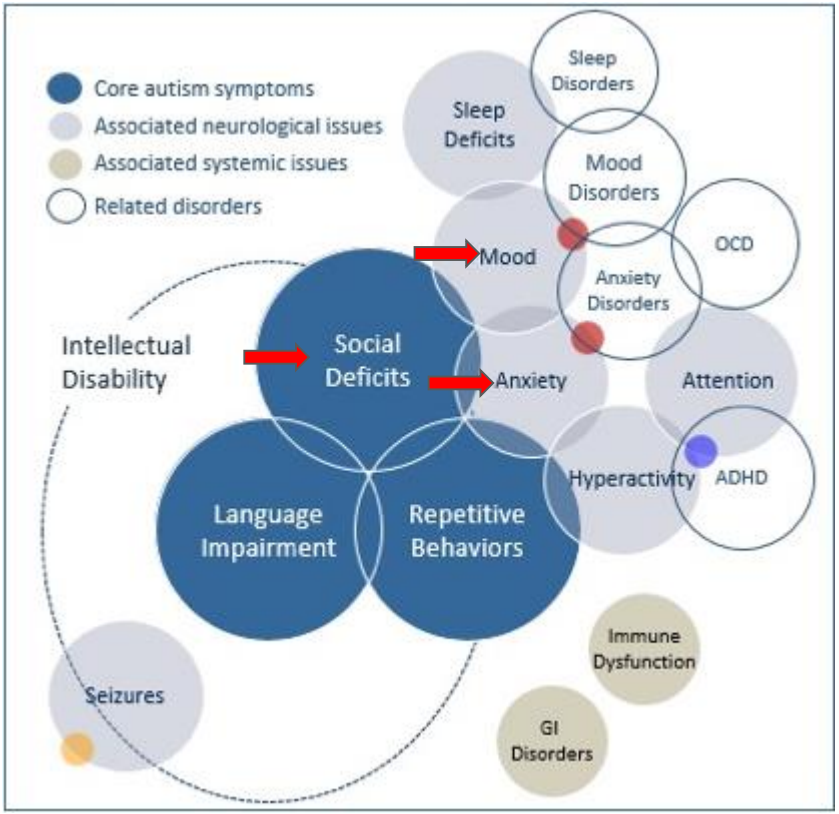
A Facial Affect Analysis System for Autism Spectrum Disorder

https://arxiv.org/abs/1904.03616

Beibin Li

Symptoms of Autism (Motivation)



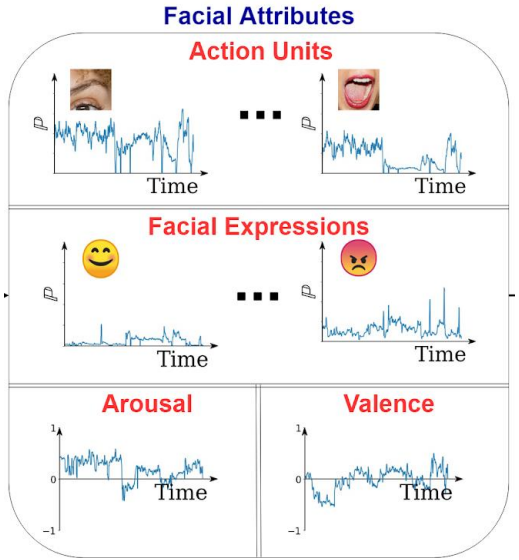
https://www.autismspeaks.org/what-are-symptoms-autism

Goal

- Create a system for **wild data**. Analyze facial images from a various setting (illumination, pose, occlusion, etc).
- Combine and analyze **four domains** of facial expression.
- Use **machine learning** and facial affect attributes to classify participants with/without autism.

