Abstract— **High fidelity myoelectric control of prostheses and orthoses is paramount to restoring lost function to amputees and neuro-muscular disease sufferers. In this study we prove that spatio-temporal imaging can be used to allow convolutional neural networks to classify sparse channel EMG samples from a consumer-grade device with over 94% accuracy. 10,572 images are generated from 960 samples of simple and complex isometric finger poses recorded from 4 fully intact subjects. Real-time classification of 12 poses is achieved with a 250ms continuous overlapping window.**

I. INTRODUCTION

Surface electromyography (EMG) is the detection and recording of electro-muscular activity through the skin. Instrumentation amplifiers and cascaded signal filters translate instantaneous voltage potentials from muscle-fiber depolarization into practicable signals [1]. EMG is a painless non-invasive procedure unlike its intra-muscular variant despite yielding less accurate results [2]. It is thus a critical avenue of research to improve signal acquisition and interpretation for such a non-invasive method to improve the lives of those who need it.

Applications for myoelectric-based controllers range from medical prosthesis [3-6] to industrial orthosis [7-11] and even entertainment [12]. However, high-fidelity control based on these multivariate quasi-random time variant signals is still yet to be achieved.

The nature of EMG signals makes them a good candidate for neural networks which can approximate any function linearly separable by an n-dimensional hyperplane. Kernel methods may be used to map data into a linearly separable space if required. Feature mapping on the other hand is used to reduce dimensionality [13] and necessary classifier complexity accordingly. The choice or combination of techniques will affect the margin of separation between each class and the hyperplane. This research aims to analyze the effectiveness of the intrinsic feature learning properties of convolutional neural networks.

Many systems base their recognition on transitional isotonic movements between resting states [6, 14, 15]. For n possible states there exists $\sum_{i=1}^{n}(n-i)$ state transitions e.g. 10 poses gives 45 transitions. Systems that operate on

R. M. Stephenson is with the School of Mechanical and Mechatronic Systems, University of Technology Sydney, Ultimo, NSW 2007 Australia (robert.stephenson@student.uts.edu.au).

R. Chai is with the Faculty of Science, Engineering & Technology, Swinburne University of Technology, Hawthorne, VIC 3122 Australia (rchai@swin.edu.au).

D. Eager is with the School of Mechanical and Mechatronic Systems, University of Technology Sydney, Ultimo, NSW 2007 Australia (david.eager@uts.edu.au).

transient motions run the risk of de-synchronization with the user which can only be recovered with forced repetition of the intended motion. The alternative is to perform recognition based on steady-state isometric poses which would allow a system to always remain synchronized. The consequent reduction in time-variant information however adds complexity to the inverse problem since the average signal power of electro-muscular activity changes very little with static poses. Prior studies have proven pre-processing in the time and frequency domains allows shallow network classifiers to achieve accurate results with transient motions [15, 16] given the high separability of unique gestures. This study will determine whether sufficient time-variant information exists in steady-state isometric EMG data to utilize spatial convolution.

II. DATA COLLECTION

A. Overview

A group of 4 healthy fully intact subjects (3M-1F) ranging from 23 to 27 years of age were asked to perform a series of isometric poses. Subjects transition from an arbitrary rest state to a target pose which is held for 3 seconds. The 12 poses include simple single flexions of digits (5 poses), complex flexions of digits with thumb (4 poses), clenched fist with and without thumb extension (2 poses) and full digit extension (1 pose).
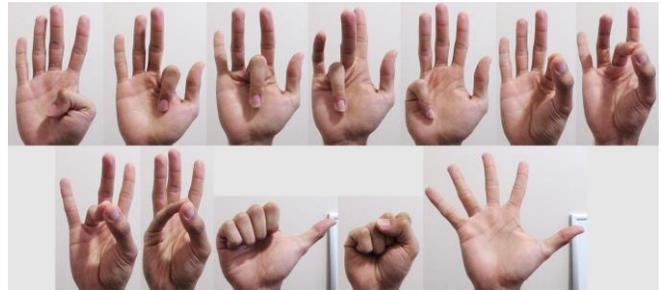
Figure 1- Poses used for classification from 1 to 12 (top left to bottom right).

