SPATIO-TEMPORAL CONVOLUTIONAL NETWORKS FOR MYOELECTRIC CONTROL

Milad Jabbari and Kianoush Nazarpour

School of Informatics, University of Edinburgh, UK

ABSTRACT

Utilising both within-channel temporal and between-channel spatial dependencies of the surface Electromyographic (sEMG) signals improves the accuracy of machine learning-based models of myoelectric control. Here, we introduce the Spatio-Temporal Convolutional Network (STCN) to decode five hand gestures from eight EMG signals recorded from the forearm of eight able-bodied subjects. We compared our proposed STCN model with a combination of a conventional convolutional neural network (CNN) and a Long Short-Term Memory (LSTM) deep learning model, as well as the Linear Discriminant Analysis (LDA). The results show that STCN model can outperform both CNN-LSTM and LDA methods at a much lower computational complexity.

INTRODUCTION

For decades, conventional machine learning models have served as the gold-standard models in EMG -based hand gesture classification [1]. Despite providing reliable results, they suffer from a lack of generalisation and robustness. Furthermore, they depend extremely on the distribution and separability of the feature set. Therefore, to achieve higher performance, they need high-quality hand-crafted and separable features. Extracting such features is not trivial and may lead to complex computations, posing difficulties in translating the system into a real-time experience [2].

Deep learning models possess the ability to extract insightful features from raw data, thereby obviating the necessity for manual hand-crafted feature extraction. Moreover, through hierarchical representation, non-linear decision boundaries, and regularization techniques, they enhance the robustness of performance [3]. These inherent advantages position such models as exceptional candidates for mapping raw EMG signals or basic low-level features to a notably enriched informative space with maximal separability.

In EMG-based hand gesture classification system, extracting both temporal and spatial features can enhance the performance in terms of accuracy and generalisation. For instance, in hybrid CNN-LSTMs models, CNN and LSTM extract spatial and temporal dependencies, respectively [4-5]. However, the substantial computational cost of this approach renders hardware implementation challenging. Recently, Temporal Convolutional Networks (TCN) have proved effective in sequence modelling. By implementing the concept of dilation, TCNs demonstrates their superiority in capturing temporal dependency, while maintaining significantly lower computational cost compared to other deep learning models [6].

In this work, we introduce the Spatio-Temporal Convolutional Network (STCN) model to enable simultaneous extraction of temporal and spatial components. We compare this network with the CNN-LSTM structure as well as the LDA method, which serves a benchmark.

METHODS

Participants and EMG Recording

A schematic experimental setup from the experiment is shown in Figure 1. The experimental protocol was in accordance with the ethical approval granted by the local committee at the University of Edinburgh (reference number: 2019/89177). Eight participants took part in the experiment after signing an informed consent. During the experiment, participants performed five hand gestures including Open, Power, Pointer, Tripod, and Rest. The EMG signals were recorded using eight OY Motion EMG electrodes at a sampling frequency of 1000 Hz. Before the start of the experiment, electrodes were placed on the participants' arm and secured using an armband. During the experiment, participants sat in front of a screen which instructed the gestures. We recorded fifteen examples for each gesture, ech 2 seconds long.

Utilised Models

In this study, to conduct spatial convolution, we aimed to use a *depthwise* convolution layer to extract spatial features from the channels. Following a spatial block, a temporal block containing 4 dilated temporal convolutional layers captured the temporal dependencies. Figure 2(A) presents the proposed STCN concept. A simplified block diagram representation of STCN model is presented in Figure 2(B). In addition to the proposed STCN, we deployed a hybrid CNN-LSTM model and a conventional LDA classifier to perform 5-movement classification. For both models, spatial part consists of two 1-Dimensional CNN layer followed by a batch normalization and a dropout layer. Utilized CNN-LSTM model benefits from an LSTM layer consists of 100 units, whereas TCN block in STCN is a single-stack basic TCN.

Feature set and statistical analysis

We fed these three models with six features called spatio-temporal feature set (STFS), which were in line with [7]. This feature set consists of integral square descriptor, normalised root square coefficient of first and second differential derivatives, mean log-kernel, an estimate of mean derivative of the higher-order moments, and a measure of spatial muscle information. To evaluate statistical significance of the obtained findings, Wilcoxon rank signed test was employed. The Bonferroni correction method was used to adjust the *p*-values and re-balance the compounding risks.

