Concordia University
Engineering and Computer Science

SURFACE EMG-BASED HAND GESTURE RECOGNITION VIA DILATED CONVOLUTIONAL NEURAL NETWORKS

Elahe Rahimian[†], Soheil Zabihi[‡], Seyed Farokh Atashzar[††], Amir Asif[‡], and Arash Mohammadi[†]

[†]Concordia Institute for Information Systems Engineering, Concordia University, Montreal, QC, Canada
[‡]Department of Electrical and Computer Engineering, Concordia University, Montreal, QC, Canada
[††] Department of Electrical and Computer Engineering and Department of Mechanical and Aerospace Engineering, New York University, USA

Email: {e_ahimia}@encs.concordia.ca

Outline

Introduction

Motivations and Contributions

Motivations

- This the paper proposes a novel deep learning-based architecture for processing surface Electromyography (sEMG) signals to classify and recognize upper-limb hand gestures via incorporation of dilated causal convolutions.
- The motivation of this paper is proposing a promising approach for hand gesture recognition to facilitate the life of hand-amputated individuals.
- The proposed approach has the potential to significantly improve the overall recognition accuracy due to the specific design of the convolutional layers.

Motivations and Contributions

Contributions

- Using dilated causal convolutions enables us to gradually increases the receptive field of the network.
- By applying Conv1D, the proposed architecture eliminates the need for readjustment of the input sequences
- Takes into account the hidden temporal correlations existing among the available set of sEMG sequences, inherently.
- The proposed architecture can provide several advantages over RNNs such as lower memory requirement and faster training.
- In the proposed architecture, instead of exhaustively concatenating all permutations of input sequences, we use the input sEMG sequences in temporal depth and then apply 1-dimensional convolutions.

The Proposed Hand Gesture Recognition Architecture

Database

- The second Ninapro database referred to as the DB2 is used in this work, which is a publicly available dataset for hand gesture recognition tasks.
- Delsys Trigno Wireless EMG System with 12 wireless electrodes (channels) is used in the NinaProject to collect electrical activities of muscles at a rate of 2 kHz.
- The DB2 consists of 50 different gestures including wrist, hand, grasping, and functional movements together with force patterns from 40 healthy (intact) subjects.
- More specifically, the dataset consists of signals collected from 28 males and 12 females with age 29.9 ± 3.9 years among which 34 are right-handed and 6 left-handed.
- The DB2 dataset is presented in three sets of exercises denoted as Exercise B, C, and D.
- The subjects repeated each movement for 6 times, each time lasted for 5 seconds followed by 3 seconds of rest.

Database (continued)

- In this paper, Exercise B is utilized for developments, which includes 17 movements, i.e., 8 isometric and isotonic hand configurations, and 9 basic movements of the wrist.
- For the sake of comparison and following the recommendation provided by the dataset, Repetitions 1, 3, 4, and 6 are considered for constructing the training set, and repetitions 2 and 5 are used for testing the proposed architecture.
- It is noteworthy to mention that the DB2 database has a second set of refined labels, where again the movements are repeated by the subjects. In this case, labels associated with the duration of the movement are refined in order to represent the real movement.
- The refined data is donated as *Posterior Data*, and the data, which is not refined, is referred to as the *Prior Data*.

Database (continued)

Figure: Hand and wrist movements (Exercise B)

Pre-processing Step

- The data derived from DB2 (Exercise B) was pre-processed using a 4^{th} order Butterworth low-pass filter with a 20Hz cutoff frequency
- Then the data was normalized based on the Z-score approach.
- To compare the proposed method with previous literature and also to satisfy the acceptable delay time (which should be under 300ms, the sEMG data for each channel is segmented by a window with a length of 100ms (200 samples per window).
- The sliding window with steps of 10ms is considered for this segmentation of the input data.

