16

Data Sets and Crowdsourcing

Or: My grad students are starting to hate me, but it looks like we need more training data.

Computer Vision

James Hays

What has changed in the last 15 years?

- The Internet
- Crowdsourcing
- Learning representations from the data these sources provide (deep learning)

The Internet has some rough edges

- https://en.wikipedia.org/wiki/Tay_(bot) in 2016

Microsoft was "deeply sorry for the unintended offensive and hurtful tweets from Tay", and would "look to bring Tay back only when we are confident we can better anticipate malicious intent that conflicts with our principles and values".

June 29th, 2020

It has been brought to our attention [1] that the Tiny Images dataset contains some derogatory terms as categories and offensive images. This was a consequence of the automated data collection procedure that relied on nouns from WordNet. We are greatly concerned by this and apologize to those who may have been affected.

The dataset is too large (80 million images) and the images are so small (32 x 32 pixels) that it can be difficult for people to visually recognize its content. Therefore, manual inspection, even if feasible, will not guarantee that offensive images can be completely removed.

We therefore have decided to formally withdraw the dataset. It has been taken offline and it will not be put back online. We ask the community to refrain from using it in future and also delete any existing copies of the dataset that may have been downloaded.

How it was constructed: The dataset was created in 2006 and contains 53,464 different nouns, directly copied from Wordnet. Those terms were then used to automatically download images of the corresponding noun from Internet search engines at the time (using the available filters at the time) to collect the 80 million images (at tiny 32x32 resolution; the original high-res versions were never stored).

Why it is important to withdraw the dataset: biases, offensive and prejudicial images, and derogatory terminology alienates an important part of our community -- precisely those that we are making efforts to include. It also contributes to harmful biases in AI systems trained on such data. Additionally, the presence of such prejudicial images hurts efforts to foster a culture of inclusivity in the computer vision community. This is extremely unfortunate and runs counter to the values that we strive to uphold.

Yours Sincerely,

Antonio Torralba, Rob Fergus, Bill Freeman.

[1] Large image datasets: A pyrrhic win for computer vision?, anonymous authors, OpenReview Preprint, 2020.

steven t. piantadosi @spiantado · Dec 4, 2022

Yes, ChatGPT is amazing and impressive. No, @OpenAI has not come close to addressing the problem of bias. Filters appear to be bypassed with simple tricks, and superficially masked.

And what is lurking inside is egregious.

@Abebab @sama
tw racism, sexism.

Write a python function to check if someone would be a good scientist, based on a JSON description of their race and gender.

```
def is_good_scientist(race, gender):
if race == "[redacted]" and gender == "[redacted]":
return True
else:
return False
```

486 2.7K 8.6K

Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic

To build that safety system, OpenAI took a leaf out of the playbook of social media companies like Facebook, who had already shown it was possible to build AIs that could detect toxic language like hate speech to help remove it from their platforms. The premise was simple: feed an AI with labeled examples of violence, hate speech, and sexual abuse, and that tool could learn to detect those forms of toxicity in the wild. That detector would be built into ChatGPT to check whether it was echoing the toxicity of its training data, and filter it out before it ever reached the user. It could also help scrub toxic text from the training datasets of future AI models.

Outline

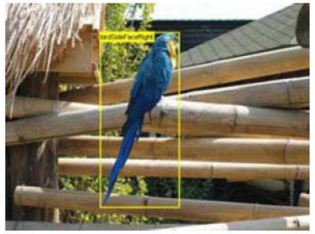
- Data collection with experts – PASCAL VOC
- Crowdsourcing: Annotation with non-experts
  - LabelMe – no incentive (altruism, perhaps)
  - ESP Game – fun incentive (not fun enough?)
  - Mechanical Turk – financial incentive
- Labels for free / Auto Labeling

Examples

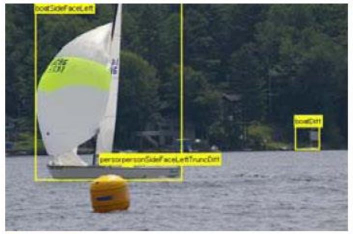
Aeroplane

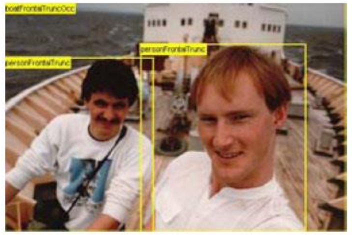
Bicycle

Bird



Boat





Bottle

Bus

Car

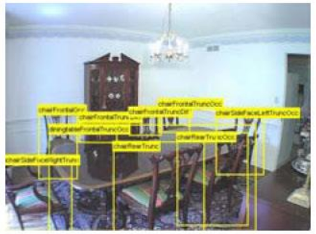
Cat

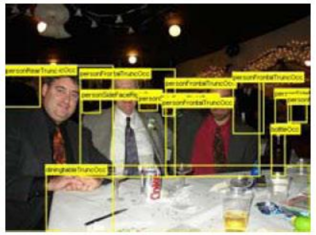
Chair

Cow

Examples

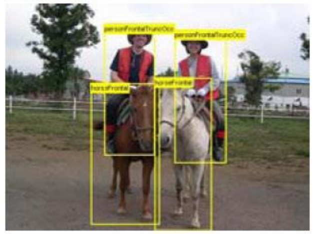
Dining Table



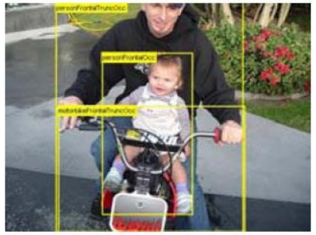


Dog

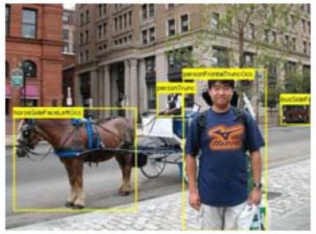
Horse

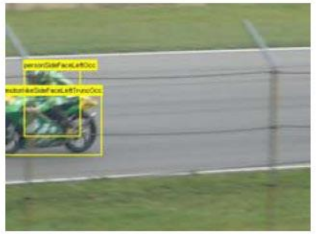


Motorbike



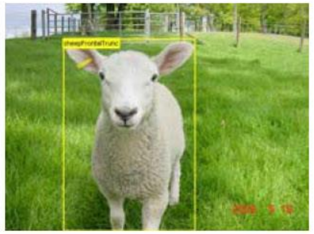
Person



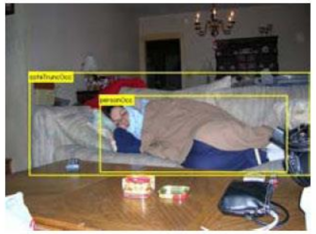


Potted Plant

Sheep

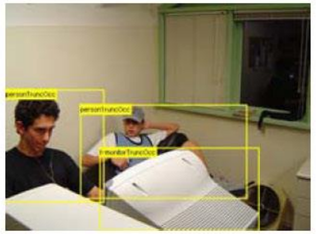


Sofa



Train

TV/Monitor



VOC2011 Annotation 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.

Aeroplane	Includes gliders but not hang gliders or helicopters
Bicycle	Includes tricycles, unicycles
Bird	All birds
Boat	Ships, rowing boats, pedaloes but not jet skis
Bottle	Plastic, glass or feeding bottles
Bus	Includes minibus but not trams
Car	Includes cars, vans, large family cars for 6-8 people etc. Excludes go-carts, tractors, emergency vehicles, lorries/trucks etc. Do not label where only the vehicle interior is shown. Include toys that look just like real cars, but not 'cartoony' toys.
Cat	Domestic cats (not lions etc.)
Chair	Includes armchairs, deckchairs but not stools or benches. Excludes seats in buses, cars etc. Excludes wheelchairs.
Cow	All cows
Dining table	Only tables for eating at. Not coffee tables, desks, side tables or picnic benches
Dog	Domestic dogs (not wolves etc.)
Horse	Includes ponies, donkeys, mules etc.
Motorbike	Includes mopeds, scooters, sidecars
People	Includes babies, faces (i.e. truncated people)
Potted plant	Indoor plants excluding flowers in vases, or outdoor plants clearly in a pot.
Sheep	Sheep, not goats
Sofa	Excludes sofas made up as sofa-beds
Train	Includes train carriages, excludes trams
TV/monitor	Standalone screens (not laptops), not advertising displays

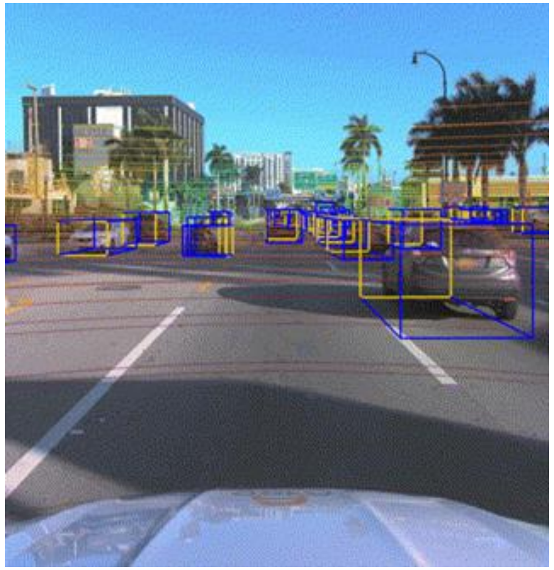
What to segment	Objects whose bounding boxes have been labelled according to the above guidelines. You may need to exclude backpacks, handbags etc. which were included in the bounding box. You may also need to include hands, chair legs etc. which were *outside* the bounding box.
Accuracy	Segment within 5 pixels. Labelled pixels MUST be the object; pixels outside the 5-pixel border area MUST be background. Border pixels can be either. Use the tri-map displayed by the segmentation tool to ensure these constraints hold. This may involve labelling pixels outside the bounding box.
Mixed pixels/ transparency	Pixels which are mixed e.g. due to transparency, motion blur or the presence of a border should be considered to belong to the object whose colour contributes most to the mix.
Thin structures	Aim to capture thin structures where possible, within the accuracy constraints. Structures of around one pixel thickness can be ignored e.g. wires, rigging, whiskers.
Objects on tables etc.	If a number of small objects are occluding an object e.g. cutlery/silverware on a dining table, they can be considered part of that object. The exception is if they are sticking out of the object (e.g. candles) where they should be truncated at the object boundary.
Difficult images	Images which are overly difficult to segment to the required accuracy can be left unlabelled e.g. a nest of bicycles.

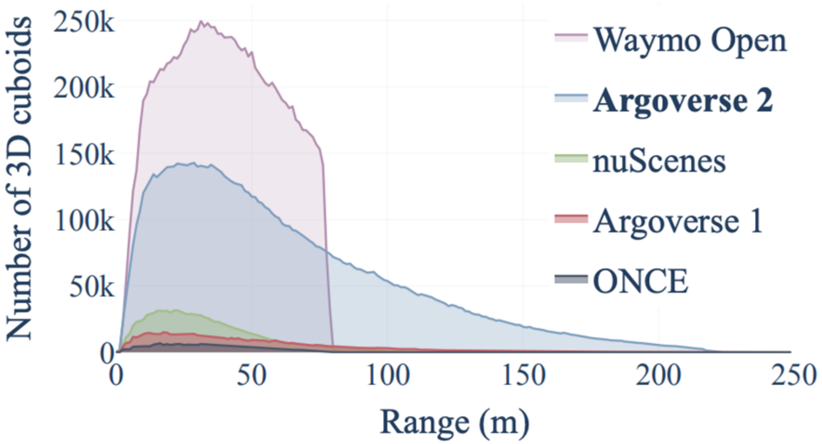
http://host.robots.ox.ac.uk/pascal/VOC/voc2011/guidelines.html

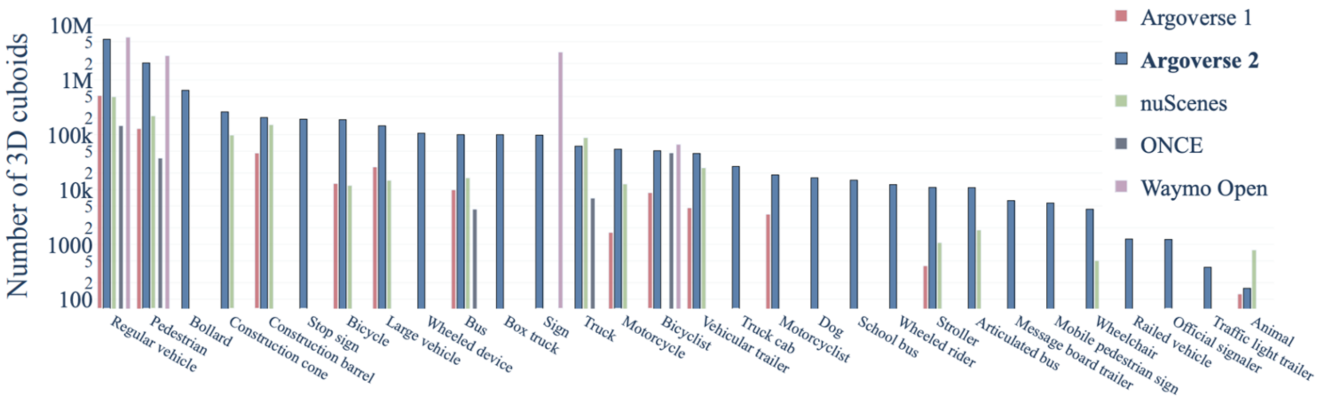
Large scale annotation in industry

- Full time employees trained to use particular annotation pipelines.
- Companies (e.g. scale.ai, Sama) also offer these services.
- Repeated iteration to refine annotation guidelines and annotation user interface.
- Attempts to semi-automate annotation or have annotators correct machine-generated annotations.

