Object Detection

Danna Gurari
The University of Texas at Austin
Fall 2019

https://www.ischool.utexas.edu/~dannag/Courses/CrowdsourcingForCV/CourseContent.html

Review

- Last week
  - Segmentation definition and applications
  - Segmentation evaluation
  - Crowdsourcing segmentations

- Assignments (Class Website & Canvas)
  - Reading assignment 5 due yesterday
  - Lab assignment 2 due in two weeks
  - Project pre-proposal due in two weeks

- Questions?

Today’s Topics

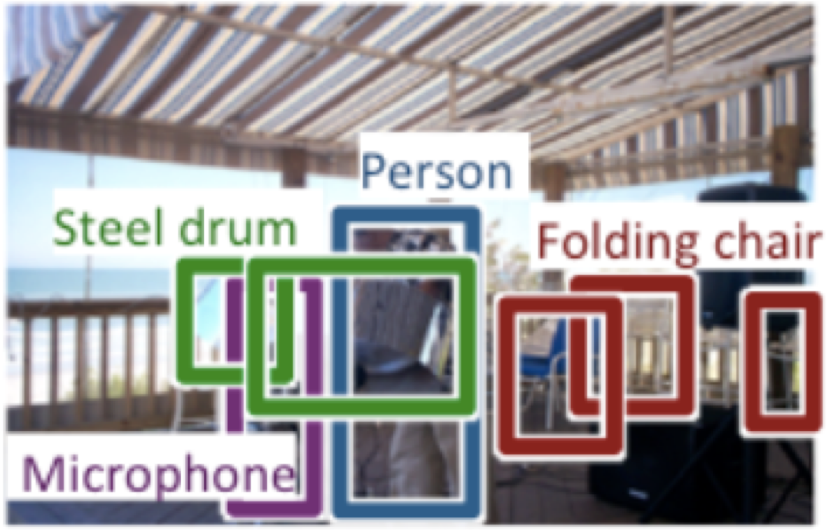
- Object detection applications
- Object detection evaluation
- Crowdsourcing object detection
- Class discussion (chosen by YOU ☺)
- Lab: drawing on images

Today’s Topics

- Object detection applications
- Object detection evaluation
- Crowdsourcing object detection
- Class discussion (chosen by YOU ☺)
- Lab: drawing on images

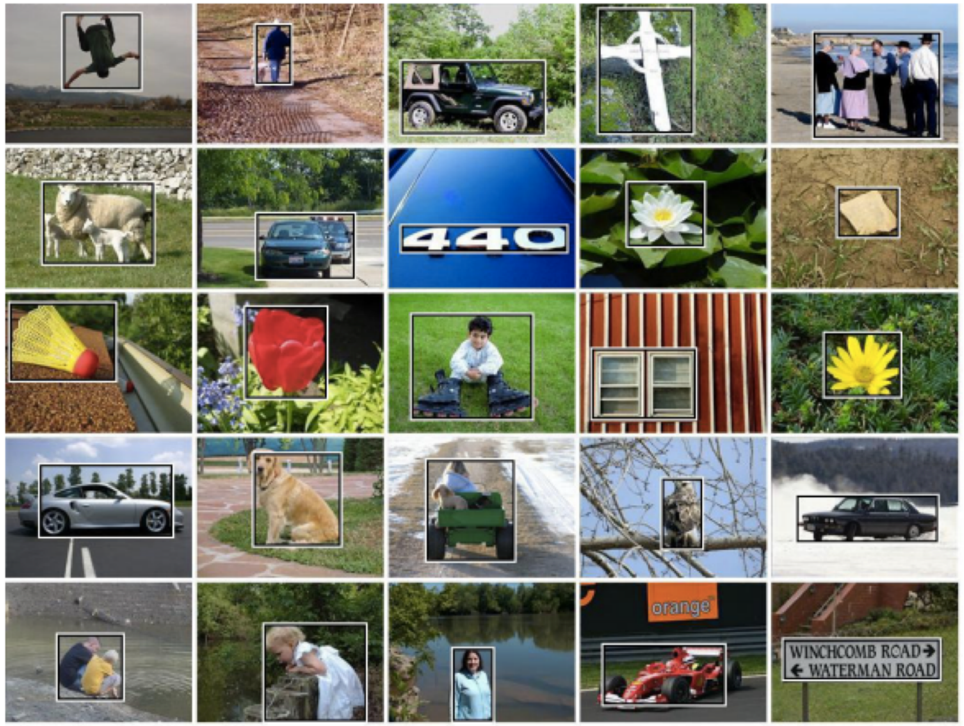
Definition

- Object detection: localize every instance of an object
  - Semantic: pre-specified categories (e.g., cat, dog, person, …)



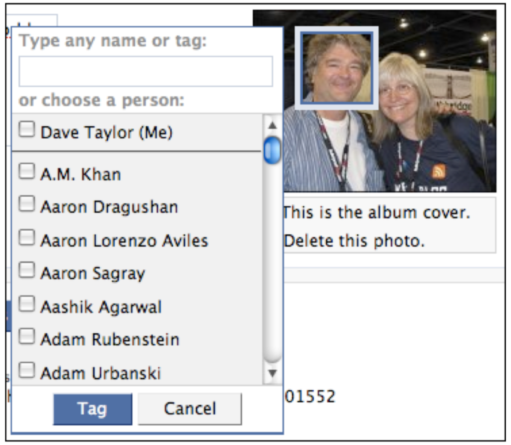
[Russakovsky et al; IJCV 2015]

  - Salient: any category

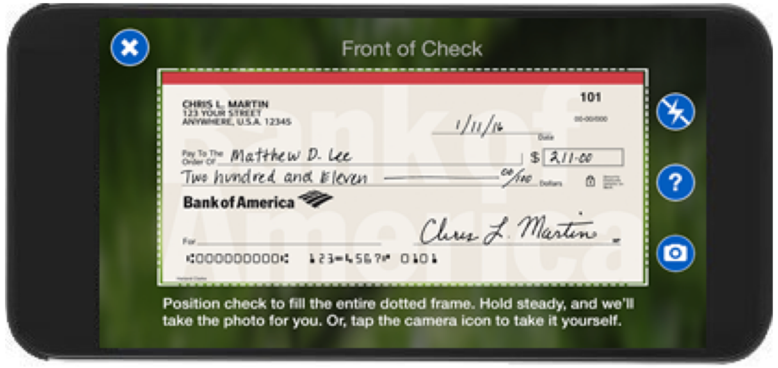


[Liu et al; CVPR 2007]

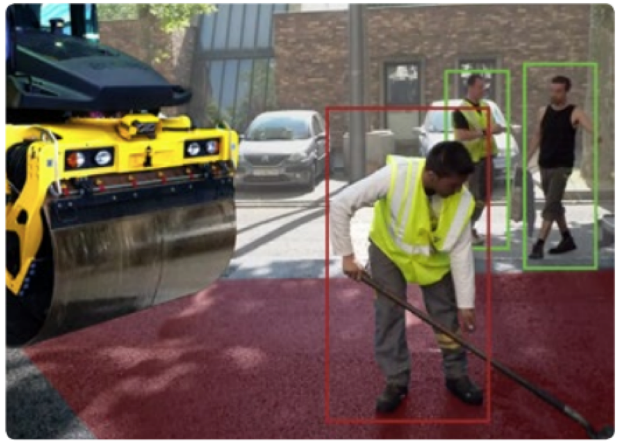
Applications: Detection



Face detection
(e.g., Facebook)



Mobile check deposit
(e.g., Bank of America)



Pedestrian Detection
(e.g., Blaxtair)
http://media.brintex.com/Occurrence/121/Brochure/3435/brochure.pdf

License Plate Detection (e.g., AllGoVision)

Applications: Counting

Counting Fish (e.g., SalmonSoft)
http://www.wecountfish.com/?page_id=143

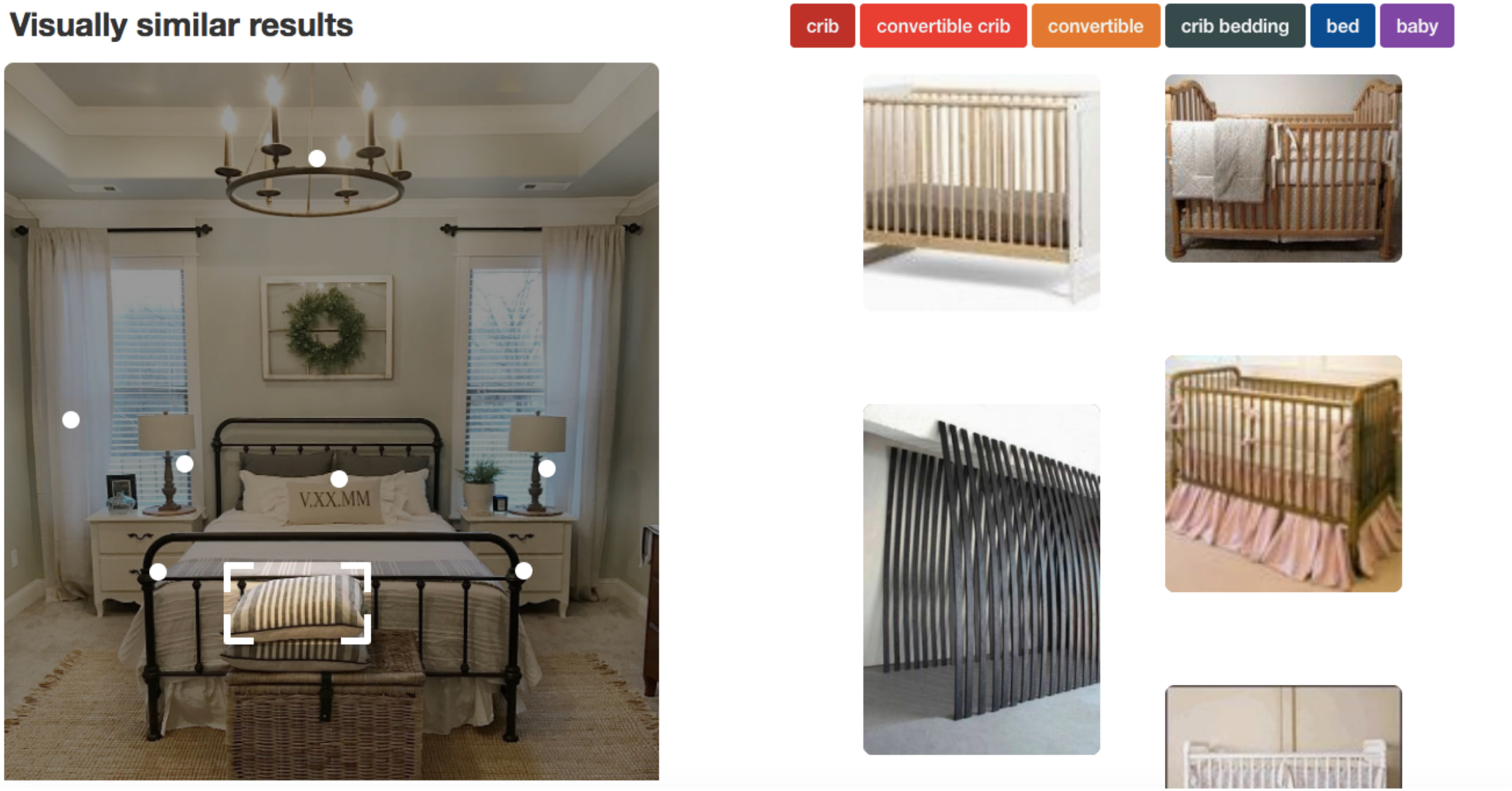
Business Traffic Analytics

Applications: Anomaly Detection

Surveillance

http://www.intel.com/content/www/us/en/embedded/digital-security-surveillance/dss-left-object-detection-demo.html

Applications: Search



Search for specific items (e.g., Pinterest)

Object Detection versus Recognition

"What is the difference between object recognition and object detection?"

- Learn appearance of object rather than its image context; e.g., giraffe

Object Detection versus Segmentation

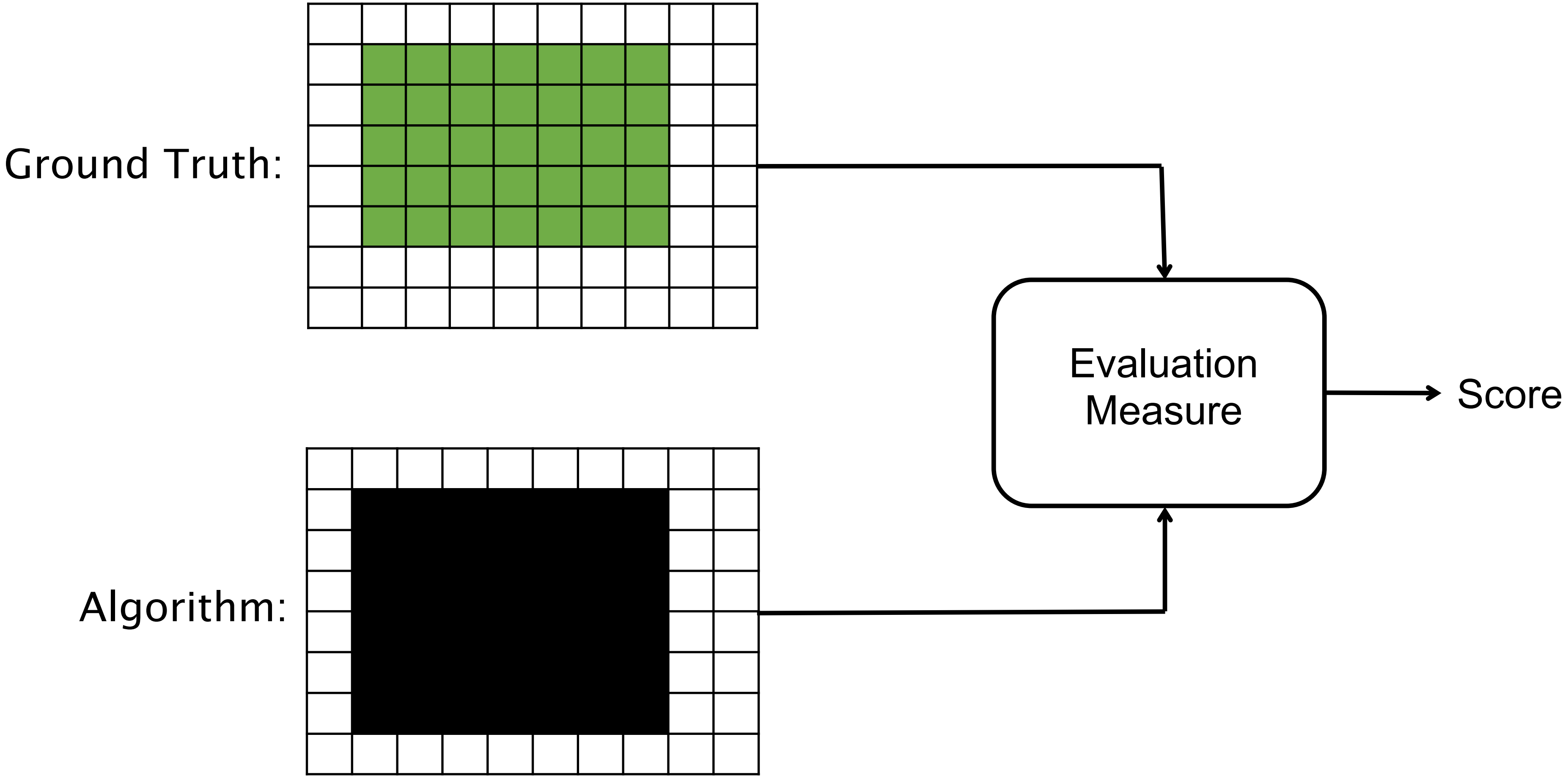
- Why choose object “segmentation” over “detection”?