The schematic of the proposed architecture

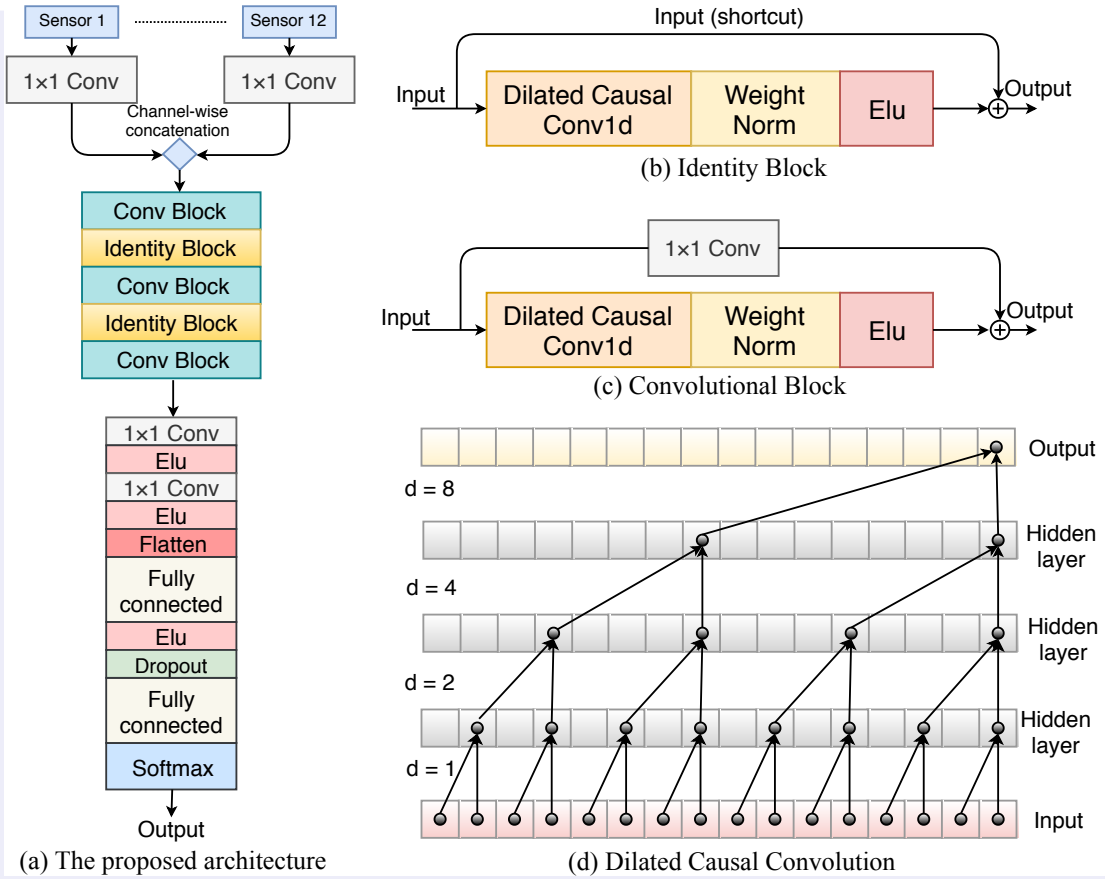


Figure: (a) The schematic of the proposed architecture. (b) The identity block. (c) The Convolutional block. (d) The dilated causal convolution (dilation factor d is equal to $[1, 2, 4, 8]$ for the layers). The receptive field of the last neuron in the output layer is shown in "pink".

The Proposed Architecture

Increasing The Number of Input Features

- sEMG signals from the recording sensors (12 sensors used in the collection of the DB2 dataset) are provided, separately, as inputs to a (1×1) convolution layer with 10 kernels.
- The features are concatenated channel-wise, which are then fed to the first block of the architecture.
- This input strategy will result in increasing the number of input features, i.e., with signals from 12 sensors and using 10 kernels we end up with 120 input features.
- In other words, this approach allows us to train 120 different features on the first layer of the network.

