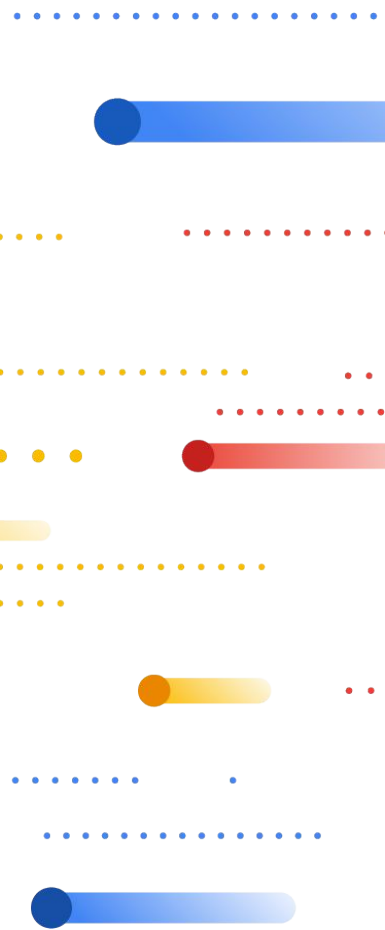


OPEN

MAGES

CHALLENGE 2018



Why Open Images?

- It's open!
- Restrict to images with CC_BY license → no copyright problems
- Can even use it commercially
- Enables legally safe crowdsourcing

Why Open Images?

- Start from Flickr, then dedupe against the web
→ biased away from simple images
- Not scraped based on class keywords (all of CC on Flickr)
→ natural class statistics, no initial design bias

