# Neuroprosthetic Grasping Generation Improvement using sEMG and Deep Generative Modeling

María Romeo Tricas    Jaume Gual Ramon    Dominik Winkelbauer

***Abstract*—Although numerous technological and scientific developments have been made in poly-articulated upper limb prostheses (ULPs) over the past 20 years [1], a recent study by Salminger et al. [2] discovered that the overall rate of ULPs abandonment was around 50% among myoelectric prosthesis users. Therefore, this work presents a two-step pipeline to improve the neuroprosthetic grasping and usage for ULP users, enhancing substantially the usage control and freedom of ULPs for the users. The first part of the pipeline involves the classification of residual sEMG signals from the user's limb using a ResNet18 classifier. In the second stage, a deep generative diffusion model is used to fine-tune the final hand pose of the grasping type, considering the desired hand pose by the user, and the object which is to be grasped. Outstanding results are obtained for the sEMG classifier, with nearly 96% classification accuracy. On the other hand, the diffusion model is capable of generating the joint angles of the hand almost perfectly; however, the position and wrist rotation of the hand pose with respect to the reference object is not always good, and can yet be improved.**

***Index Terms*—Upper Limb Prostheses, Electromyography, Grasping Generation, Deep Learning, Deep Generative modeling**

## I. Introduction

Due to difficulties in usage control and grasping selection, the abandon rate of neuroprosthetic devices for upper limb amputees is around 50 % [2]. Upper limb prostheses play a vital role in restoring functional independence and enhancing the quality of life for individuals with upper limb amputations. These devices aim to replicate natural limb movements and enable users to perform a wide range of everyday tasks with precision and control. The underlying motivation of this project is to address the urgent need of developing precise, robust, and reliable grasping pattern recognition systems. These systems are crucial, not only for amputee device control, but also in many areas such as human-computer interfaces, virtual reality applications, and sign language detection [4].

Recently, the control of upper limb prostheses has seen significant advancements through the utilization of surface electromyography (sEMG) signals [5]. EMG is the graphic recording technique of electrical activity produced by skeletal muscles and is obtained with surface electrodes on the muscles. sEMG is especially interesting, due to its cheapness and non-invasiveness for users. By capturing electrical activity from residual muscles, sEMG technology enables intuitive and real-time control of prosthesis movements, allowing users to achieve more natural and personalized interactions with their prosthetic limbs.

Frequency domain analysis of EMG signals is crucial for extracting valuable insights about muscle activation patterns and understanding the nuances of muscle contractions [6]. In this context, spectrograms are a visual representation of the frequency content of a signal over time and offer a powerful tool for analyzing EMG signals. Together, by converting the raw time-domain EMG signals from amputees remaining limbs into the frequency domain, spectrograms provide a detailed view of how different muscles contribute to complex grasping patterns and allow to identify distinct patterns associated with various grasping or movement tasks, aiding in the development of accurate and personalized control strategies.

Graspings, in the context of neuroprosthetics and robotics, refer to specific hand and limb configurations and movements designed to interact with objects, and are essential for achieving effective object manipulation and mimic the ways humans grasp and manipulate items in daily activities. Basic ULPs are capable of imitating basic hand positions and graspings [7]. However, reproducing the dexterous grasping capabilities of the human hands, with their high degree of freedom is a much more complex task.

Therefore, the main objective of this work is to develop a two-step pipeline (figure 1) for ULP users to generate a wide variety of precise grasping positions of a polyarticular prosthetic hand, using surface EMG signals and deep generative modeling. First, the sEMG signals are processed and classified into different main grasping types [3], [4], to attend to the user's intention when grasping an object. Then, the generative model optimizes the basic grasp configuration and adjusts it to the object that the user wishes to interact with [8]–[10]. Overall, this allows for enhanced usage and control of the ULP and offers a simpler way of optimizing the grasp configuration by considering the characteristics of the object to be grasped.