Main Building Blocks of the Proposed Architecture

- The Identity Block
- The Convolutional Block

The Proposed Architecture

The Identity Block

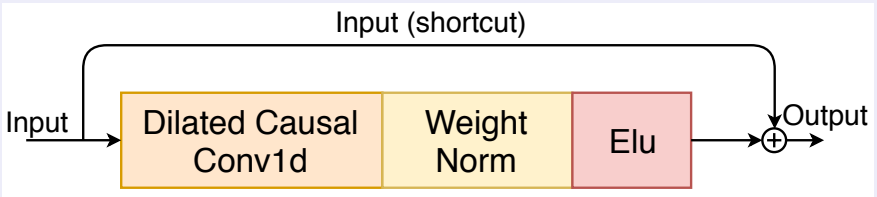


Figure: The identity block

- *Shortcut (Skip Connection)*: Inspiring from ResNets, we address the degradation problem by adding a shortcut or a skip connection to the identity block.
- *Weight Normalization*: Weight normalization beside normalizing the input signals are utilized within the Identity Block.
- *Activation Function (ELU)*: Exponential Linear Unit (ELU)is used as the nonlinear activation function and is the last component within this block.

The Proposed Architecture

The Identity Block (continued)

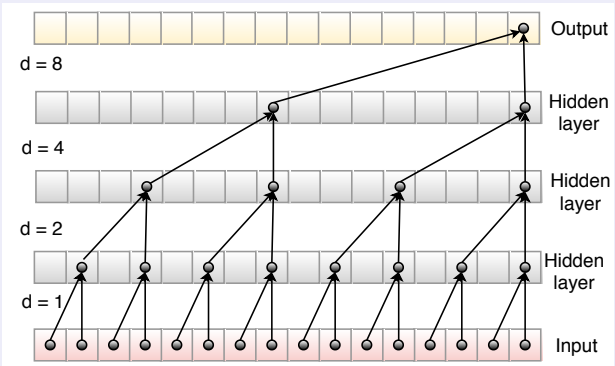


Figure: Dilated Causal Convolution

- *Dilated Causal Convolutions*: The advantage of incorporating the causal convolutions is that they do not have recurrent connections, therefore, the required training time is substantially less than that of RNNs.
- Despite having these benefit, however, the receptive field of causal convolutions is small.
- We used dilated causal convolutions within the Identity block of the proposed architecture to address the issue related to the limited receptive field of casual convolutions.