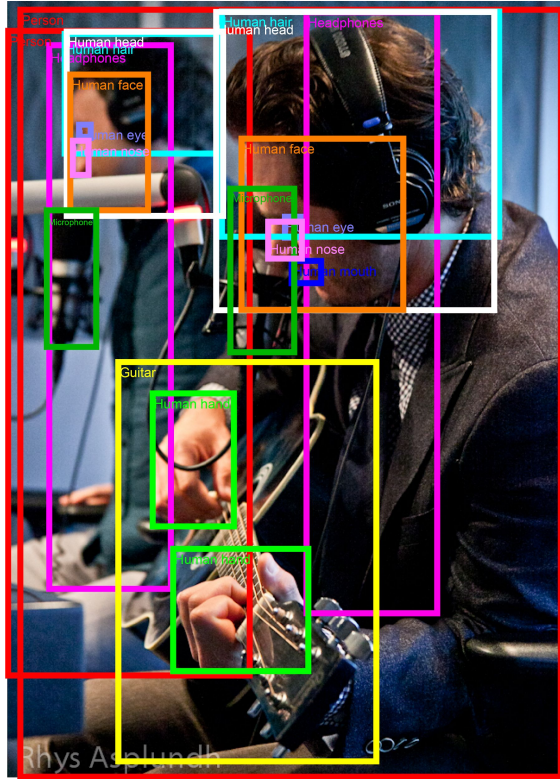


"Ball" in Google Image Search

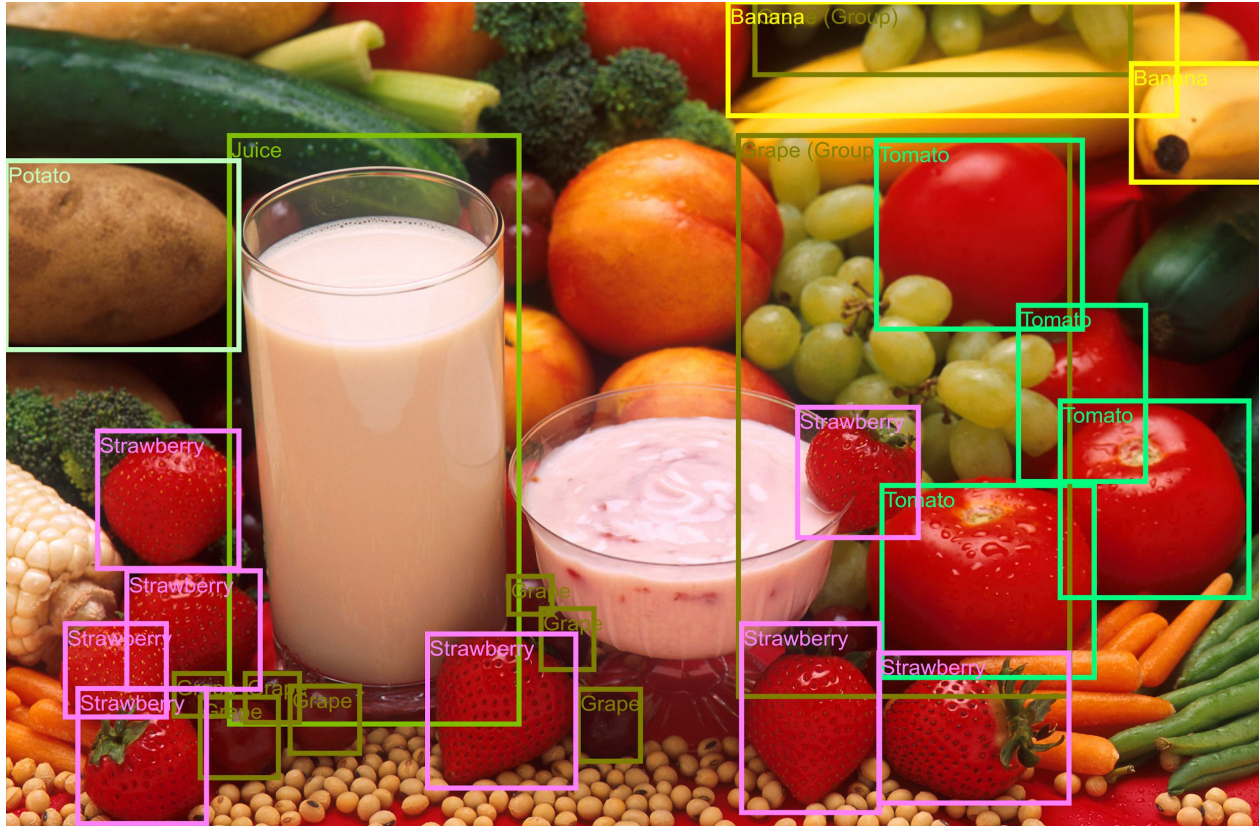


"Ball" in Open Images

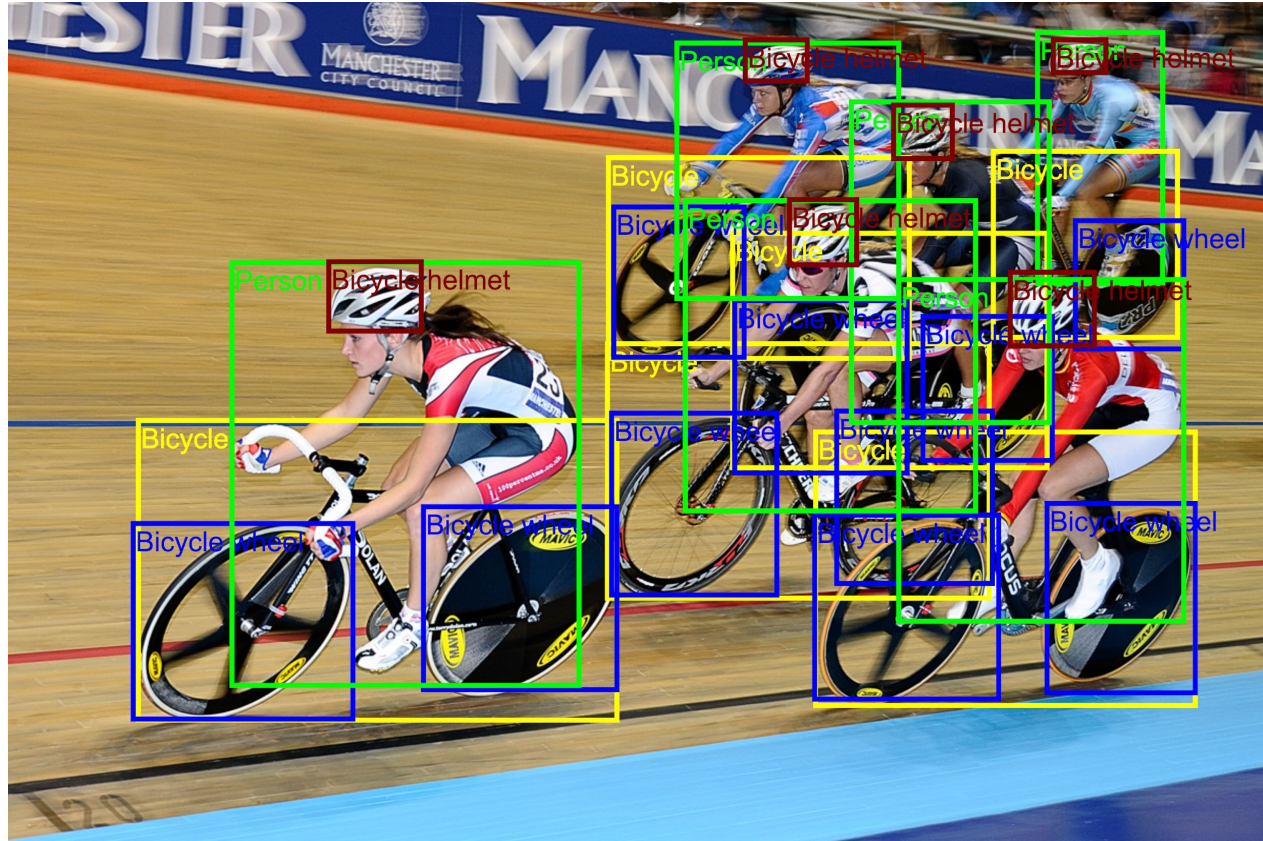
Open Images V4 dataset



Open Images V4 dataset



Open Images V4 dataset



Open Images V4 dataset (train+val+test)

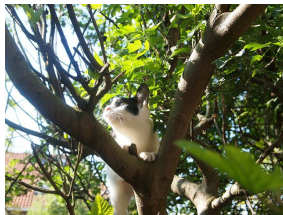
- Subset with bounding boxes:
 - 1.9 million images (1.74M train, rest val+test)
 - 15.4 million boxes
 - 8.1 boxes/image
 - 600 classes
- Subset with image-level labels:
 - 5.8 million images
 - 14.9 million labels
 - 19.8 thousand classes

Open Images V4 dataset: box annotation process

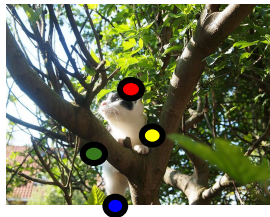
- We applied a classifier to each image and sent labels with sufficient score for human verification.
- We boxed *all* object instances of all positive labels
 - 90% using Extreme Clicking
[Papadopoulos, Uijlings, Keller, Ferrari, ICCV 2017]
 - 10% using Box Verification Series
[Papadopoulos, Uijlings, Keller, Ferrari, CVPR 2016]

→ image labels not exhaustively annotated
→ but boxed all instances for the labels we have

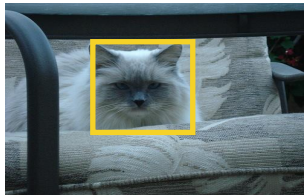
Does this image contain a cat?



