Detecting avocados to zucchinis: what have we done, and where are we going?

Olga Russakovsky\(^1\) Jia Deng\(^1\) Zhiheng Huang\(^1\) Alexander C. Berg\(^2\) Li Fei-Fei\(^1\)

\(^1\)Stanford University \(^2\)UNC Chapel Hill

**Introduction**

**Motivation**
Large-Scale Recognition is a grand goal of computer vision. Benchmarking and analysis measure progress and inform future directions.

**Goal**
The goal is to analyze and compare performance of state-of-the-art systems on large-scale recognition.

**Dataset**
The ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) 2012 is much larger and more diverse than previous datasets.

**PASCAL VOC 2005-2012**
- 20 object classes: 22,591 images
- 1000 object classes: 1,431,167 images

Classification-localization challenge (ILSVRC2012)
- Task: To determine the presence and location of an object class.

**Analysis setup**

**Why run analysis?**

**Reason 1:** Surprisingly strong performance of the winning entry.

**Reason 2:** The scale of 1000 object categories allows for unprecedented look at how objects properties affect accuracy of leading algorithms.

**State-of-the-art large-scale object localization algorithms**

**Supervision (SV)** by A. Krizhevsky, I. Sutskever, G. Hinton
- Classification: Deep convolutional neural network; 7 hidden layers, rectified linear units, one pooling, dropout, trained with SGD on two GPUs.
- Localization: regression on (x, y, w, h).

**Defend-DEEP (DDEEP)** by K. Simonyan, Y. Aytar, A. Vedaldi, A. Zisserman
- Classification: ResNet-L; color statistics, Fisher vector (1024 Gaussian), product quantization, linear SVM, one vs rest SVM, trained with Pegasos.
- Localization: Deformable parts model, root-only.

**What images are difficult?**

**Protocol**
- For every one of the 1000 classes:
  - Ask humans to annotate different properties, e.g., is this object deformable? (x)
  - Compute accuracy of algorithms on test images (y)

**Upper bound (UB)**
Currently the output of SV and VGG (using an oracle) to demonstrate the current limit of object localization accuracy.

**Where are we going?**

- Cluttered images remain very challenging for object localization
  - Proposed measure of clutter can be used for creating and evaluating datasets.
- Untextured and man-made objects are still challenging even for the best algorithms.
- Complementary advantages of SV and VGG can be used to design the next generation of detectors:
  - SV algorithm is very strong at learning object texture, and
  - VGG algorithm is less sensitive to number of instances and object scale.
- ILSVRC dataset is a promising benchmark for detection algorithms.

**Bibliography**