

Visual Object Recognition

Computational Models and Neurophysiological Mechanisms

Neurobiology 230. Harvard College/GSAS 78454

Class 1 [09/01/2021]. Introduction to Vision

Note: no class on 09/06/2021

Class 2 [09/13/2021]. Natural image statistics and the retina

Class 3 [09/20/2021]. The Phenomenology of Vision

Class 4 [09/27/2021]. Learning from Lesions

Class 5 [10/04/2021]. Primary Visual Cortex

Note: no class on 10/11/2021

Class 6 [10/18/2021]. Adventures into *terra incognita*

Class 7 [10/25/2021]. From the Highest Echelons of Visual Processing to Cognition

Class 8 [11/01/2021]. First Steps into in silico vision [Will Xiao]

Class 9 [11/08/2021]. Teaching Computers how to see

Class 10 [11/15/2021]. Computer Vision

Class 11 [11/22/2021]. Connecting Vision to the rest of Cognition

Class 12 [11/29/2021]. Visual Consciousness

FINAL EXAM, PAPER DUE 12/14/2021. No extensions.

Some computer vision problems

Classification



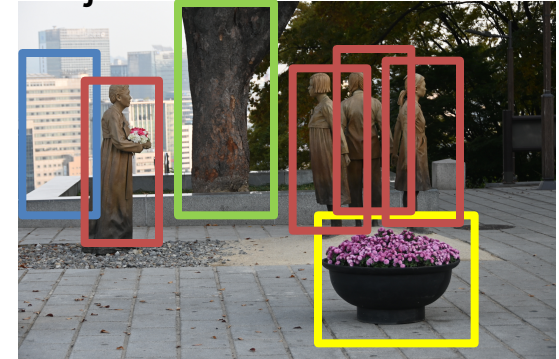
Flowers

Classification+Localization



Flowers

Object detection



Building People Tree Flowers

Instance segmentation

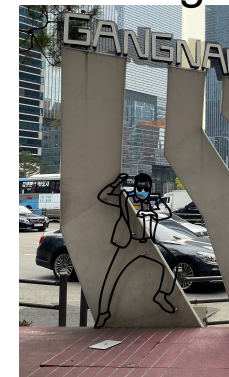


Face identification



J. Aniston
R. Whitherspoon

Action recognition



Dancing

Many more...

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

Note: lots of parameters!!!

