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

Recap

Opportunities of Scale: Data-driven methods

- Previous Lecture
  - The unreasonable effectiveness of data
  - Scene completion
  - Im2gps
  - Recognition via Tiny Images

The Internet has some rough edges

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

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".

Outline

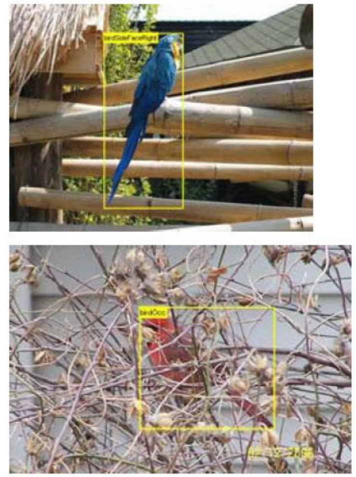
- Data collection with experts – PASCAL VOC
- Annotation with non-experts
  - LabelMe – no incentive (altruism, perhaps)
  - ESP Game – fun incentive (not fun enough?)
  - Mechanical Turk – financial incentive
- Human-in-the-loop Recognition
  - Visipedia

Examples

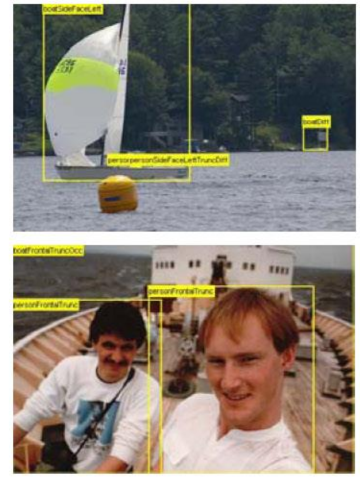
Aeroplane

Bicycle

Bird



Boat



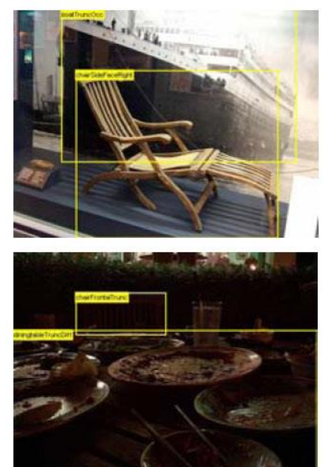
Bottle

Bus

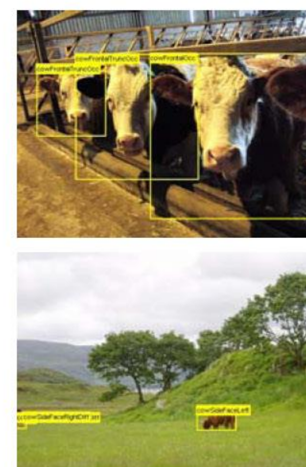
Car

Cat

Chair

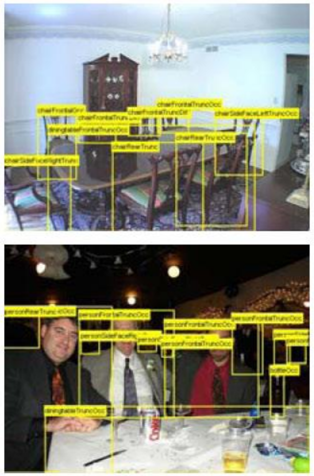


Cow



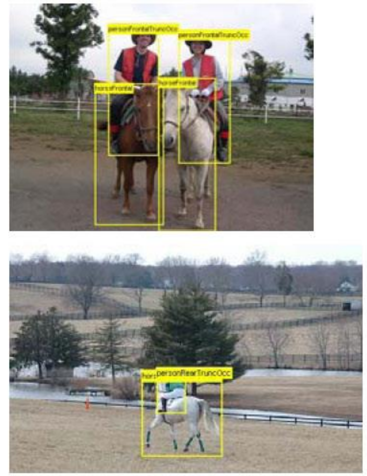
Examples

Dining Table



Dog

Horse

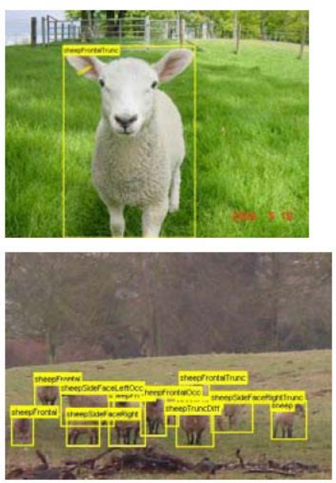


Motorbike

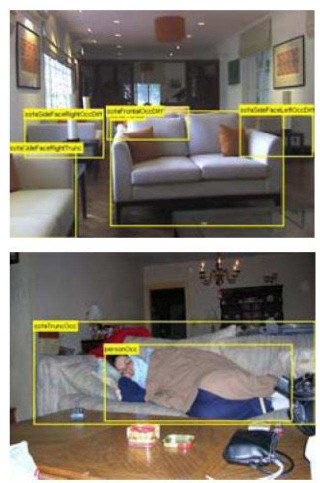
Person

Potted Plant

Sheep



Sofa



Train

TV/Monitor

Outline

- Data collection with experts – PASCAL VOC
- Annotation with non-experts
  - LabelMe – no incentive (altruism, perhaps)
  - ESP Game – fun incentive (not fun enough?)
  - Mechanical Turk – financial incentive
- Human-in-the-loop Recognition
  - Visipedia

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

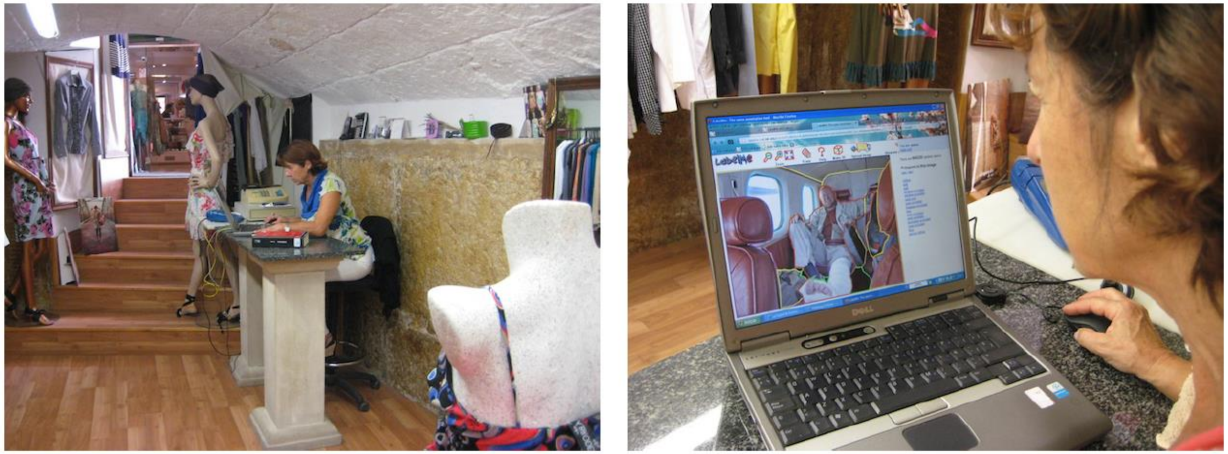


Figure 2: *The image annotation context. All the labeling was done inside a clothing shop named* Transparencia *in the heart of Palma de Mallorca, Spain.*

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

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
logged i
score
0
ESP Game
Concentrate...
time
2:56
What do you see?
taboo words
student
guesses
submit
pass
Play Anonymously
Screencast-O-Matic.com

