Real-time finger force prediction via parallel convolutional neural networks: a preliminary study

Feng Xu, Yang Zheng, and Xiaogang Hu

Abstract— **Continuous and accurate decoding of intended motions is critical for human-machine interactions. Here, we developed a novel approach for real-time continuous prediction of forces in individual fingers using parallel convolutional neural networks (CNNs). We extracted populational motor unit discharge frequency using CNNs in a parallel structure without spike sorting. The CNN parameters were trained based on two features from high-density electromyogram (HD-EMG), namely temporal energy heatmaps and frequency spectrum maps. The populational motor unit discharge frequency was then used to continuously predict finger forces based on a linear regression model. The force prediction performance was compared with a motor unit decomposition method and the conventional EMG amplitude-based method. Our results showed that the correlation coefficient between the predicted and the recorded forces of the CNN approach was on average 0.91, compared with the offline decomposition method of 0.89, the online decomposition method of 0.82, and the EMG amplitude method of 0.81. Additionally, the CNN based approach showed generalizable performance, with CNN trained on one finger applicable to a different finger. The outcomes suggest that our CNN based algorithm can offer an accurate and efficient force decoding method for human-machine interactions.**

I. Introduction

The accurate estimation of individual finger forces is a crucial part of the precise control of a fully-functional prosthetic hand. It is also a non-trivial task in driving assistive [1] or rehabilitative devices [2]. Surface electromyogram (sEMG) is a promising neural interface [1] with wide applications. In order to continuously predict finger forces, one intuitive approach is to predict the forces proportionally relative to the EMG amplitude [3], as the muscle force is correlated with the EMG amplitude. However, interference to the EMG amplitude-based method, such as EMG amplitude drift over time due to fatigue [4], or electrode shifts [5], can deteriorate the control of robotic hands.

EMG signals are composed of a large number of motor unit action potentials (MUAPs), and the motor unit (MU) discharge frequency can be a reliable source to estimate the motor output. Earlier studies [6], [7], [8] have used a MU decomposition (i.e., spike sorting) of EMG signals to extract MU discharge activities. The populational MU discharge frequency can then be calculated to predict muscle forces or joint kinematics [9], [10], [11]. In this approach, the decomposed MU spike trains are merged to a single composite train to predict the forces. Thus, the initially obtained individual MU information, with a high computational load, is lost, indicating that the spike sorting method is inefficient.

In recent years, artificial neural networks have gained considerable interest in many research fields such as computer vision [12], [13] and natural language processing [14], largely because of the outstanding performance on mapping from the input data to categorical distribution over different labels [15], and the fast learning rate on feature representations from scratch. For example, recurrent neural networks (RNNs) [16] are promising on continuous prediction of motor output because the lateral propagation structure allows them to exhibit temporal dynamic behavior. The internal state (memory) can be used to process a sequence of inputs. However, Schluter et al. [17] has shown that convolutional neural networks (CNNs) could achieve better results because of the ability to capture the temporal-frequency relation in the two-dimensional (2-D) feature images and the frequency spectrum maps of the 1-D signals. Previously, Wei et al. [18] has proposed a CNN-based model using multiple features of EMG signals to classify a limited number of hand gestures. Kim et al. [19] has also proposed a CNN-based framework to classify finite states of movement based on EMG signals in a cross-subject manner. These earlier works suggest that the learned parameters of the CNNs potentially can be generalized across subjects.

Inspired by deep learning approaches with parallel CNN models [12], [18] that are capable of extracting multiple desired features from different aspects, we implemented a CNN-based model with a paralleled structure, in order to continuously predict individual finger forces in real-time using populational MU discharge frequency without spike sorting. The implemented model was used to first extract two types of EMG features: EMG amplitude map and frequency spectrum map. After CNN layers, the features were fused together, and further used to predict the MU discharge frequency. We compared the finger force decoding performance of our model with previous spike-sorting and conventional EMG methods.

II. Methods

A. Participants

Three neurologically intact participants (age: 22-31) were recruited to the study to preliminarily evaluate the performance of our method. All participants gave informed consent via protocols approved by the Institutional Review Board of the University of North Carolina at Chapel Hill.