http://mmcheng.net/msra10k/

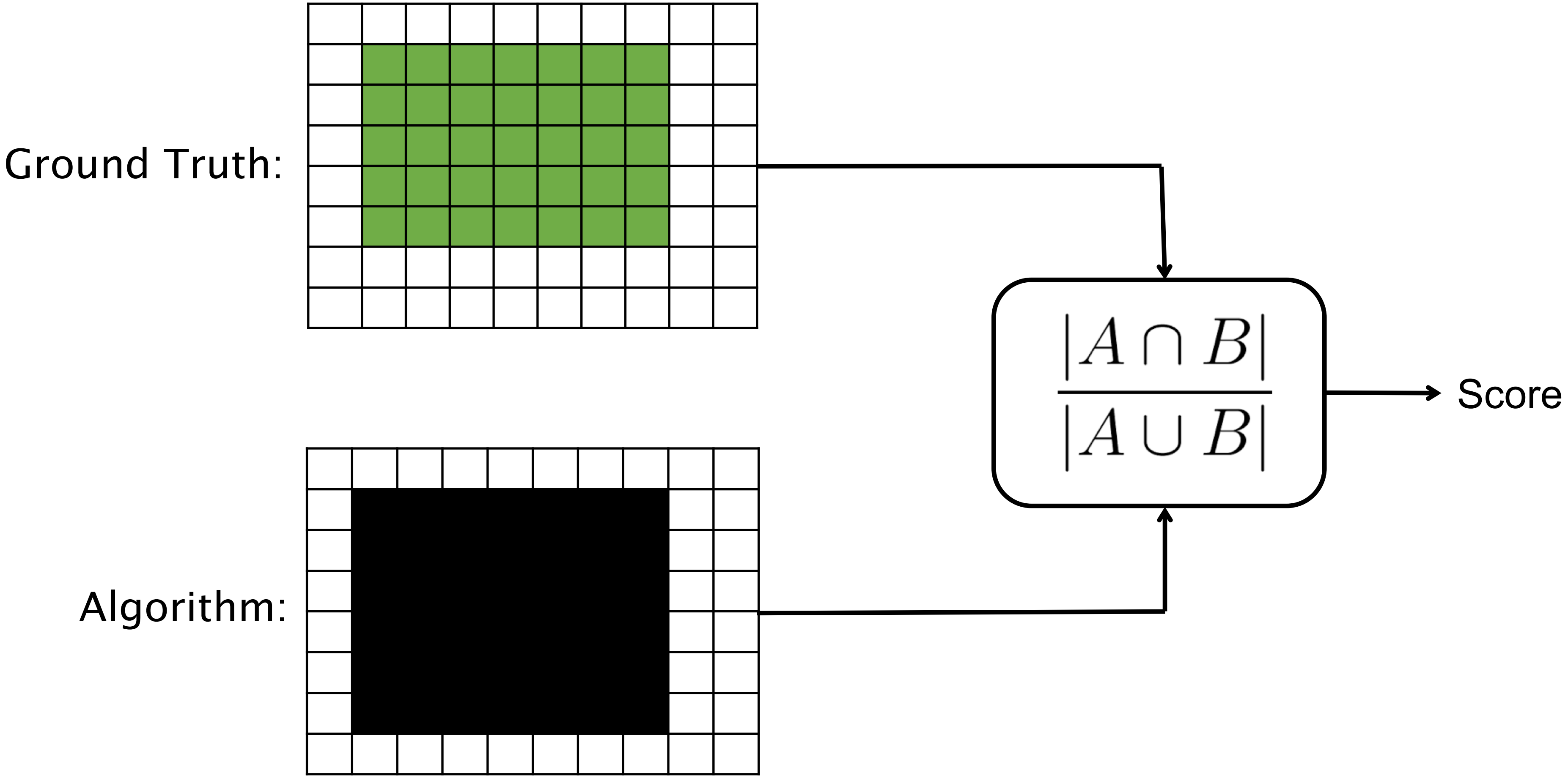
Today’s Topics

- Object detection applications
- Object detection evaluation
- Crowdsourcing object detection
- Class discussion (chosen by YOU ☺)
- Lab: drawing on images