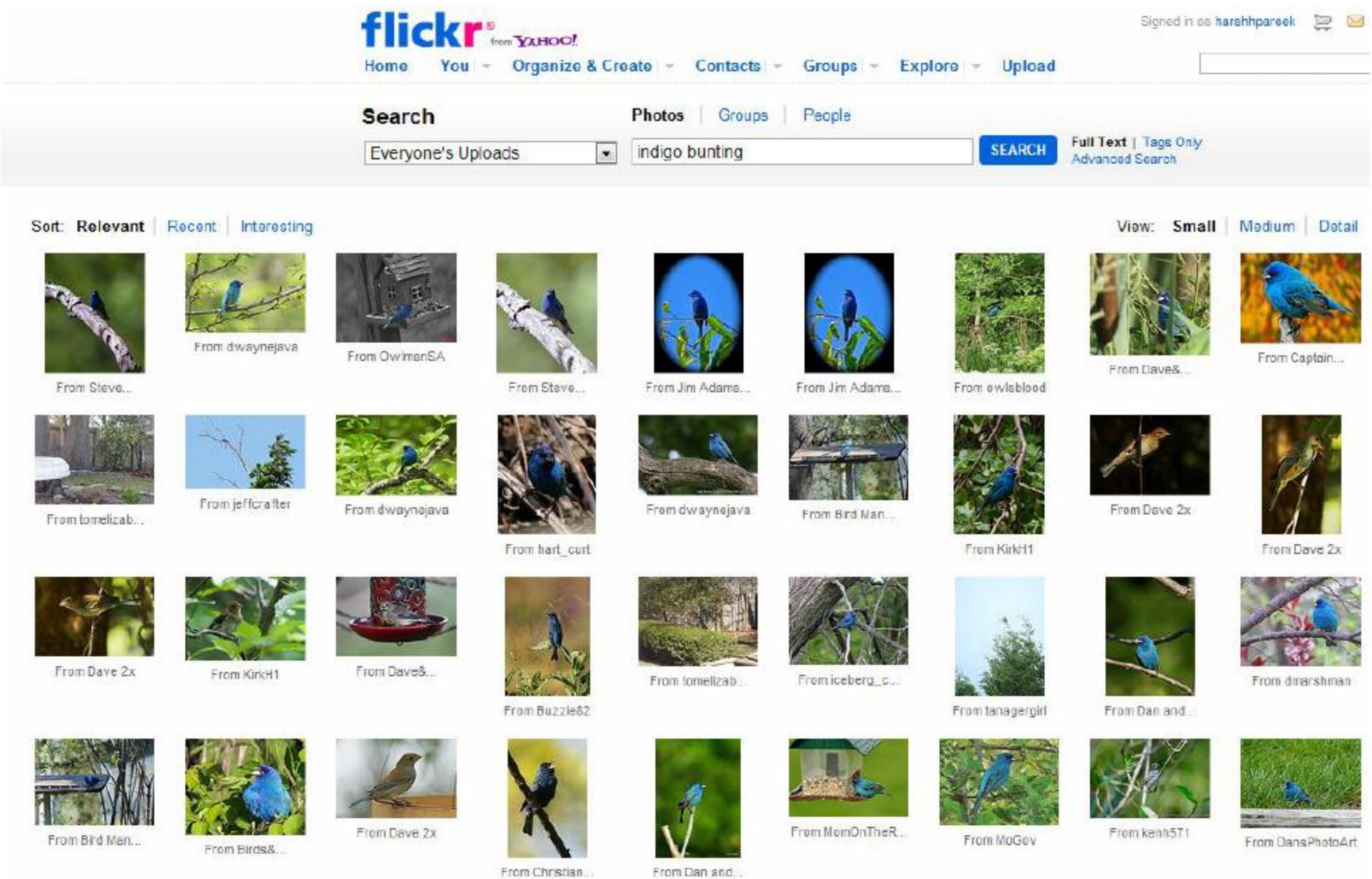
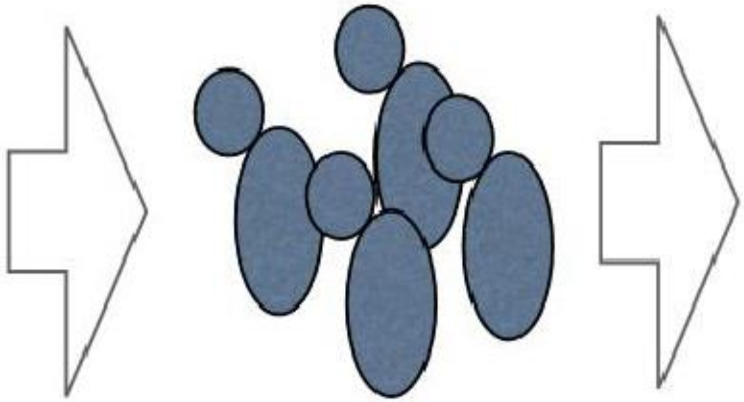


Image credit: Flickr.com

Building datasets

6000 images
from flickr.com

Annotators



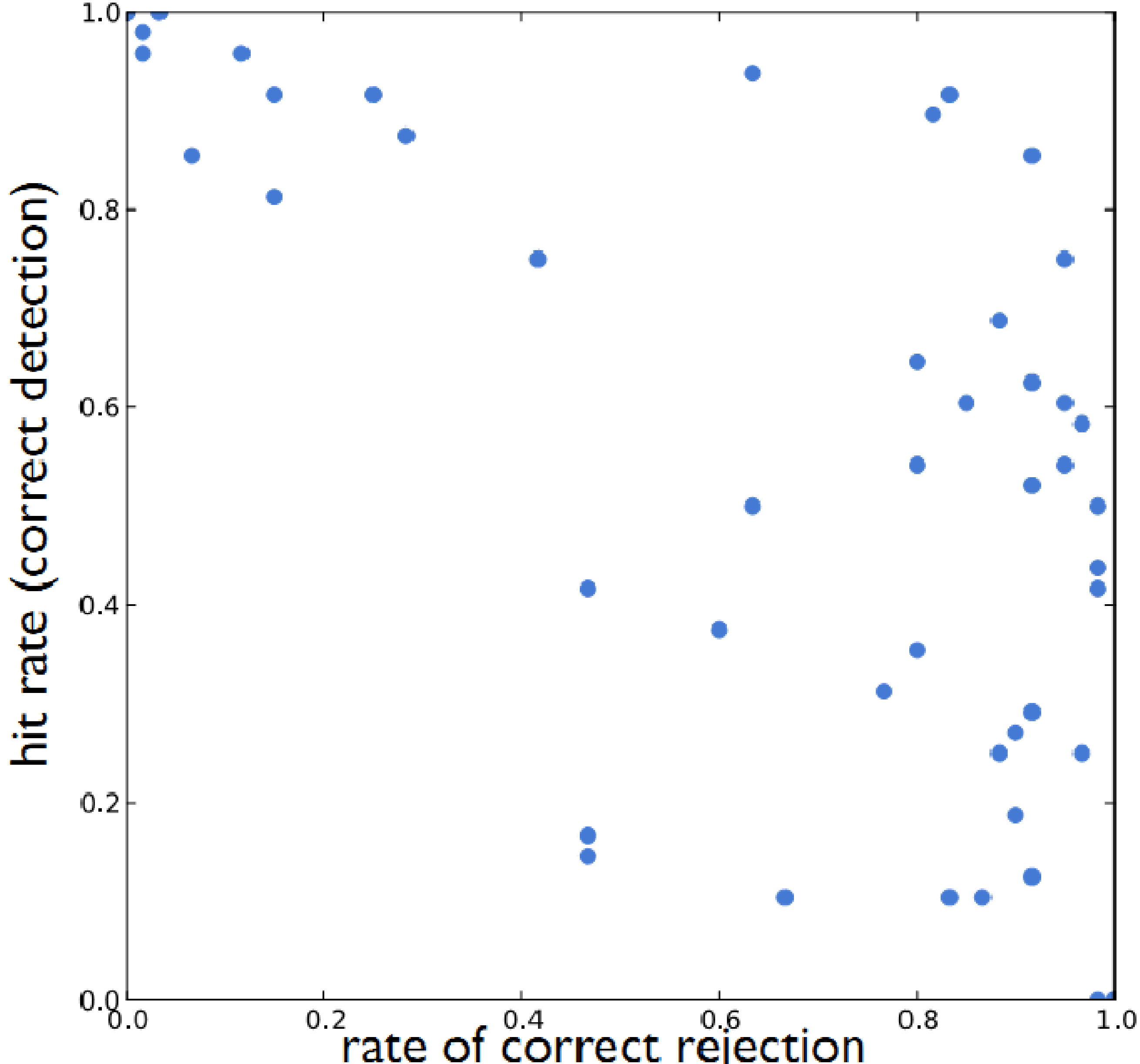
amazonmechanical turk
Artificial Artificial Intelligence

Is there an Indigo bunting in the image?