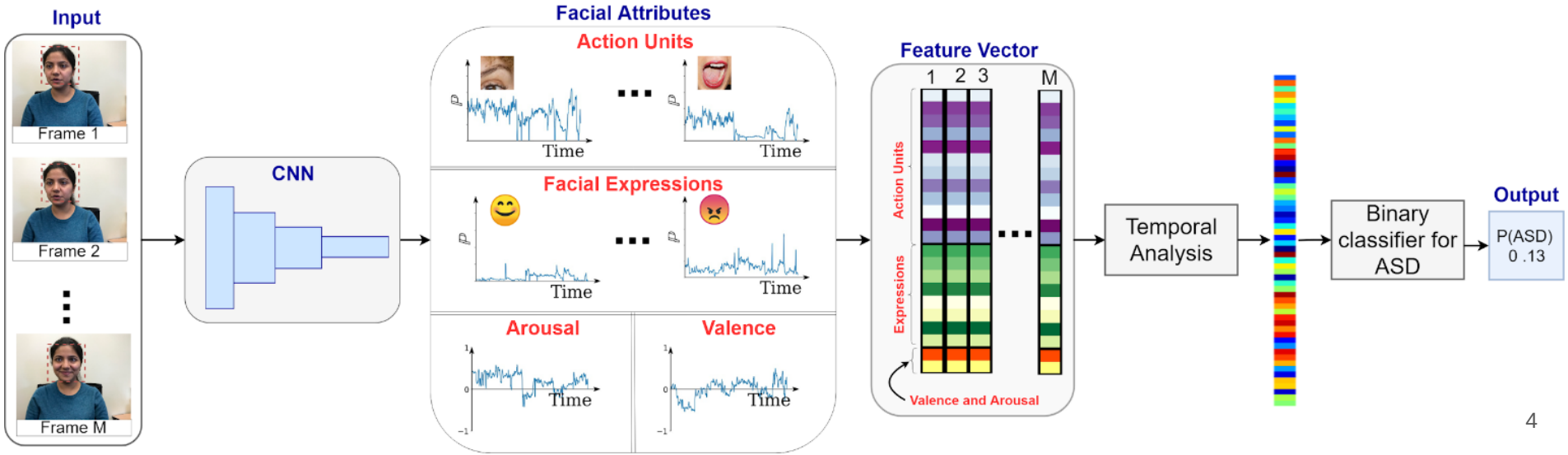
Constraint Setting

Wild Setting



Our System

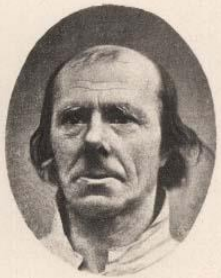
- Input: Video. No other modality (e.g. voice, EEG, heart rate, EDA, age, etc)
  - Comply with IRB/HIPAA restrictions.
  - Make the system as simple as possible.
  - More challenging.
- Output:
  - Recognized facial attributes: action units, expression, arousal, and valence.
  - Probability that the participant is influenced by autism.



Part I: Facial Affect Analysis

Facial Expression != Emotion

1



4

2

5

3

6

THE

EXPRESSION OF THE EMOTIONS

IN

MAN AND ANIMALS.

By CHARLES DARWIN, M.A., F.R.S., &c.

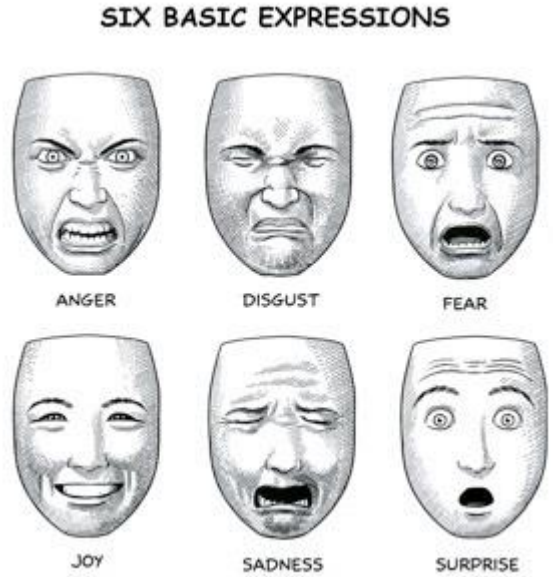
WITH PHOTOGRAPHIC AND OTHER ILLUSTRATIONS.

NEW YORK:
D. APPLETON AND COMPANY,
549 & 551 BROADWAY.
1873.