## II. Technical Outline and Results

### A. sEMG Classification

Here, the sEMG signals from [12] were used, which contained the sEMG data for 6 basic graspings (spheric, cylindric, hook, palmar, tip, and lateral), with 2 channels (anterior and posterior forearm muscles), for a total of 5 subjects and 2700 trials. Further, the sEMG signals had already been processed in the spectral domain, and epoched for the individual trials. The sampling frequency of the signals was 500 Hz. Reproducing Miften et al. [13] approach, the spectrograms of the sEMG signals for the 6 grasps were computed, using the *stft*

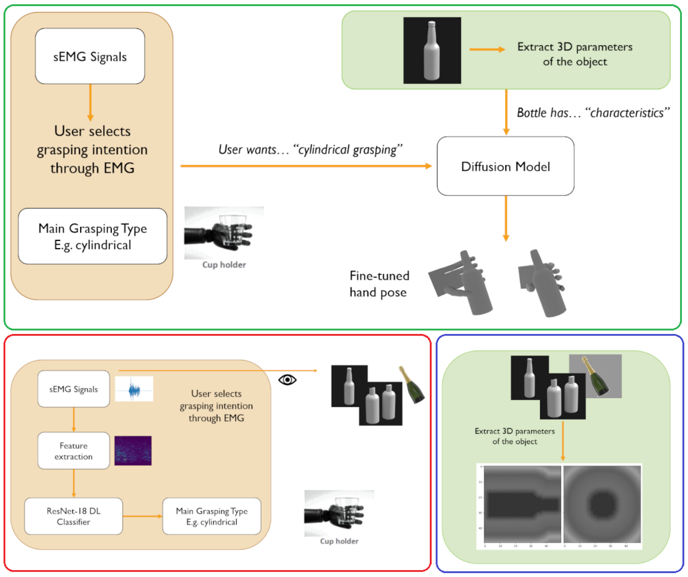


Fig. 1. Pipeline of the project.

| Batch size | Epochs | Learning rate | Momentum |
|---|---|---|---|
| 10 | 30 | 0.001 | 0.9 |

TABLE I
HYPERPARAMETERS USED TO TRAIN THE MODEL.

function, with 256-samples Hanning windows, and a 50% of overlapping. Afterward, the absolute value of each spectrogram was extracted, normalized within the same range, and plotted. Since 2 channels were contained in each trial, the two corresponding RGB spectrogram images were stacked, and stored in a *.npy* file. Altogether, the dataset was organized in 6 folders, one for each grasp, with each containing 450 tensor samples of shape $[480 \times 640 \times 6]$. To be fed into the model, the dataset was randomly split into training, validation, and testing sets, with 80%-10%-10% split values, respectively. The architecture selected to approach this classification problem was a ResNet-18 pre-trained on the ImageNet dataset. This model, a CNN architecture containing 18 layers together with residual connections to improve its optimization, was modified to receive our input images – two stacked RGB spectrograms –, as well as to output the desired number of classes – one: the main grasping type. Afterwards, the model was trained 5 times using the cross-entropy loss function and the SGD plus momentum optimizer. Besides, the hyperparameters are listed in table I. The hyperparameter selection was performed based on a coarse grid search.

The results are presented in table II and the individual learning curves for one individual training process can be observed in figure 2.

### *B. Precise Grasp Generation*

The dataset used in this second part of our work is the DexGraspNet – Large-Scale Robotic Dexterous Grasp Dataset for General Objects Based on Simulation — [16]. This dataset comprises 5355 different 3D model objects and more than 1.3 million possible stable grasps for them. Each of these grasps is defined by $\{\theta, R, T\}$: 22 joint angles ($\theta$), 3 rotation Euler angles (R), and 3 translation coordinates (T). For this stage, a custom dataset for our sEMG grasps had to be created. To do so, the first step was to carefully select three reference grasps from the DexGraspNet database, corresponding to three of our classification grasping types: cylindrical, planar, and spherical. Although 6 different grasps were classified in the first step of the project, only 3 were used in this precise grasp generation step, in order to simplify the pipeline. The reference grasps selected among all the DexGraspNet dataset are shown in figure 3. Then, these reference grasps are used to extract valid and similar grasps from the dataset by means of comparing them to all the grasps in the original dataset, using the Euclidean distance of the parameters of the grasps. If the overall Euclidean distance fell below a specific threshold – that was manually set–, the grasp was included in our dataset with its corresponding label; otherwise, it was omitted. Following this procedure, our final dataset comprised 1000 randomly selected objects and 21000 valid spherical, cylindrical and planar grasps. Afterwards, the 3D parameters of our 1000 CAD objects needed to be extracted. To do so, the objects were embedded in a 50x50x50 volume, and the Euclidean distance map was calculated for the given object, with each

| | Average (N=5) | Reference |
|---|---|---|
| **Accuracy** | 95.5 | 99.5 |
| **Recall** | 95.5 | 99.6 |
| **Precision** | 94.9 | / |
| **F1-Score** | 94.7 | 99.6 |

TABLE II
EVALUATION METRICS.

