Pose-based Body Language Recognition for Emotion and Psychiatric Symptom Interpretation
Zhengyuan Yang, Amanda Kay, Yuncheng Li, Wendi Cross, Jiebo Luo

Emotions from the body language.

- defensive, disagree
- honesty, sincerity
- ashamed, upset

Challenges

- Body language recognition with limited samples
- Interpretable emotion and psychiatric symptom inference

Contributions

- A framework to infer emotions from recognized body languages
  - Interpretable and transferable framework
  - A URMC dataset with body language, emotion, and psychiatric symptoms annotations

Pose-based Body Language Recognition

- Pose-based body language recognition: transferable and data efficient
- KNN based classification: example based and interpretable
- Emotion RNN: emotion and psychiatric symptom interpretation based on the body language sequence

Challenges

- Pose-based recognition: Requires less data and has better transferability
- Pose representation: CNN-based pose image pre-trained on action datasets

Contributions

- Interpretable and transferrable framework
- A URMC dataset with body language, emotion, and psychiatric symptoms annotations

Framework

- Pose-based body language recognition: transferrable and data efficient
- KNN based classification: example based and interpretable
- Emotion RNN: emotion and psychiatric symptom interpretation based on the body language sequence

URMC Dataset

- 144 30-second video clips cropped from 12 20-minute videos
- 9 upper body body languages, 5 lower body body languages selected from 32 body languages in the initial annotations
- 24+1 emotion labels, and 24+1 symptom labels

Quantitative Results

Results on multi-label body language recognition

<table>
<thead>
<tr>
<th>Lower Body Set</th>
<th>Acc.</th>
<th>Prec.</th>
<th>Recall</th>
<th>F1</th>
<th>Interpretable</th>
<th>Transferable</th>
<th>Required Data Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pose Stream</td>
<td>0.443</td>
<td>0.581</td>
<td>0.407</td>
<td>0.438</td>
<td>✓</td>
<td>✓</td>
<td>Small</td>
</tr>
<tr>
<td>N/T×3+VSM</td>
<td>0.322</td>
<td>0.376</td>
<td>0.409</td>
<td>0.424</td>
<td></td>
<td></td>
<td>Medium</td>
</tr>
<tr>
<td>N/T×3+SSN</td>
<td>0.483</td>
<td>0.536</td>
<td>0.625</td>
<td>0.538</td>
<td></td>
<td></td>
<td>Small</td>
</tr>
<tr>
<td>ST-CNN/Pos+SSN</td>
<td>0.384</td>
<td>0.397</td>
<td>0.608</td>
<td>0.460</td>
<td></td>
<td></td>
<td>Small</td>
</tr>
<tr>
<td>ST-CNN/Pos+VSM</td>
<td>0.520</td>
<td>0.608</td>
<td>0.554</td>
<td>0.554</td>
<td>✓</td>
<td>✓</td>
<td>Small</td>
</tr>
</tbody>
</table>

Results on emotion interpretation

<table>
<thead>
<tr>
<th>LSTMS+ST-CNN/Pos</th>
<th>Acc.</th>
<th>Prec.</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>L = 1, S = 1</td>
<td>0.408</td>
<td>0.644</td>
<td>0.630</td>
<td>0.637</td>
</tr>
<tr>
<td>L = 7, S = 3</td>
<td>0.564</td>
<td>0.775</td>
<td>0.674</td>
<td>0.721</td>
</tr>
<tr>
<td>L = 48</td>
<td>0.145</td>
<td>0.162</td>
<td>0.587</td>
<td>0.754</td>
</tr>
</tbody>
</table>

Other Methods

- LSTM+ST-CNN/Pos
- Conv1D+NT×3
- Conv1D+ST-CNN/Pos

- 0.510 0.839 0.565 0.675
- 0.490 0.788 0.565 0.658
- 0.536 0.789 0.652 0.714