Efficient inverse graphics in biological face processing

by Yildirim, Freiwald & Tenenbaum

Presented by Madis Vasser & Mansur Alizada

Why face processing?

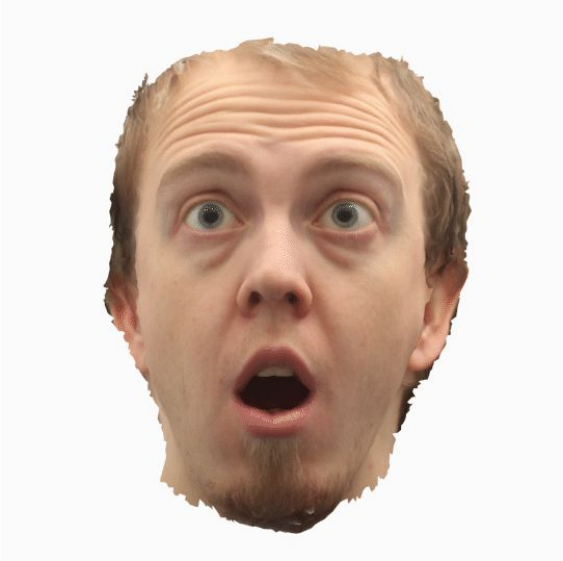
* brain imaging
* single-cell recordings
* psychophysics
* visual illusions

Abstract

Vision is not just **seeing** some objects,
but also **inferring** their sensory causes.

Inverse generative models so far have been
slow and biologically unclear.

Proposed model **explain** human and non-human
data and illusions **better than state-of-the-art**
computer vision models.



Efficient **I**nverse **G**raphics model

Main parts:

1) a probabilistic generative model for image synthesis

2) an inverse recognition function based on a DCNN

Efficient Inverse Graphics model

Exploits:

1) The observable raw image is conditionally independent of the 2.5D face components, given the ideal face image

2) The 2.5D components are conditionally independent of person identity, given the 3D scene parameters.

Efficient
Inverse Graphics
model

Novel approaches:

1) Trained to produce inputs to graphics engine, rather than class predictions

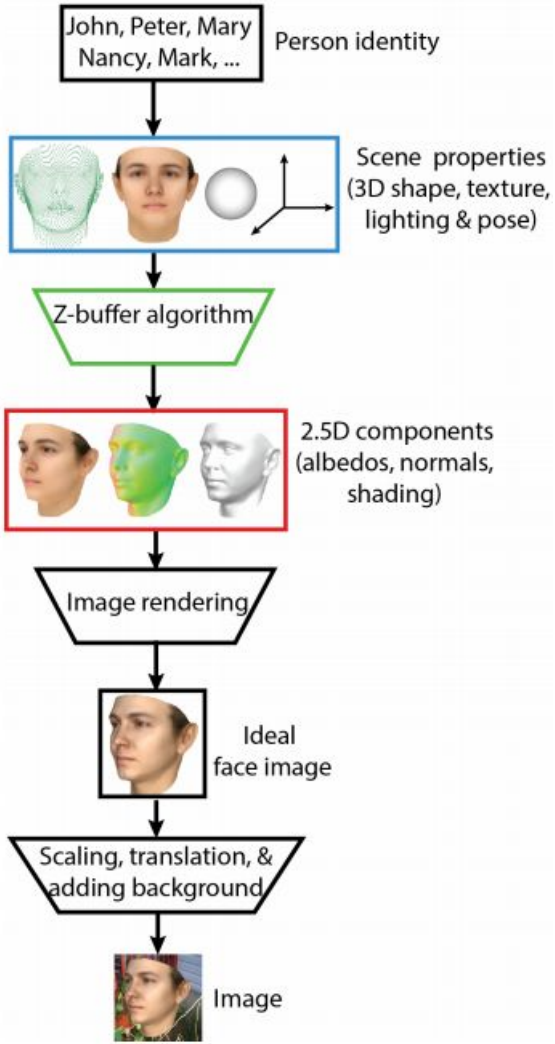
2) Trained in a completely self-supervised way, rather than by externally supervised labeled images.