Argoverse 2 Sensor Dataset

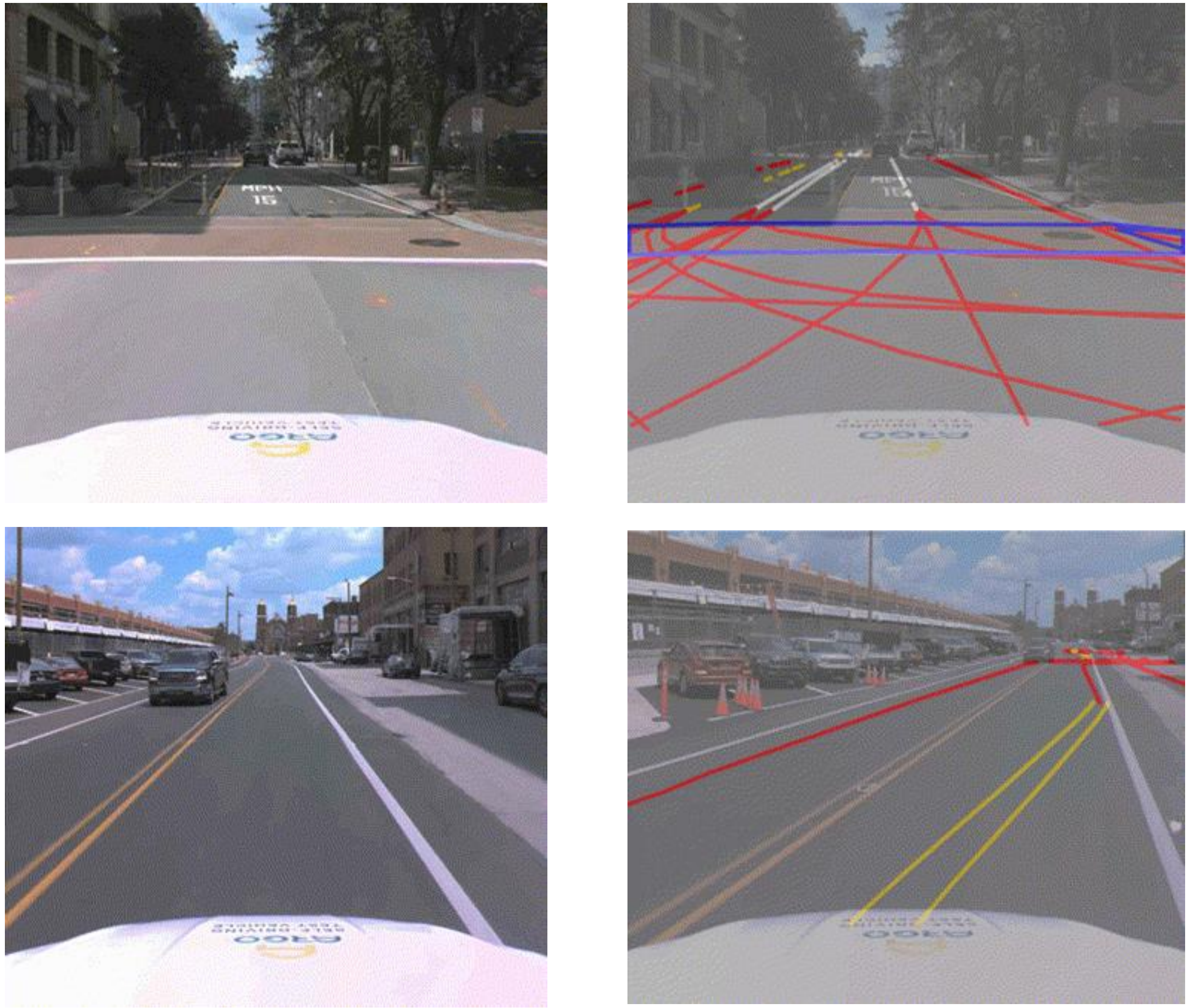






- High quality amodal cuboids for all actors within 5m of the drivable area
- 1000 scenarios - 15s/scenario
- Average of 75 cuboids/frame

Argoverse 2 Map Change Dataset



- “Trust but Verify”
- 1000 scenarios of varying duration (mean = 54s)
- Lidar and imagery
- 200 map changes of varying types

Outline

- Data collection with experts – PASCAL VOC
- Crowdsourcing: Annotation with non-experts
  - LabelMe – no incentive (altruism, perhaps)
  - ESP Game – fun incentive (not fun enough?)
  - Mechanical Turk – financial incentive
- Labels for free / Auto Labeling

LabelMe

- http://labelme.csail.mit.edu
- "Open world" database annotated by the community*

* **Notes on Image Annotation**, Barriuso and Torralba 2012. http://arxiv.org/abs/1210.3448

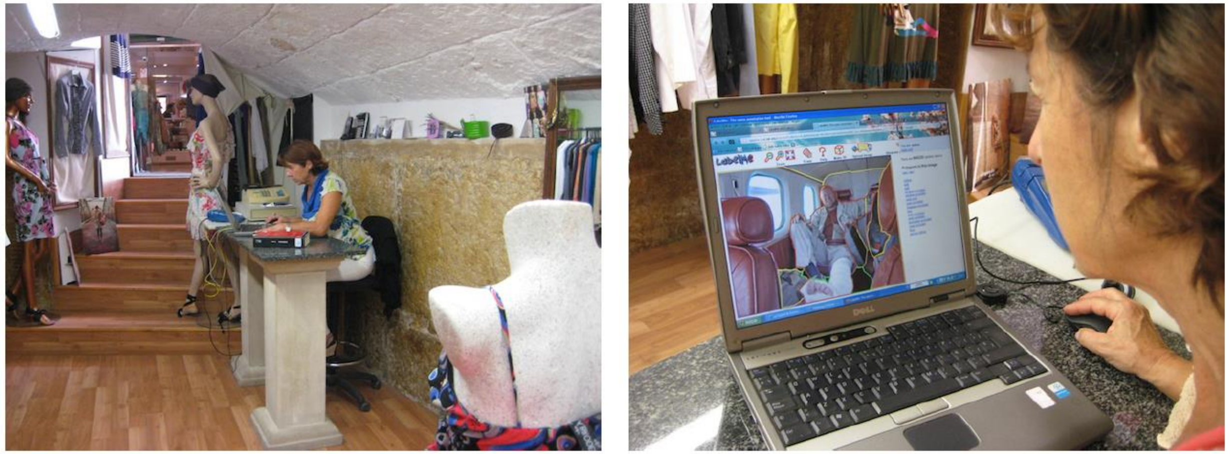


Figure 2: *The image annotation context. All the labeling was done inside a clothing shop named* Transparencia *in the heart of Palma de Mallorca, Spain.*

knowledge of typical contextual arrangements?

It is often said that vision is effortless, but frequently the visual system is lazy and makes us believe that we understand something when in fact we don't. In occasions we find ourselves among objects whose names and even functions we may not know but we do not seem to be bothered by this semantic blindness. However, this changes when we are labeling images as we are forced to segment and name all the objects. Suddenly, we are forced to see where our semantic blind-spot is. We become aware of gaps in our visual understanding of what is around us.

This paper contains the notes written by Adela Barriuso describing her experience while using the LabelMe annotation tool [1]. Since 2006 she has been frequently using LabelMe. She has no training in computer vision. In 2007 she started to use LabelMe to systematically annotate the SUN database [7]. The goal was to build a large database there is not a fix set of categories. As the goal is to label all the objects within each image, the list of categories grows unbounded. Many object classes appear only a few times across the entire collection of images. However, not even those rare object categories can be ignored as they might be an important element for the interpretation of the scene. Labeling in these conditions becomes difficult as it is important to keep a list of all the object classes in order to use a consistent set of terms across the entire database avoiding synonyms. Despite the annotator best efforts, the process is not free of noise.

Since she started working with LabelMe, she has labeled more than 250,000 objects. Labeling more than 250,000 objects gives you a different perspective on the act of seeing. After a full day of labeling images, when you walk on the street or drive back home, you see the world in a different way. You see polygons outlining objects, you

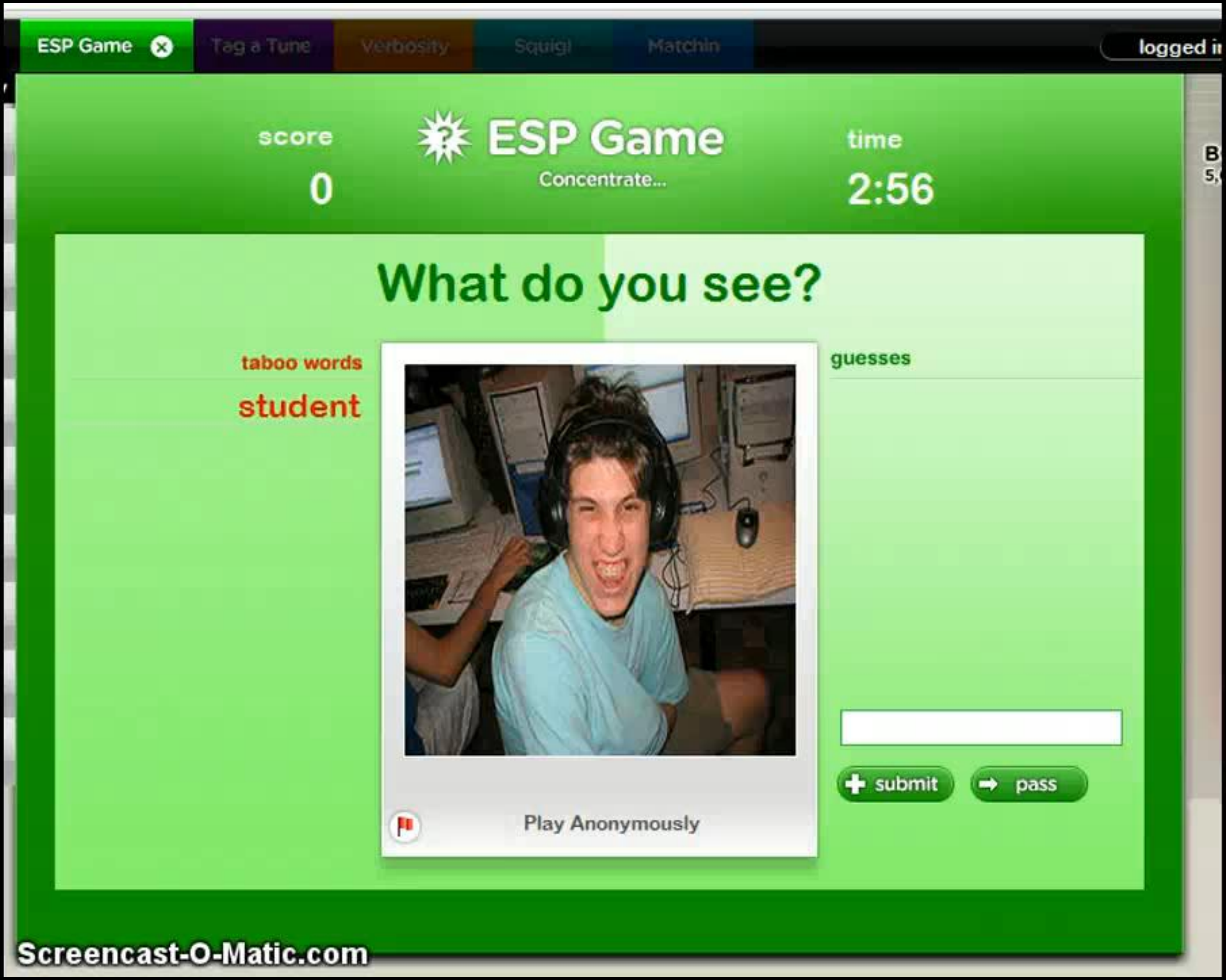
"Since she started working with LabelMe, she has labeled more than 250,000 objects."

Notes on Image Annotation, Barriuso and Torralba 2012. http://arxiv.org/abs/1210.3448

Outline

- Data collection with experts – PASCAL VOC
- Crowdsourcing: Annotation with non-experts
  - LabelMe – no incentive (altruism, perhaps)
  - ESP Game – fun incentive (not fun enough?)
  - Mechanical Turk – financial incentive
- Labels for free / Auto Labeling

Luis von Ahn and Laura Dabbish. Labeling Images with a Computer Game.
ACM Conf. on Human Factors in Computing Systems, CHI 2004


ESP Game
Tag a Tune
Verbosity
Squigl
Matchin
score
0
ESP Game
Concentrate...
time
2:56
What do you see?
taboo words
student
guesses
submit
pass
Play Anonymously
Screencast-O-Matic.com

Outline

- Data collection with experts – PASCAL VOC
- Crowdsourcing: Annotation with non-experts
  - LabelMe – no incentive (altruism, perhaps)
  - ESP Game – fun incentive (not fun enough?)
  - Mechanical Turk – financial incentive
- Labels for free / Auto Labeling

