Face Social Traits Rating Prediction & Visualization Using Deep Model

Na Zhang

What is social trait of faces?

- People often make trait judgments about unfamiliar others based on their faces in the absence of context or other information
- such as forming an impression that someone looks friendly, trustworthy, or strong

Example of social traits

Friendly

Competent

ASD

- Individuals with Autism Spectrum Disorder (ASD) often have difficulties reading social information from faces
- In this work, we try to study how do the people with ASD make facial trait judgement

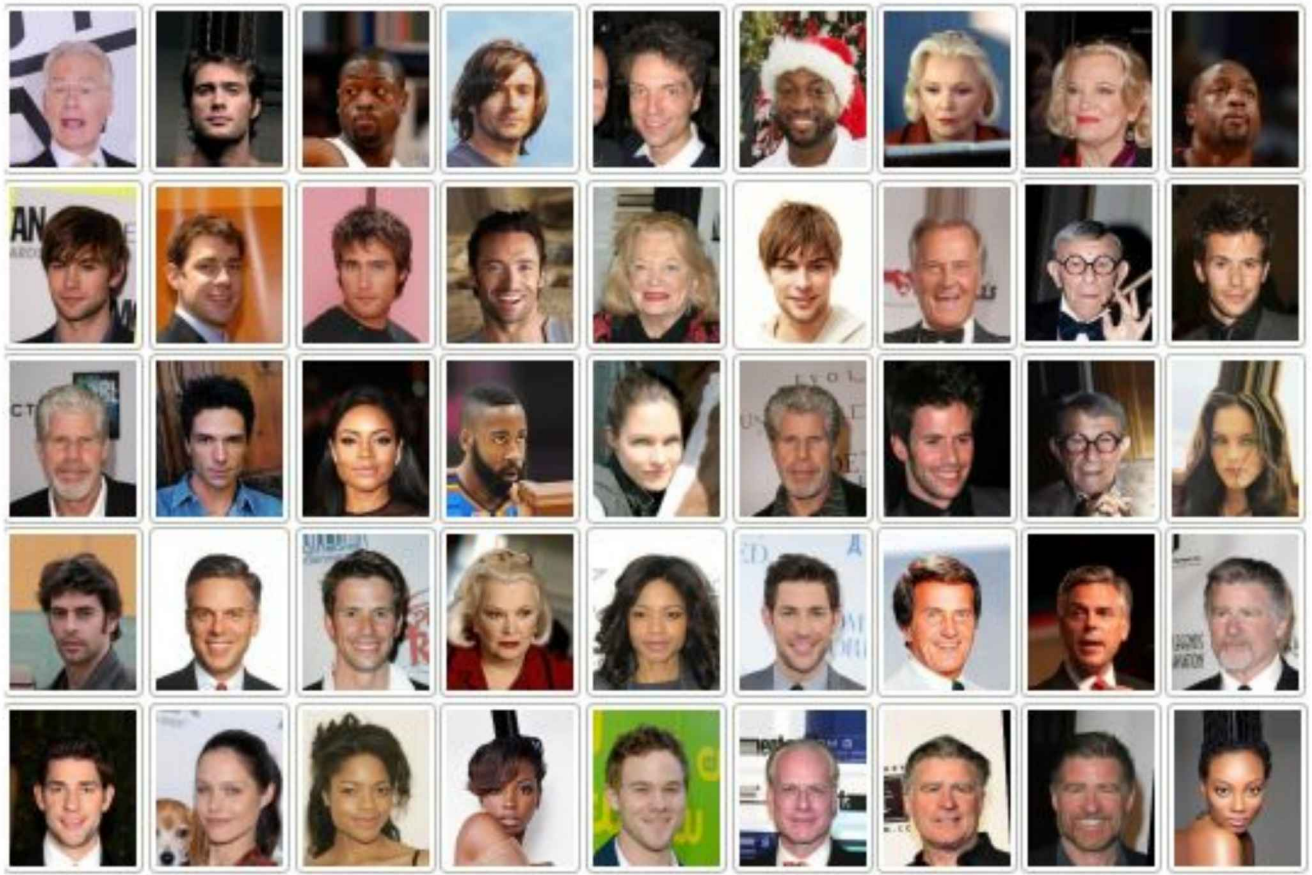
Data

- Contains 500 faces of 50 identities
- 10 face images per identity

Face Samples

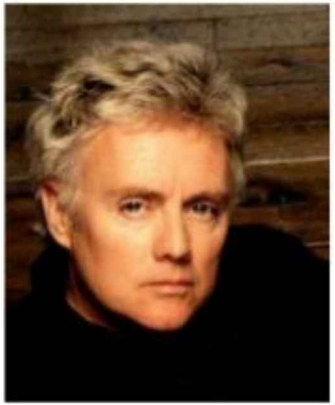
Multiple variations:

- facial expression
- pose angels
- lighting conditions
- Makeup
- occlusion (eyeglasses)
- background



Facial Ratings

Face



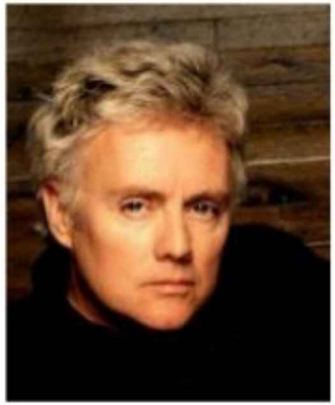
Ratings

	Warm	Critical	Competent	Practical	Feminine	Strong
ASD	3.25581395	4.7804878	5.09756098	4.43902439	3.48717949	3.43902439
NT	2.71428571	5.475	5.02564103	5.02439024	3.11111111	3.46341463