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

Data, data, data



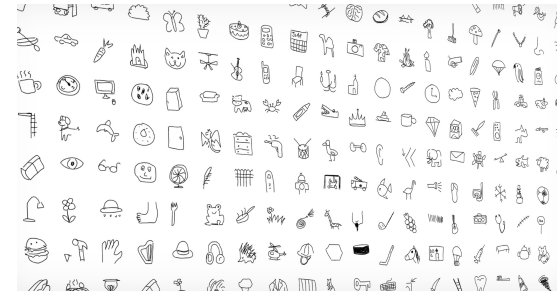
MNIST



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

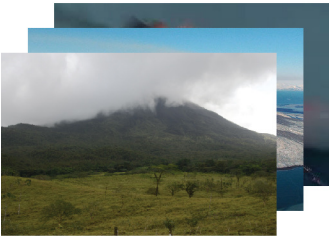


Weevil

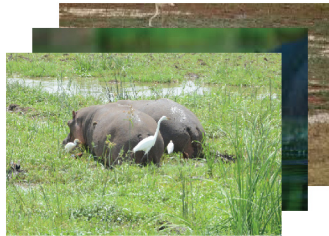


ImageNet

980: volcano



344: hippopotamus



538: dome



380: African elephant



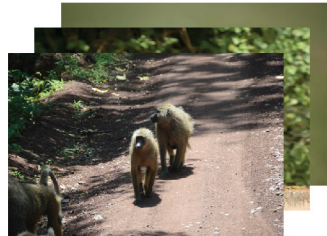
837: sunglasses