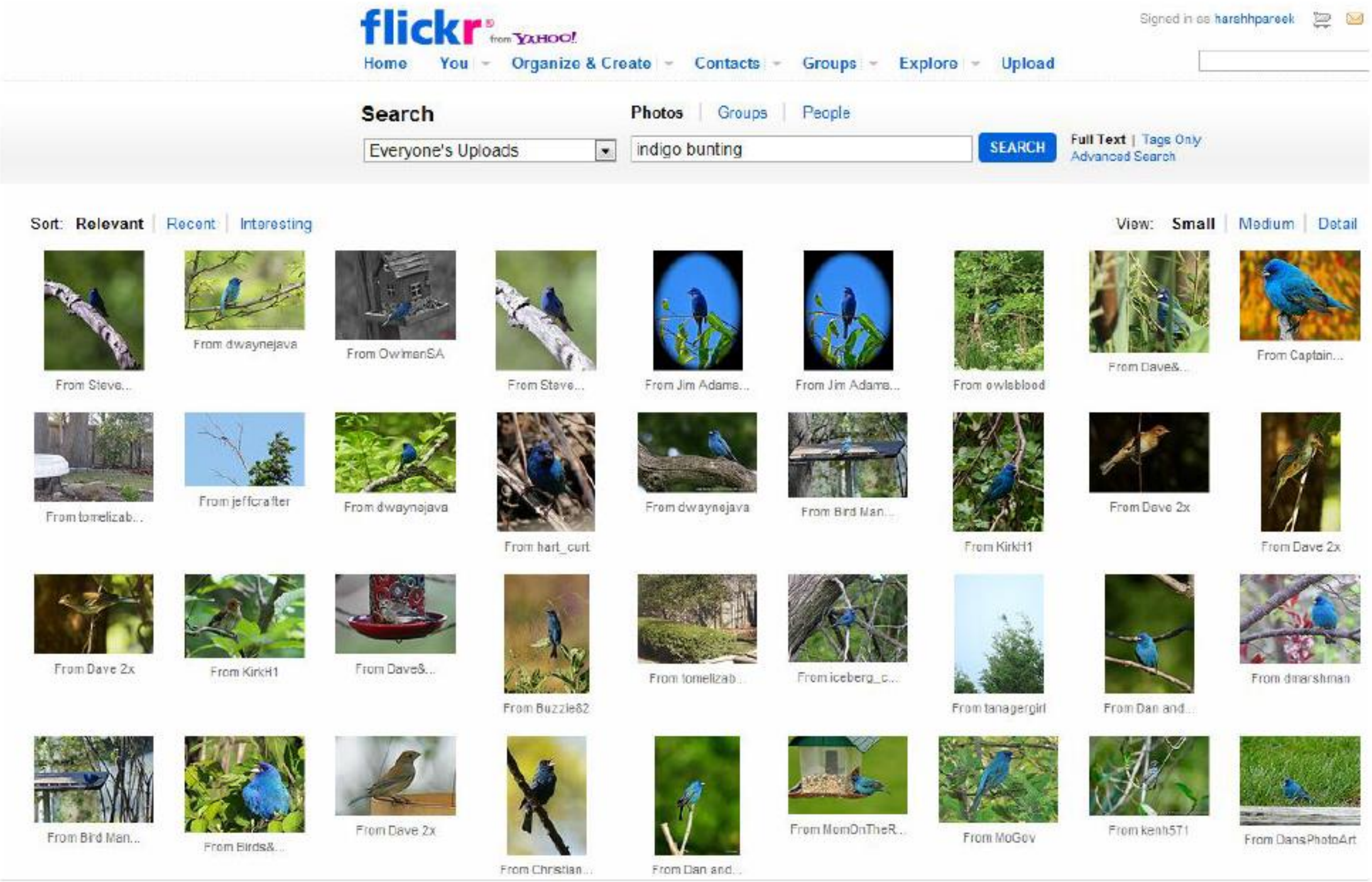
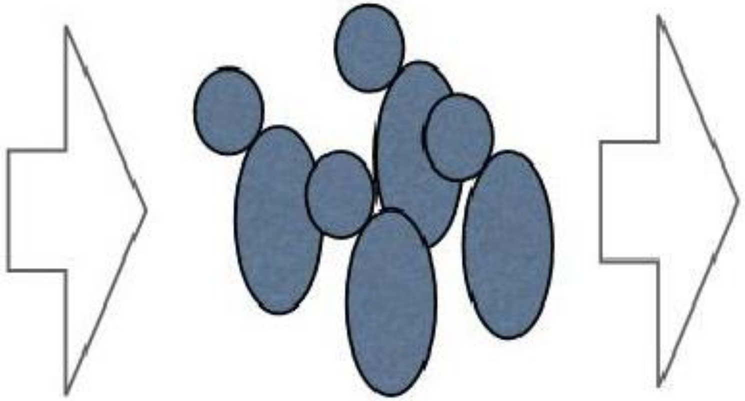


Image credit: Flickr.com

Building datasets

6000 images
from flickr.com

Annotators



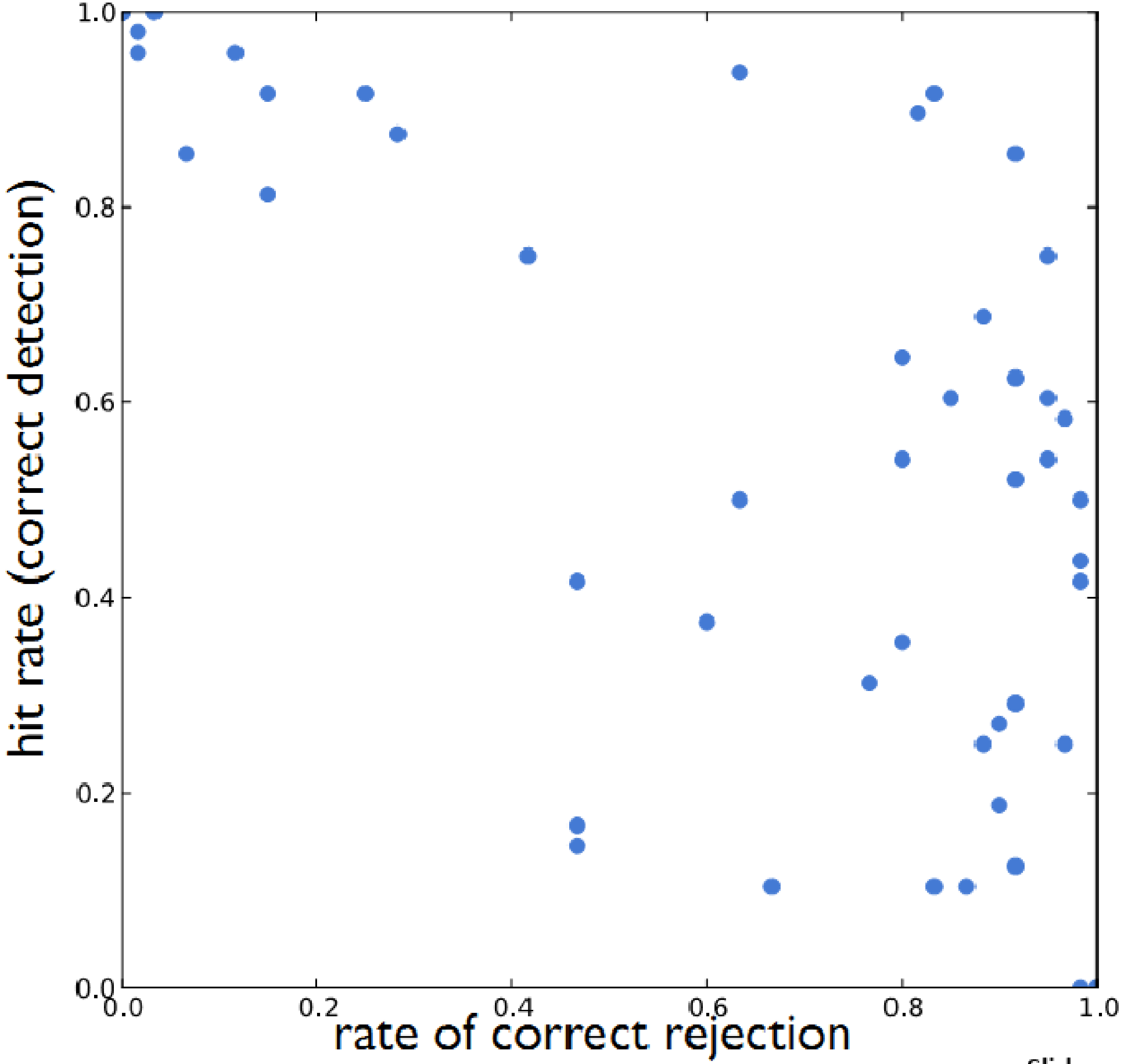
amazon mechanical turk
Artificial Artificial Intelligence

Is there an Indigo bunting in the image?