The expression of the emotions in man and animals

Charles Darwin

Cardinal Expressions



Total 8 Expression Category

- Anger, Disgust, Fear, Joy, Sad, Surprise
- Neutral
- Contempt

Pixel-wise Difference is Small

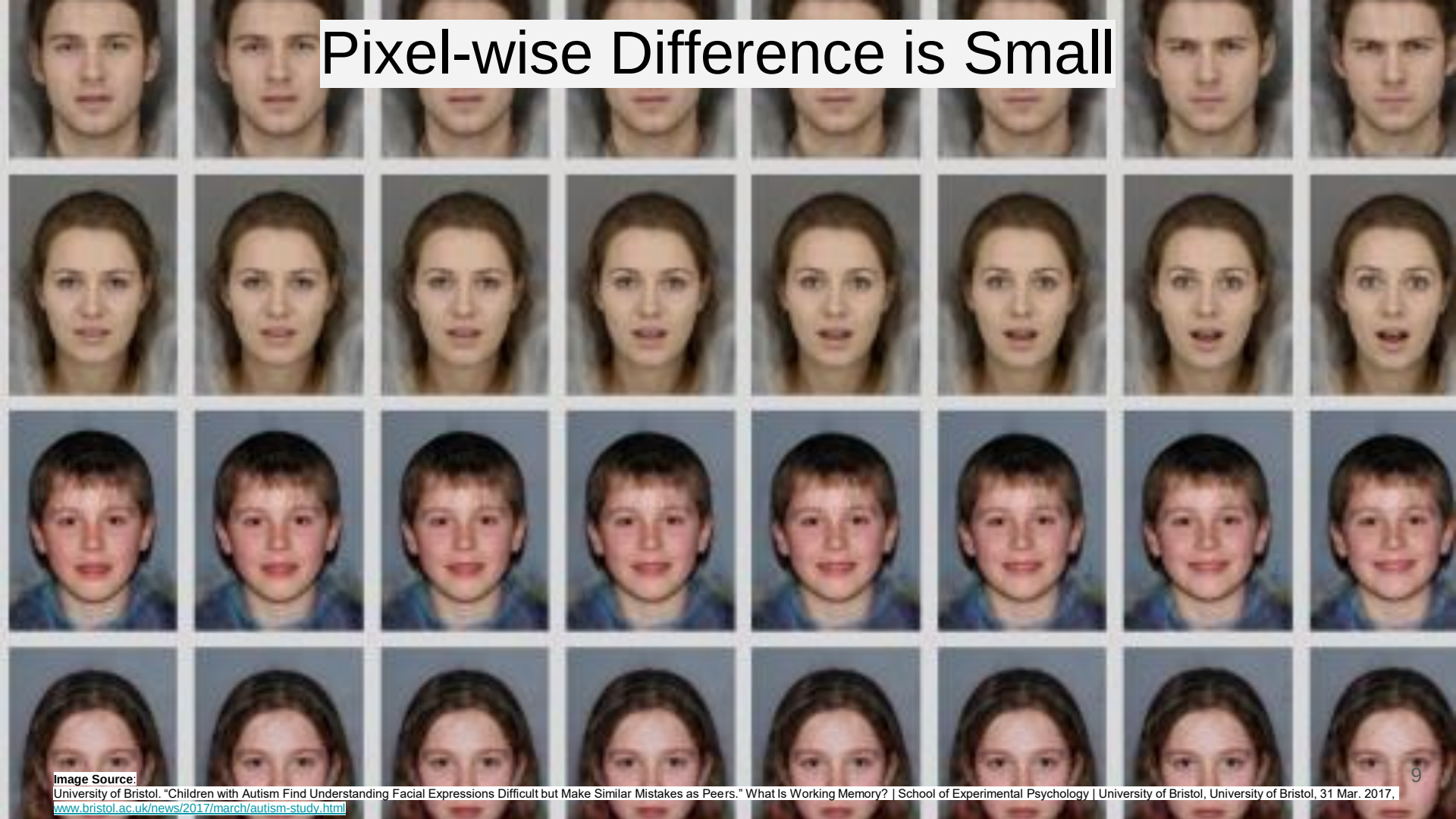


Image Source:
University of Bristol. “Children with Autism Find Understanding Facial Expressions Difficult but Make Similar Mistakes as Peers.” What Is Working Memory? | School of Experimental Psychology | University of Bristol, University of Bristol, 31 Mar. 2017, www.bristol.ac.uk/news/2017/march/autism-study.html

Even human wouldn’t agree with each other

Crowd-sourced labels on “expressions” exhibit only 65 ± 5 % accuracy

- Goodfellow, Ian J., et al. "Challenges in representation learning: A report on three machine learning contests." *International Conference on Neural Information Processing*. Springer, Berlin, Heidelberg, 2013.
- Barsoum, Emad, et al. "Training deep networks for facial expression recognition with crowd-sourced label distribution." *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. ACM, 2016.

36,000 images were annotated by two annotators... The results showed that the annotators agreed on 60.7% of the images.

- Mollahosseini, Ali, Behzad Hasani, and Mohammad H. Mahoor. "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild." *IEEE Transactions on Affective Computing* (2017).