340: zebra



372: baboon



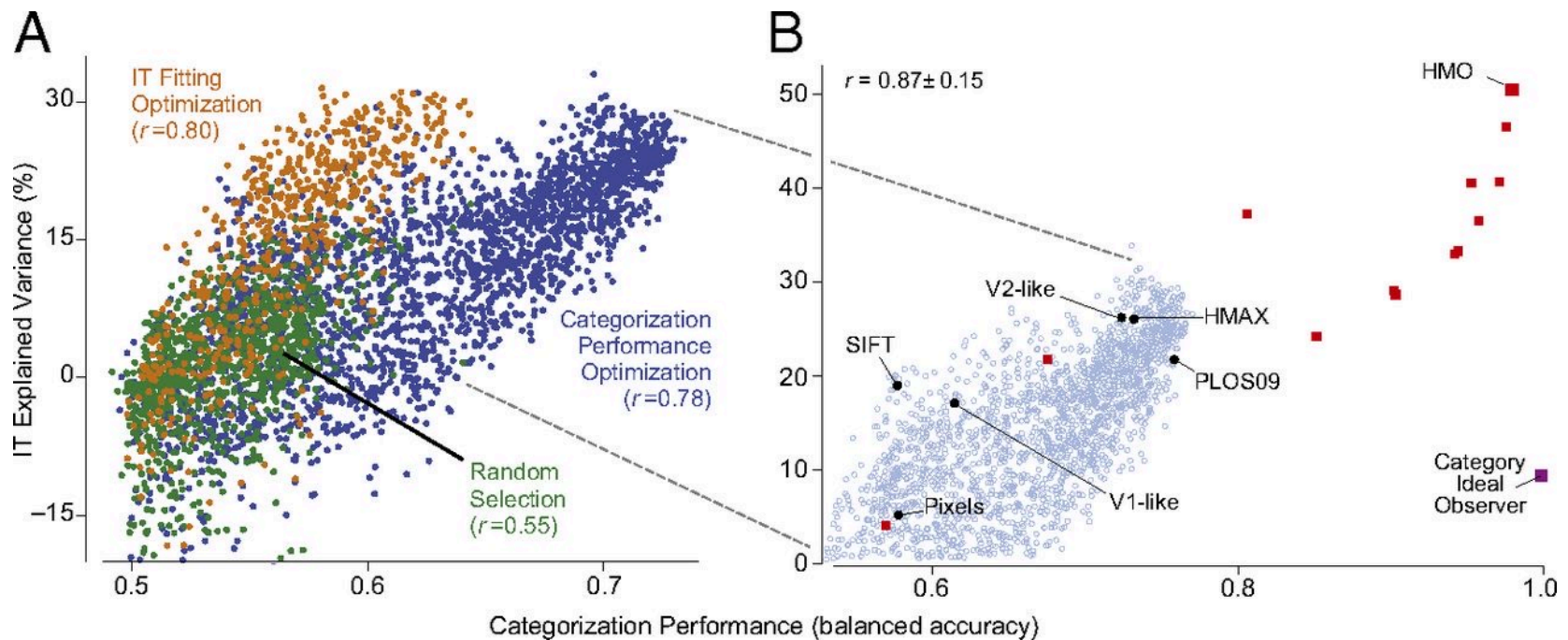
483: castle



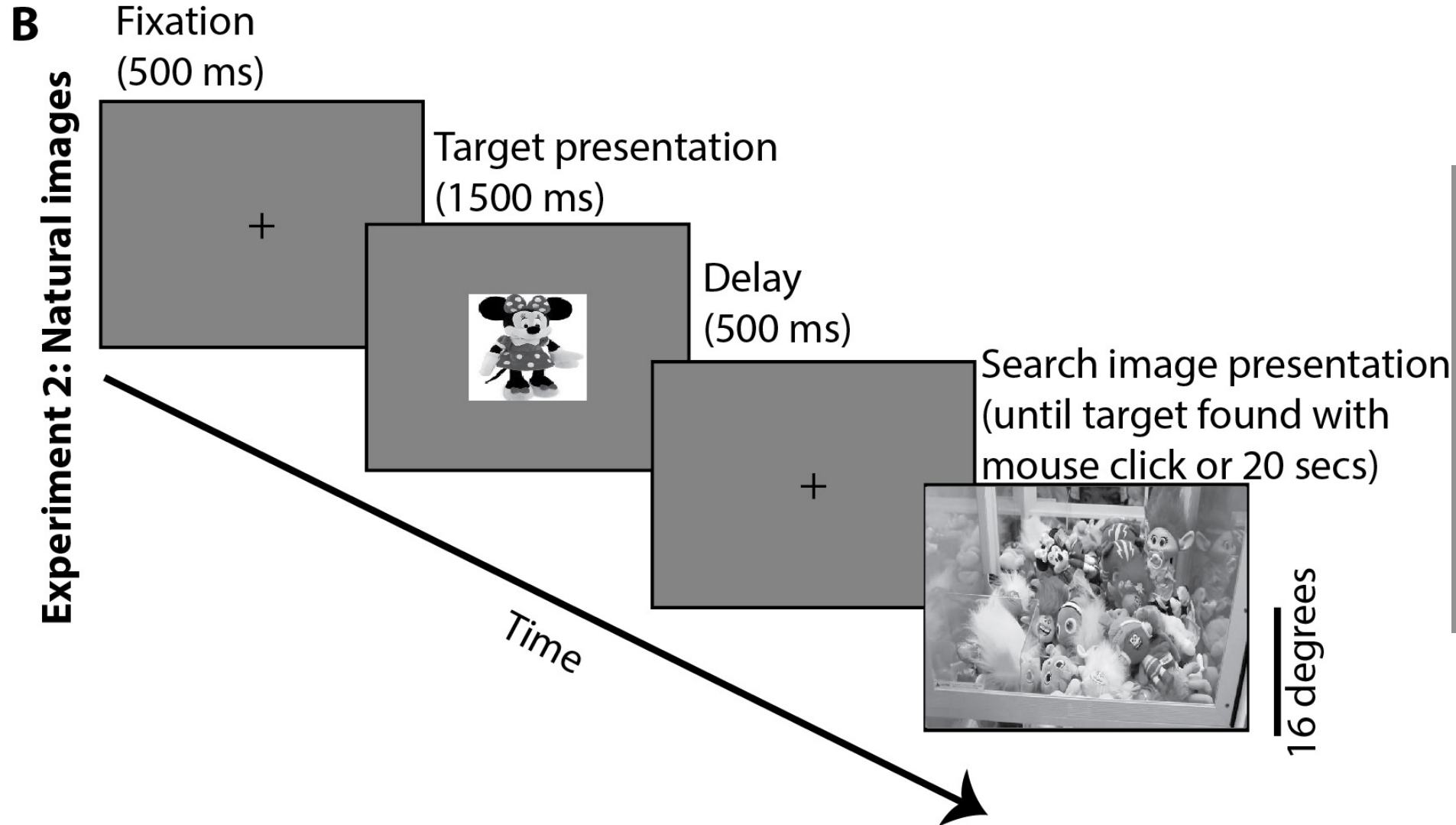
~1,000 categories

~1,000 images/category

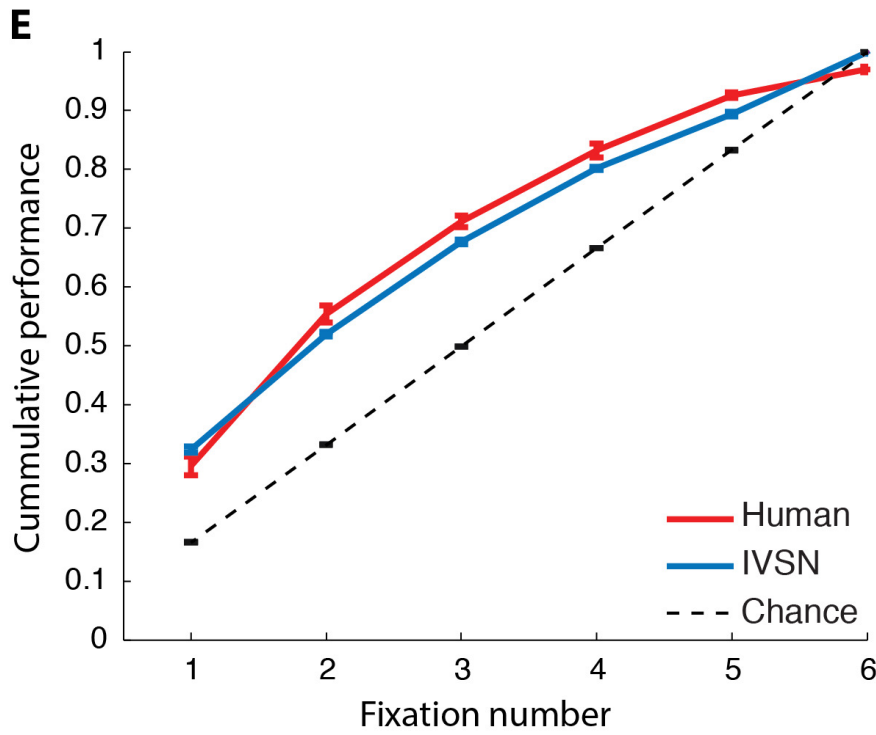
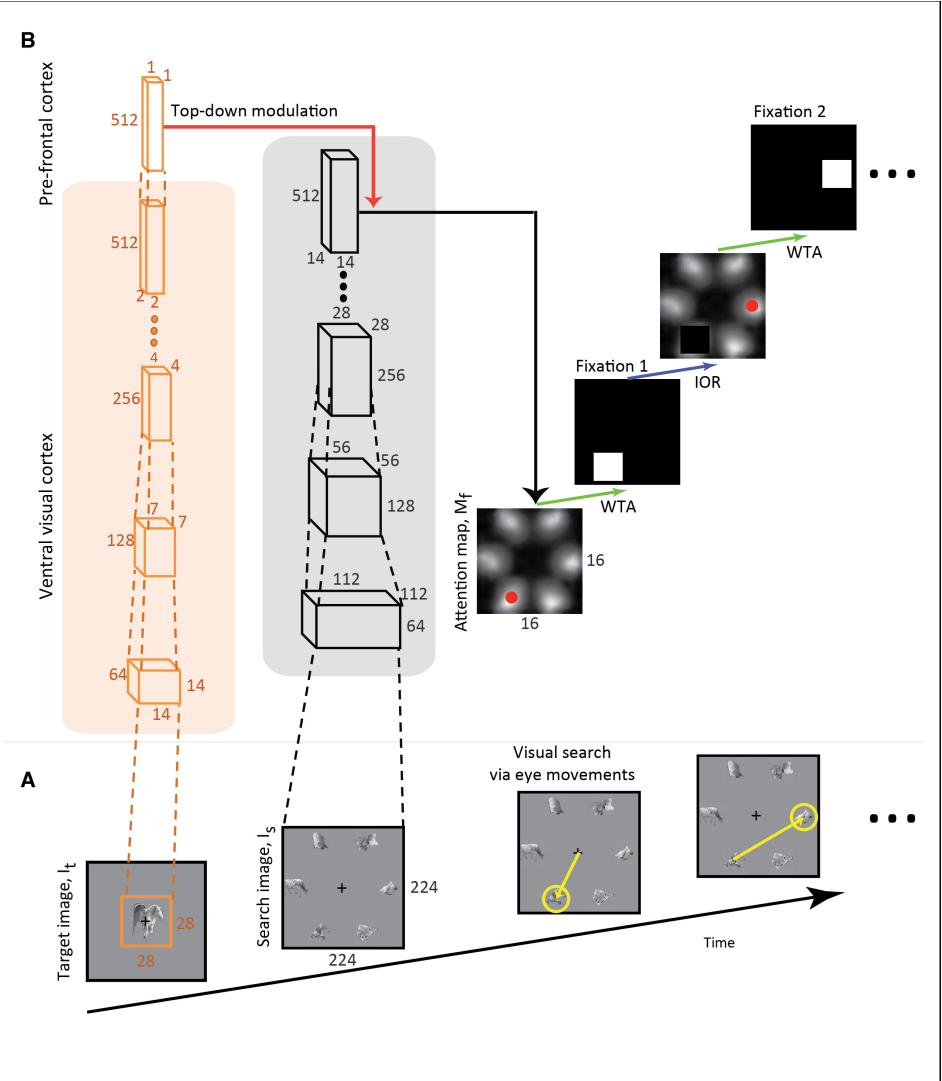
The better the biological approximation the better performance in computer vision tasks



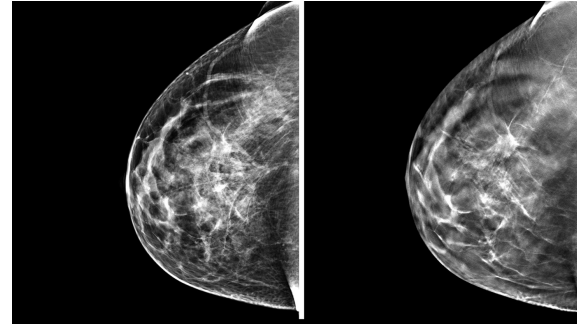
Predicting eye movements during visual search