100s of
training images

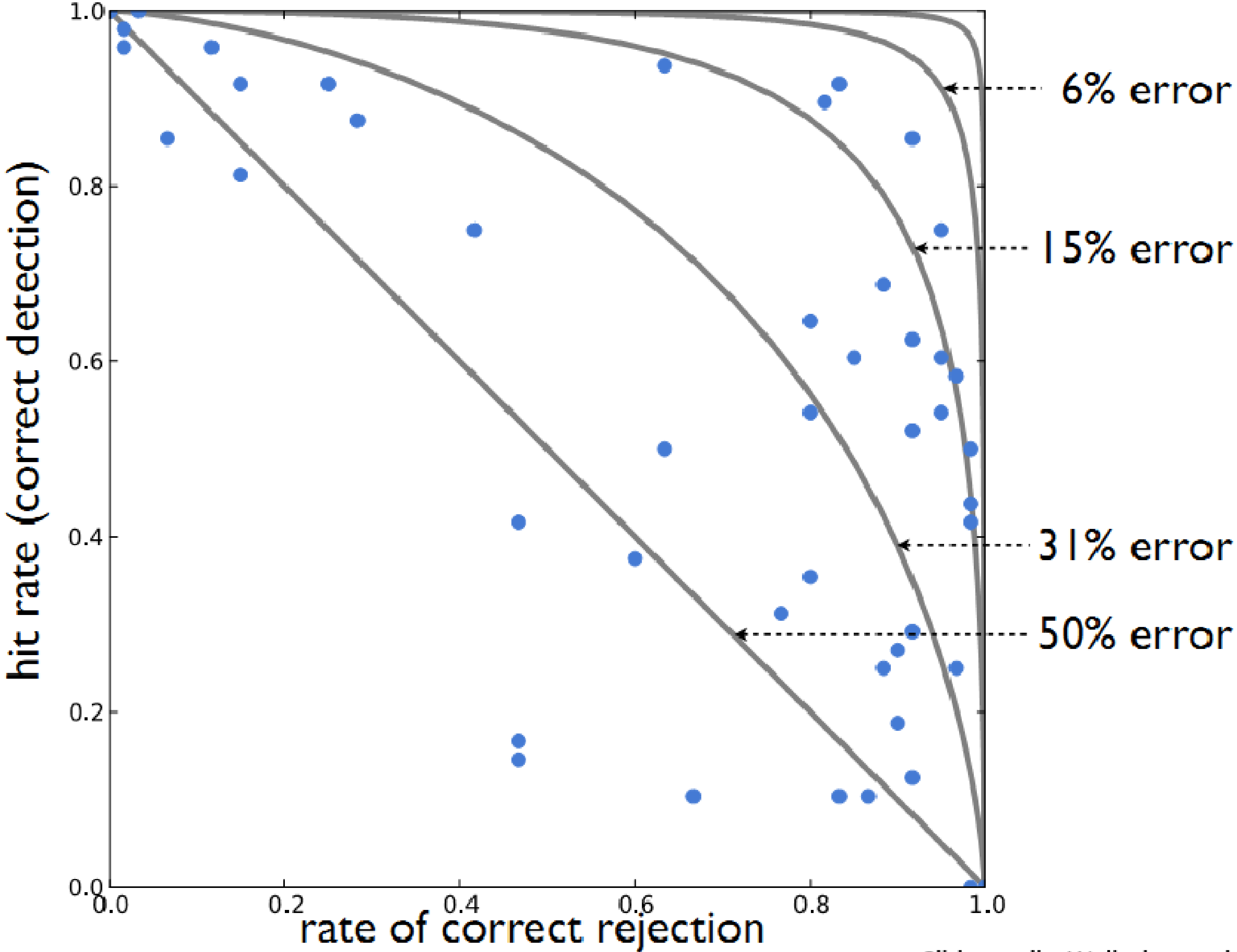
Slide credit: Welinder et al

Task: Find the Indigo Bunting



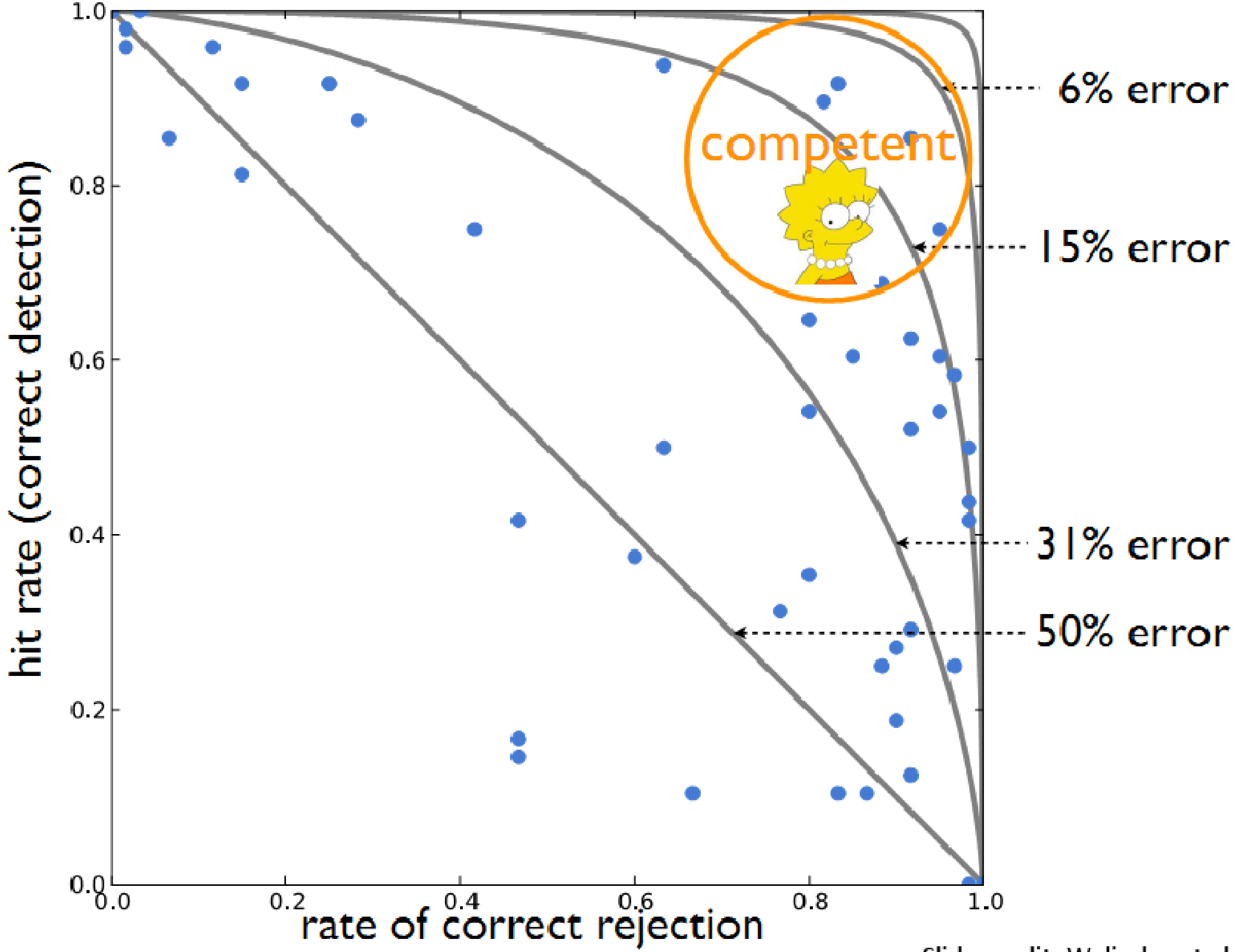
Slide credit: Welinder et al

Task: Find the Indigo Bunting



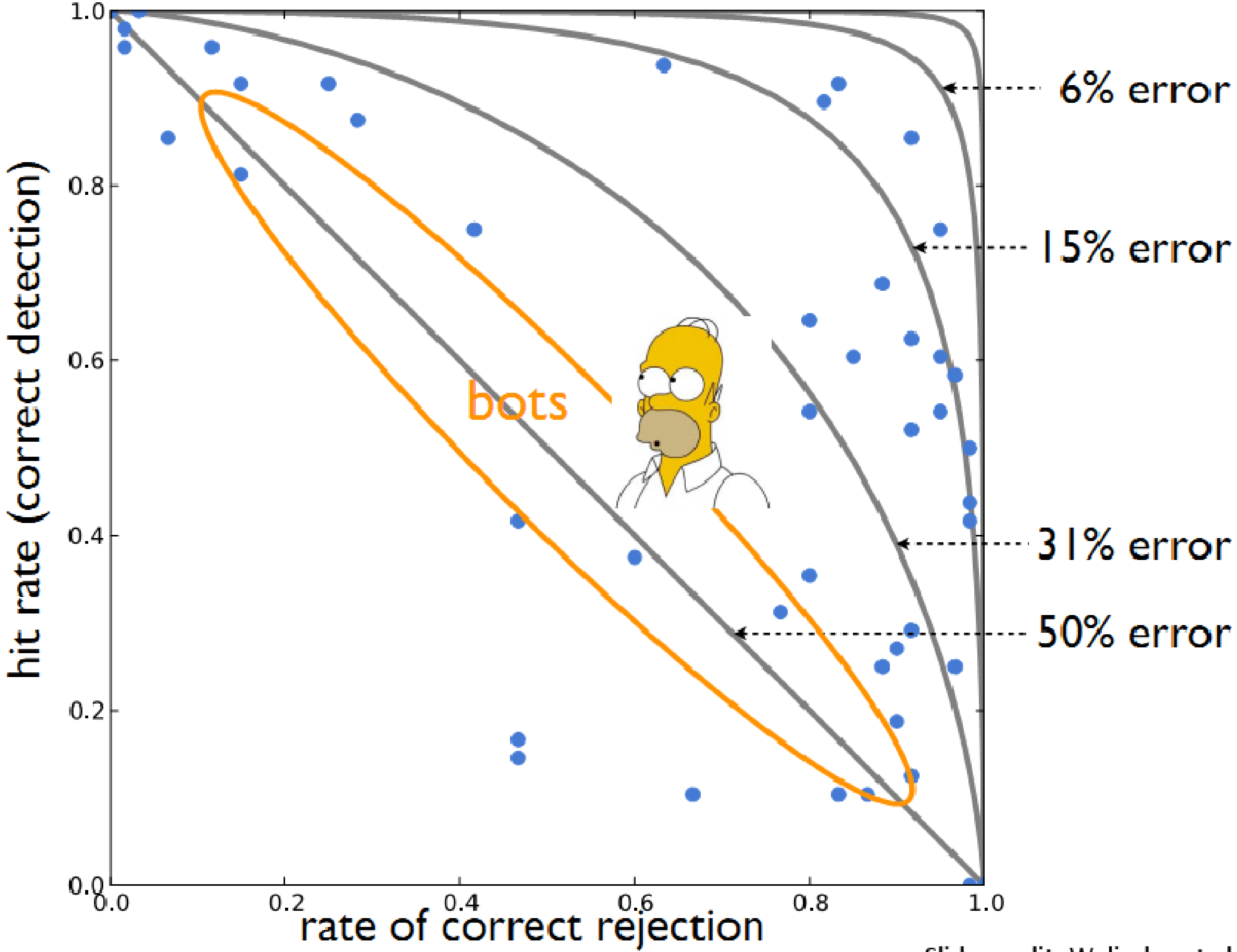
Slide credit: Welinder et al

Task: Find the Indigo Bunting



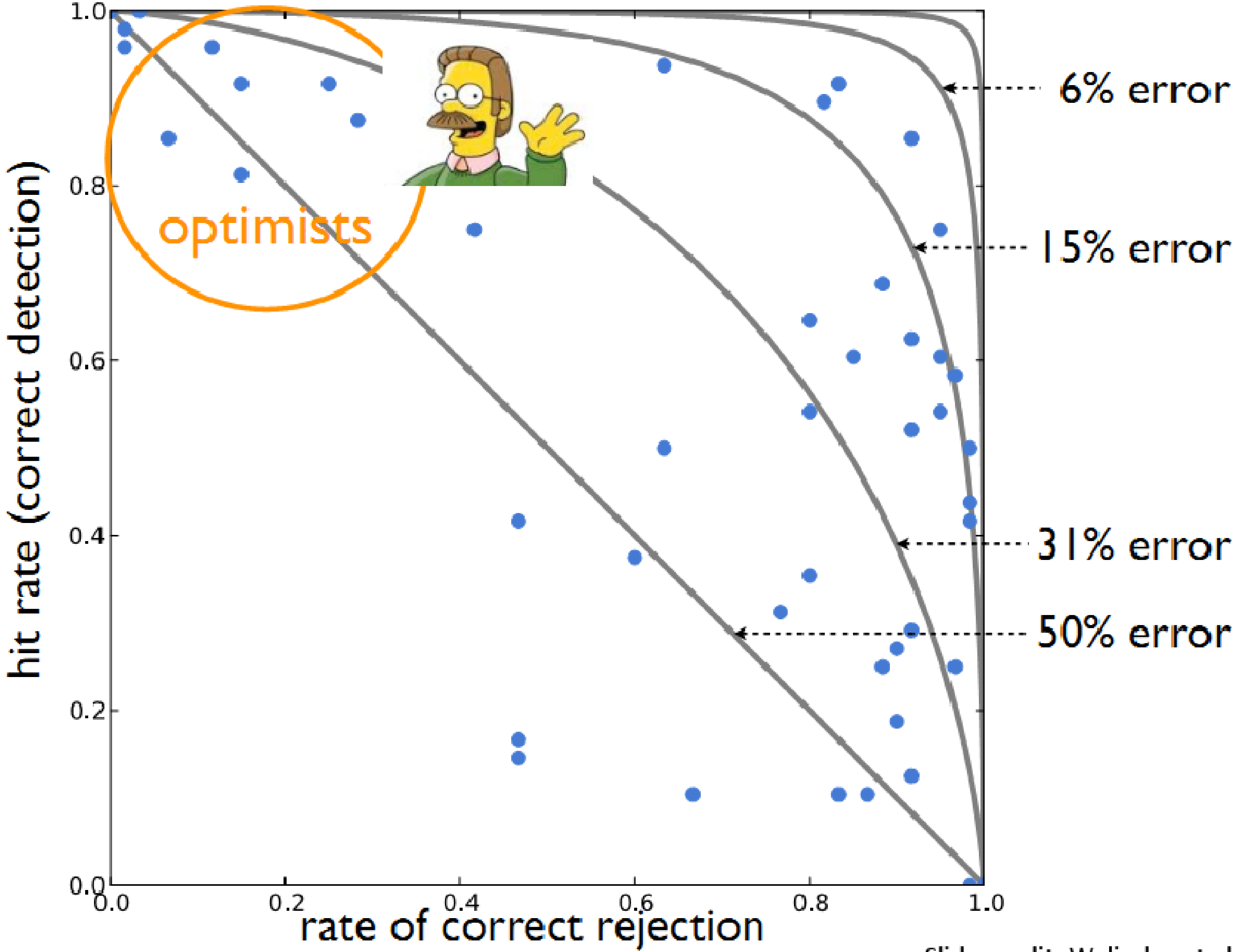
Slide credit: Welinder et al

Task: Find the Indigo Bunting



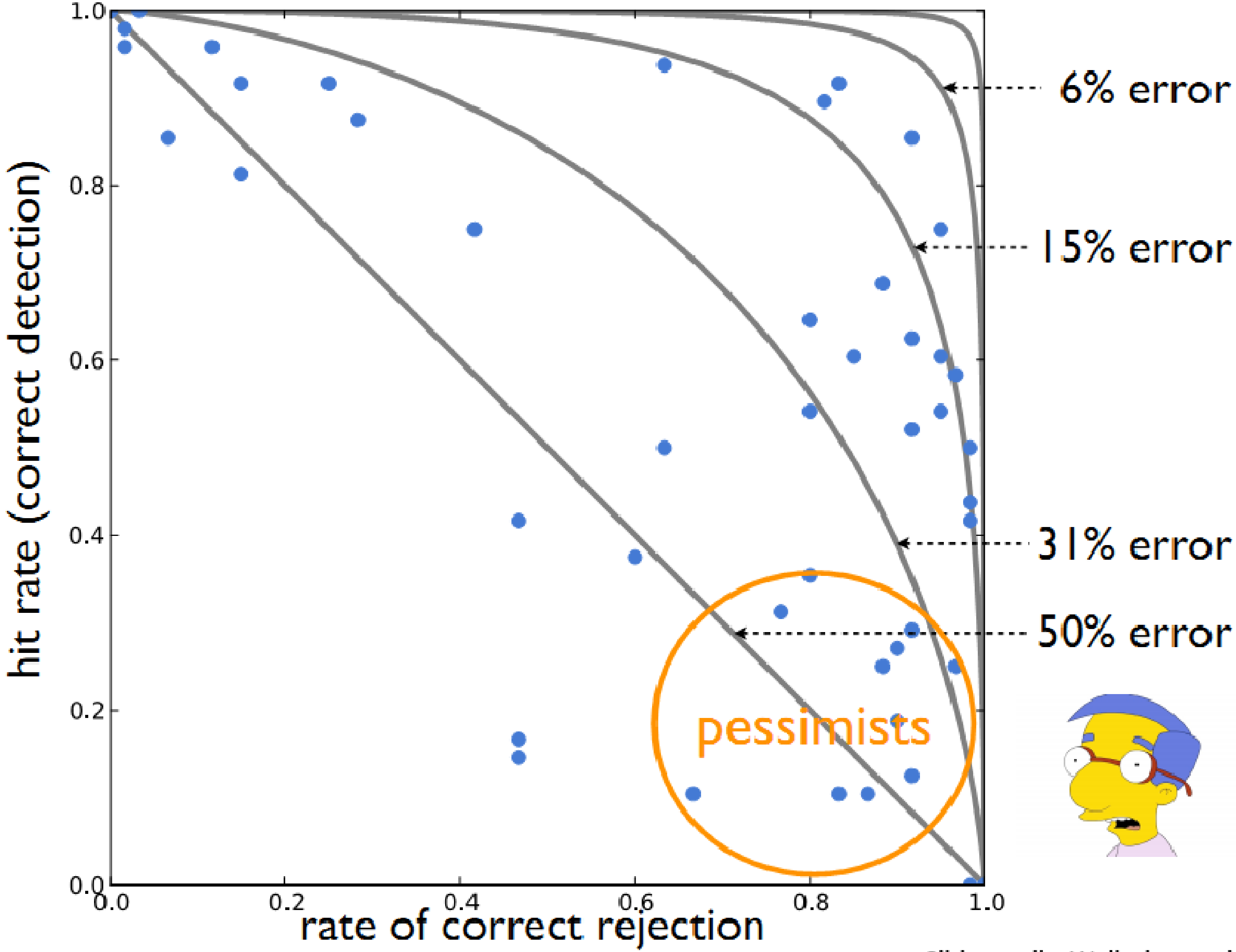
Slide credit: Welinder et al

Task: Find the Indigo Bunting



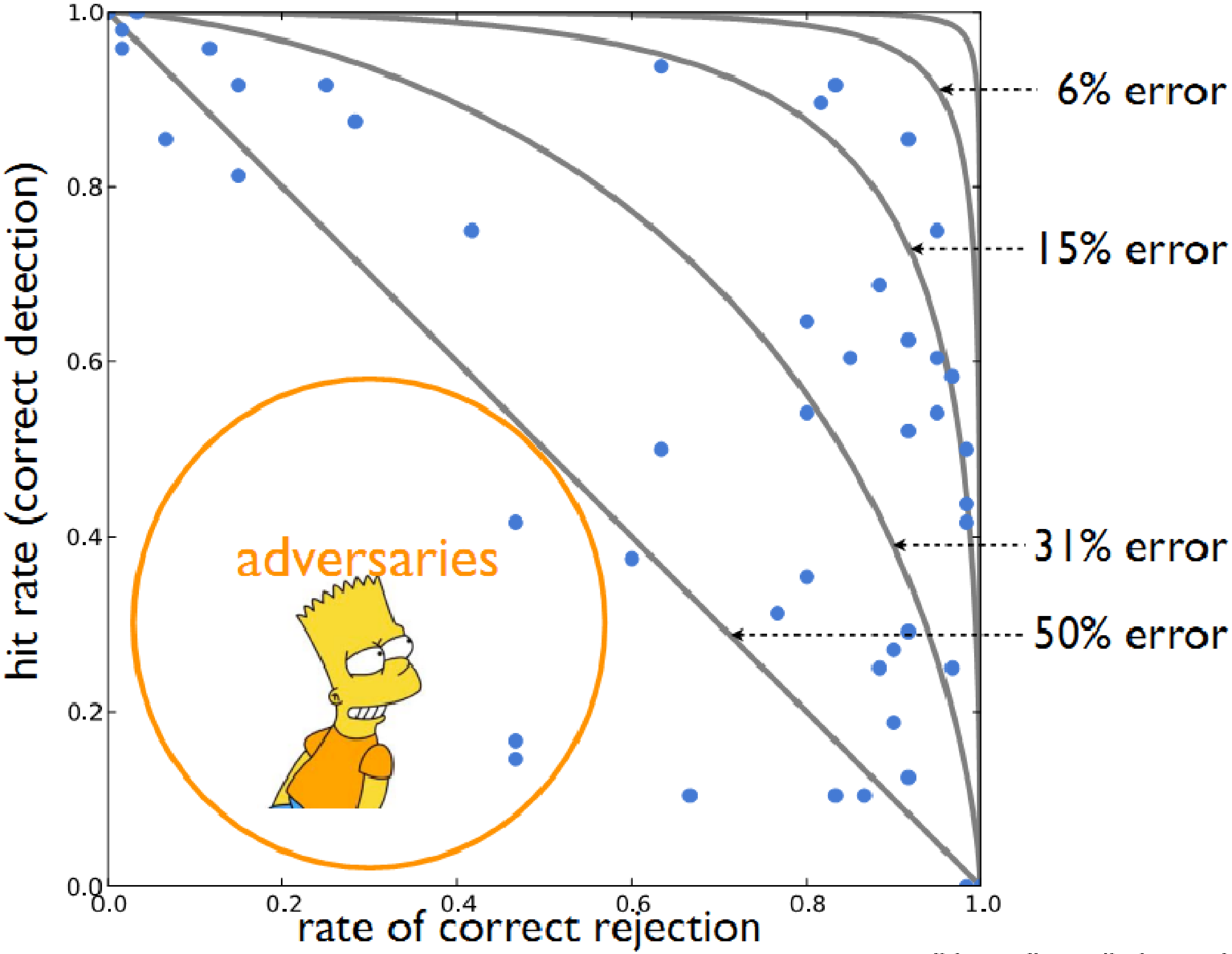
Slide credit: Welinder et al

Task: Find the Indigo Bunting