The Proposed Architecture

The Convolutional Block

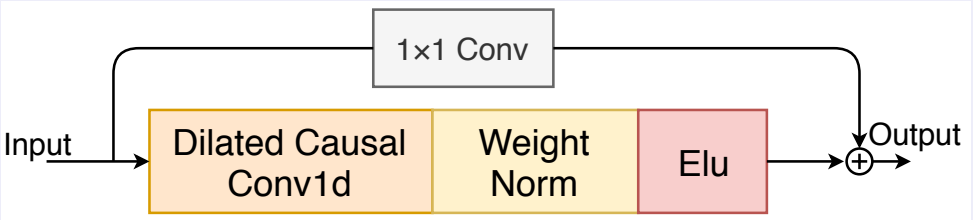


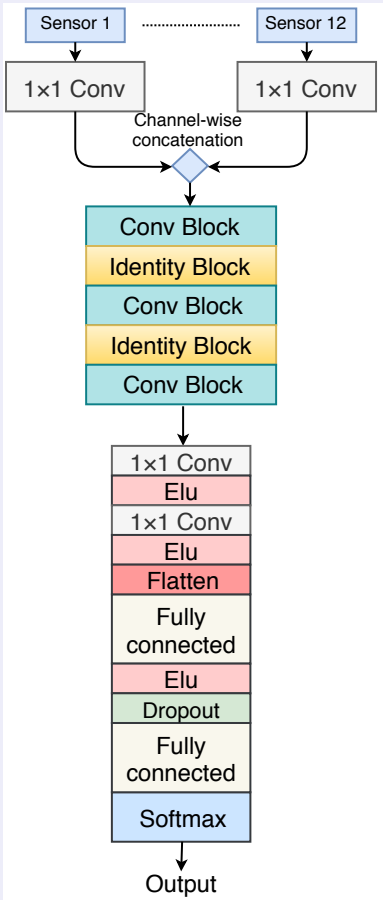
Figure: Convolutional Block

- The convolutional block is employed when the input and output dimensions do not match up.
- This happens as the number of filters used within the dilated causal component can vary from one block to another, which in turn results in a miss-match between the dimension of the input sequence and that of the output sequence.
- To deal with such scenarios, the Convolutional block is introduced.
- The difference between this block and the identity block is that there is a (1×1) convolution layer in the shortcut path, which resizes the input to a different dimension to match up with the output sequence.

The Proposed Architecture

Network Structure and Hyper-parameters Settings

The proposed architecture has 3 Convolutional Blocks and 2 Identity Blocks together 2 (1×1) Convolutions layers and 2 fully connected layers. The details of the proposed architecture are provided below:



(a) The proposed architecture

The Proposed Architecture

Network Structure and Hyper-parameters Settings

- The first Convolutional Block has 64 kernels with size 5. The dilation factor for dilated causal convolutions is set equal to 1.
- Similarly, the first Identity Block of the architecture has 64 kernels with size 5. On contrary to the previous convolutional block, here the factor for the dilated causal convolutions is set equal to 2, therefore, the output neurons in this block have extended the receptive field.
- The second Convolutional Block has 128 kernels with size 5 and increased dilation factor of 4.
- The second Identity Block has 128 kernels with size 5 with now dilation factor of 8.
- The third Convolutional Block has 256 kernels with size 5 and dilation factor of 16. The receptive fields of the blocks are increased as we move deeper into the network.

The Proposed Architecture

Network Structure and Hyper-parameters Settings (continued)

- The outputs from the third Convolutional Block, which has 256 channels, are fed to (1×1) Convolutions with 64 kernels, and then ELU activation function is applied on the output.
- There are another (1×1) Convolutions followed by ELU activation, which reduces the size of the input channels to 2.
- The first fully connected layer reduces the size of its input features to 150.
- The second fully connected reduces further reduces the size of the features, i.e., from 150 features to 17, which is equal to the number of classes.
- Adam optimizer as the optimizer algorithm with a learning rate set 0.001.
- The learning rate changes in a cycle with a length of 100 epochs. After 20 epochs, we divided the learning rate by 2, but after 100 epochs instead of dividing the learning rate by 2, we multiply it by 14.4. Therefore, the learning rate at the beginning of each cycle will be 90% of the learning rate at the beginning of the previous cycle.