Efficient Inverse Graphics **model**

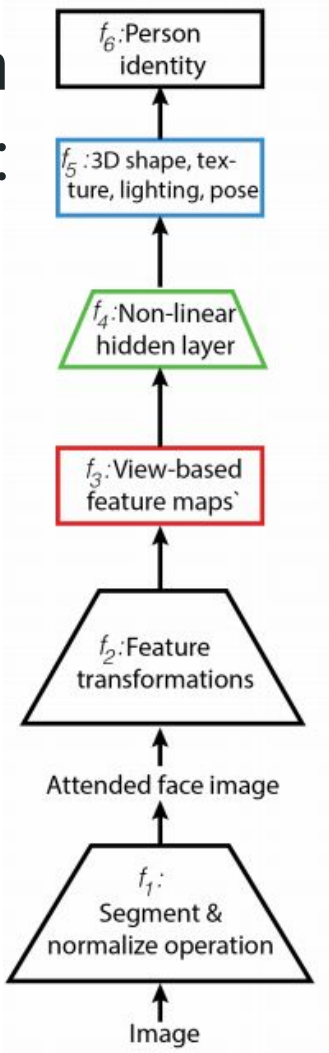
The recognition pipeline:

Stage 1:
normalize the input
Stage 2:
estimate scene properties
Stage 3:
estimate most likely face ID

Generative model:



Recognition model:



Stage 1

Segment and normalize the input image to compute the attended face image, i.e., the most probable value for the ideal image I* given the observed image O,
By maximizing Pr(I|O) using a DCNN module trained for three-dimensional face segmentation with background clutter

Stage 2

Estimate intrinsic and extrinsic scene properties {S*, T*, L*, P*} maximizing Pr(S, T, L, P|I*) from the attended face image. Four convolutional layers ending in a fifth, top convolutional feature space, followed by two fully connected layers. The second fully connected layer is trained to predict scene properties. Training images are generated by “dreaming” images from the generative model.

Stage 3

Estimate the most likely face identity label F* given the scene properties, maximizing Pr(F|S*, T*, L*, P*). Comprises a single new FCL for person identity classification, also trained on another self-generated set of simulated faces, but starting from the prior over individuals P(F), which can be specific to a particular set of faces encountered in an individual experiment

Weakness of recognition model

1. Segmentation step f1 fails

2. Model’s reconstruction accuracy may degrade

EIG vs Macaque brain (task)