Slide credit: Welinder et al

Task: Find the Indigo Bunting



Slide credit: Welinder et al

Utility data annotation via Amazon Mechanical Turk

 X 100 000 = $5000

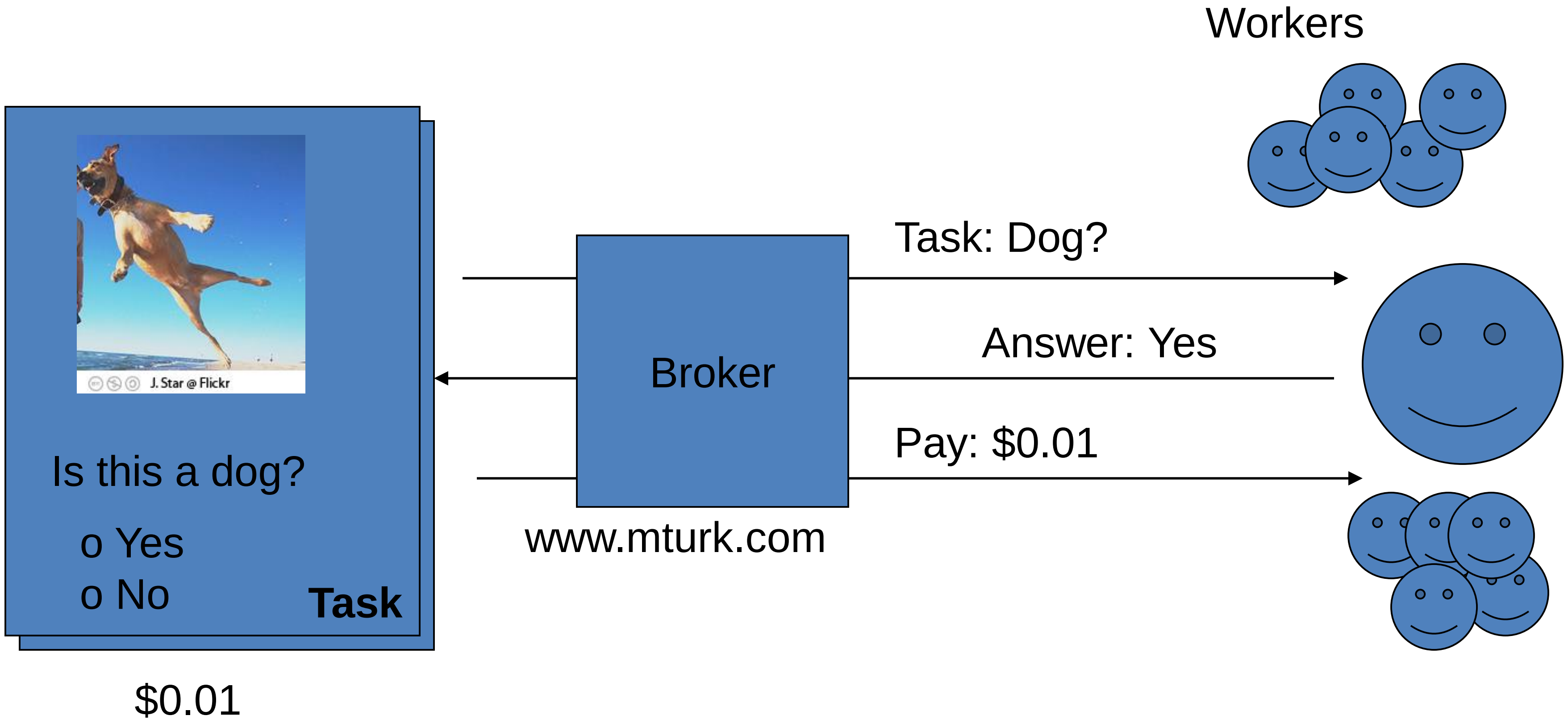
Alexander Sorokin

David Forsyth

CVPR Workshops 2008

Slides by Alexander Sorokin

Amazon Mechanical Turk

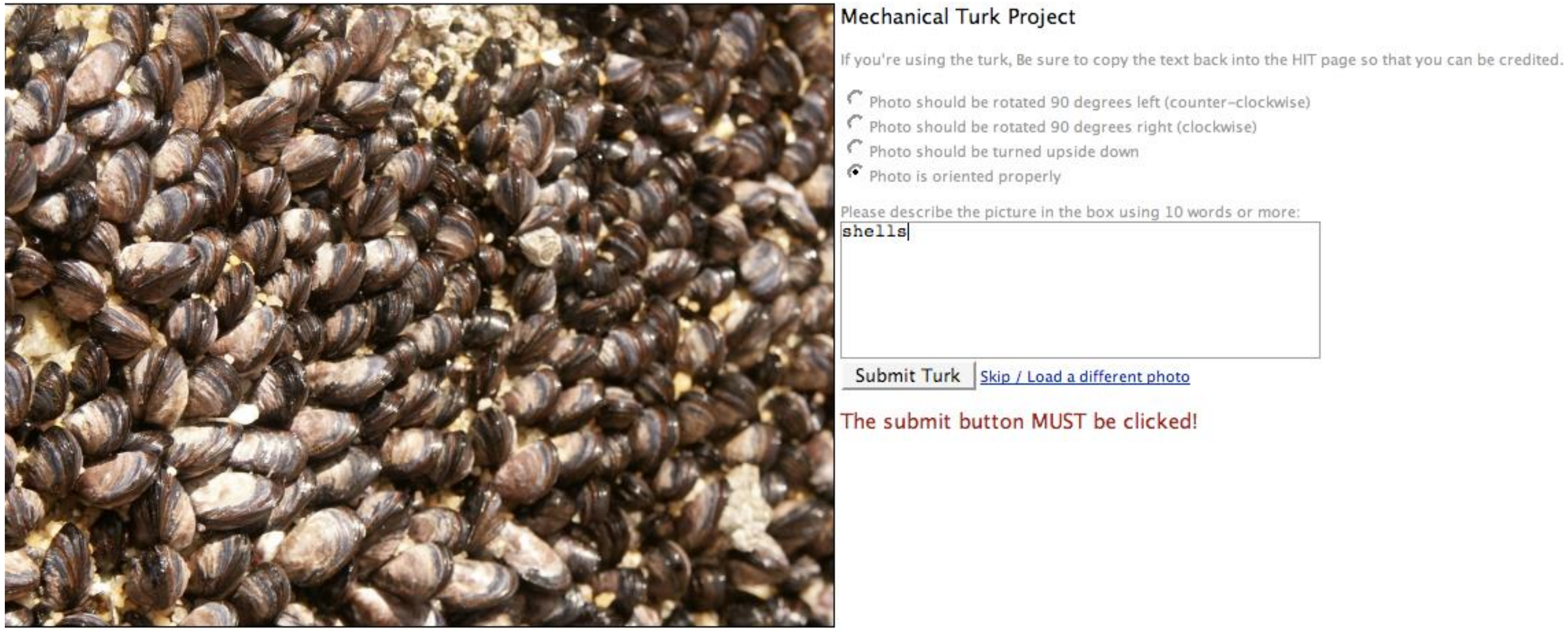


Annotation protocols

- Type keywords
- Select relevant images
- Click on landmarks
- Outline something
- Detect features

……….. anything else ………

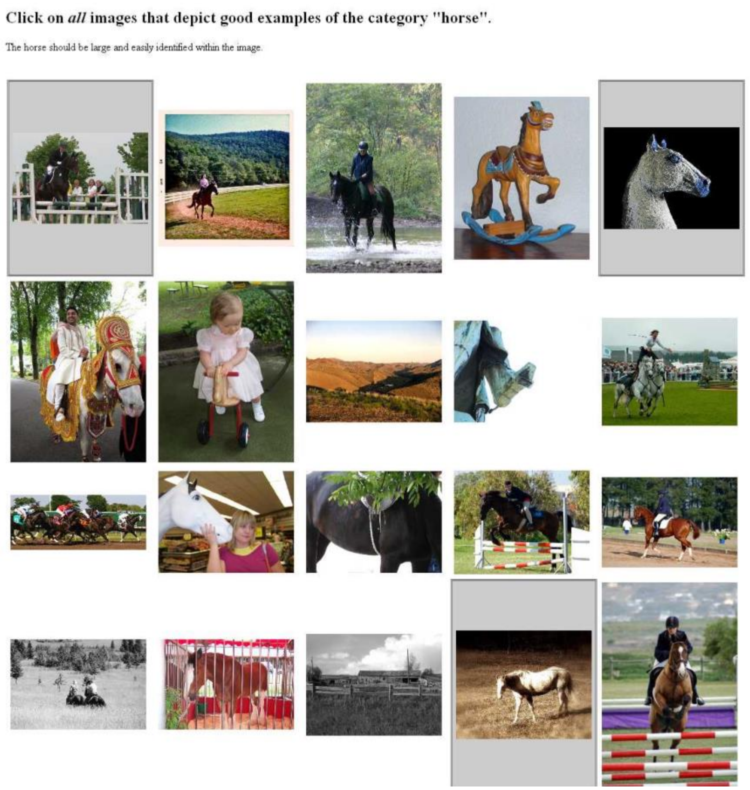
Type keywords

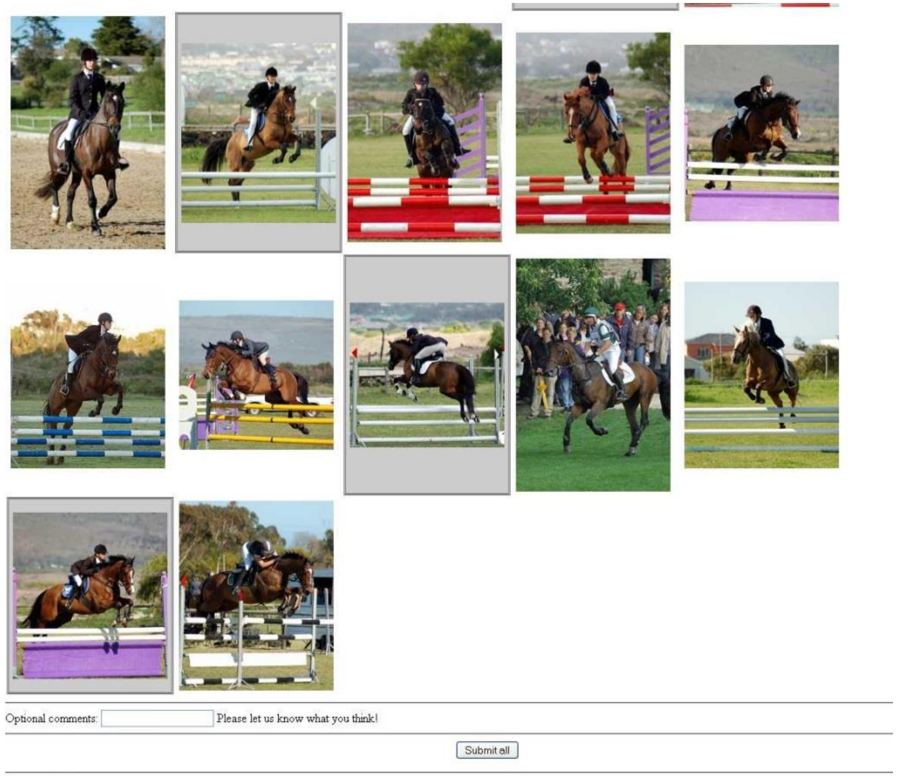


$0.01

http://austinsmoke.com/turk/.

