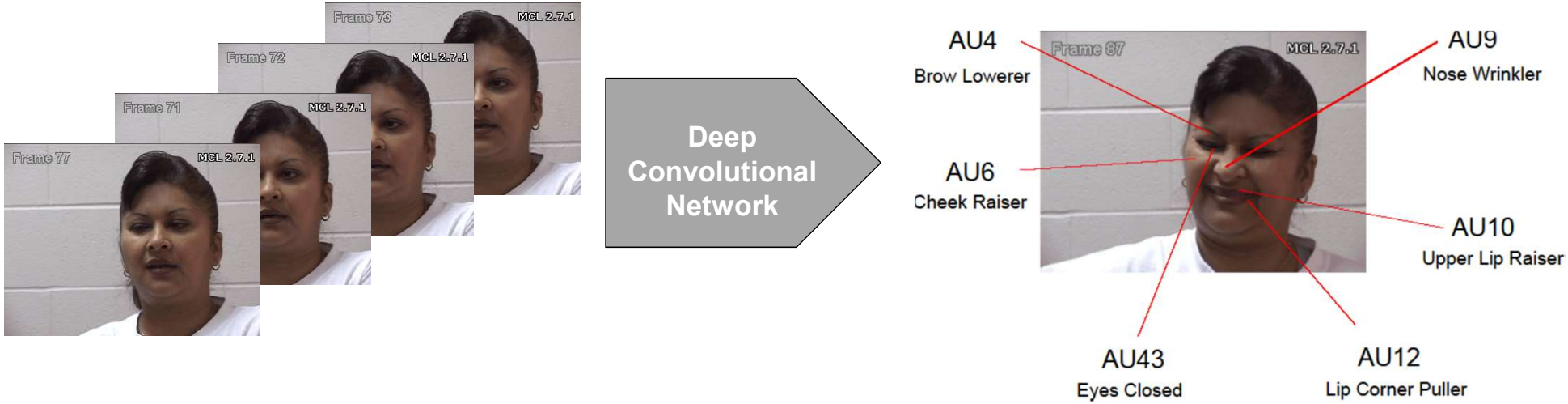
Deep Discovery of Facial Motions using a Shallow Embedding Layer



Afsaneh Ghasemi

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

Queensland University of Technoly

Brisbane, Australia

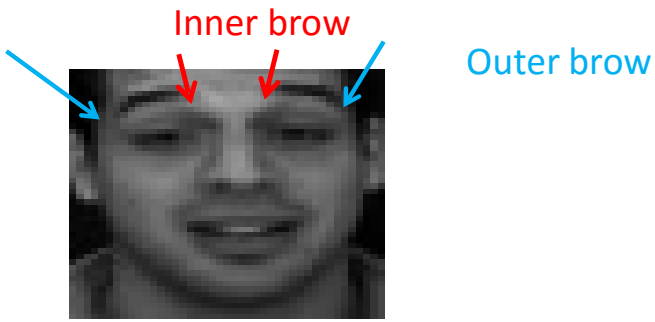
Agenda

- Introduction
  - Objectives
- Research Area and Background
- Methodology
- Experimental Results and Findings

Queensland University of Technology

Introduction

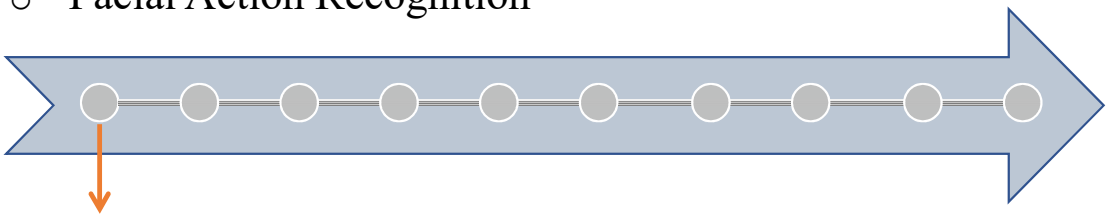
- Objectives
  - Understanding human facial actions without verbal commination



  - Understanding pain of patients since pain always accompanied with various mode of facial actions.

