the validation set consisted of repetitions 5 and 9 and the training set consisted of the remaining repetitions. For the MyoUP database, repetition 1 was used for validation, repetition 2 for testing and the remaining three for training.

B. *CNN Architecture*

After extensive testing a novel CNN architecture was designed. The CNN (Fig. 2) consists of six convolutional blocks of 32, 32, 32, 64, 64, 128 filters and 3 x 4, 3 x 3, 2 x 1, 1 x 3, 1 x 2, 2 x 2 filter kernels respectively. All convolutions are followed by a non-linear activation function (rectified linear unit – ReLU), while max pooling layers of size 2 x 2 are added after the 2nd and 4th convolutions. In addition, the 2nd, 4th and 6th convolutional blocks are followed by a dropout layer with a probability rate of 0.15. Then, three dense layers are added to the network with 512, 128, and 64 nodes respectively, followed by ReLU non-linearities and dropout layers. The final layer of the neural network is a G-way fully connected layer followed by a SoftMax activation function, where G is equal to the number of gestures. Model weights were randomly initialized using the Xavier initializer [16].

Without the dense layers, the accuracy of the classifier would be significantly lower. On the other hand, increasing the depth and accordingly the number of weights in a neural network could result in overfitting. This was avoided with dropout layers after each dense layer. To select the dropout probability rate, we experimented with a range of values from which a probability of 0.25 achieved the best results.

Fig. 1. All the hand-gestures of the MyoUP databases. The gestures are categorized into three different groups based on the type of movement.

Selecting the hyper-parameters of the network was performed by five-fold cross validation and manual tuning on a validation set composed of four subjects randomly selected from DB1. We tested 48 different combinations of hyper-parameters as can be seen in Table I. Additionally, the weight decay and the momentum were identical to the ones used in [3]. Finally, as far as the optimizer is concerned, stochastic gradient descent (SGD) was favored over the Adam optimization method, since the latter resulted in overfitting after a few epochs (Fig. 3). This way, we managed to stabilize the validation loss and allow the validation accuracy to increase by approximately 3%. The optimal values of the hyper-parameters can be seen in Table I.

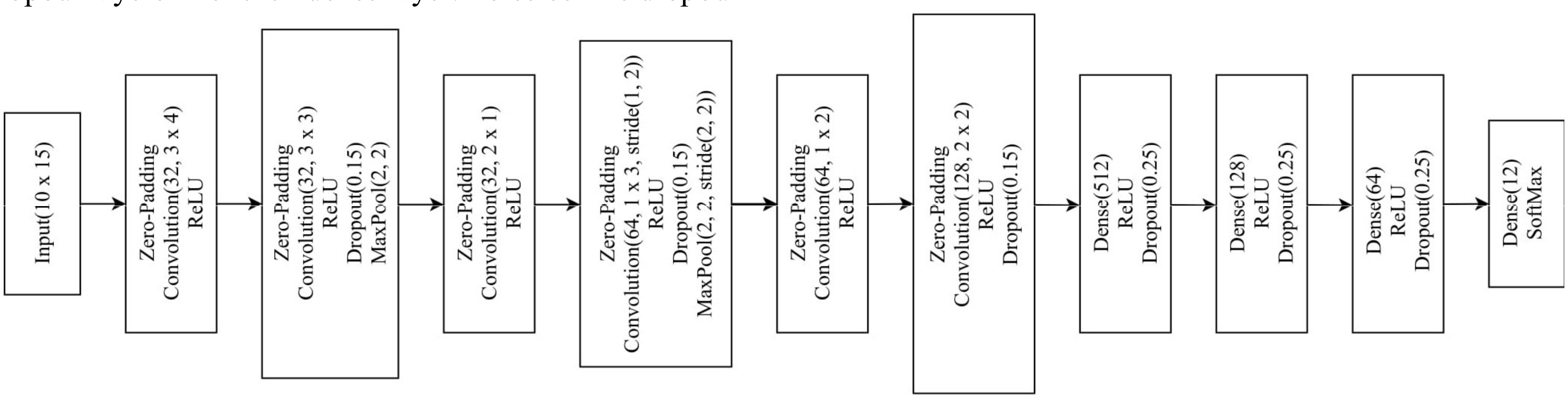

Fig. 2. CNN architecture for the classification of hand gestures.