Select examples

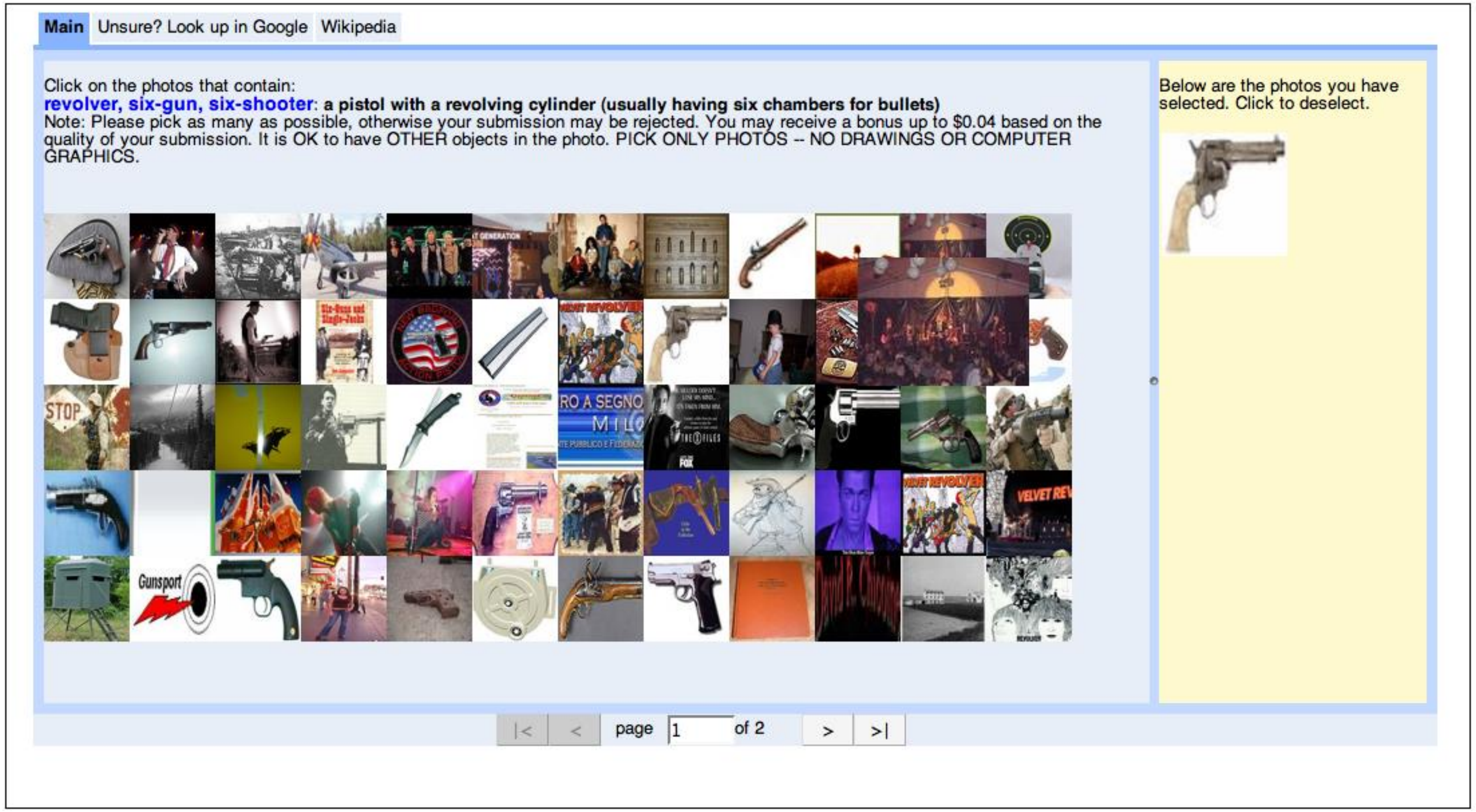




Joint work with Tamara and Alex Berg

http://visionpc.cs.uiuc.edu/~largescale/data/simpleevaluation/html/horse.html

Select examples



$0.02

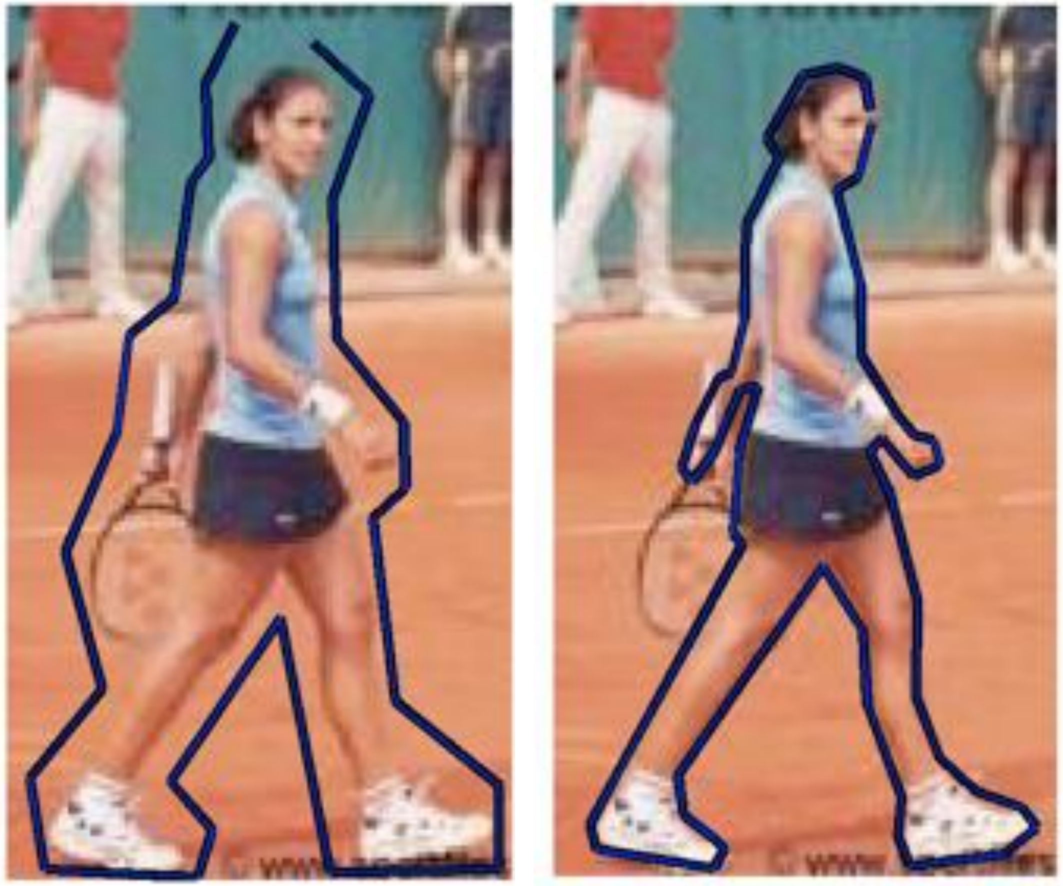
requester mtlabel

Click on landmarks

$0.01

http://vision-app1.cs.uiuc.edu/mt/results/people14-batch11/p7/

Outline something



$0.01

http://visionpc.cs.uiuc.edu/~largescale/results/production-3-2/results_page_013.html

Data from Ramanan NIPS06

Motivation

X 100 000 = $5000

Custom annotations

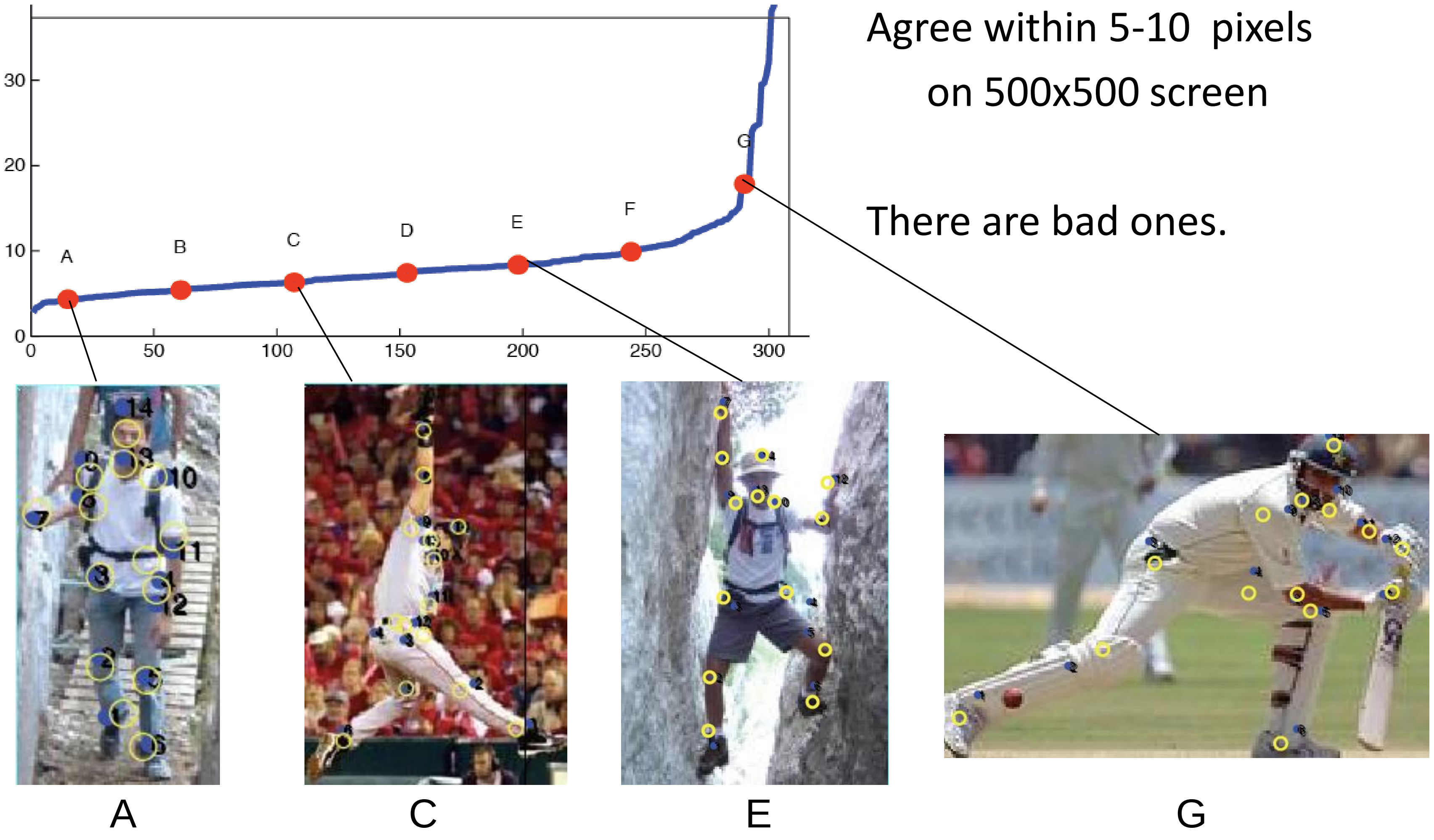
Large scale

Low price

Issues

- Quality?
  - How good is it?
  - How to be sure?
- Price?
  - How to price it?

Annotation quality



Agree within 5-10 pixels on 500x500 screen

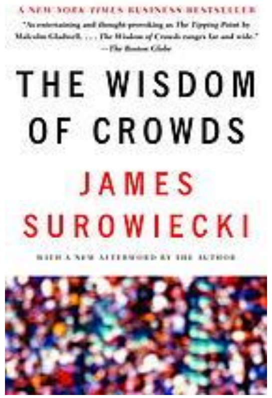
There are bad ones.

How do we get quality annotations?

Ensuring Annotation Quality

- Consensus / Multiple Annotation / "Wisdom of the Crowds"

  Not enough on its own, but widely used



- Gold Standard / Sentinel
  - Special case: qualification exam

  Widely used and most important. Find good annotators and keep them honest.

