Visual Object Recognition
Computational Models and Neurophysiological Mechanisms

Neuro 130/230. Harvard College/GSAS 78454

Which one of these images is fake?

Visual Object Recognition
Computational Models and Neurophysiological Mechanisms
Neurobiology 130/230. Harvard College/GSAS 78454

Note: no class on 09/04/2023 (Labor Day)

Class 1 [09/11/2023]. Introduction to Vision

Class 2 [09/18/2023]. The Phenomenology of Vision

Class 3 [09/25/2023]. Natural image statistics and the retina

Class 4 [10/02/2023]. Learning from Lesions

Note: no class on 10/09/2023 (Indigenous Day)

Class 5 [10/16/2023]. Primary Visual Cortex

Class 6 [10/23/2023]. Adventures into *terra incognita*

Class 7 [10/30/2023]. From the Highest Echelons of Visual Processing to Cognition

Class 8 [11/06/2023]. First Steps into in silico vision

Class 9 [11/13/2023]. Teaching Computers how to see

Class 10 [11/20/2023]. Computer Vision

Class 11 [11/27/2023]. Connecting Vision to the rest of Cognition [Dr. Will Xiao]

Class 12 [12/04/2023]. Visual Consciousness

FINAL EXAM, PAPER DUE 12/14/2023. No extensions.

Some computer vision problems

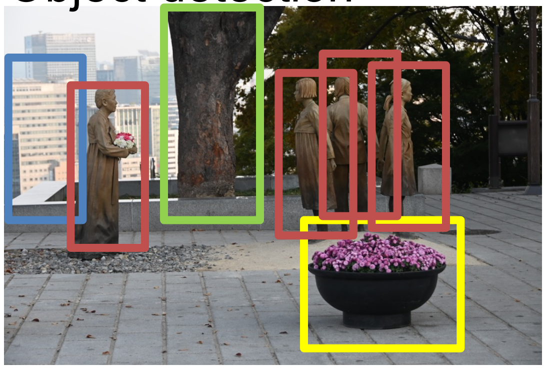
Classification

Flowers

Classification+Localization

Flowers

Object detection



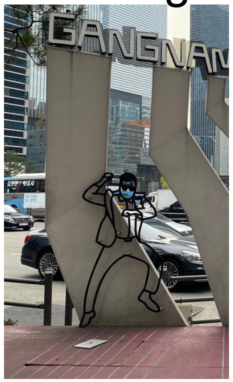
Building People Tree Flowers

Instance segmentation

Face identification

J. Aniston
R. Whiterspoon

Action recognition

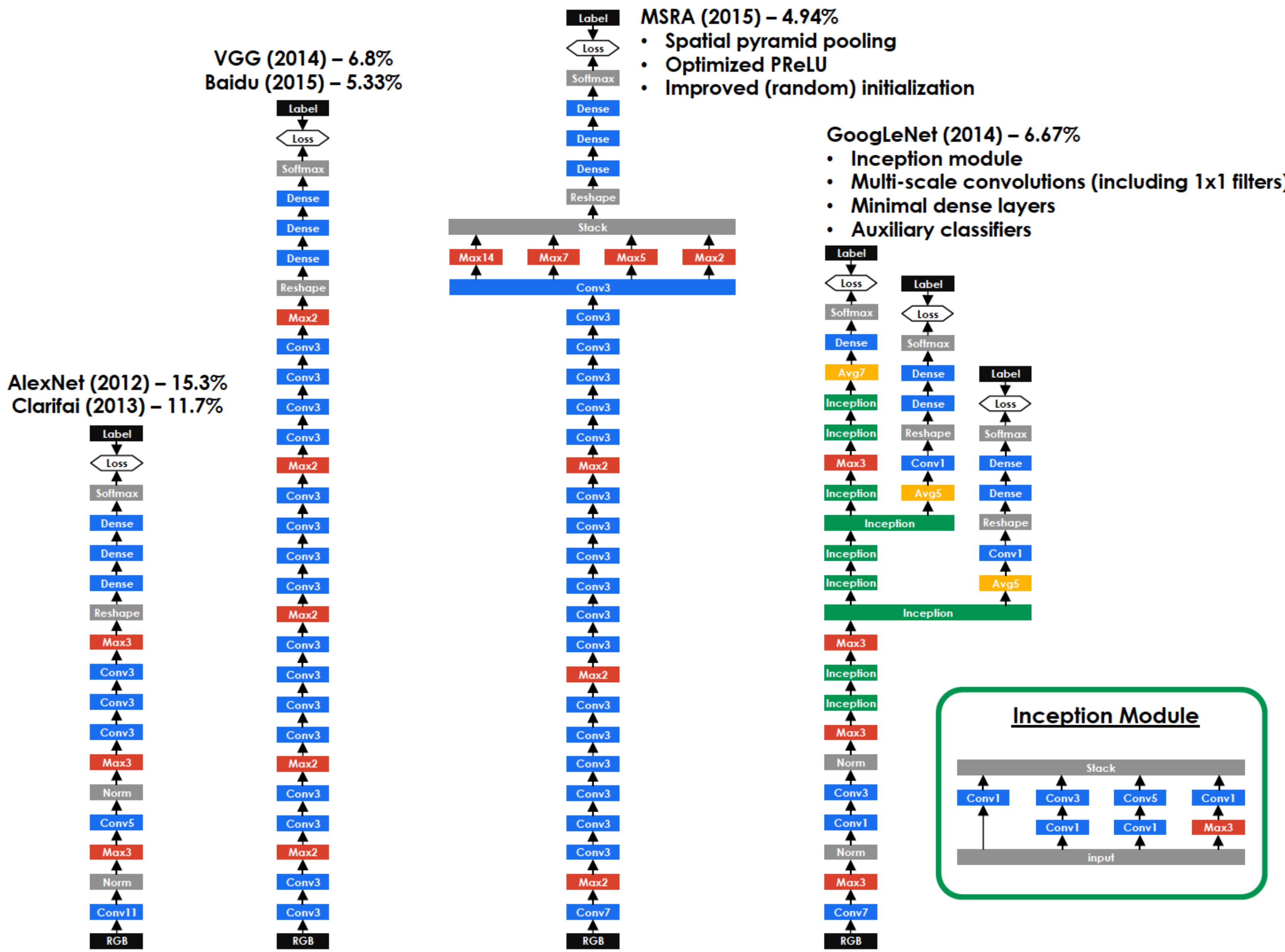


Dancing

Many more…

- Face detection
- Distance estimation
- Video prediction
- Image captioning

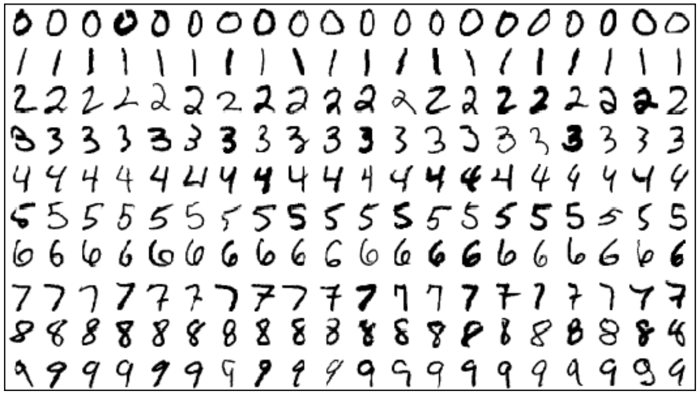
Many more architectures



Note: lots of parameters!!!