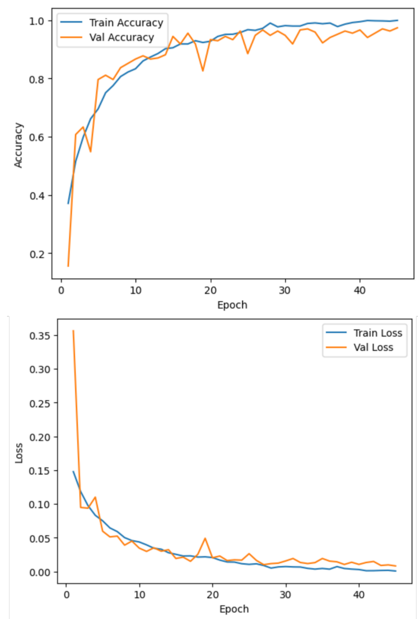


Fig. 2. Learning curves.

| Batch size | Epochs | Learning rate | Time steps |
|---|---|---|---|
| 64 | 200 | 1e−4 | 50 |

TABLE III
HYPERPARAMETERS USED TO TRAIN THE DIFFUSION MODEL.

voxel containing the distance to the closest point on the surface of the object. This Euclidean distance map is represented in figure 4. Therefore, the object was characterized and could be fed to the diffusion model. As a result, each entry of our custom dataset was composed of: a grasp (28 joint angles), an object representation (50x50x50 Euclidean distance map), and the grasp type (one-hot vector). Before feeding it to the model, all grasping data was normalized in the range of [-1,1], to ease the convergence of the model.

The precise grasping generation was implemented via deep generative diffusion modeling. First, an object embedding was obtained from the Euclidean distance map after applying a series of 3D convolutional layers to it. The size of this embedding was set to 512. Then, a positional embedding was applied to the 28-dimensional vector containing the joint angles of the grasp. The positional embedding size was 64, and the resulting joint angles embedding had a dimension of 1792 (64*28). The grasp type one hot vector was inputted into the model without any modification. Last, the time step was also embedded in a positional embedding of size 64. Finally, the diffusion model was trained by feeding the 21000 grasps custom dataset, in the form detailed above, and using the MSE loss function and the AdamW optimizers. Furthermore, the hyperparameters used are listed in table III.

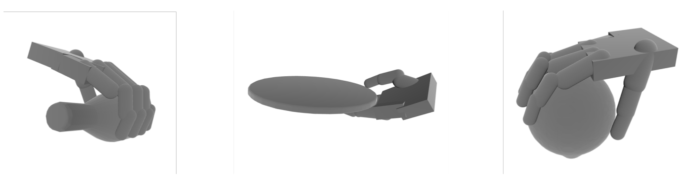

Fig. 3. Reference grasps selected: cylindrical, planar, spherical.

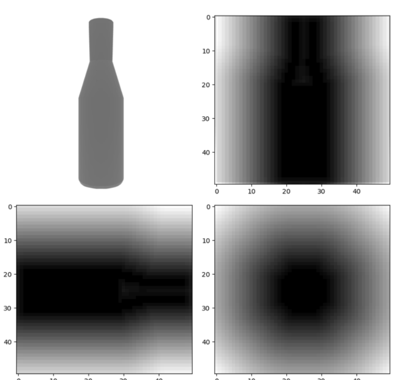

Fig. 4. Euclidean distance map of a bottle.