Predicting eye movements during visual search



Machines surpass humans in pattern recognition tasks

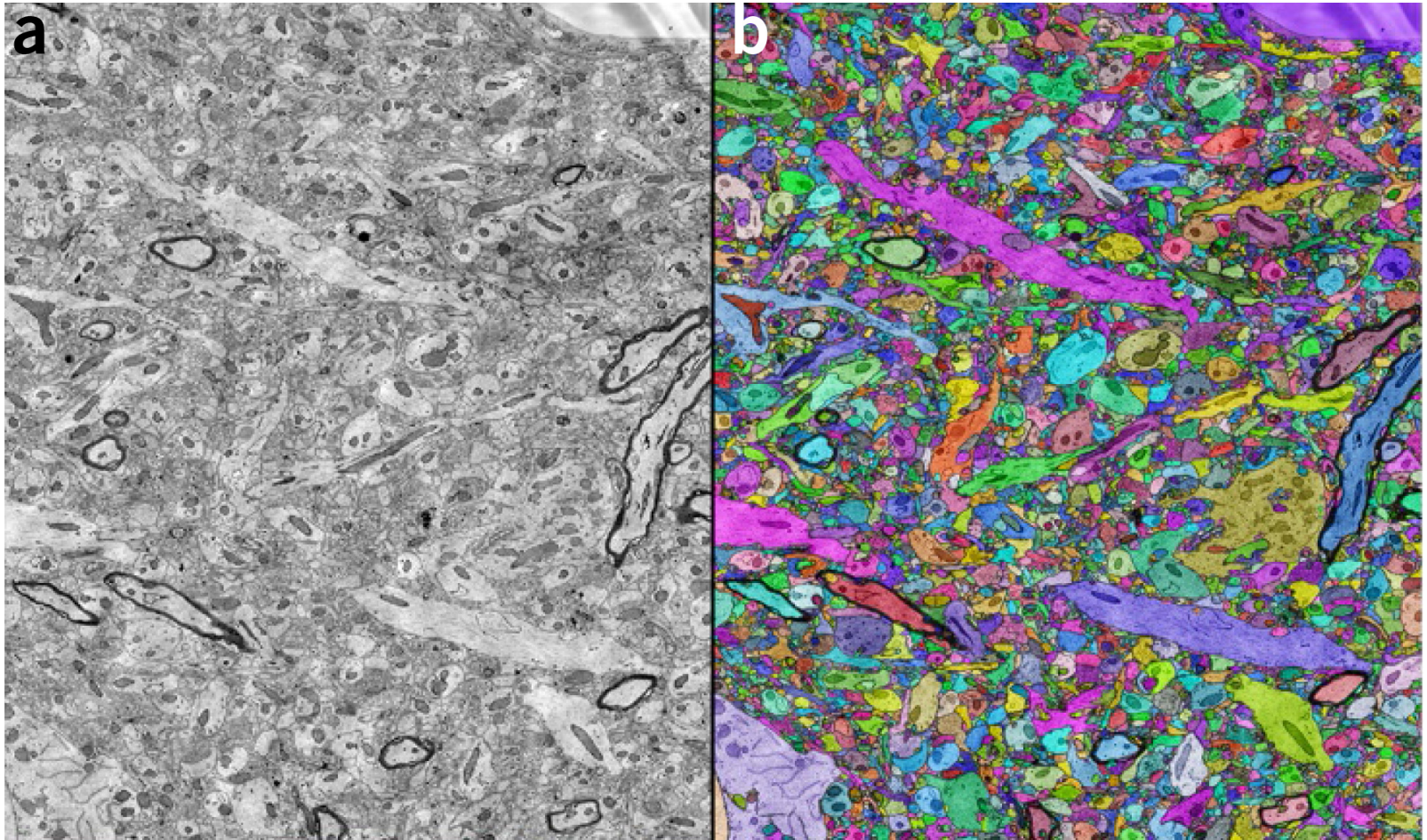


Face recognition better than forensic experts and human “superrecognizers” (Phillips et al 2018)

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)

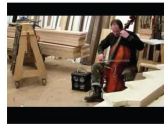
Computer vision can help segment biological images