- Image of 256 x 256 x 3 pixels = 196,608 inputs
- 1000 output categories (imagenet)
- Simplest scenario: go from pixels to outputs
- ~ 200 x 10^6 parameters
- ~ 10^6 training images in ImageNet

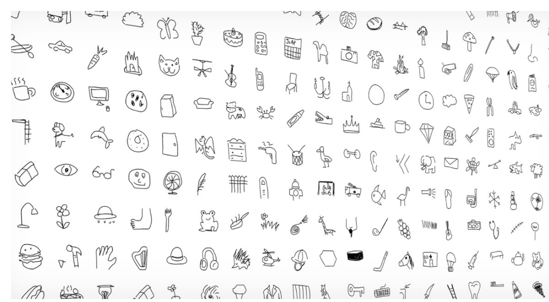
Data, data, data



MNIST

IMAGENET

MSCOCO



QUICKDRAW

Many more …

- Galaxies
- Plants
- Clinical images
- Cell types

English test

1. Briard
2. Cuirass
3. Consomme
4. Shoji
5. Busby
6. Weevil

English test

Briard

Cuirass

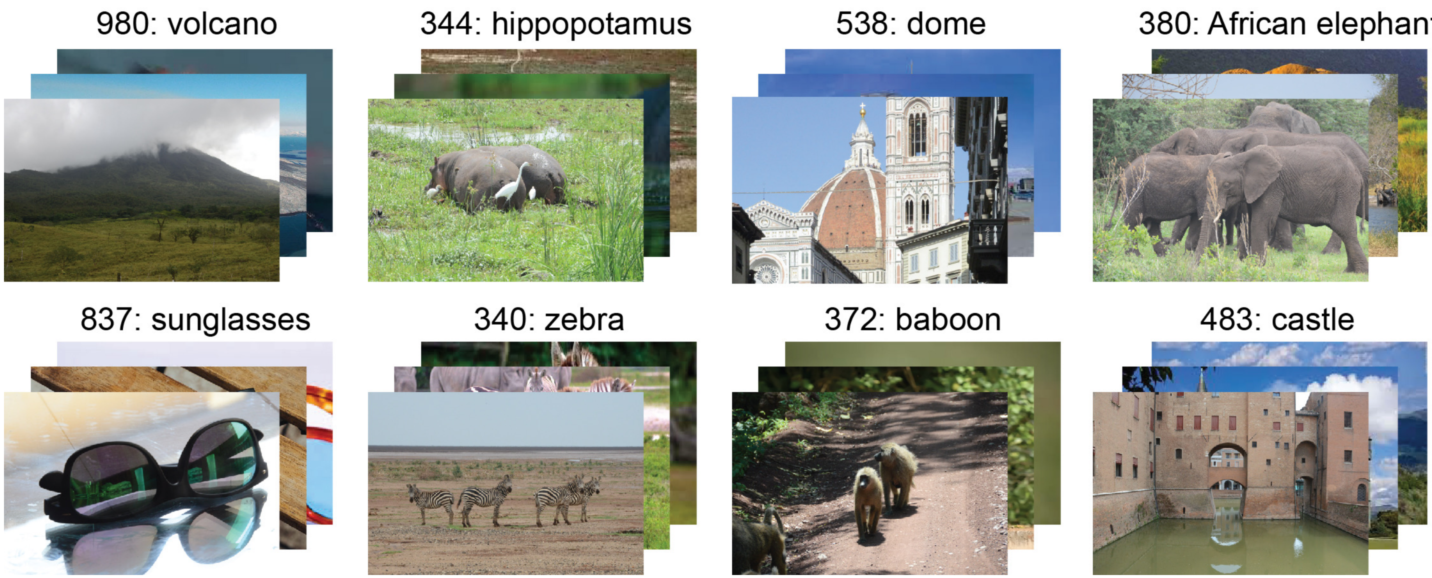
Consomme

Shoji

Busby

Weevil

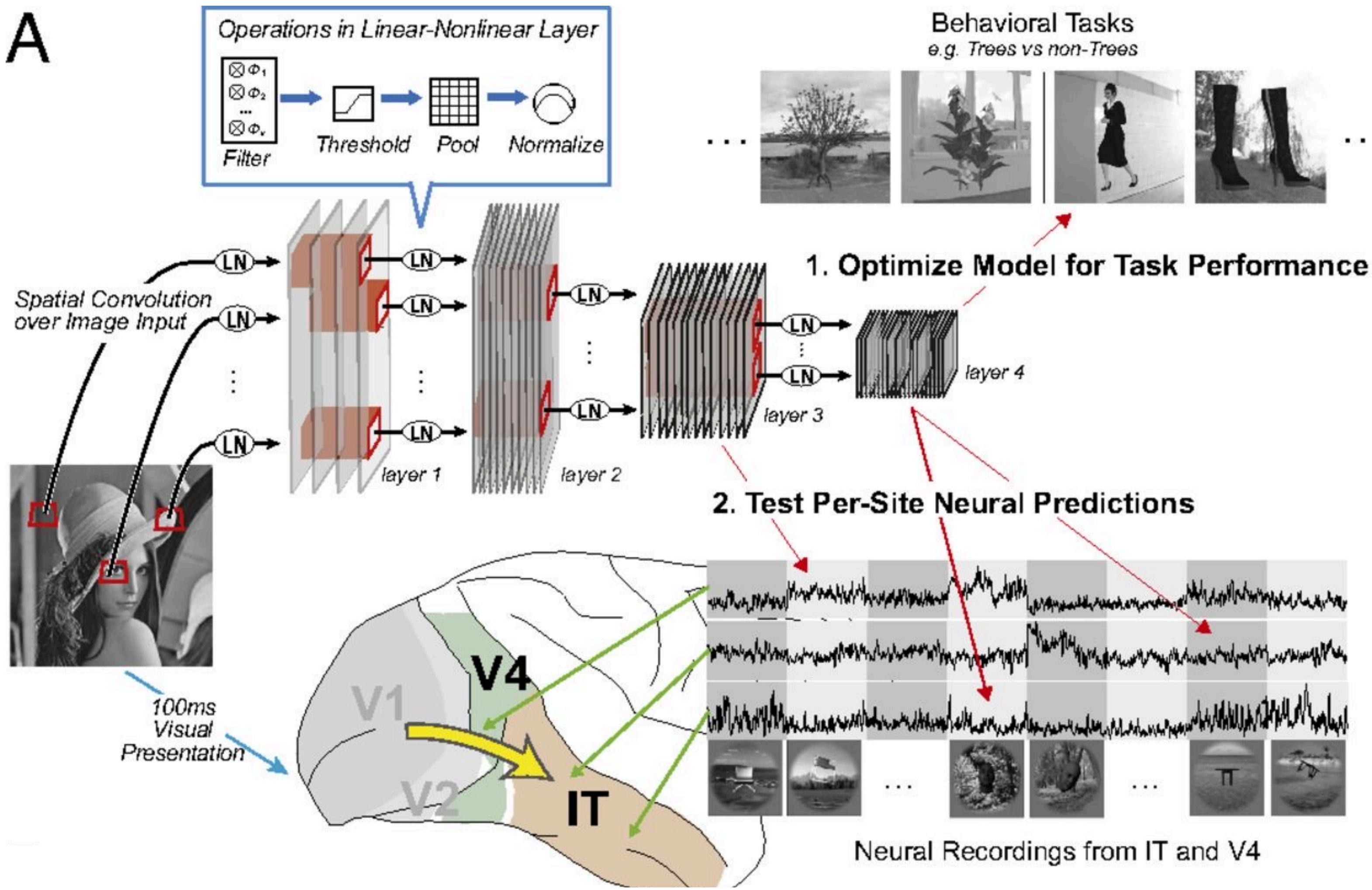
ImageNet



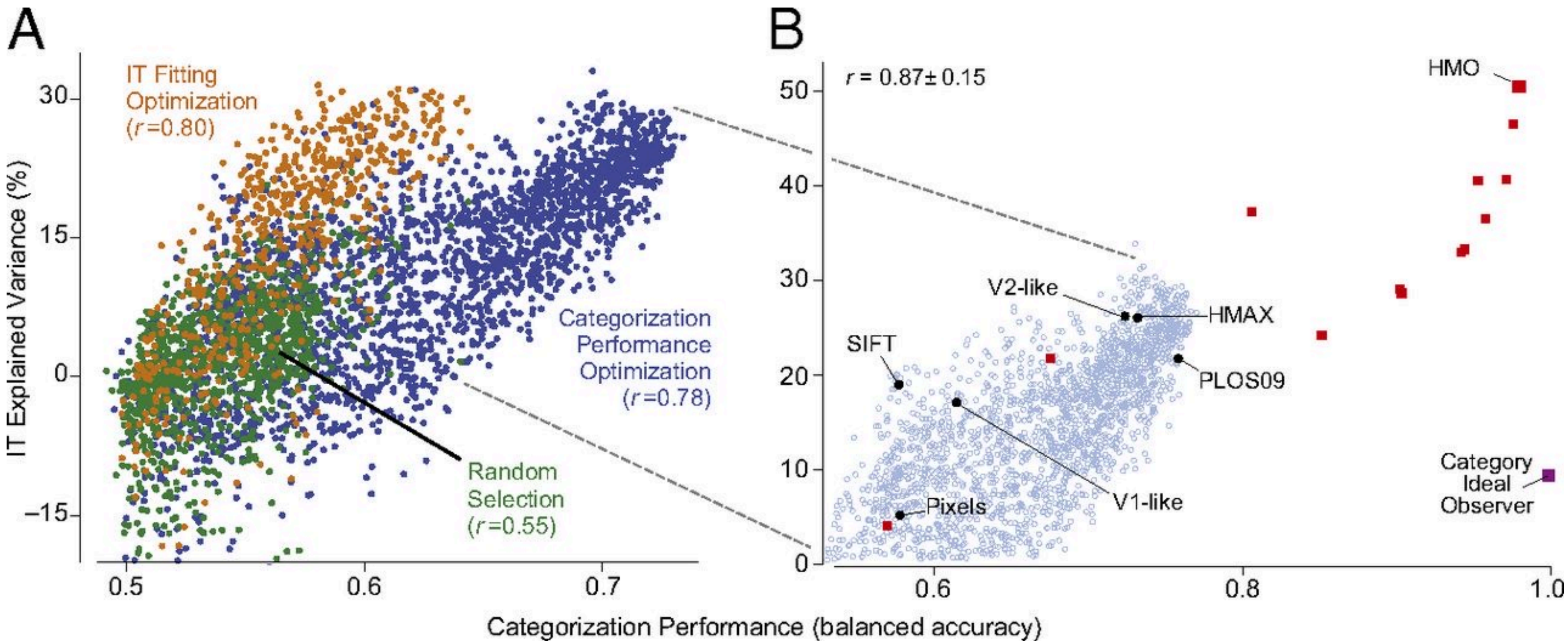
~1,000 categories

~1,000 images/category

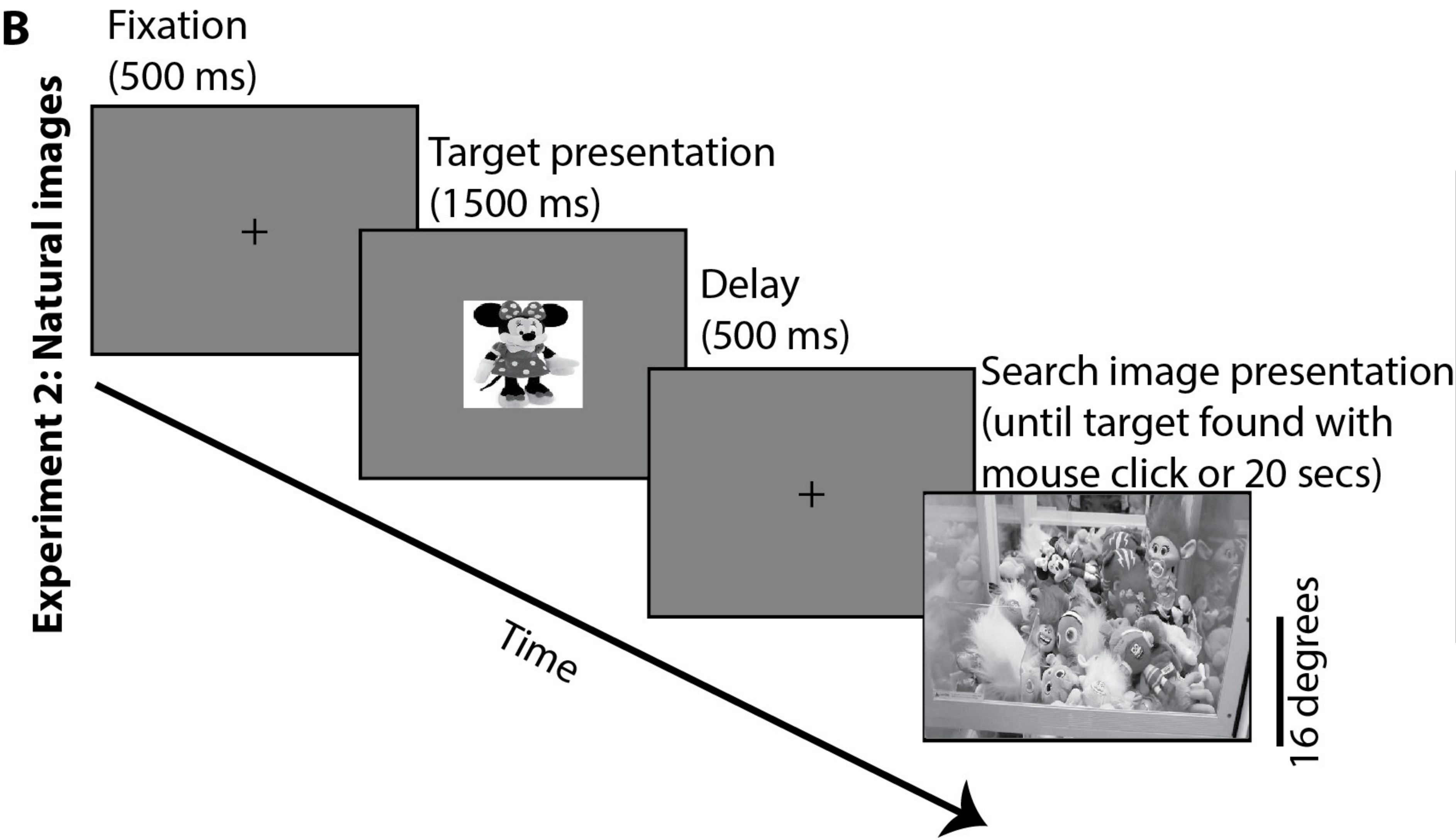
Computational models can approximate neuronal responses along the ventral visual cortex



The better the biological approximation the better performance in computer vision tasks