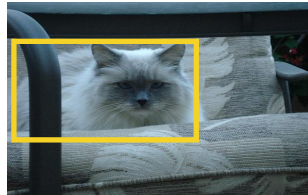
Draw extreme points on the cat



Is this a correct box for a cat?



NO



YES

Open Images V4 dataset: comparison to ILSVRC-det and COCO

	Open Images	ILSVRC-det		COCO	
Classes	600	200	3x	80	7x
Images	1,910,098	476,688	4x	123,287	15x
Boxes	15,440,132	534,309	29x	886,284	17x
Boxes per image	8.1	1.1		7.2	
		(3.0 on a 80k image subset)			

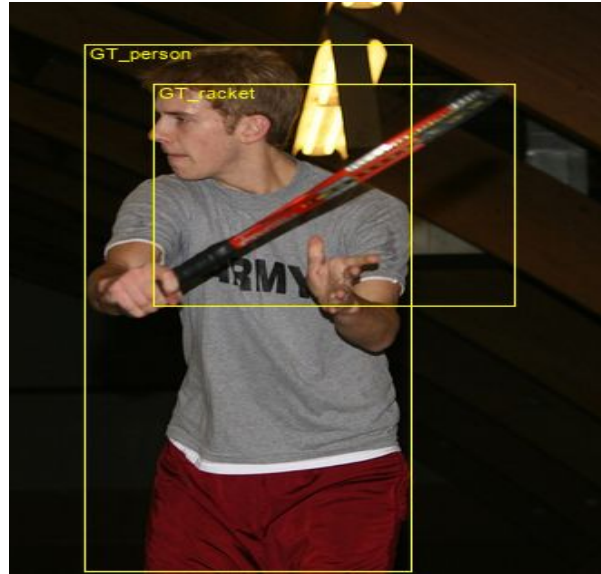
COCO has segmentations though!

Complex images (many objects per image)

Open Images V4 dataset: comparison to ILSVRC-det and COCO



Open Images



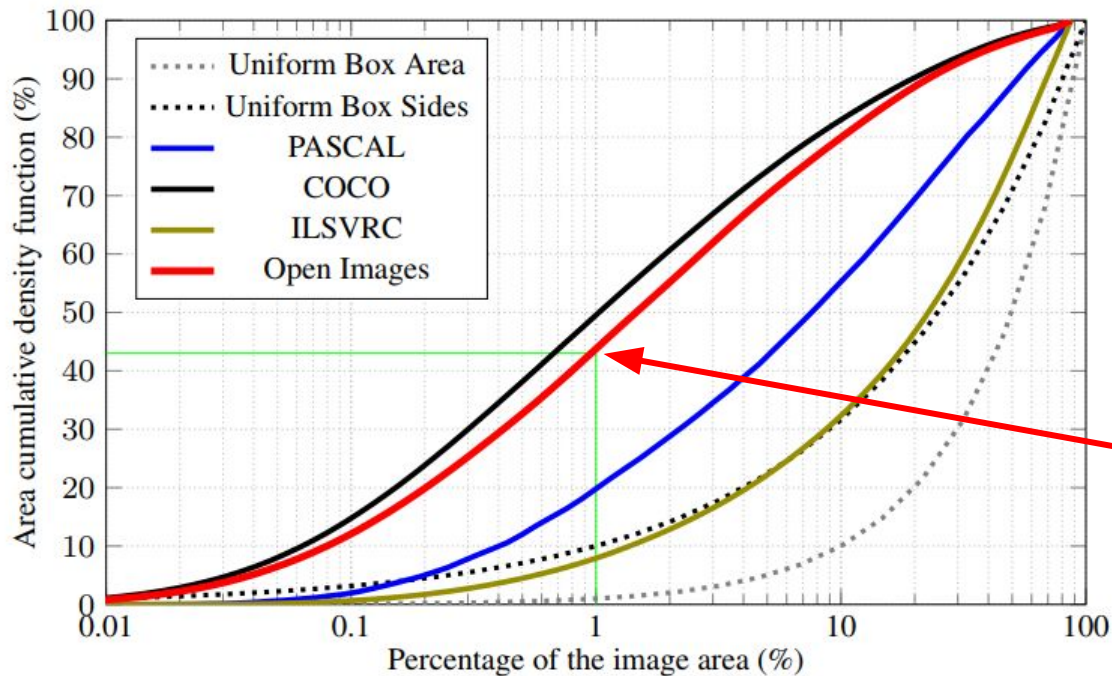
ILSVRC-Det



COCO

Open Images V4 dataset

Object size distribution similar to COCO (many small objects)



43% of the objects occupy less than 1% of the image area

Open Images Challenge 2018

Based on Open Images V4

Two tracks:

- Object Detection Track
 - Visual Relationship Detection Track
-
- 50K prize money fund
 - 4 months to further push the capabilities of computer vision



“a sandy beach, the ocean, a few people walking, some trees, grass, and buildings...a woman walking her dog right there! Oh yeah, and there is a man holding a plastic cup.”

