Object Detection

Danna Gurari
University of Colorado Boulder
Fall 2021

https://home.cs.colorado.edu/~DrG/Courses/RecentAdvancesInComputerVision/AboutCourse.html

Review

- Last lecture:
  - Scene Classification Problem
  - Scene Classification Applications
  - Scene Classification: Evolution of Datasets
  - Scene Classification Evaluation Metrics
  - Scene Classification Background: Deep Features and Fine-Tuning
  - Scene Classification Computer Vision Models
- Assignments (Canvas)
  - Reading assignment due this Wednesday
- Questions?

Object Detection: Today's Topics

- Problem
- Applications
- Datasets
- Evaluation metric
- Background: naive sliding window solution

Object Detection: Today's Topics

- Problem
- Applications
- Datasets
- Evaluation metric
- Background: naive sliding window solution

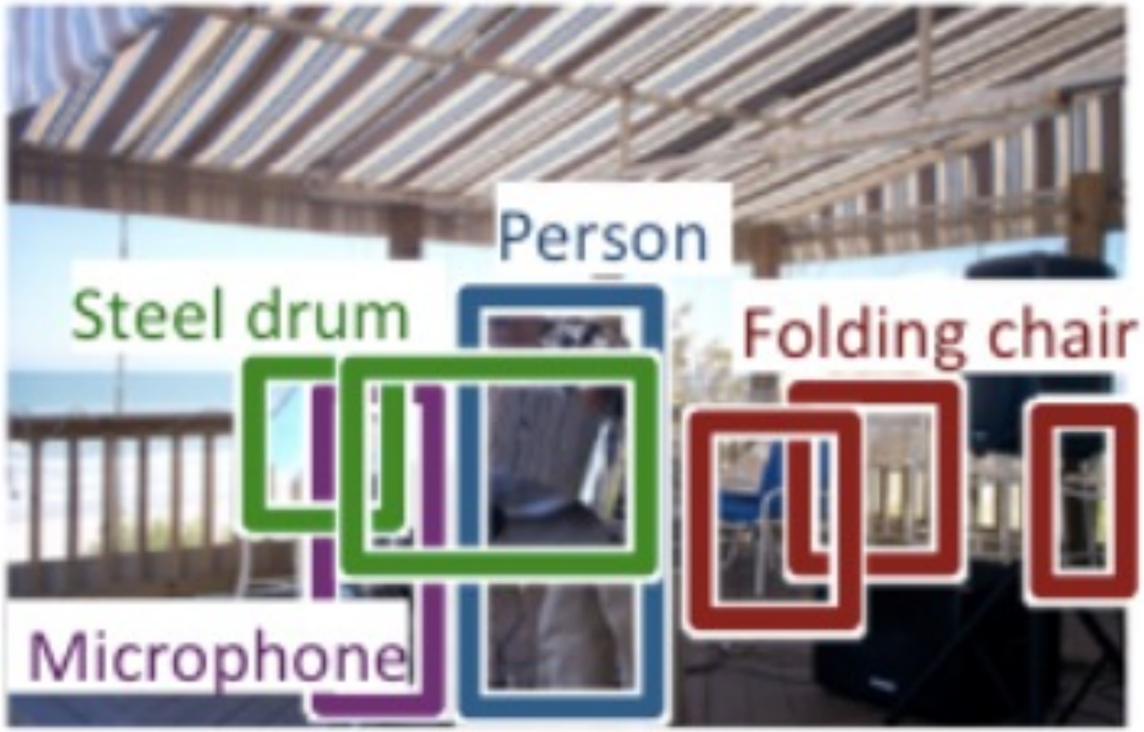
Problem Definition

- Localize with a bounding box object(s) of interest

Focus for today's lecture

Problem: **Semantic** Object Detection

- Localize with a bounding box every instance of an object from pre-specified categories



[Russakovsky et al; IJCV 2015]

To be covered in a future lecture

Problem: Salient Object Detection

- Localize with a bounding box the salient object(s)

[Liu et al; CVPR 2007]

Object Detection vs Object Recognition

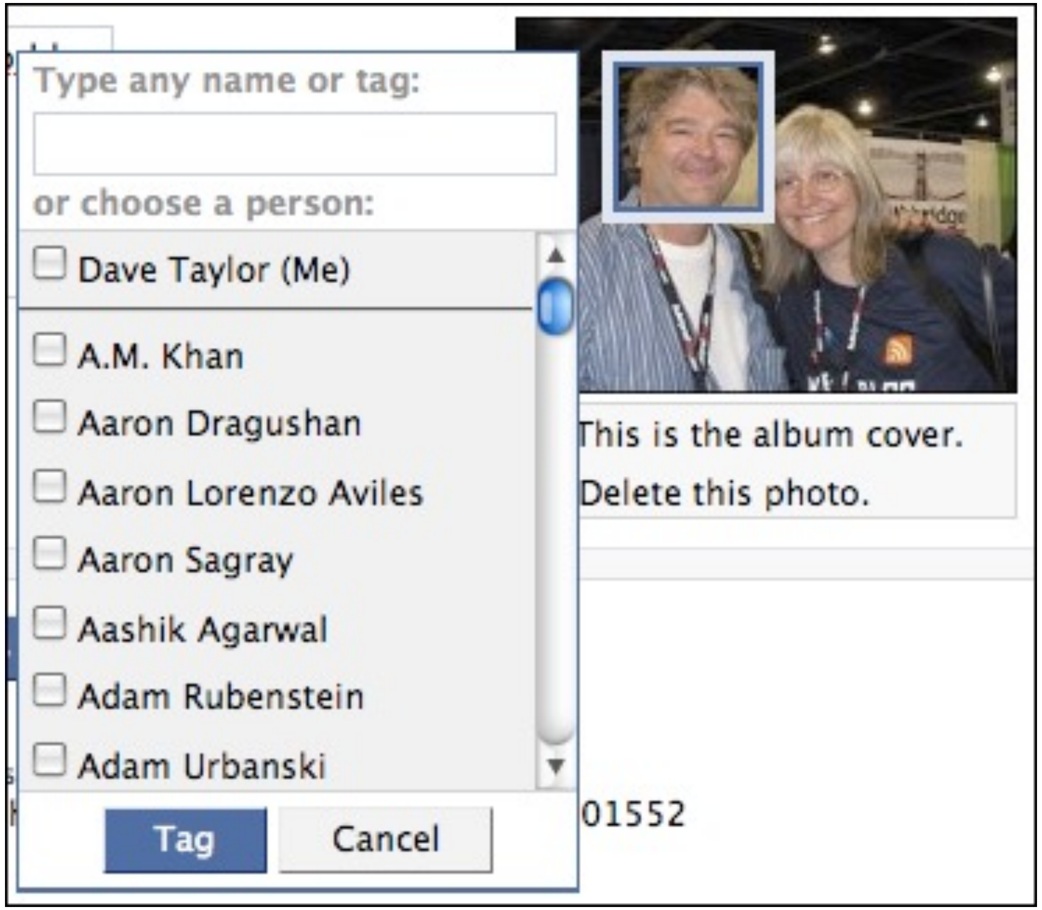
"What is the difference between (semantic) object detection and object recognition?"

- Must learn appearance of object rather than only its image context; e.g., giraffe

Object Detection: Today’s Topics

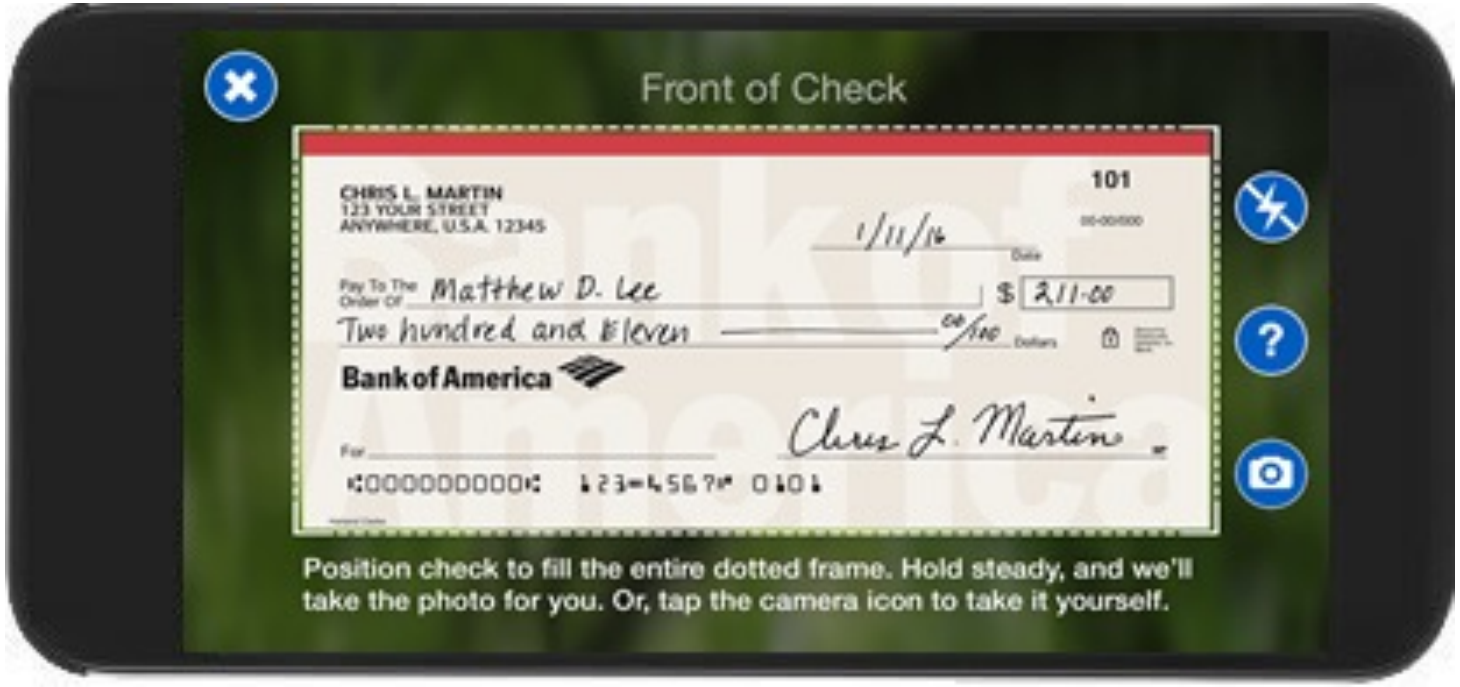
- Problem
- Applications
- Datasets
- Evaluation metric
- Background: naive sliding window solution