Each pose is repeated 20 times totaling a 960-sample 8x600 dataset which is sub-sampled into 10,572 8x50 images (II.D). The total sample pool was randomly distributed into 7400 training images, 1585 validation images and 1585 testing images (70:15:15).

B. Hardware

High-density EMG is limited in commercial applications due to size, setup complexity and expense. This research instead focuses on small and portable consumer grade EMG devices which are relatively cheap and intuitive. Though the computations in this study were performed on a computer it

is certainly possible for optimization to mobile and even embedded platforms [17].

Figure 2 – Myo armband radial electrode configuration worn on the upper forearm.

Acquisition was performed with a Myo; originally developed by Thalmic Labs as a consumer level peripheral device. It features 8 medical-grade stainless steel dry surface electrodes positioned radially in a band worn on the upper forearm. It provides a maximum EMG acquisition rate of 200Hz with signal amplitudes mapped between -1 and 1. Additionally, but not used in this experiment, it contains an inertial measurement unit (IMU) which itself can be utilized for other forms of gesture recognition [16] or fused with EMG [18, 19].

C. Procedure

The electrode band was fitted relative to the subject's forearm length (measured along the ulna). Where necessary, plastic clips were fastened into the band per manufacturer recommendations to provide tighter fitment for smaller forearm diameters. A video clip of each pose was played to provide each subject adequate time to move into position before a 3 second pause for recording. Each pose was held in isometric tension just as one would for typical daily activities such as grasping a cup. This was followed by a rest clip guiding the subject to relax and move in an intentionally arbitrary manner.

While this relaxation period is necessary for subject recovery it has the added benefit of increasing intra-class variability to promote better classifier generalization. This is furthered by having each pose recorded in sequence as single repetitions of n classes for i samples (sequential recording) rather than i repetitions of each class (batch recording).

D. Imaging

Images can be created by encoding higher-dimensional data into pixel position and color parameters. It is typical to use a single feature vector when using shallow networks like multi-layer perceptrons however multi-dimensional temporal information can be mapped to images to take advantage of the pattern recognition power of convolutional neural networks.

The use of CNN imaging has been proven with high-density EMG [20]. In that study, images were created from instantaneous samples of 2-dimensional electrode arrays and classified per temporal frame giving a total of 4 dimensional parameters: x, y, amplitude and time. However, for sparse single-dimensional electrode placement, EMG data becomes 3-dimensional. If sampled over time, the resulting $r \times t$

matrix can be encoded into the height, width and pixel luminance of a greyscale image. The hypothesis that class separable time-variant information exists within isometric EMG data is tested by classification with a CNN.
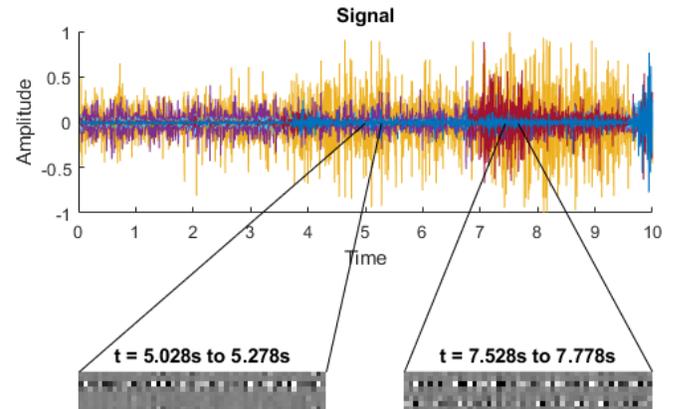
Figure 3 - The process of imaging the EMG signal involves moving the signal mean from 0 to 0.5 and scaling the normalized amplitude for 8-bit single channel pixel depth.

