Recap: HOGgles

- Data, Representation, and Learning matter.
  - This work looked just at representation
- By creating a human-understandable HoG visualization, we can see why detectors make certain mistakes.
  - False positives from overly strong normalization are common
- Visualization isn’t perfect! Missing high freq.

Project 4 steps

- Train a face-like classifier
  – How to sample negative training data?
- Test classifier at a single scale
- Add non-maximum suppression
- Test classifier at multiple scales
  – *One* call to non maximum suppression per image.

Data Sets and Crowdsourcing

CS143 Computer Vision

James Hays, Brown University

24 hours of Photo Sharing

installation by Erik Kessels

And sometimes Internet photos have useful labels

Im2gps. Hays and Efros. CVPR 2008

But what if we want more?

Image Categorization

Training

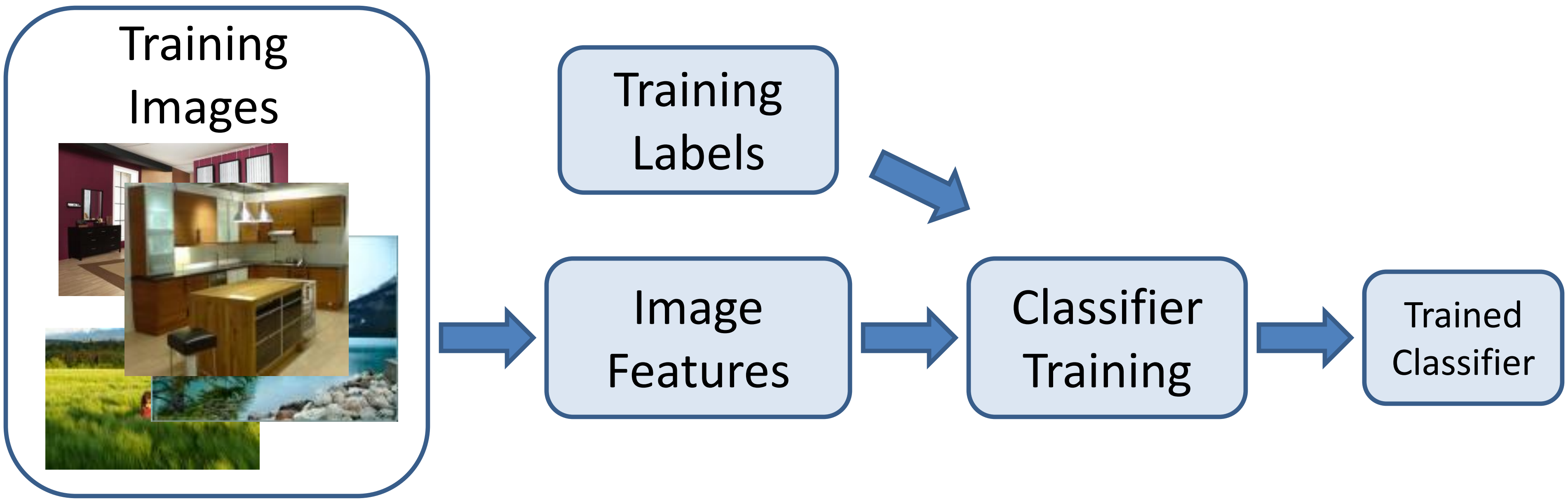
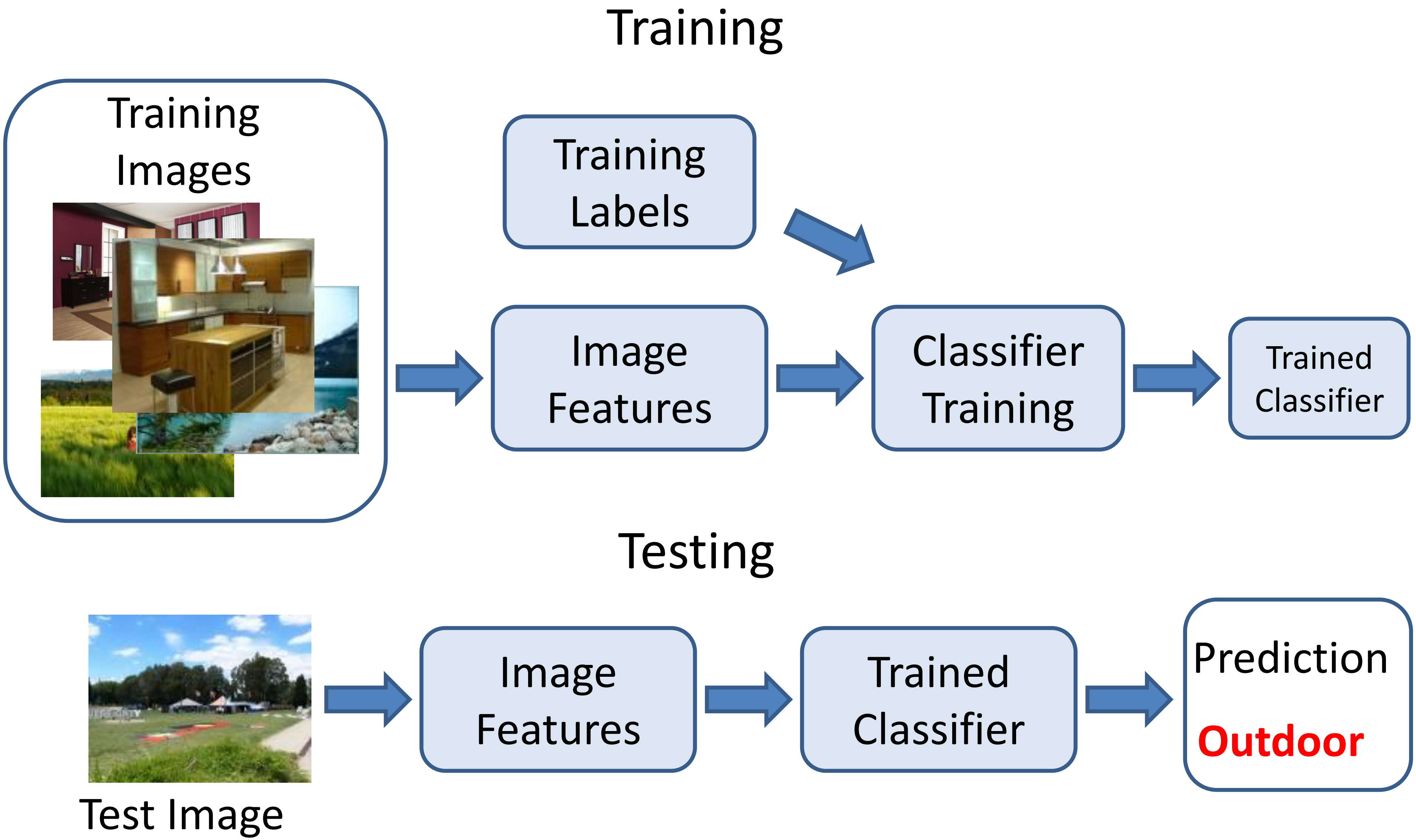
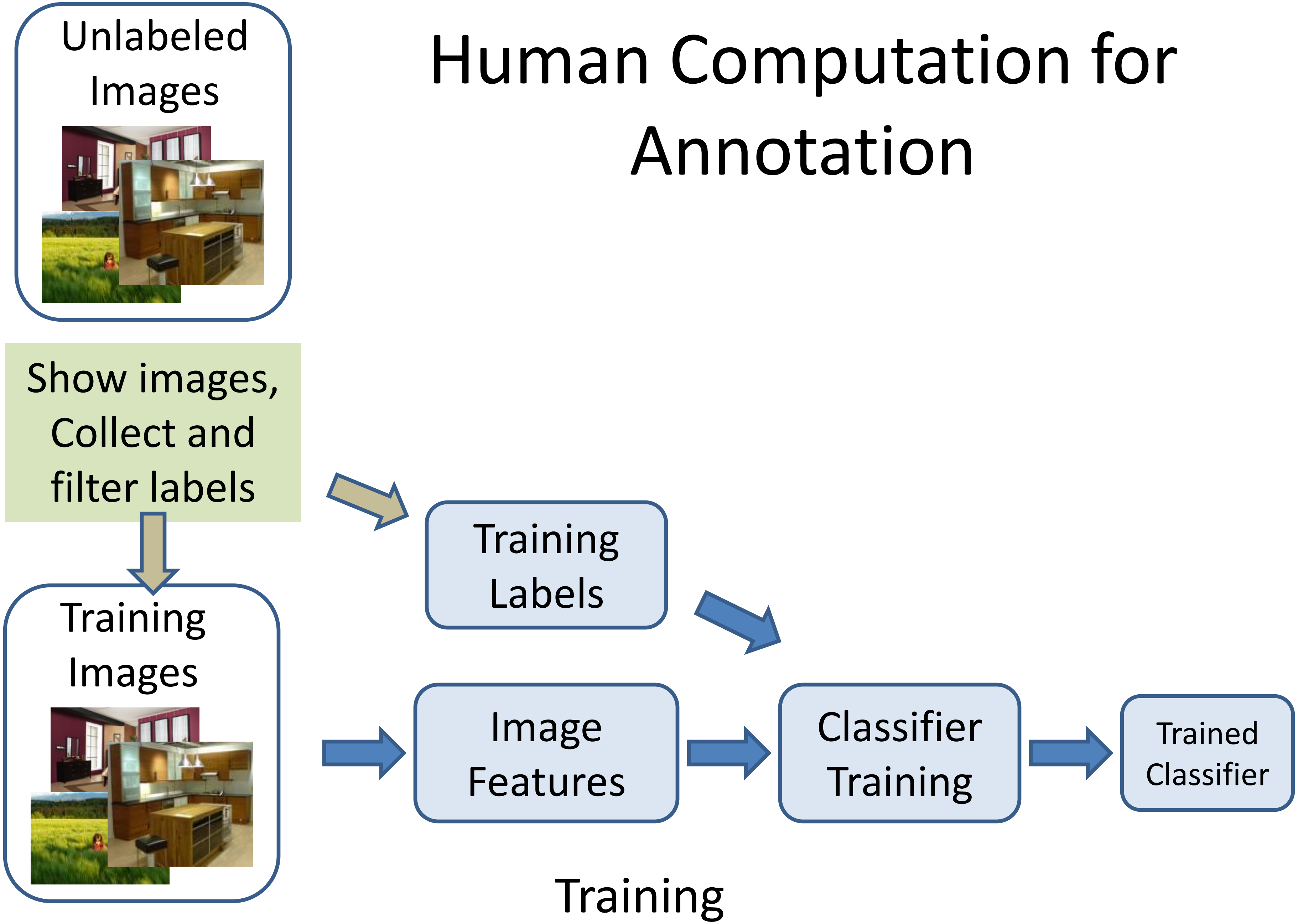


