The Unreasonable Effectiveness of Noisy Data for Fine-Grained Recognition

Jonathan Krause¹, Benjamin Sapp², Andrew Howard², Howard Zhou², Alexander Toshev², Tom Duerig², James Philbin², Li Fei-Fei¹

¹Stanford University, ²Google

Problem
-Fine-grained recognition works well with labels
-But fine-grained labels are expensive
-There are too many fine-grained categories in the world to annotate by hand: 14k birds, 278k butterflies and moths, 941k insects
-How can we scale up fine-grained recognition?

Data

Categories
Birds: 10,982 species
Butterflies: 14,553 species (+moths)
Aircraft: 409 varieties
Dogs: 515 breeds
-Images from Google Image Search

Noise
Cross-domain noise: portion of images that are not of any fine-grained category in a given domain. Measure by hand.

Cross-category noise: portion of images that have the wrong fine-grained label. Hard to estimate.

Filtering: Proposed technique to reduce noise. Simply remove images with multiple web labels!

Experiments

-Use Inception-v3 CNN classifier
-Extensive dedup with ground truth test datasets via [2]
-YFCC100M data for active learning

Prior work on GT datasets:
CUB: 84.6% (Xu et al. ICCV’15)
Birdsnap: 66.6% (Berg et al. CVPR’14)
FGVC: 84.1% (Lin et al. ICCV’15)
Stanford Dogs: 76.8% (Sermanet et al. ICLR’15)

Very Large-Scale Fine-Grained Recognition
-Test on Flickr images w/exact category name matches, deduped with other web images.

Categories
Birds: 10,982 species
Butterflies: 14,553 species (+moths)
Aircraft: 409 varieties
Dogs: 515 breeds

-Images from Google Image Search

Noise
Cross-domain noise: portion of images that are not of any fine-grained category in a given domain. Measure by hand.

Cross-category noise: portion of images that have the wrong fine-grained label. Hard to estimate.

Filtering: Proposed technique to reduce noise. Simply remove images with multiple web labels!

Active Learning
Alternative approach for collecting large quantities of fine-grained data.

We use confidence-based sampling: Label the most confident images. Can still fix false positives, and uncertain images are hard to label!

Contributions
-Demonstrate feasibility of training models of fine-grained with noisy data from the web and simple, generic, models of recognition.
-Greatly improved recognition performance on four fine-grained datasets without using ground truth training data.
-Scale fine-grained recognition to over 10,000 species of birds and 14,000 species of butterflies and moths.

References
[1] Szegedy et al. Rethinking the Inception Architecture for Computer Vision. CVPR 2016
[2] Wang et al. Learning Fine-Grained Image Similarity with Deep Ranking. CVPR 2014
[3] Wall et al. The Caltech-UCSD Birds-200-2011 Dataset. Tech. Report 2011
[4] Berg et al. Birdsnap: Large-Scale Fine-Grained Visual Classification of Birds. CVPR 2014
[5] Maji et al. Fine-Grained Visual Classification of Aircraft. Tech. Report 2013
[6] Koles et al. Novel Dataset for Fine-Grained Classification. FGVC 2011

4,224 (+1) categories recognized in this work