III. CLASSIFIER ARCHITECTURE

A. Convolutional Neural Network

CNNs operate on the principle of spatial convolution:

$$g(i,j) = \sum_{k,l} f(i+k, j+l) \cdot h(k,l) \qquad (1)$$

This involves sliding a kernel mask over an input image and writing the summation of the element-wise product to each pixel of a new image. This process has the effect of extracting features that can be defined by the kernel parameters. This linear process can be stacked as is shown in Fig. 4.
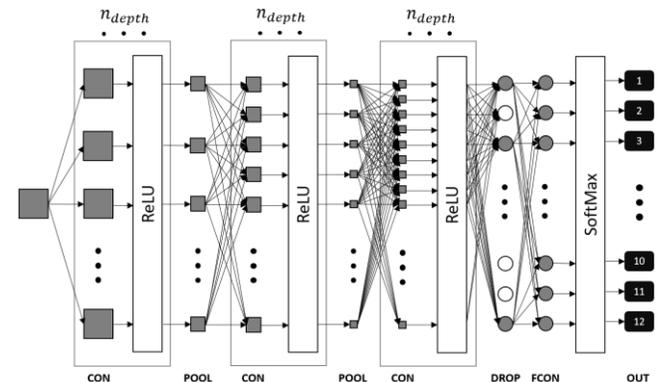
Figure 4 – Structure of convolutional neural network used in this study with adjustable network depth.

The architecture of the CNN used for this study is shown above. Training stops when the validation error gradient condition $\dot{E}_v > 0$ is reached m times (initially set to 5). Critically, a *dropout* layer is added between the final non-linear and fully connected layers. The dropout layer simulates random neuron disconnection with a probability $p = 0.5$ thereby reducing neuron dependency and the risk of over fitment [21].

The CNN structure is governed by a variable network depth. Each level contains a convolution layer, normalization layer and non-linearization layer. Non-linearization is performed with rectified linear units[22]:

$$R(x) = \max(0, x) \quad (2)$$

These have the added advantage of reducing computational requirements. Between each level is a pooling layer to down-sample kernel features. Increasing the network depth increases the amount of convolution before down-sampling occurs. To keep the computational load balanced between network levels (i.e. after pooling) the number of filters in subsequent levels is increased (by a factor of 2 for a down-sampling ratio of 0.5). Finally, the output neuron firing values are converted to probabilities with a SoftMax function layer:

$$\sigma(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}} \quad (3)$$

Classification is performed on a 250ms window of data for every n^{th} point chosen based on processing speed. This continuous overlapping window allows for fluid recognition in real-time.

IV. RESULTS

A. Preliminary

Initial hyperparameters produced 79.02% classification accuracy which was a promising start. Confusion was concentrated between classes 2, 6 and 7 (all of which have shared tendon activity).
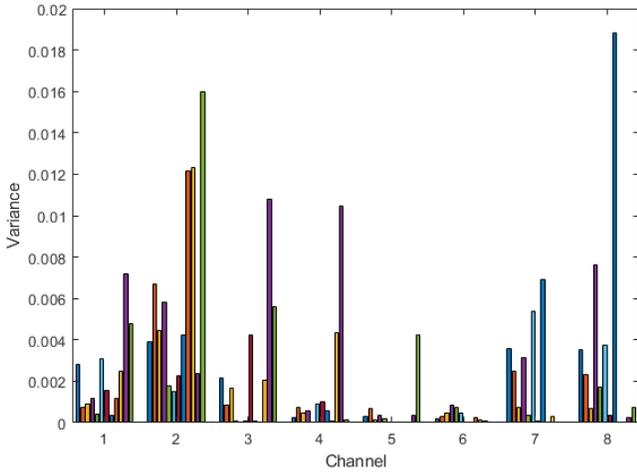
Figure 5 – Average RMS signal power variance across each class grouped by channel. Color represents class.

One of the problems with a fixed sparse electrode arrangement is the effect small variances in placement can have on observed signal energy. The average signal power was calculated for each channel and compared between subjects. Channels 5 and 6 which were located over the medial-anterior area of the upper ulna observed relatively low but stable electro muscular activity across all subjects. This was in contrast with channels 2 and 8 which observed relatively high intra-class variance between subjects. Since electrode placement is fixed with respect to the armband

itself we are not able to place electrodes in the ideal innervation zones.