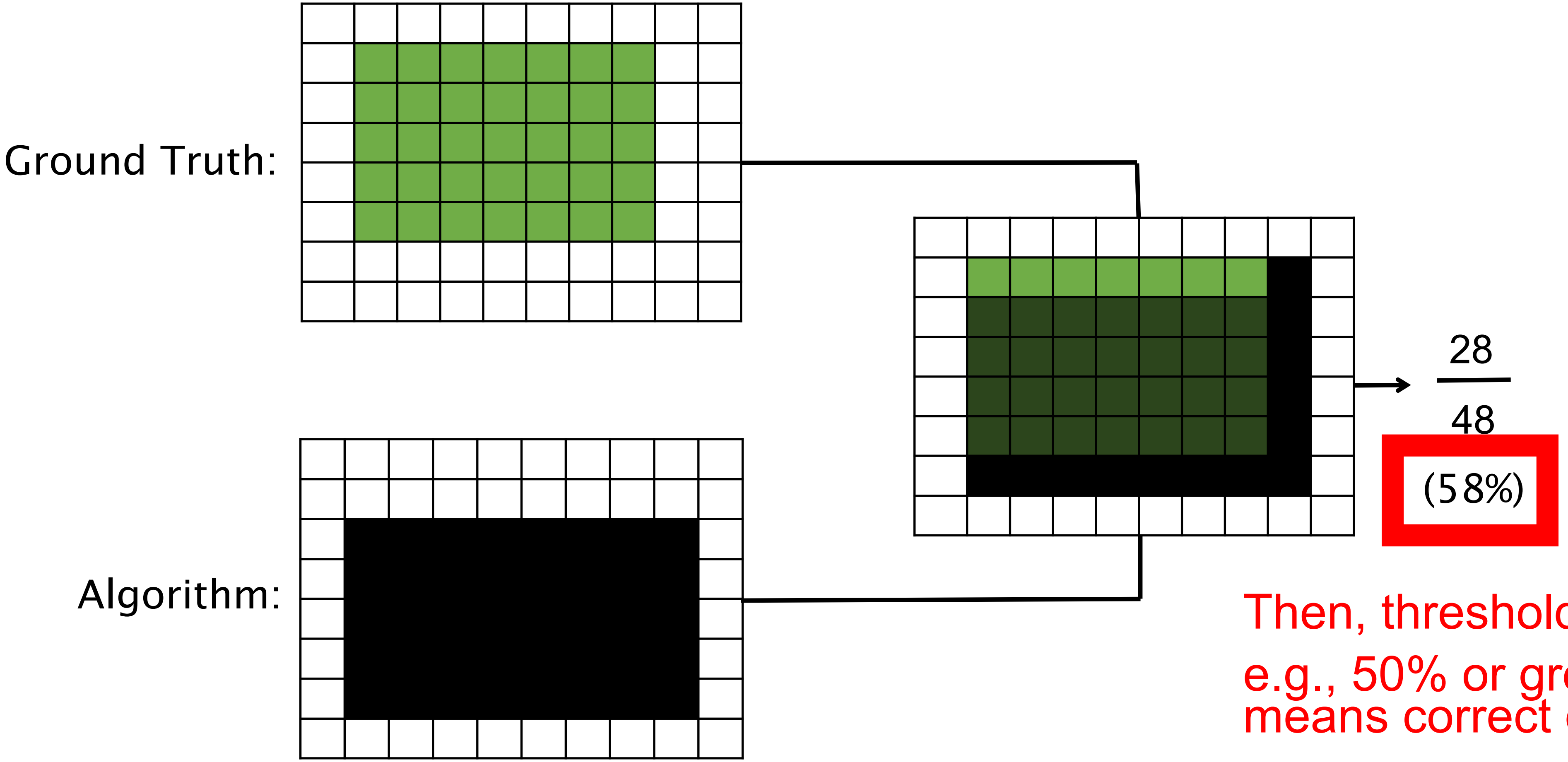
Object Detection – Single Object



Object Detection – Single Object



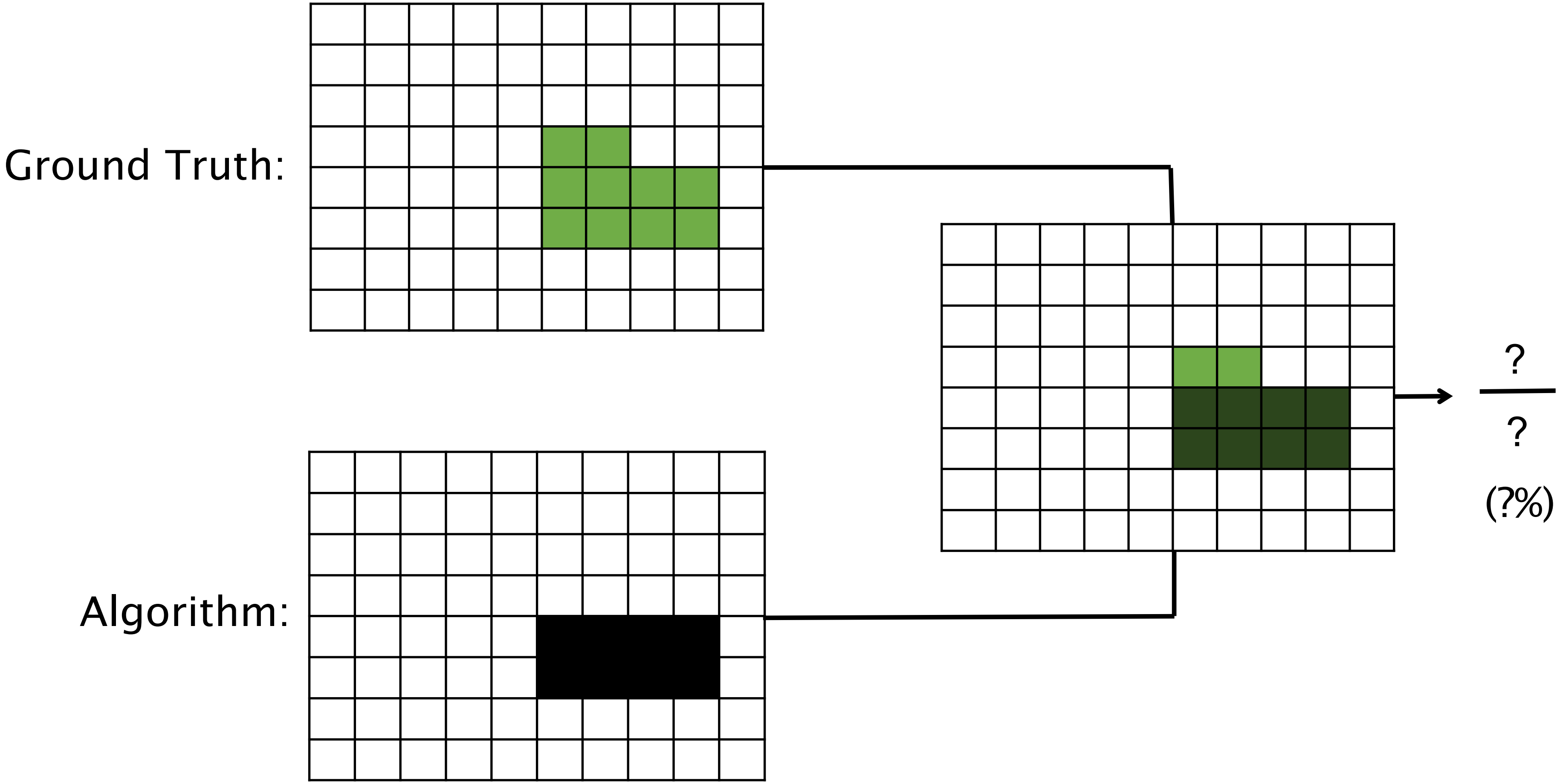
Object Detection – Single Object



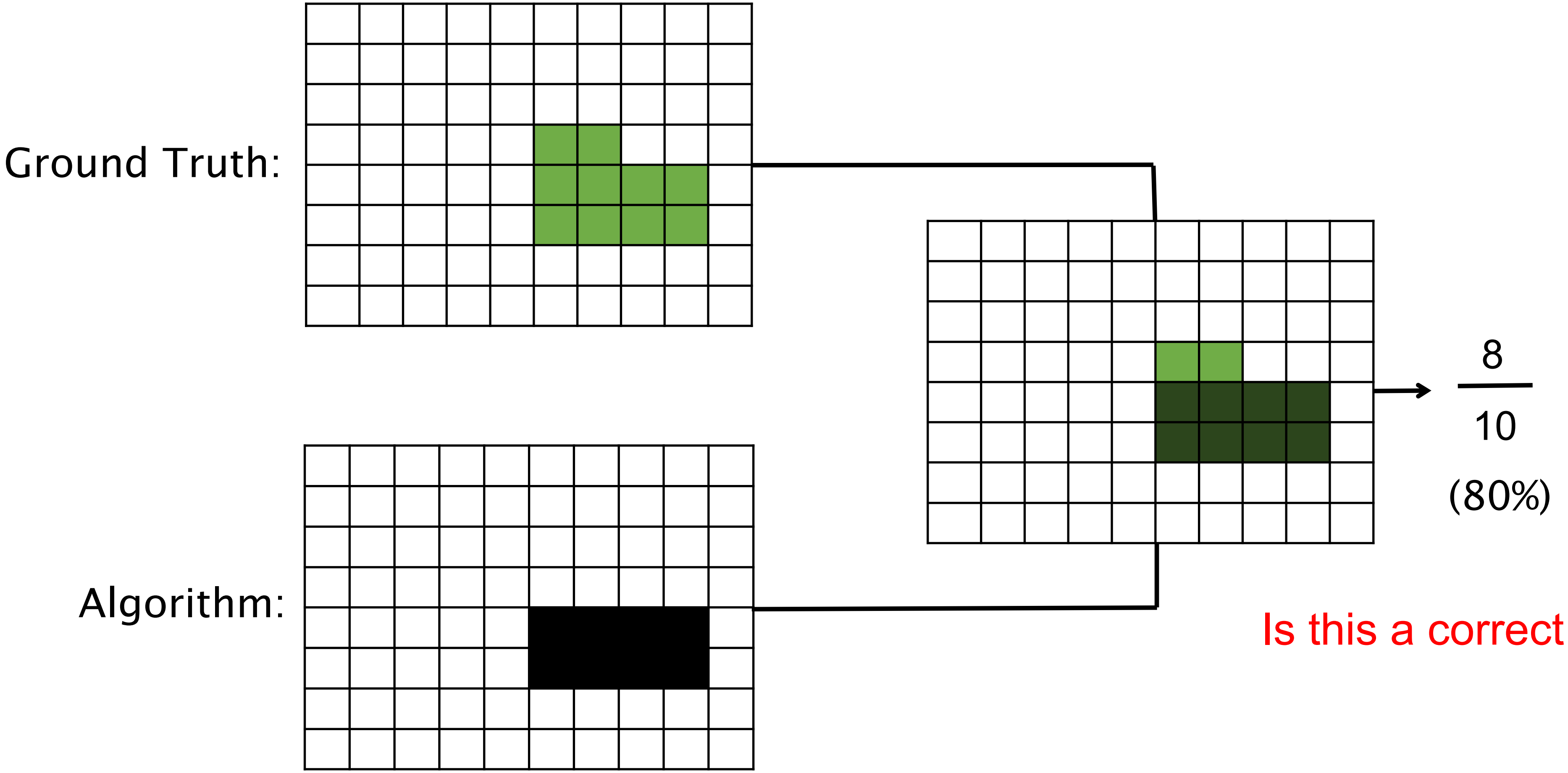
Then, threshold:

e.g., 50% or greater means correct detection!

Object Detection – Single Object



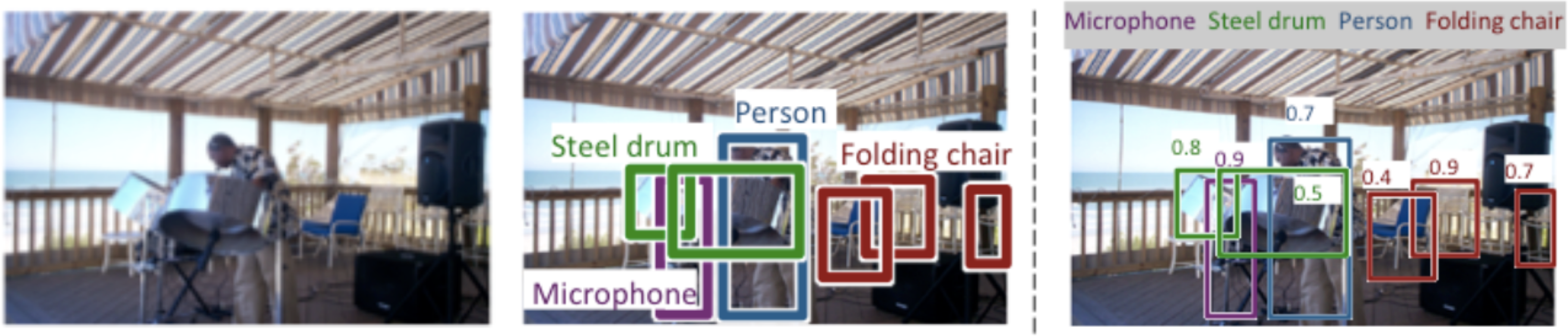
Object Detection – Single Object



Is this a correct detection?

Object Detection – Multiple Objects

- For each object class (e.g., cat, dog, …), compute:
  - Precision: fraction of correct detections out of all algorithm detections



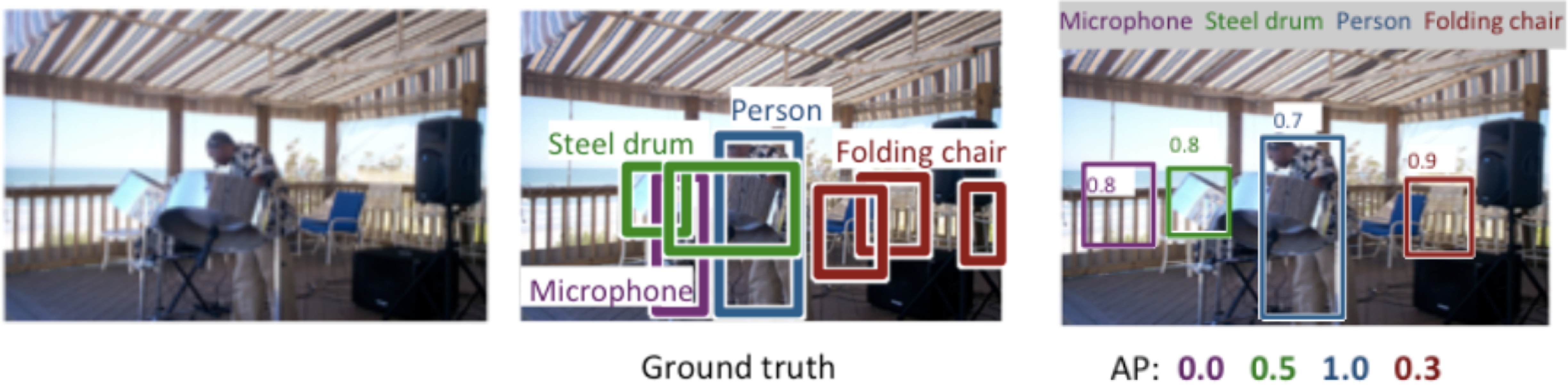
Ground truth

Algorithm BB + its Confidence

[Russakovsky et al; IJCV 2015]

Object Detection – Multiple Objects

- For each object class (e.g., cat, dog, …), compute:
  - Precision: fraction of correct detections out of all algorithm detections

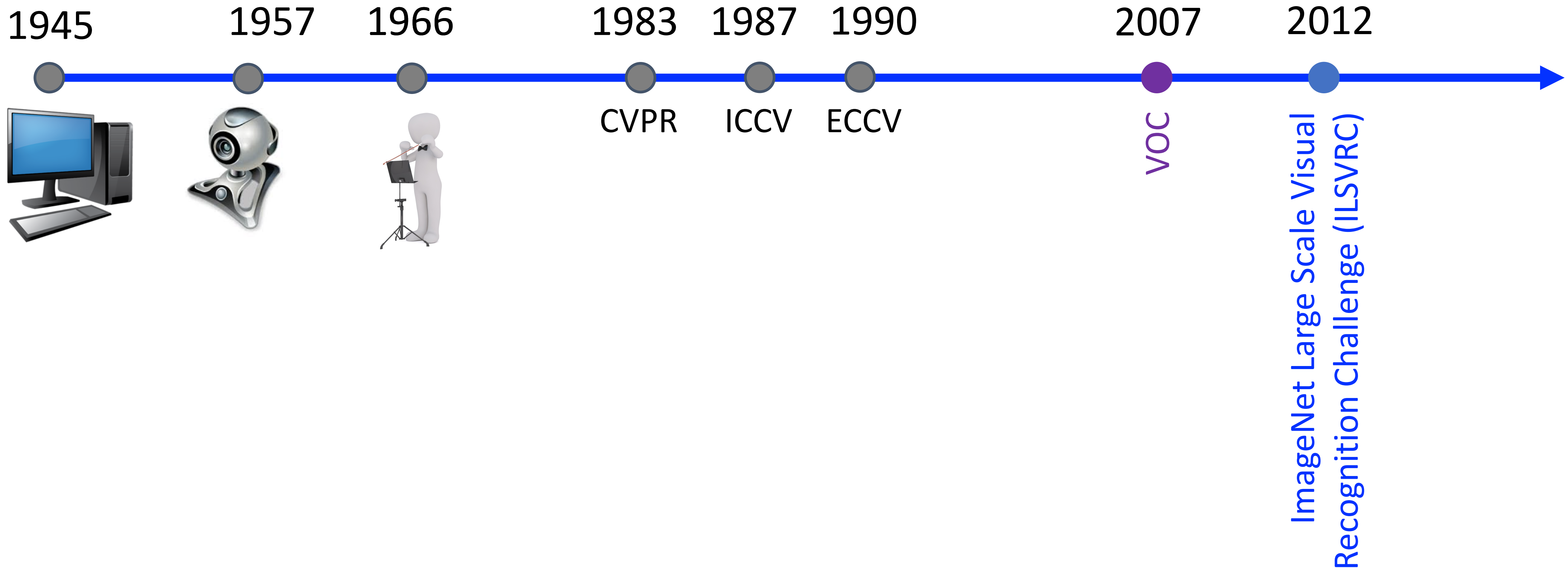


[Russakovsky et al; IJCV 2015]

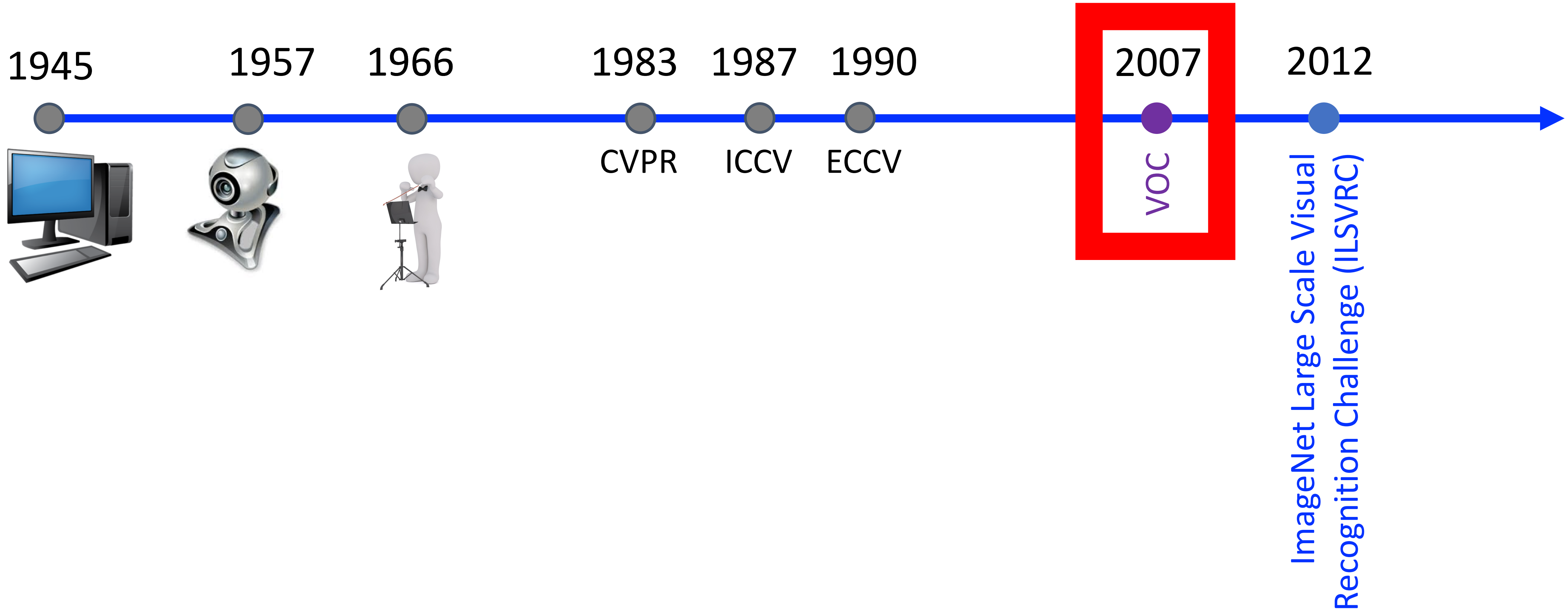
Today’s Topics

- Object detection applications
- Object detection evaluation
- Crowdsourcing object detection
- Class discussion (chosen by YOU ☺)
- Lab: drawing on images

Object Detection Datasets



Object Detection Datasets



Object Detection Datasets: VOC

1. Category Selection

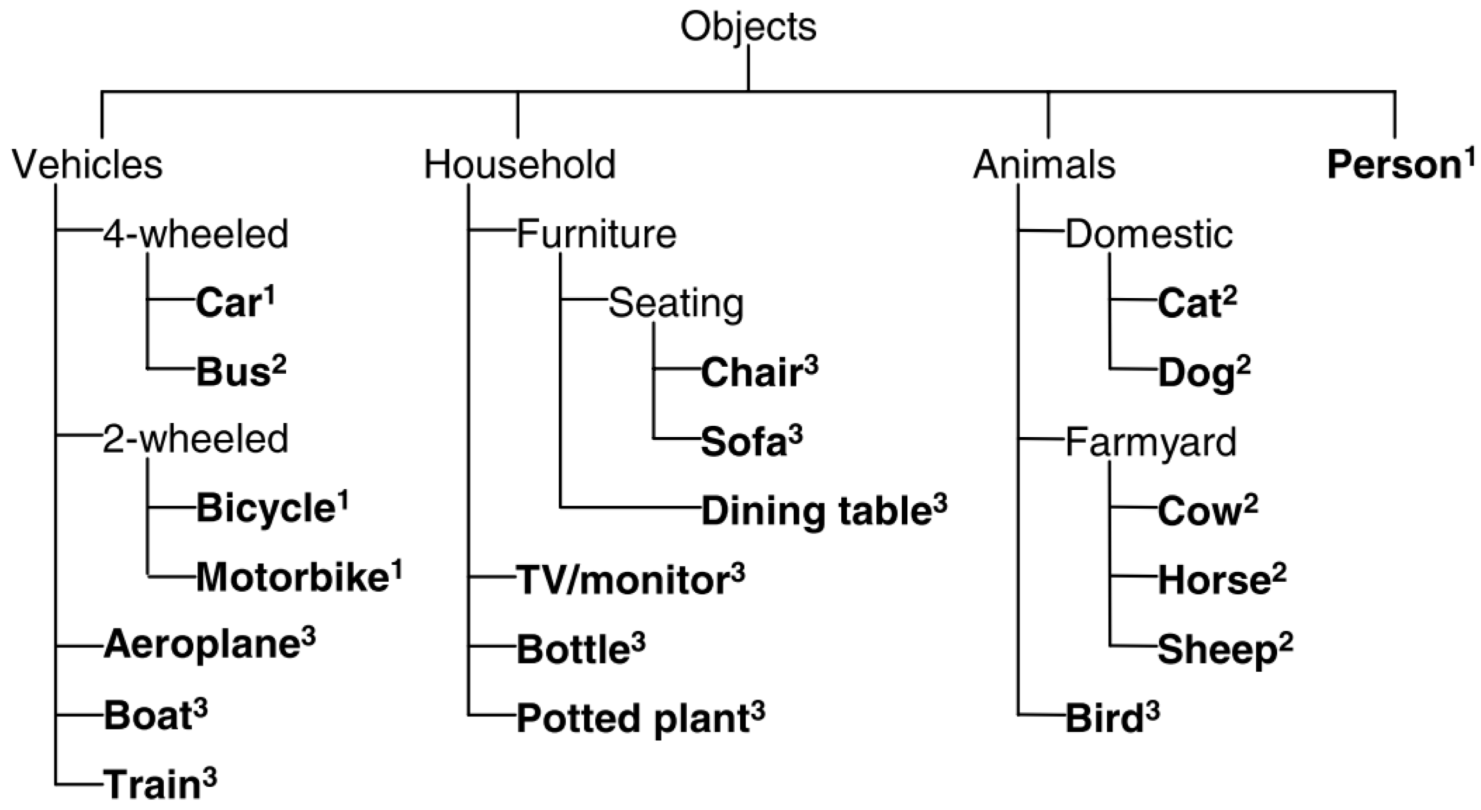
- 20 categories chosen:

1) Initial 4 categories stem from existing dataset

2) 2006: added 6 classes

3) 2007: added 10 classes

- Additional categories provide a broader domain and finer-grained categories, including visually similar things



(superscript indicates year of inclusion in the challenge: 2005[1], 2006[2], 2007[3])

Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman. The PASCAL Visual Object Classes (VOC) Challenge. IJCV 2010.