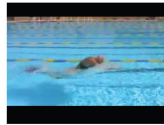
Computer vision for action recognition

A

PlayingCello 65



BreastStroke 62



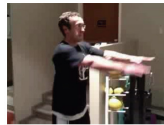
BenchPress 56



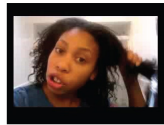
BrushingTeeth 91



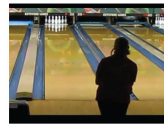
BodyWeightSquats 101



BlowDryHair 8 5



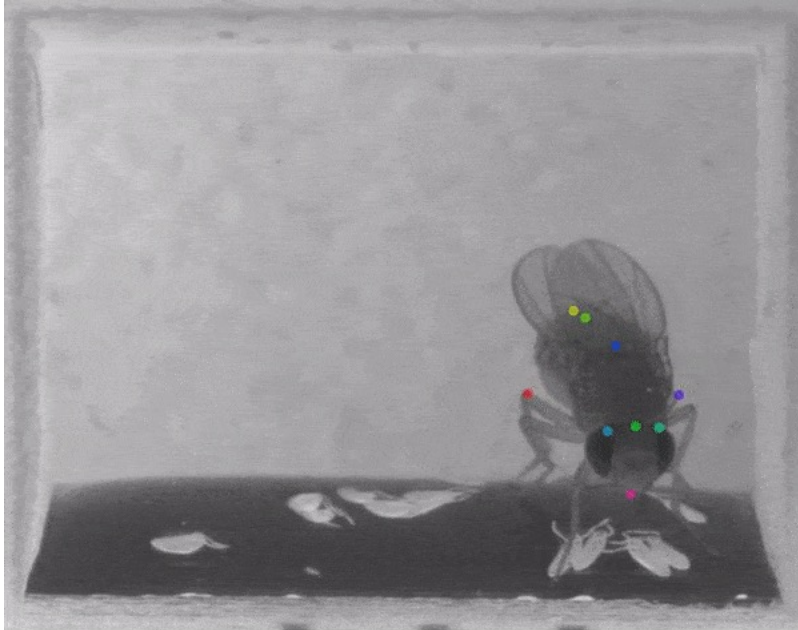
Bowling 13



SoccerJuggling 62



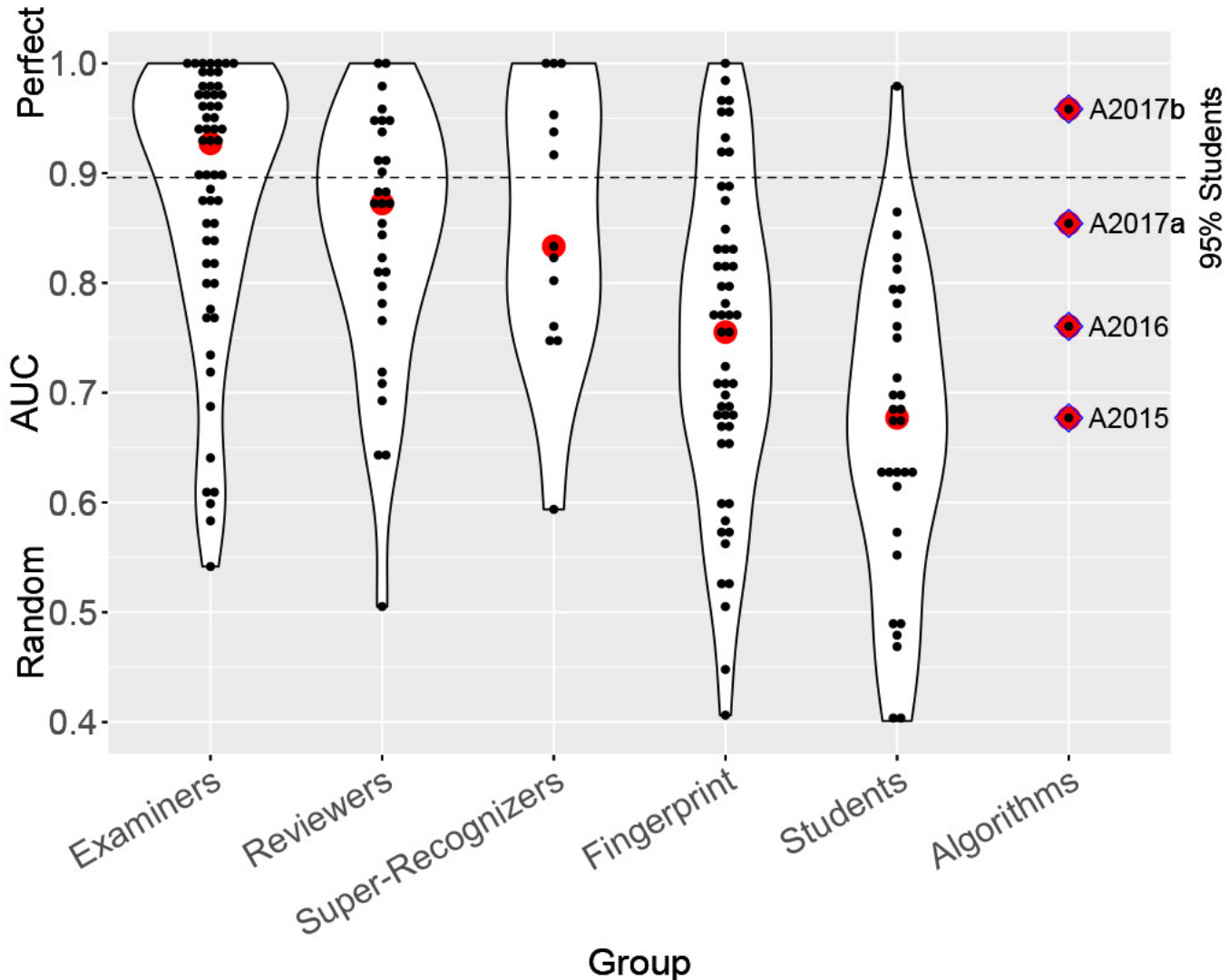
Automatic pose estimation for ethology research



Face recognition by computer vision

Same or different?

Face recognition by computer vision



Species classification and detection



Species classification and detection

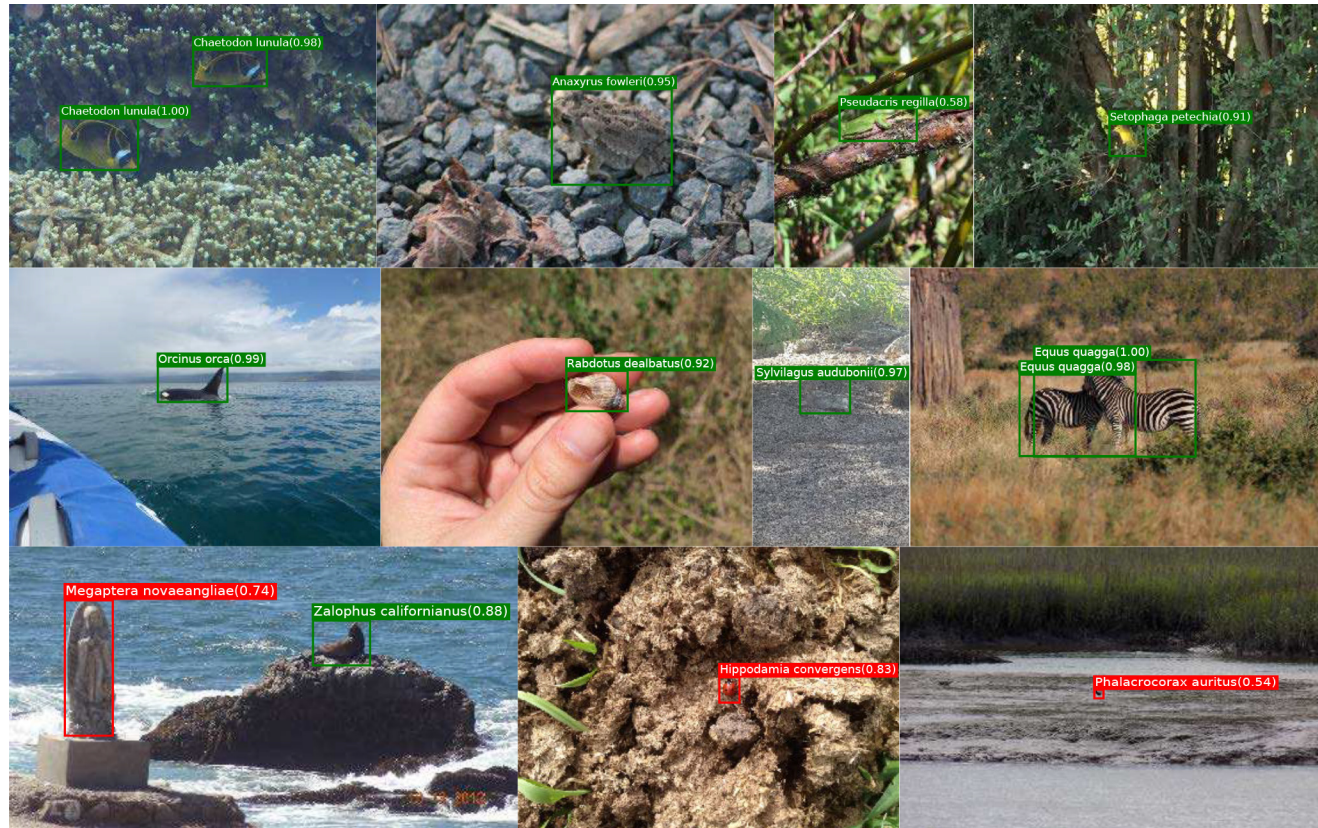
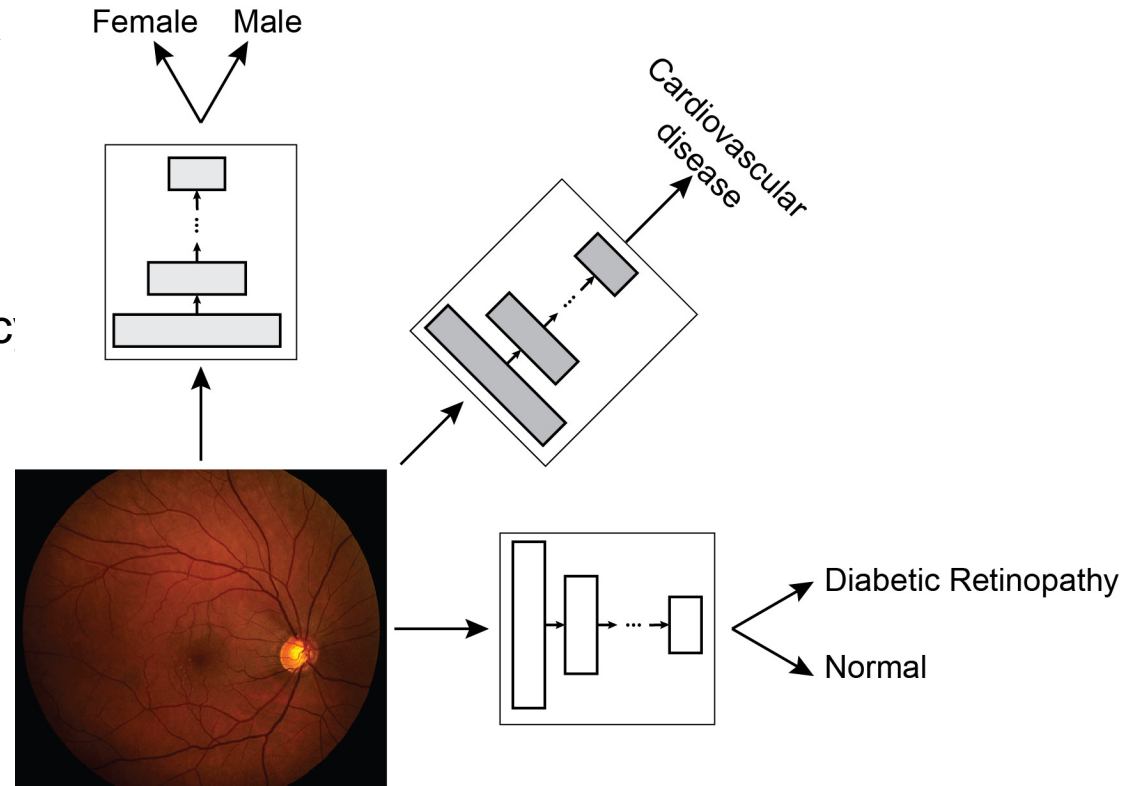