B. Bayesian Optimization

While it is possible to iterate over a predefined range for each hyperparameter in a typical grid search, it's an unfeasible approach for functions with significant cost. Instead a Bayesian optimization strategy is used to determine the appropriate network depth, initial learning rate, momentum and regularization.

Bayesian optimization [23] is a non-parametric approach that bases subsequent hyperparameter testing point selection on an expected improvement function which maximizes either exploration of unknown regions or cost minimization based on prior information. This results in maximum exploitation of a minimal number of test values. For our network this process was continued for 100 function evaluations, eventually settling on a network depth of $n_{depth} = 2$.

D. Final Result

Figure 6 - Optimized confusion matrix for test set.

A final classification accuracy of **94.9%** was achieved on a test dataset consisting of 1587 unseen images. Using this result, the optimized network's real-time performance was evaluated with a 20 second continuous sample of a subject performing 5 different poses. Fig. 7 shows the raw signal plotted against the classifier probability for each class. There is a clear output step shown for the duration of each pose with some marginal false activity between transitional states which is to be expected.

In future it would be worthwhile testing the effect of the aforementioned isotonic movements on CNN accuracy. A state recognition or other control system would need to be implemented to switch between transitional movement and steady state pose classification; this is where IMU fusion would be useful. Additionally, image augmentation techniques such as pixel translation or image flipping should be used to test the CNNs spatio-temporal invariance.
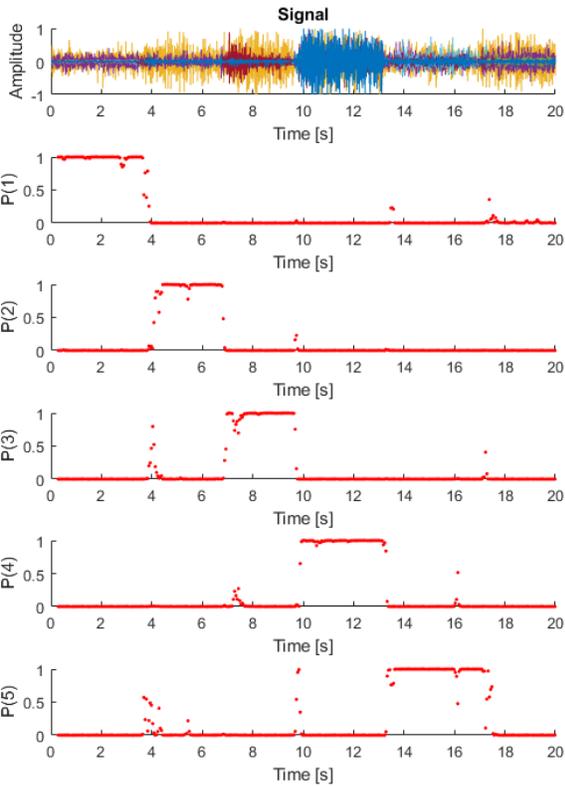
Figure 7 – Simultaneous classifier probability outputs from 0 to 1 for classes 1 to 5 during a real-time session of a subject performing different poses. Transient states (between poses) are reflected by step transitions such as those ocurring at 4s, 7s and 10s.

V. CONCLUSION

This study has put forward and tested the hypothesis that adequate time-variant information exists to classify isometric EMG samples spatio-temporally. Moreover, it proves that sparse channel consumer grade EMG devices can still be used for relatively accurate classification of static finger poses. This is in contrast with high-density arrangements and isotonic or dynamic hand gestures. Real-time classification was achieved with an overlapping 250ms processing window.