Social Media



Face detection
(e.g., Facebook)

Banking

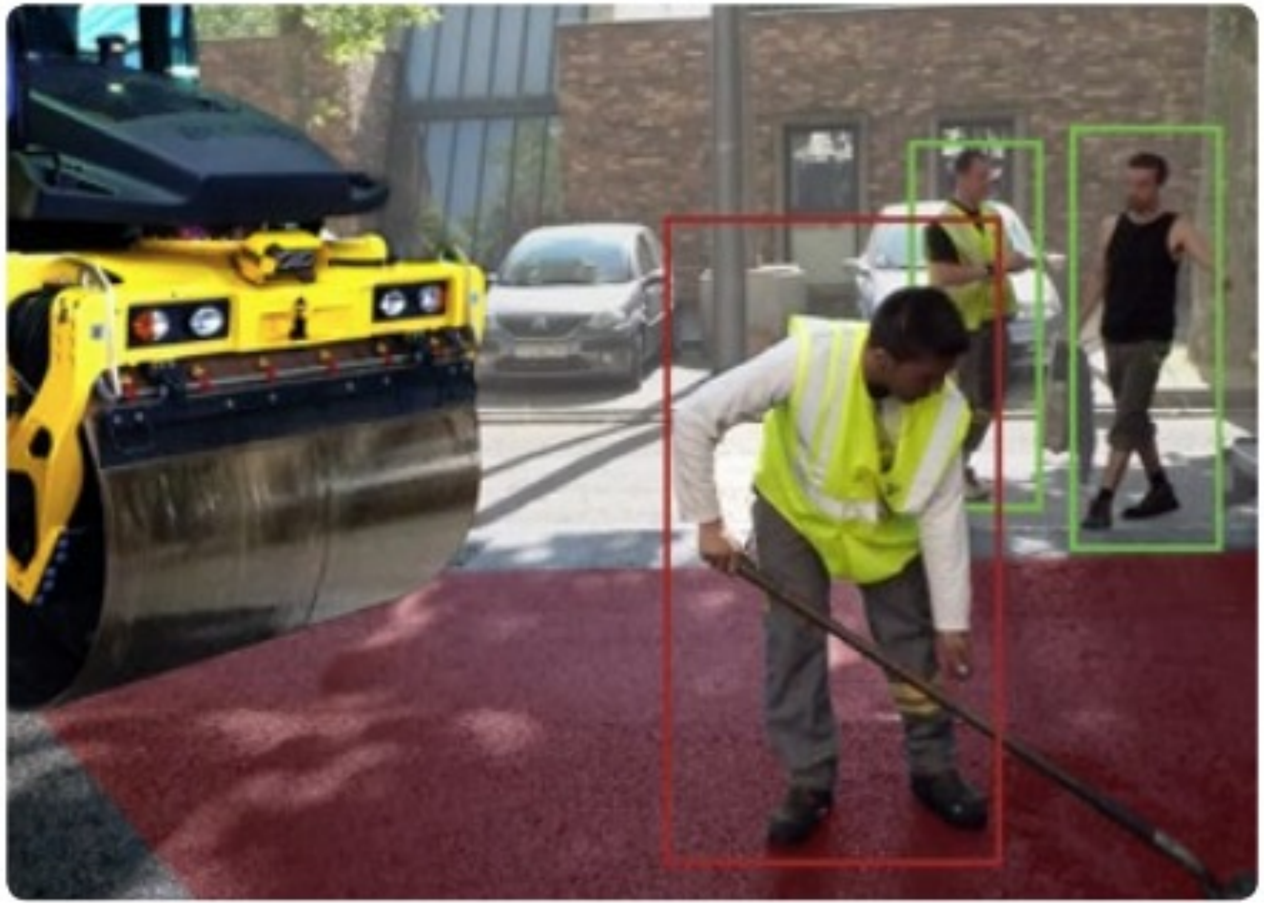


Mobile check deposit
(e.g., Bank of America)

Transportation

License Plate Detection (e.g., AllGoVision)

Construction Safety



Pedestrian Detection
(e.g., Blaxtair)

http://media.brintex.com/Occurrence/121/Brochure/3435/brochure.pdf

Counting

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

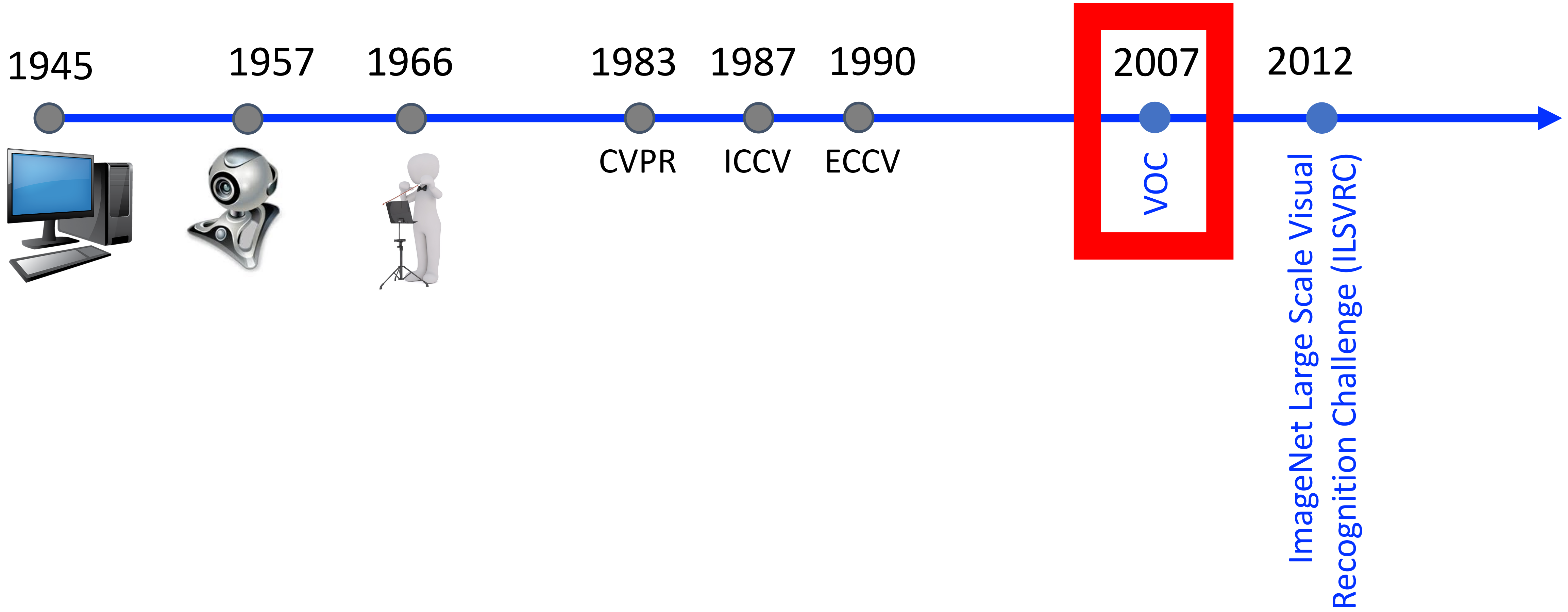
Business Traffic Analytics

Can you think of any other potential applications?

Object Detection: Today's Topics

- Problem
- Applications
- Datasets
- Evaluation metric
- Background: naive sliding window solution

Object Detection Datasets



VOC

1. Category Selection

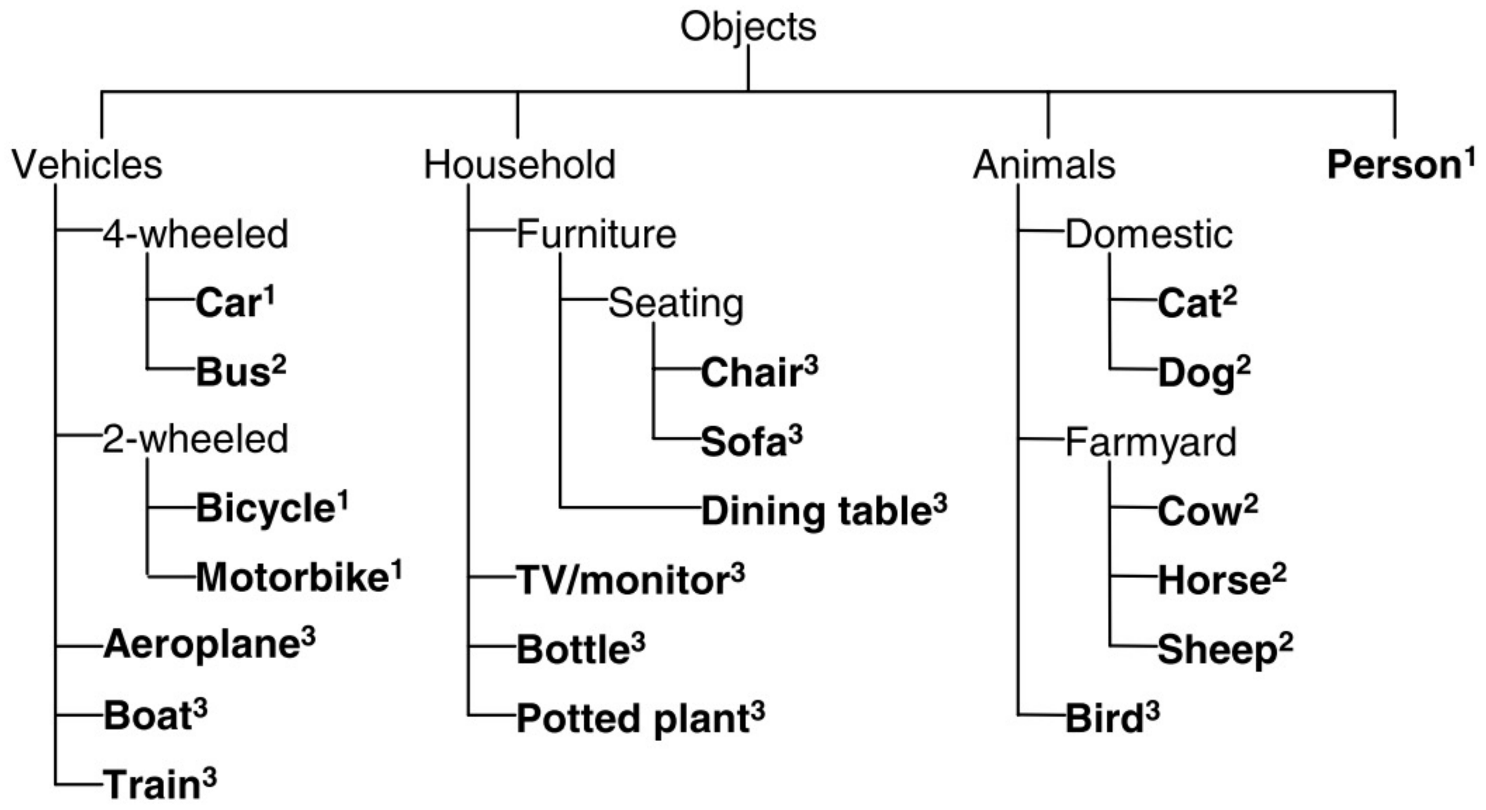
- 20 categories chosen:

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.

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.

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

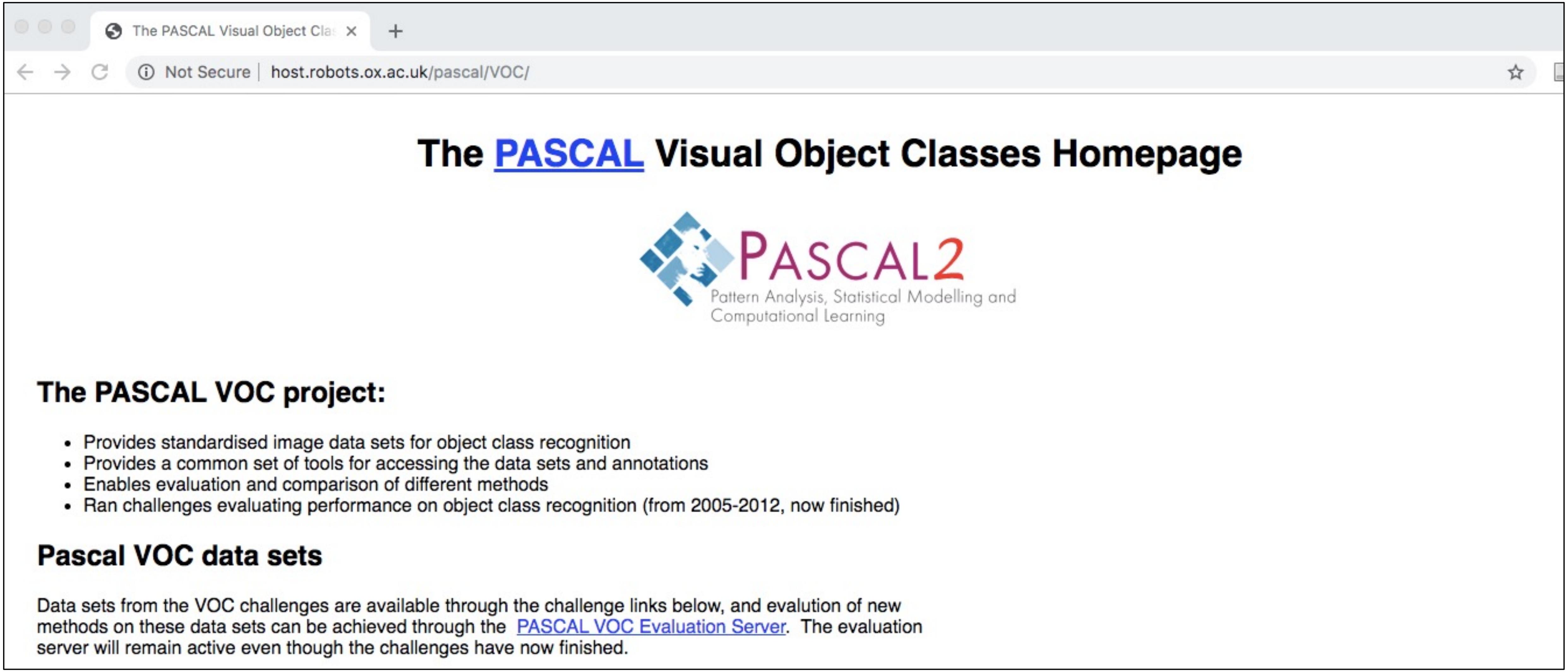
- Annotate only “minority” classes at end of party to increase the count of them

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 Annual Workshop



The PASCAL Visual Object Cla
Not Secure | host.robots.ox.ac.uk/pascal/VOC/

The PASCAL Visual Object Classes Homepage

PASCAL2
Pattern Analysis, Statistical Modelling and Computational Learning

The PASCAL VOC project:

- Provides standardised image data sets for object class recognition
- Provides a common set of tools for accessing the data sets and annotations
- Enables evaluation and comparison of different methods
- Ran challenges evaluating performance on object class recognition (from 2005-2012, now finished)

Pascal VOC data sets

Data sets from the VOC challenges are available through the challenge links below, and evalution of new methods on these data sets can be achieved through the PASCAL VOC Evaluation Server. The evaluation server will remain active even though the challenges have now finished.

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

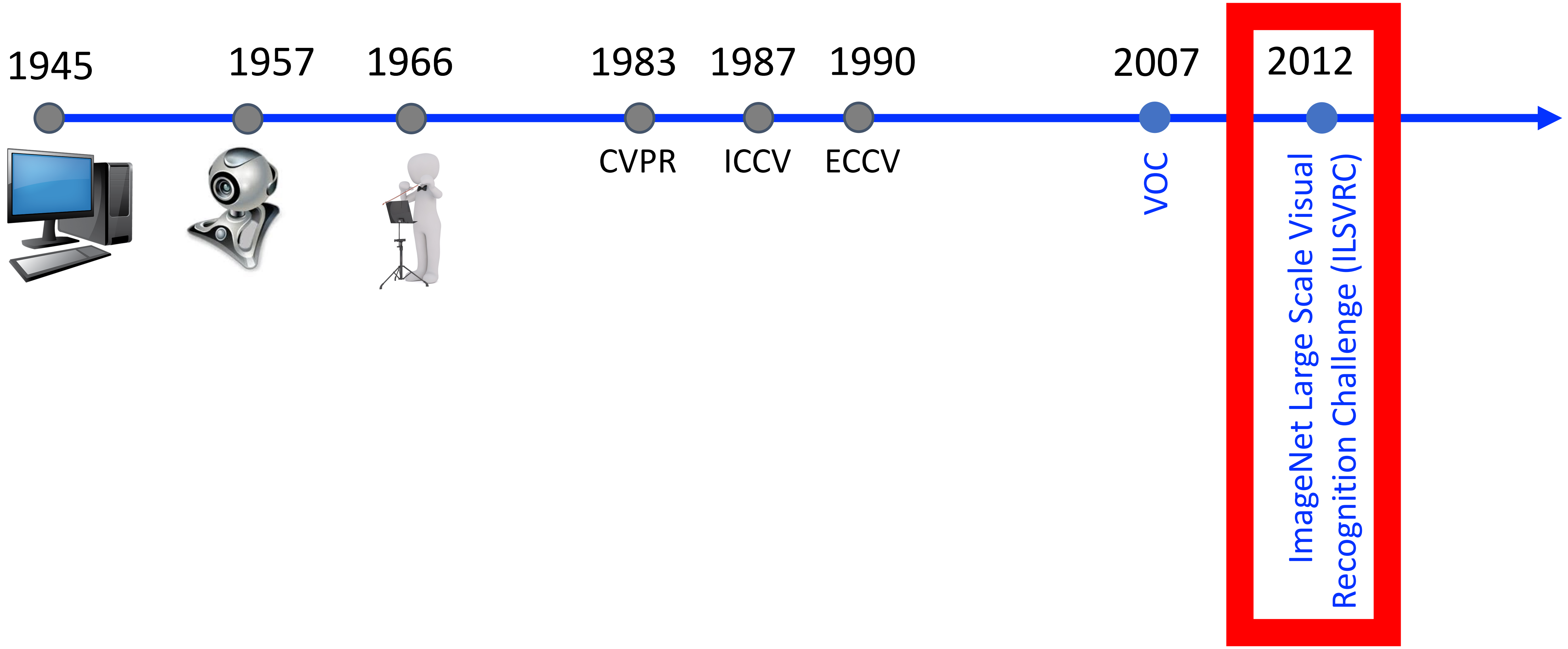
VOC: Datasets Evolved

The table below gives a brief summary of the main stages of the VOC development.

Year	Statistics	New developments	Notes
2005	Only 4 classes: bicycles, cars, motorbikes, people. Train/validation/test: 1578 images containing 2209 annotated objects.	Two competitions: classification and detection	Images were largely taken from exising public datasets, and were not as challenging as the flickr images subsequently used. This dataset is obsolete.
2006	10 classes: bicycle, bus, car, cat, cow, dog, horse, motorbike, person, sheep. Train/validation/test: 2618 images containing 4754 annotated objects.	Images from flickr and from Microsoft Research Cambridge (MSRC) dataset	The MSRC images were easier than flickr as the photos often concentrated on the object of interest. This dataset is obsolete.