Action Units

Friesen, E., and P. Ekman. "Facial action coding system: a technique for the measurement of facial movement." Palo Alto (1978).

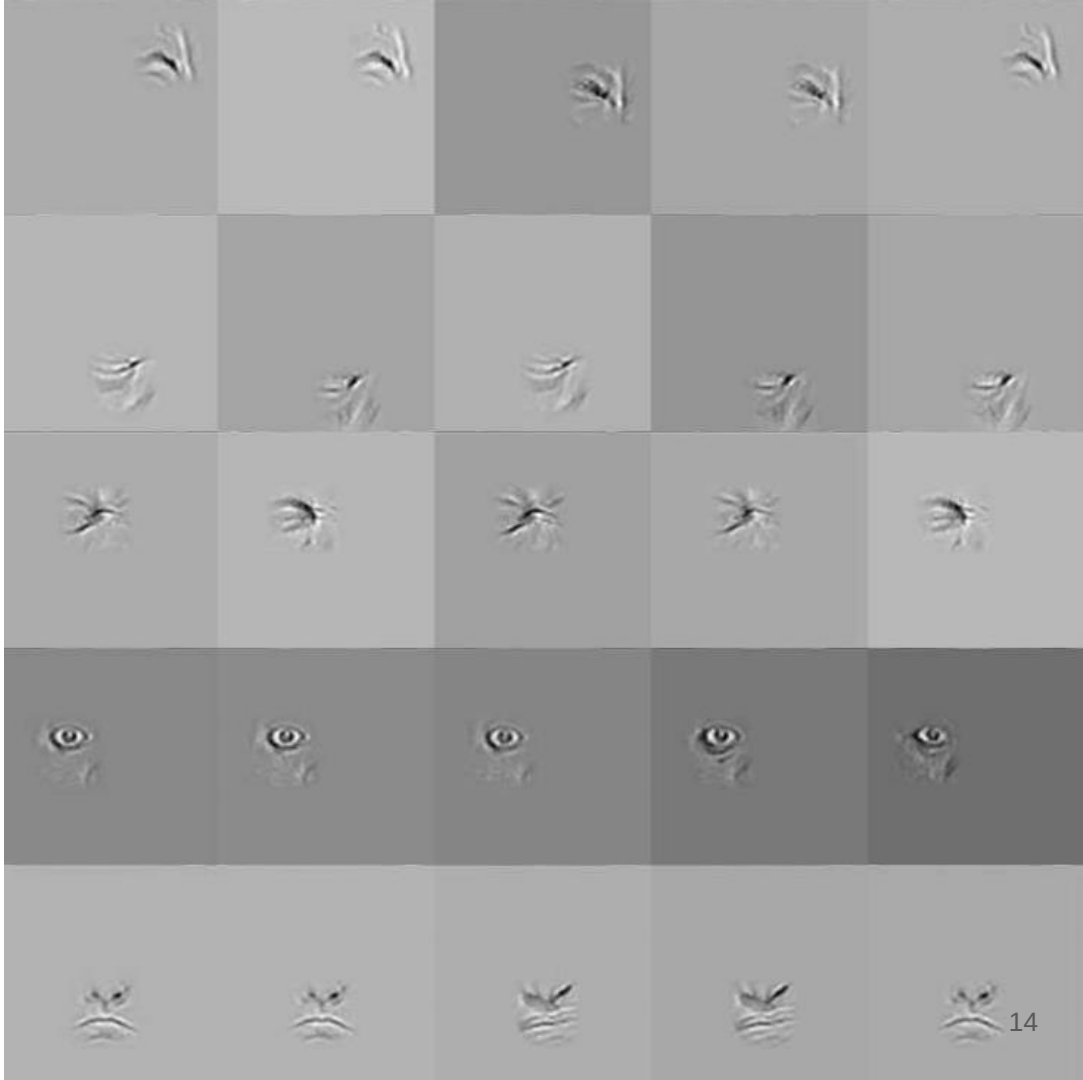
Image Source: https://www.pinterest.com.au/pin/82472236907938447/

Cons of Action Units

- Expensive to annotate
- Small muscle movements (winkles) can make a difference
- Same AU can result to different expression

