FACIAL ANALYSIS IN THE WILD WITH LSTM NETWORKS

Sarasi Kankanamge, Clinton Fookes, Sridha Sridharan

Image and Video lab, SAIVT Research Program

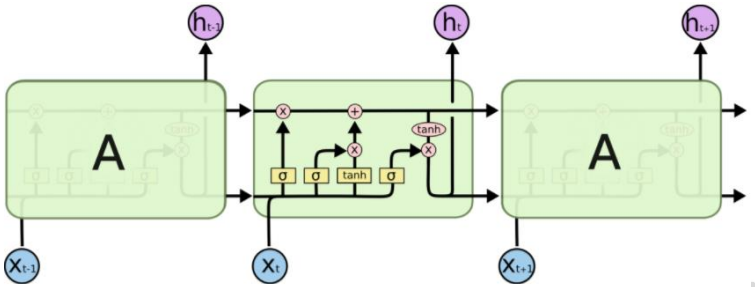
Queensland University of Technology, Brisbane, Australia

Overview

- Introduction
- Proposed Models
  - FaceVideoModel
  - ExpModel
- Experimental Results
- Conclusion

Introduction

- **LSTM:**
  - flexible model to handle a variable-length sequential data in computer vision applications with lower computation cost.
  - powerful tool for facial analysis with fundamental explanations of their ability to capture sequential patterns.
  - more effective than conventional CNNs for several classifi-cation tasks.



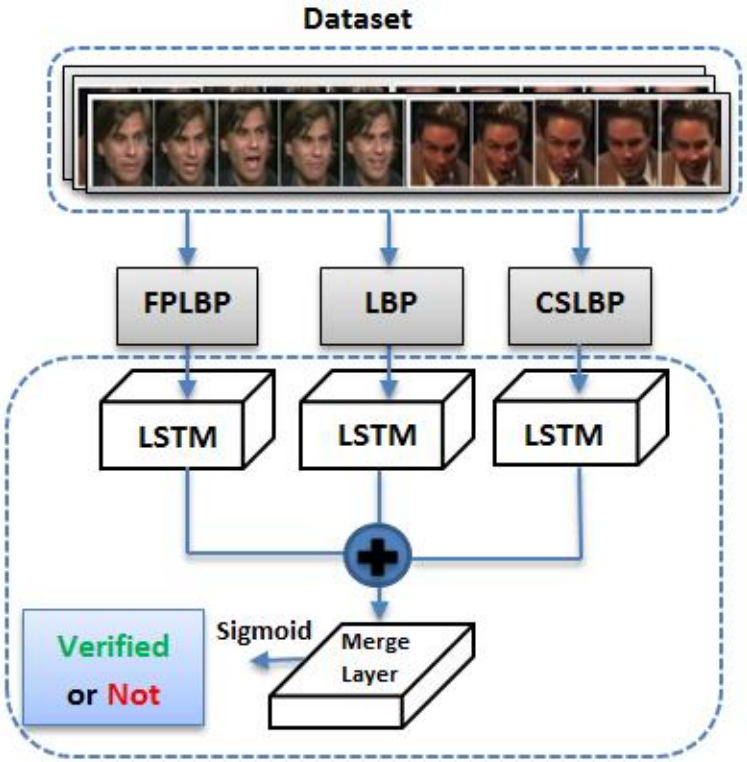
Introduction

Contributions:

1. The development of a LSTM model for video-based face verification in the wild that achieves verification accuracy that outperforms state-of-the-art results on the recently introduced challenging face video database (Youtube faces).

2. The development of a combined deep CNN model and LSTM model architecture to obtain improved spontaneous expression performance demonstrated on the challenging FER2013 facial expression dataset.