Object Detection Datasets: VOC

1. Category Selection

- 20 categories chosen:

1) Initial 4 categories stem from existing dataset

2) 2006: added 6 classes

3) 2007: added 10 classes

- Additional categories provide a broader domain and finer-grained categories, including visually similar things

→

2. Image Collection

- 500,000 images retrieved from Flickr by querying with a number of keywords

(many query terms per category)

- **aeroplane**, airplane, plane, biplane, monoplane, aviator, bomber, hydroplane, airliner, aircraft, fighter, airport, hangar, jet, boeing, fuselage, wing, propellor, flying
- **bicycle**, bike, cycle, cyclist, pedal, tandem, saddle, wheel, cycling, ride, wheelie
- **bird**, birdie, birdwatching, nest, sea, aviary, birdcage, bird feeder, bird table
- **boat** ship, barge, ferry, canoe, boating, craft, liner, cruise, sailing, rowing, watercraft, regatta, racing, marina, beach, water, canal, river, stream, lake, yacht
- **bottle**, cork, wine, beer, champagne, ketchup, squash, soda, coke, lemonade, dinner, lunch, breakfast
- **bus**, omnibus, coach, shuttle, jitney, double-decker, motorbus, school bus, depot, terminal, station, terminus, passenger, route
- **car**, automobile, cruiser, motorcar, vehicle, hatchback, saloon, convertible, limousine, motor, race, traffic, trip, rally, city, street, road, lane, village, town, centre, shopping, downtown, suburban
- **cat**, feline, pussy, mew, kitten, tabby, tortoiseshell, ginger, stray
- **chair**, seat, rocker, rocking, deck, swivel, camp, chaise, office, studio, armchair, recliner, sitting, lounge, living room, sitting room
- **cow**, beef, heifer, moo, dairy, milk, milking, farm
- **dog**, hound, bark, kennel, heel, bitch, canine, puppy, hunter, collar, leash
- **horse**, gallop, jump, buck, equine, foal, cavalry, saddle, canter, buggy, mare, neigh, dressage, trial, racehorse, steeplechase, thoroughbred, cart, equestrian, paddock, stable, farrier
- **motorbike**, motorcycle, minibike, moped, dirt, pillion, biker, trials, motorcycling, motorcyclist, engine, motocross, scramble, sidecar, scooter, trail
- **person**, people, family, father, mother, brother, sister, aunt, uncle, grandmother, grandma, grandfather, grandpa, grandson, granddaughter, niece, nephew, cousin
- **sheep**, ram, fold, fleece, shear, baa, bleat, lamb, ewe, wool, flock
- **sofa**, chesterfield, settee, divan, couch, bolster
- **table**, dining, cafe, restaurant, kitchen, banquet, party, meal
- **potted plant**, pot plant, plant, patio, windowsill, window sill, yard, greenhouse, glass house, basket, cutting, pot, cooking, grow
- **train**, express, locomotive, freight, commuter, platform, subway, underground, steam, railway, railroad, rail, tube, underground, track, carriage, coach, metro, sleeper, railcar, buffet, cabin, level crossing
- **tv/monitor**, television, plasma, flatscreen, flat screen, lcd, crt, watching, dvd, desktop, computer, computer monitor, PC, console, game

Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman. The PASCAL Visual Object Classes (VOC) Challenge. IJCV 2010.

Object Detection Datasets: VOC

1. Category Selection

- 20 categories chosen:

1) Initial 4 categories stem from existing dataset

2) 2006: added 6 classes

3) 2007: added 10 classes

- Additional categories provide a broader domain and finer-grained categories, including visually similar things

→

2. Image Collection

- 500,000 images retrieved from Flickr by querying with a number of keywords

→

3. Image Verification + Image Annotation

- University of Leeds annotation party to recruit annotators

- Annotation guidelines & real-time assistance

- Review of every annotation

- Annotate only “minority” classes at end of party to increase the count of them

Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman. The PASCAL Visual Object Classes (VOC) Challenge. IJCV 2010.

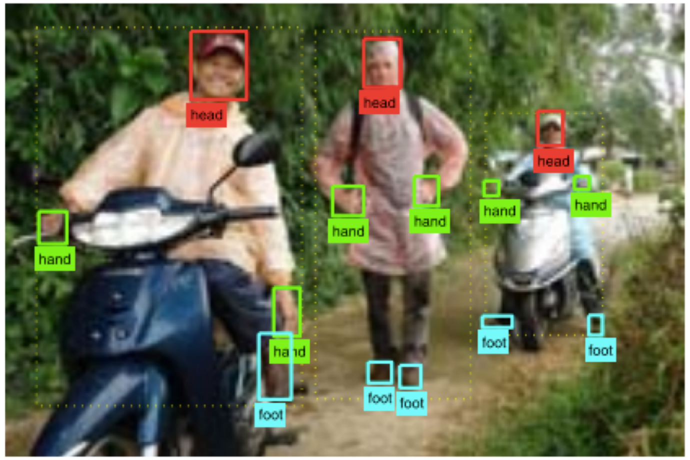
VOC Guidelines:

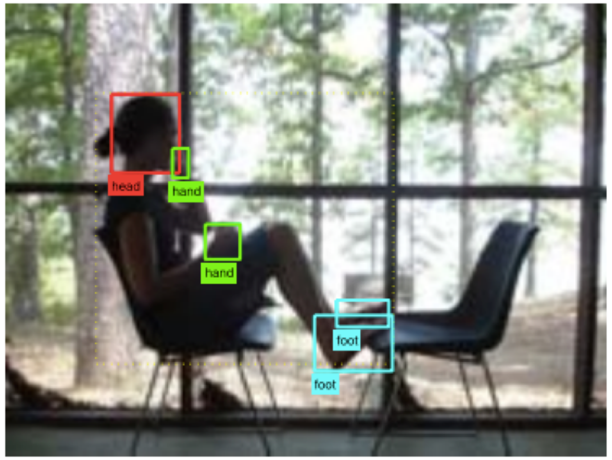
What to label	*All objects of the defined categories*, unless: • you are unsure what the object is. • the object is very small (at your discretion). • less than 10-20% of the object is visible, *such that you cannot be sure what class it is*. e.g. if only a tyre is visible it may belong to car or truck so cannot be labelled car, but feet/faces can only belong to a person. If this is not possible because too many objects, mark image as bad.
Viewpoint	Record the viewpoint of the 'bulk' of the object e.g. the body rather than the head. Allow viewpoints within 10-20 degrees. If ambiguous, leave as 'Unspecified'. Unusually rotated objects e.g. upside-down people should be left as 'Unspecified'.
Bounding box	Mark the bounding box of the visible area of the object (*not* the estimated total extent of the object). Bounding box should contain all visible pixels, except where the bounding box would have to be made excessively large to include a few additional pixels (<5%) e.g. a car aerial.
Truncation	If more than 15-20% of the object lies outside the bounding box mark as Truncated. The flag indicates that the bounding box does not cover the total extent of the object.
Occlusion	If more than 5% of the object is occluded within the bounding box, mark as Occluded. The flag indicates that the object is not totally visible within the bounding box.
Image quality/ illumination	Images which are poor quality (e.g. excessive motion blur) should be marked bad. However, poor illumination (e.g. objects in silhouette) should not count as poor quality unless objects cannot be recognised. Images made up of multiple images (e.g. collages) should be marked bad.
Clothing/mud/ snow etc.	If an object is 'occluded' by a close-fitting occluder e.g. clothing, mud, snow etc., then the occluder should be treated as part of the object.
Transparency	Do label objects visible through glass, but treat reflections on the glass as occlusion.
Mirrors	Do label objects in mirrors.
Pictures	Label objects in pictures/posters/signs only if they are photorealistic but not if cartoons, symbols etc.

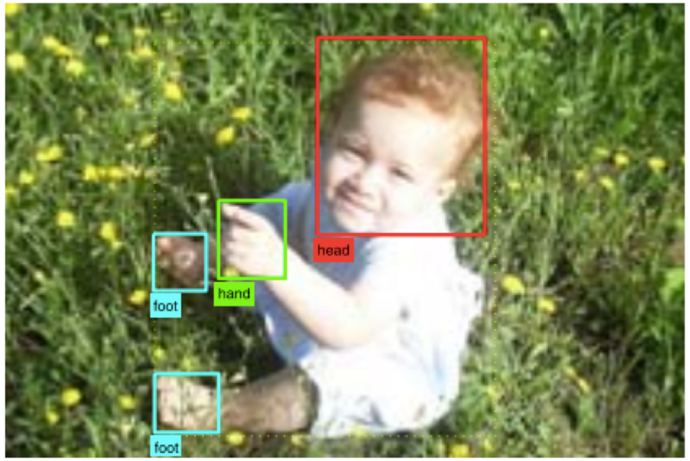
VOC Parts Detection Guidelines:

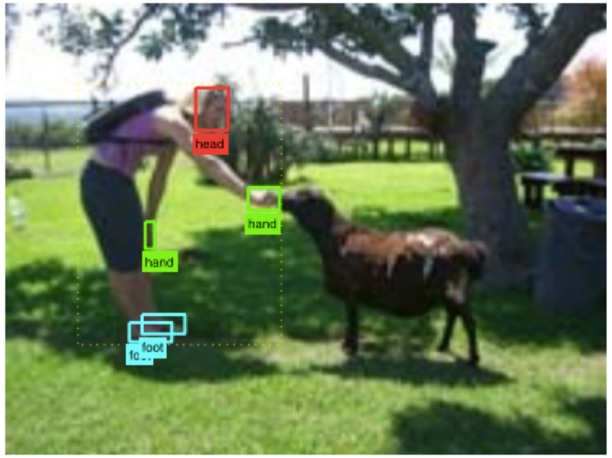
* For each person, annotate 3 types of "parts": head, hands, and feet

* Only annotate images of sufficient size where (1) there is no uncertainty in the position of the parts (2) the head and at least one other part are visible









Object Detection Datasets: VOC

1. Category Selection

- 20 categories chosen:

1) Initial 4 categories stem from existing dataset

2) 2006: added 6 classes

3) 2007: added 10 classes

- Additional categories provide a broader domain and finer-grained categories, including visually similar things

→

2. Image Collection

- 500,000 images retrieved from Flickr by querying with a number of keywords

→

3. Image Verification + Image Annotation

- University of Leeds annotation party to recruit annotators

- Annotation guidelines & real-time assistance

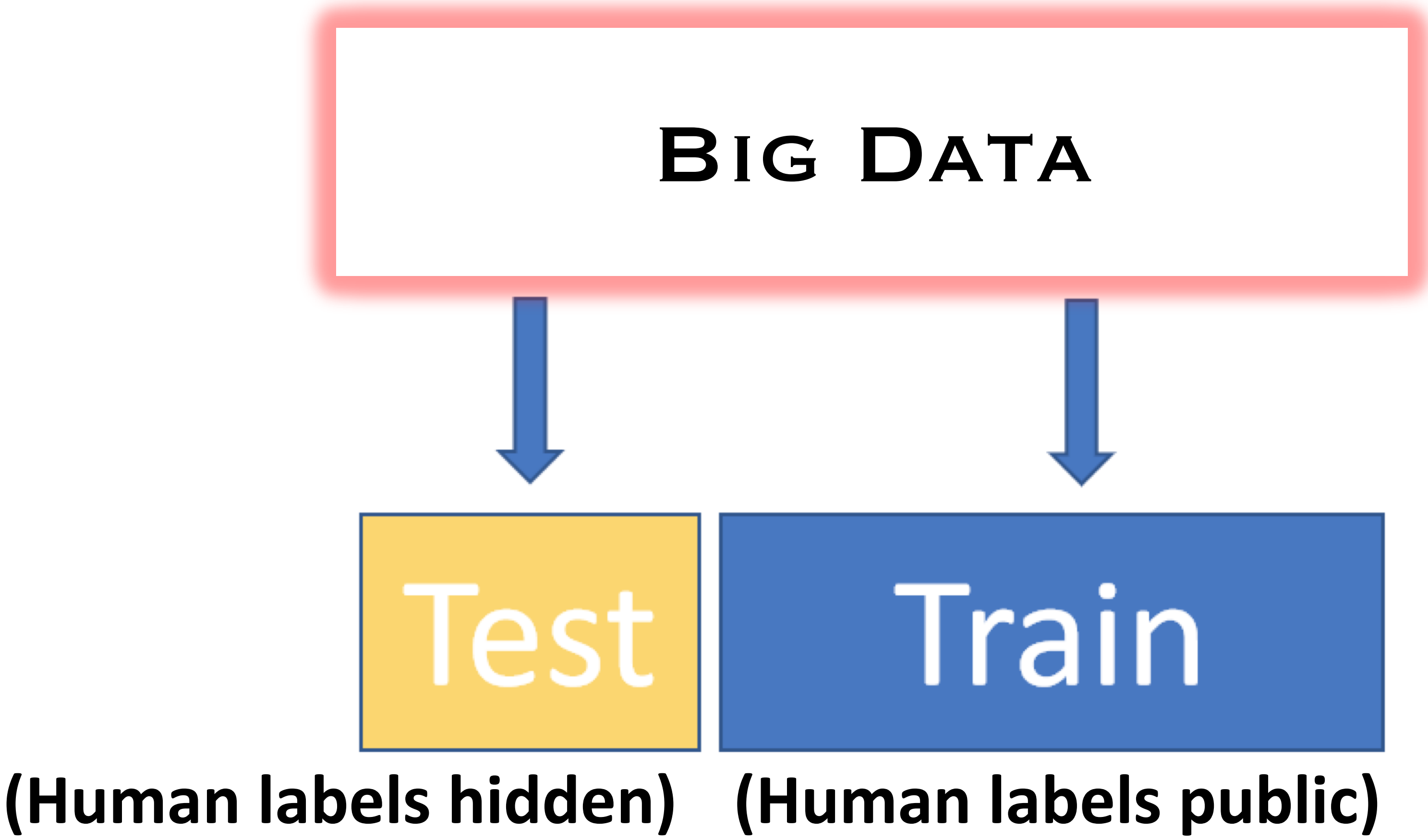
- Review of every annotation

- Annotate only "minority" classes at end of party to increase the count of them

Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman. The PASCAL Visual Object Classes (VOC) Challenge. IJCV 2010.