Research Area and Background

- Facial Action Recognition

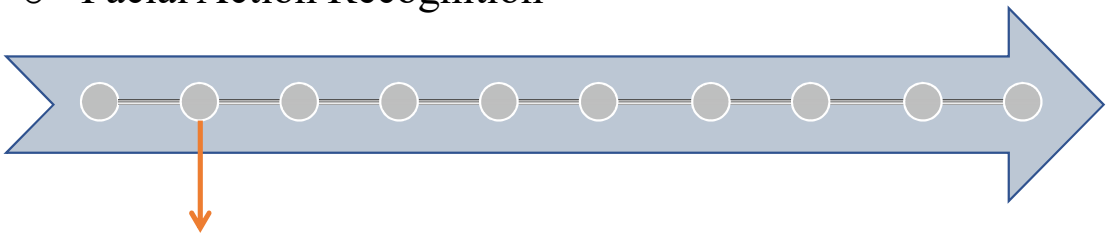


1978, Ekman at. el.
(AFER: Automatic Facial Expression Recognition using FACS encoding)

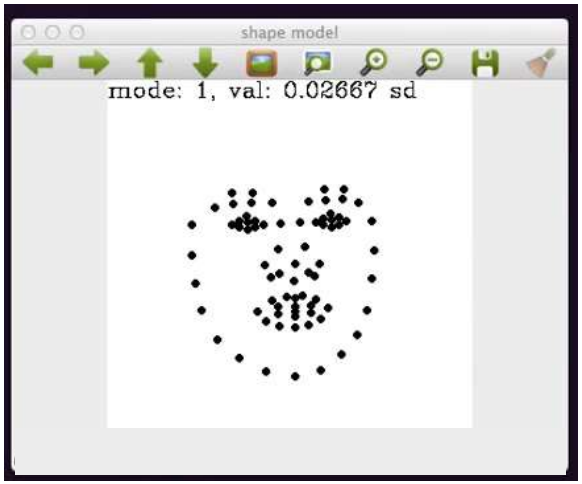
- FACS: measure uniquely facial muscle movements
- Emotions can be modelled by combination of AUs