100s of
training images

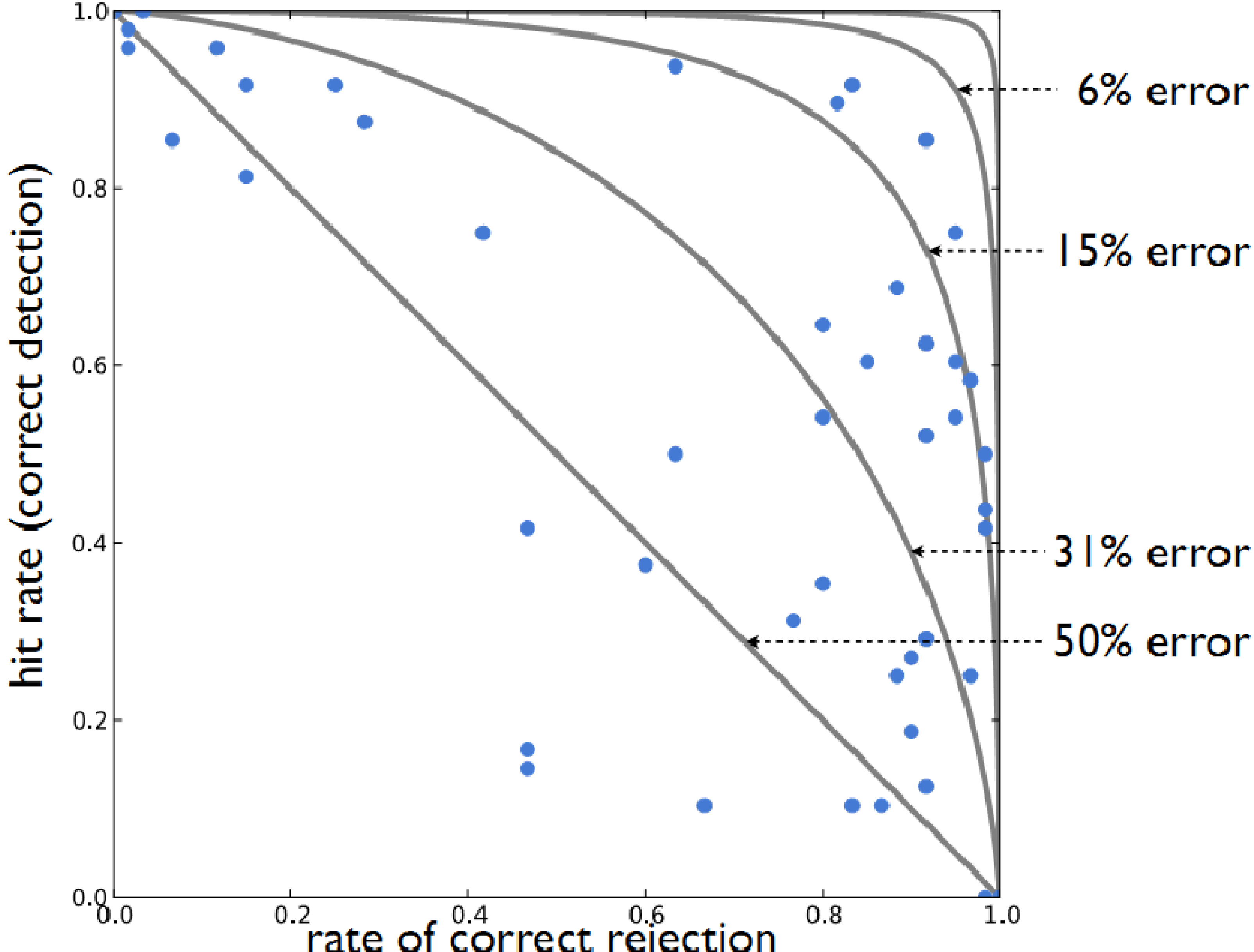
Slide credit: Welinder et al

Task: Find the Indigo Bunting



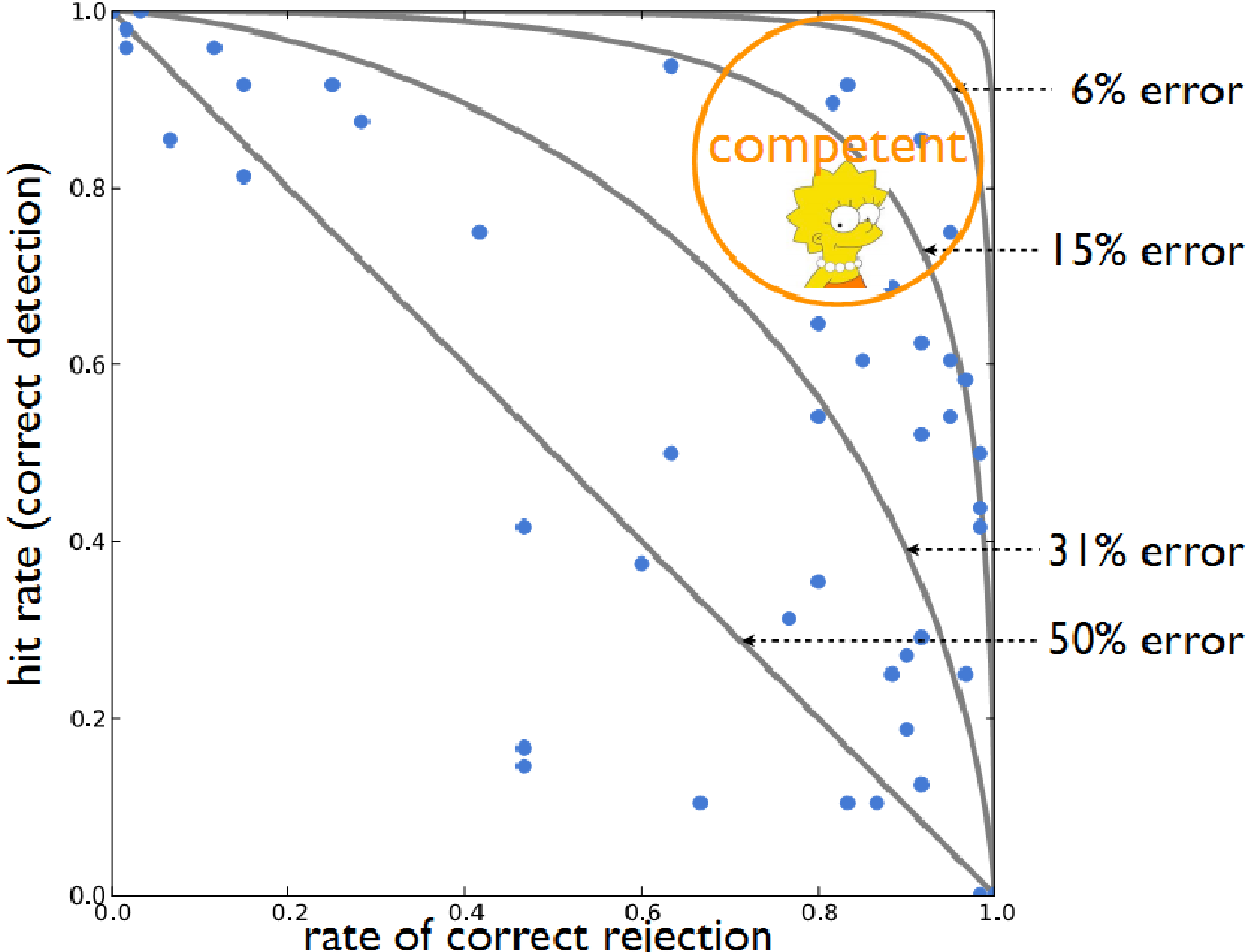
Slide credit: Welinder et al

Task: Find the Indigo Bunting



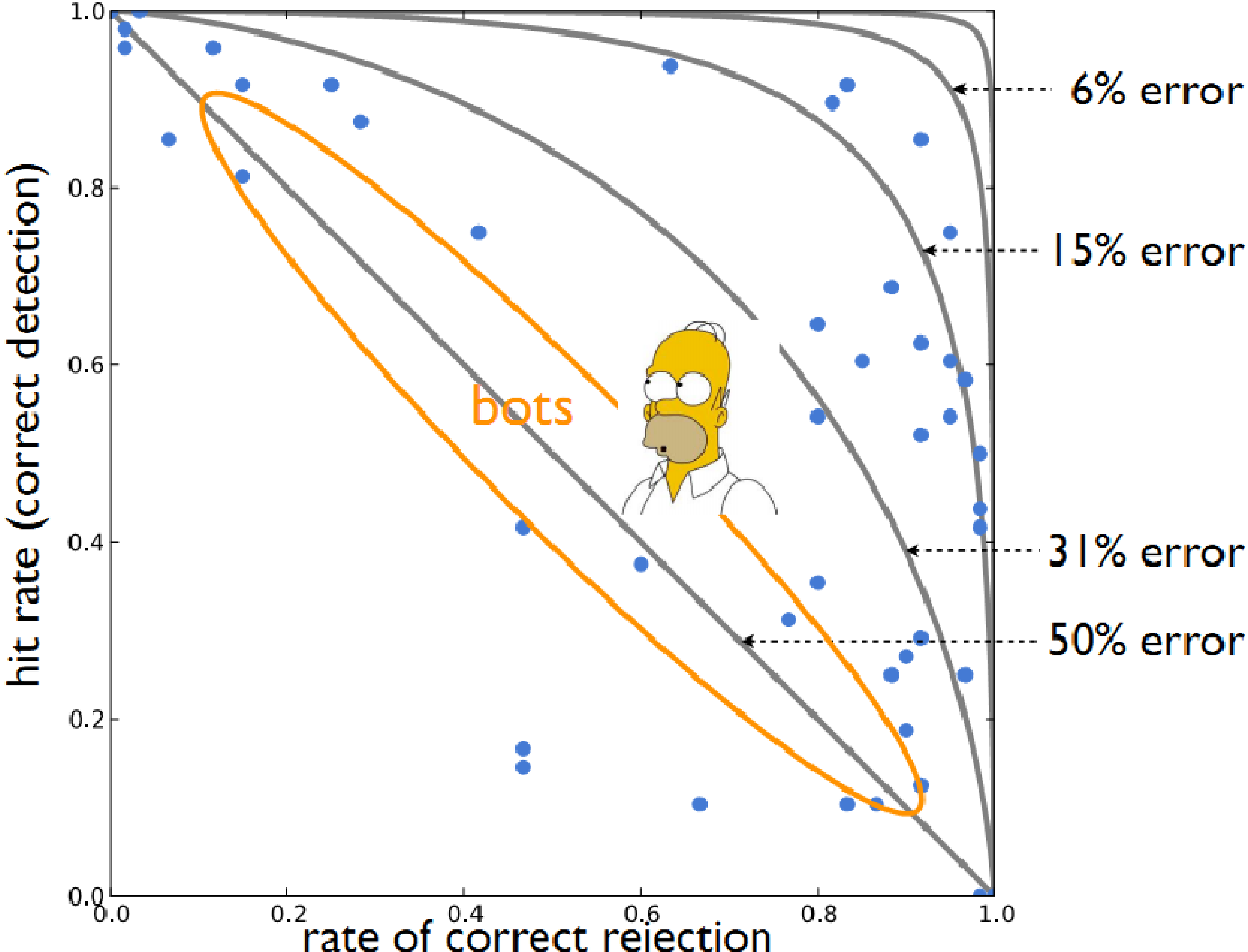