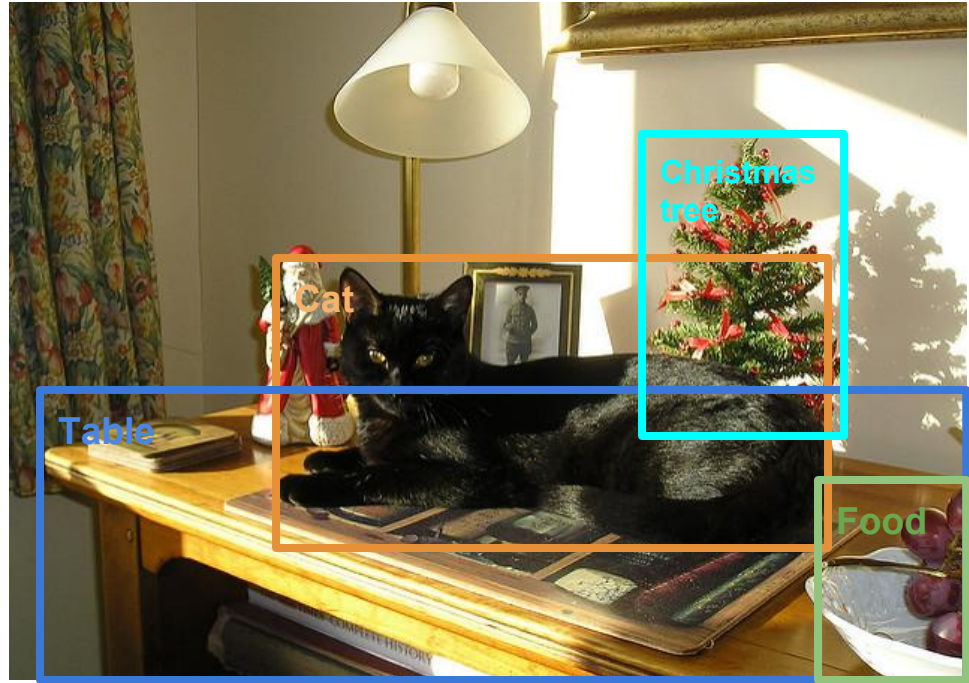
Track 1: Object Detection

Task: detect objects on an image

- object bounding box
- class label

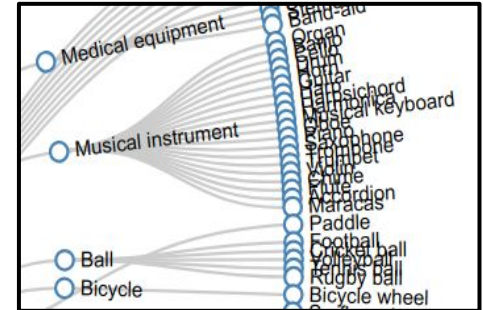
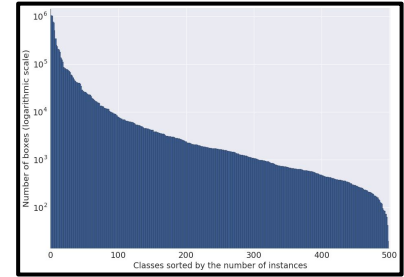
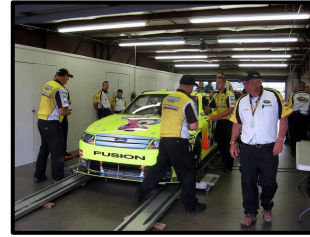
Training set:

- 12,195,144 bounding boxes
- 1,743,042 images
- 500 classes
removed some broad (e.g. "clothing") and some infrequent classes (e.g. "paper cutter")



Track 1: Object Detection - why challenging

- Challenging images
- Many classes (500)
- Imbalanced class distribution
- Class hierarchy
- Non-exhaustive annotations for classes
(but all objects annotated for those we have)



Track 2: Visual Relationships Detection

Task: detect relations between objects

- Two objects boxes and classes
- Relationship between them

Training set:

- 374,768 relationship annotations
- 3,290,070 boxes
- 1,743,042 images
- 329 distinct relation triplets
- 62 different object classes

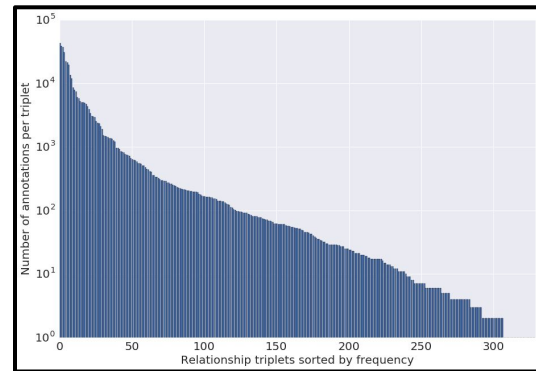
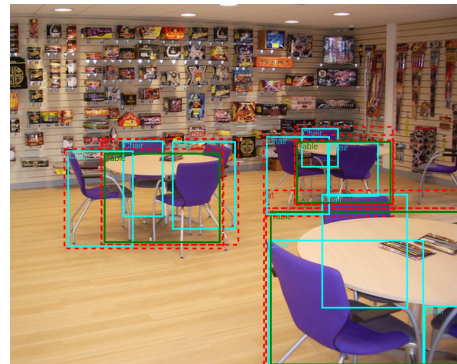
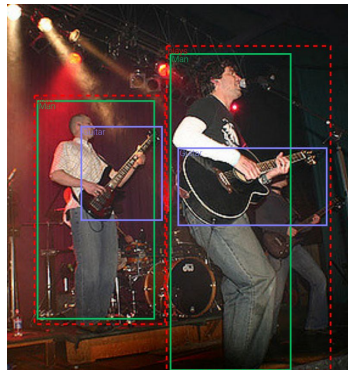


Must identify which of the two men holds the microphone!