Figure 7. Sample detection results for the 2,854-class model that was evaluated across all validation images. Green boxes represent correct species level detections, while reds are mistakes. The bottom row depicts some failure cases. We see that small objects pose a challenge for classification, even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

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, accuracy
- 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?

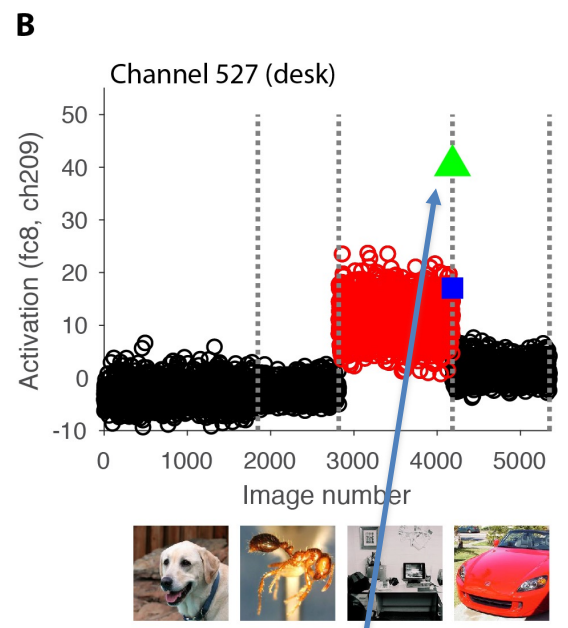
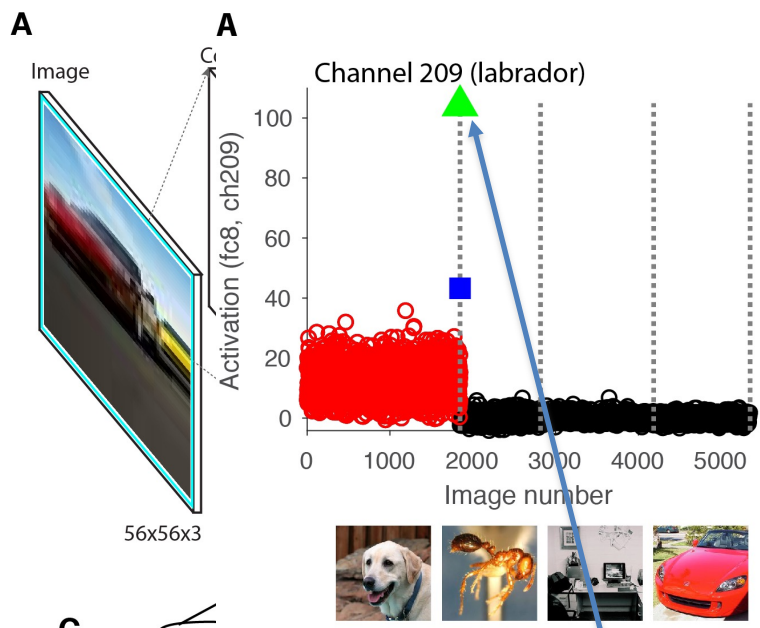


Generative adversarial networks (GANs)

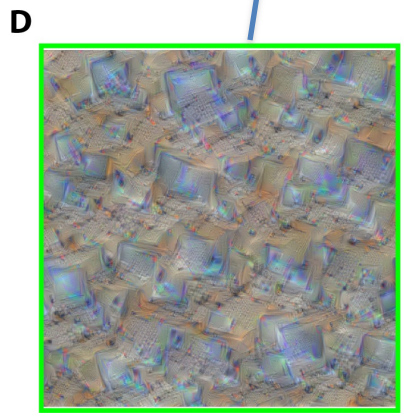
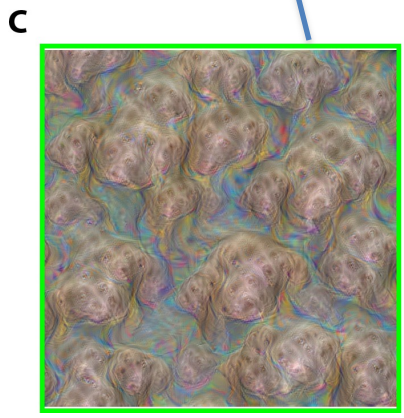