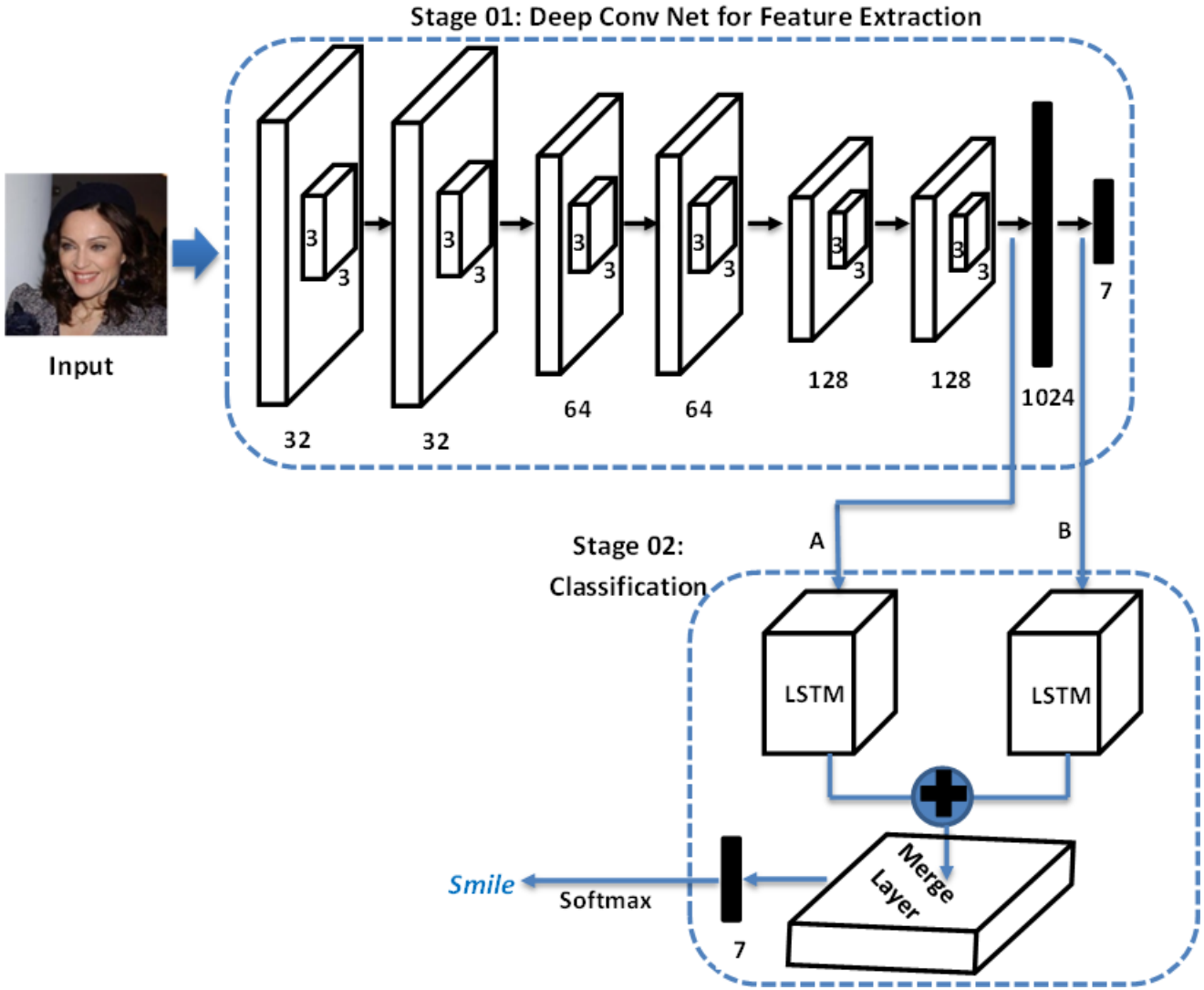
The proposed FaceVideoModel



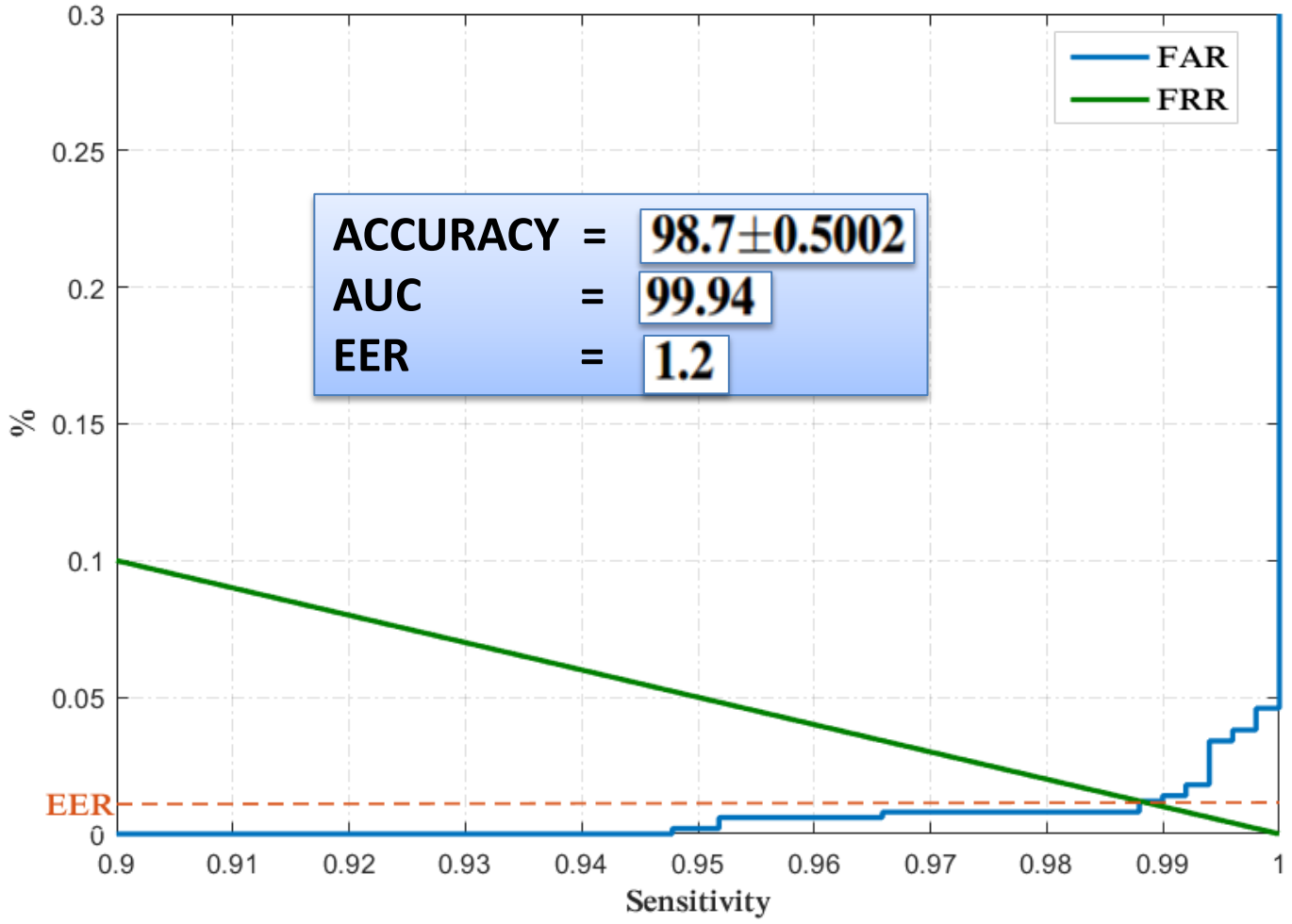
FaceVideoModel

Accuracy curve on YTF

The Proposed ExpModel



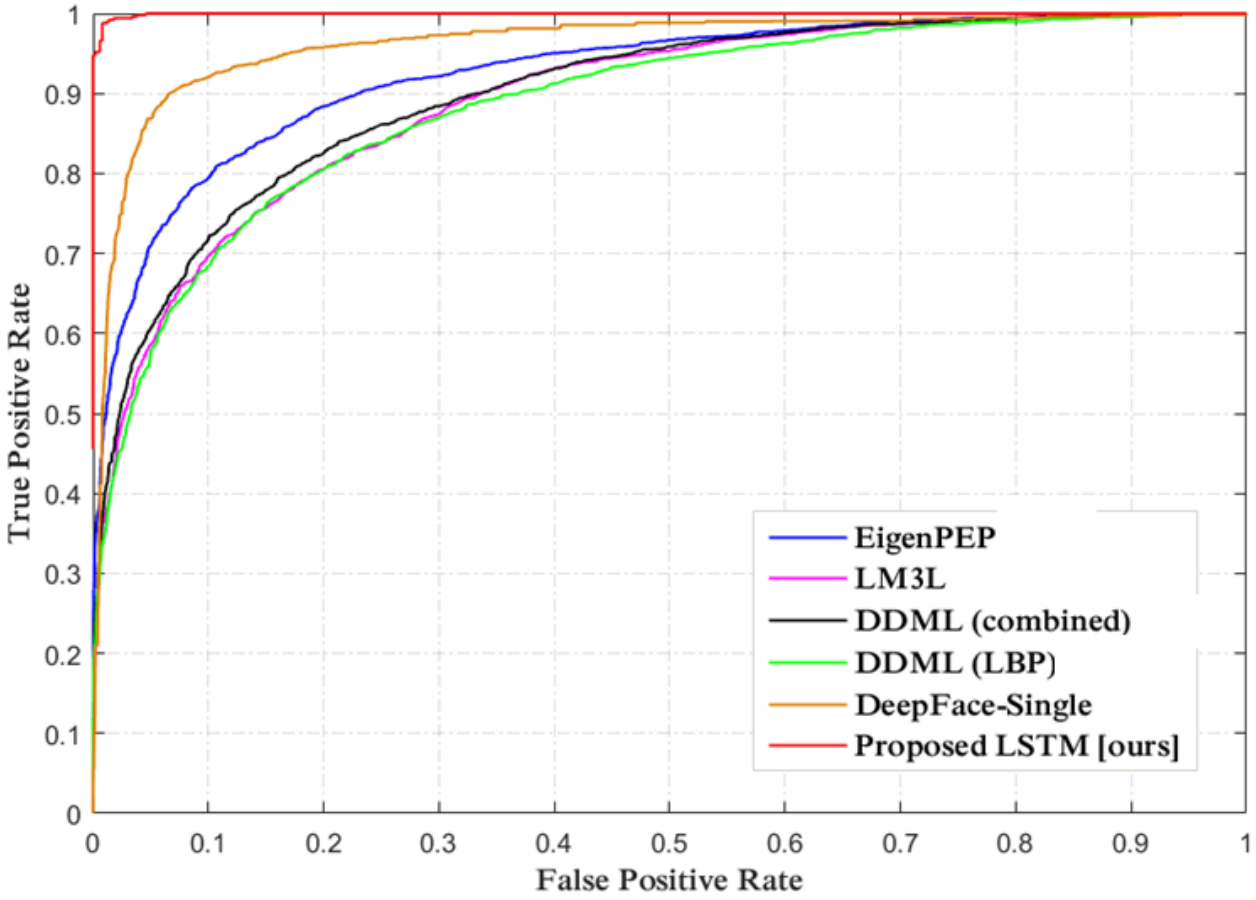