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

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)
aeroplane	airplane
bicycle	bicycle
bird	bird
boat	*watercraft*
bottle	*wine bottle*
bus	bus
car	car
cat	domestic cat
chair	chair
cow	*cattle*
dining table	*table*
dog	dog
horse	horse
motorbike	motorcyle
person	person
potted plant	*flower pot*
sheep	sheep
sofa	sofa
train	train
tv/monitor	tv or monitor

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

→

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

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

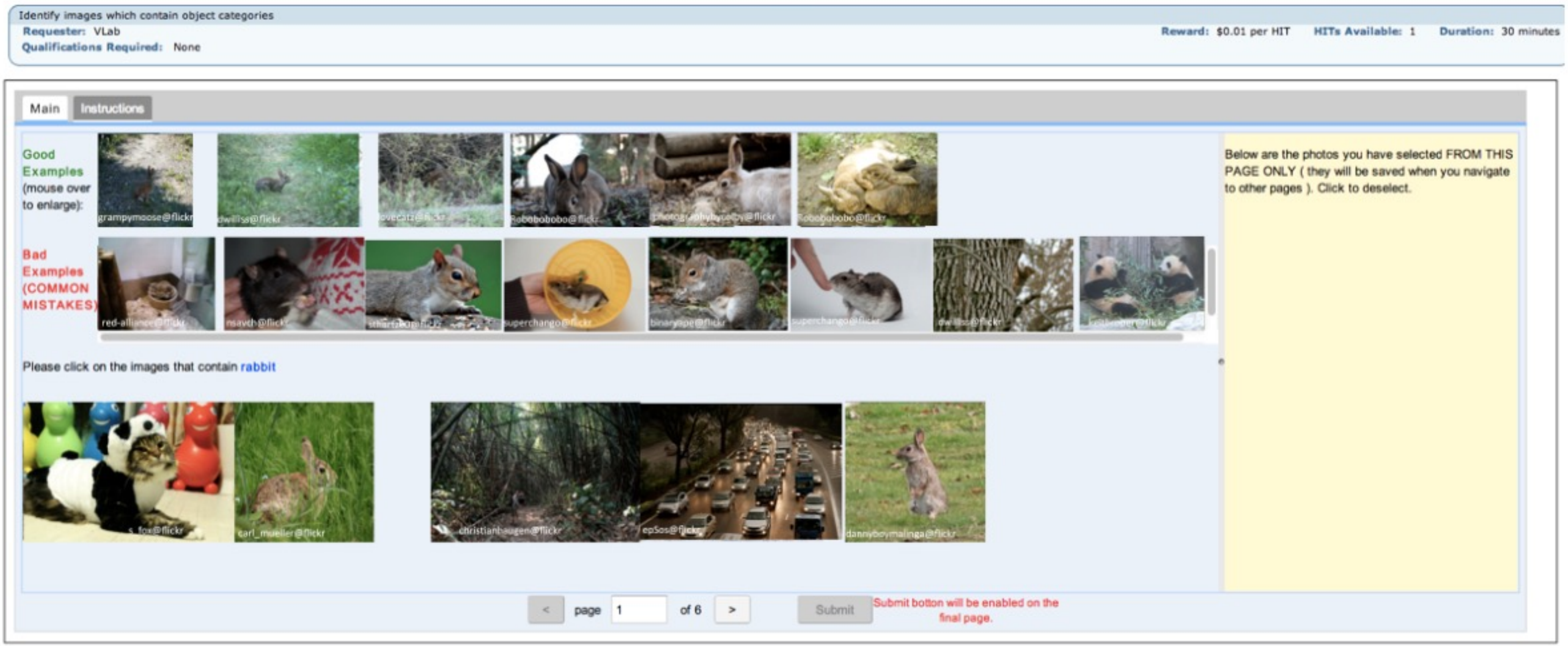
- Subset of images from ImageNet
- Additional images from Flickr

3. Image Verification

- Crowdsource assigning 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: 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

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

- Crowdsource assigning all relevant categories from 200 object categories to each image

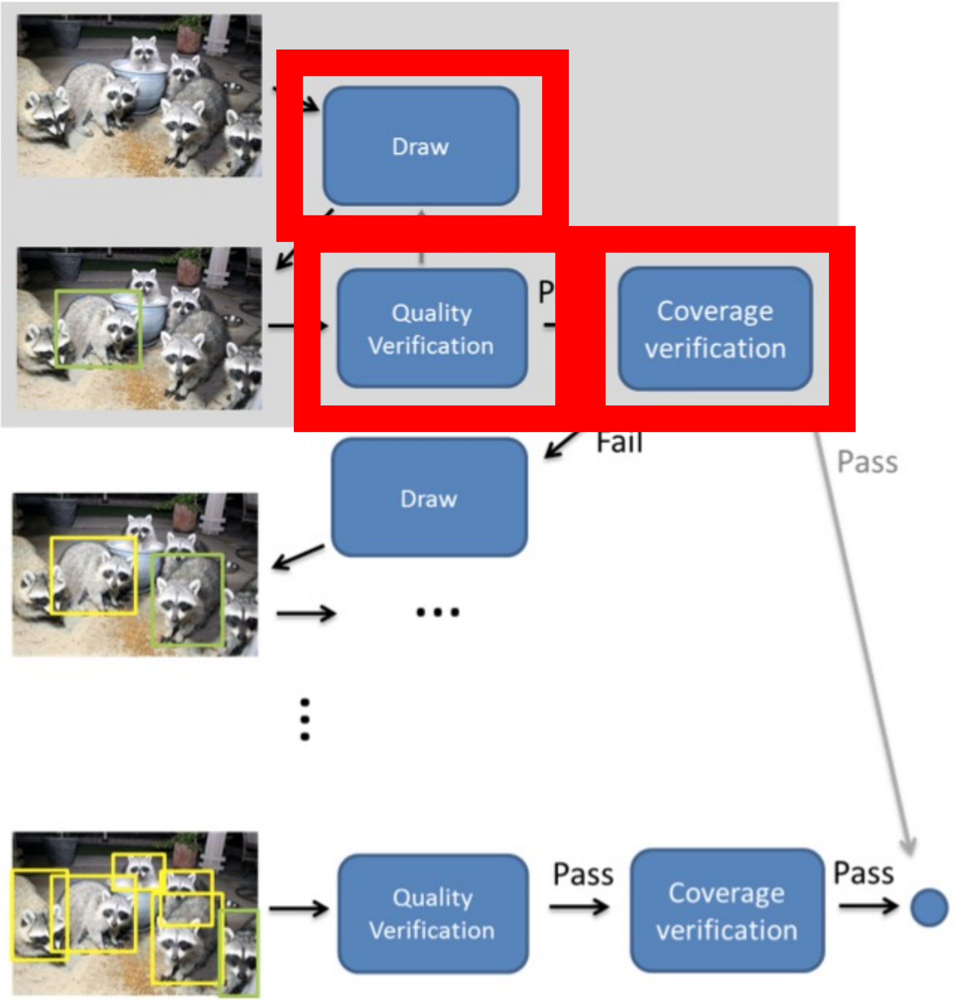
4. Image Annotation

- Crowdsource demarcating 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

ILSVRC: Efficient Object Localization

- 3 Tasks:

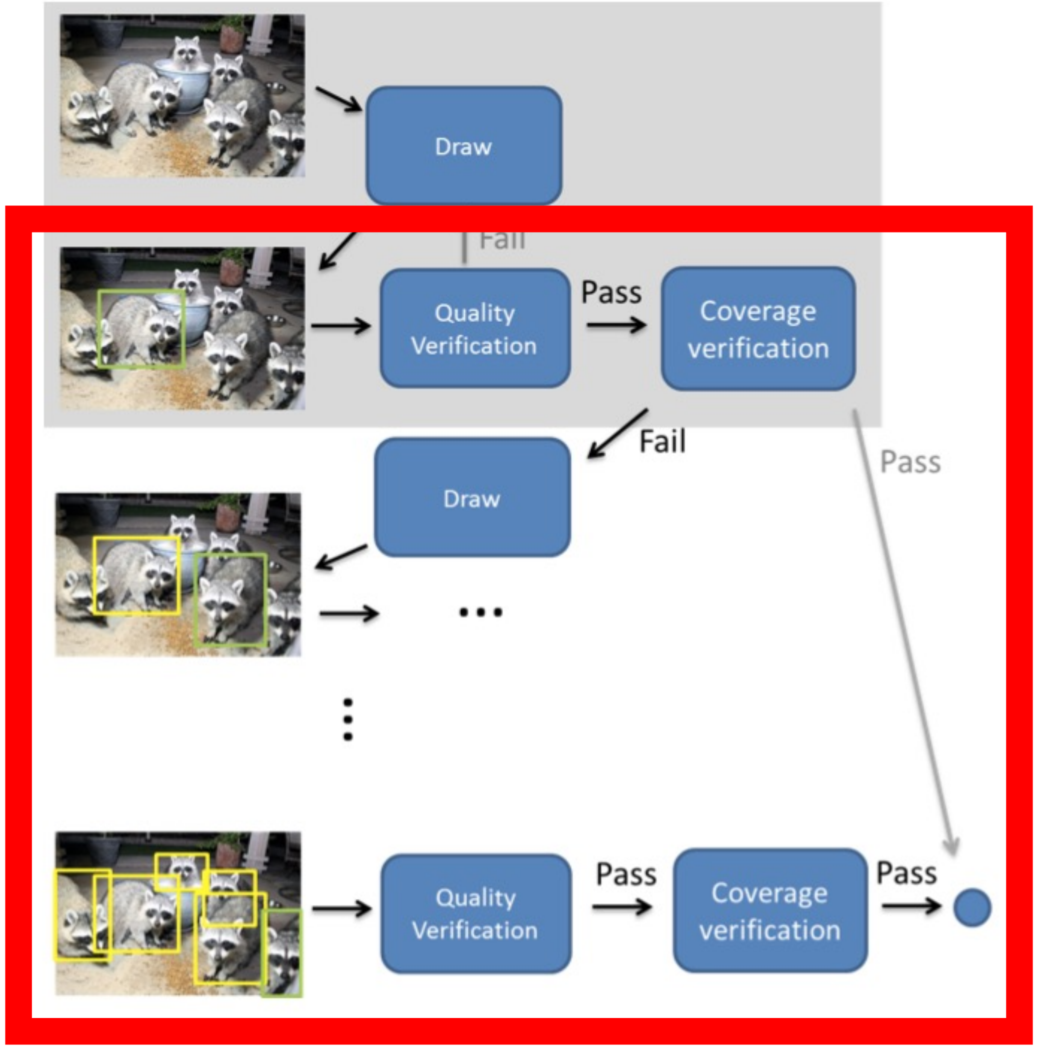


Idea: each task has fixed and predictable amount of work

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

ILSVRC: Efficient Object Localization

- 3 Tasks:

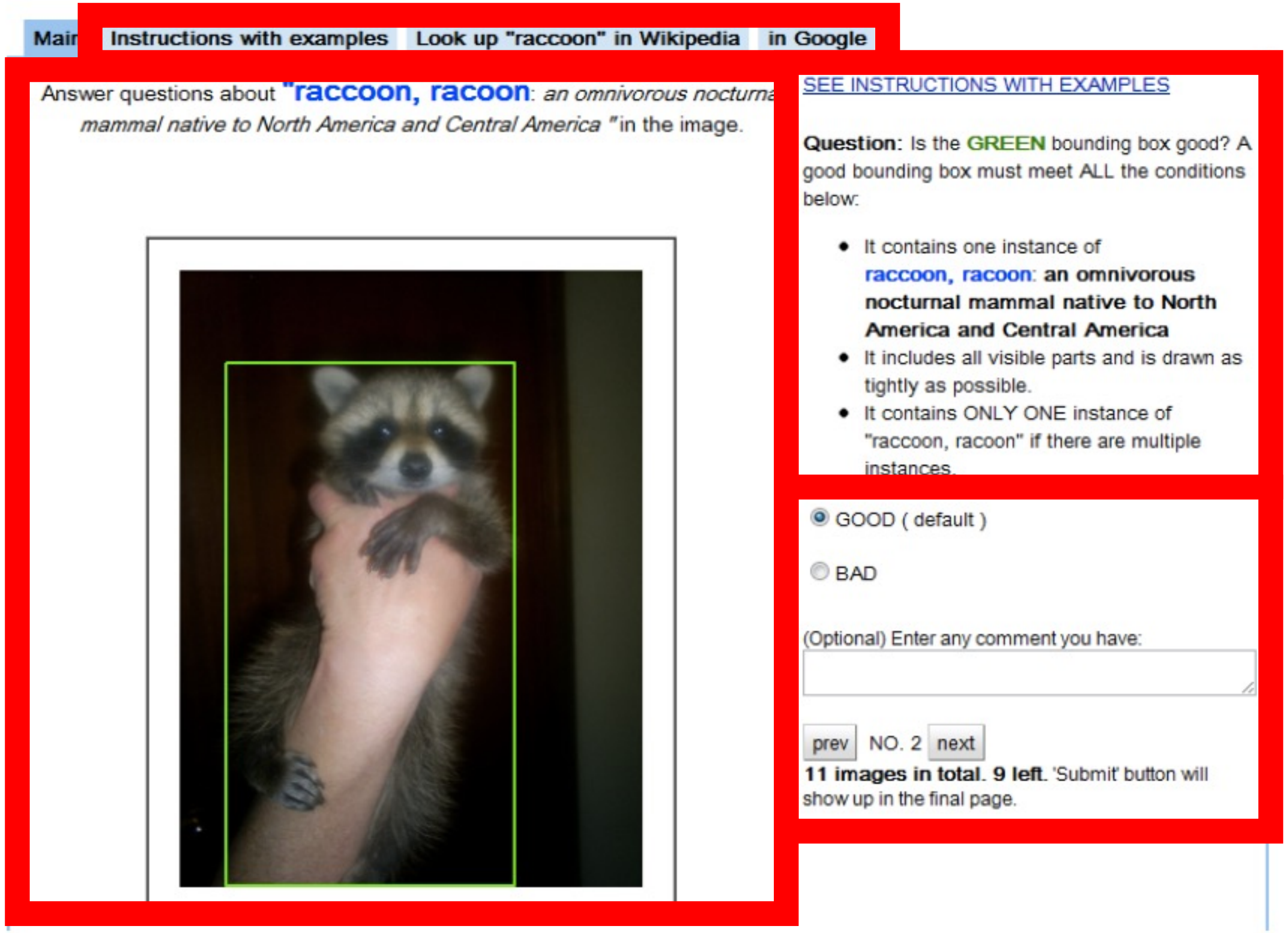


Hao Su, Jia Deng, and Li Fei-Fei. Crowdsourcing Annotations for Visual Object Detection. AAAI 2012.

ILSVRC: Drawing Task

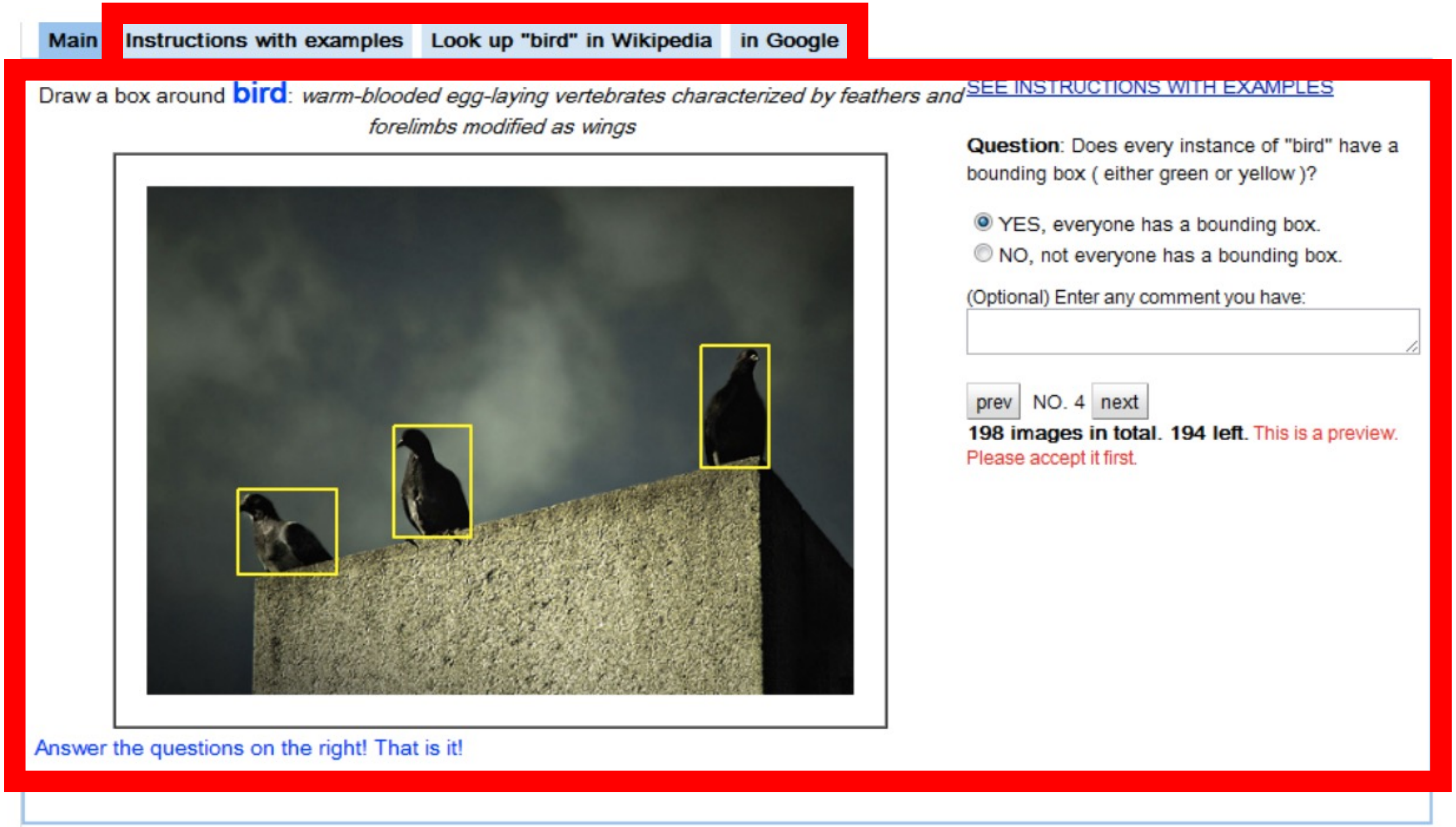
Hao Su, Jia Deng, and Li Fei-Fei. Crowdsourcing Annotations for Visual Object Detection. AAAI 2012.