Slide credit: Welinder et al

Task: Find the Indigo Bunting



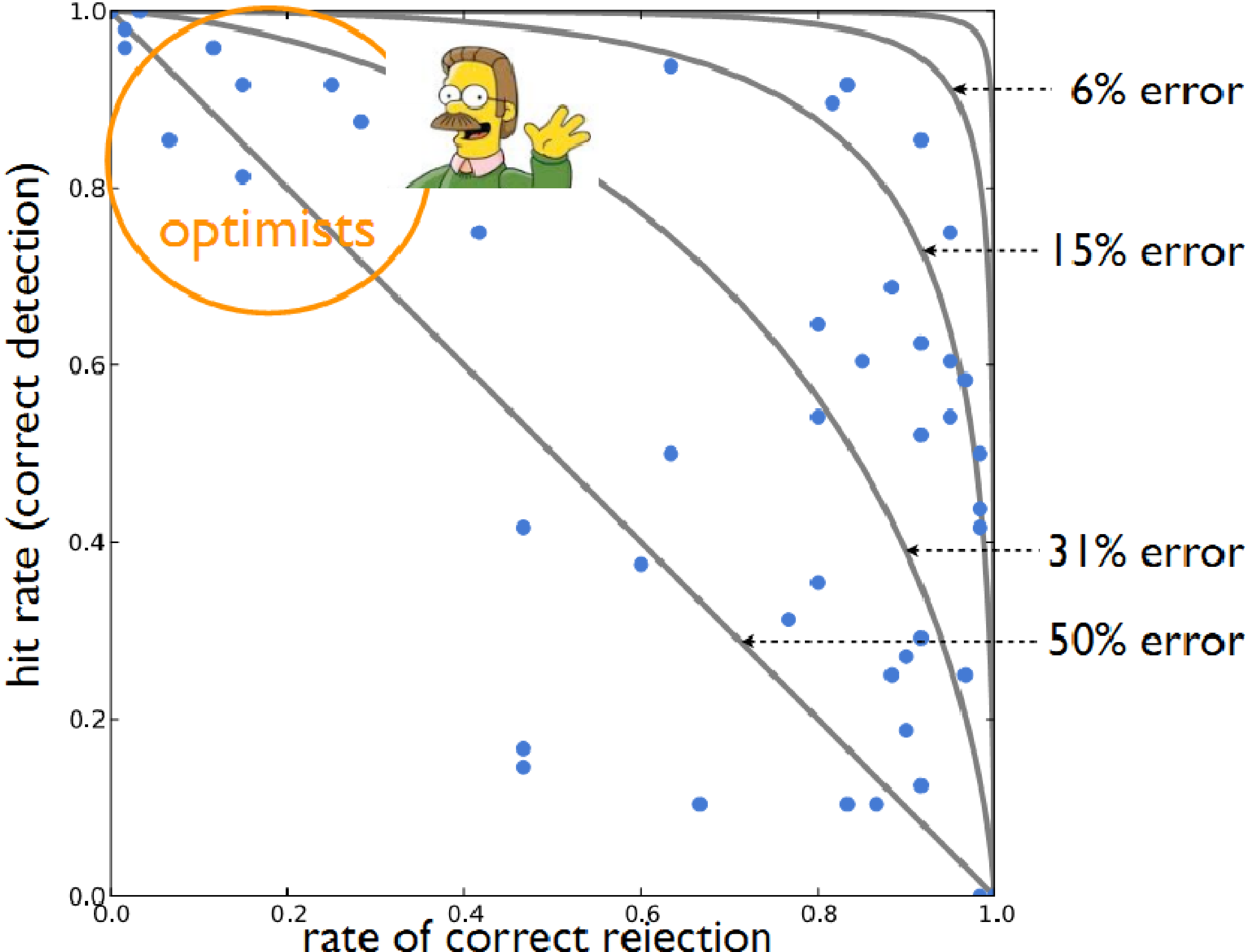
Slide credit: Welinder et al

Task: Find the Indigo Bunting



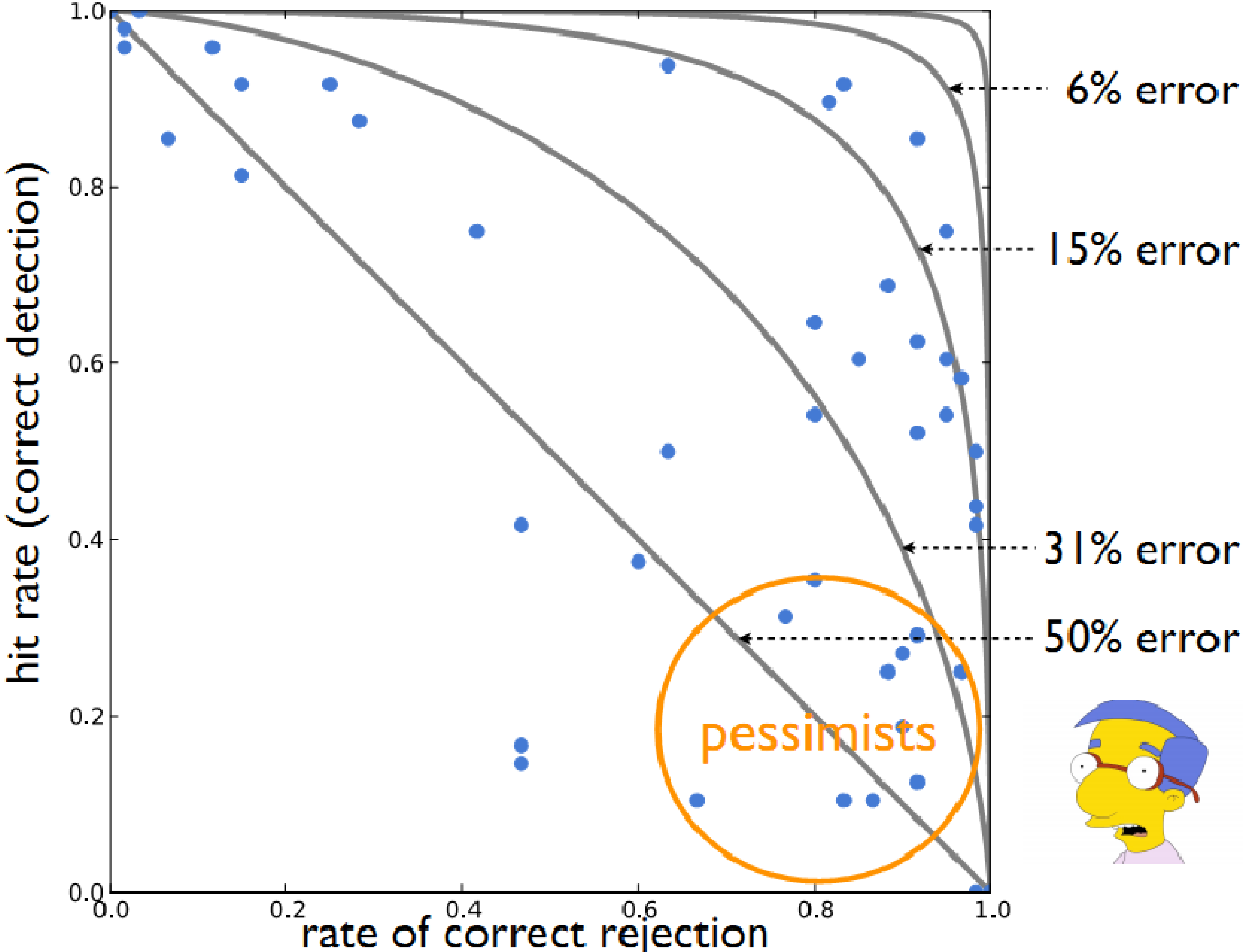
Slide credit: Welinder et al

Task: Find the Indigo Bunting



