A Six-unit Network Is All You Need To Discover Happiness

Matthew N. Dailey, Garrison W. Cottrell
Gary's Unbelievable Research Unit (GURU)
Computer Science and Engineering Department
Institute for Neural Computation
UCSD

Ralph Adolphs
Department of Neurology
University of Iowa
Questions About Face Processing

- What kinds of *representations* are involved in face perception?
- How do specific *tasks* recruit those representations?
- Does categorization interact with perception? Conflicting data:
  - Sometimes facial expression perception looks continuous.
  - Sometimes facial expression data indicate *categorical perception*:
    - Sharp boundaries between categories.
    - Improved discrimination of pairs of stimuli near category boundaries.
- Our computational model of facial expression recognition begins to answer some of these questions.
Outline

An overview of our facial expression recognition system.

- Classification of prototypical facial expressions.
- Perception/classification of expression morph sequences.
- Conclusions: the model provides a simple explanation for a variety of human behavioral phenomena.
Facial Expression Database

- Ekman and Friesen quantified muscle movements (facial actions) involved in prototypical portrayals of happiness, sadness, fear, anger, surprise, and disgust.
  - Result: the pictures of facial affect (1976).
  - >70% agreement on emotional content by naive human subjects (average 92%).
- 110 images, 14 actors, 7 expressions.

"JJ" portraying 6 “basic” emotions and Neutral. JJ’s faces are the easiest for humans and our model to classify.
The Facial Expression Classifier

- Gabor Filtering
- PCA
- 1-layer NNet

Levels:
- Pixel (Retina) Level
- Perceptual (V1) Level
- Object (IT) Level
- Category Level

Emotions:
- Happy
- Sad
- Afraid
- Angry
- Surprised
- Disgusted
The Gabor Lattice Representation

- Basic feature: the 2-D Gabor wavelet filter (Daugman, 85):

- Combine two filters to get phase insensitivity, modeling complex cell responses in primary visual cortex.

- Subsample in a 29x36 grid:
Principal Components Analysis (PCA) for unsupervised dimensionality reduction:

- 50 element reduced representation suitable for supervised neural network training.
- Input/Output: 40,600-element Gabor Lattice
- (PCA can be learned by a Hebbian network)
- The resulting 50 inputs are fed to the category layer: a 6-unit softmax network (no hidden layer).
The Final Layer: Classification

- Trained based on putative emotions happy, sad, afraid, angry, surprised, or disgusted.
- The single layer, 6-unit softmax network is trained with the delta rule.
- Visualization of faces optimally activating each unit in one trained net:

(The inverse mapping from the 50-dimensional Gabor/PCA representation back to image space is crudely approximated here by linear regression.)

- Small number of parameters (no hidden layer) helps the network generalize well to unseen actors’ faces.
An overview of our facial expression recognition system.

Classification of prototypical facial expressions:
- Generalization performance is at the level of naïve human subjects.
- The human emotional *similarity structure* and the human *confusion patterns* are emergent properties of the model.

Perception/classification of expression morph sequences:

Conclusions: the model provides a simple explanation for a variety of human behavioral phenomena.
Prototype Generalization Results

<table>
<thead>
<tr>
<th>Expression</th>
<th>Network % Correct</th>
<th>Human Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happiness</td>
<td>100.0%</td>
<td>98.7%</td>
</tr>
<tr>
<td>Surprise</td>
<td>100.0%</td>
<td>92.4%</td>
</tr>
<tr>
<td>Disgust</td>
<td>100.0%</td>
<td>92.3%</td>
</tr>
<tr>
<td>Anger</td>
<td>89.1%</td>
<td>88.9%</td>
</tr>
<tr>
<td>Sadness</td>
<td>83.3%</td>
<td>89.2%</td>
</tr>
<tr>
<td>Fear</td>
<td>67.2%</td>
<td>87.7%</td>
</tr>
<tr>
<td>Average</td>
<td><strong>90.0%</strong></td>
<td><strong>91.6%</strong></td>
</tr>
</tbody>
</table>

- Average network/human performance is equivalent \((t=0.587, p=0.279)\).
- Kendall’s tau (rank order correlation): \(0.667, p=0.0441\).
- The correlation is an emergent property of the model.
- This implies that humans have difficulty classifying fear because the expression is *intrinsically difficult*. 
Comparing Similarity Structures

- Multidimensional scaling (MDS) helps visualize similarity ratings. The technique makes facial expression space look continuous.

