Optimizing Gesture Recognition Using a 3D Convolutional Neural Network for High-Density Surface EMG

Vasupradha Ramji, Dr. Yue Wen

University of Central Florida, College of Engineering and Computer Science

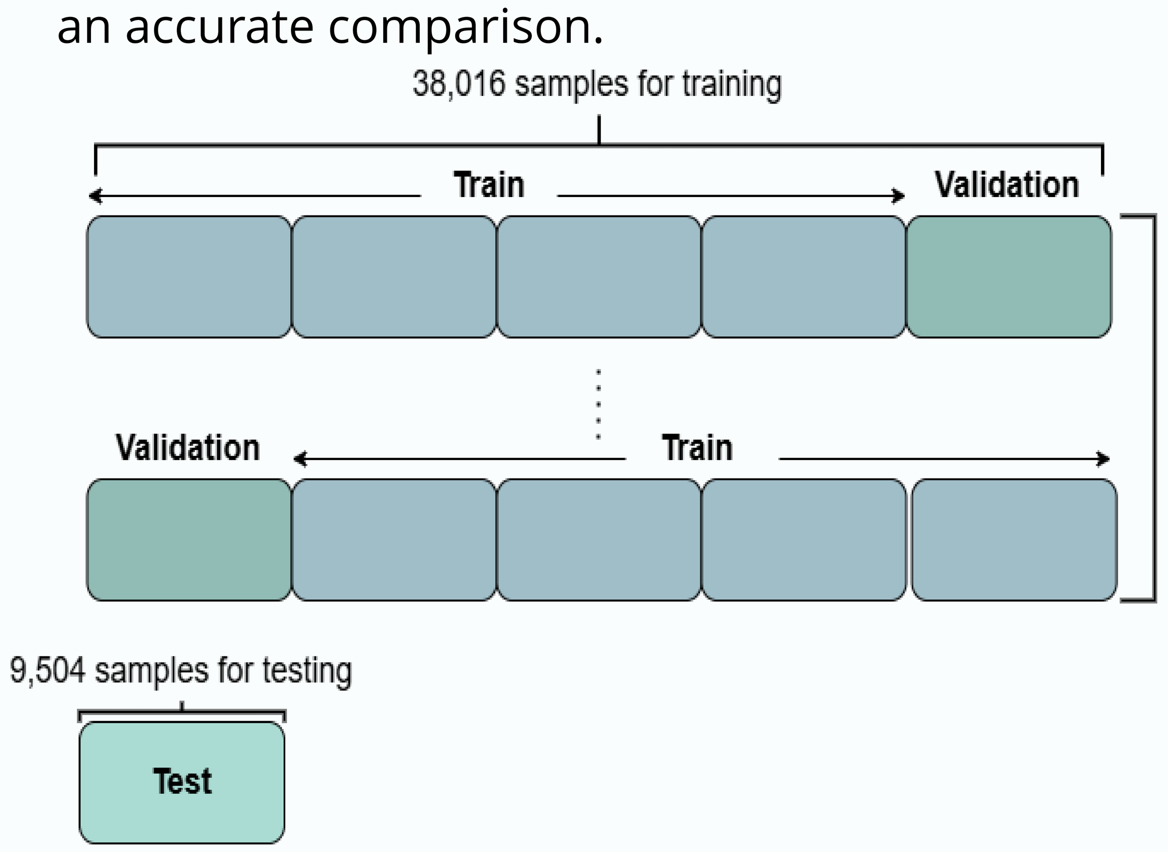
Abstract

The processing of high-density surface electromyography (HD-sEMG) signals through deep learning has significantly propelled the accuracy and responsiveness in gesture recognition. In previous studies, 3D convolutional neural networks (3D CNNs) have shown superior recognition accuracy compared to 2D CNNs. However, this improvement comes at the cost of higher computational costs and longer inference latencies. For real-time applications such as prosthetic control systems, minimizing inference latency is crucial. This study aimed to optimize a 3D CNN architecture to reduce inference latency without compromising accuracy.

Methodology

The open-sourced CapgMyo-Dba database, which is composed of HD-sEMG signals for 8 distinct hand gestures obtained from 18 healthy subjects, was incorporated as data for this research. The datasets were preprocessed by segmenting the HD-sEMG signals into smaller time windows to allow for spatiotemporal analysis in the 3D CNN.

In specific, a series of datasets across six subjects were then loaded and split into 80% training and validation through K-fold sampling and 20% testing. Due to the initial model performing the best with subjects 2, 6, 9, 12, 17, and 18 during training, it was decided that these subjects would be used to train both model variations to allow for an accurate comparison.