	Youthful	Charismatic	Trustworthy	Dominant	Recognize
ASD	2.88372093	3.71794872	4.0	4.11363636	1.97674419
NT	2.18421053	3.75609756	3.7027027	4.66666667	2.42857143

Choose two traits as examples for analysis

Face

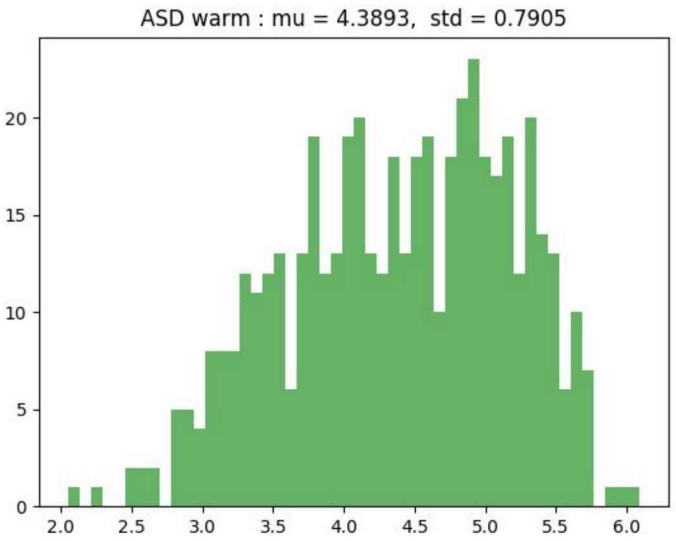


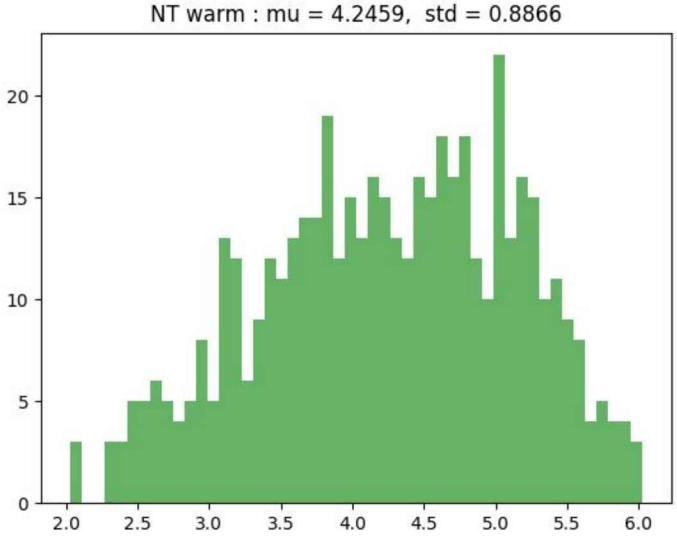
Ratings

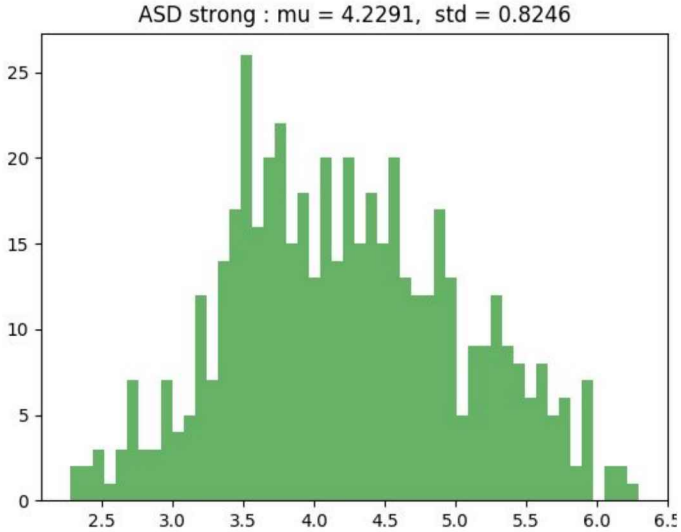
	Warm	Critical	Competent	Practical	Feminine	Strong
ASD	3.25581395	4.7804878	5.09756098	4.43902439	3.48717949	3.43902439
NT	2.71428571	5.475	5.02564103	5.02439024	3.11111111	3.46341463

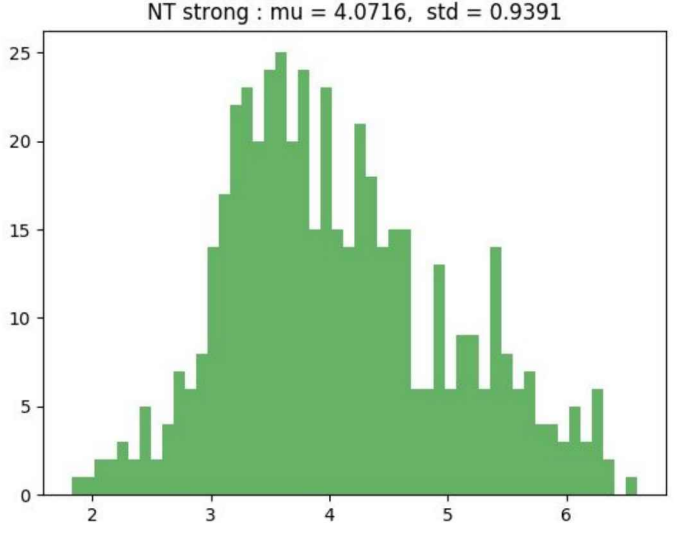
	Youthful	Charismatic	Trustworthy	Dominant	Recognize
ASD	2.88372093	3.71794872	4.0	4.11363636	1.97674419
NT	2.18421053	3.75609756	3.7027027	4.66666667	2.42857143