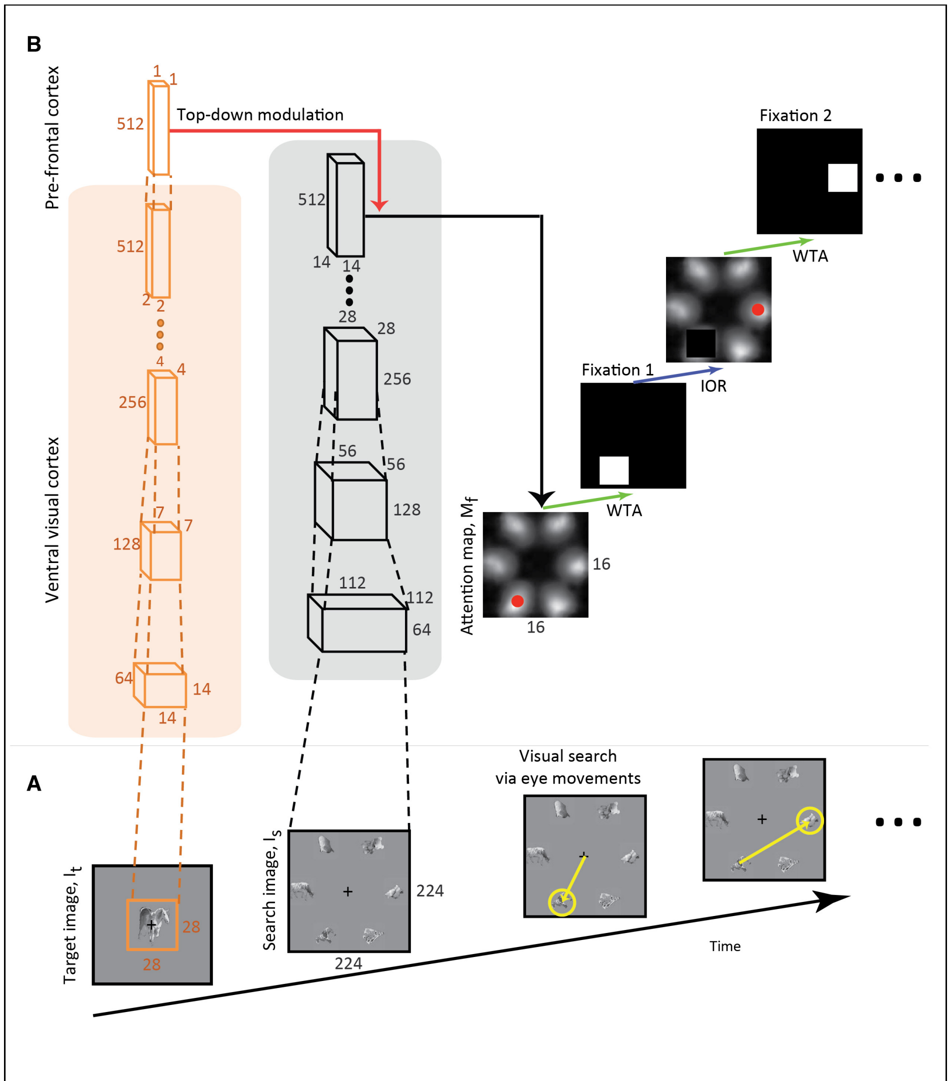


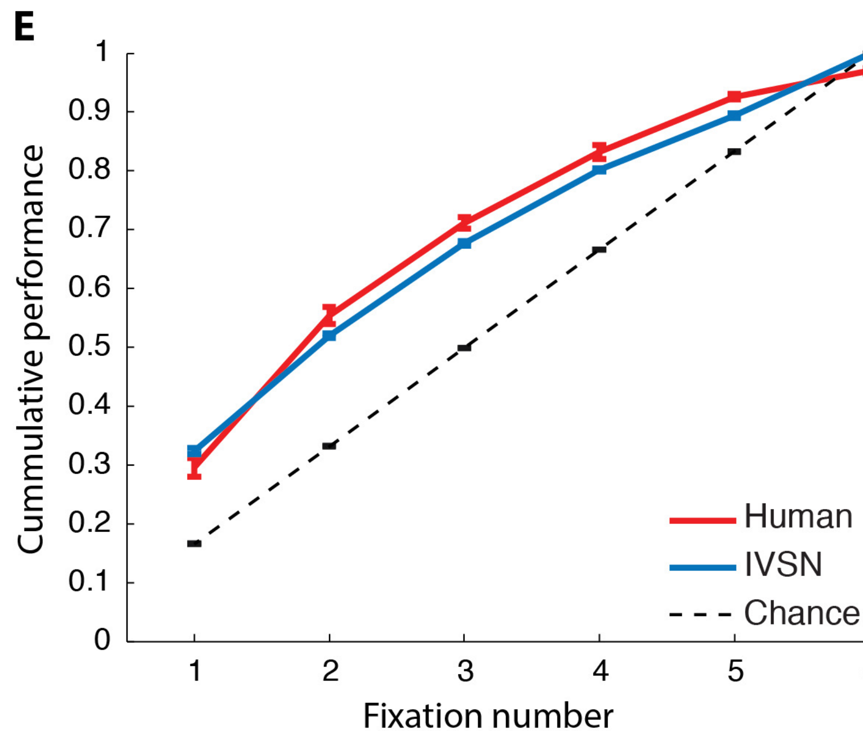
Predicting eye movements during visual search