Distances are derived from correlation of human 6-way forced choice responses.

Configuration: H-S-F-M-A-D
Network Similarity Structure

- MDS configuration derived from training set similarity at the network’s output layer has the same order around the configuration:

  Same arrangement of emotions (H-S-F-M-A-D) about the origin.
Correlation of Net/Human Errors

- MDS informally indicates that the network induces some of the similarity structure present in the data.
- Correlation of network and human responses can quantify this more formally.
- Networks and humans have a 6x6 confusion matrix for the stimulus set. The off-diagonal terms are the errors.
- Correlation of off-diagonal terms: training set: $r = 0.4818 \ (p = 0.007)$; test set: $r = 0.6614 \ (p<0.001)$.
- This correlation is another *emergent property* of the model: it is never told which expressions it should confuse.
Outline

- An overview of our facial expression recognition system.
- Classification of prototypical expressions:

  Perception/classification of expression morph sequences:
  - *Decision* based on network output layer accounts for human 6-way forced choice identification data.
  - *Output layer uncertainty* explains human response times.
  - *Input (Object) layer similarity* accounts for human discrimination data.
  - Model detects secondary expression in a morph as well as humans do.

- Conclusions: the model provides a simple explanation for a variety of human behavioral phenomena.
Morph Transition Perception

- Morphs help psychologists study categorization behavior in humans.
- Example: JJ fear--sadness morph:

We use data from Young et al.’s (1997) “megamix” study for comparison. They presented images from morphs between all 6 emotions (15 sequences) to subjects in random order, with a 6-way forced choice task.
Six-way Forced Choice

- Human data: sharp transitions, few unrelated intrusions.

- Only 6 of 15 possible transitions are shown here, but the rest are similar (only 2 intrusions across all 15 transitions). This is one of the criteria for categorical perception.

- We model the forced-choice button push as a probabilistic coin-toss according to the probabilities at the network output.
Modeling the Six-way Button Push

- Correlation of $r=0.9416$ over all 15 transitions,
- Even though the model is never trained on JJ or morph images!
Human RT patterns are inconsistent with predictions of “perfect” CP.

Model nevertheless accounts for data: $r=0.6771$ ($p<0.001$).
A necessary condition for “categorical perception” is better discrimination of two stimuli a fixed distance apart near category boundaries.

Indeed, Young et al. found that subjects were significantly better at discrimination near boundaries than near the prototypes.
Young et al. found that subjects were significantly better at discrimination near boundaries than near the prototypes.
Modeling Discrimination

- Is the improved discrimination near the boundaries due to the influence of category boundaries?
- Discrimination is naturally modeled as the flip side of similarity:
  - We model discrimination as $1-r$ (correlation) between pairs.
- Prediction of CP: best fit should occur at category level of the model.
Surprise! The model fits the data best at a precategorical level, the “object layer,” NOT at the category level!

- Facial expression CP may be a result of recruiting a tuned representation.
Detecting 2\textsuperscript{nd} Expression in a Morph

- When subjects are asked to rate 1\textsuperscript{st}, 2\textsuperscript{nd}, 3\textsuperscript{rd} most apparent expressions, subjects reliably detect the mixed-in expression at the 30\% morph level.

- These data seem more consistent with \textit{continuous} perception.

- But they arise naturally in the network when its 1\textsuperscript{st}, 2\textsuperscript{nd}, and 3\textsuperscript{rd}-largest responses are scored the same way.
Discussion

- Some data on facial expression perception seems continuous:
  - MDS facial expression circumplex.
  - Categorization RT cost when moving away from prototypical faces.
  - Ability to detect the 2nd expression mixed in to a morph image.

- Other data show CP:
  - Sharp category boundaries.
  - Better discrimination near category boundaries.

- The model naturally accounts for all these results.
  - Some expressions are inherently more difficult than others.
  - Some expressions are inherently more confusable than others.
  - Any good classifier will give you sharp category boundaries.
  - Better discrimination near the boundaries is pre-categorical with our representation.
Conclusions

- Our model simultaneously fits data supporting both categorical and continuous theories of expression perception.
- We believe the fits are due to:
  - The way expression categories slice up perceptual face space,
  - The inherent similarity of facial expressions in that space,
  - And tuning the representation to facial identity/expression via PCA.
- The model shows that discrete categories and innate CP are unnecessary to explain the data.
- We have demonstrated in detail how experiments can tap either a continuous perceptual representation or a discrete decision process.
- We believe these results will generalize to other domains where CP appears.