Ratings Distribution

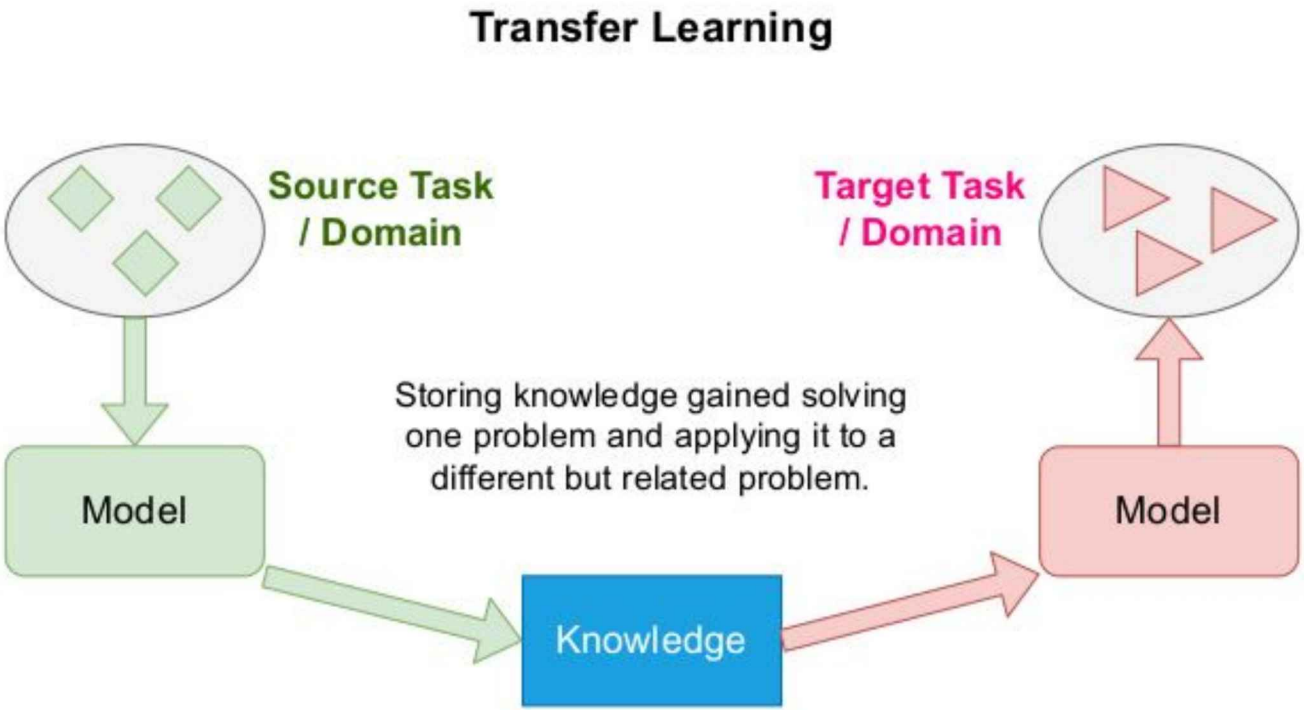


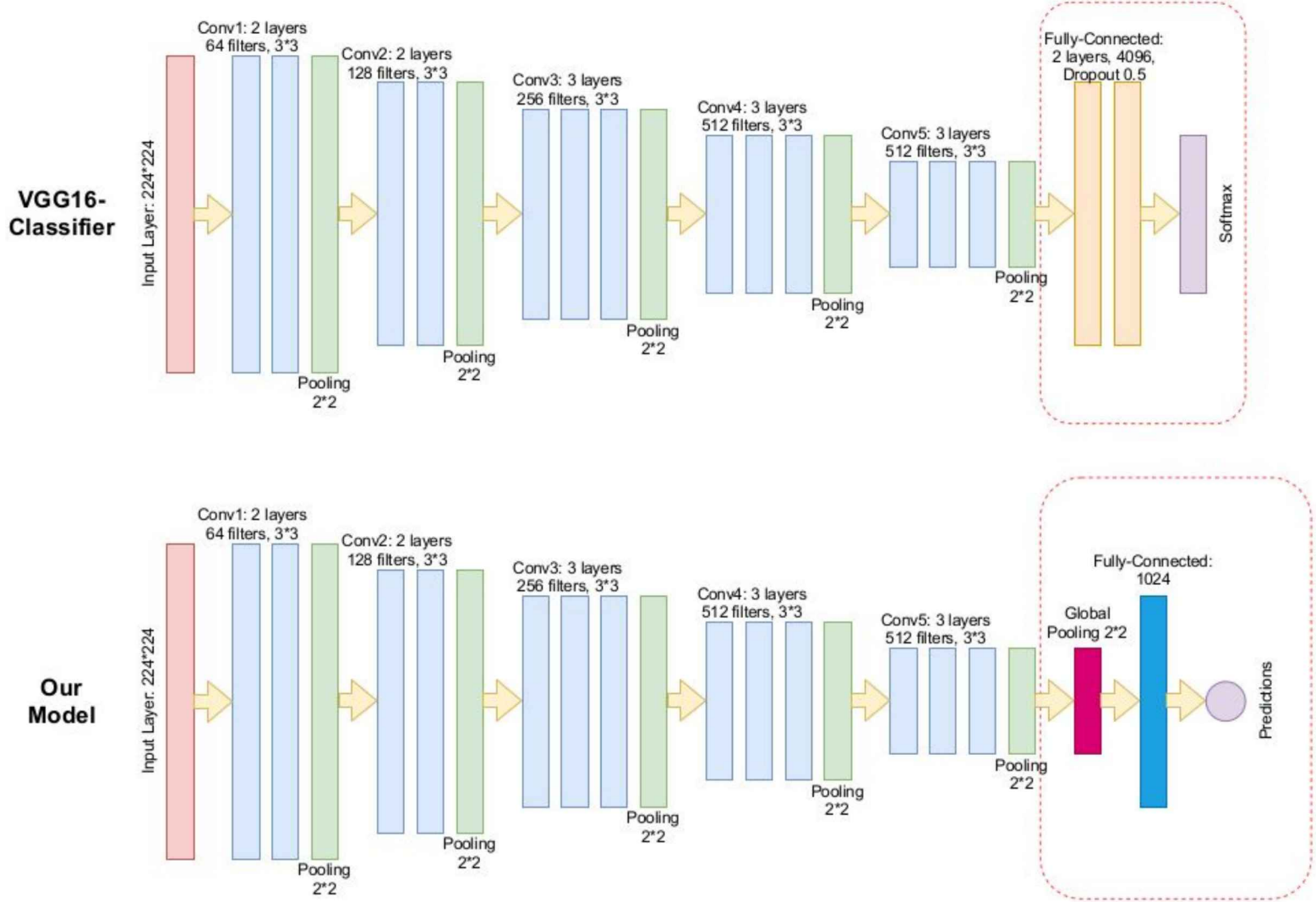




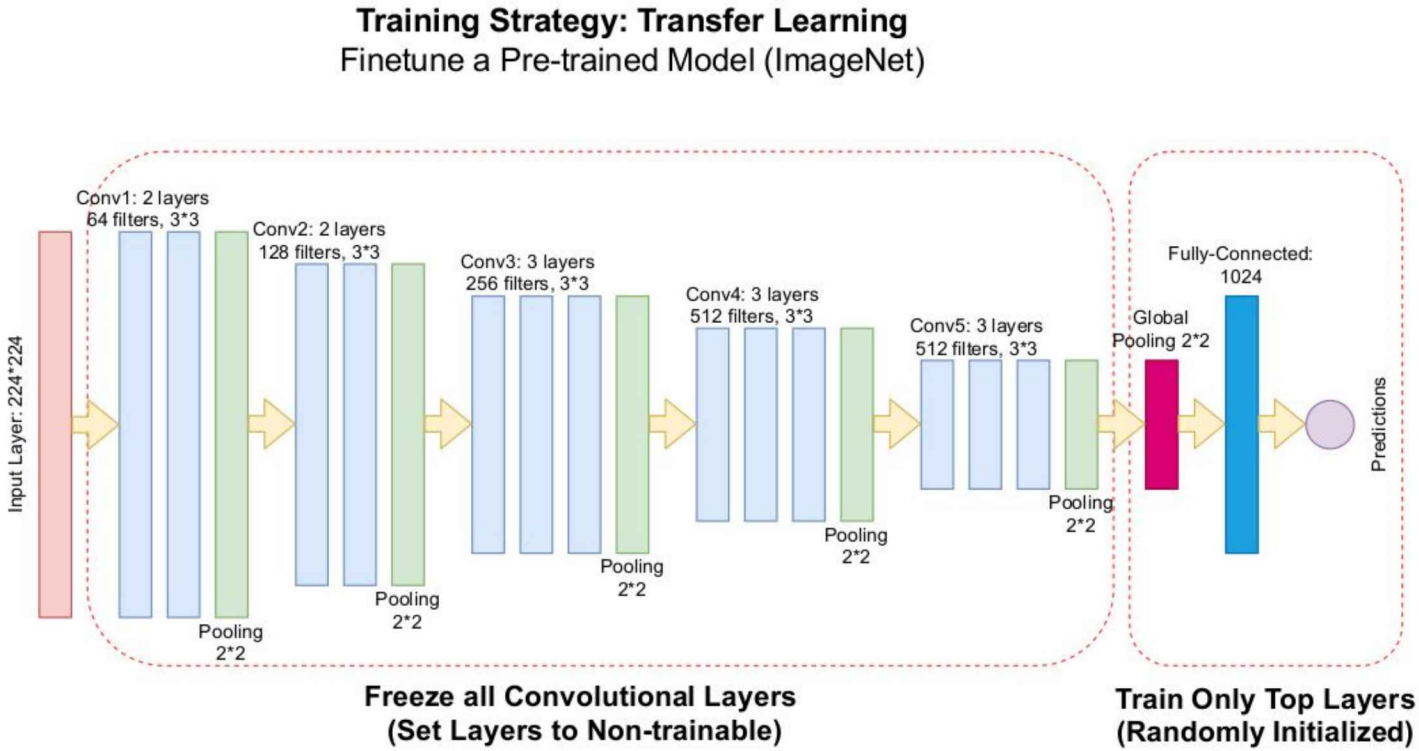


Model

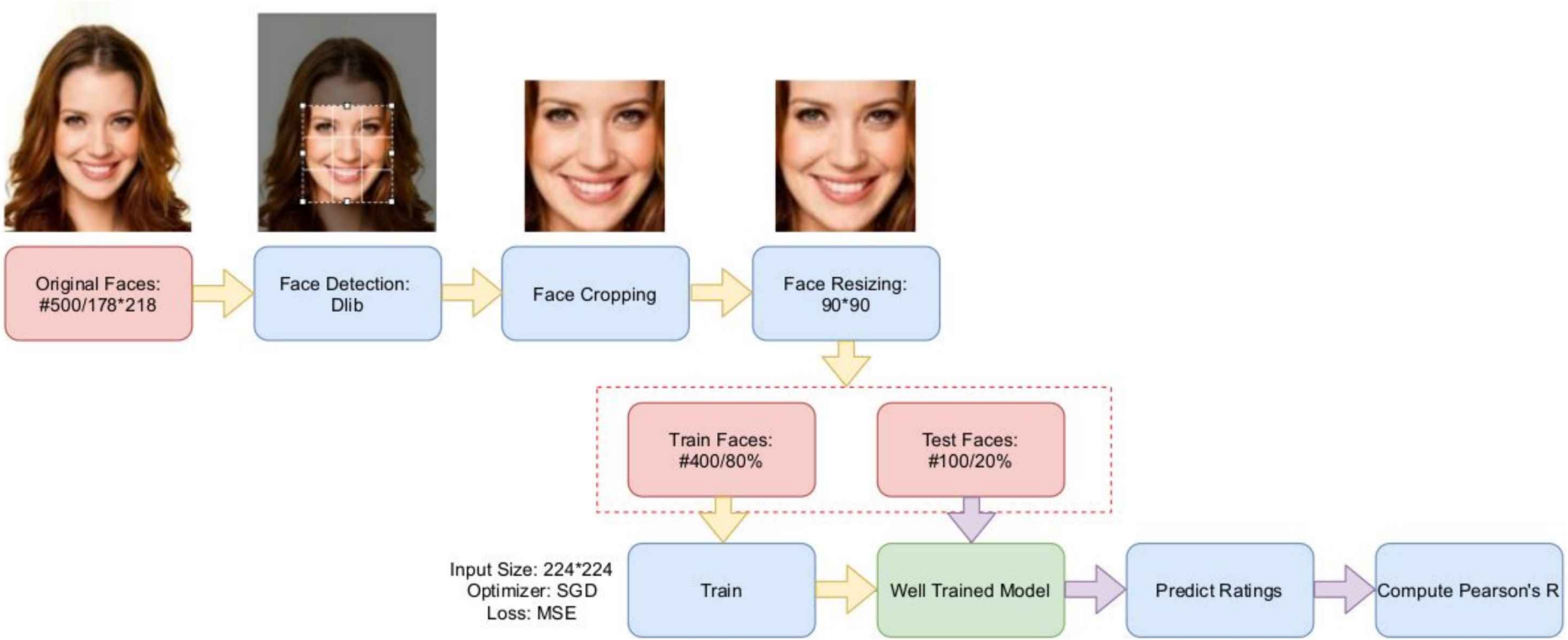

Transfer Learning
Source Task
/ Domain
Target Task
/ Domain
Storing knowledge gained solving one problem and applying it to a different but related problem.
Model
Model
Knowledge


VGG16-Classifier
Input Layer: 224*224
Conv1: 2 layers
64 filters, 3*3
Pooling
2*2
Conv2: 2 layers
128 filters, 3*3
Pooling
2*2
Conv3: 3 layers
256 filters, 3*3
Pooling
2*2
Conv4: 3 layers
512 filters, 3*3
Pooling
2*2
Conv5: 3 layers
512 filters, 3*3
Pooling
2*2
Fully-Connected:
2 layers, 4096,
Dropout 0.5
Softmax
Our Model
Input Layer: 224*224
Conv1: 2 layers
64 filters, 3*3
Pooling
2*2
Conv2: 2 layers
128 filters, 3*3
Pooling
2*2
Conv3: 3 layers
256 filters, 3*3
Pooling
2*2
Conv4: 3 layers
512 filters, 3*3
Pooling
2*2
Conv5: 3 layers
512 filters, 3*3
Pooling
2*2
Global
Pooling 2*2
Fully-Connected:
1024
Predictions