Zhang et al, 2018

Predicting eye movements during visual search

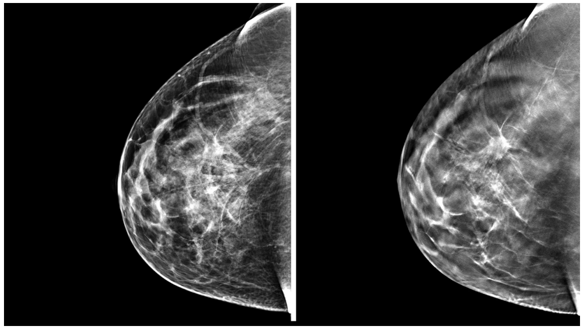




Zhang et al 2018

Machines surpass humans in pattern recognition tasks

(01)09312345678907

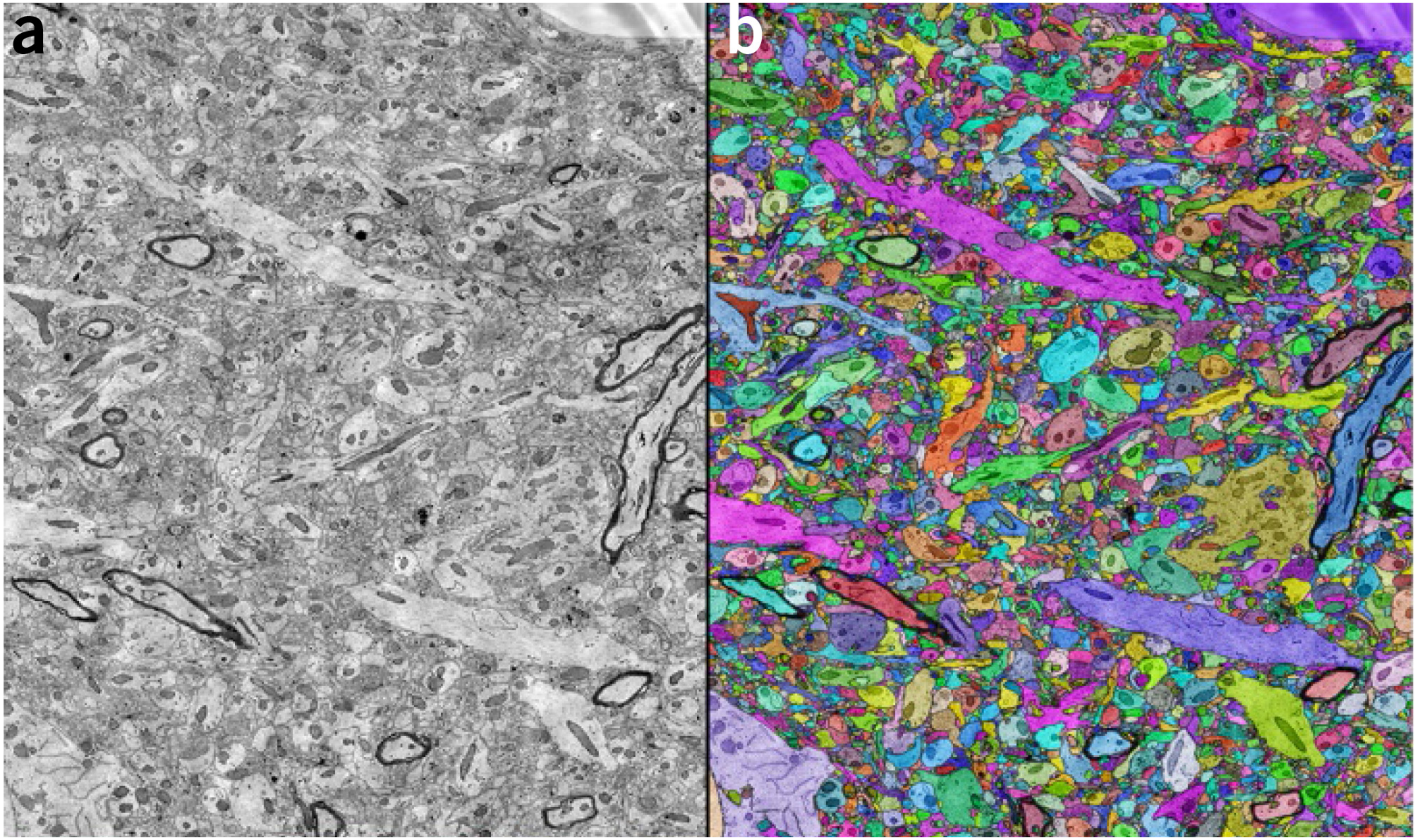


Face recognition better than forensic experts and human "superrecognizers" (Phillips et al 2018)

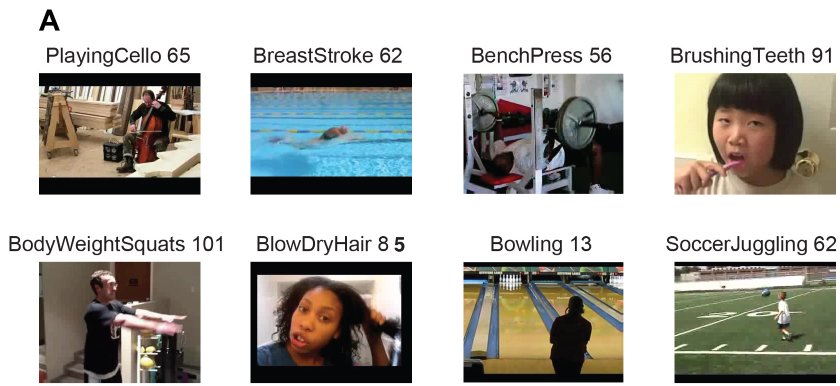
Plant and animal classification (iNaturalist, Van Horn et al. 2018) ~ 1M photos from 5,089 taxa and 13 "super-classes": expert human levels and better than naïve observers

Pose tracking in animal biomedical research (Matthis 2018)