*This work was supported in part by the National Science Foundation CBET-1847319

Feng Xu (email: fengxu@unc.edu), Yang Zheng (email: yang1127@email.unc.edu), and Xiaogang Hu (email: xiaogang@unc.edu) are with the Joint Department of Biomedical Engineering at University of North Carolina at Chapel Hill and NC State University (correspondence email: xiaogang@unc.edu)

978-1-7281-1990-8/20/$31.00 ©2020 IEEE

B. Data acquisition

The subjects were seated in front of a desk resting their forearms on a soft foam pad on the desk during the experiment. The elbow extended approximately 135°. Stiff foam pads were placed to hold the hand and to reduce hand movement and force transmission from the wrist. To record individual finger extension forces, all four fingers (index, middle, ring, and pinky) were respectively secured (Figure 1) to four miniature load cells (SM-200N, Interface), and the forces were sampled at 1000 Hz. An 8×20 high-density electromyogram (HD-EMG) electrode array was attached to each subject's forearm to obtain activities of the extensor digitorum communis (EDC) muscle. Each electrode was in 3 mm diameter and 10 mm spaced apart from each other. The HD-EMG recordings were sampled at 2048 Hz using EMG-USB2+ (OT Bioelettronica) with a gain of 1000 and band-passed at 10-900 Hz. The force data were upsampled to 2048 Hz linearly to match the dimension of the EMG data. During the data acquisition period, the subjects were requested to perform isometric finger extension and followed a preset pseudorandom target trajectory ranging from 0% to 40% MVC after the MVC has been measured for individual fingers.

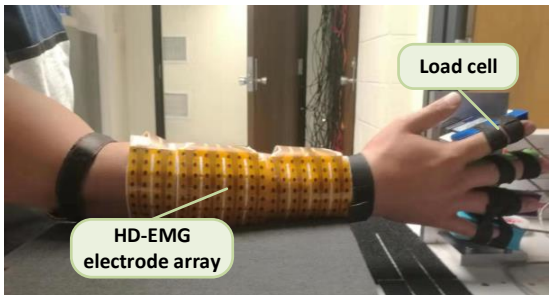

Fig. 1. Experimental setups for the data acquisition.

C. Data processing

Ground truth of composite firing frequency as the target output of the CNNs was calculated using the FastICA-based decomposition algorithm as mentioned in [4]. The output of the algorithm was the binary firing events of individual MUs. The composite firing frequency was then calculated by summing up the firing events of all the retained MUs.

We then calculated the two features within a sliding window, and rearranged the ground truth in the same time range as well. The length of the sliding window was 64 samples and the step length was 16 samples, both at 2048 Hz, as a shorter window length and step length can provide shorter delay for the real-time calculation. For example, a trial of 25 seconds will be separated into $n\_section = \frac{trial\_length-(win\_length-step\_length)}{step\_length}$ = 3197 sections.

First, the feature of frequency spectrum maps were calculated using Fourier transform (FT) including all the sampled EMG points within the window for each channel. As the spectrum was calculated channel-wise, there were 160 spectrum vectors within a single window.

Then, the feature of energy heatmaps were calculated based on the mean square (MS) value within the window of each channel. As a result, the dimension of the energy heatmap feature was 20×8 (height×width) for a single window.

D. Neural network model

Inspired by a multi-view neural network study [20] that input multiple features into separate CNN pathways, our implemented model (Figure 2) had four convolutional layers for each pathway before three fully connected layers for the output of the populational discharge rate.

After the raw HD-EMG signals were passed through the data processing module (shown as a yellow cycle with annotation 'DP'), the two features were fed into the CNNs. One feature was $\mathbf{F_{engy}} \in \mathbb{R}^{T\times H\times W}$ (Figure 2(A)), where T was the number of consecutive time frames, which was also the number of heatmaps. H and W represented the height and the width of each energy heatmap. The second feature was $\mathbf{F_{spec}} \in \mathbb{R}^{N\times M\times T}$, where $N = H \times W$ denoted the number of HD-EMG channels, T denoted the number of consecutive time frames, and M was the length of the frequency components after FT. A value of T larger than 1 denotes the number of consecutive windows that used for separate FT calculation.

Figure 2(B) illustrates the two pathways of CNNs, with kernel sizes indicated in the blocks. The upper pathway extracted high-level representations from the feature of energy heatmaps. The lower pathway extracted high-level representations from the feature of frequency spectrum maps. Each convolutional layer was followed by a max-pool module and activated by the ReLU function [21], and the Dropout (p = 0.5) was also adopted to reduce the possibility of overfitting [22]. Afterwards, the extracted representations were viewed as a feature vector and were fed to the fully connected layers (Figure 2). The number of neurons of the output layers was set to 8. Therefore, a higher discharge frequency (close to the class 7) represented a higher force level, while a lower discharge frequency (close to the class 0) indicated a lower force level.