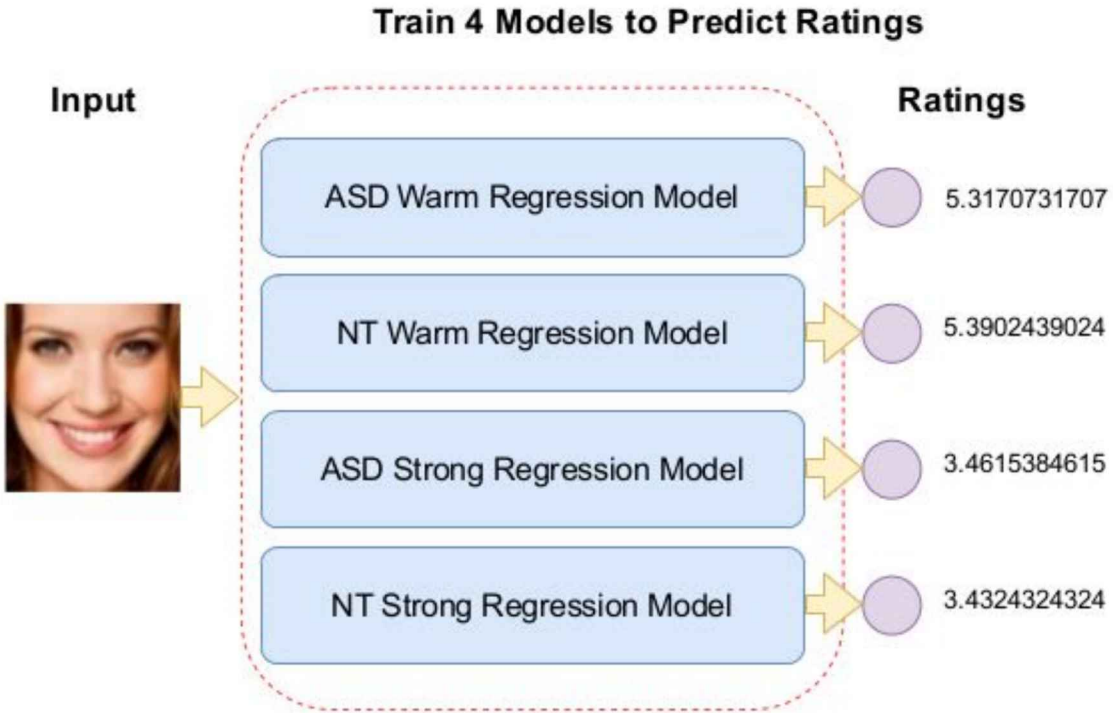
Training Settings



Pipeline


Original Faces:
#500/178*218
Face Detection:
Dlib
Face Cropping
Face Resizing:
90*90
Train Faces:
#400/80%
Test Faces:
#100/20%
Input Size: 224*224
Optimizer: SGD
Loss: MSE
Train
Well Trained Model
Predict Ratings
Compute Pearson's R

4 models are trained



Measure Metrics

- Pearson's r value

Given a pair of random variables (X, Y), the formula for ρ is:

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} \quad \text{(Eq.1)}$$

where:

cov is the covariance

σ_X is the standard deviation of X

σ_Y is the standard deviation of Y

values in [-1, 1]

- 0: no correlation
- <-0.5/>0.5: notable correlation
- others: less notable correlation

5-fold Cross Validation

- Pearson's r is calculated

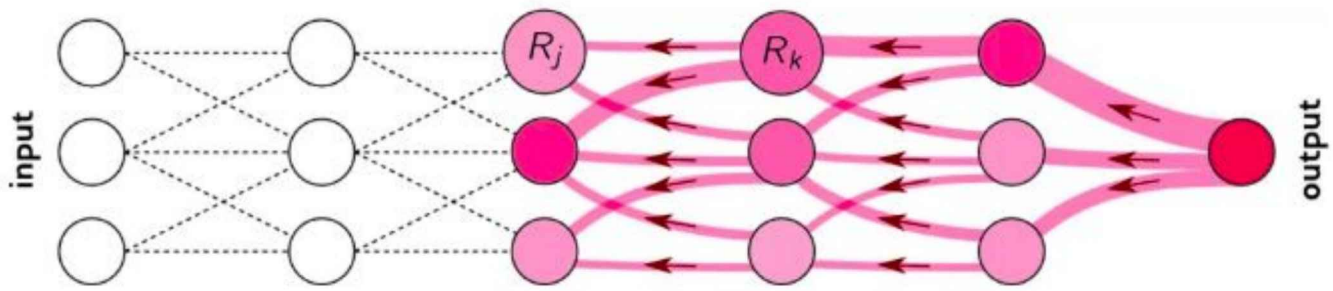
5-Fold	Warm		Strong	
	ASD	NT	ASD	NT
1	0.525	0.528	0.7751	0.8303
2	0.7689	0.7733	0.5413	0.657
3	0.6535	0.6129	0.5129	0.7819
4	0.7688	0.7714	0.5233	0.5214
5	0.5278	0.5376	0.5056	0.5598

- verify (fold 1)

times	Warm		Strong	
	ASD	NT	ASD	NT
1	-0.0544	0.1718	-0.0134	0.0812
2	0.011	0.0756	-0.0899	-0.0843
3	-0.0866	0.1113	0.1526	-0.1095
4	0.1852	0.0447	0.0112	-0.0927
5	0.0516	-0.0871	-0.0627	0.121