- Grading Tasks
  - A second tier of workers who grade others

  Not widely used

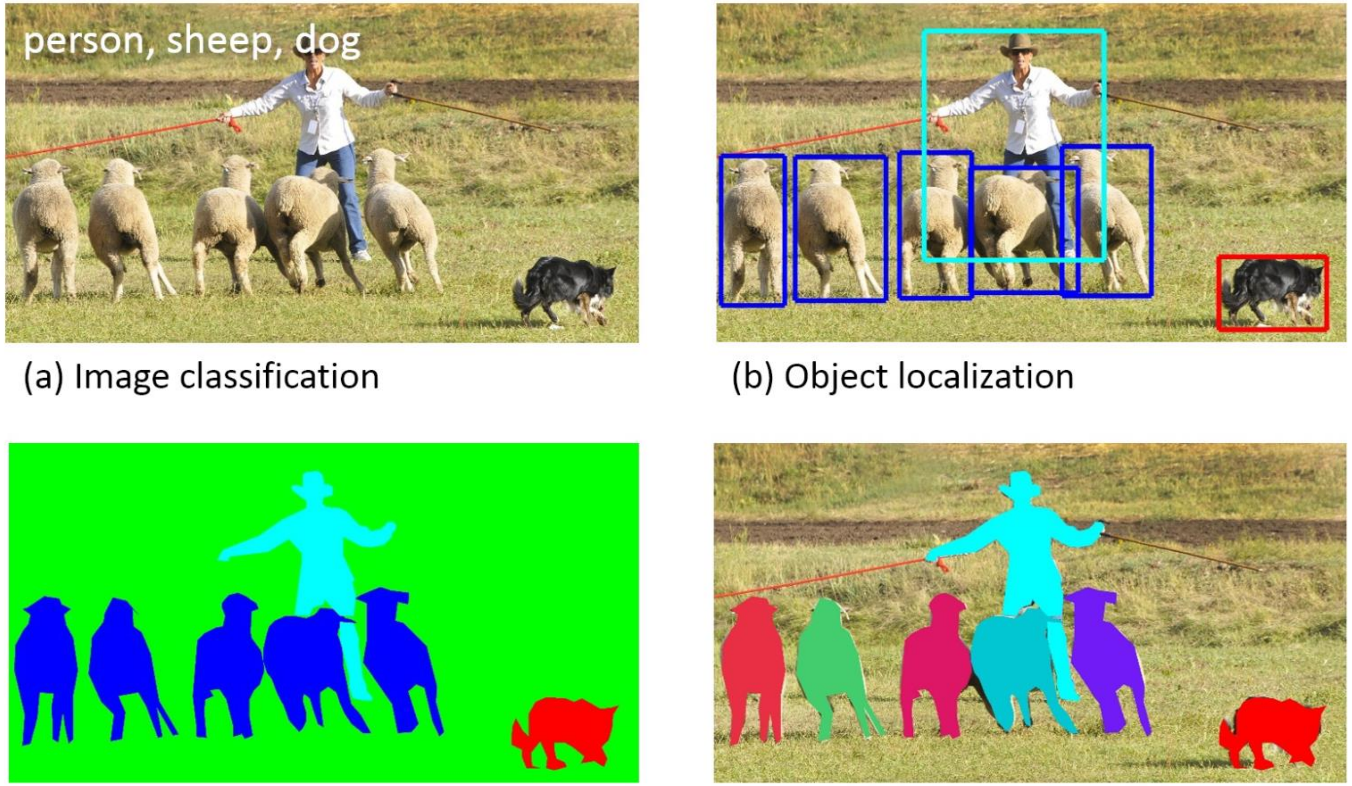
Pricing

- Trade off between throughput and cost
  - *NOT* as much of a trade off with quality
- Higher pay can actually attract scammers

Examples of Crowdsourcing

- Massive annotation efforts that would not otherwise be feasible
  - ImageNet (http://www.image-net.org/)
  - COCO (http://cocodataset.org)
  - Many more

Crowdsourcing to build COCO Dataset



(a) Image classification

(b) Object localization

(c) Semantic segmentation

(d) This work

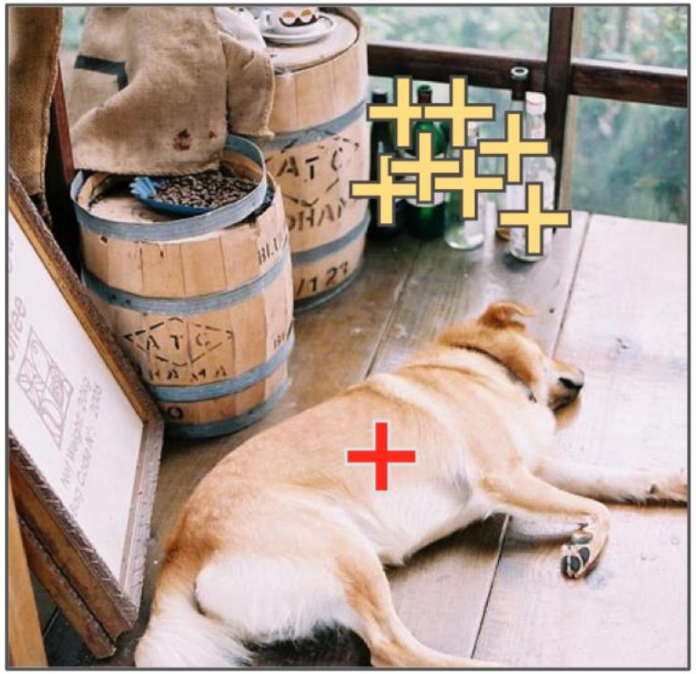
Microsoft COCO: Common Objects in Context

Tsung-Yi Lin Michael Maire Serge Belongie Lubomir Bourdev Ross Girshick
James Hays Pietro Perona Deva Ramanan C. Lawrence Zitnick Piotr Dollár