Computer vision can help segment biological images

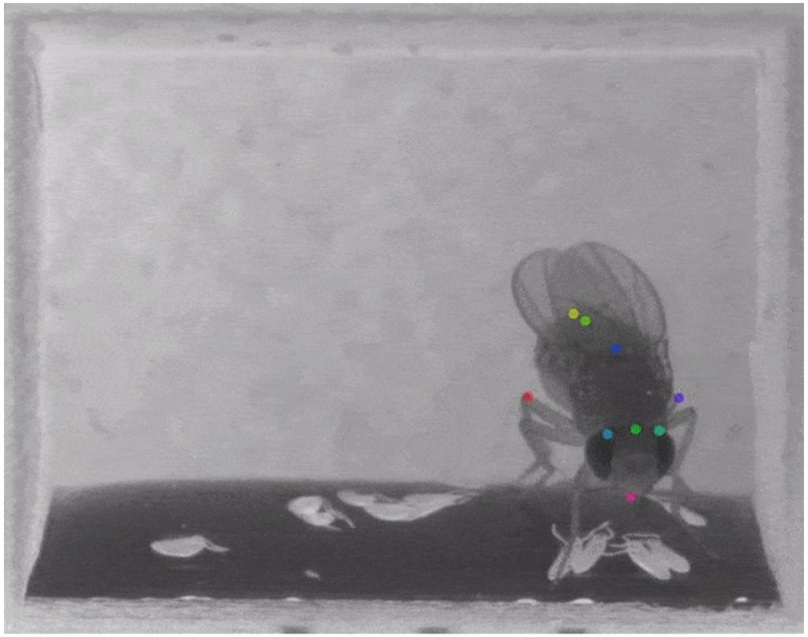


Computer vision for action recognition



Jacquot et al CVPR 2020

Automatic pose estimation for ethology research



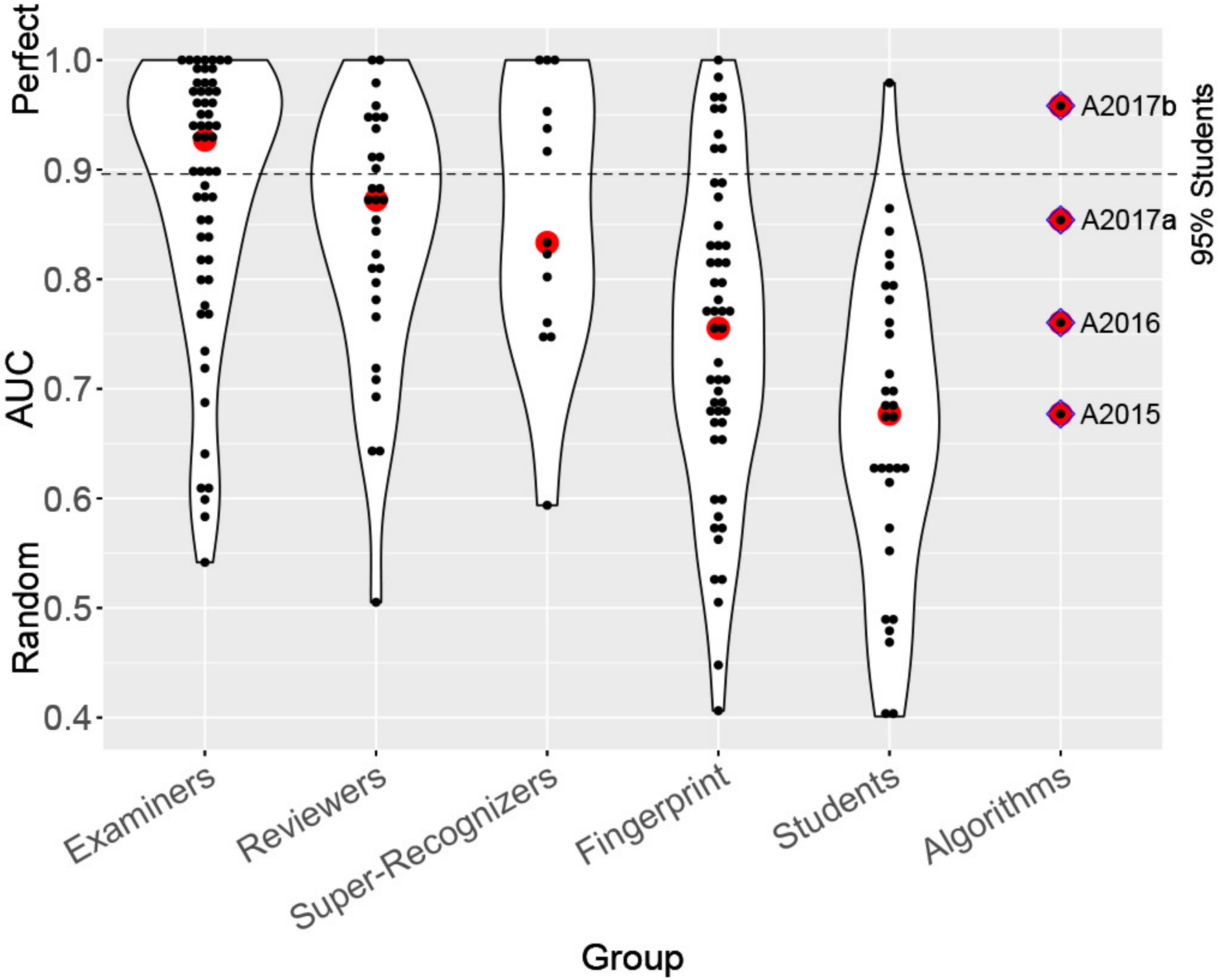
Mathis et al 2018

Face recognition by computer vision

Same or different?

Phillips et al 2018

Face recognition by computer vision



Phillips et al 2018

Species classification and detection

Van Horn et al 2018

Species classification and detection

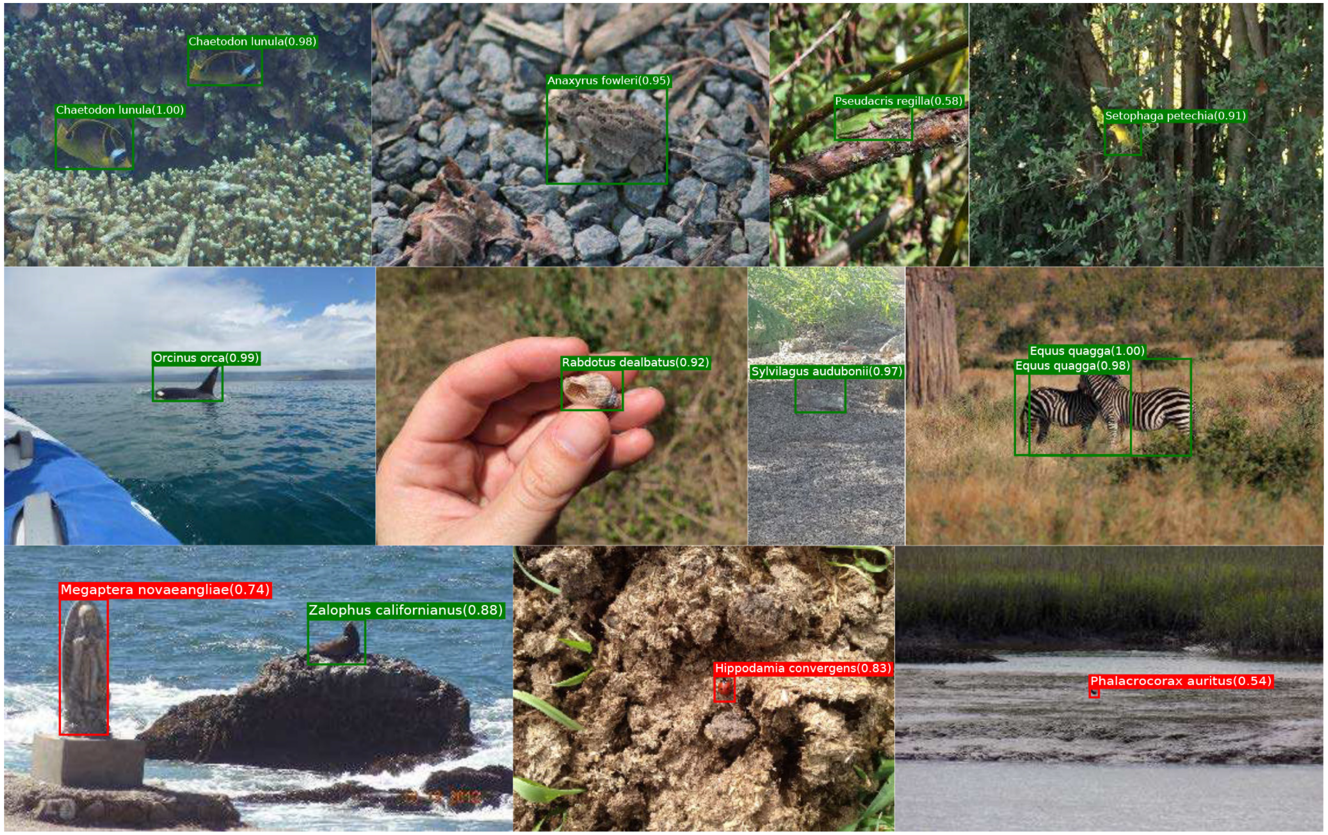
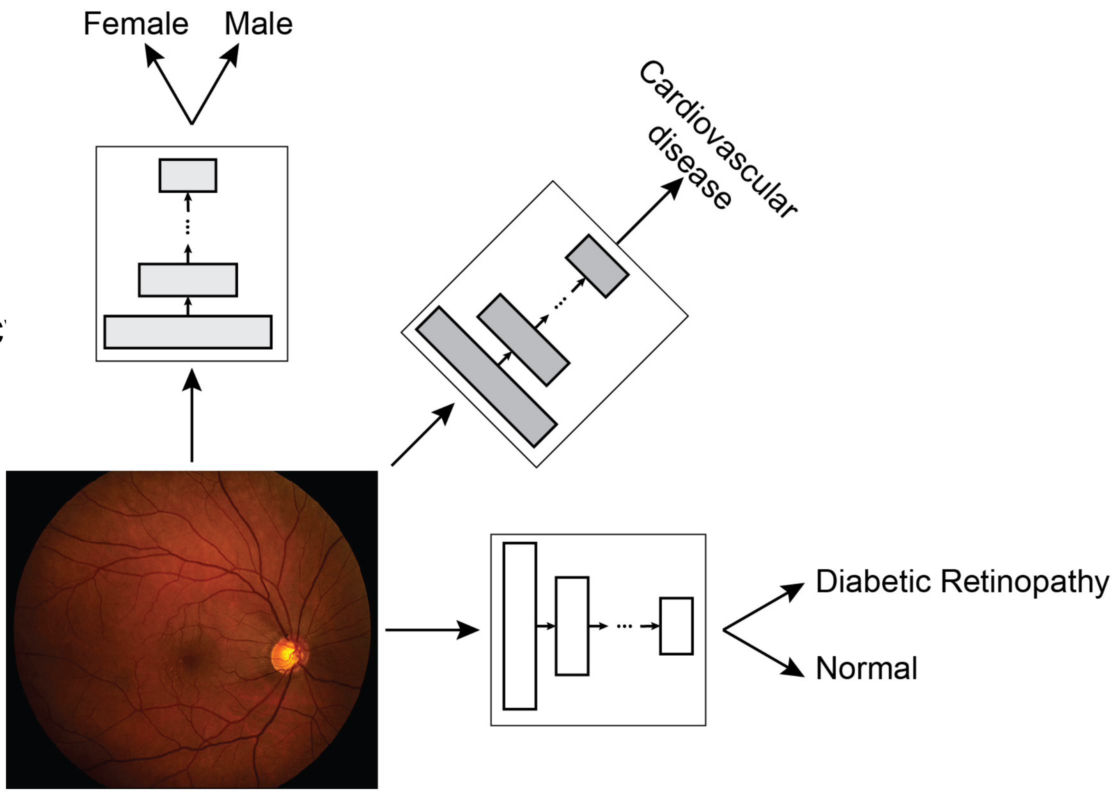


Figure 7. Sample detection results for the 2,854-class model that was evaluated across all validation images. Green boxes represent correct species level detections, while reds are mistakes. The bottom row depicts some failure cases. We see that small objects pose a challenge for classification, even when localized well.