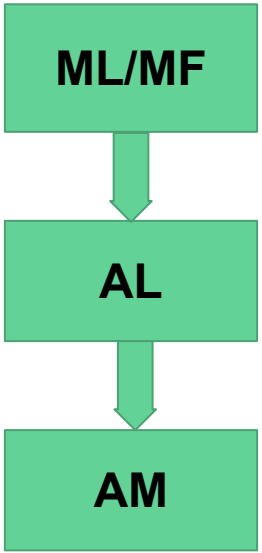
Three level hierarchy

Middle Lateral

Middle Fundus

Anterior Lateral

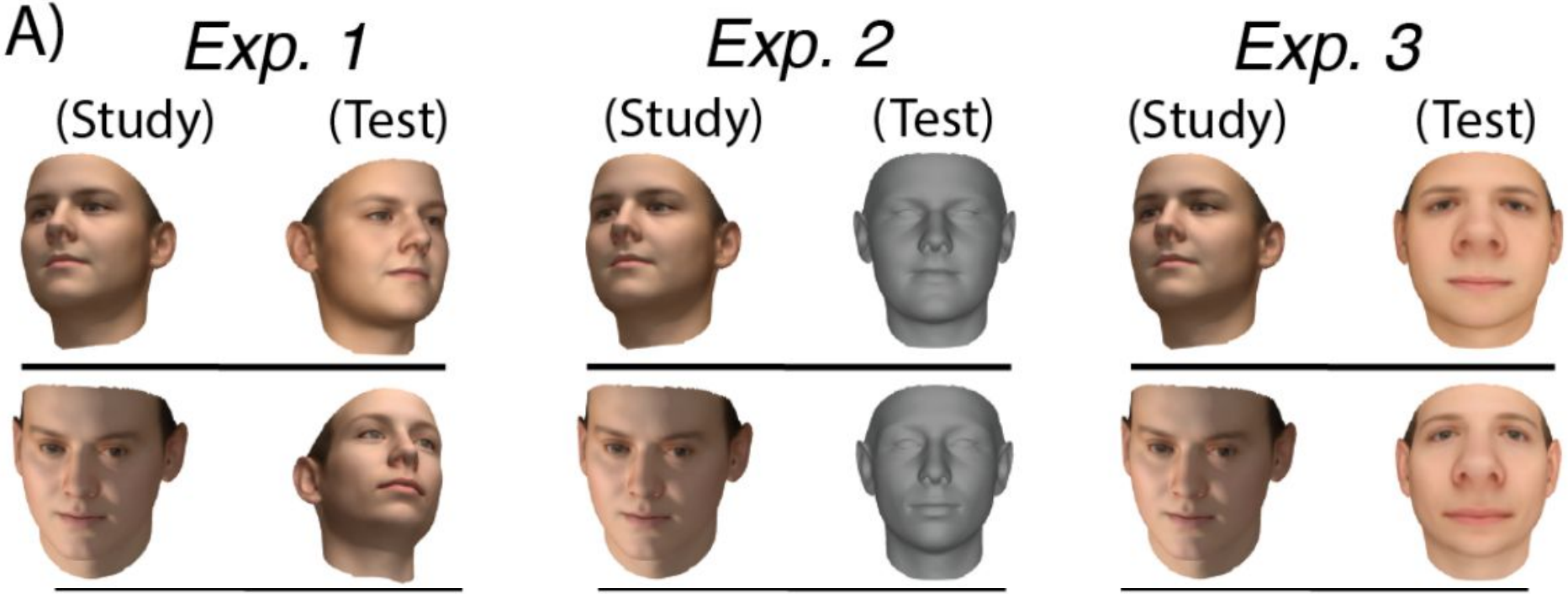
Anterior Medial



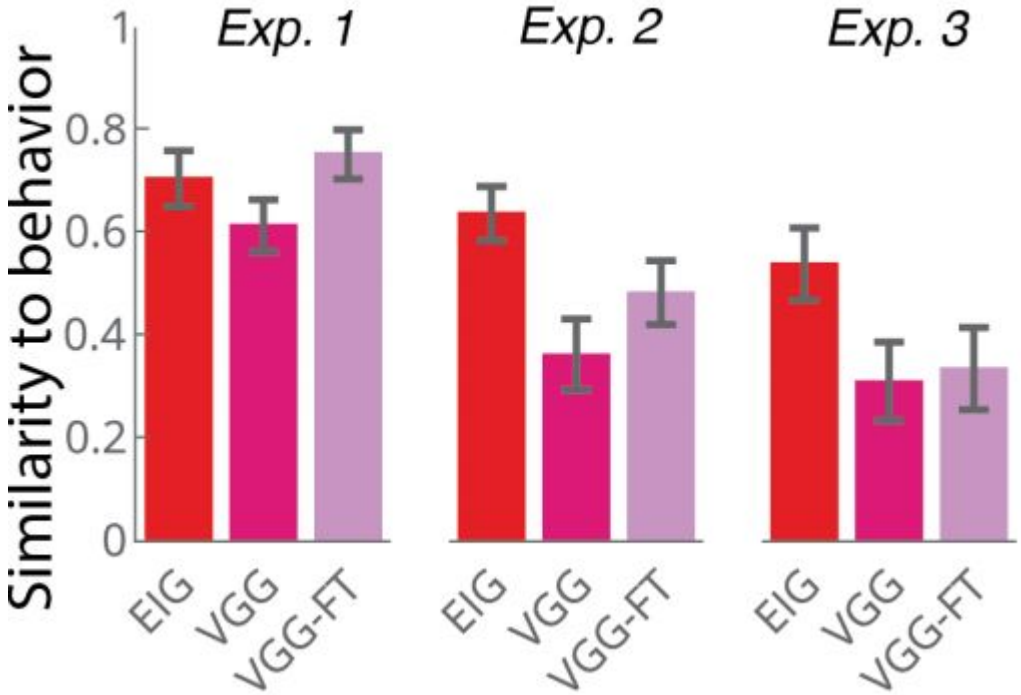
EIG vs Macaque brain (results)