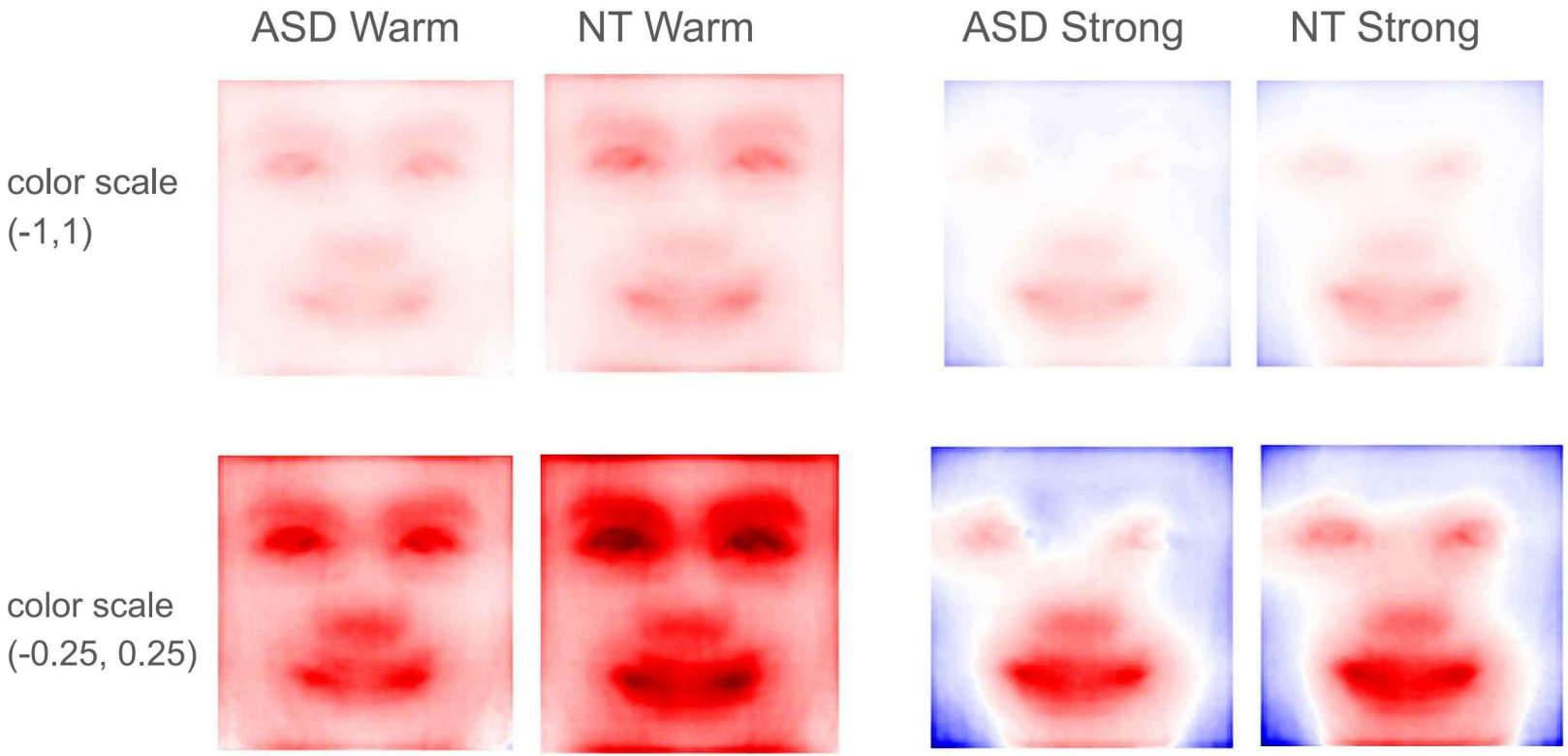
LRP Analysis

- Layer-wise Relevance Propagation (LRP) technique [1]
- highlight which input features deep neural network uses to support its output



[1] Alber, Maximilian, et al. "iNNvestigate neural networks!." *J. Mach. Learn. Res.* 20.93 (2019): 1-8.

Mouth, nose, eyes make big positive contributions



More Examples

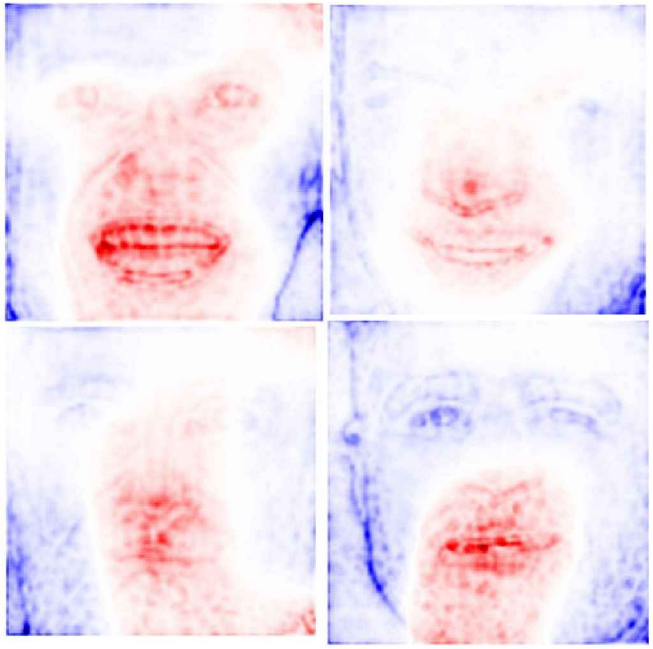
For each model

- we choose several face images with highest ratings and several images with lowest ratings
- to show the LRP result

ASD Strong

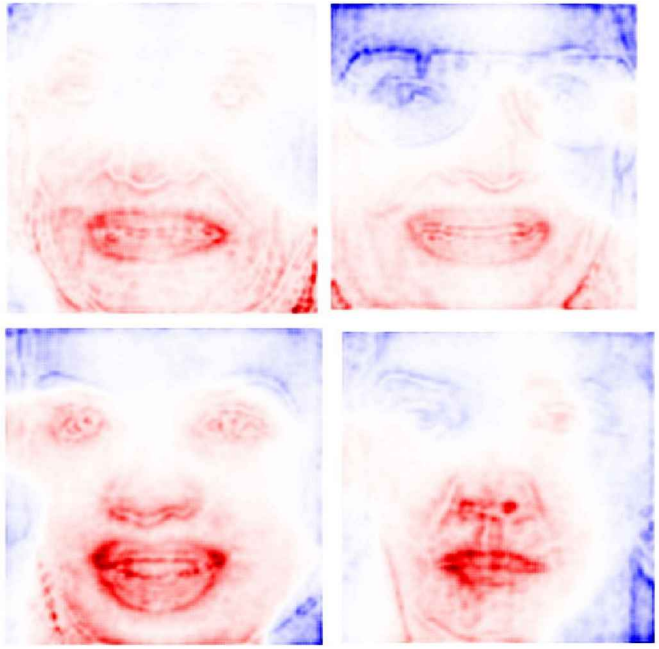
highest ratings

5.783783784 5.769230769 5.756756757 5.666666667



lowest ratings

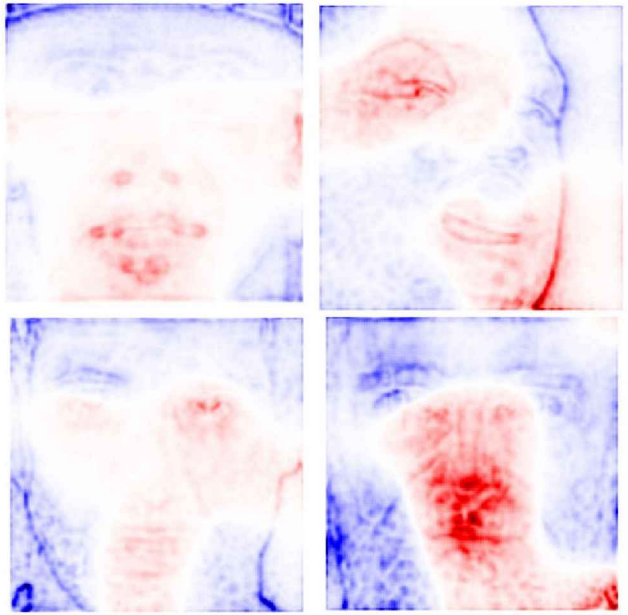
2.297297297 2.690476191 2.711111111 2.72972973