Overall, the diffusion generative model provided satisfactory results. The model proved to be capable of reproducing the joint angles of the polyarticular hand for cylindrical, grasping, and planar grasps, attending to the object shape and dimensions, and to the user's input. Nevertheless, the model didn't provide consistent results for the wrist rotation angles, and the hand position (translation) with respect to the object. Also, for every generated grasp, the diffusion model had to predict 28 joint angles, which was thought to be a too high-dimensional output for the grasp generation, and over the capacity of the selected diffusion model. As a result, to assess this hypothesis, two lower-dimensional models were trained. First, a model was trained to predict only the 6-dimensional wrist rotation and translation values. Whereas the generation of these rotation and translation values didn't perform well in the first model, this alternative model provided consistent and good results, thus confirming that a 28-dimensional output was to over dimensioned in the original model. Similarly, to reduce this over-dimensionality, a PCA encoding of the joint angles was applied. In such way, the 28-dimensional output was reduced to a 22-dimensional, and then uncompressed with the PCA inverse transform. Here as well, the PCA model provided better results, although these were not consistent at all, especially for the rotation and translation of the wrist. The results of the experiments conducted and explained in this section are shown in the generated graspings presented in figure 5.

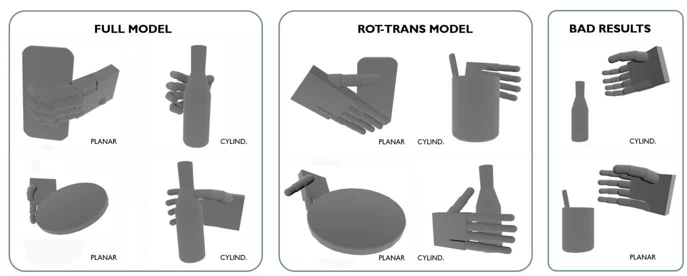


Fig. 5. Graspings generated with the diffusion model.

## III. DISCUSSION AND FUTURE WORK

In regard to the sEMG results, we can conclude that the classification results are very good since the accuracy achieved by the model is almost 96%. However, some modifications and improvements can be introduced in future work. On the one hand, there is still a 3% difference in testing accuracy between our work, and the reference paper [12]. Further hyperparameter tuning could be done in order to improve the performance of the model. Moreover, an extension of the current model — using a ResNet-50 — could be included, since that is the difference between the work presented in [4] and ours.Also, depending on the length of the signals considered, the spectrogram computation can become quite expensive. Since the ultimate objective of this project is to create a pipeline that can be used for online purposes, we would like to consider simpler features — or even feed the raw sEMG data to the model — to simplify the computation, and improve its performance. Further, we would like to extend the classification extent to more hand positions — in order to obtain a broader grasping spectrum —, providing a finer final selection for the user.

Regarding the precise grasp generation, we can conclude that the results accomplished with the initial scope of the project but can yet be strongly improved. Whereas the full model is completely able to generate precise grasps adapted to a given object and grasp type from scratch –especially regarding the joint angles–, it still struggles when trying to generate valid wrist rotations and translations, thus approaching the object for grasping. As mentioned in the results section, this shortcoming in the rotation and translation generation is due to an over-dimensionality of the output data to be generated by the diffusion model. Although different dimensionality reduction methods were tested, PCA did not accomplish satisfactory results, and this is a field yet to be explored. In the future, further methods to reduce the dimensions of the joint angles and thus improve the diffusion model prediction will be studied. Also, a diffusion model with higher capacities could be implemented, so it has higher capabilities to handle high-dimensional inputs and outputs with ease. Similarly, methodologies to include the reach-to-grasp process to the model will be explored. In such a way, considering the user position and approach to the object, enhanced precise graspings could be attained.

## References


[1] Marinelli, A., Boccardo, N., Tessari, F., Di Domenico, D., Caserta, G., Canepa, M., Gini, G., Barresi, G., Laffranchi, M., De Michieli, L. (2023). Active upper limb prostheses: a review on current state and upcoming breakthroughs. Progress in Biomedical Engineering, 5(1), 012001. doi:10.1088/2516-1091/acac57

[2] Salminger S, Stino H, Pichler LH, Gstoettner C, Sturma A, Mayer JA, Szivak M, Aszmann OC. Current rates of prosthetic usage in upper-limb amputees - have innovations had an impact on device acceptance? Disabil Rehabil. 2022 Jul;44(14):3708-3713. doi: 10.1080/09638288.2020.1866684. Epub 2020 Dec 30. PMID: 33377803.