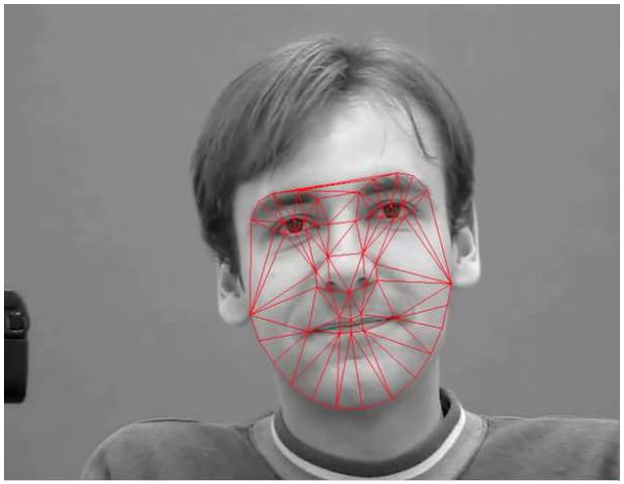
Research Area and Background

- Facial Action Recognition



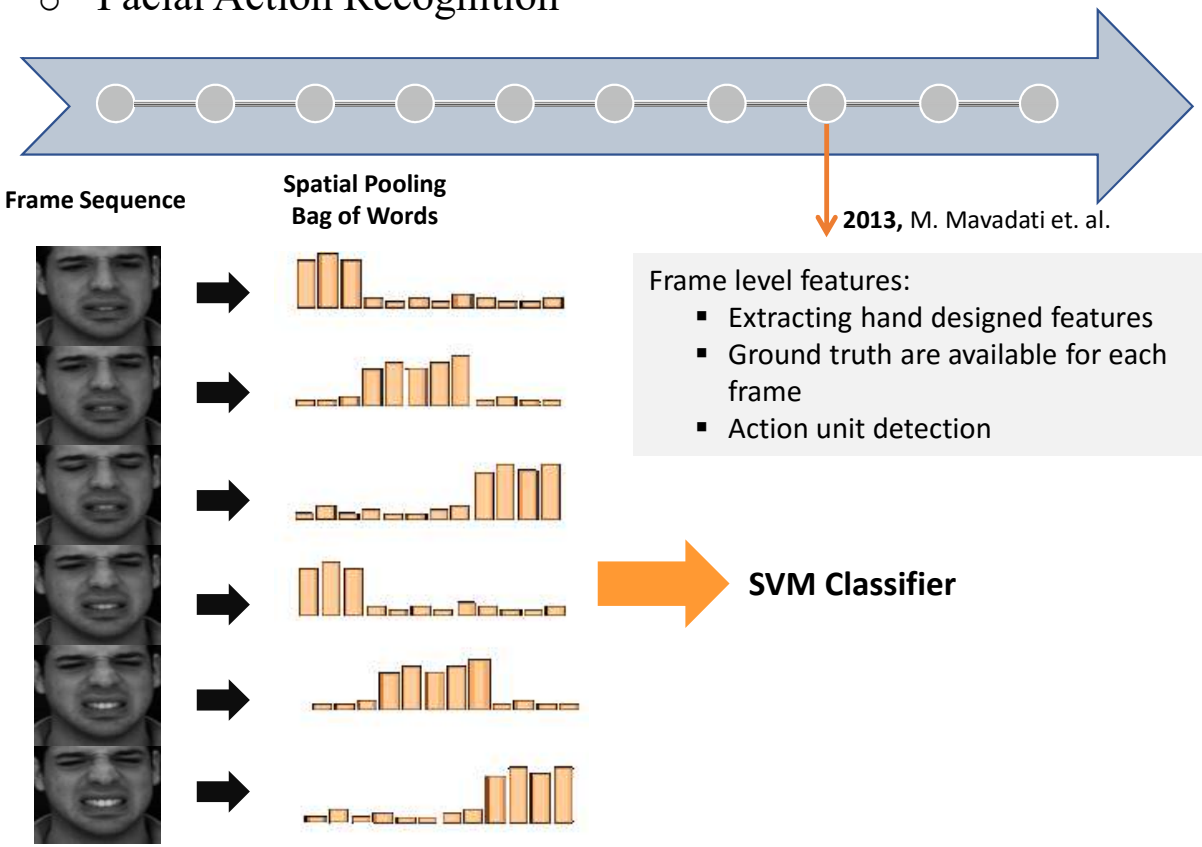
[1995-2001], Cootes at. el.
Active Shape Models (ASM), Active Appearance Models (AAM)