The optimizer Adam set [23] was selected at an initial learning rate of 3e-4. The training, validation, and test epochs were 25-second of non-overlapped raw data randomly chosen from a 5-minute trial for the index, middle, and ring-pinky fingers, respectively. We also evaluated our model for the cross-finger conditions, where the 25-second training epochs were from a particular finger, but the validation and test epochs were from a different finger. The targets for training were derived from the spike trains calculated by the FastICA-based method, and summed 2-dimensionally (summed ones for all spike trains within a single window's timespan) for the two input features, leading to a 1-D target array in a chronological order. Each target in the array was divided by a constant and floored to the nearest integer to ensure that each target represents a class from 0 to 7. The constant was decided by the largest number in the target array which could be scaled to 7. The number of classes equaled to the number of neurons at the output layer of the model. An inadequate number of classes would reduce the accuracy in

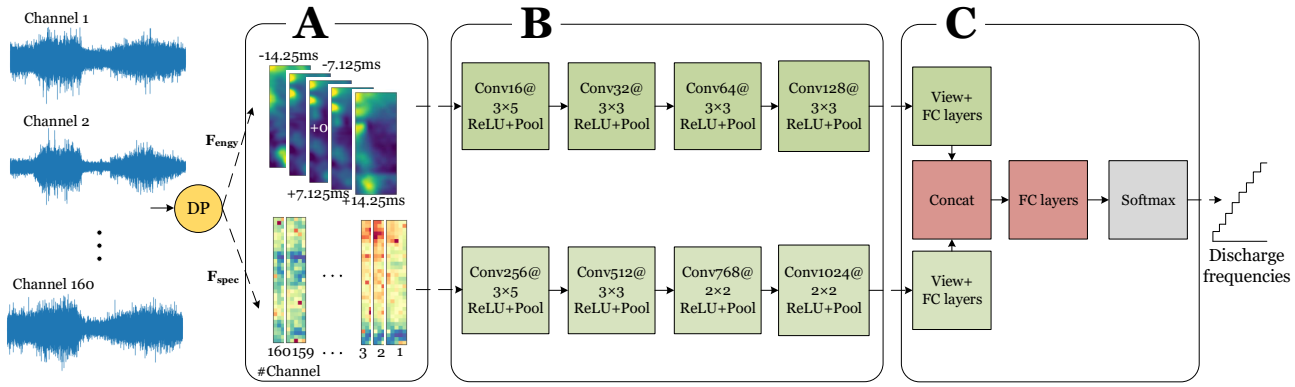

Fig. 2. Illustration of the proposed paralleled Convolutional Neural Networks model for finger force prediction. DP denotes the data processing module. The upper part of panel A illustrates the amplitude map feature from five consecutive windows. In the lower part of panel A, #Channel represents the index of EMG channels from which the frequency spectrum map is extracted. For each map, five time segments are used to calculate the spectrum, leading to a map of 5 × frequency bands. In panel B, each block represents a convolutional layer with the ReLU activation function and the Pooling layer. The number after 'Conv' denotes the number of kernels, and the numbers after '@' denotes the size of the kernels. In panel C, the blocks represent tensor viewing operation, fully-connected layers for each feature, tensor concatenation, fully-connected layers for all features, and the softmax function for calculating the probabilities for each class.

predicting discharge frequency, while an overly large number would exert difficulty for the neural network to fit to the target. During the training session, training data fed to the model were shuffled among the windows. The model was saved and evaluated using the validation epoch every 500 iterations. The optimal number of iterations was determined in the validation session, in which the model output (populational discharge frequency) achieved the highest correlation coefficient with the force trajectory. In both the validation and test sessions, the processed data from each period were fed to the model in a chronological order.

We low-pass filtered the model-predicted discharge frequencies to a 3^{rd} order Butterworth filter with a cutoff frequency of 10 Hz. Lastly, a linear regression model was used to predict the forces based on the smoothed discharge frequency. The regression model parameters were obtained in the validation session of the optimal number of iterations. The criteria used are the correlation coefficient between the measured force and the estimated force by the FastICA-based method, EMG amplitude-based method, and our developed methods, respectively. For the FastICA-based method, the firing event train of a single MU was obtained through spike sorting, in both offline [24] and online [4] manner. The calculation of the EMG amplitude-based method was performed as the RMS of the EMG signals. The RMS was averaged across all 160 channels to represent the overall EMG amplitude. We used a low-pass filtered with the same parameters aforementioned for both the EMG amplitude-based method and the FastICA-based method. Two linear regression models for the FastICA-based method and the EMG amplitude-based method were applied after initialization.

