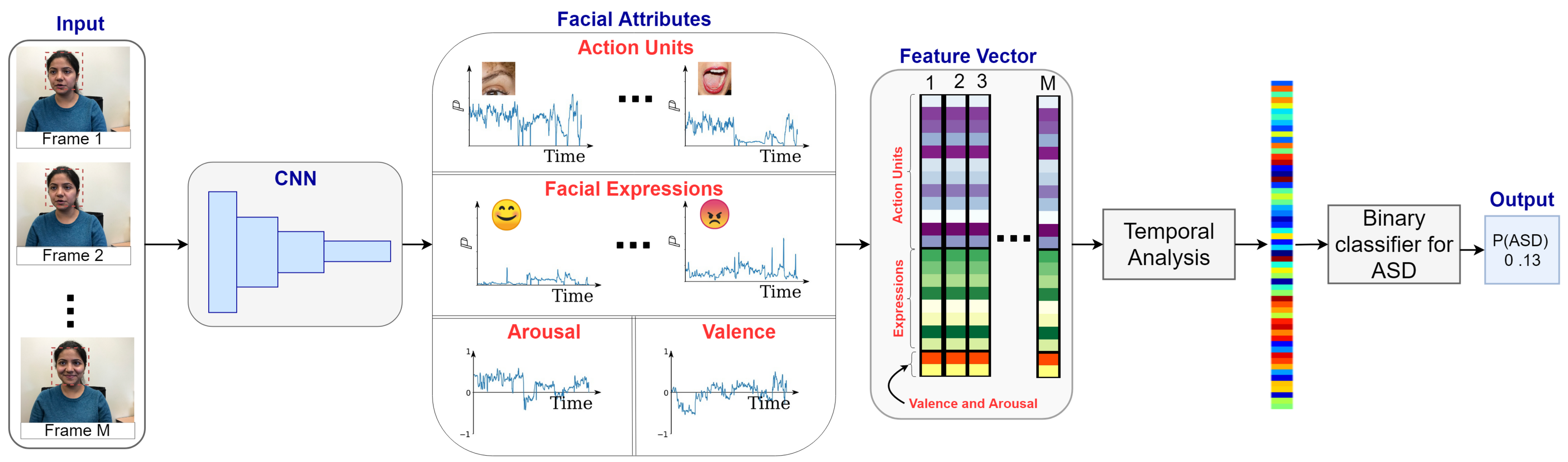
A Facial Affect Analysis System for Autism Spectrum Disorder

Beibin Li, Sachin Mehta, Deepali Aneja, Claire Foster, Pamela Ventola, Frederick Shic, Linda Shapiro

Contact: beibin@uw.edu



Overview of our end-to-end system for autism spectrum disorder (ASD) classification using facial affect attributes.

- **Multi-task learning improves facial affect attributes recognition.**
- **The addition of different affect attributes improves ASD/nonASD classification.**

CNN Unit	# Params	FLOPs	Expr (F1)	AU (mF1Acc)	Val (CC)	Aro (CC)
Single-task						
Bottleneck [1]	25.9 M	3.4 B	0.56	**0.78**	0.63	0.54
MobileNet [2]	24.8 M	3.1 B	0.57	0.77	0.64	0.52
EESP [3]	9.7 M	1.2 B	0.57	0.76	0.64	0.52
Multi-task						
Bottleneck [1]	6.5 M	0.85 B	**0.58**	0.75	0.68	0.61
MobileNet [2]	6.2 M	0.78 B	**0.58**	0.75	0.68	**0.62**
EESP [3]	**2.4 M**	**0.29 B**	**0.58**	0.75	**0.69**	0.61

Performance of our Affect Analysis system on AffectNet and EmotioNet.