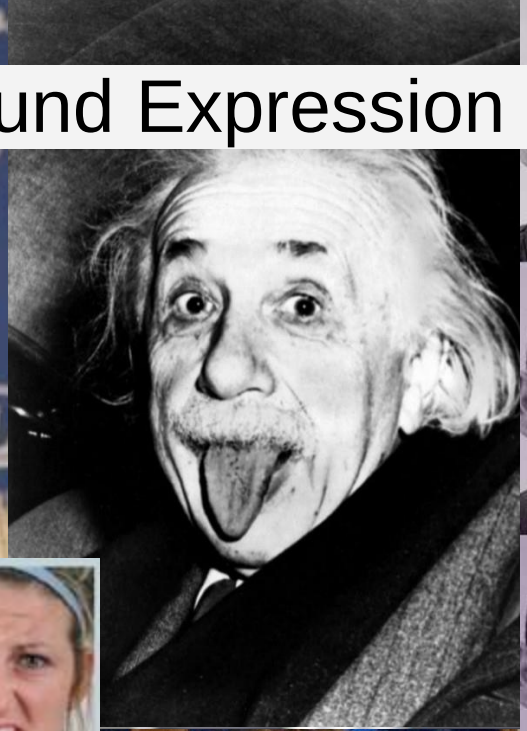
Same Action Units, but ...

Classify Expression Directly

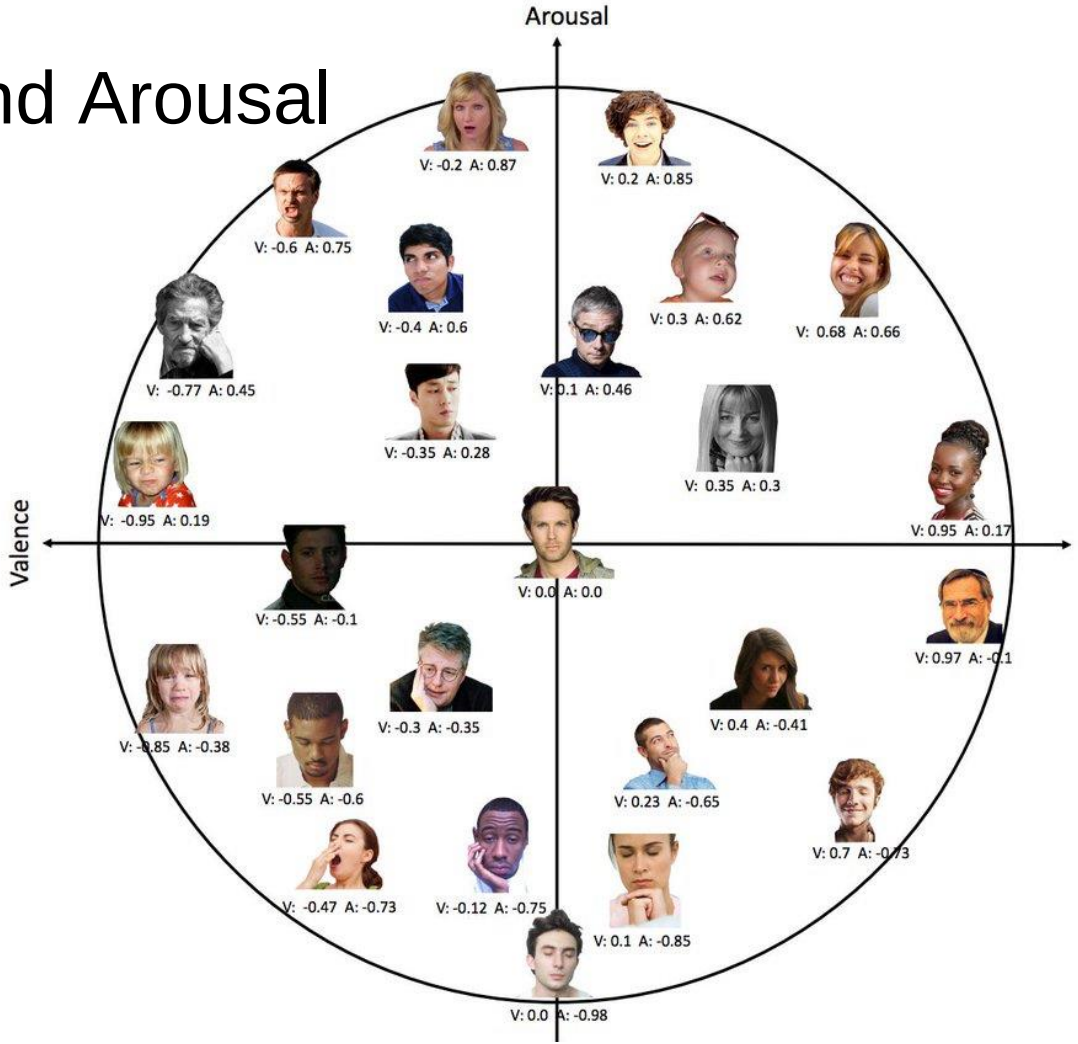


Khorrami, Pooya, Thomas Paine, and Thomas Huang. "**Do deep neural networks learn facial action units when doing expression recognition?**." *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2015.