RESULTS

To assess the proposed STCN model and provide comparison with CNN-LSTM and LDA methods, we calculated classification accuracy using:

$$\text{CA} = \frac{correctly\ classified\ samples}{total\ classified\ samples} \times 100\% \tag{1}$$

Furthermore, averaged confusion matrixes over all participants were calculated to demonstrate between class variation of the performance. Figure 3(A) illustrates boxplot presentation of CA values for STCN, CNN-LSTM, and LDA methods. The results show that using STCN can significantly outperform both CNN-LSTM and LDA with $p<0.05$ and $p<0.001$ and average CA about 95.21±0.01%. Moreover, averaged confusion matrixes over all participants for LDA and STCN is shown in figure 3(B) and figure 3(C), respectively. It is visible that performance of STCN in terms of between classification variation is better than LDA.

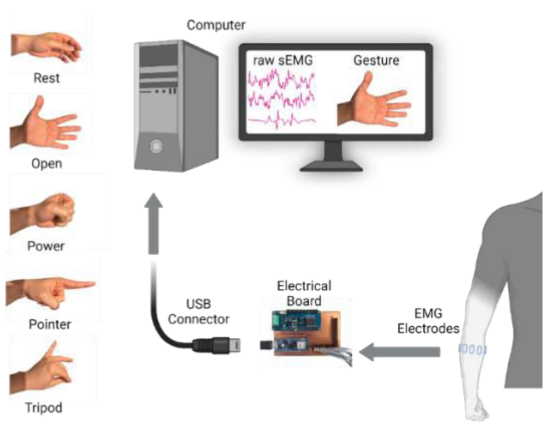

Figure 1: Schematic experimental setup. Participant generates the illustrated gestures including Rest, Open, Power, Pointer, and Tripod, while is sitting on a chair and with a fixed position arm.

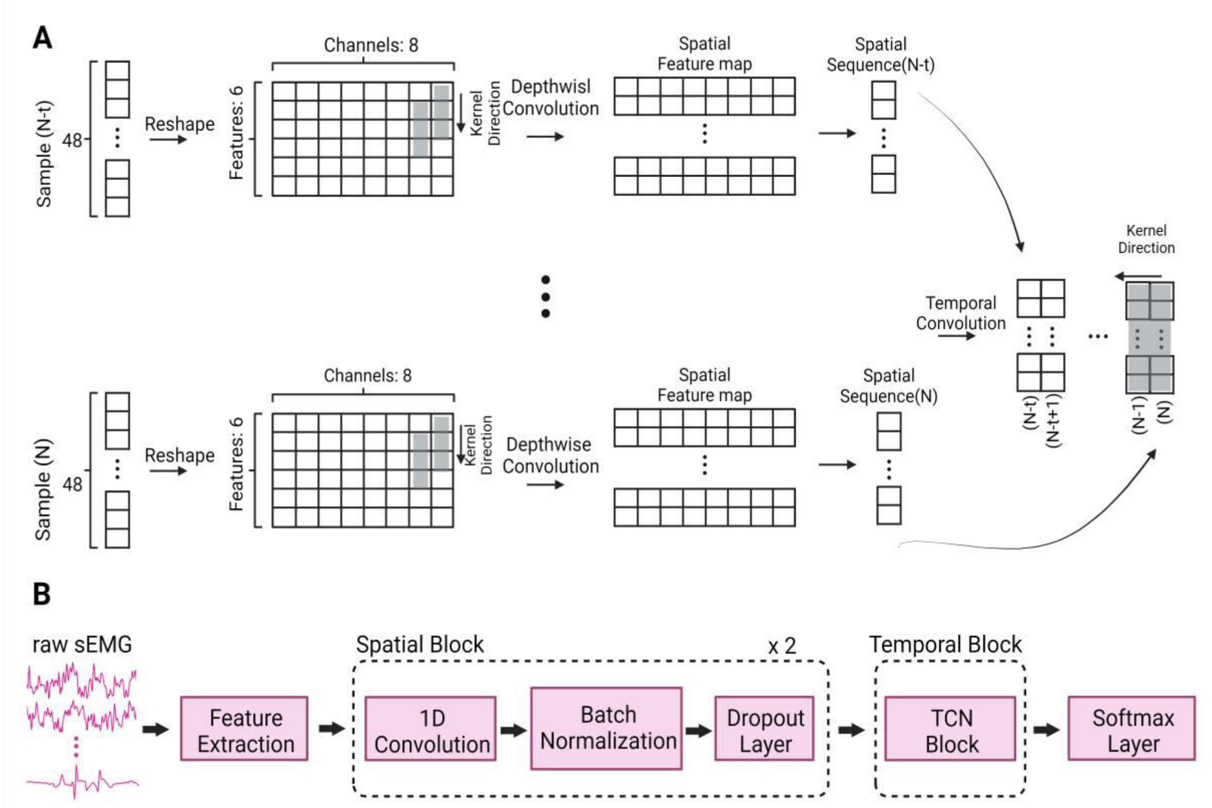

Figure 2: Schematic representation of the utilized Spatio-Temporal approach and overall block diagram of the STCN model shown in A and B.