Track 2: Visual Relationships Detection

Challenges:

- Challenging images
- Class co-occurrence not sufficient to predict the relationship
- High relationship triplets imbalance



Today's program

Time	Section
13:30 - 13:50	Overview of Open Images and the challenge
13:50 - 14:10	Object Detection track - settings, metrics, winners, analysis
14:10 - 15:00	Presentations by three selected Object Detection track participants
15:00 - 15:40	Keynote: Devi Parikh Title: "A-STAR: Towards Agents that See, Talk, Act, and Reason"
15:40 - 16:20	Break and Poster session
16:20 - 16:40	Visual Relationship Detection track - settings metrics, winners, analysis
16:40 - 17:10	Presentations by two selected Visual Relationship Detection track participants
17:10 - 17:20	Concluding remarks and plans for future of Open Images

Challenge organizers



Vittorio
Ferrari



Alina
Kuznetsova



Jordi
Pont-Tuset



Matteo
Malloci



Jasper
Uijlings



Jake
Walker



Rodrigo
Benenson

All contributors to the Open Images Dataset

Project Lead & Coordination

Vittorio Ferrari
Tom Duerig
Victor Gomes

Image collection

Ivan Krasin
David Cai

Image-level labels

Neil Alldrin
Ivan Krasin
Shahab Kamali
Tom Duerig
Zheyun Feng
Anurag Batra
Alok Gunjan

Bounding boxes

Hassan Rom
Alina Kuznetsova
Jasper Uijlings
Stefan Popov
Matteo Mallocci
Sami Abu-El-Haija
Vittorio Ferrari

Website & visualizer

Jordi Pont-Tuset

Classes & hierarchy

Chen Sun
Kevin Murphy
Tom Duerig
Vittorio Ferrari

Challenge

Vittorio Ferrari
Alina Kuznetsova
Jordi Pont-Tuset
Matteo Mallocci
Jasper Uijlings
Jake Walker
Rodrigo Benenson

Advisers

Andreas Veit
Serge Belongie
Abhinav Gupta
Dhyanesh Narayanan
Gal Chechik

Thanks to ...



Common Visual Data Foundation

For hosting the data through AWS
Special thanks to Tsung-Yi Lin



Figure Eight

For hosting the data



Kaggle

For hosting the competition

Thanks to our sponsors



Materials after this slide

Open Images Challenge 2018

<https://storage.googleapis.com/openimages/web/index.html>



Track 2: Visual Relationships Detection

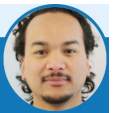
1. Select relationships: co-occurrences analysis on image-level labels level + common sense
2. Annotations: annotators verification



The next slide I usually show as a big summary of the key stats and facts

Slides used by Neil in Google Research Conference
2017 below this point

People



Teams



VALE



Project Loco

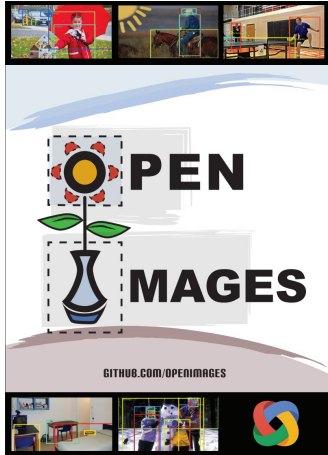


Image
Understanding



Crowdsource





- **Large scale** image dataset, as of V3:
 - 9M images, ~20M labels, ~4M boxes
- Safe for **commercial use** (CC-BY)
- Foster **industry + academic collaboration**

Vision Tasks We Want To Enable



Image Classification



Object Detection


In the future:

- Segmentation
- OCR
- Captioning
- VQA
- ...

Bicycle wheel
 Clothing
 Bicycle
 Human arm
 Human body
 Person
 Wheel
 Sports equipment
 Footwear

Image-Level Labels:

- Bicycle wheel
- Person
- Land vehicle
- Human body
- Clothing
- Footwear
- Arm
- Bike
- Bicycle frame
- Extreme sport
- Cycle sport
- Sports equipment
- Freestyle bmx
- Sports
- Vehicle
- Wheel
- Bicycle motocross



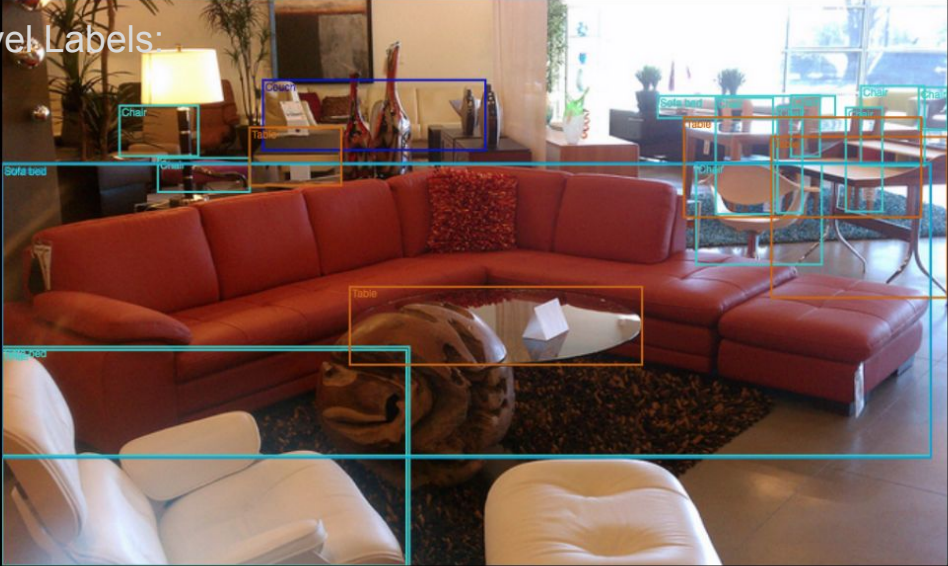
The image shows a person in a yellow shirt and dark pants performing a bicycle motocross trick on a ramp. The person is in mid-air, with the bike tilted. Bounding boxes are drawn around various parts of the person and the bike, labeled with text. The labels include: 'Person body' (yellow), 'Clothing' (teal), 'Bicycle' (brown), 'Human arm' (olive), 'Human body' (red), 'Person' (yellow), 'Wheel' (brown), 'Sports equipment' (magenta), and 'Footwear' (yellow). The background is a dark, indoor setting with a ramp.