Research Area and Background

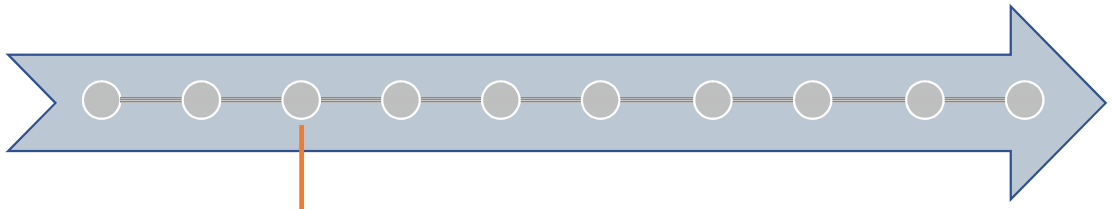
- Facial Action Recognition



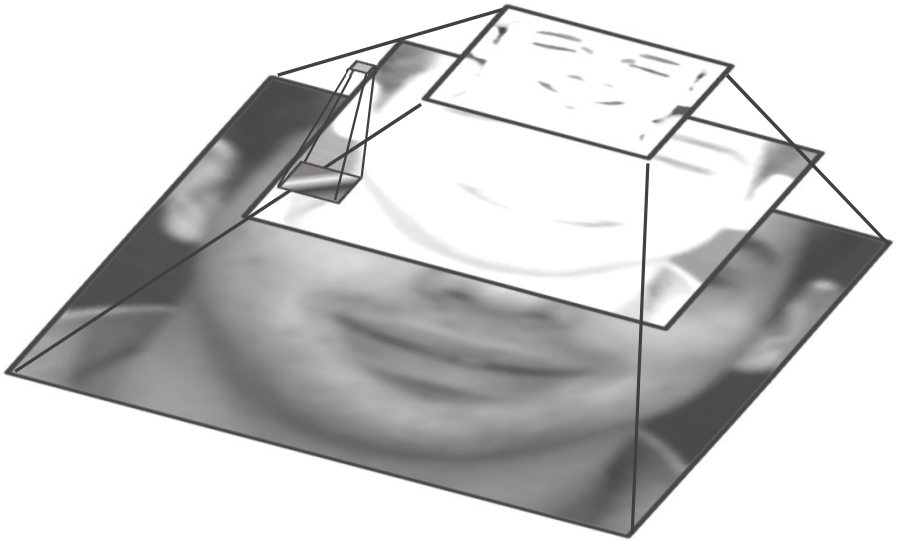
Queensland University of Technology

Research Area and Background

- Facial Action Recognition



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



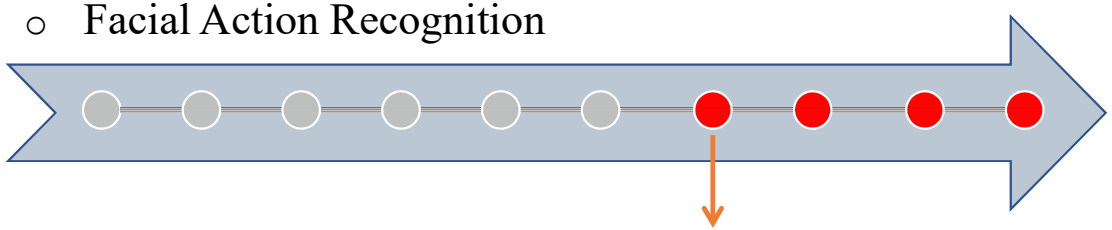
Pooling

Convolution

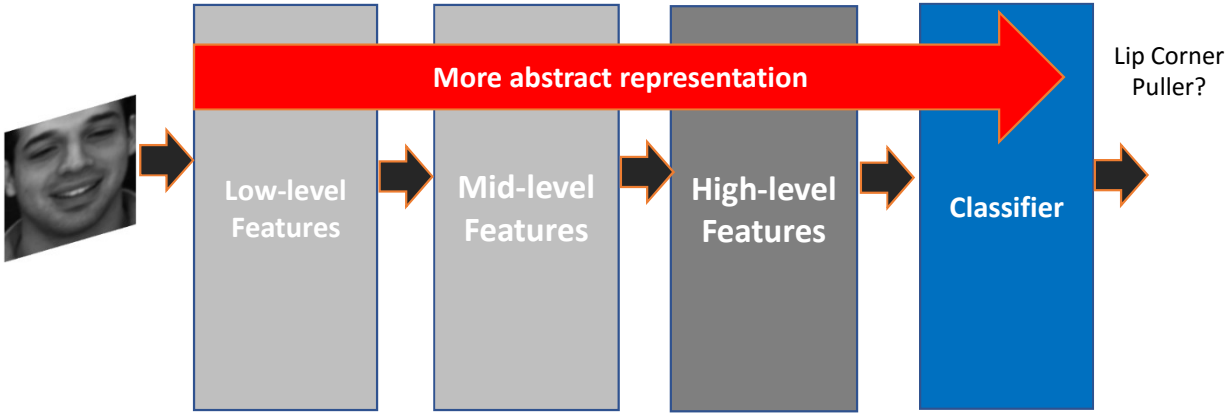
Input Image

Research Area and Background

- Facial Action Recognition



Recent advances in Deep learning

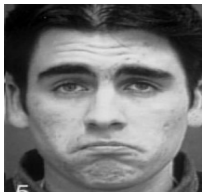


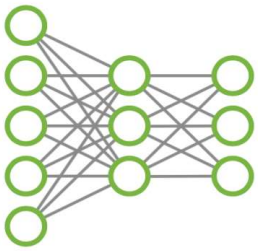
Queensland University of Technology

- Learning non-linear relationships

Research Area and Background

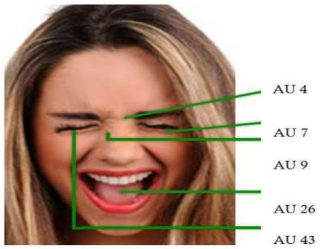
Recent developments in deep learning shows promising results for the recognition



