HATRNet: Human Activity/Transition Recognition using Deep Neural Networks
Nicholas Gaudio, Akash Levy, Jonas Messner, (ngaudio, akashl, messner)@stanford.edu
Department of Electrical Engineering, Stanford University

BACKGROUND and MOTIVATION
- Human Activity Recognition has great potential for customized healthcare
- Smartphones incorporate sensors (accelerometer, gyroscope etc.)
- Sensor data can be used to classify human activities and transitions
- Improvements compared to state of the art:
  1. Advanced preprocessing including data augmentation
  2. End-to-end deep learning solution (no feature extraction)
  3. Improved architecture enabling accurate classification of transitions

1. DATASET
- SBHAR dataset of 6 activities and 6 postural transitions from the Galaxy S5
- 3-axial linear acceleration and 3-axial angular velocities traces at 50 Hz
- Augmented to 3,640 examples with no feature selection (beyond FFT)

2. NEURAL NETWORK ARCHITECTURE

Siamese (non-weight sharing) CNN
- Left subnetwork takes time traces as input (6 zero-padded channels)
- Right subnetwork takes frequency and phase traces as input (12 interpolated channels)
- Late sensor fusion employed for encoded, efficient feature extraction
- Conv1D filter size: 1x14  Conv2D filter size: 3x42  Filter #60  learning rate: 0.0026

Sequence (LSTM) Model
- Takes time traces as input (6 channels of variable length)
- Two LSTM layers (128 to 32 output channels)

3. RESULTS and DISCUSSION

Literature comparison (grouped postural transitions):
<table>
<thead>
<tr>
<th>SVM, 561 features extracted [1]</th>
<th>CNN (2.5 second traces) [2]</th>
</tr>
</thead>
<tbody>
<tr>
<td>WA</td>
<td>WC</td>
</tr>
<tr>
<td>WA</td>
<td>18</td>
</tr>
<tr>
<td>WC</td>
<td>10</td>
</tr>
<tr>
<td>WD</td>
<td>0</td>
</tr>
<tr>
<td>SI</td>
<td>0</td>
</tr>
<tr>
<td>ST</td>
<td>0</td>
</tr>
<tr>
<td>LD</td>
<td>0</td>
</tr>
<tr>
<td>PT</td>
<td>0</td>
</tr>
</tbody>
</table>

Siamese (non-weight sharing) CNN
- Architecture Comparison:

<table>
<thead>
<tr>
<th>Number of categories</th>
<th>Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN1</td>
<td>3.29%</td>
</tr>
<tr>
<td>CNN2</td>
<td>3.02%</td>
</tr>
<tr>
<td>LSTM</td>
<td>10.11%</td>
</tr>
<tr>
<td>SVM [1]</td>
<td>3.22%</td>
</tr>
<tr>
<td>Perceptron [2]</td>
<td>2.75%</td>
</tr>
</tbody>
</table>

FUTURE WORK
- Model ensembling using data representations from sequence models
- Implement the model on a smartphone for real-time inference
- Incorporate frequency and phase traces into the sequence model

REFERENCES