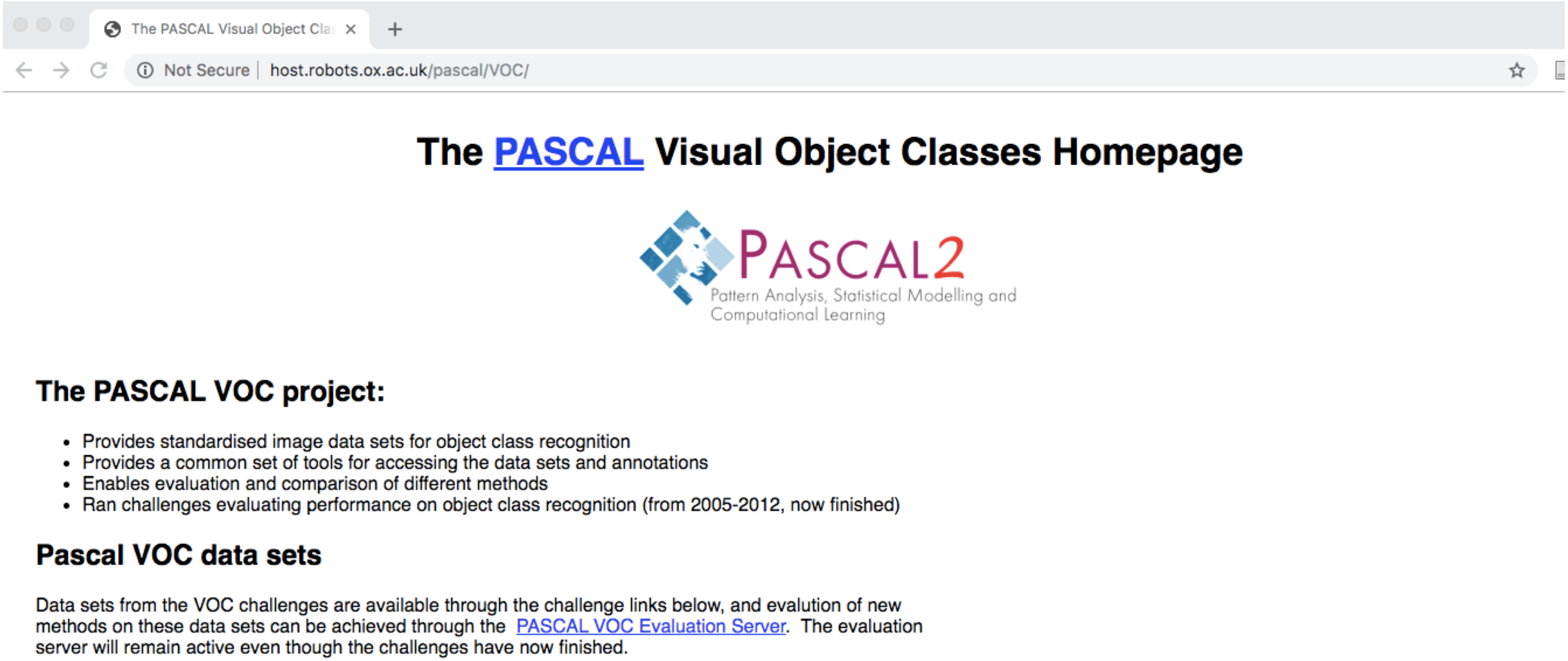
Object Detection: VOC Challenge

Object Detection: VOC Challenge



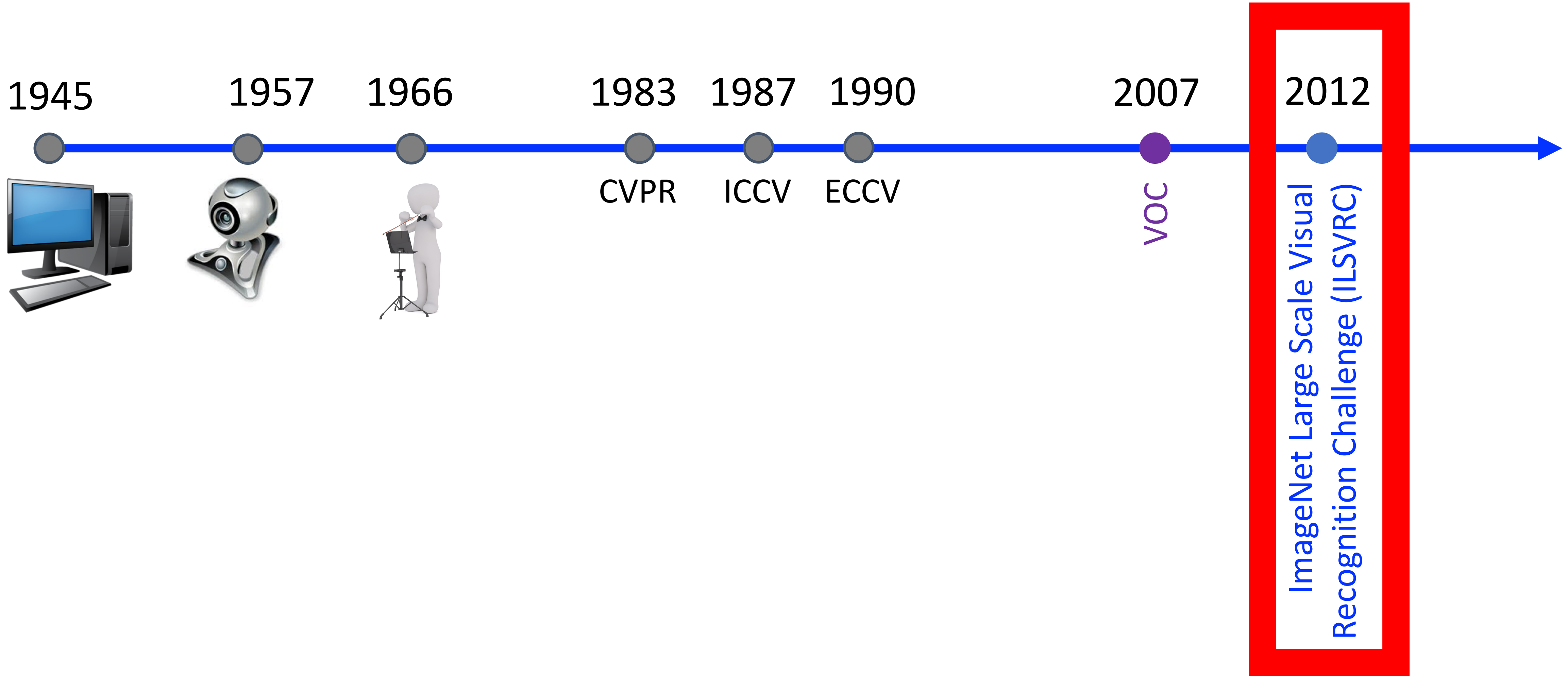
Winner: highest scoring method on the hidden test set

Object Detection: VOC Annual Workshop



http://host.robots.ox.ac.uk/pascal/VOC/

Object Detection Datasets



Object Detection Datasets: ILSVRC

"ILSVRC follows in the footsteps of the PASCAL VOC challenge… which set the precedent for standardized evaluation of recognition algorithms in the form of yearly competitions."

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, Li Fei-Fei , IJCV 2015

ILSVRC

1. Category Selection

- 200 ImageNet classes which:
1) exclude synset overlap
2) exclude object classes too "big" in the image
3) are basic-level categories
4) backward compatible: VOC

Class name in PASCAL VOC (20 classes)	Closest class in ILSVRC-DET (200 classes)	Avg object scale (%)	
		PASCAL VOC	ILSVRC-DET
aeroplane	airplane	29.7	22.4
bicycle	bicycle	29.3	14.3
bird	bird	15.9	20.1
boat	*watercraft*	15.2	16.5
bottle	*wine bottle*	7.3	10.4
bus	bus	29.9	22.1
car	car	14.0	13.4
cat	domestic cat	46.8	29.8
chair	chair	12.8	10.1
cow	*cattle*	19.3	13.5
dining table	*table*	29.1	30.3
dog	dog	37.0	28.9
horse	horse	29.5	18.5
motorbike	motorcyle	32.0	20.7
person	person	17.5	19.3
potted plant	*flower pot*	12.3	8.1
sheep	sheep	12.2	17.3
sofa	sofa	41.7	44.4
train	train	35.4	35.1
tv/monitor	tv or monitor	14.6	11.2

ILSVRC is more difficult!

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, Li Fei-Fei , IJCV 2015

Object Detection Datasets: ILSVRC

1. Category Selection

- 200 ImageNet classes which:
1) exclude synset overlap
2) exclude object classes too "big" in the image
3) are basic-level categories
4) backward compatible: VOC

→

2. Image Collection

- Subset of images from ImageNet

- Additional images from Flickr

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, Li Fei-Fei , IJCV 2015

Object Detection Datasets: ILSVRC

1. Category Selection

- 200 ImageNet classes which:
1) exclude synset overlap
2) exclude object classes too “big” in the image
3) are basic-level categories
4) backward compatible: VOC

→

2. Image Collection

- Subset of images from ImageNet

- Additional images from Flickr

→

3. Image Verification

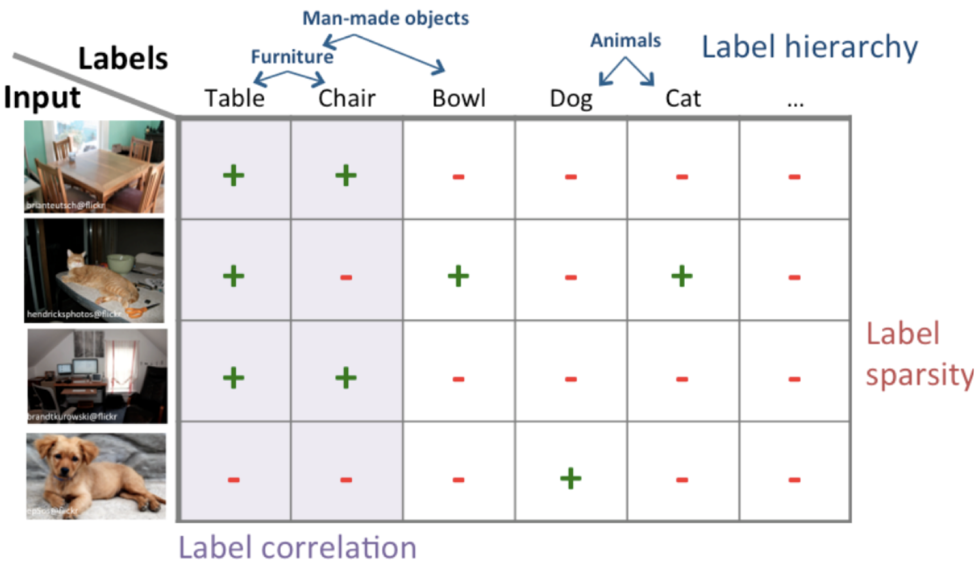
- Assign all relevant categories from 200 object categories to each image

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, Li Fei-Fei , IJCV 2015

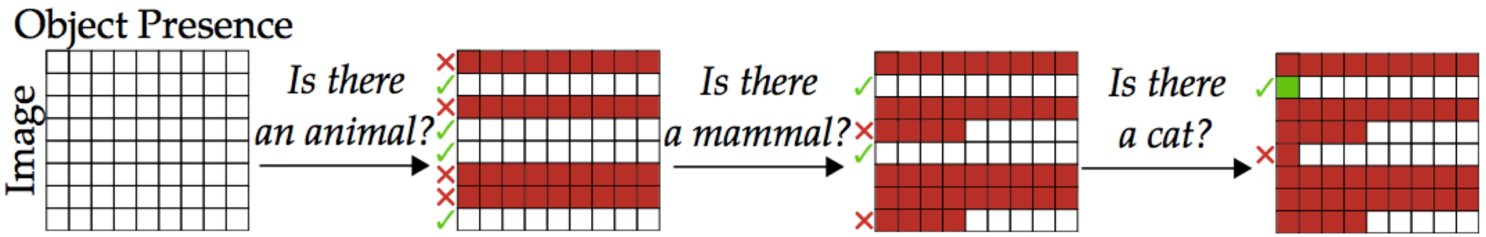
Recall from ImageNet: Efficient Object Presence Labeling

- Brute force approach:
  - 200 categories x 500,000 images = ??? queries
  - 100,000,000 queries; inefficient!
- Proposed approach

Idea:

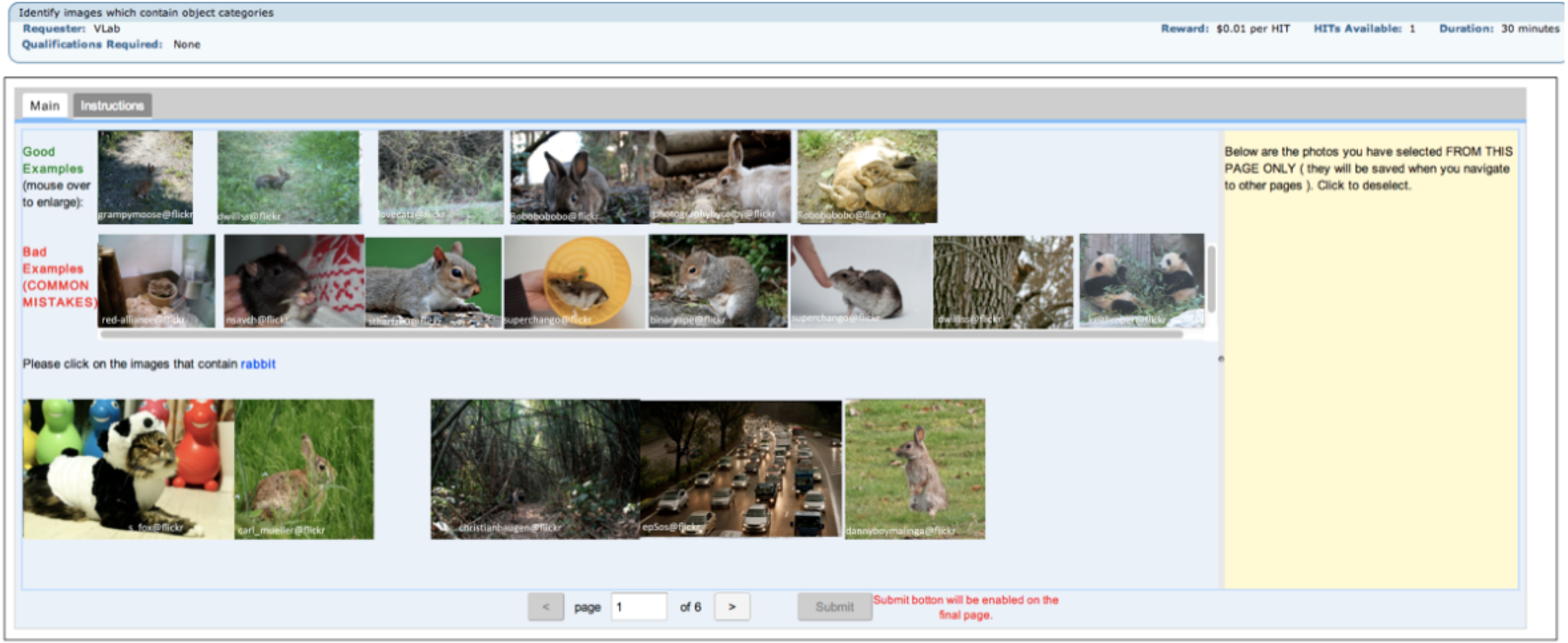


Approach:



Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, Li Fei-Fei , IJCV 2015

Recall from ImageNet:
Efficient Object Presence Labeling



Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, Li Fei-Fei , IJCV 2015

Object Detection Datasets: ILSVRC

1. Category Selection

- 200 ImageNet classes which:
1) exclude synset overlap
2) exclude object classes too “big” in the image
3) are basic-level categories
4) backward compatible: VOC

2. Image Collection

- Subset of images from ImageNet
- Additional images from Flickr

3. Image Verification

- Assign all relevant categories from 200 object categories to each image

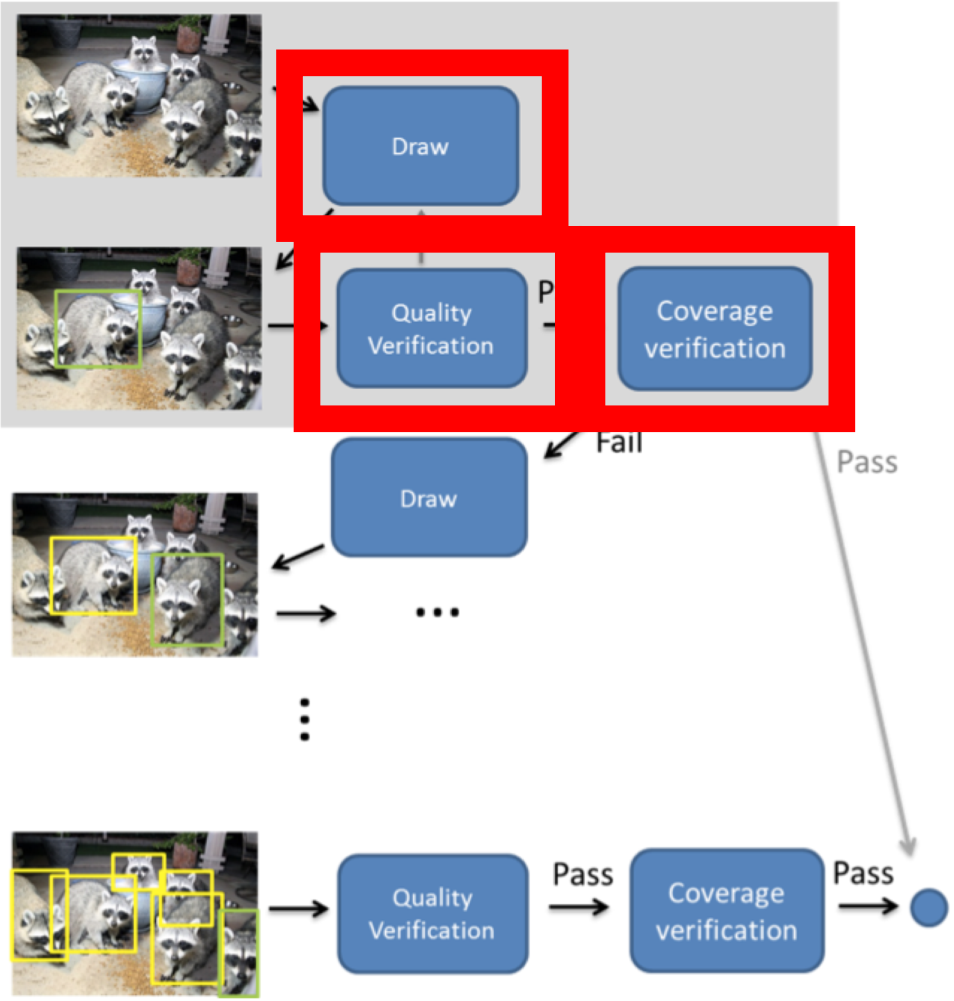
4. Image Annotation

- Demarcate a bounding box around EVERY instance of every object category

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, Li Fei-Fei , IJCV 2015

Efficient Object Localization

- 3 Tasks:

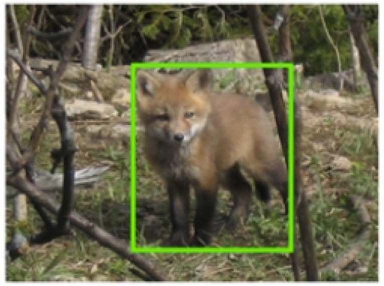


Idea: each task has fixed and predictable amount of work

Hao Su, Jia Deng, and Li Fei-Fei. Crowdsourcing Annotations for Visual Object Detection. AAAI 2012.

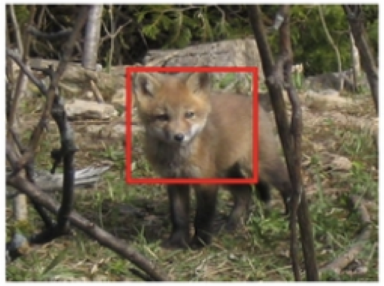
Drawing Training

Rule 1: Include all visible part and draw as tightly as possible.

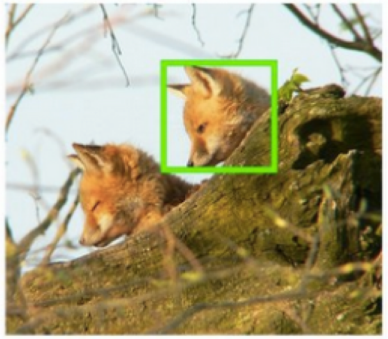


CORRECT

WRONG: must be as tight as possible!



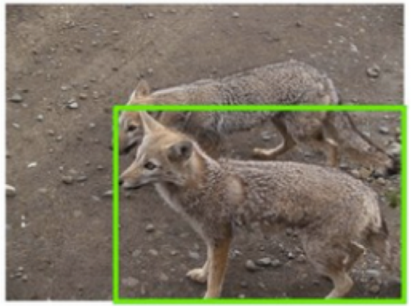
WRONG: must include all visible parts!



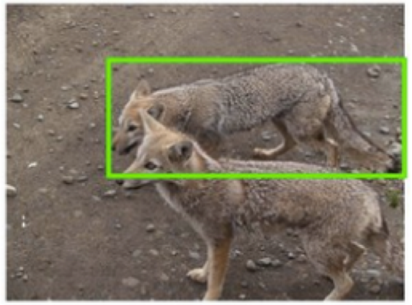
CORRECT

WRONG: occluded parts do not matter as long as all visible parts are included.

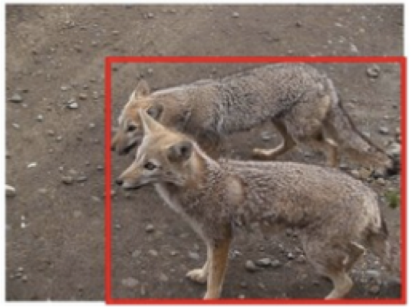
Rule 2: If there are multiple instances, include only ONE (any one).



CORRECT



CORRECT



WRONG: should include only one instance.

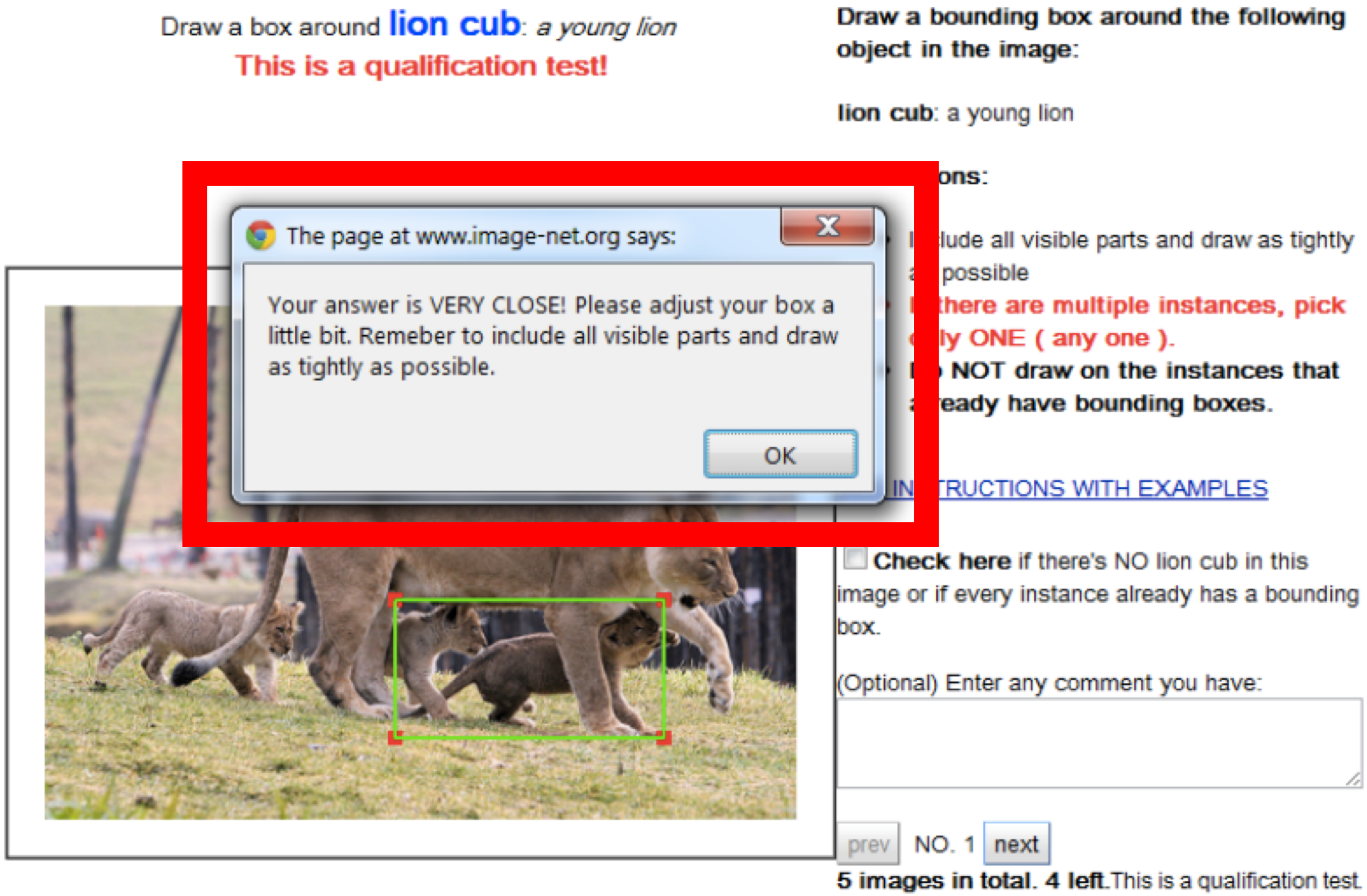
Rule 3: DO NOT draw on an instance that already has a bounding box, as shown below in yellow. Draw on a new instance.

Hao Su, Jia Deng, and Li Fei-Fei. Crowdsourcing Annotations for Visual Object Detection. AAAI 2012.

Drawing Task

Hao Su, Jia Deng, and Li Fei-Fei. Crowdsourcing Annotations for Visual Object Detection. AAAI 2012.