Author: [Tony Hisgett \(flickr\)](https://www.flickr.com/photos/hisgett/6390303655)
 Photo URL: <https://www.flickr.com/photos/hisgett/6390303655>
 License: [CC BY 2.0](https://creativecommons.org/licenses/by/2.0/)

Chair
Sofa bed
Loveseat
Couch
Table

Image-Level Labels:

- Sofa bed
- Loveseat
- Chair
- Lobby
- Couch
- Interior design
- Living room
- Suite
- Table
- Property
- Room
- Arab cuisine
- Furniture



Author [Scott Allen \(flickr\)](#)

Photo URL <https://www.flickr.com/photos/scottandjayne/6812100728>

License [CC BY 2.0](#)

Vegetable Tart Knife

Image-Level Labels:

Knife
Weapon
Vegetable
Quiche
Cuisine
Dessert
Dish
Tart
Baked goods
Baking
Sweet potato pie
Breakfast
Pie



Author [Tom Purves \(flickr\)](#)

Photo URL <https://www.flickr.com/photos/thomaspurves/265753037>

License [CC BY 2.0](#)

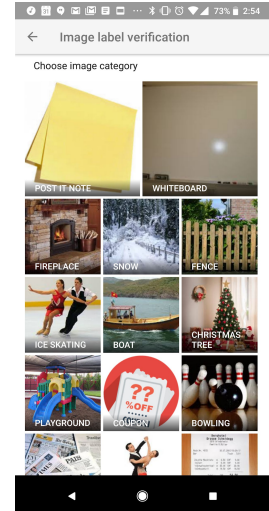
Dataset Motivation and Origins

Problem: Other datasets are **copy-encumbered**

Solution: restrict to images with CC_BY license

Enables Crowdsourcing

- Google Village Crowdsourcing App
- ReCaptcha

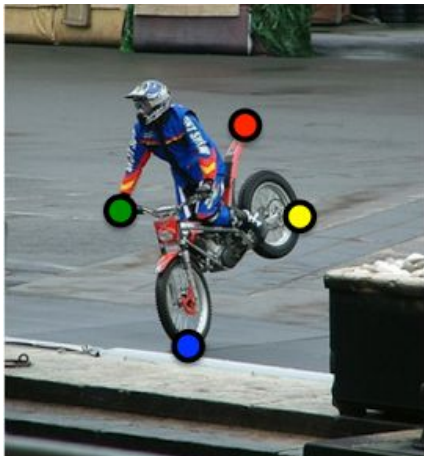


Dataset Motivation and Origins

Problem: **Not enough boxes** in existing datasets

Solution: Concerted box collection effort (LOCO team)

MOST BOXES IN ANY PUBLIC DATASET!



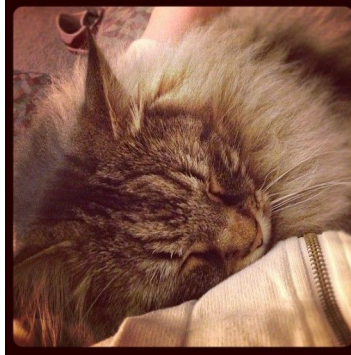
Dataset Motivation and Origins

Problem: **Web bias**



ne web

Typical "Cat" in Image Search.



"Cat" Images in OLD dataset.



Image-Level Annotations

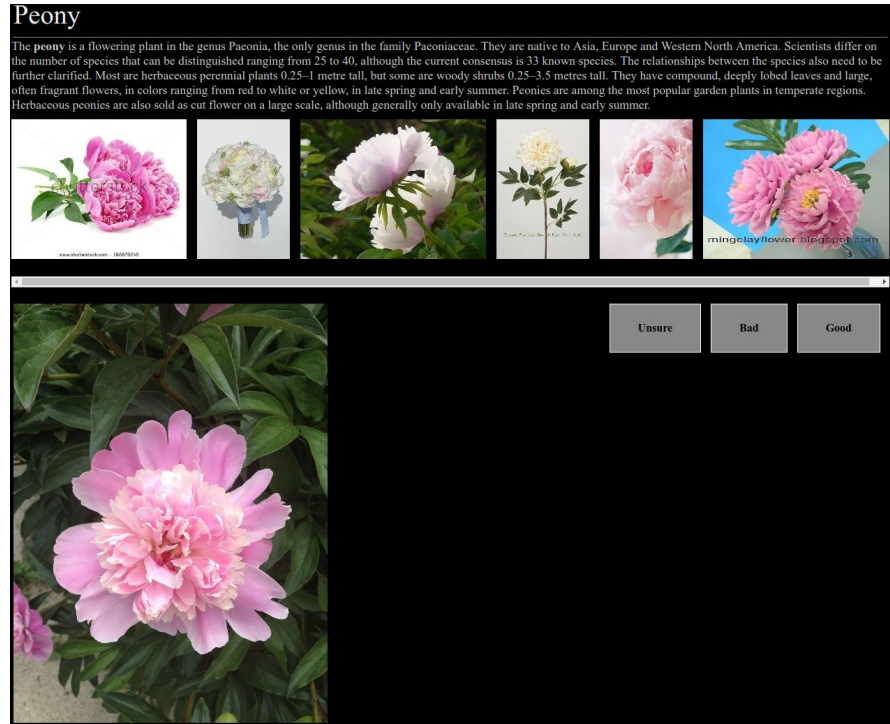
1. **Machine Generated Predictions:** ICA ENet model
2. **Paid Raters:** Crowd-compute
3. **Crowdsourcing:** Crowdsourcing App, ReCaptcha