Goodfellow 2014

Deep Dreaming

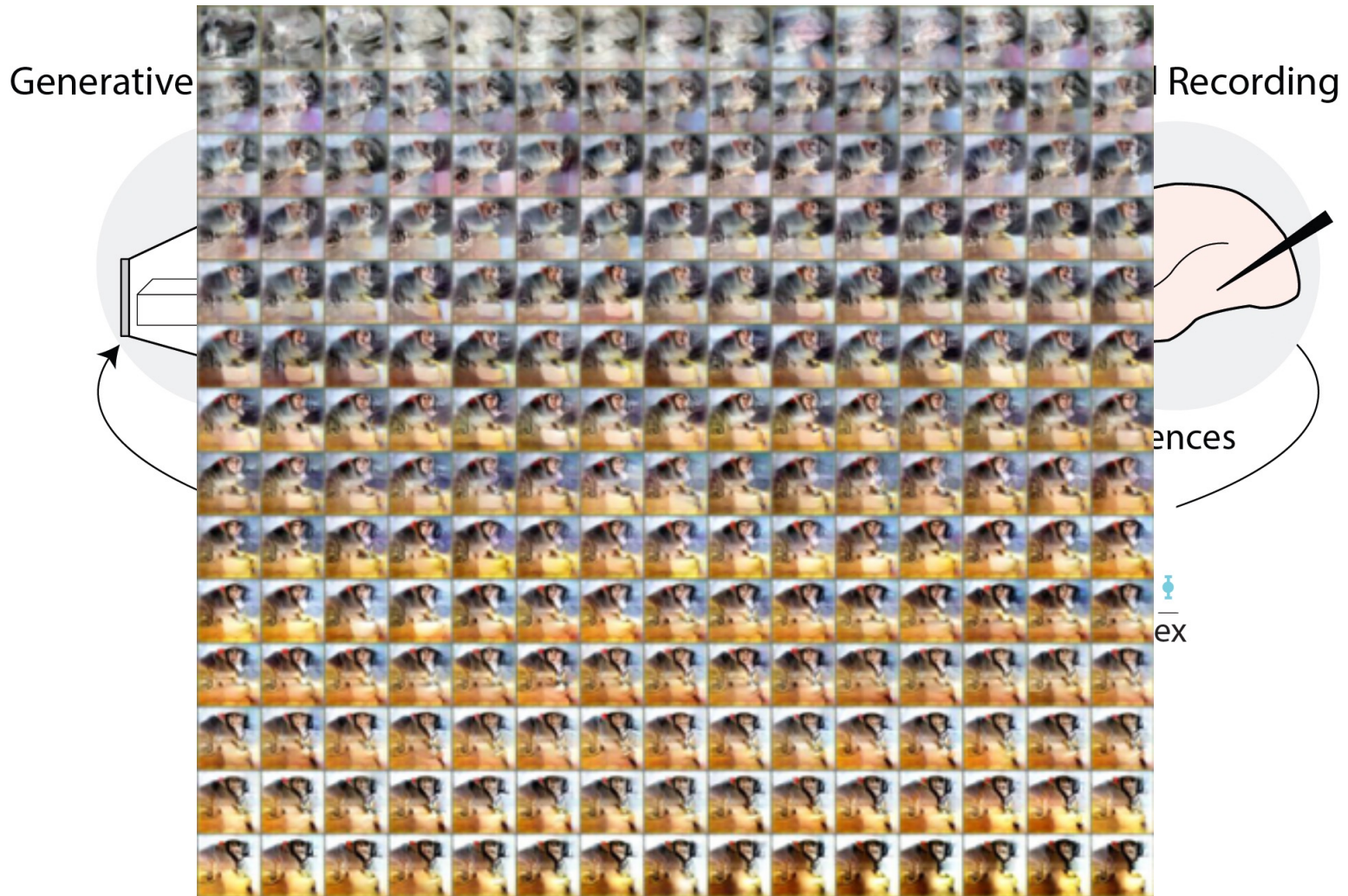


C

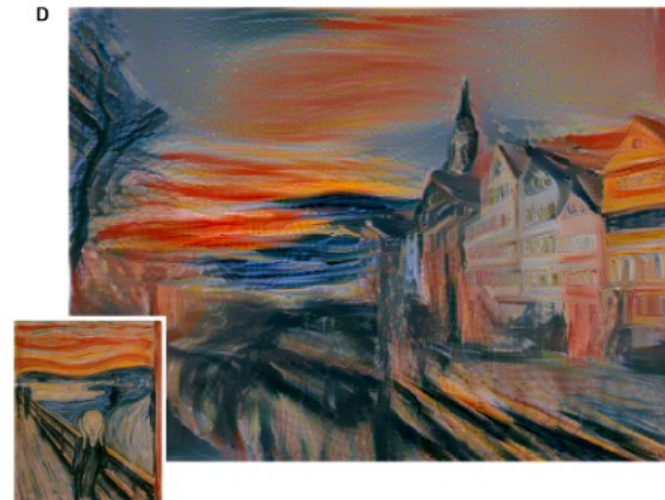


r)