Drawing Qualification Test



- Verify rules are understood with test images

- Train with 3 types of feedback messages:

1) bounding box is not sufficiently tight

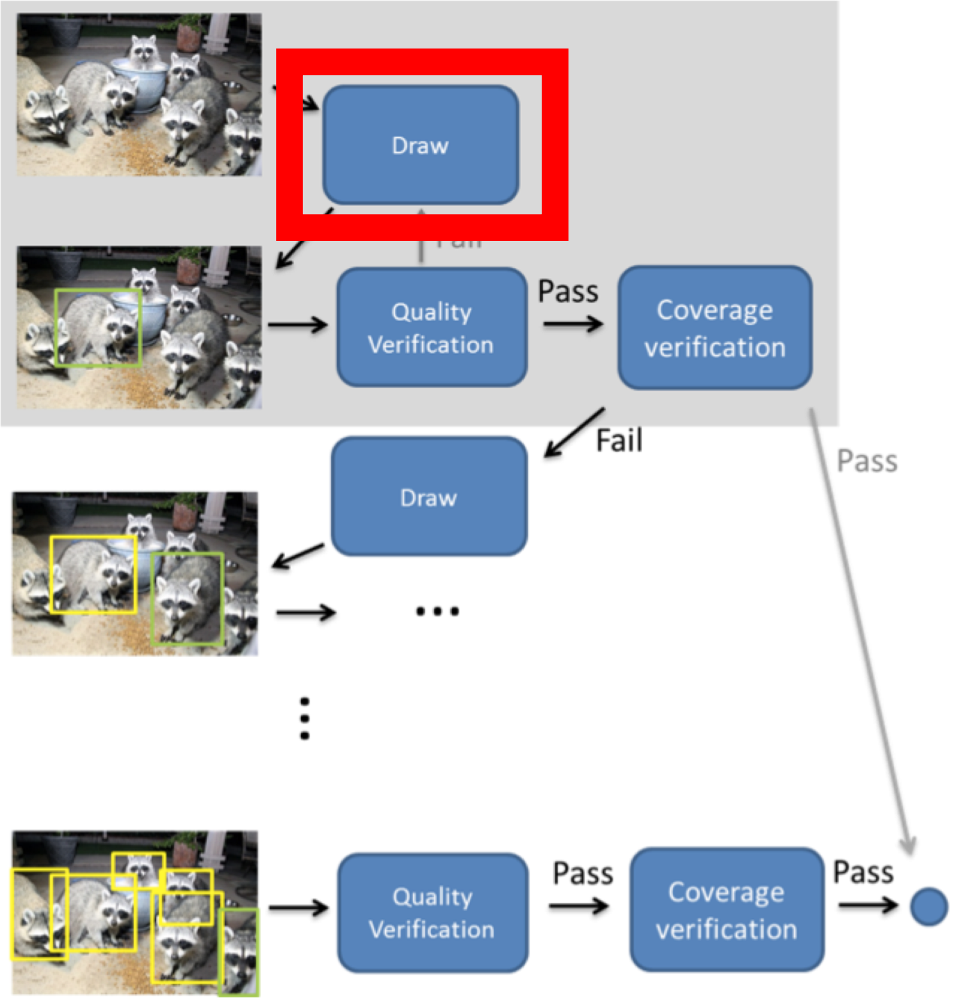
2) object selected is not the solicited object

3) object selected already has bounding box

Hao Su, Jia Deng, and Li Fei-Fei. Crowdsourcing Annotations for Visual Object Detection. AAAI 2012.

Efficient Object Localization

- 3 Tasks:



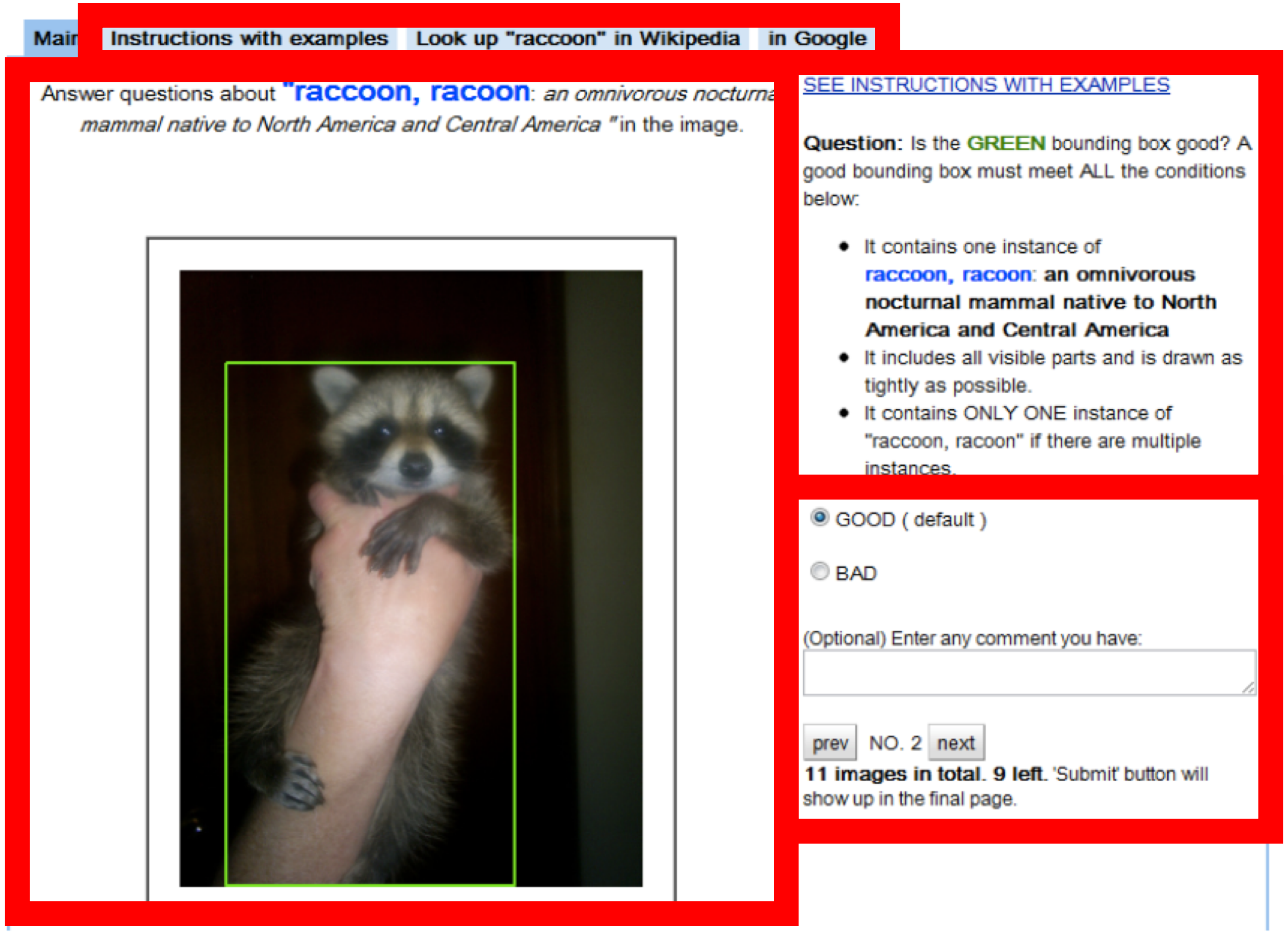
Hao Su, Jia Deng, and Li Fei-Fei. Crowdsourcing Annotations for Visual Object Detection. AAAI 2012.

Quality Verification **Training**

- Instructions with illustrations explaining 3 rules:
  - Rule 1: Good bounding box must include instance of the required object.
  - Rule 2: Good bounding box must include all visible parts and be as tight as possible.
  - Rule 3: If there are multiple instances, a good bounding box must include only ONE (any one).

Hao Su, Jia Deng, and Li Fei-Fei. Crowdsourcing Annotations for Visual Object Detection. AAAI 2012.

Quality Verification Task



Hao Su, Jia Deng, and Li Fei-Fei. Crowdsourcing Annotations for Visual Object Detection. AAAI 2012.

Quality Verification **Qualification Test**

Successfully rate test images known to have good and bad bounding boxes

Hao Su, Jia Deng, and Li Fei-Fei. Crowdsourcing Annotations for Visual Object Detection. AAAI 2012.

Quality Verification **Quality Control**

Trust work if worker does well on validation images in the task batch:

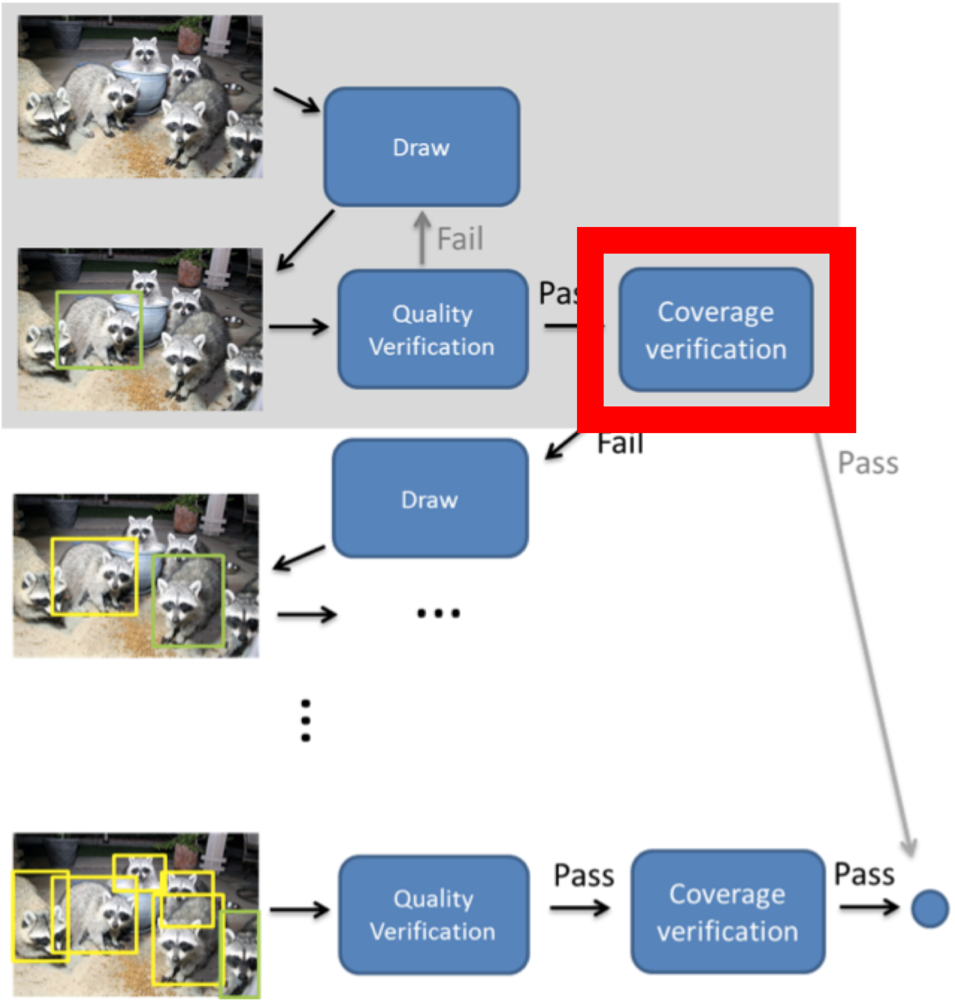
- Good bounding boxes: bounding boxes rated by multiple workers with strong consensus as ``good”

- Bad bounding boxes: perturb good bounding boxes

Hao Su, Jia Deng, and Li Fei-Fei. Crowdsourcing Annotations for Visual Object Detection. AAAI 2012.

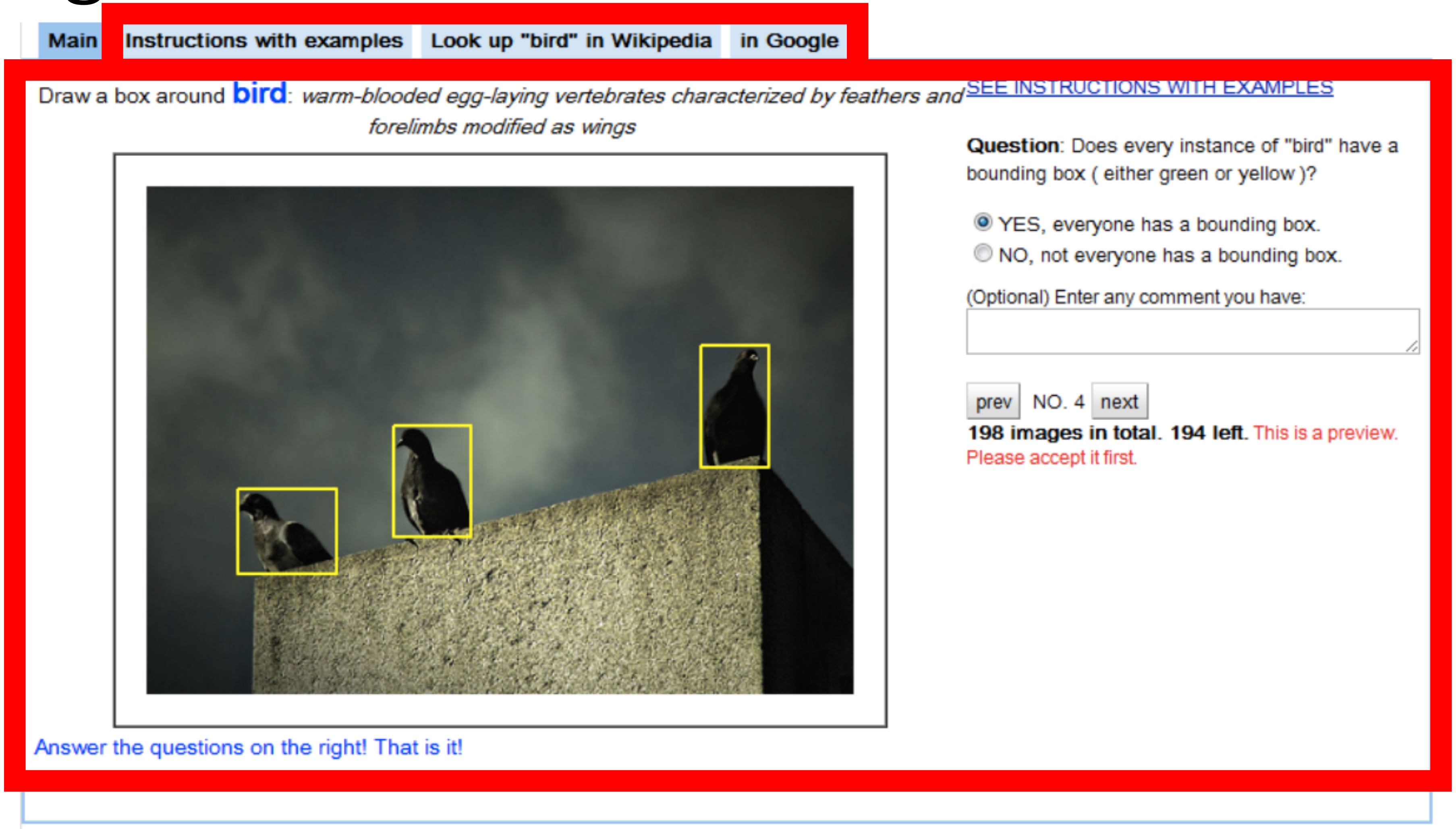
Efficient Object Localization

- 3 Tasks:



Hao Su, Jia Deng, and Li Fei-Fei. Crowdsourcing Annotations for Visual Object Detection. AAAI 2012.