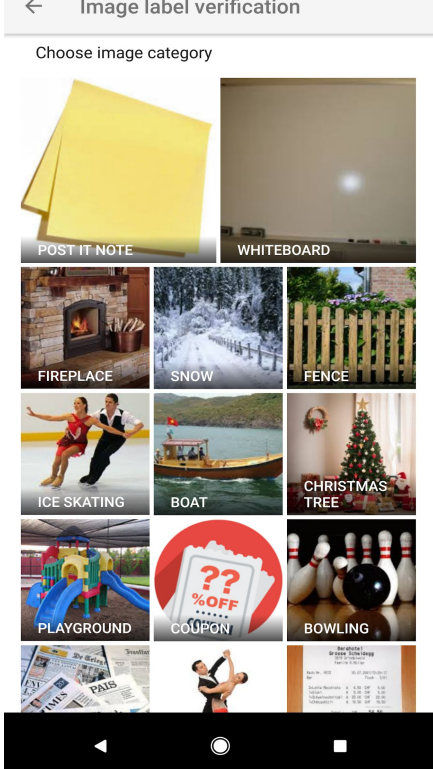
Winter
Snow
Sled
Sledding
Headgear

Crowd Compute "Verify Entity" Task

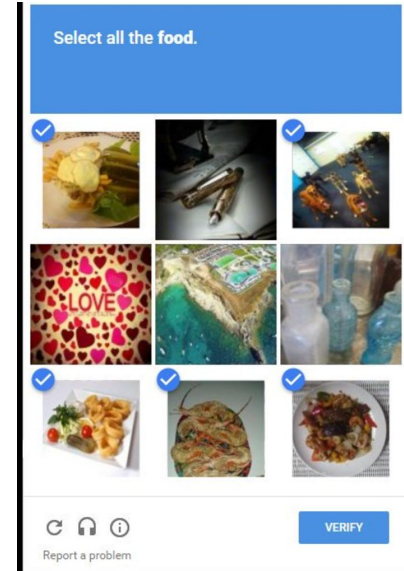
- Verification of ICA classifier output
- >20k unique labels
- Expert raters for fine-grained



Crowd Sourced Labels





Crowdsource App



Recaptcha

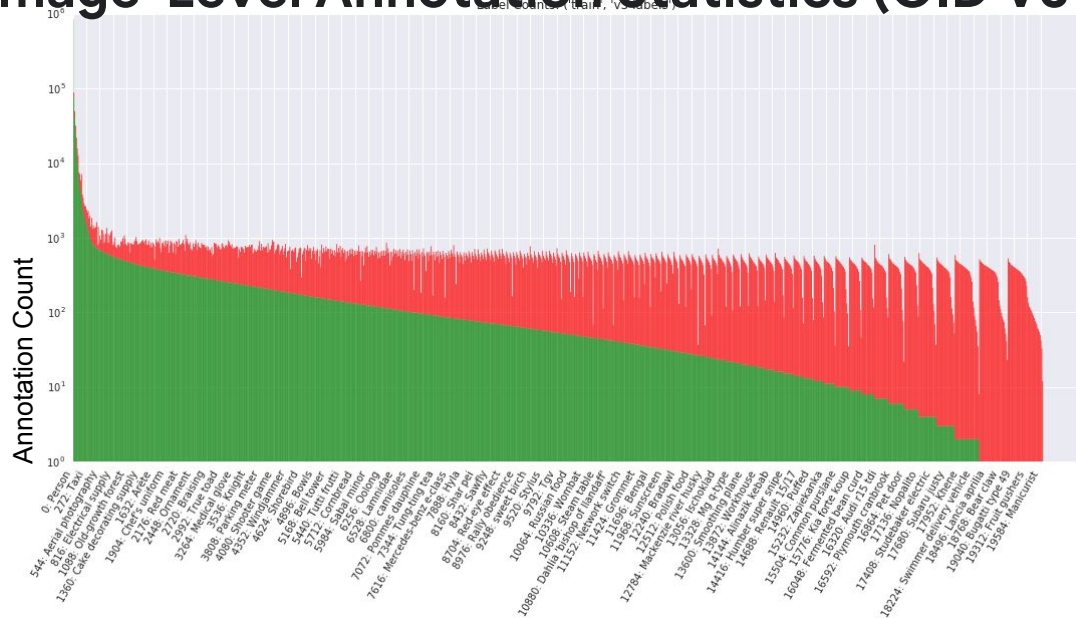
Image-Level Annotation Statistics (OID V3)

	Train	Test	Validation	# Classes	# Trainable Classes
Images	9M	125K	42K	-	-
Machine Generated Labels	79M	1.5M	512K	8K	5K
Human Verified Labels	20M 	1.6M	551K	20K 	5K

Compare to other datasets

- **Imagenet:** 14M labels, 22K classes

Image-Level Annotation Statistics (OID V3 Train Set)

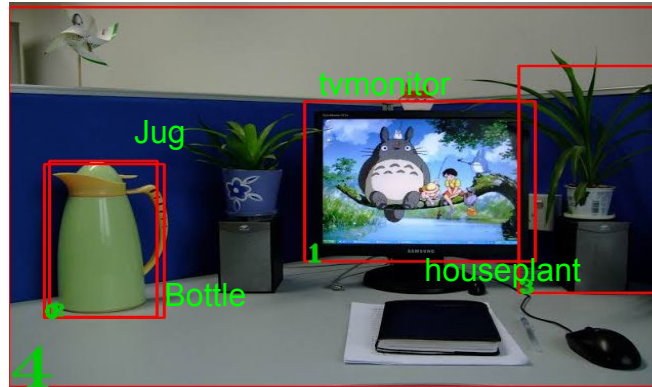


Negative Verifications

Positive Verifications

Boxes

1. **Freeform** box drawing
2. **XClick**: Faster box drawing method
3. **ActiveMIL**