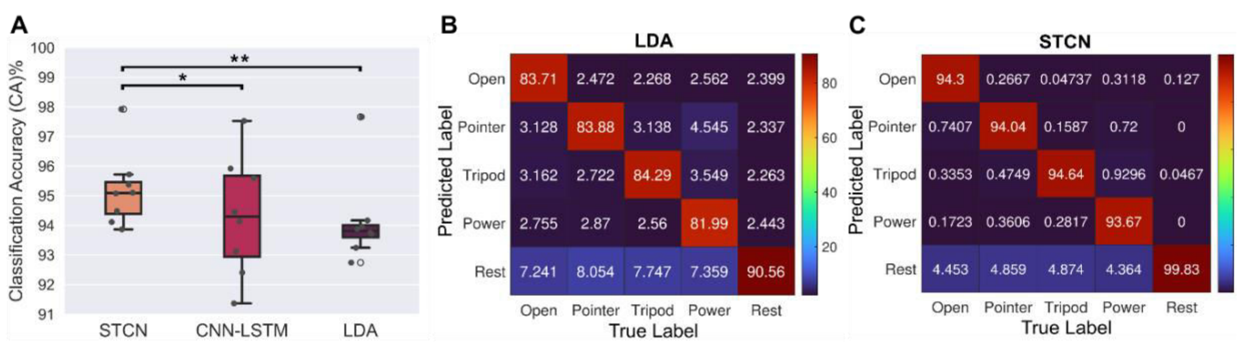

Figure 3: Boxplot representing of averaged CA values over all subjects for STCN, CNN-LSTM, and LDA and averaged confusion matrices of LDA and STCN are illustrated in A, B, and C, respectively.

DISCUSSION

This study presents a novel spatio-temporal model designed to capture both the spatial and temporal dependencies of sEMG signals. Inspired by the convolution concept and moving the kernel in spatial and temporal directions in separated blocks, we show that a reliable performance can be achieved. Deploying a spatial convolutional layer by using CNNs reveals the spatial flow of information and connectivity between muscles signals. These spatially distributed components of information can be fed into a TCN block after appropriately reshaping of timesteps and result in higher accuracy compared to conventional ML models. Previous studies have demonstrated the excellence of hybrid CNN-LSTM models in conducting EMG based hand gesture classification, however, in this study we show that using dilated convolution may lead to better performance compared to cell state concept in LSTM.

Our findings demonstrate that employing a deep learning-based model with low-level representation of sEMG signal as feature may lead to reliable classification. An important consideration in the STCN model is the choice of kernel shape and size. In this study we utilised 1D kernels. An alternative approach can involve switching to 2D kernels of different size. Considering how electrodes are distributed in distal/proximal and lateral directions, selecting the kernel type and size will also be a crucial parameter.

ACKNOWLEDGEMENTS

This work is supported by funding from Engineering and Physical Sciences Research Council (EPSRC) under grant (EP/R004242/2).

REFERENCES

[1] E. Scheme, and K. Englehart, "Electromyogram pattern recognition for control of powered upper-limb prostheses: state of the art and challenges for clinical use," *J. Rehabil. Res. Dev.*, vol. 46, no. 6, 2011.

[2] M. Jabbari, R. N. Khushaba, and K. Nazarpour, "Spatio-temporal warping for myoelectric control: an offline, feasibility study," *J. Neural Eng.*, vol. 18, no. 6, p. 066028, 2021.

[3] J. Ye, N. Jiangqun, and Y. Yang, "Deep learning hierarchical representations for image analysis," *IEEE Trans. Inf. Forensics Secur.*, vol. 12, no. 11, pp. 2545-2557, 2017.

[4] Y. Hu, *et al.*, "A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition," *PLoS One.*, vol. 13, no. 10, p. e0206049, 2018.

[5] T. Bao, *et al.*, "A CNN-LSTM hybrid model for wrist kinematics estimation using surface electromyography," *IEEE Trans. Instrum. Meas..*, vol. 70, pp. 1-9, 2020.

[6] U. Cote-Allard, *et al.*, "A transferable adaptive domain adversarial neural network for virtual reality augmentation EMG-based gesture recognition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 546-555, 2021.

[7] S. O. Williams, *et al.*, "Spatio-temporal based descriptor for limb movement-intent characterization in EMG-pattern recognition system," in *proceeding IEEE EMBC*, pp. 2637-2640, 2019.