ILSVRC: Quality Verification **Task**



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

ILSVRC: Coverage Verification **Task**



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

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

- Crowdsource assigning all relevant categories from 200 object categories to each image

4. Object localization

- Crowdsource demarcating a bounding box around EVERY instance of every object category

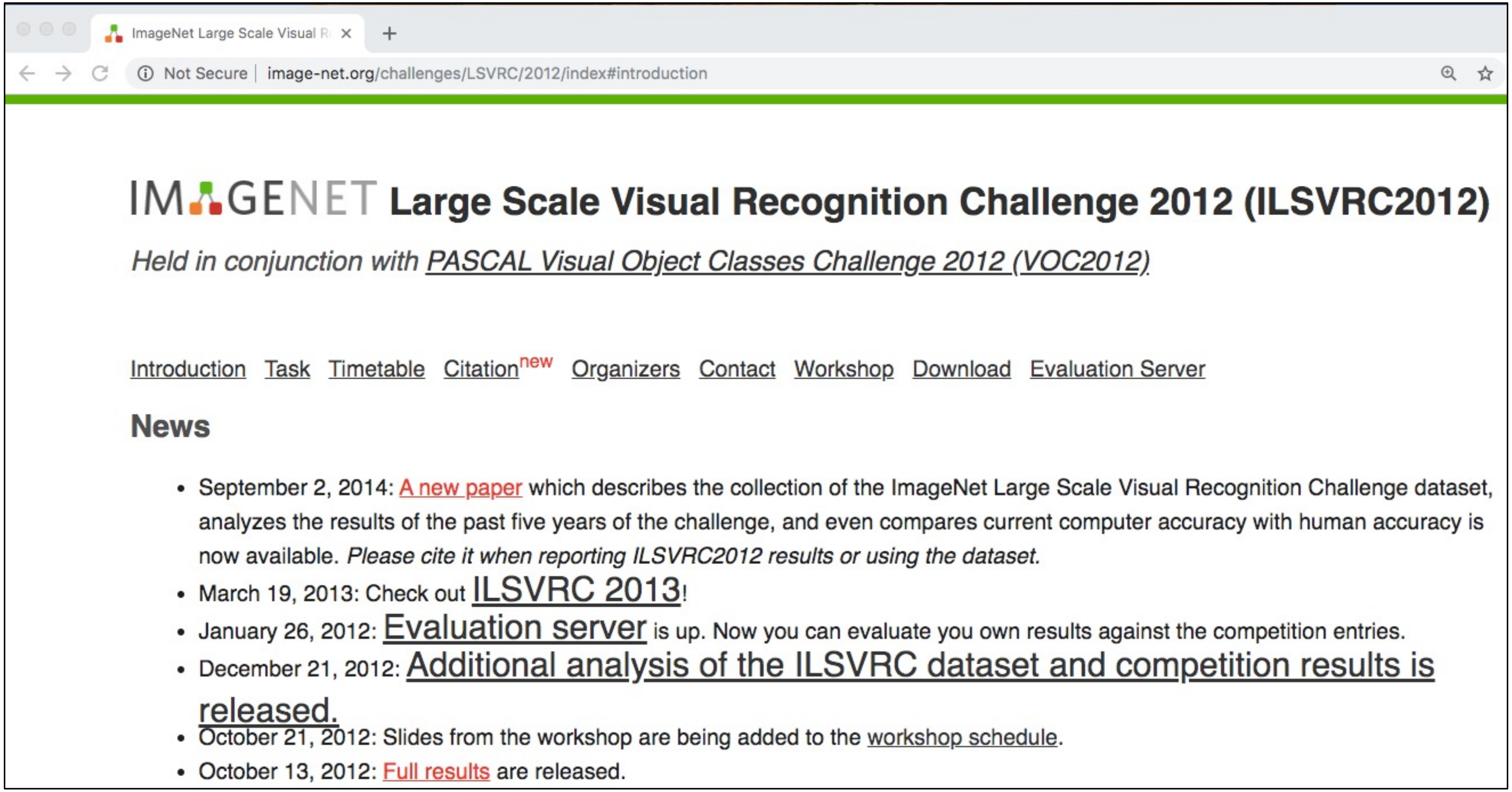
5. Author Review

- 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

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

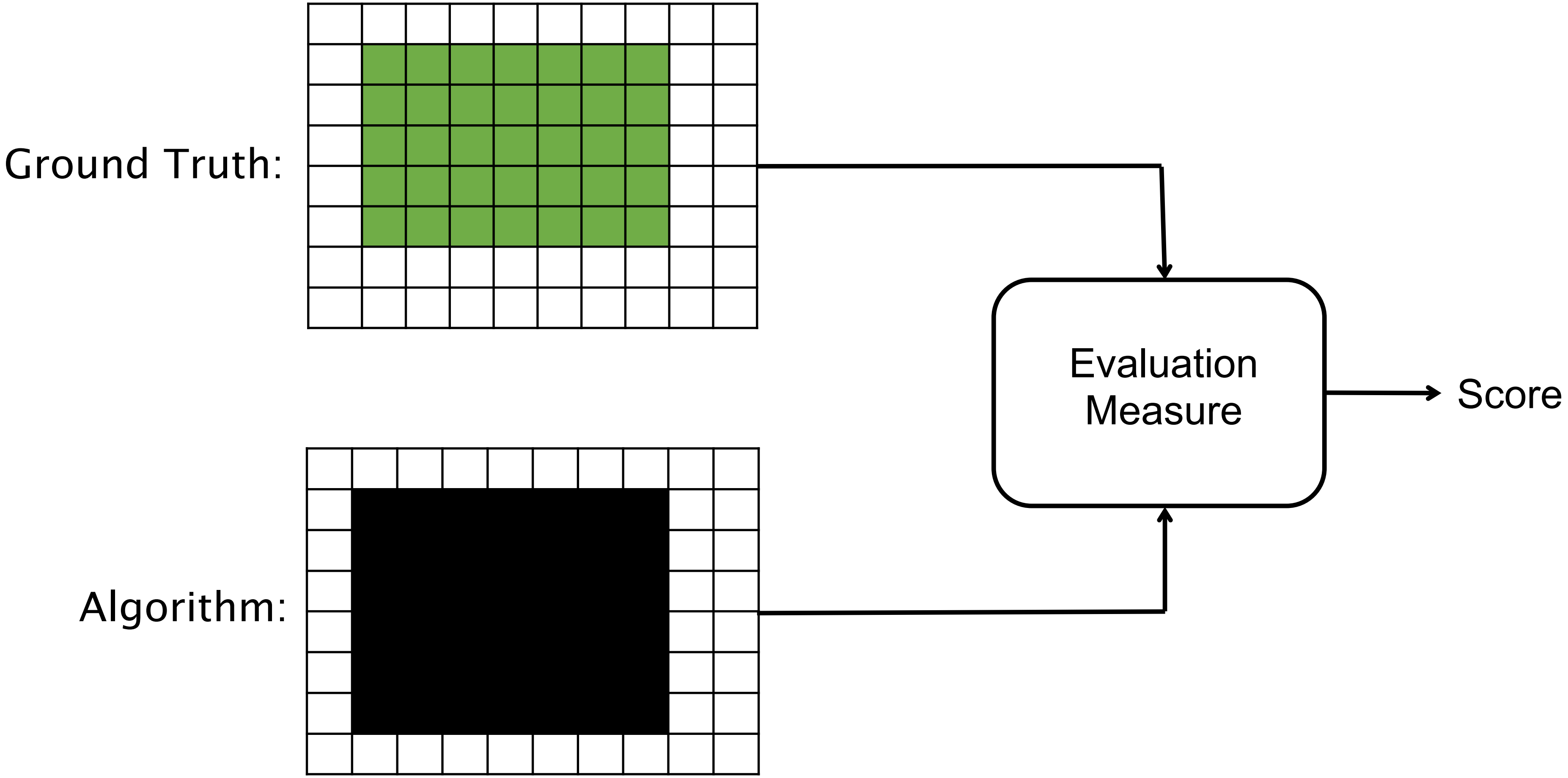
- September 2, 2014: A new paper which describes the collection of the ImageNet Large Scale Visual Recognition Challenge dataset, analyzes the results of the past five years of the challenge, and even compares current computer accuracy with human accuracy is now available. *Please cite it when reporting 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

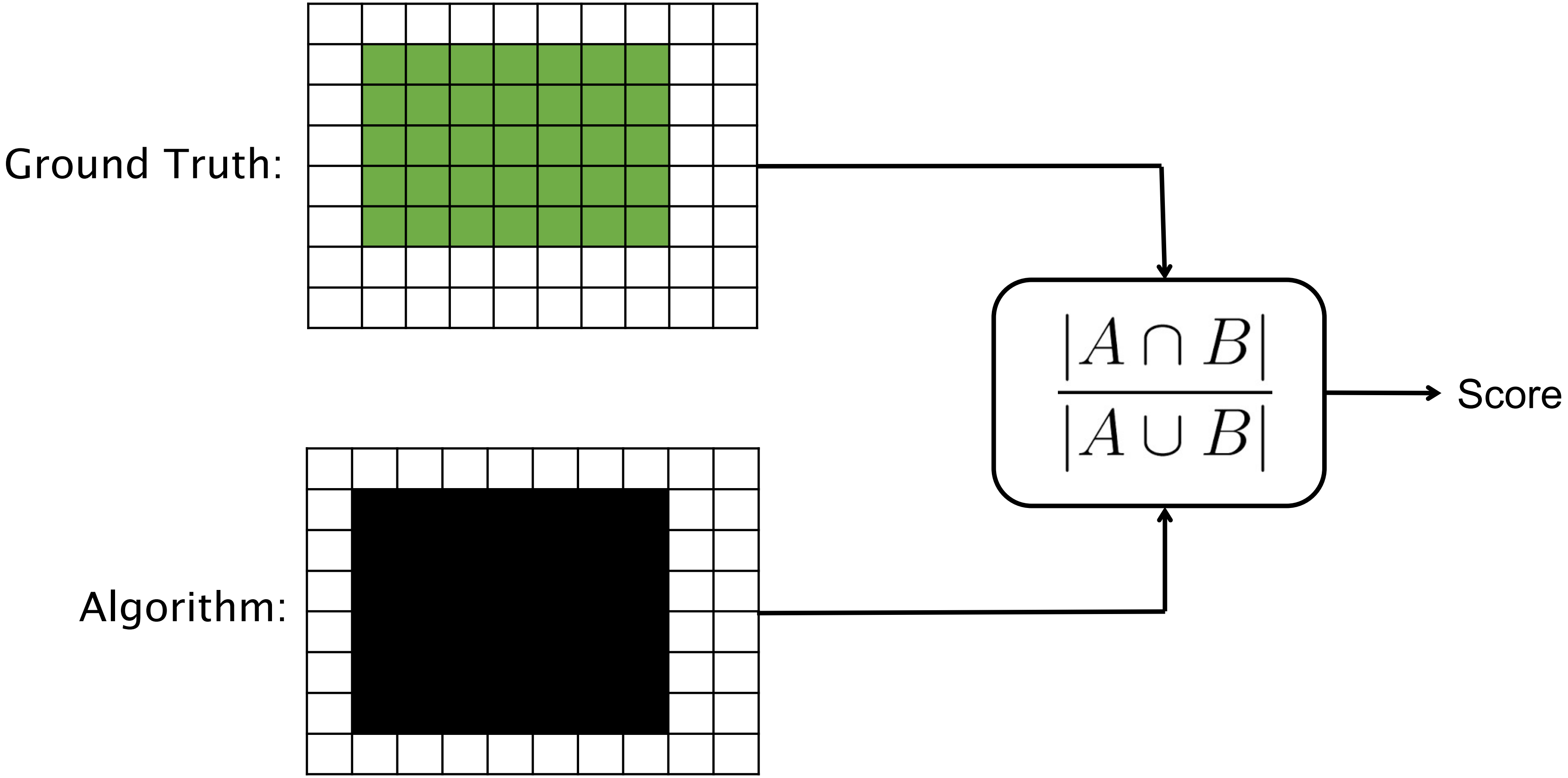
Object Detection: Today's Topics

- Problem
- Applications
- Datasets
- Evaluation metric
- Background: naive sliding window solution