IV. Results And Discussion

To evaluate our model, a series of experiments were performed using both a benchmark dataset (Ninapro DB1 and DB5) and the database we developed (MyoUP). With the Ninapro DB1, we assessed the performance of the network on three representative groups of hand gestures and we also compared the classification accuracy with the accuracy of a shallower CNN [3]. Then, experiments with the Ninapro DB5 and the MyoUP were conducted to measure the behaviour of the network when sEMG data are recorded with a low-cost commercial device.

A. Ninapro DB1

Firstly, we tested our architecture using the DB1 dataset from the Ninapro database. We trained our model with the recordings from the three different hand-gesture groups (E1, E2 and E3) for each individual subject. The first group contained 12 basic movements of the fingers, the second contained 8 isometric and isotonic hand configurations and 9 basic movements of the wrist, while the third group consisted of 23 grasping and functional movements. Consequently, we trained a new CNN model for every one of the 27 subjects using a different set of gestures each time. The assessment of the results is based on the average accuracy on the testing set across subjects. From the results (Table II – A), we see that it is easier for the network to distinguish between gestures of the first group (average accuracy 80.04%), which contains simple finger movements, than to correctly recognize complex grasps (average accuracy 64.42%).

Afterwards, we trained our model using the DB1 dataset, but this time we used the recordings from all subjects to train the network once. The results (shown in Table II – B), are very close to the average values calculated in the previous experiment. However, in DL, accuracy is expected to rise as the amount of data increases, which was not the case with this experiment. An explanation to this is given in [15], were a confusion matrix is shown for the classification of all the movements in DB1. It is explained that similar gestures, like the adduction and flexion of the thumb, may lead to misclassification. Fig. 4 presents the confusion matrix for the classification of hand-gestures calculated for a random subject. We repeated this process five times in order to understand which combinations of movements cause misclassification systematically. Indeed, all the confusion matrices proved that parts of sEMG signals from certain gestures are very similar for the classifier to recognize accurately. This phenomenon applies mostly to similar hand-gestures, like the "thumb opposing the base of the little finger" (E2, gesture 4) with the "abduction of all the fingers" (E2, gesture 5).

Finally, regarding our previous hypothesis, even though Table II suggests that we opt for a selection of hand-gestures with diverse sEMG signals, our architecture still managed to achieve better results than the ones presented in [3]. In Table III, a comparison can be seen between the results from our CNN, the CNN in [3] and the most successful algorithm in [3]. Our model surpasses in accuracy the CNN of [3], as we expected, taking into consideration that our CNN was deeper, more complex and trained with more data, since we used 60% overlap as opposed to not using overlap at all. On the other hand, the performance of our model was inferior to the algorithm with the highest results in [3], which was a Random Forest.

TABLE I. Hyper-Parameter Range and Optimal Values.

Hyper-parameter	Search range	Optimal value
Learning Rate	0.1, 0.01, 0.001	0.01
Batch Size	64, 256, 1024, 2048	1024
Epochs	25, 50, 100, 150	50
Weight Decay	-	1e-6
Momentum	-	0.9

B. Ninapro DB5 and MyoUP

The reason we decided to use the Myo Armband is that even though the performance in classification tasks was not high enough when the Ninapro Myo Armband recordings were used, the perks, that the Myo Armband possesses, demand that we investigate whether it can perform better. For this purpose, we opted for hand-gestures that have both diverse sEMG signals and practical applications, such as the "medium diameter grasp".

We tested how accurately our architecture can classify 12 hand gestures from the DB5 dataset compared to how well it can classify the exact same hand gestures from the MyoUP database. The most accurately classified gestures by our network, when trained with the DB1 dataset, were selected. The DB5 dataset, contains data from two Myo Armbands that are recorded simultaneously. We made our comparison using the Myo Armband that produced the highest accuracy. As is shown in Table IV, our sEMG database can produce much better results with the same architecture. For this experiment, we trained our network with the same number of images in each case. Thus, 130K images were extracted from 8 subjects and 5 repetitions (randomly selected in the case of DB5).