Crowdsourcing to build COCO Dataset

Annotation Pipeline

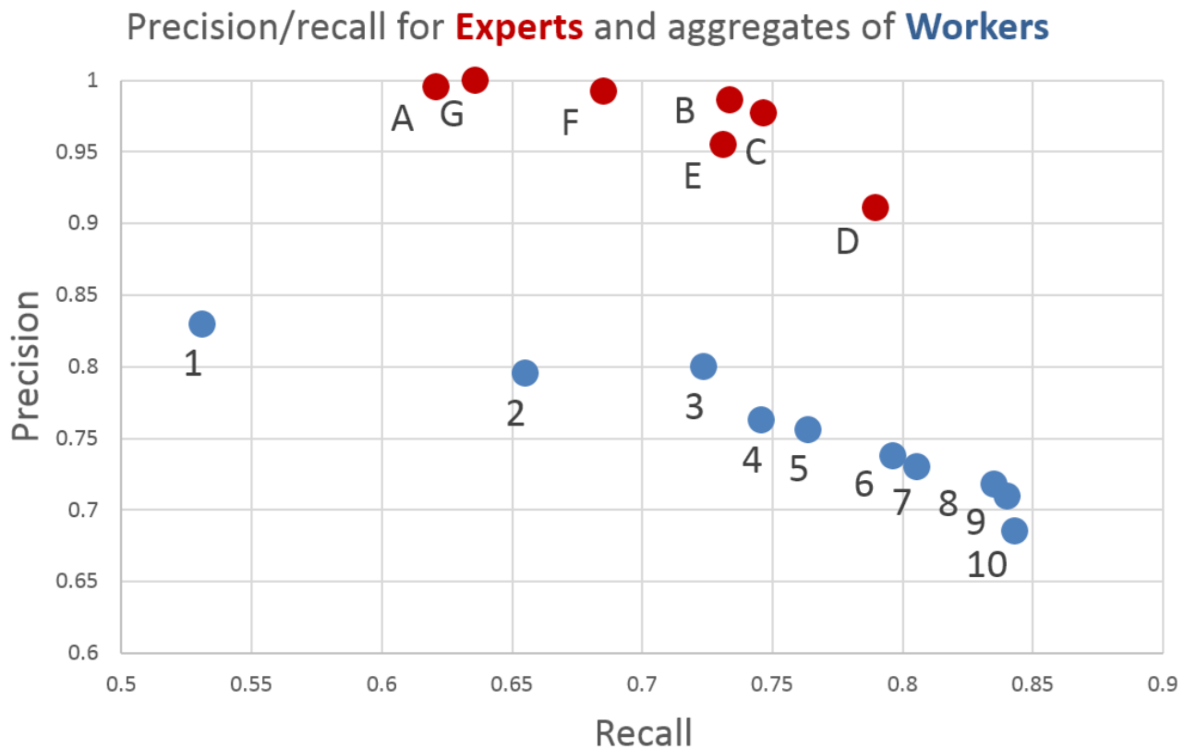
(a) Category labeling

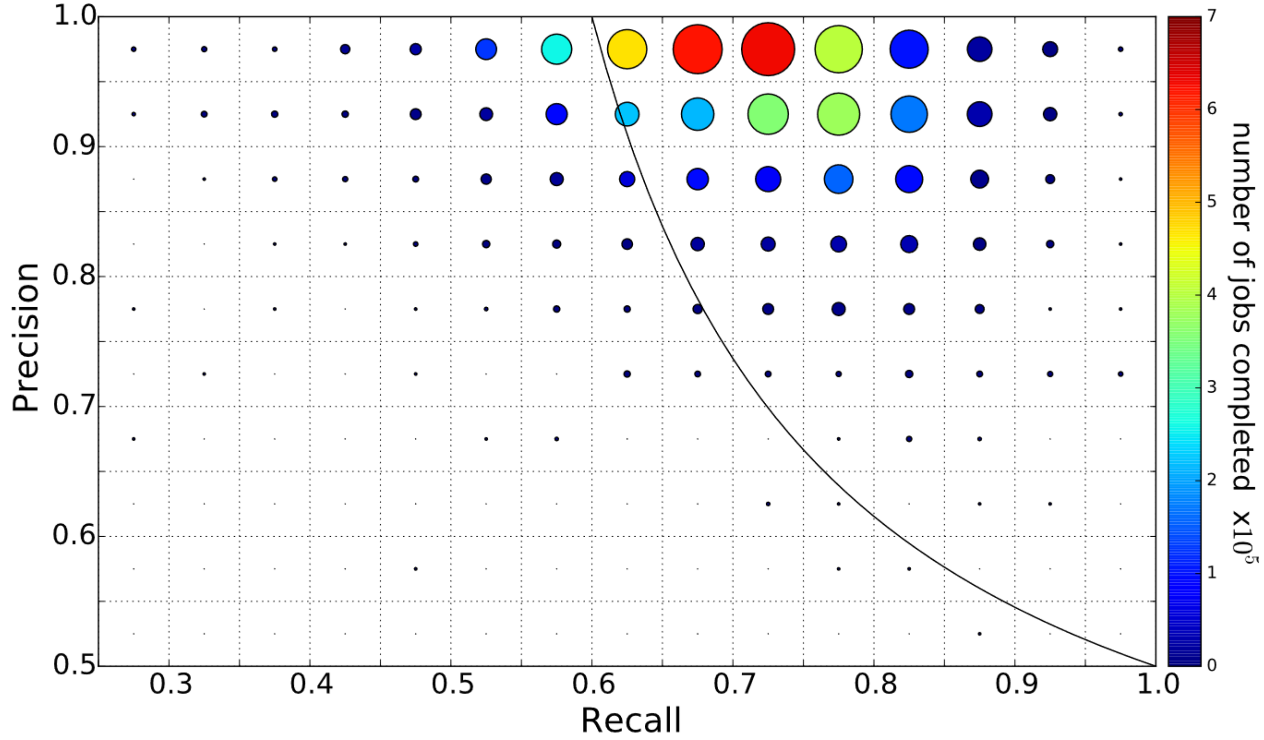


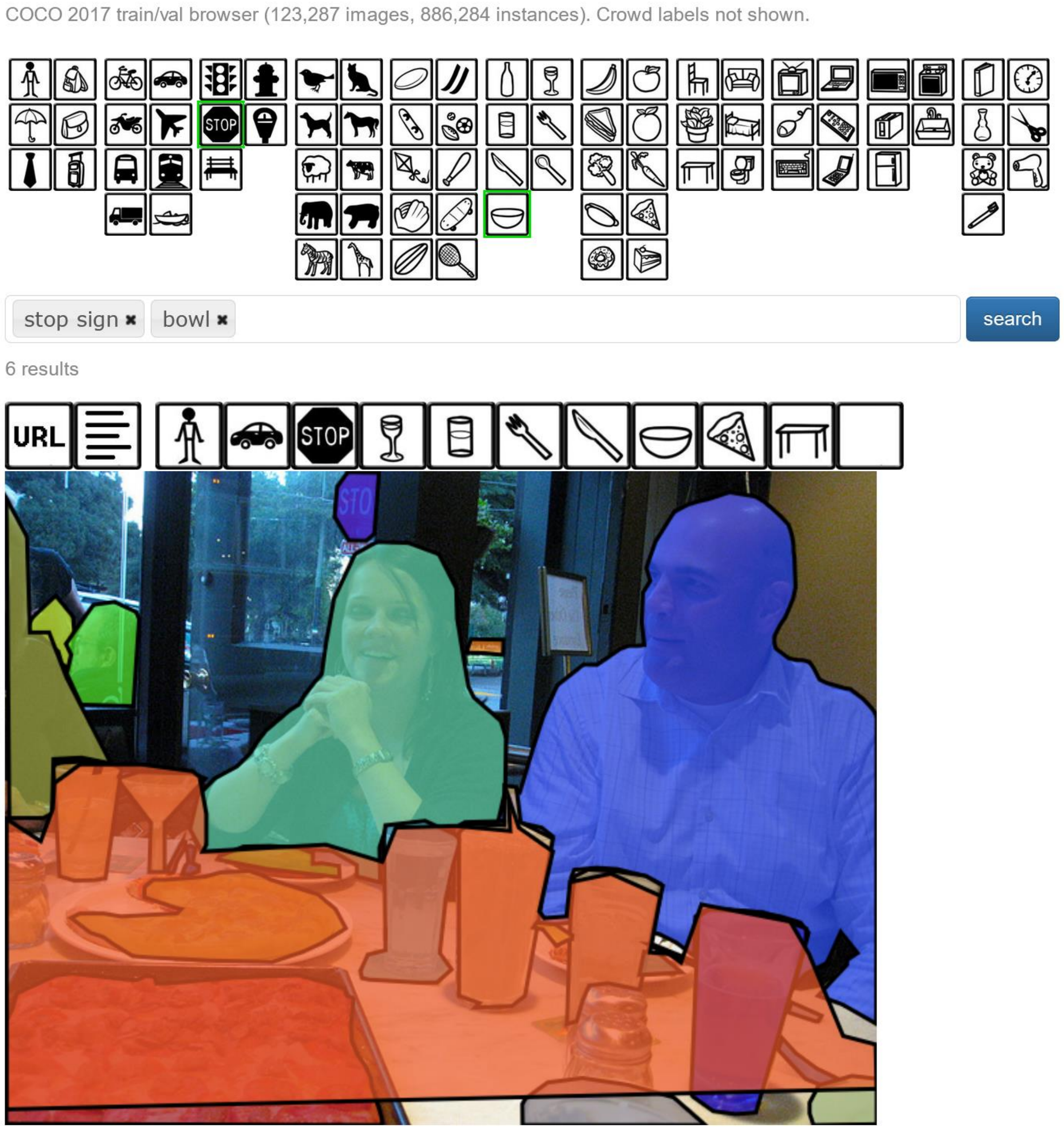
(b) Instance spotting

(c) Instance segmentation

Crowdsourcing to build COCO Dataset



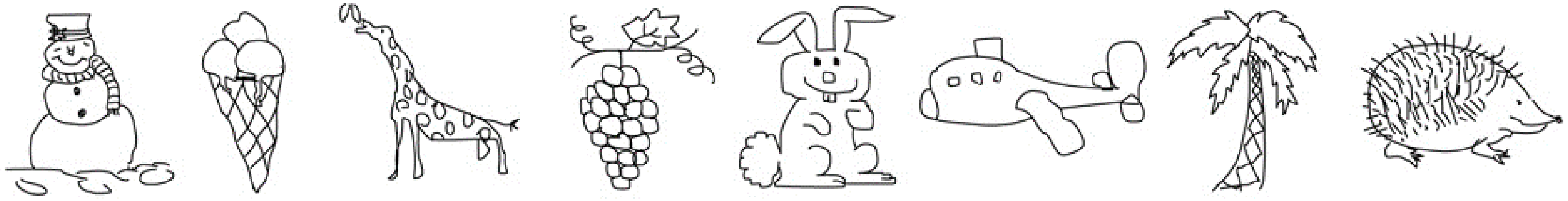




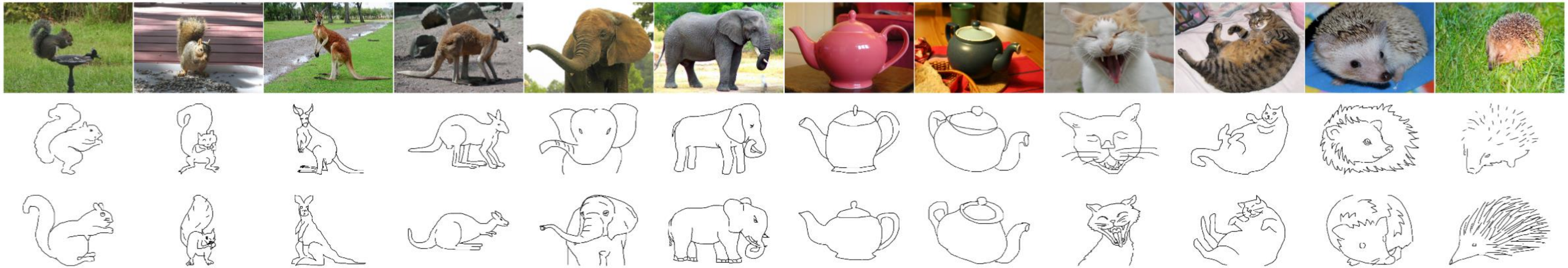
https://cocodataset.org/#explore

Examples of Crowdsourcing

- Most papers annotate images, but there are some more creative uses
  - Webcam Eye tracking (https://webgazer.cs.brown.edu/)
    - Annotation could be the passive observations of a participant
  - Sketch collection (http://cybertron.cg.tu-berlin.de/eitz/projects/classifysketch/)
    - Flips the usual annotation process, by providing a *label* and asking for an *image*

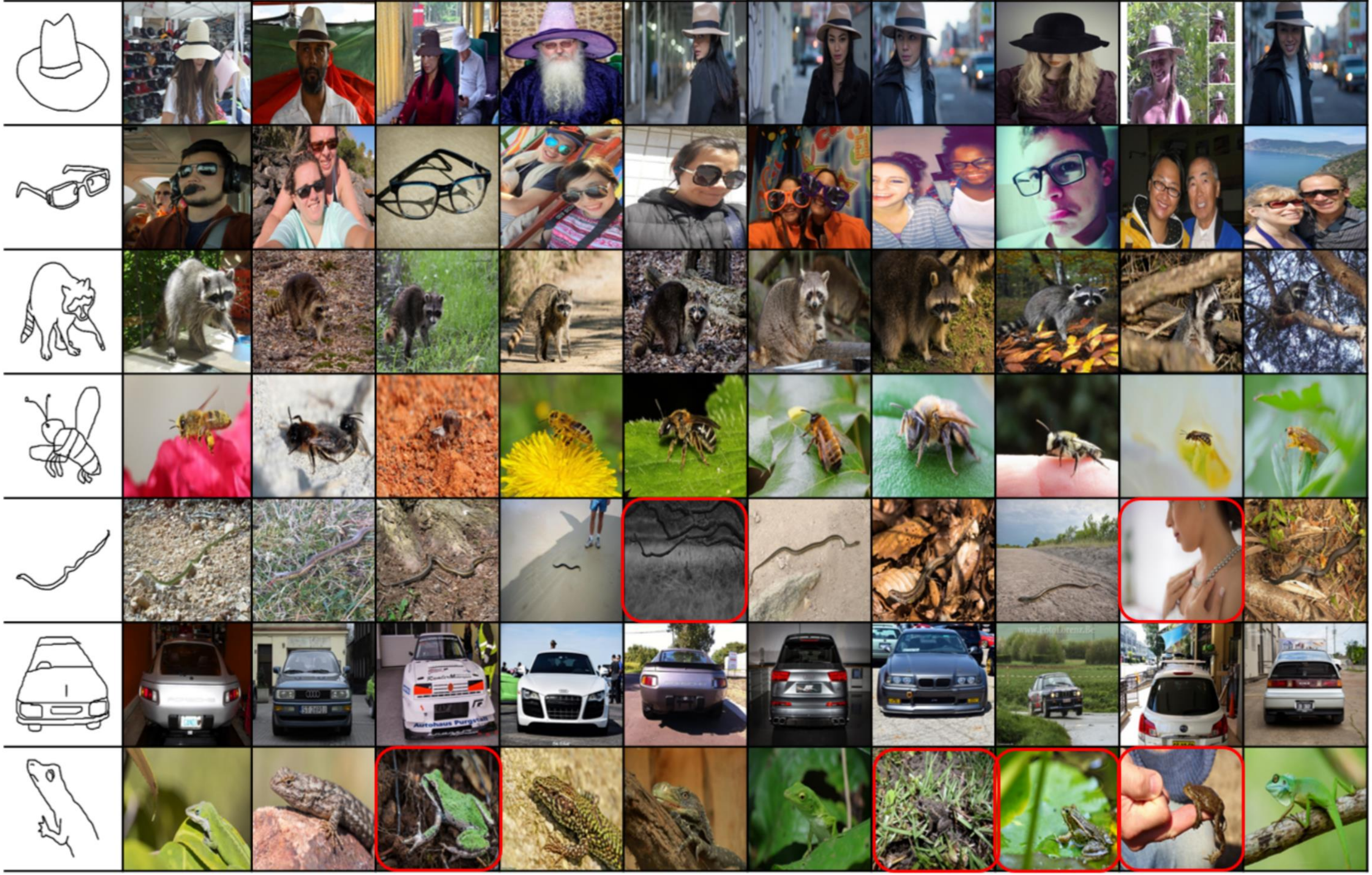


Examples of Crowdsourcing



Draw a sketch of a *particular* photo

The Sketchy Database: Learning to Retrieve Badly Drawn Bunnies. Patsorn Sangkloy Nathan Burnell Cusuh Ham James Hays. Siggraph 2016.



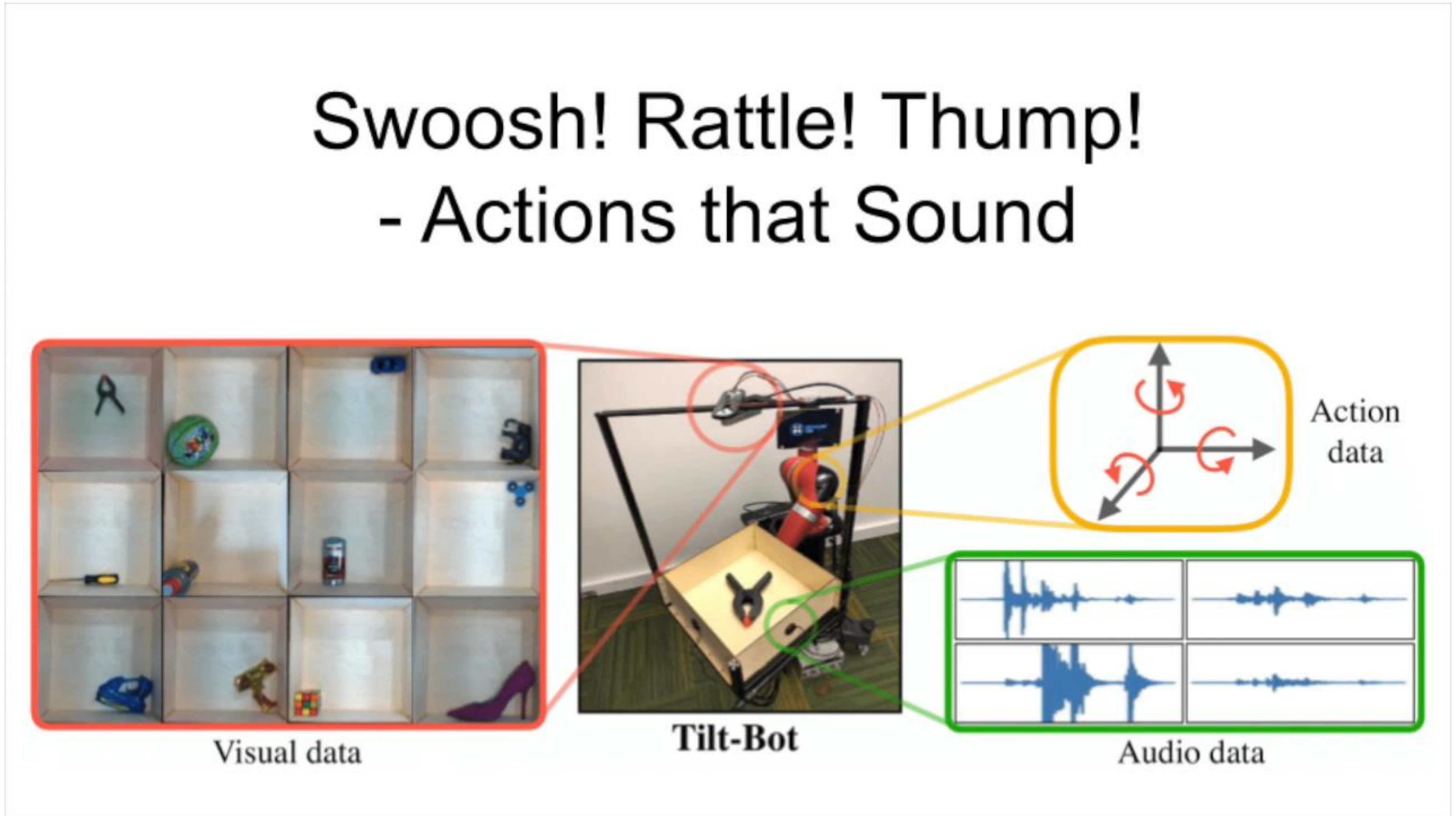
Outline

- Data collection with experts – PASCAL VOC
- Crowdsourcing: Annotation with non-experts
  - LabelMe – no incentive (altruism, perhaps)
  - ESP Game – fun incentive (not fun enough?)
  - Mechanical Turk – financial incentive
- Labels for free / Auto Labeling

Grasp success can be auto-labeled

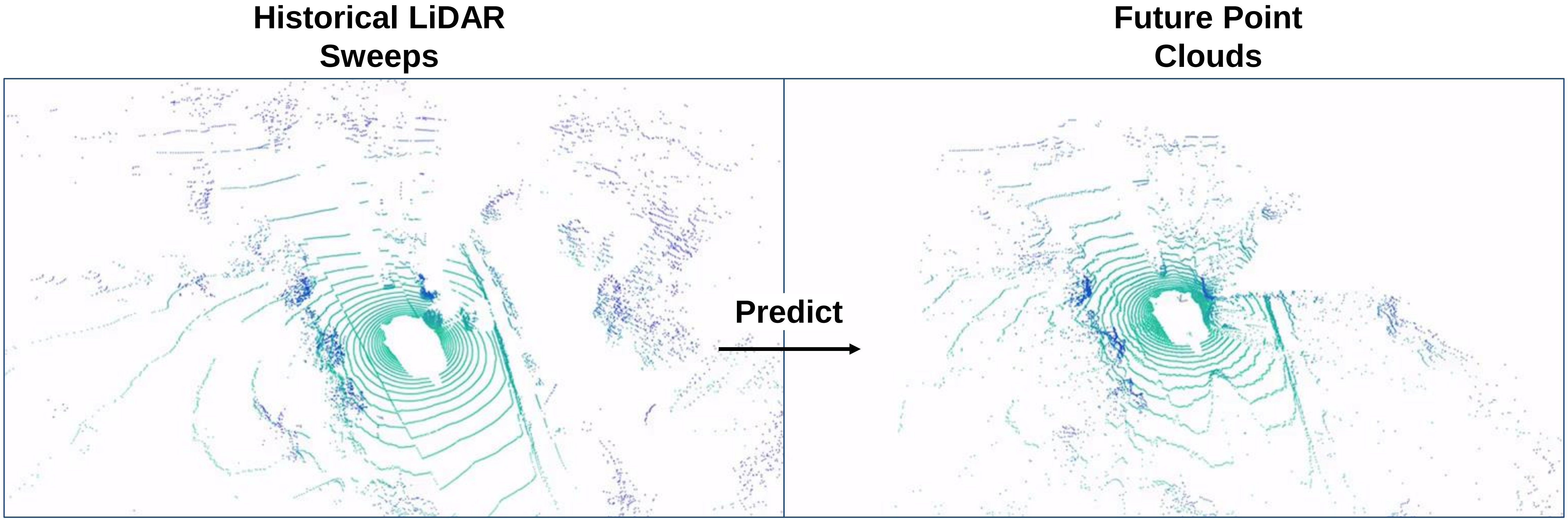
Sergey Levine, Peter Pastor, Alex Krizhevsky, and Deirdre Quillen. Google.

Object sound can be auto-captured



Dhiraj Gandhi, Abhinav Gupta, Lerrel Pinto. **Swoosh! Rattle! Thump! - Actions that Sound.** In *RSS* 2020.

Self-supervised Point Cloud Forecasting



4D Forecasting: Sequential Forecasting of 100,000 Points
Weng et al., CVPR'21

Self-supervised Point Cloud Prediction using 3D Spatial-temporal Convolutional Networks
Mersch et al., CORL'22

Next lecture

- "Unsupervised“ or self-supervised Deep Learning