III. Results

An example force prediction is shown in Figure 3. The results were averaged across three subjects and all fingers for each subject. It showed that the CNN-based method (Correlation coefficient = 0.91, SD = 0.069) achieved similar performance compared with the FastICA-based method (Correlation coefficient = 0.89, SD = 0.052) (Figure 4). Both showed better performance than the EMG amplitude-based method (Correlation coefficient = 0.81, SD = 0.095). The results showed that the CNN-based method tended to show a slightly higher correlation coefficient than the FastICA-based method. It also showed a higher correlation coefficient than the real-time calculation of the FastICA-based method [4] (Correlation coefficient = 0.82, SD = 0.073). Specifically, the CNN-based method achieved higher correlation in the same finger conditions (Correlation coefficient = 0.93, SD = 0.045), compared with the cross-finger conditions (Correlation coefficient = 0.89, SD = 0.077). The delay for real-time decoding was 15.6 ms for the CNN-based method, because two windows were needed in advance to calculate the features.

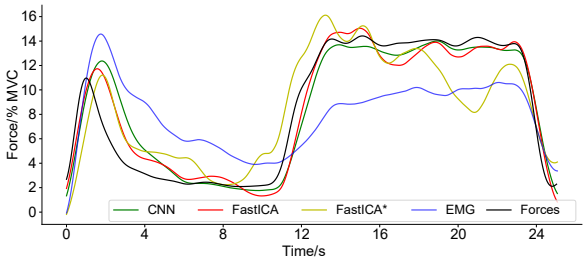

Fig. 3. Force estimation by CNN-based method (Real-time), FastICA-based method, and EMG amplitude-based method. *Online FastICA.

IV. Discussion

The current study developed a continuous and generalizable finger force prediction method based on parallel CNNs. As input signals to the CNN, frequency spectrum map and energy heatmap features were extracted from HD-EMG signals. We then used the CNN-based model to estimate populational MU discharge frequency without spike sorting. The obtained discharge frequency was used to continuously predict individual finger forces via a linear model.

Convolution with kernels is the key operation of the convolutional layers and our model. The CNN kernel is applied to a small portion of the entire map to explore the local pixels' relationships. Particularly, it explores the regional characteristics of the amplitude map feature, while also explores the correlations among frequency and time in

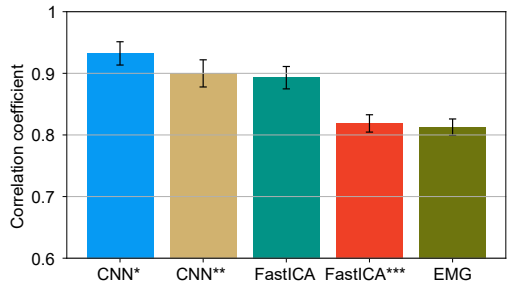

Fig. 4. The correlation coefficient of the three methods. *Same finger conditions. **Cross finger conditions. ***Online FastICA.

the frequency spectrum map. The explored correlation in the form of activation maps, after one or several convolutional layers, can be utilized by fully connected layers to predict the global discharge frequency. Although square kernels and square pooling are commonly used in computer vision rectangular shapes with adjustable width and length are more effective, when dealing with signal feature images extracted from short-time windows. Another advantage of CNN is that it can identify the profile of the activated area, and capture the spatiotemporal relation among channels and the global energy, which promoted the cross-finger force predictions. Additionally, a key setting of our method is that the populational MU discharge frequency is set as the training output of CNN, instead of directly training on forces. Our rationale is that the finger forces are positively correlated with the global MU discharge frequency. Training on the discharge frequency can enable the model to be generalizable to other tasks, such as the prediction of individual finger flexion forces or finger joint angles.

There are several limitations on the developed method. First, we observed a low prediction accuracy of the discharge frequency in the near-0 and near-7 classes, which may be caused by the similarity of feature patterns at such low or high force levels. Second, the performance of our model relies on the accuracy of the decomposition results of the FastICA-based method. During the initial training phase of the CNNs, the decomposition results were used as lost/cost functions. Inaccurate decomposition may lead to CNN parameters converge to regions that can lead to large errors in the estimation of discharge frequency, and subsequently lead to large force prediction errors.

