Datasets overstate real-world performance

Throughout machine learning disciplines performance on datasets is far higher than the performance a user can expect in the real world. Transferring models between datasets and recollecting datasets leads to a large performance drop. While models have changed radically, dataset collection methods have not. It’s time that they do.

Object detectors fail in the real world

On benchmarks like ImageNet object detectors beat humans, but they aren’t anywhere close to that. The number of false positives and false negatives is huge in the frames below (YOLO).

The consequences can be severe!

Debiased data is much more varied!

We tell users how to orient objects, where to put them, and how to image them, randomly varying these properties.

ObjectNet has no training set!

ImageNet vs ObjectNet: bottle opener

Much more challenging for detectors!

A new feedback signal to make progress in object detection, with no training set! Generalization to new datasets should be our main concern.

Fine-tuning doesn’t save you!

Using 16 images per class, which saturates most existing datasets, only recovers 15% of the performance.

The numbers are wildly optimistic, we only do transfer learning for 113 classes which overlap ImageNet.

A fairer comparison would transfer all of the classes, 28% top-1 when transferring on all classes with fine-tuning on 16 images per class.

Data collection with controls

ObjectNet is hard because of the controls, object detectors aren’t able to learn good representations.

Preliminary results show human performance is about 95% Which features do humans use, how different is neural activity for easy and hard images?

Can we use human data to regularize the representations detectors learn?

Let’s take cross-dataset generalization, safe & predictable vision, with fewer object & cultural biases seriously. What new detectors will this lead to?

This work was supported, in part, by the Center for Brains, Minds and Machines, NSF STC award 1231216, the MIT-IBM Watson AI Center, the Toyota Research Institute, and the CBMM-Siemens Graduate Fellowship. The views and conclusions contained in this document are those of the authors and do not represent the official policies, either express or implied, of the U.S. Government or any other funding agency. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes, notwithstanding any copyright notation herein.