ObjectNet: Understanding object detection performance at scale in humans and machines

Faculty: Boris Katz1, Josh Tenenbaum2, Dan Gutfreund3
Research Scientists / Postdocs: Andrei Barbu1, Guy Ben-Yosef1,2
Graduate students: David Mayo1, Battushig Myanganbayar1, Cristina Mata1, Gil Dekel1
1. Computer Science and AI Laboratory, Massachusetts Institute of Technology — 2. Dept. of Brain and Cognitive Sciences, Massachusetts Institute of Technology — 3. IBM Watson Cambridge

1. Top of the line object detector today
YOLO9000: Better, Faster, Stronger. Joseph Redmon and Ali Farhadi, Received an award at CVPR 2017. ≈ 90% top-5 accuracy on ImageNet (75% top 1)
Nowhere near human-level vision. Benchmarks tell us we’re done but experience shows we’re just starting.

2. Human-level object recognition back in 2015!
Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification, Shaoqing Ren, Kaiming He, Ross Girshick, Xiangyu Zhang, Jian Sun, ICCV 2015. 4.94% top-5 machine error on ImageNet
5.1% human-level performance on ImageNet
Today performance is at or below 3% error!

3. Evaluate object detection like a psychologist
Image

Human

Machine

What object do you see?
Classify this image

Which of these two were describing this image?
This is an easy, automatic, natural task for humans. It really tests object detection not labeling. No need for top-N.

4. ObjectNet, 100k images (50k now), 300 object classes

5. Arbitrary correlations make the task too easy
Preferred rotations
Similar lighting and imaging conditions
Similar occlusions
Similar backgrounds
Bad data
Graduate student descent
Similar context

6. Preliminary results
We control for object class, object orientation, camera orientation, background, etc.

7. Collecting natural data while controlling for biases

8. Next steps in understanding human vision
Accuracy(Foreground, Background, Time)

9. Hands: A new dataset for investigating occlusion and deformable objects
Machine on previous datasets (5px)
Human on our dataset (10px)
Machine on our dataset (20px)

10. Contributions to Science and Engineering of Intelligence
The first large-scale measurements of human and machine accuracy. Datasets are critical to ML, we are working to adopt modern standards for this ad-hoc discipline. Knowing how well models works is critical for safety, dataset performance should be predictive. Human performance can help explain the representations used and the organization of the brain. We can use human data to help constrain network architecture search, discovering new NNs. Non-human primate and animal performance can point toward principles behind object classes.

This work is funded, in part, by the National Science Foundation, award CCF-1231216.