Coverage Verification Task



Hao Su, Jia Deng, and Li Fei-Fei. Crowdsourcing Annotations for Visual Object Detection. AAAI 2012.

Coverage Verification Qualification Test

Successfully rate test images known to show all bounding boxes demarcated and some bounding boxes missing

Hao Su, Jia Deng, and Li Fei-Fei. Crowdsourcing Annotations for Visual Object Detection. AAAI 2012.

Coverage Verification **Quality Control**

Trust work if worker does well on validation images in the task batch:

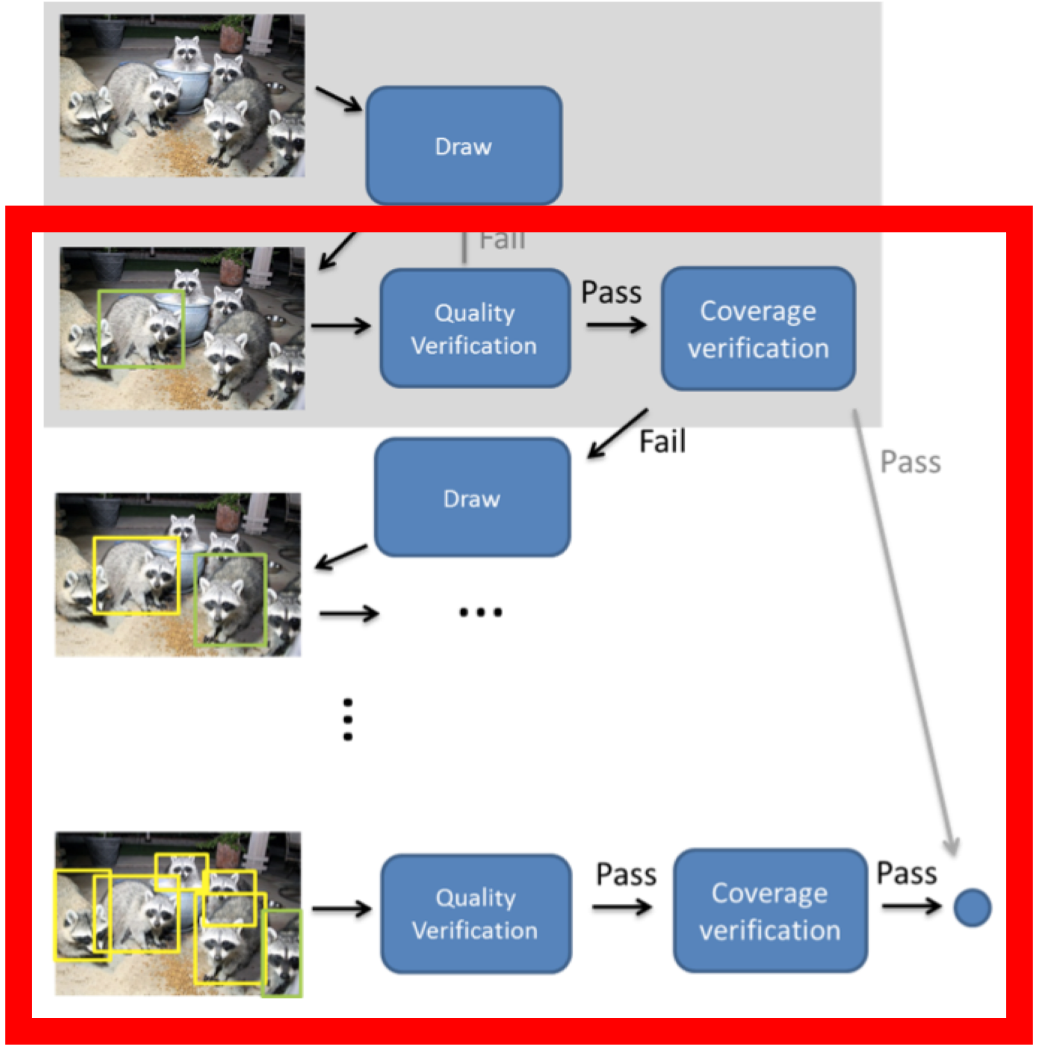
- Good coverage: coverage rated by multiple workers with strong consensus as ``good”

- Bad coverage: remove bounding boxes from images with ``good” coverage

Hao Su, Jia Deng, and Li Fei-Fei. Crowdsourcing Annotations for Visual Object Detection. AAAI 2012.

Efficient Object Localization

- 3 Tasks:



Hao Su, Jia Deng, and Li Fei-Fei. Crowdsourcing Annotations for Visual Object Detection. AAAI 2012.

Analysis of Task Decomposition Task

Task Name	Time per b.box	
	Median	Mean
Drawing	25.5s	50.8s
Quality Verification	9.0s	21.9s
Coverage Verification	7.8s	15.3s
Total	42.4s	88.0s

Proposed system: ~88.0 seconds per BB
BB Consensus: ~116.9 seconds (50.8×2+15.3)

Consensus approach is at least 32.8% more expensive than proposed approach!

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, Li Fei-Fei , IJCV 2015

Object Detection Datasets: ILSVRC

1. Category Selection

- 200 ImageNet classes which:
1) exclude synset overlap
2) exclude object classes too “big” in the image
3) are basic-level categories
4) backward compatible: VOC

→

2. Image Collection

- Train dataset: 3 sources

- Val & test datasets: 2 sources

→

3. Object presence labeling

- Assign all relevant categories from 200 object categories to each image

→

4. Object localization

- Demarcate a bounding box around EVERY instance of every object category

→

5. Author Review

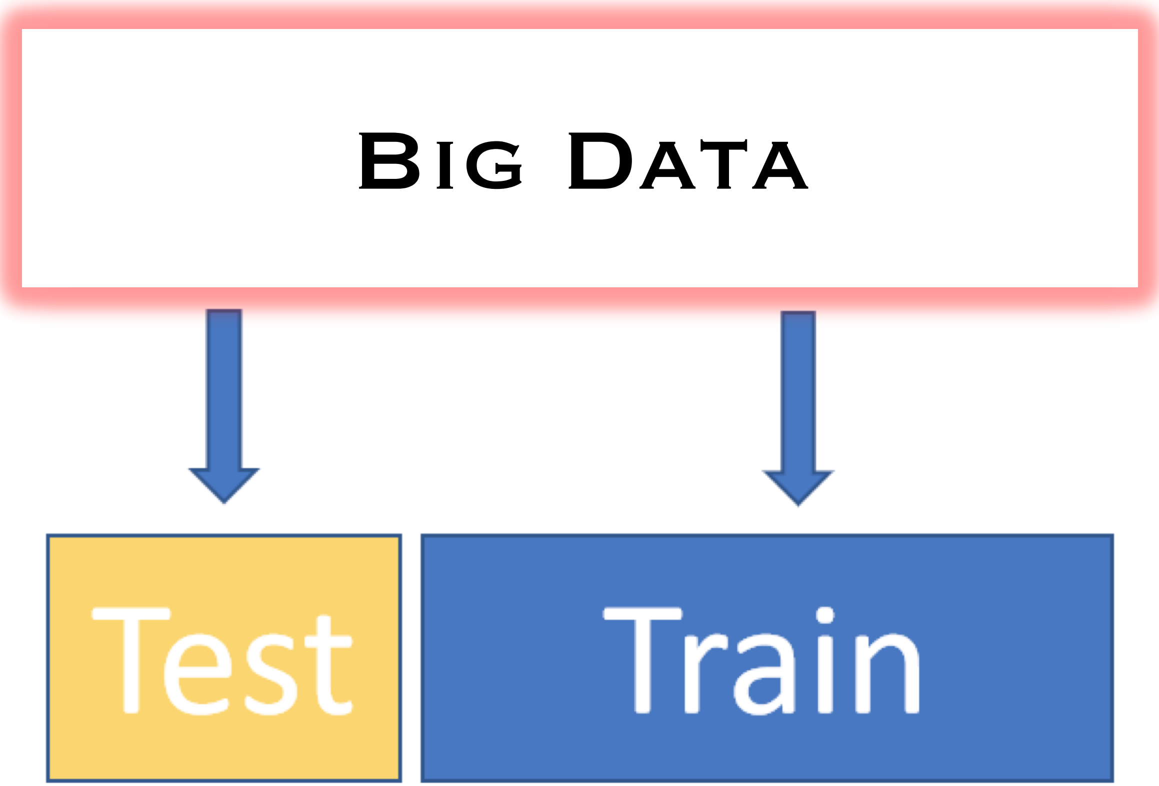
- Ambiguous objects: BB for two categories with large overlap(~3%).

- Duplicates: >50% overlap for same object (~1%).

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, Li Fei-Fei , IJCV 2015; *1,955 citations in 2/17*

Object Detection: ILSVRC Challenge

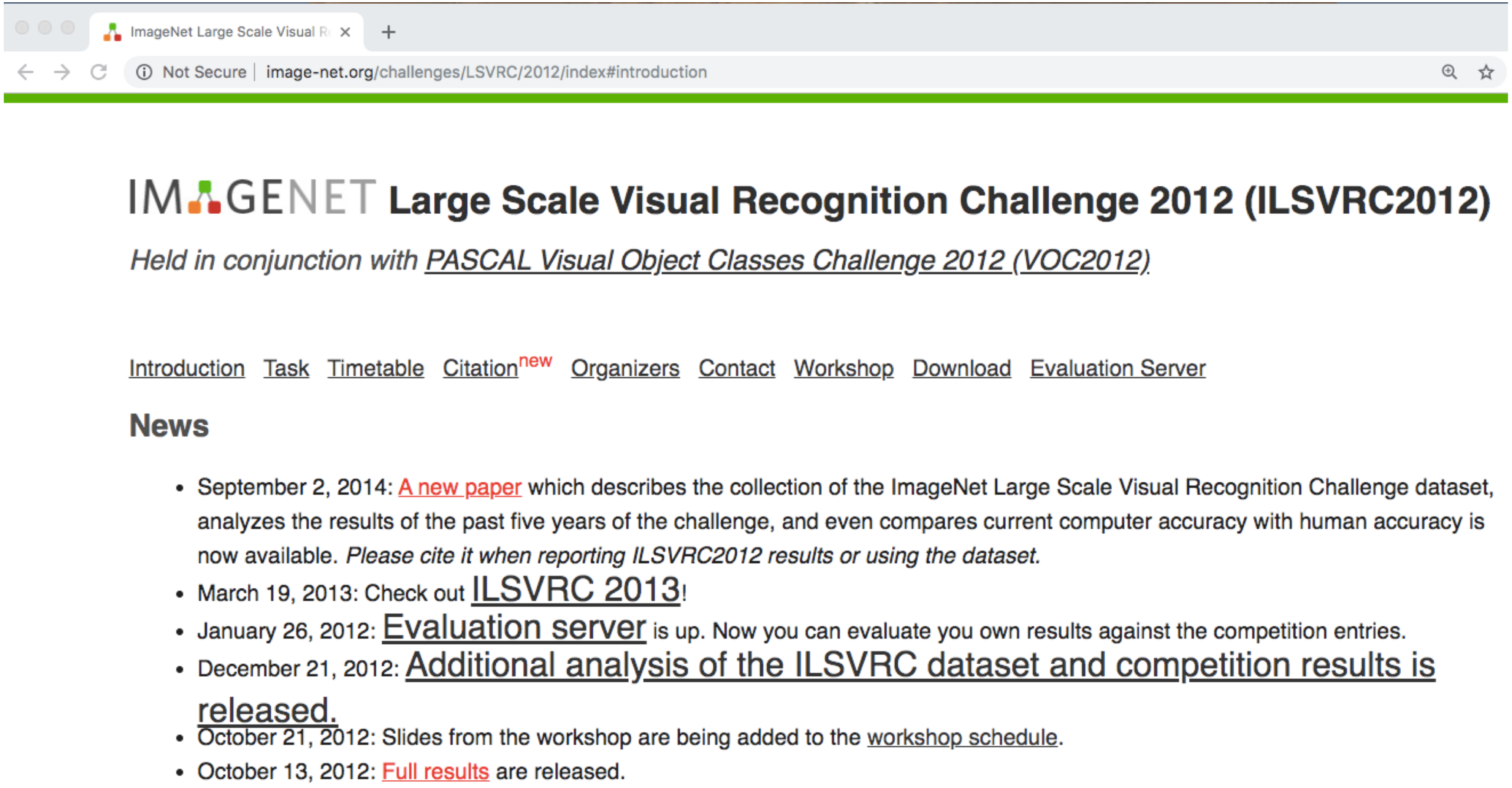
Object Detection: ILSVRC Challenge



(Human labels hidden) (Human labels public)

Winner: highest scoring method on the hidden test set

Object Detection: ILSVRC Annual Workshop



ImageNet Large Scale Visual R

Not Secure | image-net.org/challenges/LSVRC/2012/index#introduction

IMAGENET **Large Scale Visual Recognition Challenge 2012 (ILSVRC2012)**

Held in conjunction with PASCAL Visual Object Classes Challenge 2012 (VOC2012)

Introduction Task Timetable Citation[new] Organizers Contact Workshop Download Evaluation Server

News

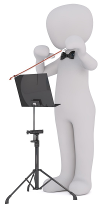
- September 2, 2014: A new paper which describes the collection of the ImageNet Large Scale Visual Recognition Challenge dataset, analyzes the results of the past five years of the challenge, and even compares current computer accuracy with human accuracy is now available. *Please cite it when reporting ILSVRC2012 results or using the dataset.*
- March 19, 2013: Check out ILSVRC 2013!
- January 26, 2012: Evaluation server is up. Now you can evaluate you own results against the competition entries.
- December 21, 2012: Additional analysis of the ILSVRC dataset and competition results is released.
- October 21, 2012: Slides from the workshop are being added to the workshop schedule.
- October 13, 2012: Full results are released.

http://image-net.org/challenges/LSVRC/2012/index#introduction

Object Detection Datasets

1945 — 1957 — 1966 — 1983 — 1987 — 1990 — 2007 — 2012





1983: CVPR
1987: ICCV
1990: ECCV
2007: VOC
2012: ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

	VOC	ILSVRC
# Categories for Final Version of Dataset:	20	200
# Images for Final Version of Dataset:	21,738	516,840

Today’s Topics

- Object detection applications
- Object detection evaluation
- Crowdsourcing object detection
- Class discussion (chosen by YOU ☺)
- Lab: drawing on images

Today's Topics

- Object detection applications
- Object detection evaluation
- Crowdsourcing object detection
- Class discussion (chosen by YOU ☺)
- Lab: drawing on images