We introduce an end-to-end machine learning-based system for classifying autism spectrum disorder (ASD) using facial affect attributes such as expressions, action units, arousal, and valence. Our system classifies ASD using representations of different facial affect attributes from convolutional neural networks, which are trained on *images in the wild*. Our experimental results show that different facial affect attributes used in our system are statistically significant and improve sensitivity, specificity, and F1 score of ASD classification by a large margin.

Introduction

Along with two widely used categorical facial affect attributes (facial expressions and AUs) for natural images, our system also predicts two continuous facial *affect* attributes (arousal and valence) that have been found to be effective in autism related clinical studies [4]. For simplicity, we use facial affect attributes to represent **facial expression**, **AUs**, **arousal**, and **valence**. Since there are no publicly available datasets for autism with *all* of these different attributes, we learn representations for these attributes by leveraging two large-scale facial datasets of natural images that are collected in a wide variety of settings, including age, gender, race, pose, and lighting variations. The contributions of this work are:

1. Present an ASD classification system based on facial affect attributes
2. Show the importance of these attributes in improving the performance of our system through statistical analysis
3. Analysis of single vs. multi-task learning for facial affect attribute recognition

Training the Affect Recognition System

- We use public datasets collected in the wild to train our facial affect recognition system. We choose **AffectNet** and **EmotioNet** with 1.2 million images combined.
- MTL loss function: $\arg\min \sum_{t=1}^{T} \sum_{i=1}^{N} l(y_i^t, f(\boldsymbol{x}_i^t; \boldsymbol{w}^t))$

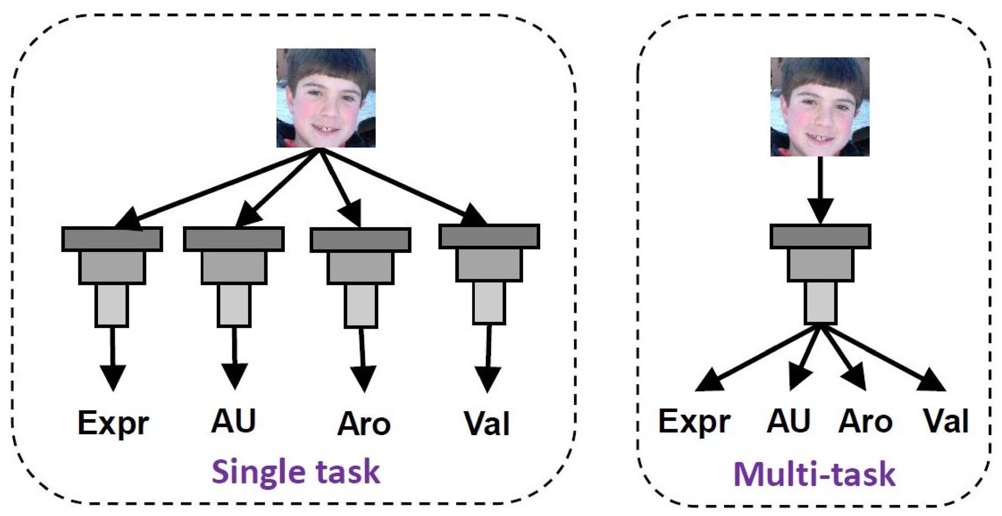


Figure 1: Single v.s. **Multi-Task Learning**
Expr = Expression, AU = Action Units,
Val = Valence, Aro = Arousal.

Data Collection from iPad Experiment

- We collect a video dataset of 105 children with one 720p 24 Hz video per participant using an iPad application; 88 of these children (ASD: 49; non-ASD: 39) finish the experiment and consent to use their data for our research.
- Each participant watches an expert-designed video stimulus on an iPad. While the participant watches a video, our application captures and records the participant's facial response using the iPad's front camera. The video recorded using the iPad application is about 6 minutes and 35 seconds (9,575 valid frames) per participant.

