Which one of these images is fake? Type your answer in the chat

1
2
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


FINAL EXAM, PAPER DUE 12/14/2021. No extensions.
Some computer vision problems

- Classification
- Classification + Localization
- Object detection
- Instance segmentation
- Face identification
- Action recognition

Many more...
- Face detection
- Distance estimation
- Video prediction
- Image captioning
Many more architectures

- VGG (2014) – 6.8%
  - Baidu (2015) – 5.33%

- MSRA (2015) – 4.94%
  - Spatial pyramid pooling
  - Optimized PReLU
  - Improved (random) initialization

- GoogleNet (2014) – 6.67%
  - Inception module
  - Multi-scale convolutions (including 1x1 filters)
  - Minimal dense layers
  - Auxiliary classifiers

- AlexNet (2012) – 15.3%
  - Clarifai (2013) – 11.7%

Inception Module
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

- Galaxies
- Plants
- Clinical images
- Cell types

Many more …
- MNIST
- IMAGENET
- MSCOCO
- QUICKDRAW
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

Experiment 2: Natural images

- Fixation (500 ms)
- Target presentation (1500 ms)
- Delay (500 ms)
- Search image presentation (until target found with mouse click or 20 secs)

16 degrees

Zhang et al, 2018
Predicting eye movements during visual search

Zhang et al 2018
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

<table>
<thead>
<tr>
<th>Action</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>PlayingCello</td>
<td>65</td>
</tr>
<tr>
<td>BreastStroke</td>
<td>62</td>
</tr>
<tr>
<td>BenchPress</td>
<td>56</td>
</tr>
<tr>
<td>BrushingTeeth</td>
<td>91</td>
</tr>
<tr>
<td>BodyWeightSquats</td>
<td>101</td>
</tr>
<tr>
<td>BlowDryHair</td>
<td>5</td>
</tr>
<tr>
<td>Bowling</td>
<td>13</td>
</tr>
<tr>
<td>SoccerJuggling</td>
<td>62</td>
</tr>
</tbody>
</table>
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, 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)

D-dimensional noise vector

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

- schoolbus
- add this “noise”
- ostrich

Szegedy 2013
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

Web site: http://tinyurl.com/visionclass
Class notes, Class slides, Readings Assignments
Location: Biolabs 2062
Time: Mondays 03:00 – 05:00
Lectures: Faculty: Gabriel Kreiman (and invited guests)
TA: Will Xiao
Contact information: Gabriel Kreiman
gabriel.kreiman@tch.harvard.edu
Will Xiao
xiaow@fas.harvard.edu
617-919-2530
Office Hours: Before class (Mondays 2pm), after class (Mondays 5pm). By appointment