Van Horn et al 2018

Applications of computer vision to clinical diagnosis

- Excellent performance in many clinical diagnosis tasks
  E.g. breast tumor detection
  E.g. diabetic retinopathy
- Reliability, consistency, accurac
- Machines can discover properties in the data that humans never even thought of before
  E.g. cardiovascular disease risk from fundus photographs
- Beware of incidental findings
- Beware of biases in training data

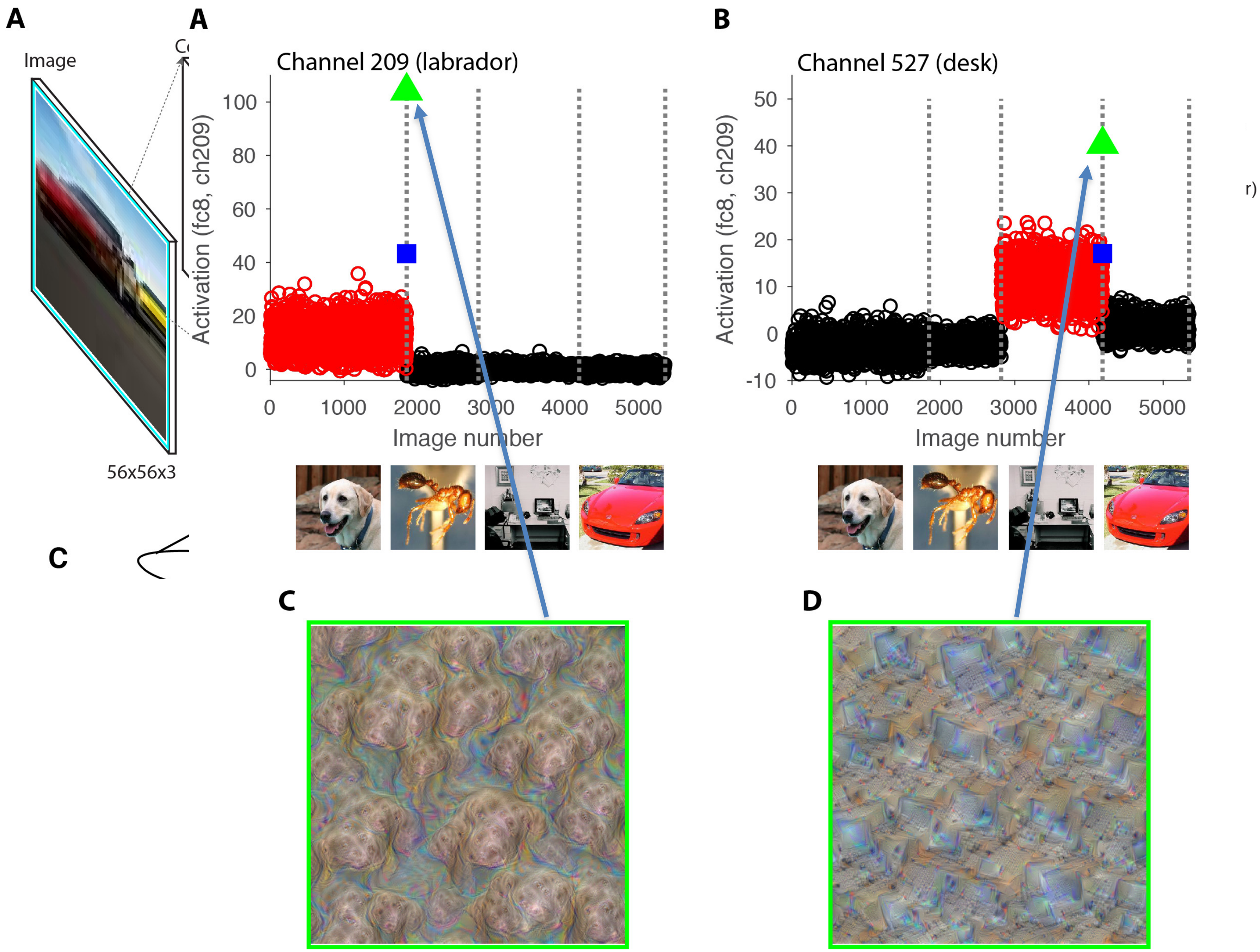


What is common to all these faces?

Generative adversarial networks (GANs)