Inference Pipeline

1. Use Multi-Task Learning (MTL) and Convolutional Neural Network (CNN) to extract facial affect attributes (expression, facial action unit, arousal, valence) for each frame of recorded videos.
2. Extract temporal features on each facial affect attribute for each participant.
3. Use temporal features on these affect attributes to classify children with and without autism.

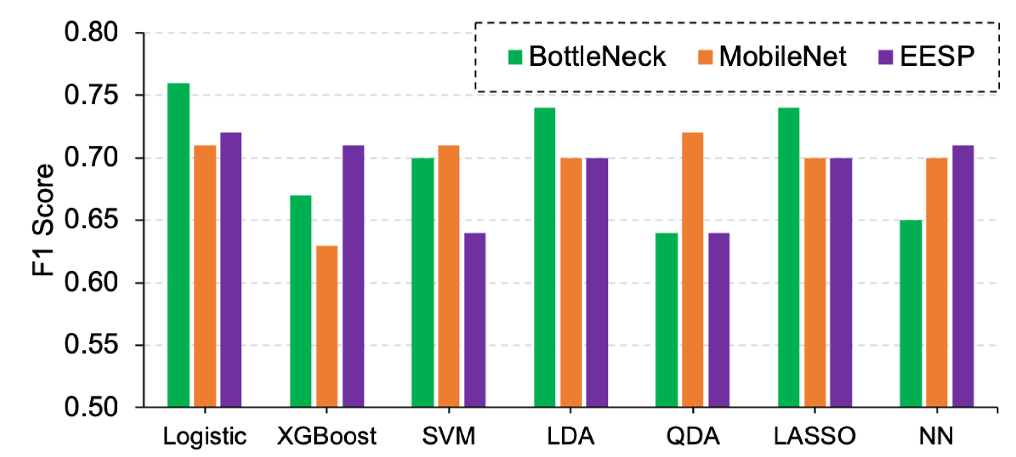


Figure 2: Machine Learning Performance for **ASD/nonASD Classification**

Facial affect attributes				F1	Sensitivity	Specificity
AU	Aro	Val	Expr			
✓				0.69	0.69	0.62
✓	✓			0.72	0.71	0.67
✓	✓	✓		0.69	0.67	0.67
✓	✓	✓	✓	**0.76**	**0.76**	**0.69**

Table 1: Results from Logistic Regression to classify ASD and nonASD with different combination of facial affect attributes

Limitation

- Expression $\neq$ Emotion. Expression recognition is a hard task, and even human would not agree with each other.
- This is only a proof-of-concept study with only 88 valid participants. More clinical research are needed to fully study the facial behavior for children with autism.
- There are $2^4 = 64$ different combinations of these 4 facial affect attributes, and only analyzing 4 is not comprehensive. Using all 4 domains might overfit our dataset because n (the number of participants) is small.
- Interpretability and uncertainty are important in clinical application, but they are not fully studied in CNN.

Takeaway Message

- **MTL can improve the performance for facial affective attribute recognition.**
- **Our system outperform SOTA in arousal and valence prediction.**
- **The affect attributes are statistically significant between children with and without ASD.**
- **Representations of different facial affect attributes improve the ASD classification performance by about 7% with F1 score of** 76%.

References

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[2] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[3] Sachin Mehta, Mohammad Rastegari, Linda Shapiro, and Hannaneh Hajishirzi. Espnetv2: A light-weight, power efficient, and general purpose convolutional neural network. *arXiv preprint arXiv:1811.11431*, 2018.

[4] Ognjen Rudovic, Jaeryoung Lee, Miles Dai, Bjorn Schuller, and Rosalind Picard. Personalized machine learning for robot perception of affect and engagement in autism therapy. *Science. 3. 10.1126/scirobotics.aao6760.*, 2018.

Acknowledgements

This work is supported by NIH awards K01 MH104739, R21 MH103550; the NSF Expedition in Socially Assistive Robotics #1139078; and Simons Award #383661. Special thanks to Shun-Tak Leung and his Faithful Steward Endowed Fellowship.