NT Strong

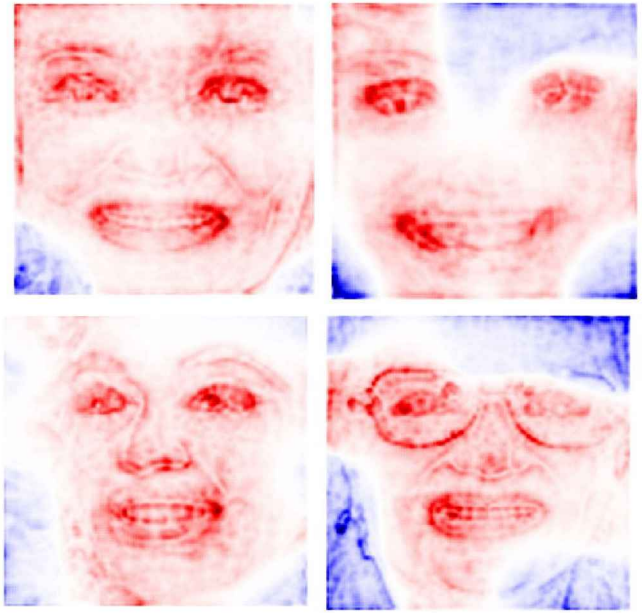
highest ratings

5.976190476 5.775 5.608695652 5.861111111



lowest ratings

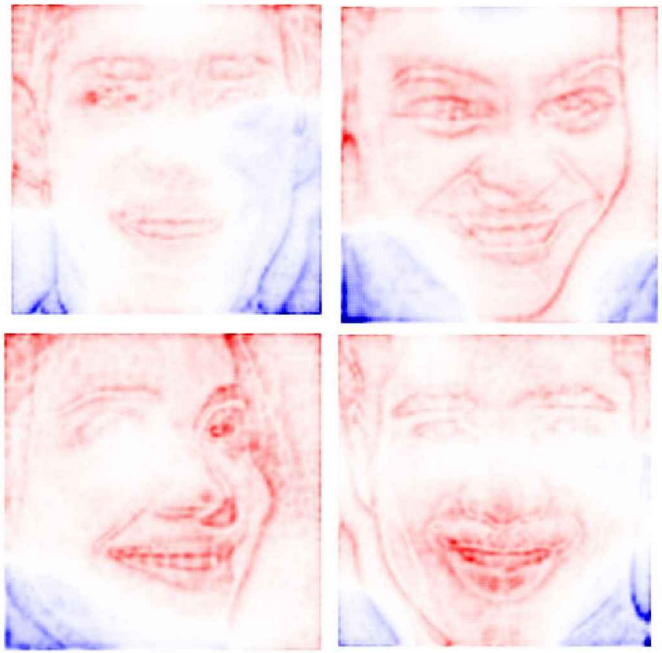
1.825 2.102564103 2.45945946 2.5



ASD Warm

highest ratings

5.675 5.648648649 6.093023256 5.720930233



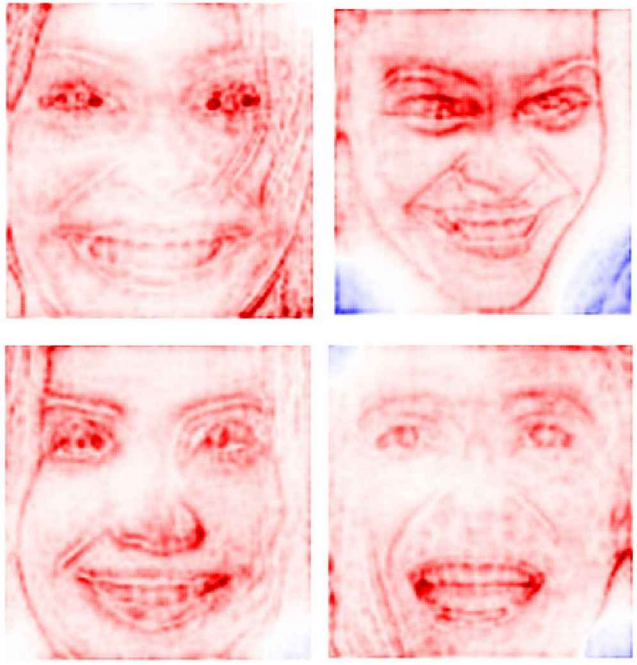
lowest ratings

2.047619048 2.282051282 2.564102564 2.666666667

NT Warm

highest ratings

6.027027027 5.956521739 5.947368421 5.880952381



lowest ratings

2.266666667 2.275 2.418604651 2.447368421

Next Step

- Use projection method to study how faces are clustered and separated in the t-SNE space
- Try other facial traits

Summary

- We have studied how do the people with ASD make facial trait judgements
- We find that mouth, eyes make the largest positive contributions to facial trait ratings