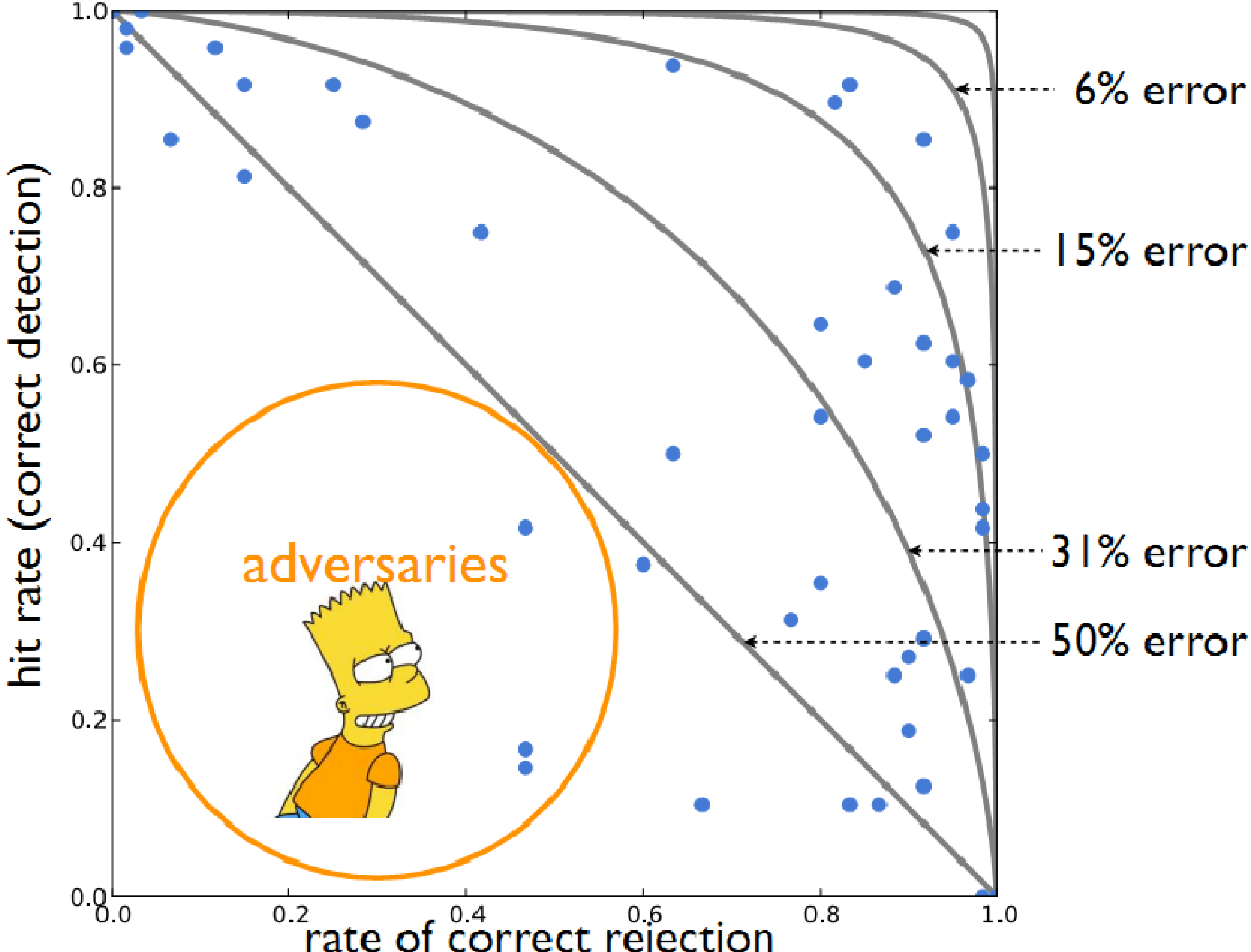
Slide credit: Welinder et al

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

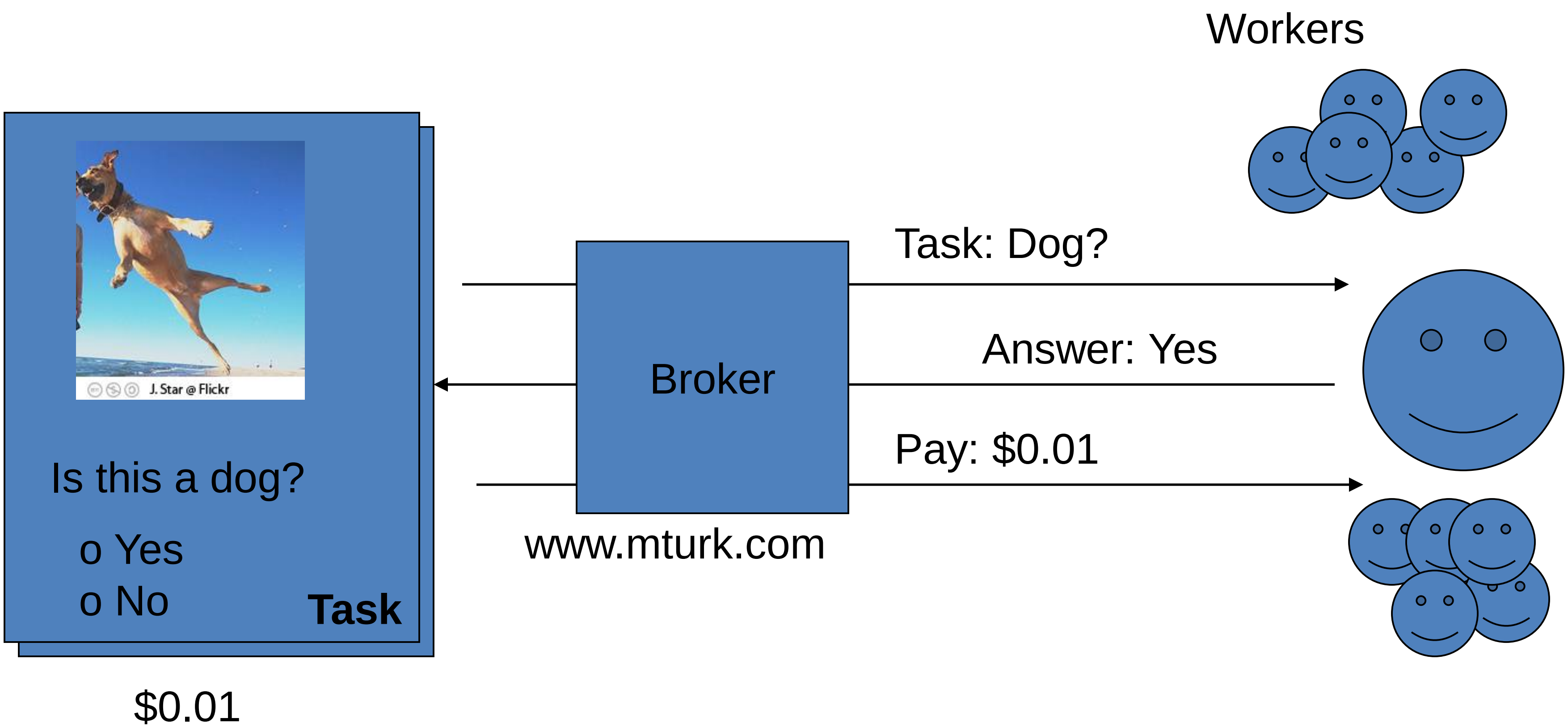
Alexander Sorokin

David Forsyth

CVPR Workshops 2008

Slides by Alexander Sorokin