Afterwards, we examined how accurately our model can classify all the available hand gestures from the MyoUP

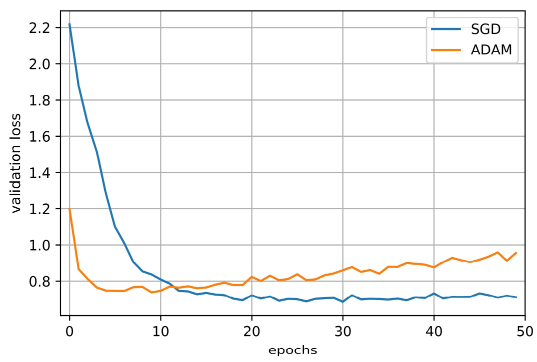

Fig. 3. The loss graphs on the validation set for different optimizers.

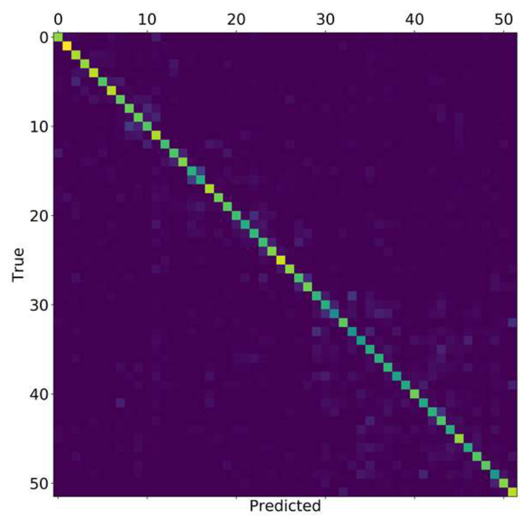

Fig. 4. Average confusion matrix constructed by the predictions of five CNNs trained with the recordings of five random subjects. Lighter colors correspond to higher values

TABLE II. COMPARISON BETWEEN THE RESULTS OF OUR CNN AFTER BEING TRAINED: A) WITH THE RECORDINGS FROM ONE SUBJECT AND B) WITH ALL THE AVAILABLE DATA.

Experiment	E1	E2	E3
A*	80.04% ± 6.30%	73.44% ± 6.01%	64.42% ± 8.36%
B	77.01%	71.02%	61.25%

* values represent the average and standard deviation across the subjects

TABLE III. COMPARISON BETWEEN OUR RESULTS AND THE RESULTS FROM [3], FOR THE DB1 DATASET.

	CNN [3]	Best [3]	Our CNN
Accuracy*	66.59% ± 6.40%	75.32% ± 5.69%	71.85% ± 5.56%

* values represent the average and standard deviation across the subjects

TABLE IV. COMPARISON BETWEEN THE RESULTS FROM THE MYOUP DATABASE AND THOSE FROM THE NINAPRO, DB5 DATASET.

	Ninapro	MyoUP
Test set acc.	55.31%	78.98%

TABLE V. CLASSIFICATION RESULTS ON THE MYOUP.

	Training Accuracy	Training Loss	Testing Accuracy	Testing Loss
All Gestures	79.57%	0.6418	72.83%	0.7102
11 Gestures	91.26%	0.2725	86.02%	0.4915

database. Even though the accuracy was almost 73%, the confusion matrix suggested again that the network struggles to classify correctly similar movements. Consequently, we chose the 11 most successfully recognized gestures, which resulted to increasing the accuracy, by approximately 14%, as it can be observed in Table V.

V. CONCLUSIONS

In this paper we presented our novel CNN architecture for classifying hand-gestures via sEMG signals. Our network, after being trained with the DB1 dataset from the Ninapro database, managed to surpass in accuracy the CNN model presented in [3], by approximately 5.5% on average. However, taking into consideration the fact that our model was much deeper and more complex, better results were expected. Our hypothesis is that some hand gestures produce similar sEMG recordings, which could lead to misclassification. This speculation is based on the confusion matrix from [15], were it is clearly shown that certain hand movements are often mistaken for others, e.g. the flexion and the adduction of the thumb. This is further supported by the fact that after training our network 27 different times, each time with the recordings from a different subject, the average accuracy was similar to the one produced after training our model with the recordings from every subject at once. Since in the second experiment the amount of data used for training was far greater than the first time, the DL algorithm theoretically should have performed better.