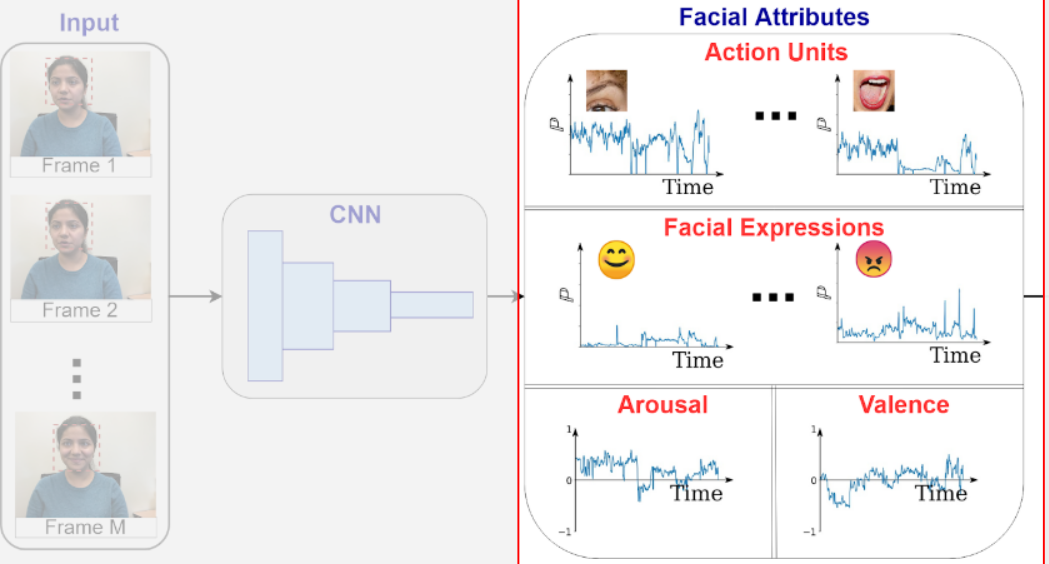
Compound Expression & Micro Expression



Valence and Arousal



Kensinger, Elizabeth A. "Remembering emotional experiences: The contribution of valence and arousal." *Reviews in the Neurosciences* 15.4 (2004): 241-252.



Four Domains of Facial Affect Attributes

- **Action Units**: major muscle movements.
- **Expression**: 8 classes
- **Arousal**: how intense an expression is
- **Valence**: how pleasant an expression is

Dataset

- ***EmotioNet***
  - 975,000 images in the wild
- 12 Important Action Units:
  - AU1 Inner Brow Raiser
  - AU2 Outer Brow Raiser
  - AU4 Brow Lowerer
  - AU5 Upper Lid Raise
  - AU6 Cheek Raise
  - AU9 Nose Wrinkler
  - AU12 Lip Corner Puller
  - AU17 Chin Raiser
  - AU20 Lip stretcher
  - AU25 Lips Part
  - AU26 Jaw Drop
  - AU43 Eyes Closed