Amazon Mechanical Turk

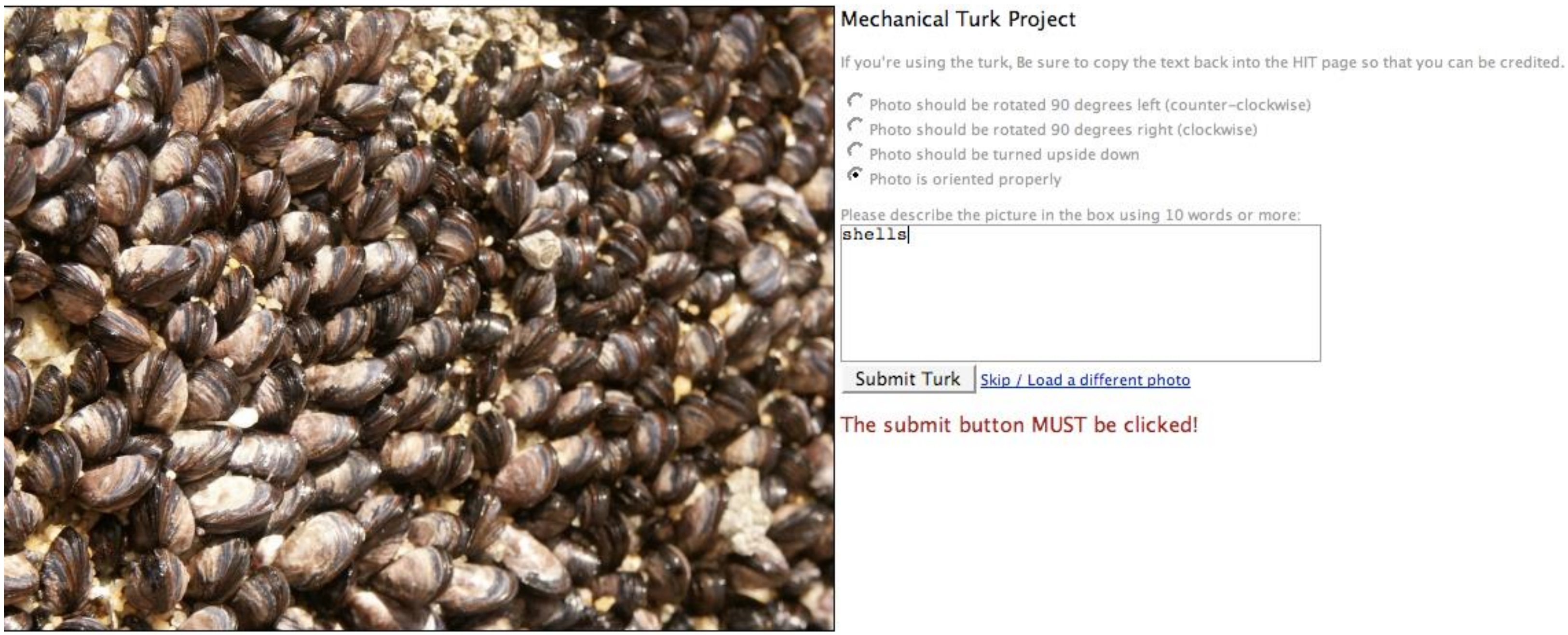


Annotation protocols

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

……….. anything else ………

Type keywords



$0.01

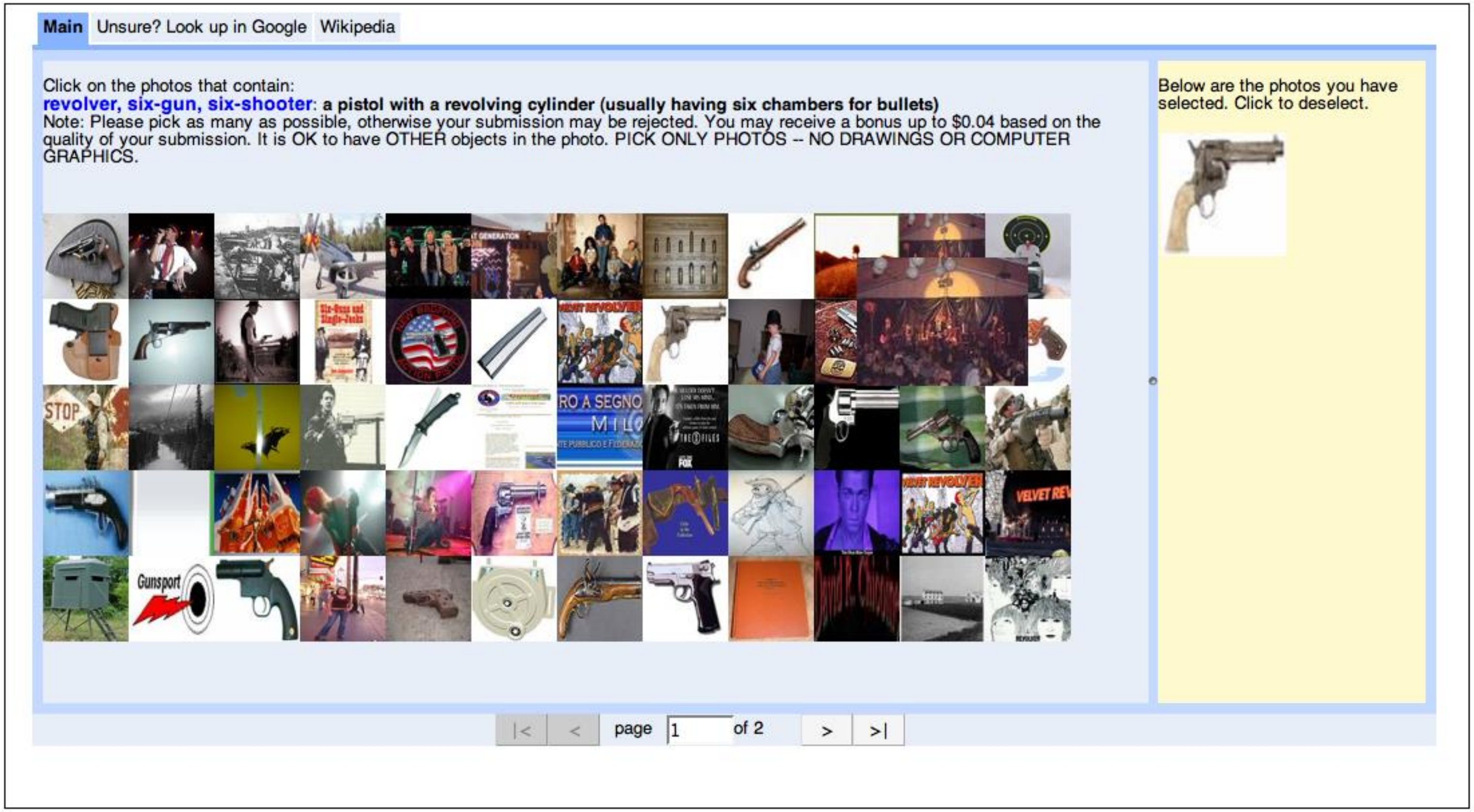
http://austinsmoke.com/turk/.