Experimental Results: FaceVideoModel



Experimental Results: FaceVideoModel

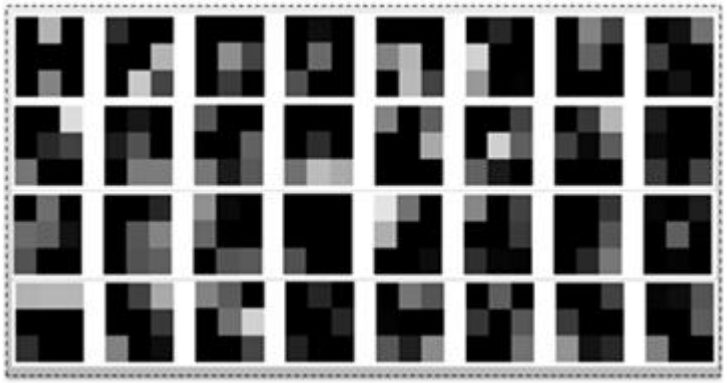
Method	Accuracy±SE	AUC	EER
LM3L *	81.3±1.2	89.3	19.7
DDML (LBP) *	81.3±1.6	88.7	19.7
DDML (combined) *	82.3±1.5	90.1	18.5
EigenPEP *	84.8±1.4	92.6	15.5
MMMF Fusion *	-	93.9	12.6
DeepFace-Single	91.4±1.1	96.3	8.6
AlexNet+LSTMs (ours) *	93.2±0.6136	-	-
FaceNet	95.12±0.39	-	-
Embedding Learning	97.3	-	-
FaceVideoModel (ours) *	**98.7±0.5002**	**99.94**	**1.2**