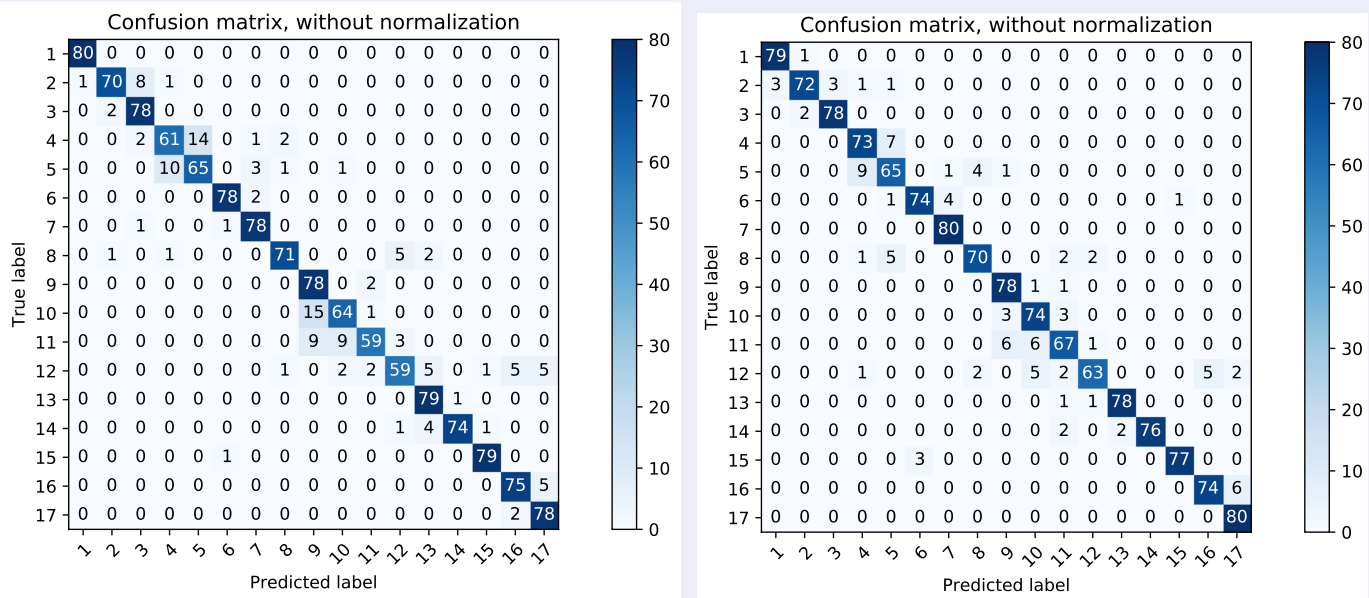
Experiments and Results

The Results Comparison

Table: The Accuracy between our model and Reference [5].

	Data	Method	Accuracy (%)
Ours	Prior data	Proposed model	90.14
	Posterior data	Proposed model	**92.5**
[5]	Prior data	–	–
	Posterior data	C-B1B2	**83.79**
		CC	77.96
		C-2B1	82.16
		C-2B2	82.49
		C-DK	79.23
		C-SK	78.52
		C-SK2	81.3

The Confusion Matrix



(a) (b)

Figure: (a) Confusion matrix of the model for prior data. (b) Confusion matrix of the model for posterior data.

Analysis of confusion matrix

Table: Analysis of confusion matrix for *Posterior Data*.

Class	Precision	Recall	F1-score
Class 1	0.96	0.99	0.98
Class 2	0.96	0.9	0.93
Class 3	0.96	0.97	0.97
Class 4	0.86	0.91	0.88
Class 5	0.82	0.81	0.81
Class 6	0.96	0.93	0.94
Class 7	0.94	1	0.97
Class 8	0.92	0.88	0.89
Class 9	0.89	0.97	0.92
Class 10	0.86	0.93	0.89
Class 11	0.86	0.84	0.84
Class 12	0.94	0.79	0.85
Class 13	0.97	0.97	0.97
Class 14	1	0.95	0.97
Class 15	0.99	0.96	0.97
Class 16	0.94	0.93	0.92
Class 17	0.91	1	0.95

Conclusion

Unique Characteristics of the Proposed Architecture

- The surface electromyography (sEMG)-based gesture recognition via deep learning solutions is considered as a promising approach to facilitate the life of hand-amputated individuals.
- The paper proposed a novel deep architecture based on dilated causal convolutions, which increases the *receptive field* or *field of view* of the network.
- The training step of the proposed architecture is considerably faster than that of RNNs due to the absence of recurrent connections.
- Different from the existing deep learning methods for hand gesture recognition, in this paper, by applying Conv1d, the network itself is able to extract the hidden correlations existing between different signal sequences.
- It is observed that the proposed architecture can produce promising results with high accuracy, which is 8.71% greater than the state-of-the-art results.

Thanks for Your Attention