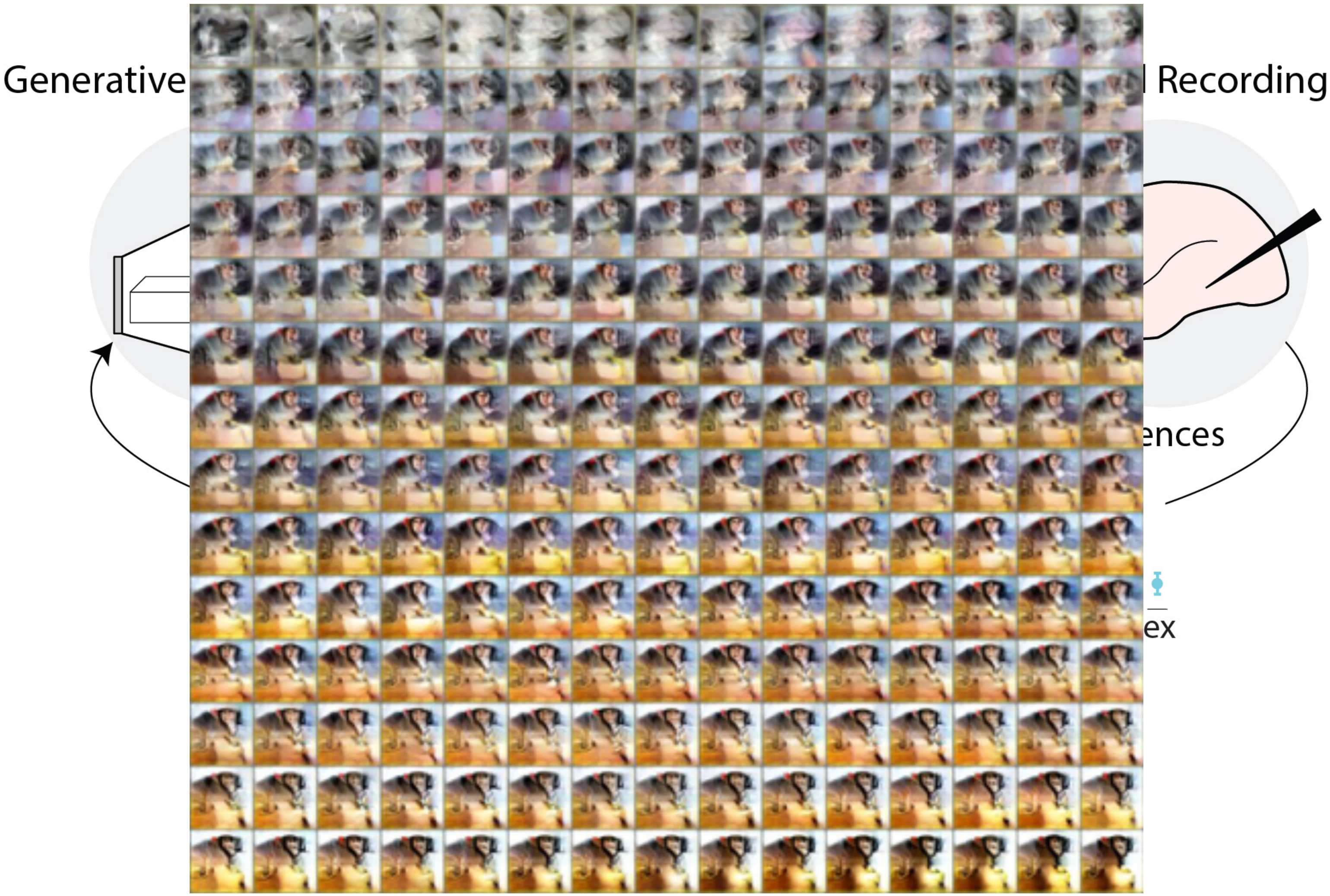
Goodfellow 2014

Deep Dreaming



nonyan et al 2014
Kreiman 2019

Xdream: Discovering neuronal tuning preferences



Ponce, Xiao, et al 2019

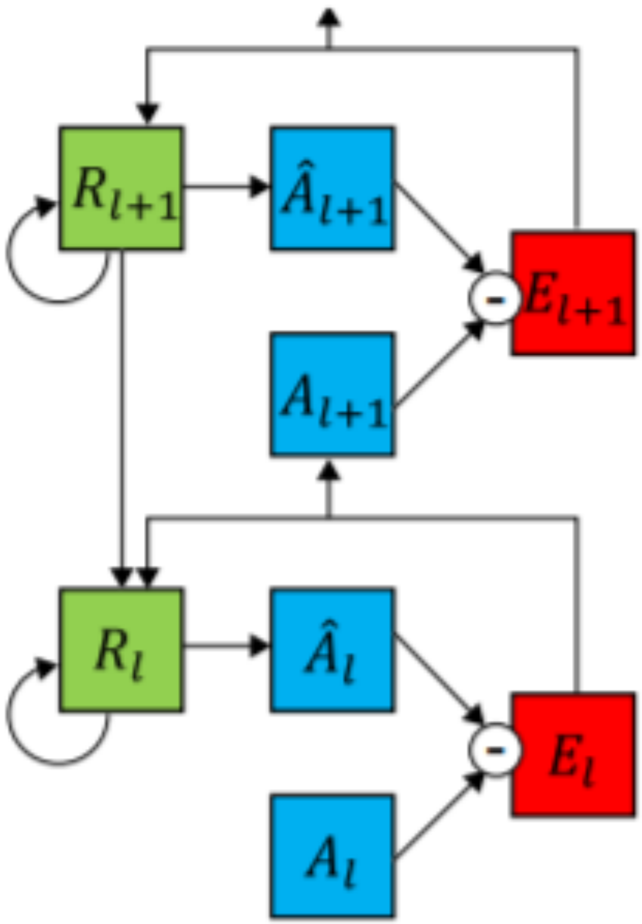
Style transfer

Gatys
2015

The portrait of Edmond de Belamy

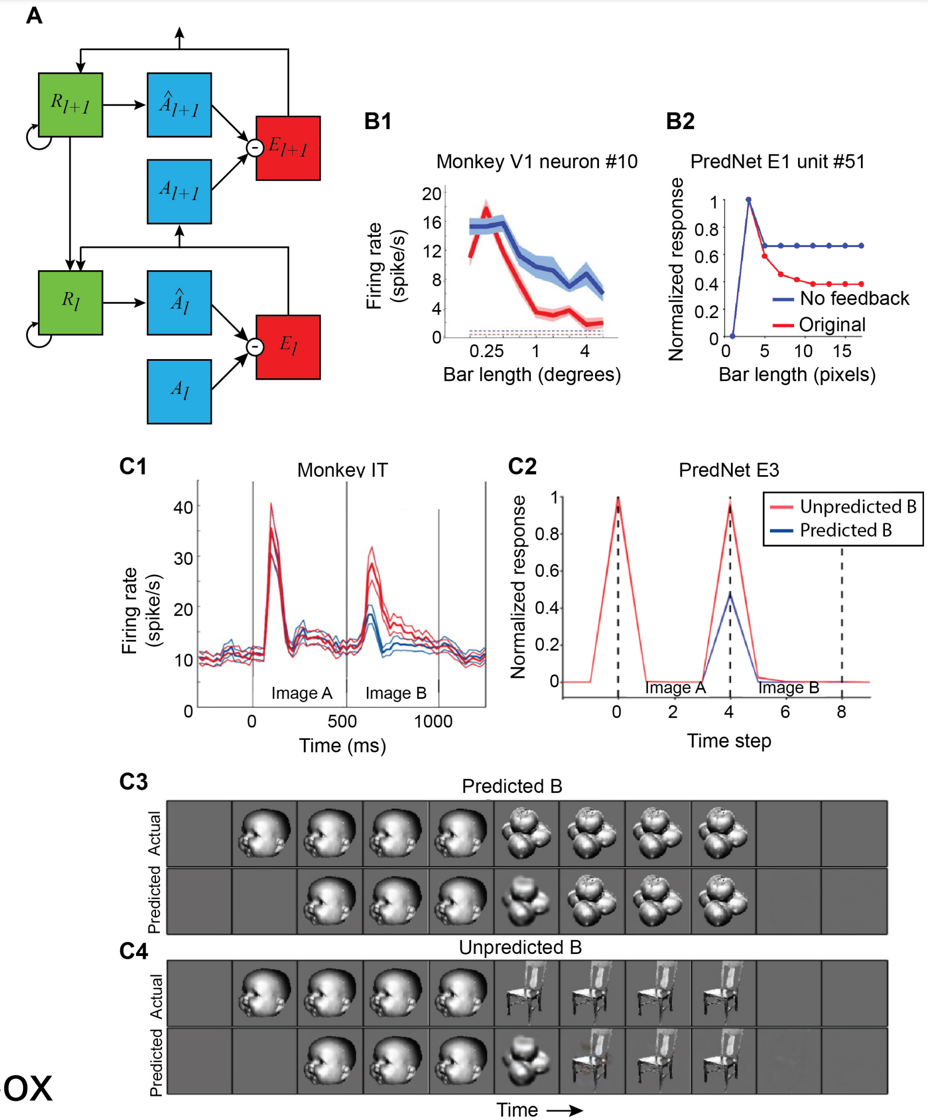
Sold at Christie’s auction: $432,500

Predicting the next video frames



William Lotter, David Cox

PredNet captures neurophysiological properties!



William Lotter, David Cox

Adversarial examples

add this “noise”

schoolbus

ostrich

Szegedy 2013

Summary

Models of ventral visual cortex provide a first order approximation to visual behavior (e.g., recognition, eye movements)

Models of ventral visual cortex provide a first-order approximation to neural responses

Computer vision has shown major strides in the last decade in many applications

- Face recognition
- Clinical diagnosis
- Object segmentation
- Tracking behavior
- Action recognition

Inverting recognition models yields powerful image generators

A model that predicts what will happen next can learn in a self-supervised manner and captures fundamental responses in visual cortex

Visual Object Recognition

Computational Models and Neurophysiological Mechanisms

Neuro 130/230. Harvard College/GSAS 78454