Select examples

Joint work with Tamara and Alex Berg

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

Select examples



$0.02

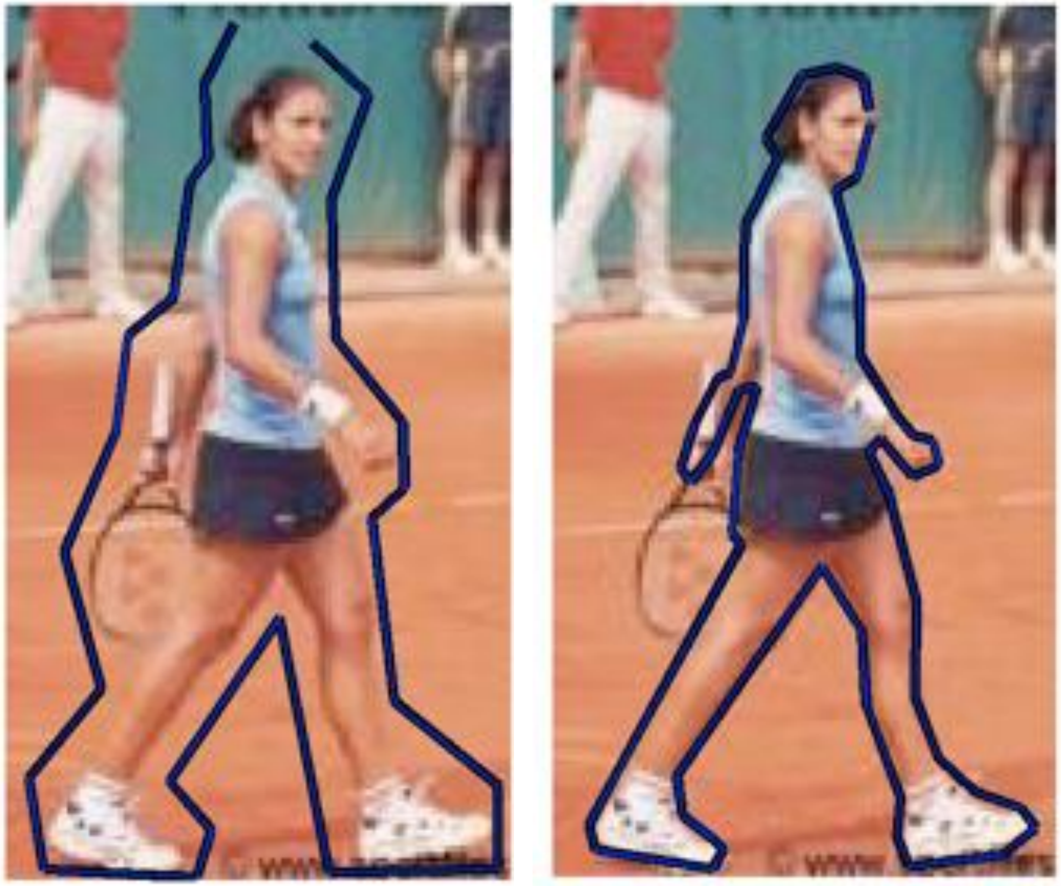
requester mtlabel

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

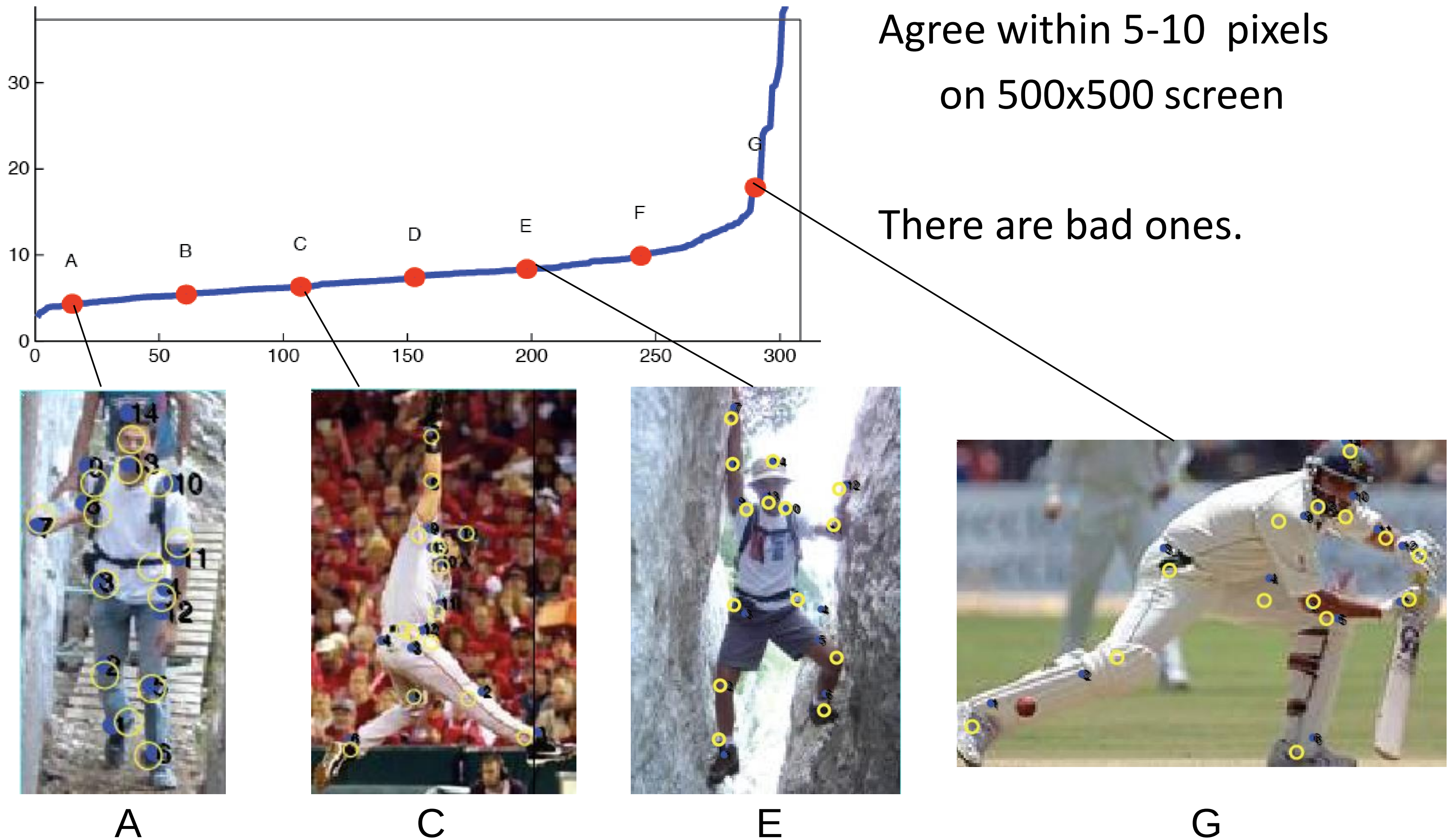
Large scale

Low price

Issues

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

Annotation quality

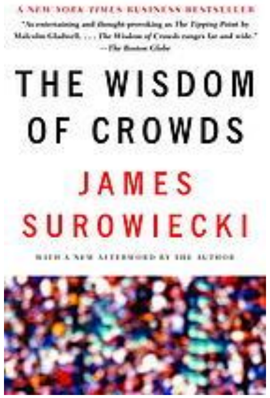


Agree within 5-10 pixels
on 500x500 screen

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

Examples of Crowdsourcing

- Most papers annotate images, but there are some more creative uses
  - Webcam Eye tracking (https://webgazer.cs.brown.edu/)
  - 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*

Outline

- Data collection with experts – PASCAL VOC
- Annotation with non-experts
  - LabelMe
  - ESP Game
  - Mechanical Turk
- Human-in-the-loop Recognition
  - Visipedia

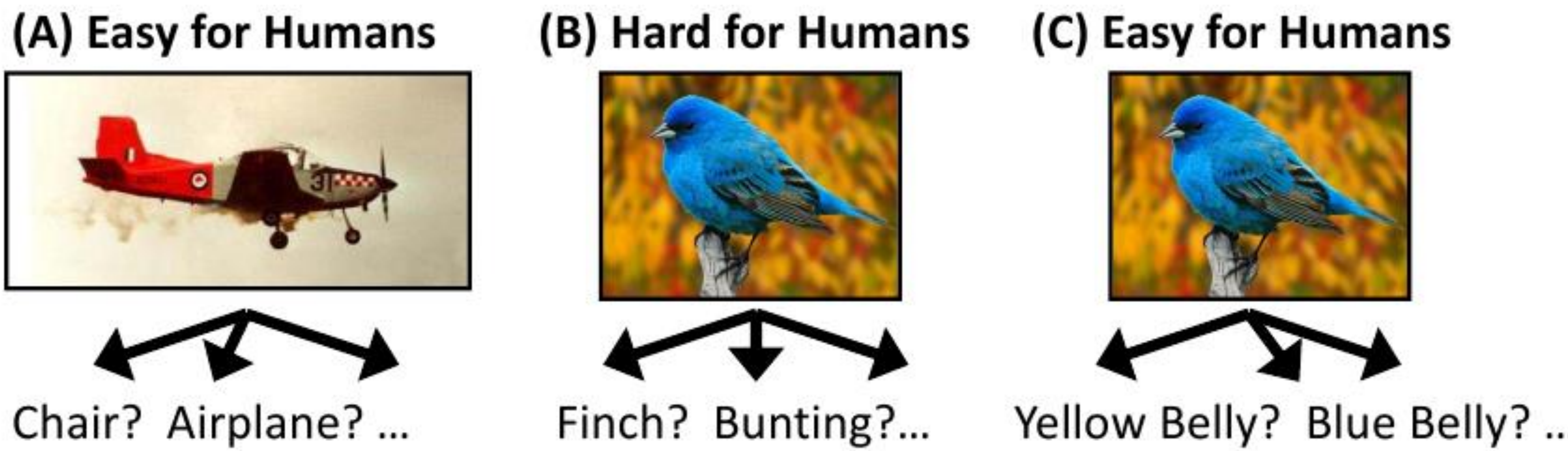
Visual Recognition with Humans in the Loop

Steve Branson, Catherine Wah, Florian Schroff, Boris Babenko, Peter Welinder, Pietro Perona, Serge Belongie

Part of the [Visipedia project]()

Slides from Brian O'Neil

Introduction:



Computers starting to get good at this.

If it's hard for humans, it's probably too hard for computers.

Semantic feature extraction difficult for computers.

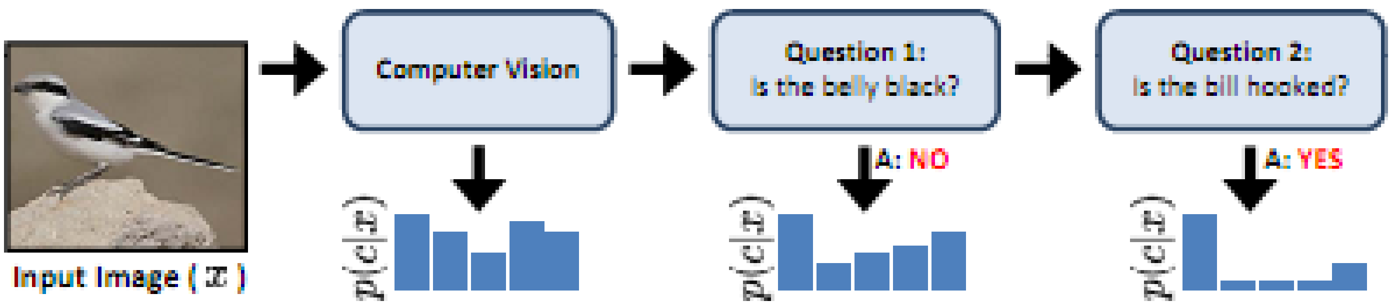
Combine strengths to solve this problem.

The Approach: What is progress?

- Supplement visual recognition with the human capacity for visual feature extraction to tackle difficult (fine-grained) recognition problems.
- Typical progress is viewed as increasing data difficulty while maintaining full autonomy
- Here, the authors view progress as reduction in human effort on difficult data.

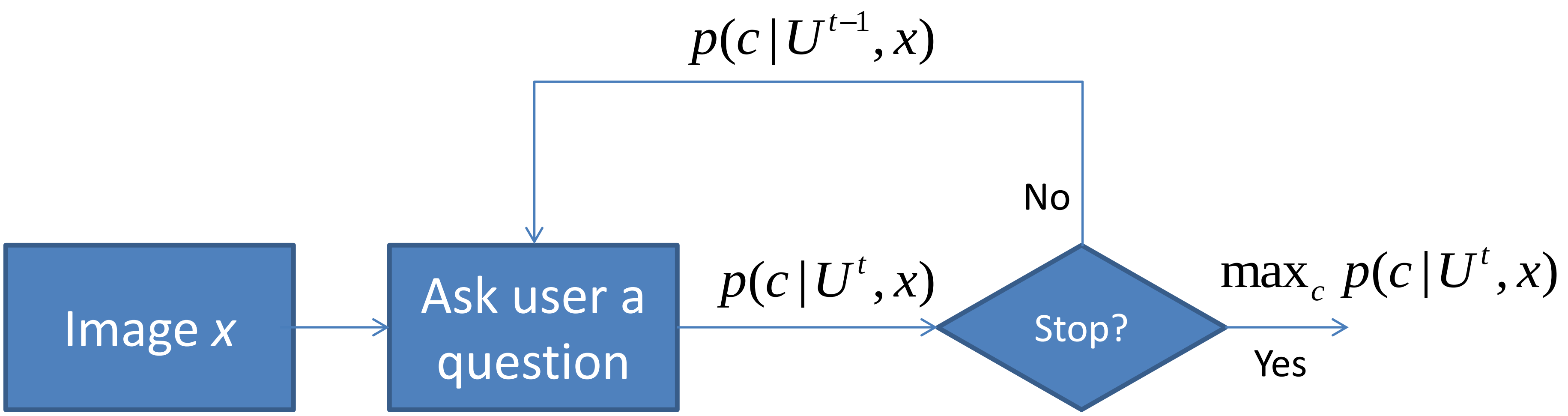
The Approach: 20 Questions

- Ask the user a series of discriminative visual questions to make the classification.



Which 20 questions?

- At each step, exploit the image itself and the user response history to select the most informative question to ask next.



Which question to ask?

- The question that will reduce entropy the most, taking into consideration the computer vision classifier confidences for each category.

The Dataset: Birds-200

- 6033 images of 200 species





Implementation

amazonmechanical turk

- Assembled 25 visual questions encompassing 288 visual attributes extracted from www.whatbird.com
- Mechanical Turk users asked to answer questions and provide confidence scores.

User Responses.