In conclusion, we developed a novel approach for the real-time prediction of individual finger extension forces using a CNN-based model. The output of the model was the global MU discharge frequency. The finger forces were then predicted using a linear regression of the global discharge frequency. Our method outperforms the FastICA-based method in terms of generalizability across fingers, and is more accurate compared with the conventional EMG amplitude-based method.

REFERENCES

[1] P. K. Artemiadis and K. J. Kyriakopoulos, "EMG-based position and force estimates in coupled human-robot systems: Towards EMG-controlled exoskeletons," in *Exp. Robot.* Springer, 2009, pp. 241–250.

[2] C. S. Klein, S. Li, X. Hu, and X. Li, "Editorial: Electromyography (EMG) Techniques for the Assessment and Rehabilitation of Motor Impairment Following Stroke," *Front. Neurol.*, vol. 9, p. 1122, 2018.

[3] A. Fougner, Ø. Stavdahl, P. J. Kyberd, Y. G. Losier, and P. A. Parker, "Control of upper limb prostheses: Terminology and proportional myoelectric control—A review," *IEEE Trans. neural Syst. Rehabil. Eng.*, vol. 20, no. 5, pp. 663–677, 2012.

[4] Y. Zheng and X. Hu, "Real-time isometric finger extension force estimation based on motor unit discharge information," *J. Neural Eng.*, vol. 16, no. 6, p. 066006, 2019.

[5] I. Kyranou, S. Vijayakumar, and M. S. Erden, "Causes of performance degradation in non-invasive electromyographic pattern recognition in upper limb prostheses," *Front. Neurorobot.*, vol. 12, p. 58, 2018.

[6] M. Chen and P. Zhou, "A novel framework based on FastICA for high density surface EMG decomposition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 1, pp. 117–127, 2015.

[7] A. Holobar and D. Zazula, "Multichannel blind source separation using convolution kernel compensation," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4487–4496, 2007.

[8] C. Dai and X. Hu, "Independent component analysis based algorithms for high-density electromyogram decomposition: Systematic evaluation through simulation," *Comput. Biol. Med.*, vol. 109, pp. 171–181, 2019.

[9] C. Dai and X. Hu, "Finger joint angle estimation based on motoneuron discharge activities," *IEEE J. Biomed. Heal. Informatics*, vol. 24, no. 3, pp. 760–767, 2020.

[10] C. Dai, Y. Cao, and X. Hu, "Prediction of Individual Finger Forces Based on Decoded Motoneuron Activities," *Ann. Biomed. Eng.*, vol. 47, no. 6, pp. 1357–1368, 2019.

[11] D. Farina, I. Vujaklija, M. Sartori, T. Kapelner, F. Negro, N. Jiang, K. Bergmeister, A. Andalib, J. Principe, and O. C. Aszmann, "Man/machine interface based on the discharge timings of spinal motor neurons after targeted muscle reinnervation," *Nat. Biomed. Eng.*, vol. 1, no. 2, p. 25, 2017.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, 2016, pp. 770–778.

[14] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *Adv. NIPS*, pp. 3104–3112, 2014.

[15] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[16] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[17] J. Schlüter and S. Böck, "Improved musical onset detection with convolutional neural networks," in *2014 ieee Int. Conf. Acoust. speech signal Process.* IEEE, 2014, pp. 6979–6983.

[18] 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 Recognit. Lett.*, vol. 119, pp. 131–138, 2019.

[19] K.-T. Kim, C. Guan, and S.-W. Lee, "A Subject-Transfer Framework based on Single-Trial EMG Analysis using Convolutional Neural Networks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, pp. 94–103, 2019.

[20] L. Ge, H. Liang, J. Yuan, and D. Thalmann, "Robust 3d hand pose estimation in single depth images: from single-view cnn to multi-view cnns," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, 2016, pp. 3593–3601.

[21] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 807–814.

[22] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv Prepr. arXiv1412.6980*, 2014.

[24] C. Dai and X. Hu, "Independent component analysis based algorithms for high-density electromyogram decomposition: Experimental evaluation of upper extremity muscles," *Comput. Biol. Med.*, vol. 108, pp. 42–48, 2019.