Xdream: Discovering neuronal tuning preferences



Style transfer



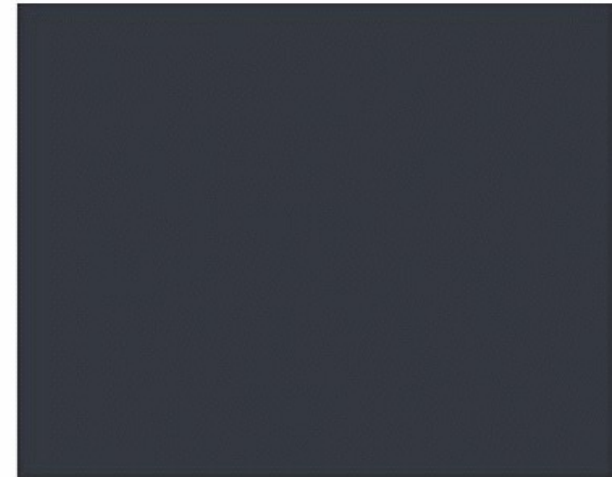
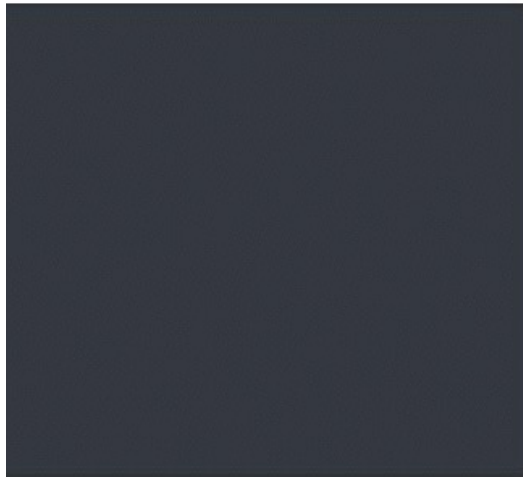
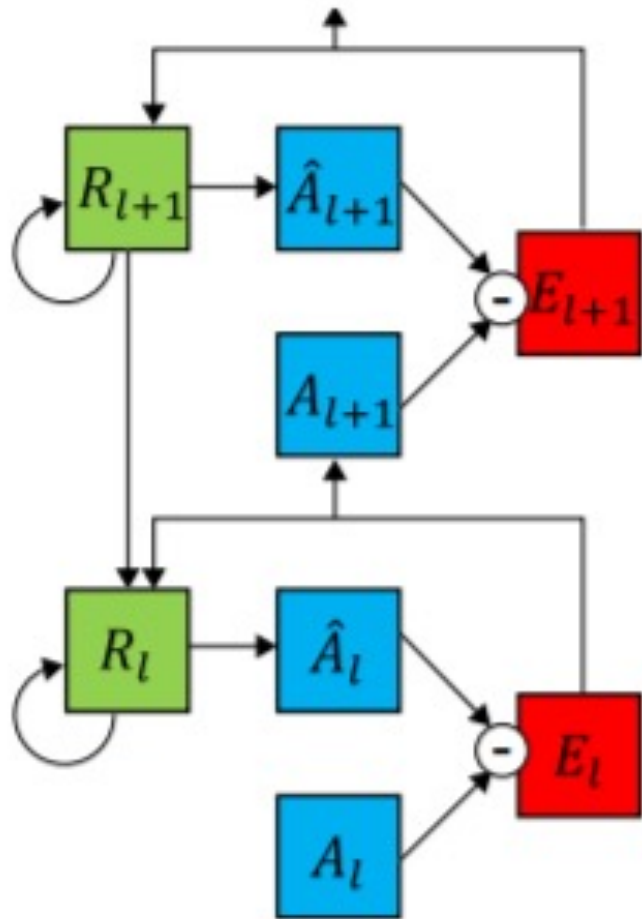
Gatys
2015

The portrait of Edmond de Belamy

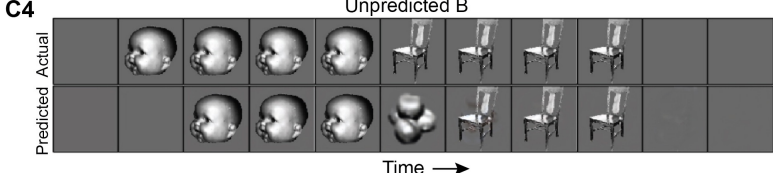
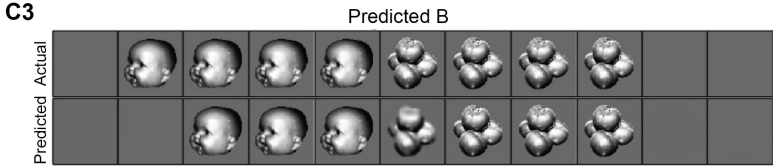
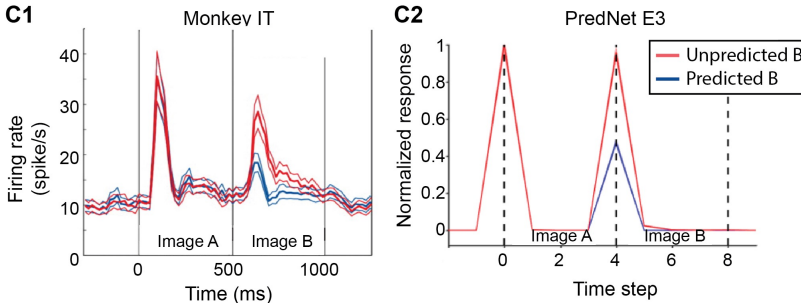
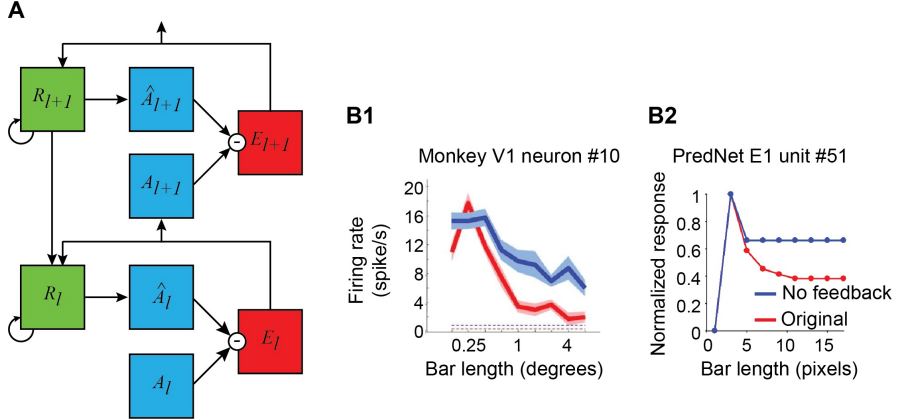


Sold at Christie's auction: \$432,500

Predicting the next video frames



PredNet captures neurophysiological properties!



Adversarial examples

schoolbus

add this "noise"

ostrich



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