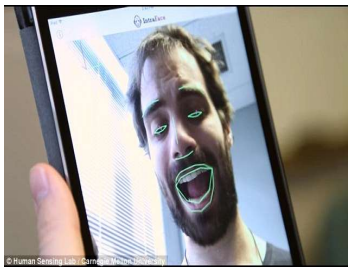


❑Current Challenges

- Facial action appear in various complex combinations



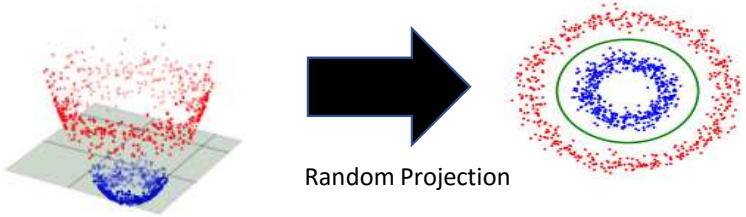
- Extensive efforts have been devoted to exploring the use of various hand-crafted but still using them in real time scenario is ambiguous



Methodology

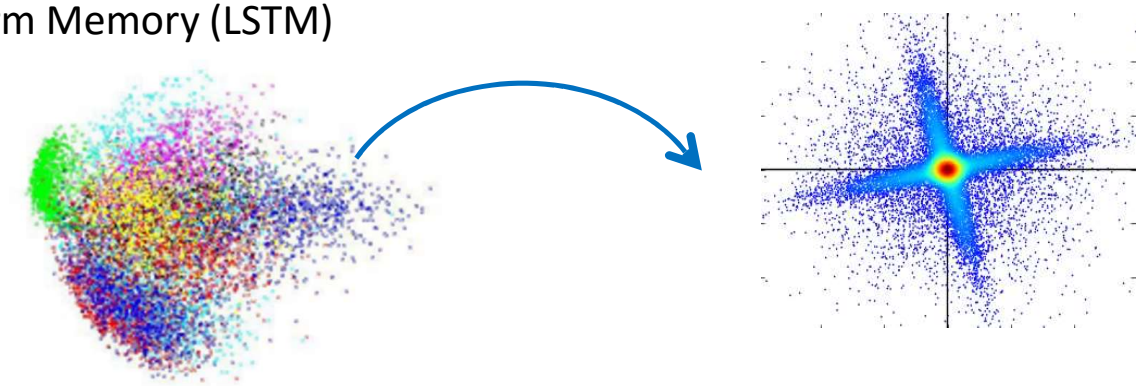
- Although Deep learning shows promising results
  - Training a complex model is computationally expensive
  - Fully connected layers has 95% of the network parameters
- Solution?
  - We use **Random Projection** to decrease parameters
  - Powerful method for dimensionality reduction
  - The original 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

Queensland University of Technology



Methodology

- 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)