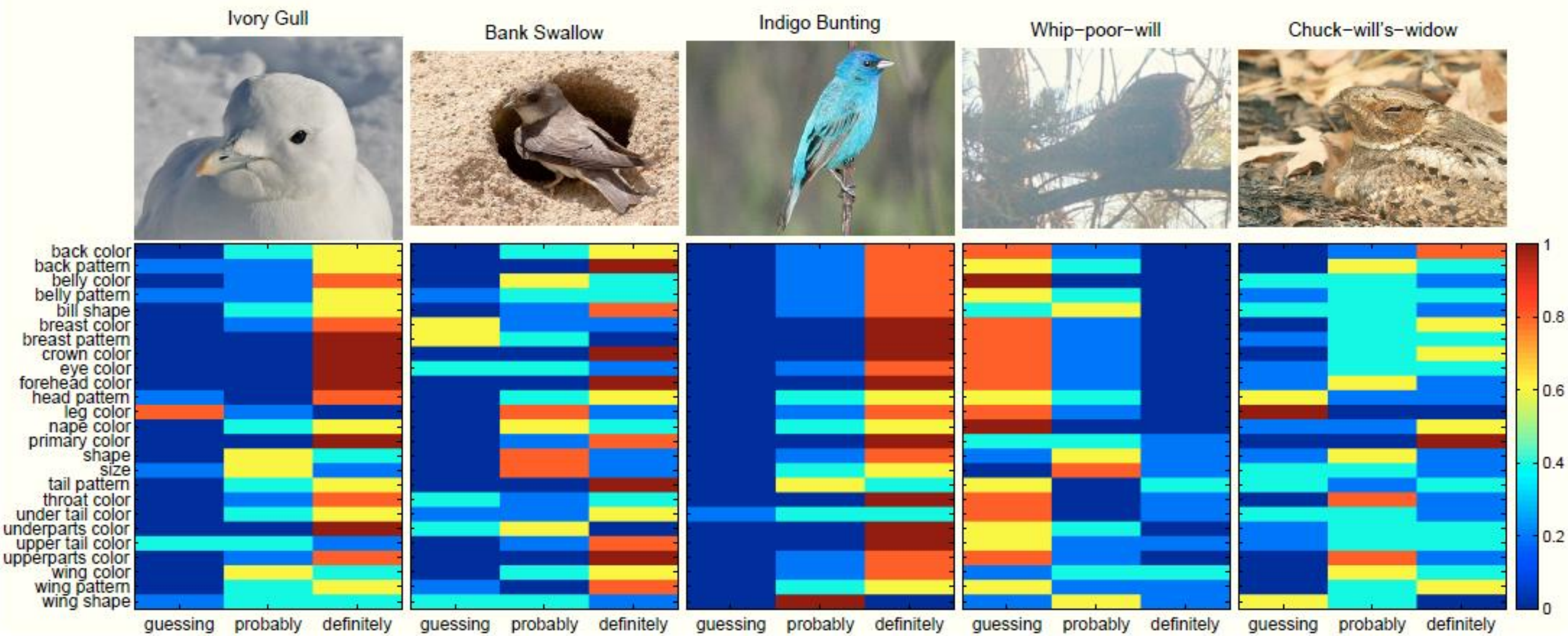
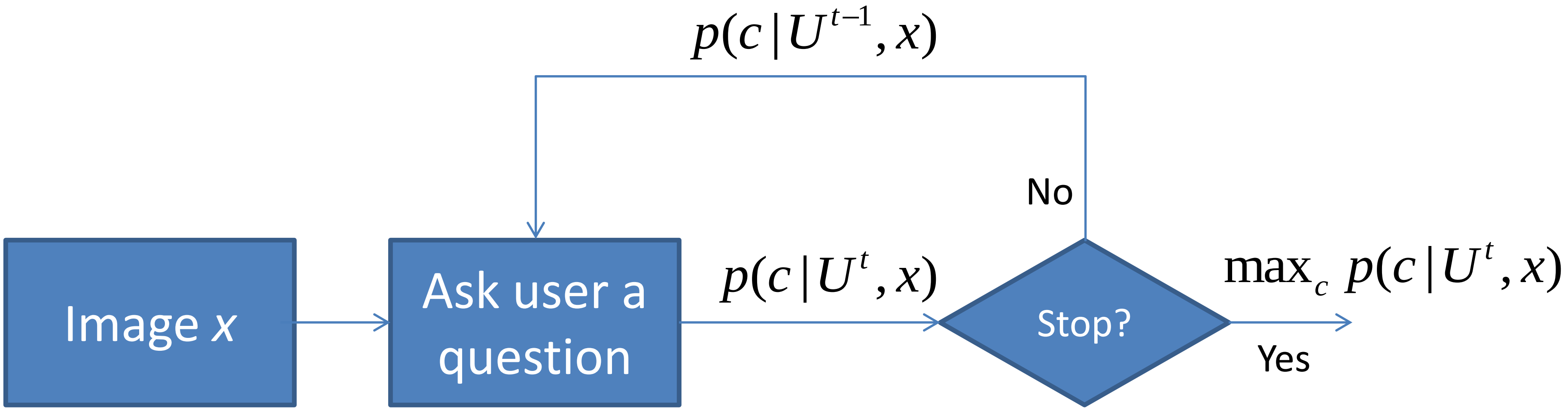


Fig. 4. Examples of user responses for each of the 25 attributes. The distribution over $\{Guessing, Probably, Definitely\}$ is color coded with blue denoting 0% and red denoting 100% of the five answers per image attribute pair.

Visual recognition

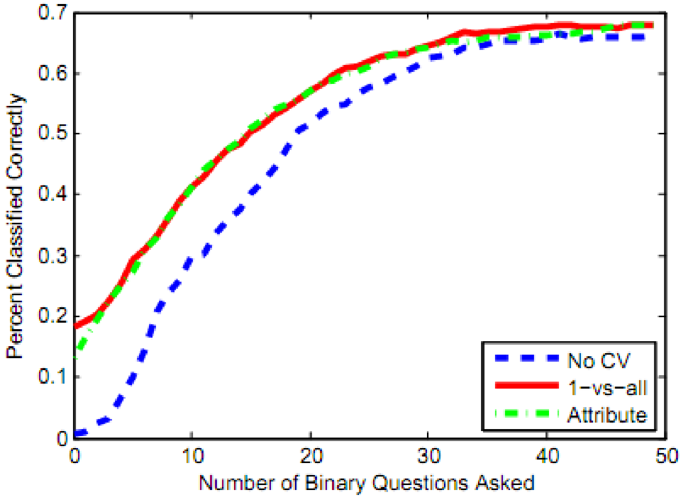
- Any vision system that can output a probability distribution across classes will work.
- Authors used Andrea Vedaldis's code.
  - Color/gray SIFT
  - VQ geometric blur
  - 1 v All SVM
- Authors added full image color histograms and VQ color histograms

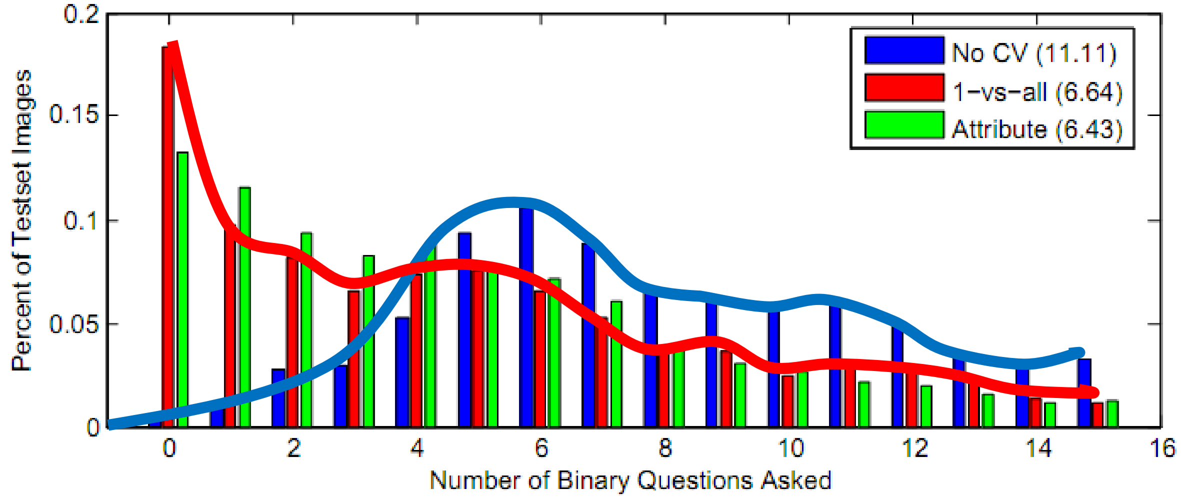
Experiments



- 2 Stop criteria:
  - Fixed number of questions – evaluate accuacy
  - User stops when bird identified – measure number of questions required.

Results





- Average number of questions to make ID reduced from 11.11 to 6.43
- Method allows CV to handle the easy cases, consulting with users only on the more difficult cases.

Key Observations

- Visual recognition reduces labor over a pure "20 Q" approach.
- Visual recognition improves performance over a pure "20 Q" approach. (69% vs 66%)
- User input dramatically improves recognition results. (66% vs 19%)

Strengths and weaknesses

- Handles very difficult data and yields excellent results.
- Plug-and-play with many recognition algorithms.
- Requires significant user assistance
- Reported results assume humans are perfect verifiers
- Is the reduction from 11 questions to 6 really that significant?

Next lecture(s)

- Human-in-the-loop
- Attributes
- More crowdsourcing (ImageNet, MS COCO)