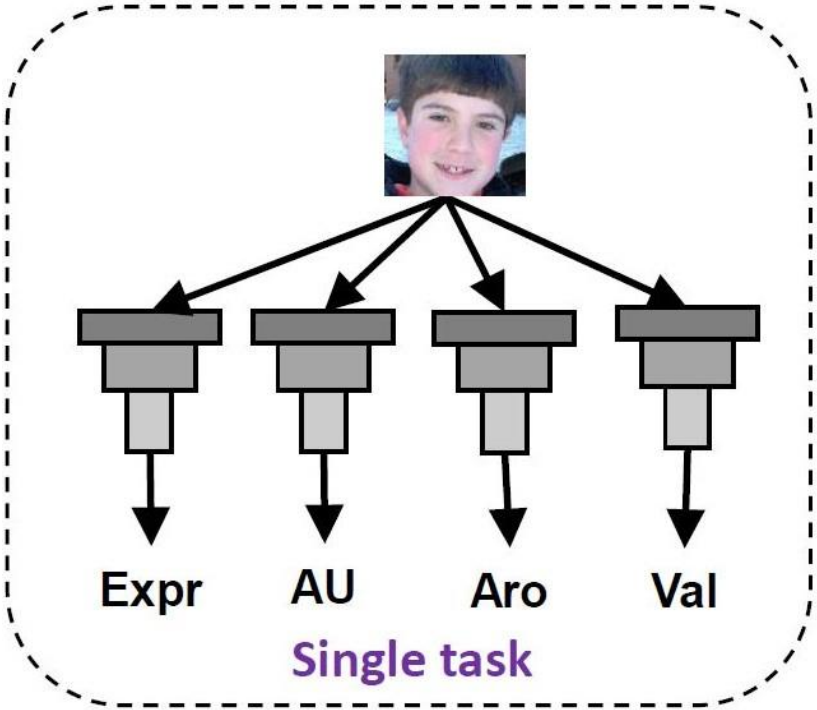
- ***AffectNet***
  - 450,000 images
- Eight Expressions
- Arousal
- Valence

Total: **1.4 million** images

Four Blocks of Convolution

Layer/ Stride	Repeat	Output	
		Size	Channels
Conv-3/2	1	112×112	32
CU/2	1	56×56	32
CU/1	1	56×56	32
CU/2	1	28×28	64
CU	3	28×28	64
CU/2	1	14×14	128
CU	7	14×14	128
CU/2	1	7×7	256
CU/1	3	7×7	256
DWConv-3/1	1	7×7	512
Avg. pool		1×1	512
Linear $\times 4$		$C_{expr}, C_{au}, C_{val}, C_{aro}$	

Multi-Task Learning



Engineering Details

- Missing Labels
  - Weight loss with zero
- Loss
  - Action Units Recognition: Weighted Binary Cross Entropy Loss
  - Expression Classification: Weighted Cross Entropy Loss
  - Arousal / Valence: Euclidean + Manhattan Distance
- Regularization
  - 20% Dropout before last layer.
- Stochastic Gradient Descent
  - Initial Learning Rate: 0.01; 10% decay per epoch.
  - Momentum 0.9.
  - Total 30 epochs.

Results

CNN Unit	# Params	FLOPs	Expression (F1)	AU (mF1Acc)	Valence (CC)	Arousal (CC)
			Single-task			
Bottleneck	25.9 M	3.4 B	0.56	**0.78**	0.63	0.54
MobileNet	24.8 M	3.1 B	0.57	0.77	0.64	0.52
EESP	9.7 M	1.2 B	0.57	0.76	0.64	0.52
			Multi-task			
Bottleneck	6.5 M	0.85 B	**0.58**	0.75	0.68	0.61
MobileNet	6.2 M	0.78 B	**0.58**	0.75	0.68	**0.62**
EESP	**2.4 M**	**0.29 B**	**0.58**	0.75	**0.69**	0.61
			Literature			
SOTA	-	-	0.58	-	0.66	0.54
Human Performance	-	-	0.61	*	0.82	0.57

0.82 kappa for CK+ dataset

- Benitez-Quiroz, Carlos Fabian, Yan Wang, and Aleix M. Martinez. "Re with Deep Nets and a New Global-Local Loss." *ICCV*. 2017.
- Benitez-Quiroz, C. Fabian, et al. "EmotioNet Challenge: Recognition of facial expressions of emotion in the wild." *arXiv preprint arXiv:1703.01210* (2017).
- Mollahosseini, Ali, Behzad Hasani, and Mohammad H. Mahoor. "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild." *IEEE Transactions on Affective Computing* (2017).