Afterwards, we presented our sEMG database, MyoUP, which was recorded with the Myo Armband. After training the same network with both the DB5 dataset, from the Ninapro database, and our database, the latter resulted to much higher accuracy. More specifically, the MyoUP database produced an accuracy of 78.98%, which was approximately 23.7% higher than the one produced by the DB5 dataset.

The presented results suggest that both our network architecture and our sEMG database are reliable for hand-gesture classification tasks. However, further research must be conducted in order to determine whether the classification of certain groups of hand-gestures could result to higher accuracy.

ACKNOWLEDGMENTS

This research was supported by the Laboratory of Digital Signal and Image Processing, Department of Electrical and Computer Engineering, University of Patras, Greece. Additionally, the authors would like to thank Prof. Konstantinos Moustakas, head of the Visualization and Virtual Reality Research (VVR) Group of the Wire Communications Laboratory (WCL), Department of Electrical and Computer Engineering, University of Patras, Greece for providing the Myo Armband used for the sEMG signal acquisition.

REFERENCES

[1] B. Hudgins, P. Parker, and R.N. Scott. “A new strategy for multifunction myoelectric control,” IEEE Transactions on Biomedical Engineering. 40, 1, 82–94, 1993.

[2] K. Park, and S. Lee. “Movement intension decoding based on deep learning for multiuser myoelectric interfaces,” In Proceedings of the 4th International Winter Conference on Brain-Computer Interface (BCI), 1–2, 2016.

[3] M. Atzori, M. Cognolato, and H. Müller. “Deep Learning with Convolutional Neural Networks applied to electromyography data: A resource for the classification of movements for prosthetic hands,” Frontiers in Neurorobotics, 10, 2016.

[4] Y. LeCun, et al. “Comparison of learning algorithms for hand written digit recognition,” In Proceedings of the International Conference on Artificial Neural Networks, Paris, 53–60, 1995.

[5] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu, and J. Li. “Gesture recognition by instantaneous surface EMG images,” Scientific Reports, 6(36571), 2016.

[6] W. Wei, Y. Wong, Y. Du, Y. Hu, M. Kankanhalli, and W. Geng. “A multi-stream Convolutional Neural Network for sEMG-based gesture recognition in muscle-computer interface,” Pattern Recognition Letters, 2017.

[7] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. “Dropout: A simple way to prevent Neural Networks from overfitting,” Journal ofMachine Learning Research, 15:1929–1958, 2014.

[8] I. Sergey, and C. Szegedy. “Batch Normalization: Accelerating Deep Network training by reducing in- ternal covariate shift,” ArXiv e-prints, 2015.

[9] Y. Du, W. Jin, W. Wei, Y. Hu, and W. Geng. “Surface EMG-Based Inter-Session Gesture Recognition Enhanced by Deep Domain Adaptation,” Sensors (Basel, Switzerland), 17(3), 458, 2017.

[10] M. Atzori, et al. “Electromyography data for non-invasive naturally-controlled robotic hand prostheses,” Scientific Data, 1(140053), 2014.

[11] U. Côté-Allard, et al. “Deep Learning for electromyographic hand gesture signal classification using transfer learning,” ArXiv e-prints, Jan. 2018.

[12] M. Zia ur Rehman, et al. “Multiday EMG-based classification of hand motions with Deep Learning techniques,” Sensors, 18, 8, 2497, 2018.

[13] Y. Zhang, Y. Chen, H. Yu, X. Yang, W. Lu, and H. Liu. “Wearing-independent hand gesture recognition method based on EMG armband,” Pers. Ubiquitous Comput., 22, 3, 511–524, 2018.

[14] S. Pizzolato, et al. “Comparison of six electromyography acquisition setups on hand movement classification tasks,” PLOS ONE 12(10): e0186132, 2017.

[15] P. Tsinganos, B. Cornelis, J. Cornelis, B. Jansen, and A. Skodras. “Deep Learning in EMG-based Gesture Recognition,” In Proceedings of the 5th International Conference on Physiological Computing Systems (PhyCS), Seville, 107-114, 2018.

[16] X.Glorot and Y.Bengio. “Understanding the difficulty of training deep feedforward neural networks,” In Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS), Sardinia, 249-256, 2010.