Experimental Results: FaceVideoModel

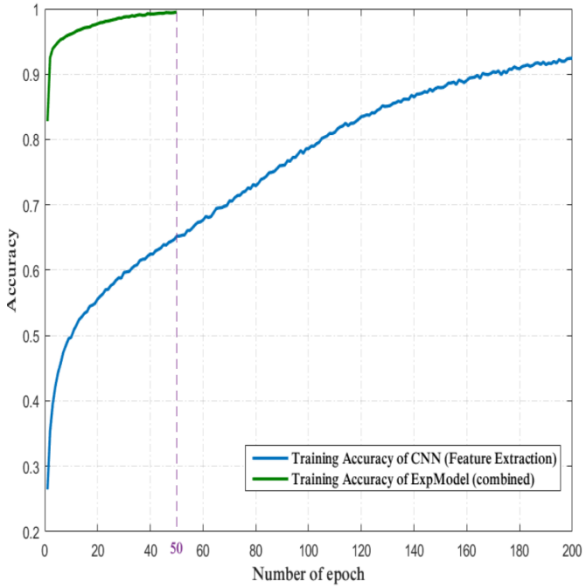


Experimental Results: ExpModel

Feature Maps from first Conv layer

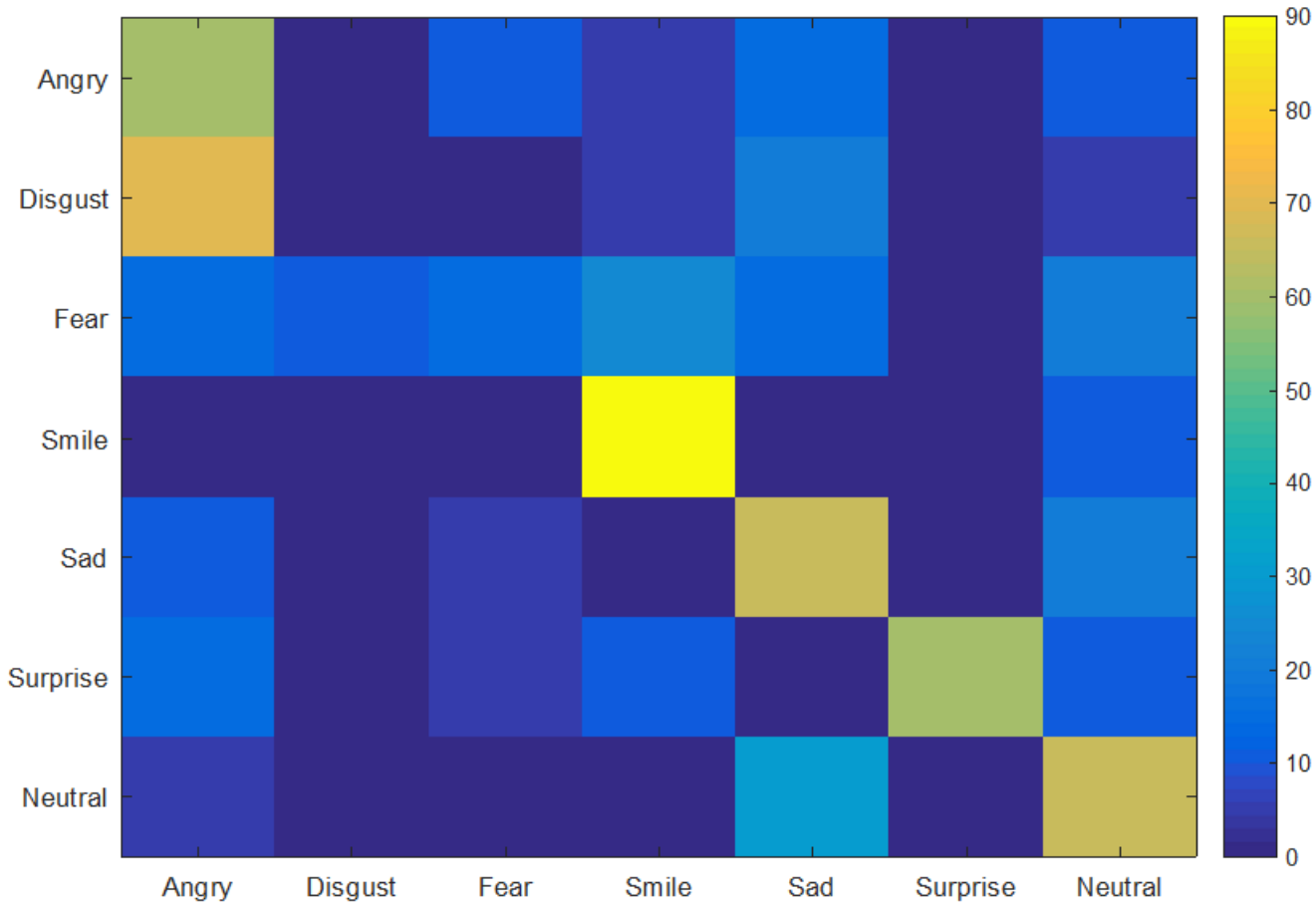


Filter output from first Conv layer



Approach	Validation	Test
DLSVM	0.694	0.712
MNL	≈**0.7**	≈**0.72**
CNN [our]	0.650	-
ExpModel (dense output) [ours]	0.683	-
ExpModel (last conv output) [ours]	0.667	-
ExpModel (combined) [ours]	**0.700**	0.715

Experimental Results: ExpModel



Cross-database Validation: tested on CK+

Conclusion

- We have shown the power of LSTM Networks to exploit sequential information for facial analysis in the wild.
- FaceVideoModel achieved 98.70% face verification on YTF database which is the best performance in the benchmarking exercises.
- ExpModel reported effective performance for spontaneous facial expression recognition on the FER2013, and our it can yield good results on controlled CK+ database even with more diverse wild training set.
- The proposed systems have the potential value within the computer vision community for more effectively managing unconstrained facial analysis applications.