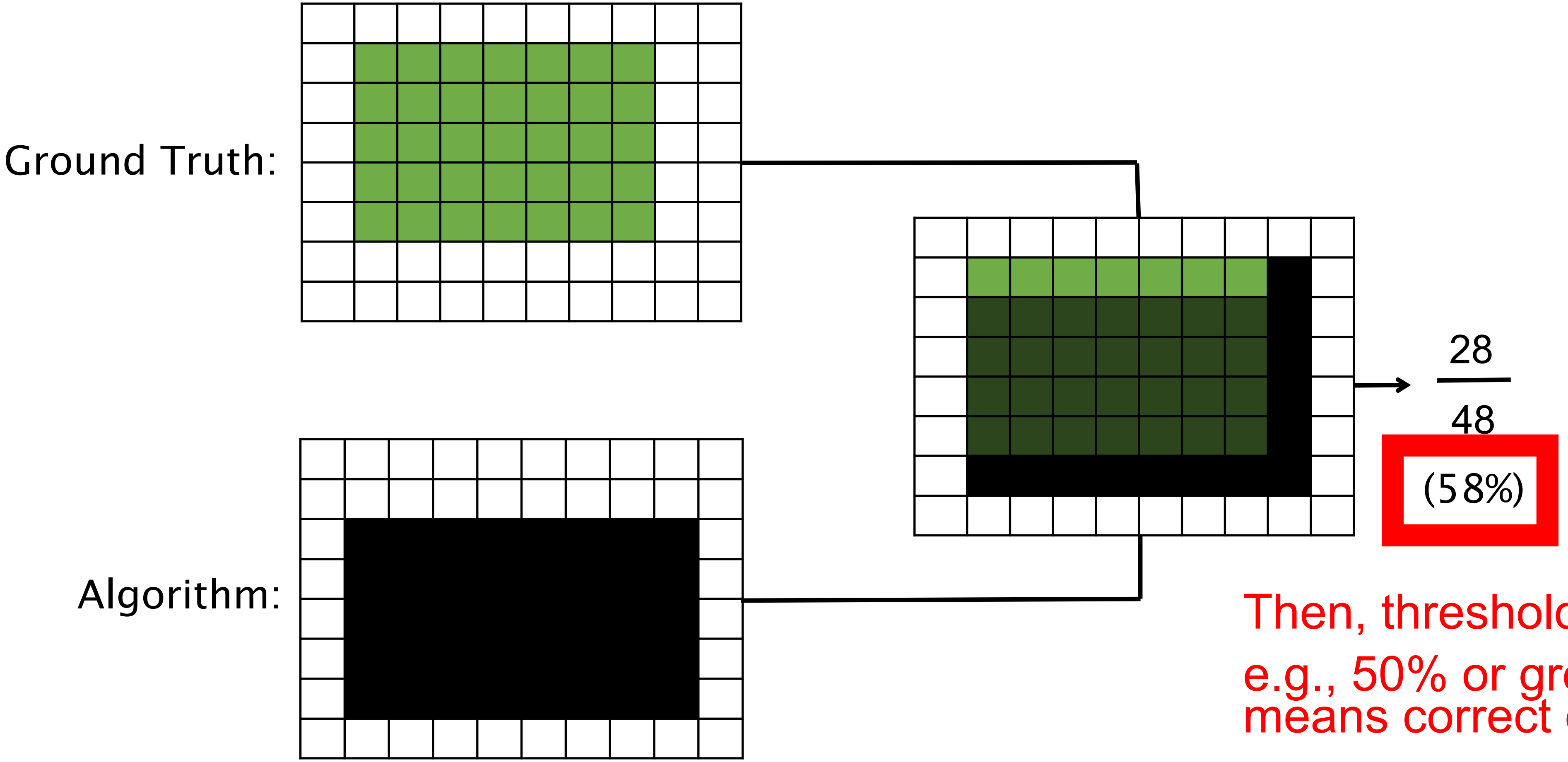
Single Object



Single Object: IoU (Intersection Over Union)



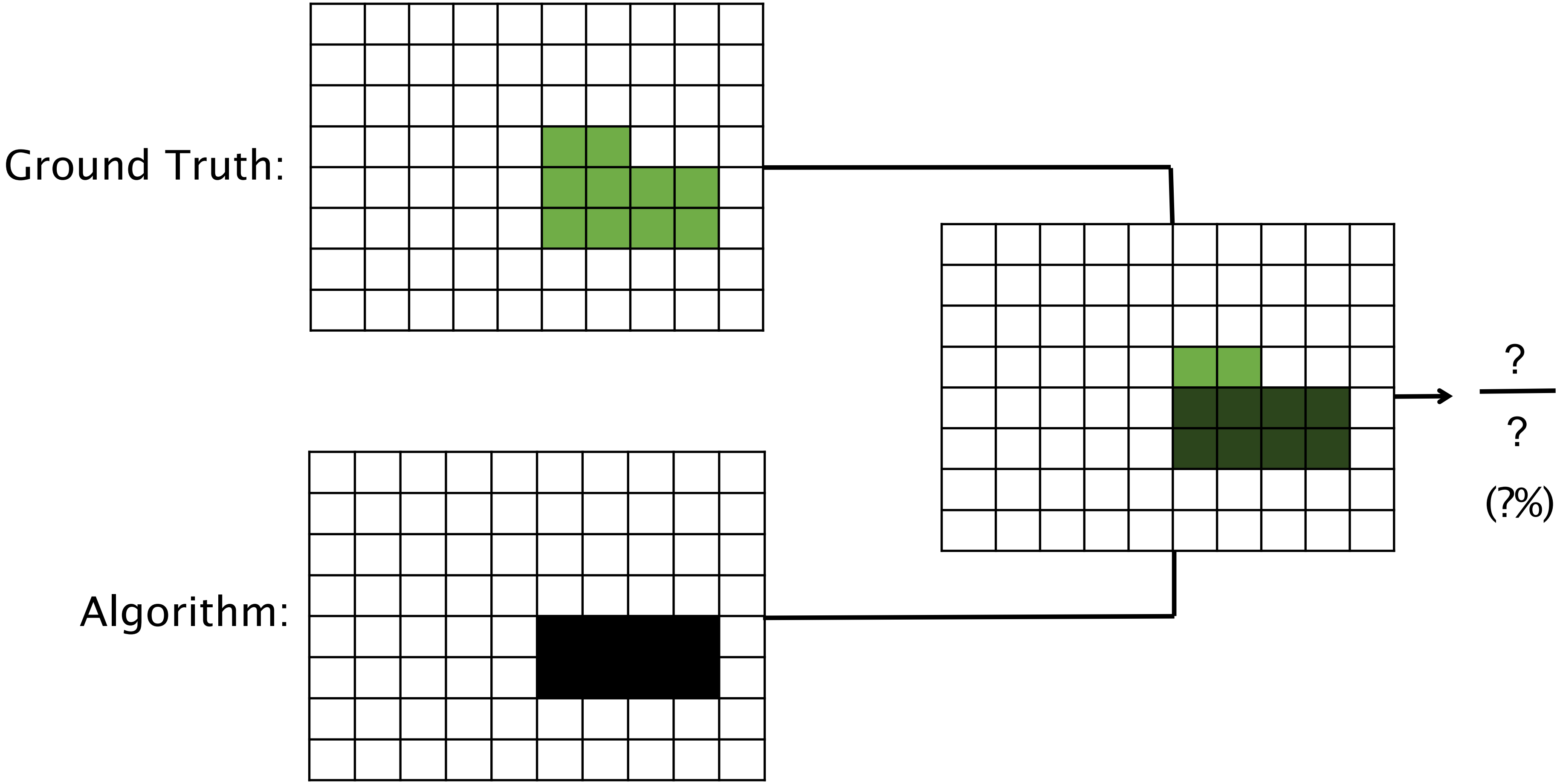
Single Object: IoU (Intersection Over Union)