3D CNN Model Analysis

After careful experimentation, it was determined that the model learned best and converged efficiently with:

- An initial learning rate of 0.01.
- A batch size of 90.
- A patience of 10 for early stopping with a threshold of 0.01.

Moreover, experimentation with optimizing the network structure for lower inference latency was most efficient with a combination of:

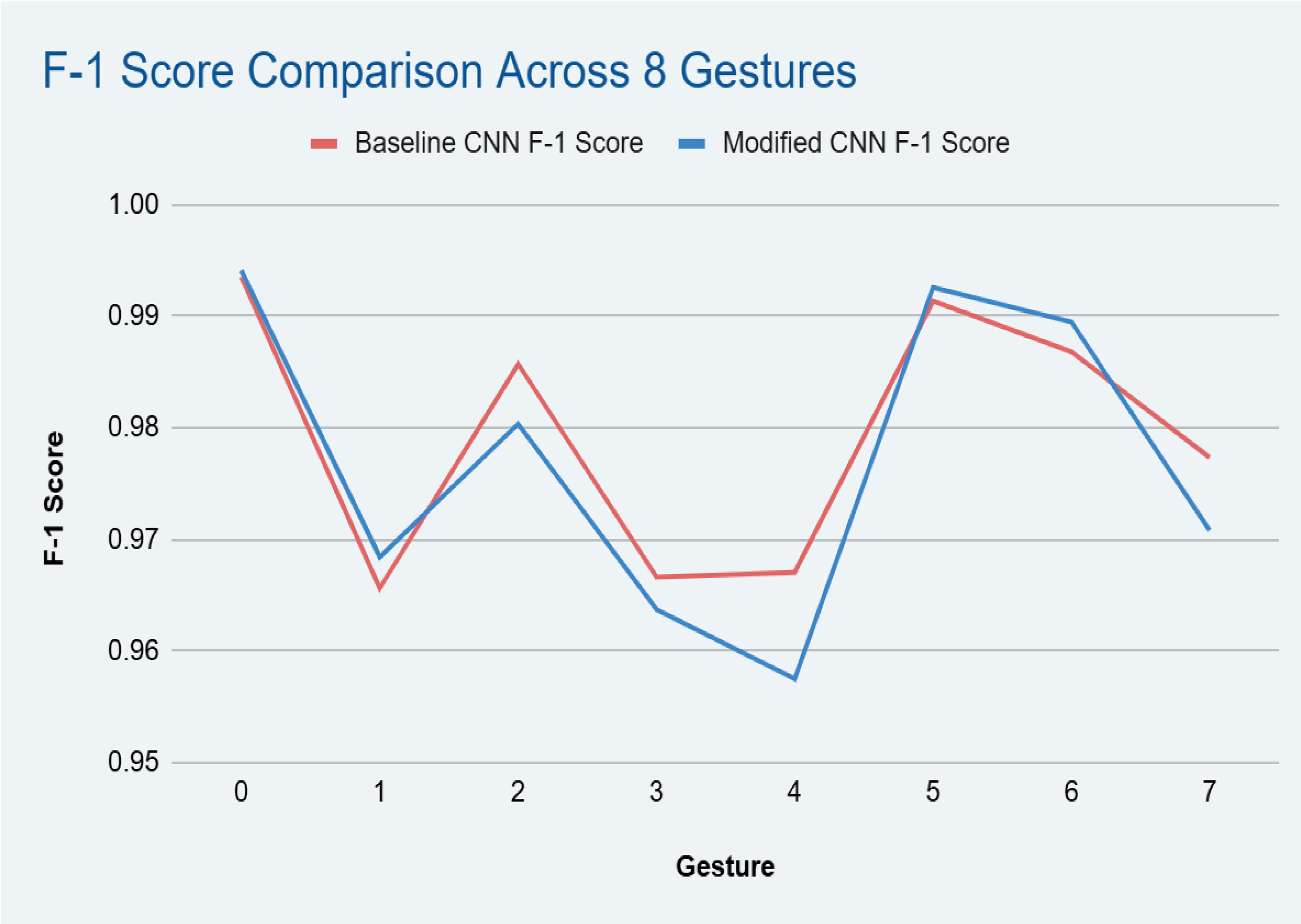
- 10% of the filters with the lowest L1 Norms being pruned.
- Decreasing the dropout from 0.5 to 0.3.
- Cutting down the filter sizes in half to simplify the network while retaining vital information.

Trainable Parameters: **5.74M → 2.86M parameters.**

Below is a summary of the Modified CNN's resulting metrics in comparison to the Baseline CNN's results.