REFERENCES

1. Merletti, R. and P. Philip, *Electromyography: Physiology, Engineering, and Non-Invasive Applications*. 2004: John Wiley & Sons, Inc.
2. Hargrove, L.J., K. Englehart, and B. Hudgins, *A Comparison of Surface and Intramuscular Myoelectric Signal Classification.* IEEE Transactions on Biomedical Engineering, 2007. **54**(5): p. 847-853.
3. Jiang, N., et al., *Extracting Simultaneous and Proportional Neural Control Information for Multiple-DOF Prostheses From the Surface Electromyographic Signal.* IEEE Transactions on Biomedical Engineering, 2009. **56**(4): p. 1070-1080.
4. Cipriani, C., et al., *On the Shared Control of an EMG-Controlled Prosthetic Hand: Analysis of User–Prosthesis Interaction.* IEEE Transactions on Robotics, 2008. **24**(1): p. 170-184.
5. Li, G., A.E. Schultz, and T.A. Kuiken, *Quantifying Pattern Recognition—Based Myoelectric Control of Multifunctional Transradial Prostheses.* IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2010. **18**(2): p. 185-192.
6. Chu, J.U., I. Moon, and M.S. Mun, *A Real-Time EMG Pattern Recognition System Based on Linear-Nonlinear Feature Projection for a Multifunction Myoelectric Hand.* IEEE Transactions on Biomedical Engineering, 2006. **53**(11): p. 2232-2239.
7. Kiguchi, K. and Y. Hayashi, *An EMG-Based Control for an Upper-Limb Power-Assist Exoskeleton Robot.* IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2012. **42**(4): p. 1064-1071.
8. Fleischer, C. and G. Hommel. *Calibration of an EMG-Based Body Model with six Muscles to control a Leg Exoskeleton.* in *Proceedings 2007 IEEE International Conference on Robotics and Automation*. 2007.
9. Lenzi, T., et al., *Intention-Based EMG Control for Powered Exoskeletons.* IEEE Transactions on Biomedical Engineering, 2012. **59**(8): p. 2180-2190.
10. Artemiadis, P.K. and K.J. Kyriakopoulos. *EMG-based position and force control of a robot arm: Application to teleoperation and orthosis.* in *2007 IEEE/ASME international conference on advanced intelligent mechatronics*. 2007.
11. Fleischer, C., C. Reinicke, and G. Hommel. *Predicting the intended motion with EMG signals for an exoskeleton orthosis controller.* in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2005.
12. Bitar, F., et al. *A Portable MIDI Controller Using EMG-Based Individual Finger Motion Classification.* in *2007 IEEE Biomedical Circuits and Systems Conference*. 2007.
13. Nazmi, N., et al., *A Review of Classification Techniques of EMG Signals during Isotonic and Isometric Contractions.* Sensors (Basel), 2016. **16**(8).
14. Tenore*, F.V.G., et al., *Decoding of Individuated Finger Movements Using Surface Electromyography.* IEEE Transactions on Biomedical Engineering, 2009. **56**(5): p. 1427-1434.
15. Jun-Uk, C., et al. *Control of multifunction myoelectric hand using a real-time EMG pattern recognition.* in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2005.
16. Stephenson, R.M., G.R. Naik, and R. Chai. *A system for accelerometer-based gesture classification using artificial neural networks.* in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2017.
17. Silva, C.F. and C.A. Siebra. *An investigation on the use of convolutional neural network for image classification in embedded systems.* in *2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*. 2017.
18. Bai, O., et al. *Characterization of lower limb activity during gait using wearable, multi-channel surface EMG and IMU sensors.* in *2017 International Electrical Engineering Congress (iEECON)*. 2017.
19. Wu, J., L. Sun, and R. Jafari, *A Wearable System for Recognizing American Sign Language in Real-Time Using IMU and Surface EMG Sensors.* IEEE Journal of Biomedical and Health Informatics, 2016. **20**(5): p. 1281-1290.
20. Geng, W., et al., *Gesture recognition by instantaneous surface EMG images.* Sci Rep, 2016. **6**: p. 36571.
21. Srivastava, N., et al., *Dropout: A Simple Way to Prevent Neural Networks from Overfitting.* Journal of Machine Learning Research, 2014. **15**.
22. Nair, V. and G.E. Hinton. *Rectified Linear Units Improve Restricted Boltzmann Machines.* in *International Conference on Machine Learning*. 2010. Haifa, Israel.
23. Snoek, J., H. Larochelle, and R.P. Adams. *Practical bayesian optimization of machine learning algorithms.* in *Advances in neural information processing systems*. 2012.