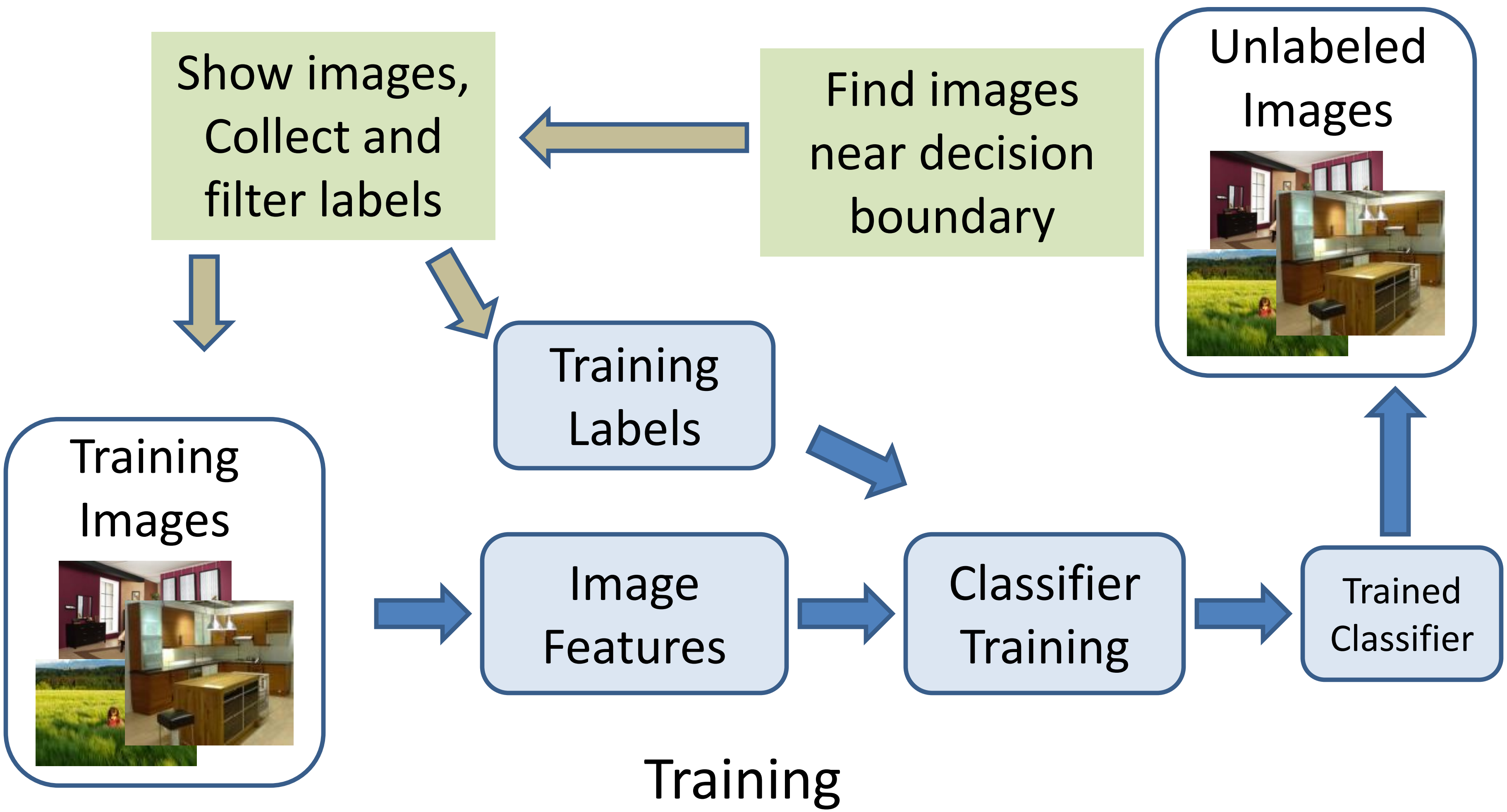
Image Categorization



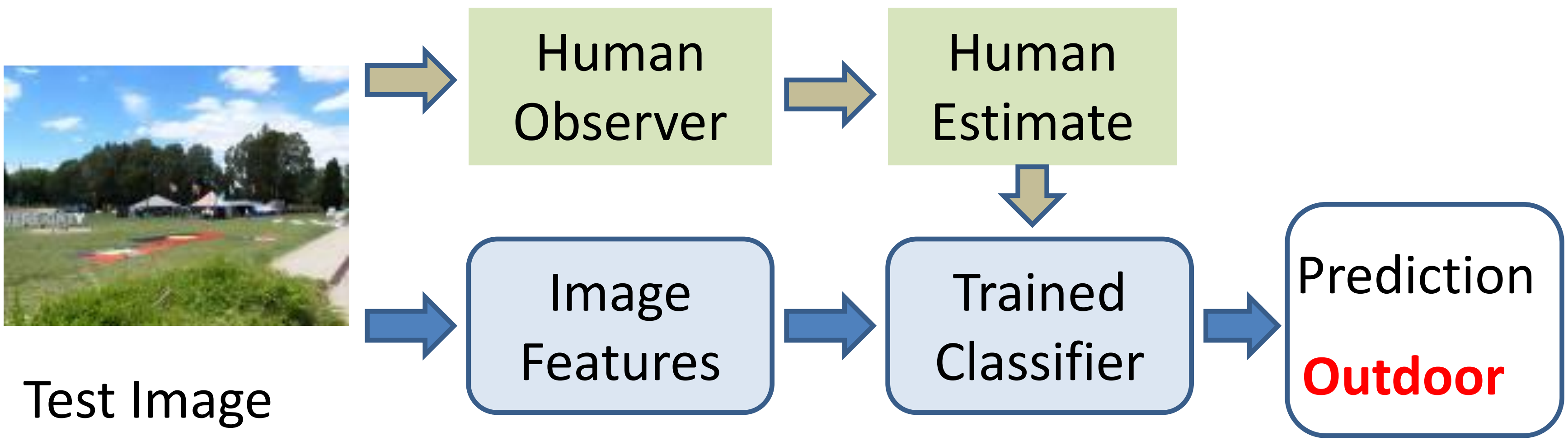
Human Computation for Annotation



Active Learning



Human-in-the-loop Recognition



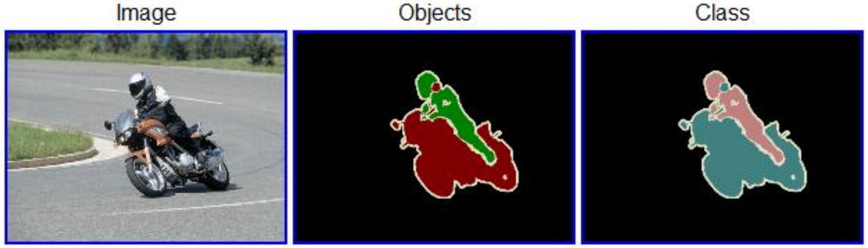
Testing

Outline

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

The PASCAL Visual Object Classes Challenge 2009 (VOC2009)

- Twenty object categories (aeroplane to TV/monitor)
- Three challenges:
  – Classification challenge (is there an X in this image?)
  – Detection challenge (draw a box around every X)
  – Segmentation challenge



Slides from Noah Snavely

Dataset: Collection

- Images downloaded from **flickr**
  - 500,000 images downloaded and random subset selected for annotation

Dataset: Annotation

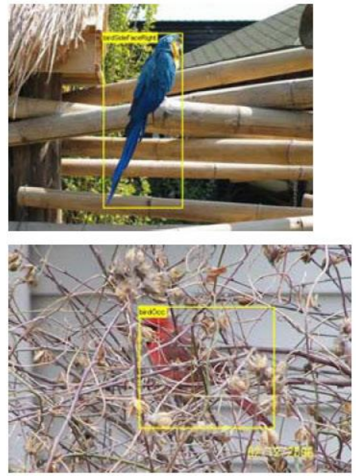
- Complete annotation of all objects
- Annotated over web with <u>written guidelines</u>
  - High quality (?)