- EIG faithfully reproduced all patterns in the neural data
- EIG model closely tracked the functional compartmentalization of the brain
- EIG was better than alternatives (EIG- or VGG)
- Suggests that the face processing network begins with face segmentation and culminates in targets that encode 3D scene properties rather than features optimized for identity discrimination.

EIG vs Human brain (task)



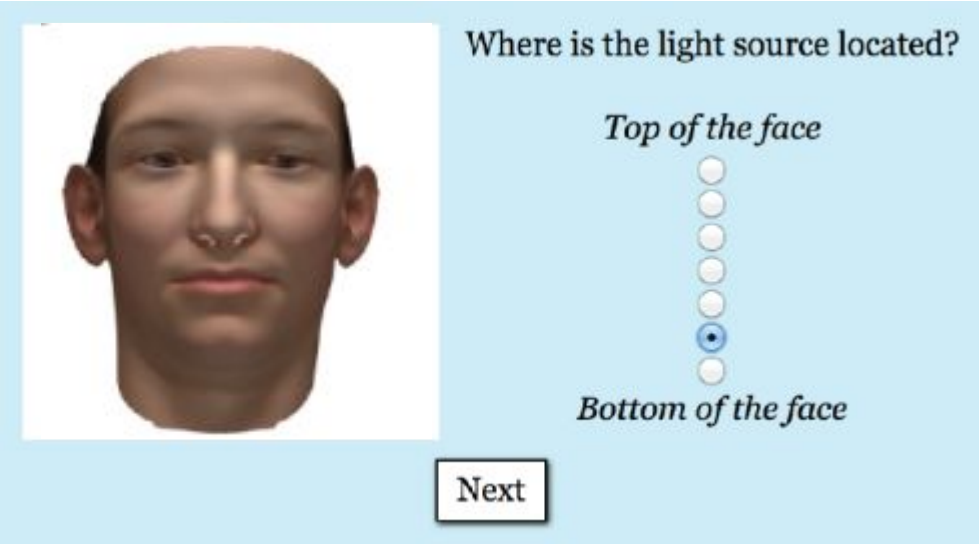
EIG vs Human brain (results)



- EIG predicted human error patterns, with r values 0.70 / 0.64 / 0.54
- EIG exhibited better generalization than alternative models when test images were distorted

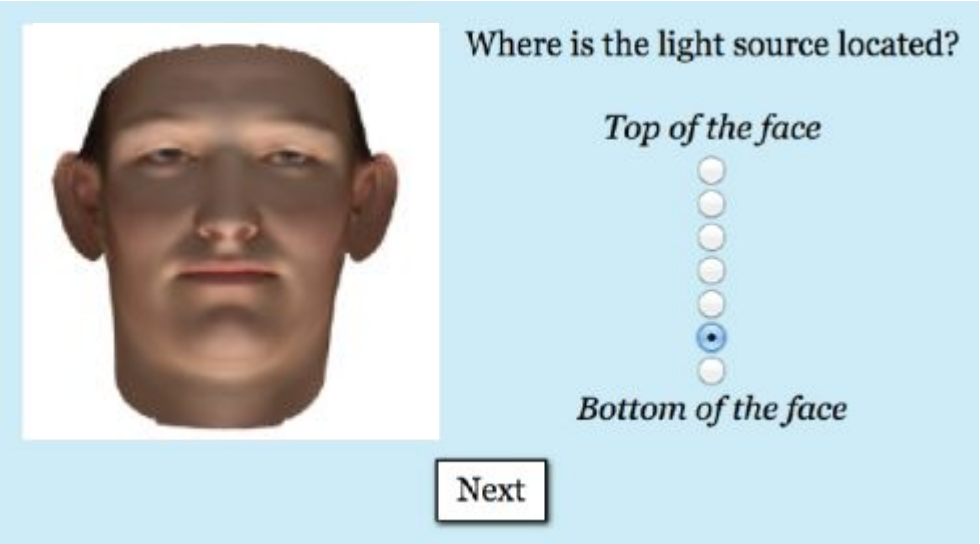
EIG vs Human brain (results)