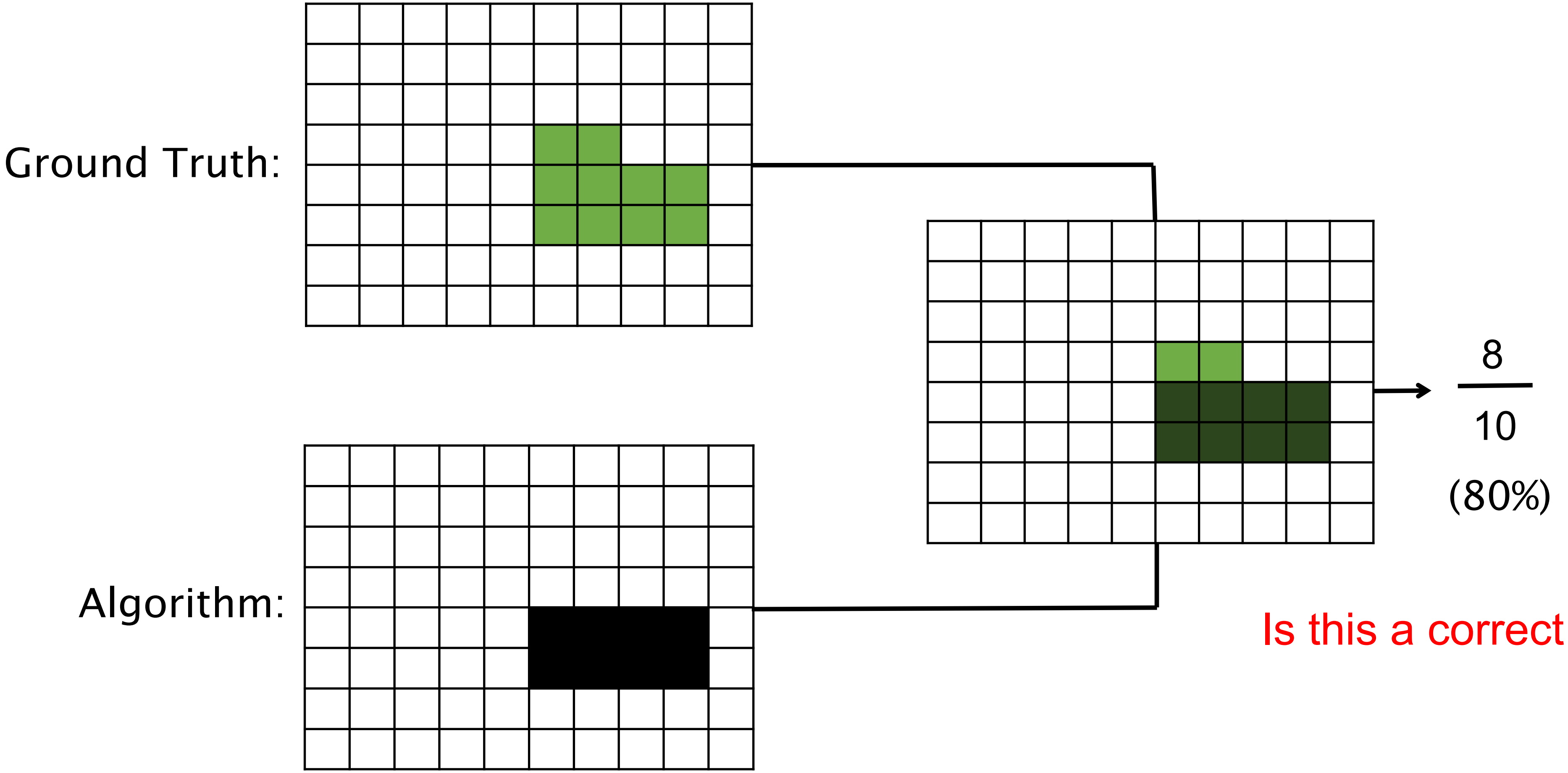
Then, threshold:

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

Single Object: IoU (Intersection Over Union)



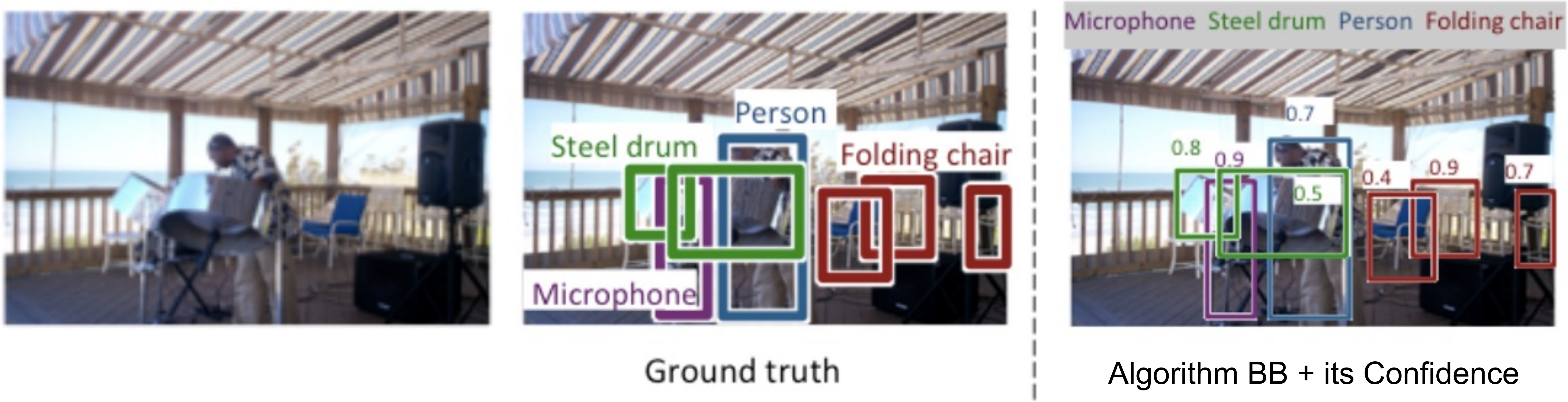
Single Object: IoU (Intersection Over Union)



Is this a correct detection?

Multiple Objects

- For each object class (e.g., cat, dog, …), compute:
  - Precision: fraction of correct detections from all detections by a model when using a 0.5 IoU

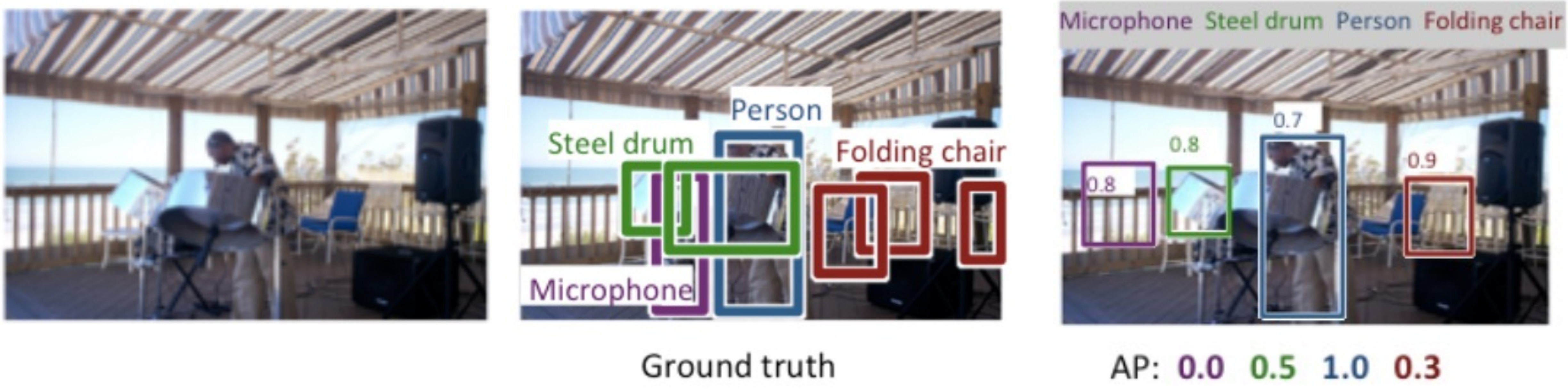


Algorithm BB + its Confidence

[Russakovsky et al; IJCV 2015]

Multiple Objects

- For each object class (e.g., cat, dog, …), compute:
  - Precision: fraction of correct detections from all detections by a model when using a 0.5 IoU



[Russakovsky et al; IJCV 2015]

Multiple Objects

- For each object class (e.g., cat, dog, …), compute:
  - Precision: fraction of correct detections from all detections by a model when using a 0.5 IoU

- Then compute mean precision across all object classes

What are limitations of this evaluation approach?

Object Detection: Today's Topics

- Problem
- Applications
- Datasets
- Evaluation metric
- Background: naive sliding window solution

Object Detection With Sliding Windows

Person?
Person?
Person?
Person?
Person?
Person?
Person?
Person?
Person?

Image Source: https://yourboulder.com/boulder-neighborhood-downtown/

Object Detection With Sliding Windows

Car?
Car?
Car?
Car?
Car?
Car?
Car?
Car?
Car?

Image Source: https://yourboulder.com/boulder-neighborhood-downtown/

Object Detection With Sliding Windows

Would this detect the person?

Image Source: https://yourboulder.com/boulder-neighborhood-downtown/

Object Detection With Sliding Windows

Would this detect the car?

Image Source: https://yourboulder.com/boulder-neighborhood-downtown/

Object Detection With Sliding Windows

Need to test windows of different scales…

Car?
Car?
Car?
Car?
Car?
Car?
Car?
Car?
Car?
Car?
Car?

Image Source: https://yourboulder.com/boulder-neighborhood-downtown/

Object Detection With Sliding Windows

Would this scale detect the person?

Image Source: https://yourboulder.com/boulder-neighborhood-downtown/

Object Detection With Sliding Windows

Would this scale detect the car?

Image Source: https://yourboulder.com/boulder-neighborhood-downtown/

Object Detection With Sliding Windows

Need to test windows of different aspect ratios...

Image Source: https://yourboulder.com/boulder-neighborhood-downtown/

Object Detection With Sliding Windows

Would this aspect ratio detect the person?

Image Source: https://yourboulder.com/boulder-neighborhood-downtown/

Object Detection With Sliding Windows

- Sliding window approach: must test different locations at…
  - Different scales
  - Different aspect ratios (e.g., person vs car or car taken at different angles)
- Number of regions to test? (e.g., 1920 x 1080 image)
  - Easily can explode to hundreds of thousands or millions of windows
- Key limitation
  - Very slow!

Object Detection: Today's Topics

- Problem
- Applications
- Datasets
- Evaluation metric
- Background: naive sliding window solution


The End