[3] P. Tsinganos, B. Cornelis, J. Cornelis, B. Jansen and A. Skodras, ”Improved Gesture Recognition Based on sEMG Signals and TCN,” ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 2019, pp. 1169-1173, doi: 10.1109/ICASSP.2019.8683239.

[4] M. A. Ozdemir, D. H. Kisa, O. Guren, A. Onan and A. Akan, ”EMG based Hand Gesture Recognition using Deep Learning,” 2020 Medical Technologies Congress (TIPTEKNO), Antalya, Turkey, 2020, pp. 1-4, doi: 10.1109/TIPTEKNO50054.2020.9299264.

[5] Nisheena V. Iqbal, Kamalraj Subramaniam Shaniba Asmi P. (2018) A Review on Upper-Limb Myoelectric Prosthetic Control, IETE Journal of Research, 64:6, 740-752, DOI: 10.1080/03772063.2017.1381047

[6] Chen B, Chen C, Hu J, Nguyen T, Qi J, Yang B, Chen D, Alshahrani Y, Zhou Y, Tsai A, Frush T, Goitz H. A Real-Time EMG-Based Fixed-Bandwidth Frequency-Domain Embedded System for Robotic Hand. Front Neurorobot. 2022 Jun 30;16:880073. doi: 10.3389/fnbot.2022.880073. PMID: 35845759; PMCID: PMC9280080.

[7] Belić JJ, Faisal AA. Decoding of human hand actions to handle missing limbs in neuroprosthetics. Front Comput Neurosci. 2015 Feb 26;9:27. doi: 10.3389/fncom.2015.00027. PMID: 25767447; PMCID: PMC4341559.

[8] Kushwaha, V., Shukla, P. Nandi G.C. Generating quality grasp rectangle using Pix2Pix GAN for intelligent robot grasping. Machine Vision and Applications 34, 15 (2023). https://doi.org/10.1007/s00138-022-01362-2

[9] L. T. Taverne, M. Cognolato, T. Bützer, R. Gassert and O. Hilliges, ”Video-based Prediction of Hand-grasp Preshaping with Application to Prosthesis Control,” 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 2019, pp. 4975-4982, doi: 10.1109/ICRA.2019.8794175.

[10] D. Watkins-Valls, J. Varley and P. Allen, ”Multi-Modal Geometric Learning for Grasping and Manipulation,” 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 2019, pp. 7339-7345, doi: 10.1109/ICRA.2019.8794233.

[11] M. Romeo, J. Gual., ” EMG and Computer Vision-Based Deep Generative Modeling for Neuroprosthetic Grasping Patterns”, 2023, Proposal ADLR.

[12] Sapsanis, Christos. (2013). Recognition of basic hand movements using Electromyography.

[13] Miften FS, Diykh M, Abdulla S, Siuly S, Green JH, Deo RC. A new framework for classification of multi-category hand grasps using EMG signals. Artif Intell Med. 2021 Feb;112:102005. doi: 10.1016/j.artmed.2020.102005. Epub 2020 Dec 28. PMID: 33581825.

[14] D. Winkelbauer, B. Bäuml, M. Humt, N. Thuerey and R. Triebel, ”A Two-stage Learning Architecture that Generates High-Quality Grasps for a Multi-Fingered Hand,” 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Kyoto, Japan, 2022, pp. 4757-4764, doi: 10.1109/IROS47612.2022.9981133.

[15] Ozdemir MA, Kisa DH, Guren O, Akan A. Dataset for multi-channel surface electromyography (sEMG) signals of hand gestures. Data Brief. 2022 Feb 4;41:107921. doi: 10.1016/j.dib.2022.107921. PMID: 35198693; PMCID: PMC8844426.

[16] Wang, R., Zhang, J., Chen, J., Xu, Y., Li, P., Liu, T., Wang, H. ”DexGraspNet: A Large-Scale Robotic Dexterous Grasp Dataset for General Objects Based on Simulation,” 2022 arXiv preprint arXiv:2210.02697.

[17] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diusion probabilistic models. In Neural Information Processing Systems, 2020.

[18] Atzori, M., Gijsberts, A., Castellini, C. et al. Electromyography data for non-invasive naturally-controlled robotic hand prostheses. Sci Data 1, 140053 (2014). https://doi.org/10.1038/sdata.2014.53