	Average Inference Latency per Batch (s)	Final Best Test Accuracy	Average F-1 Score
Baseline CNN	0.1117	97.93%	0.9793
Modified CNN	0.0318	97.73%	0.9771

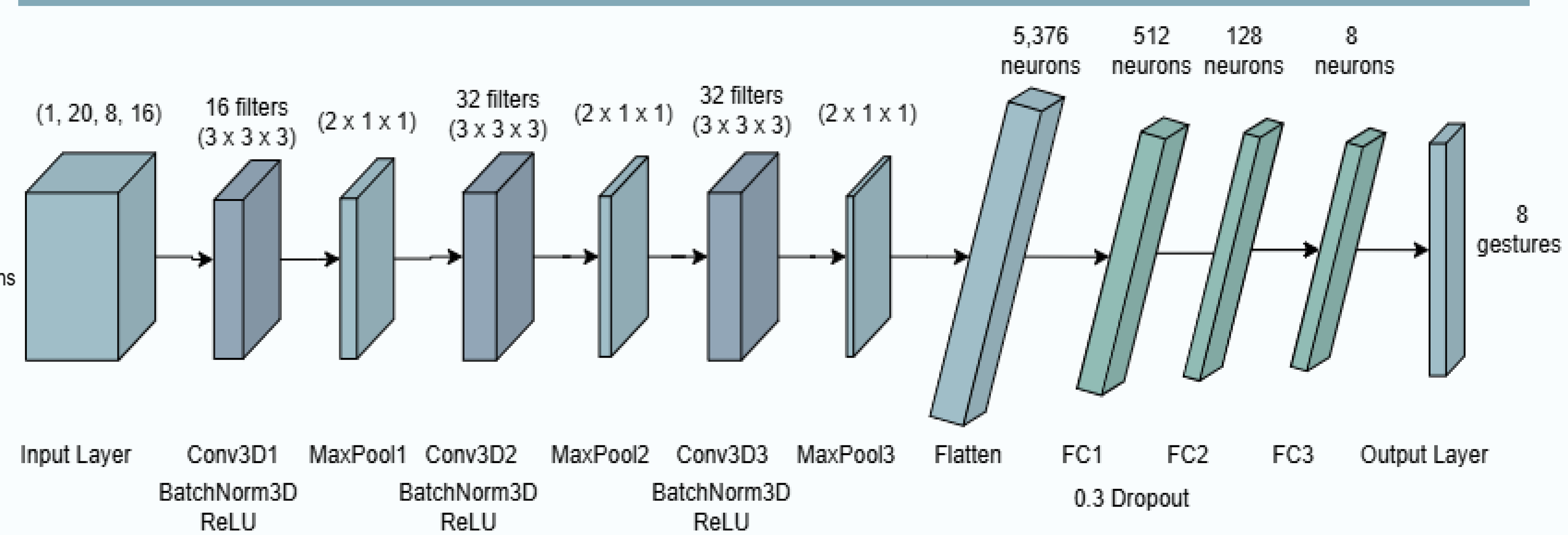
- Percent Reduction in F-1 Score: **0.20%**
- Percent Reduction in Inference Latency: **51.57 %**
- Percent Reduction in Final Test Accuracy: **0.20%**



As shown by the graph, the F-1 Scores across the 8 gestures for each model only varied by a miniscule percentage. Both models had high F-1 Scores above 0.95.

With subjects 1, 5, and 6, the modified CNN evidently outperforms the baseline CNN.

Proposed 3D CNN Architecture



Discussion and Conclusions

While there was a slight reduction in average F-1 Score in the modified CNN model, this was outweighed by the significant reduction in the time taken in inference per batch. This in turn suggests that the trade-off between inference latency and model performance was effectively minimized without compromising model performance.

However, further research is necessary to ensure generalization across unseen subjects – a crucial component in real-time gesture recognition for aiding individuals with disabilities and refining prosthetic limbs.

Due to variations in HD-sEMG signals across individuals, achieving perfect generalization remains a challenge. Implementing Leave-One-Out Cross Validation (LOOCV) and additional data augmentation techniques will be the key next steps toward improving model robustness and moving towards constructing a generalizable model.

Nevertheless, the proposed neural network architecture will allow for enhanced efficiency in real-time gesture recognition, allowing for smoother control and swift responsiveness in myoelectric control systems and prosthetics.

References

Geng W, Du Y, Jin W, et al. Gesture recognition by instantaneous surface EMG images[J]. Scientific reports, 2016, 6: 36571.

Draw.io software was used for the architecture and k-fold sampling visualizations.

Acknowledgments

Thank you to Dr. Yue Wen for his expert guidance and constant encouragement.

UCF Office of Undergraduate Research
UNIVERSITY OF CENTRAL FLORIDA