Extreme Clicking (XClick)



Idea: Click **physical points** on object

- E.g., leftmost point on the motorcycle

Same quality as drawing
7-10s (vs 30s traditional)

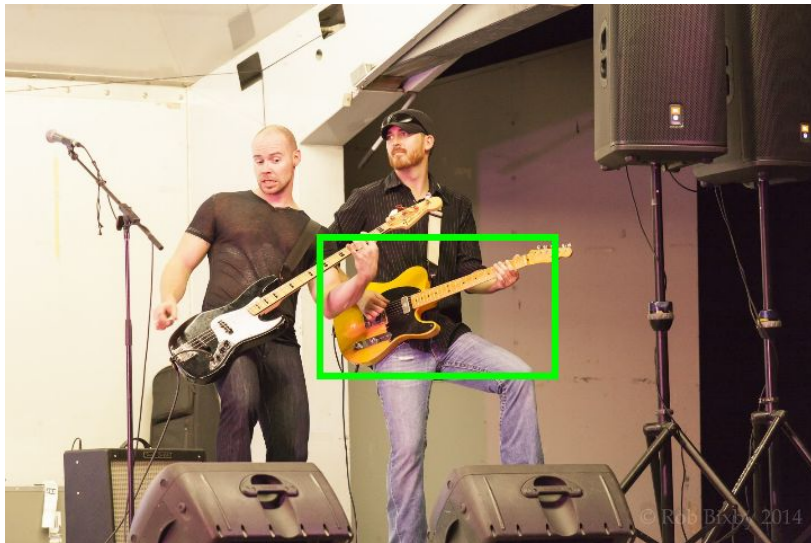
ActiveMIL Verification Sequence



ActiveMIL Verification Sequence



ActiveMIL Verification Sequence



Box Statistics (OID V3)

	Train	Test	Validation	# Classes	# Trainable Classes
Images	1.6M	108K	36K	-	-
Boxes	3.7M	625K	205K	600	545

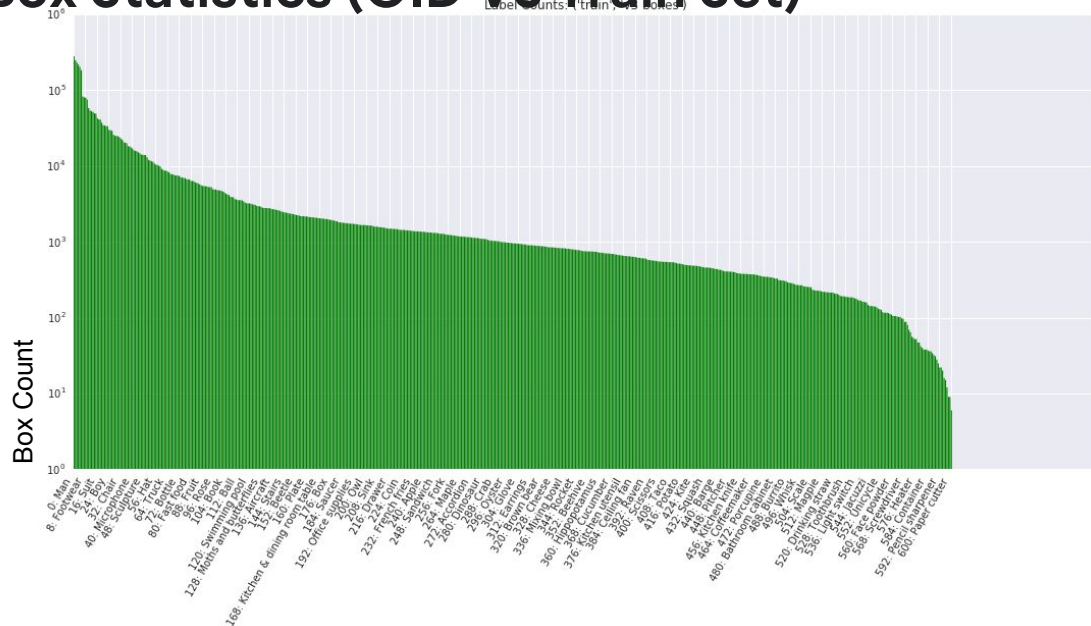
MOST BOXES IN ANY PUBLIC DATASET!

Compare to other datasets

- **Imagenet**: 350K* boxes, 200 classes
- **COCO**: 600K boxes, 80 classes

*Up to 1M boxes if one really digs

Box Statistics (OID V3 Train Set)



Pre-trained models

- Image classifiers
 - Available on externally on github
- Object detection models
 - Available externally via Object Detection API

Next Steps

- **More boxes**
 - XClick, internally ~9M almost ready
- **More labels**
 - Crowdsourcing
- **Other types of groundtruth**
 - Segmentation
 - OCR
 - ...
- **Challenge?** - We could use 20%ers!



Thank you!



go/openimages
github.com/openimages

Comparison to Other Datasets

	Images	Categories	Image Annotations	Boxes
Open Images V3	9M	20K trainable: 5K	20M	3.7M
ImageNet	14M	22K	14M	*350K
COCO	330K	objects: 80 stuff: 91	N/A	600K
PASCAL VOC 2012	11.5K	20	N/A	27.5K

* Up to 1M boxes if one really digs

COCO: cocodataset.org, ImageNet: image-net.org, PASCAL VOC 2012: host.robots.ox.ac.uk/pascal/VOC/voc2012