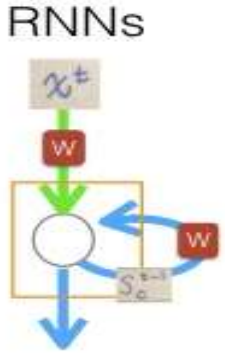


Methodology

- Our Deep Network consists of,

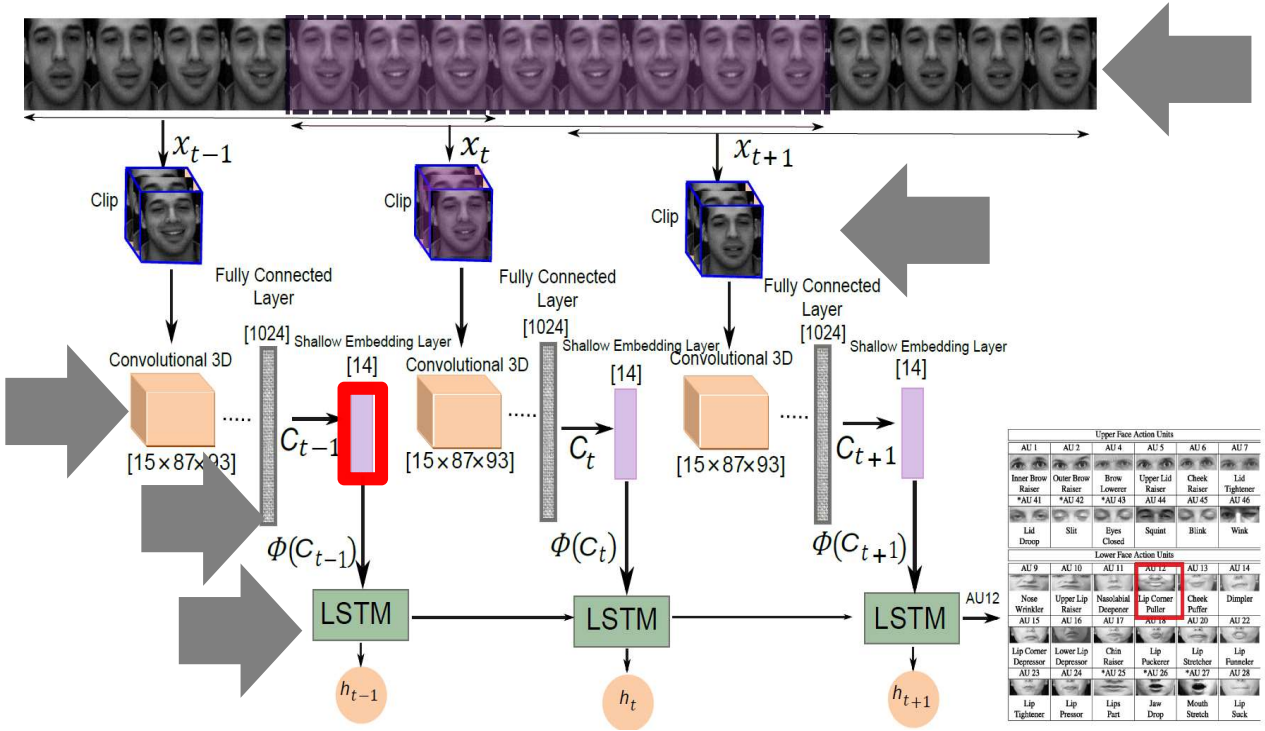
  Convolutional 3D (C3D)
  - As a result of C3D, we extract spatio-temporal features

- 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



Queensland University of Technology

Experimental Results

Datasets

130,000
images
27 subjects

DISFA

200 videos
25 subjects

Pain Archive (UNBC-McMaster)

Queensland University of Technology

Experimental Results

- Accuracy of facial action detection using sequence of frames:

DISFA	Average AU detection Accuracy
Original approach (D = 1024)	74%
Φ_{32}	68%
Φ_{16}	72%
Φ_{14}	76%
Φ_{10}	54%

Pain Archive	Average AU detection Accuracy
Original approach (D = 1024)	94%
Φ_{64}	83%
Φ_{32}	88%
Φ_{16}	91%
Φ_{8}	94%

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

	Processing time	
	Train (sec)	**Test (sec)**
Original approach D = 1024	**262.26**	**0.37**
Proposed approach Φ_{14}	34.26	0.02
Φ_{16}	**35.61**	**0.03**
Φ_{32}	**36.47**	**0.03**

Queensland University of Technology

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