- EIG is also fooled by the hollow face illusion!



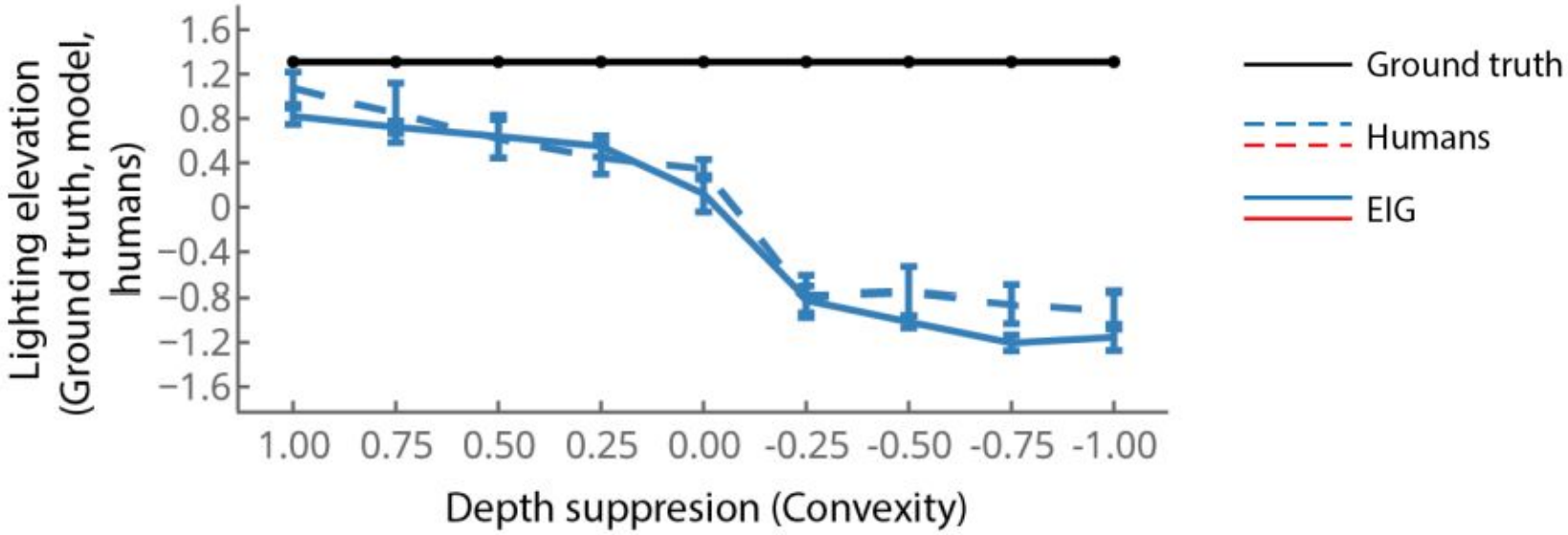
EIG vs Human brain (results)

- EIG is also fooled by the hollow face illusion!



EIG vs Human brain (results)

- EIG is also fooled by the hollow face illusion!



EIG drawbacks and possibilities

- Feedback and other non-hierarchical connectivity is absent in the model
- “Semi-interpretable”
- Could parse multiple or occluded objects in theory
- Flexible cross-modal transfer, e.g. things seen can be recognized by touch