Part II: Application to Autism

Data Collection

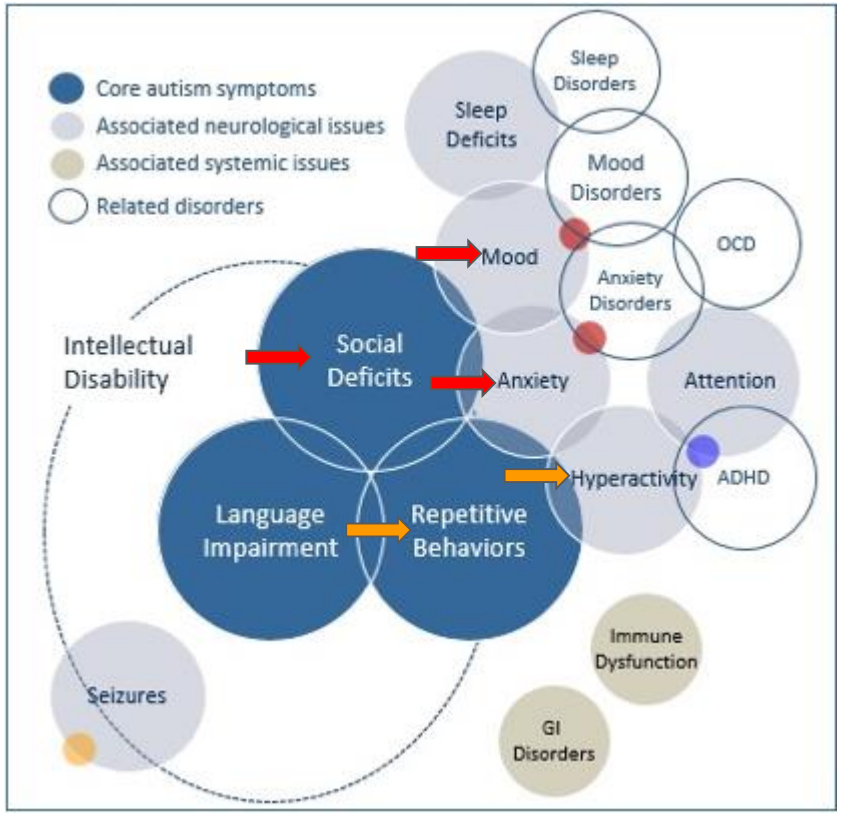
- About 24 Frames per second for recording
- Pause experiment if no face detected by iPad
- Valid Data: 88 participants finished the experiment and signed consent.

Statistical Analysis

- ASD has more joy but less neutral compared to TD.
  - ASD may like the stimulus more than the TD.
- Expression/Arousal/Valence/Head movements are more various for children with ASD.

Facial attributes	**p-value**
Action Units (AUs)	0.223
Arousal (Aro)	0.007
Valence (Val)	0.001
Expression (Expr)	0.006

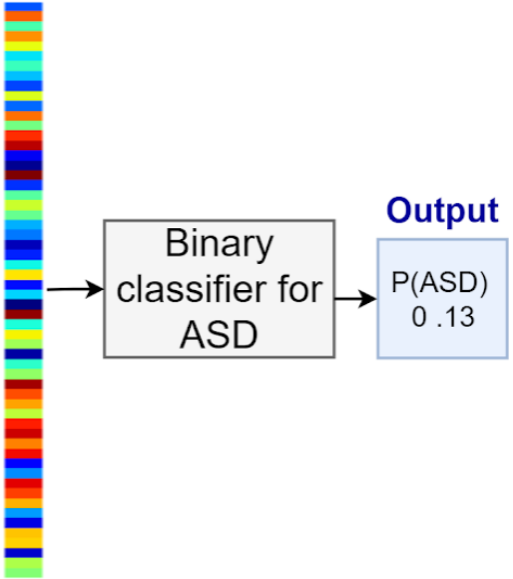
Symptoms of Autism (Revisit)



https://www.autismspeaks.org/what-are-symptoms-autism

ASD/TD Classification

- Use the 58-dimensional vector to classify ASD/TD
- Seven binary classifiers
  - Logistic regression, LASSO, LDA, QDA, SVM-rbf, XGBoost, and a two-hidden-layer neural network (NN).
- Challenge:
  - We have 88 valid participants, and 58-dimensional vector might overfit our data.
- Solution:
  - Use PCA to reduce dimension first, then apply classifiers.
  - Use default hyperparameters for all models.



Classification Result v.s. Affect Domain

Facial attributes				F1	Sensitivity	Specificity
AU	Arousal	Valence	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**

Limitation

- < 100 participants, and participants are not i.i.d from the population.
- The analysis is restricted by the training data and public facial images
- Lack of testing data from other source, and results might be **too optimistic**!
- Need to correlate affective attributes to the ADOS score.

Takeaway

- Create a system for **wild data** (facial images from a various setting).
- Combine and analyze **four domains** of facial expression.
- Extracted features are **statistically significant** for ASD/TD groups.
- Even with simple features, our machine learning results showed there is **potential** to use facial affect analysis to help classify ASD/TD in the future.