Deep Discovery of Facial Motions using a Shallow Embedding Layer

Afsaneh Ghasemi

Mahsa Baktashmotlagh, Simon Denman, Sridha Sridharan, Dung Nguyen Tien, Clinton Fookes

Queensland University of Technoly
Brisbane, Australia
Agenda

- Introduction
  - Objectives

- Research Area and Background

- Methodology

- Experimental Results and Findings
Introduction

- Objectives

  - Understanding human facial actions without verbal communication
  - Understanding pain of patients since pain always accompanied with various mode of facial actions.
Research Area and Background

- Facial Action Recognition

  1978, Ekman et al.
  (AFER: Automatic Facial Expression Recognition using FACS encoding)

- FACS: measure uniquely facial muscle movements
- Emotions can be modelled by combination of AUs
Research Area and Background

- Facial Action Recognition

[1995-2001], Cootes at. el.
Active Shape Models (ASM), Active Appearance Models (AAM)
Facial Action Recognition

- Extracting hand designed features
- Ground truth are available for each frame
- Action unit detection

2013, M. Mavadati et al.

Frame level features:
- Extracting hand designed features
- Ground truth are available for each frame
- Action unit detection
Research Area and Background

- Facial Action Recognition

2002, B. Fasel at. el.
Facial expression using Conv Net.

Input Image
Convolution
Pooling
Input Image
Research Area and Background

- Facial Action Recognition

- Recent advances in Deep learning

- Learning non-linear relationships
Recent developments in deep learning shows promising results for the recognition of facial actions.

- **Current Challenges**
  - Facial actions appear in various complex combinations.
  - Extensive efforts have been devoted to exploring the use of various hand-crafted features, but their application in real-time scenarios is ambiguous.
Although Deep learning shows promising results
  - Training a complex model is computationally expensive
  - Fully connected layers has 95% of the network parameters

Solution?
- We use Random Projection to decrease parameters
  - Powerful method for dimensionality reduction
  - The original high-dimensional data is projected onto a lower-dimensional subspace using a Gaussian random projection
- In comparison with PCA
  - Fast, computationally efficient
  - Preserves distances quite nicely in lower dimensional subspace
Contributions

- We propose a shallow embedding layer using Gaussian Random projection to reduce the number of parameters of Deep learning framework.

- We learn dependencies of temporal segments from neutral to apex of facial actions using Long Short-Term Memory (LSTM).
Methodology

- Our Deep Network consists of,
  - Convolutional 3D (C3D)
    - As a result of C3D, we extract spatio-temporal features
  - Recurrent Neural Network
    - Incorporate salient temporal aspects of expression in hidden states of LSTM network, enabling to learn temporal evolution of action from neutral to onset in a video
Methodology

Video Analysis

- Temporal Classification of Facial Actions using the proposed shallow embedded layer
Datasets

- **Pain Archive (UNBC-McMaster)**
  - 130,000 images
  - 27 subjects
  - 200 videos
  - 25 subjects

- **DISFA**
  - 27 subjects

Experimental Results
## Experimental Results

- **Accuracy of facial action detection using sequence of frames:**

<table>
<thead>
<tr>
<th>DISFA</th>
<th>Average AU detection Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original approach (D = 1024)</td>
<td>74%</td>
</tr>
<tr>
<td>$\Phi_{32}$</td>
<td>68%</td>
</tr>
<tr>
<td>$\Phi_{16}$</td>
<td>72%</td>
</tr>
<tr>
<td>$\Phi_{14}$</td>
<td>76%</td>
</tr>
<tr>
<td>$\Phi_{10}$</td>
<td>54%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pain Archive</th>
<th>Average AU detection Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original approach (D = 1024)</td>
<td>94%</td>
</tr>
<tr>
<td>$\Phi_{64}$</td>
<td>83%</td>
</tr>
<tr>
<td>$\Phi_{32}$</td>
<td>88%</td>
</tr>
<tr>
<td>$\Phi_{16}$</td>
<td>91%</td>
</tr>
<tr>
<td>$\Phi_{8}$</td>
<td>94%</td>
</tr>
</tbody>
</table>

- **Processing time of LSTM using proposed approach vs. the original:**

<table>
<thead>
<tr>
<th>Processing time</th>
<th>Train (sec)</th>
<th>Test (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original approach D = 1024</td>
<td>262.26</td>
<td>0.37</td>
</tr>
<tr>
<td>Proposed approach $\Phi_{14}$</td>
<td>34.26</td>
<td>0.02</td>
</tr>
<tr>
<td>$\Phi_{16}$</td>
<td>35.61</td>
<td>0.03</td>
</tr>
<tr>
<td>$\Phi_{32}$</td>
<td>36.47</td>
<td>0.03</td>
</tr>
</tbody>
</table>
Conclusions and Future work

- **Conclusions,**
  - Effectiveness of using Deep learning based structure in facial action recognition
  - Embedding a shallow layer into the LSTM for dimensionality reduction using random projection

- **For future work,**
  - Dimensionally reduction at convolutional layers rather than fully connected layers, enabling them to run on smart phones
Thank you and Questions