Examples

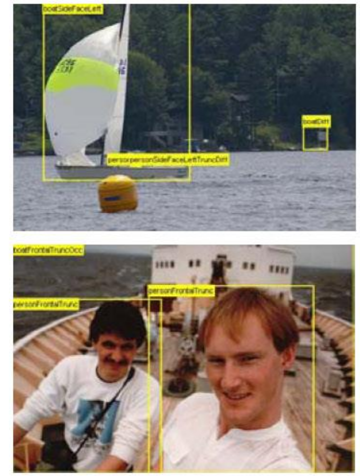
Aeroplane

Bicycle

Bird



Boat



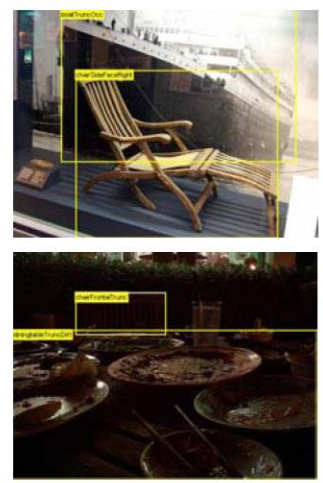
Bottle

Bus

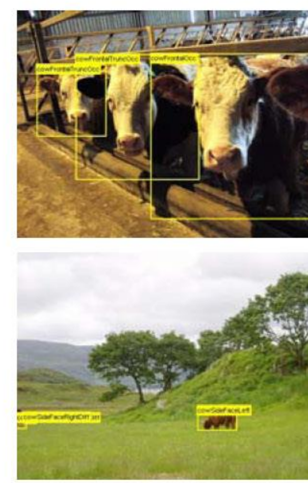
Car

Cat

Chair

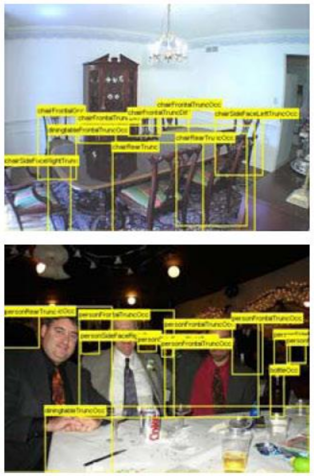


Cow



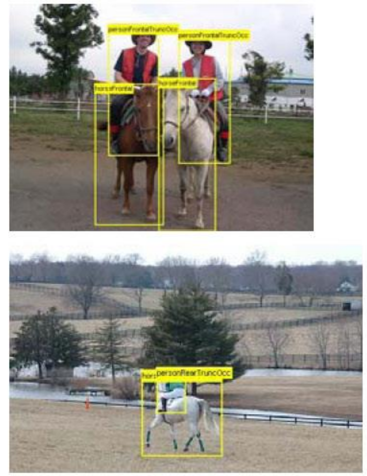
Examples

Dining Table



Dog

Horse

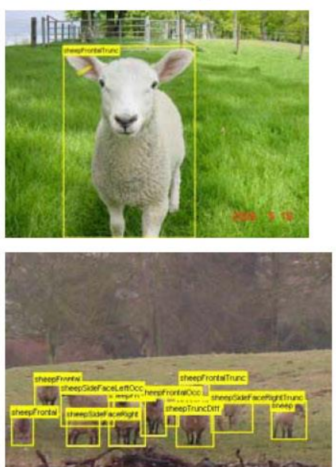


Motorbike

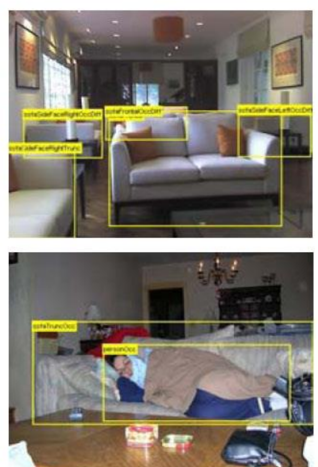
Person

Potted Plant

Sheep



Sofa



Train

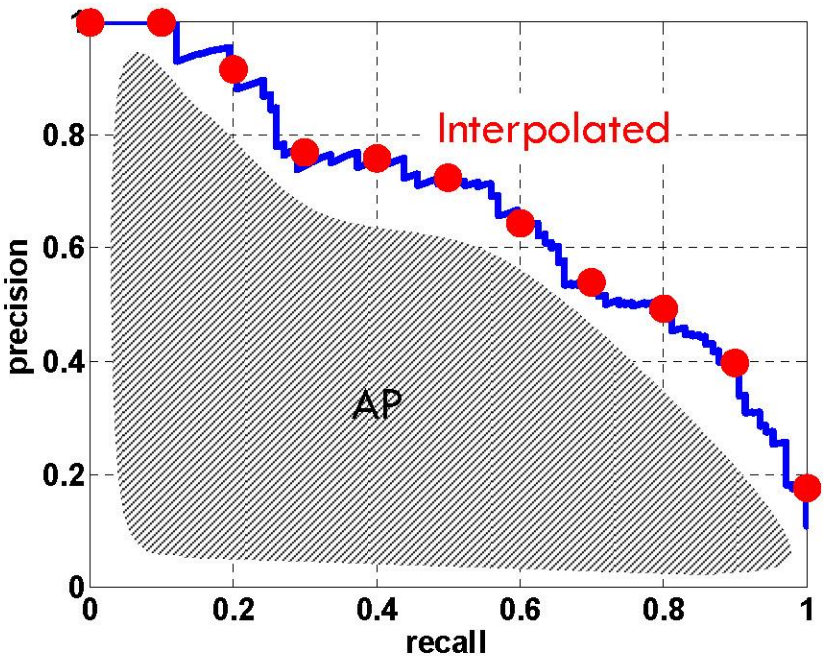
TV/Monitor

Detection Challenge

- Predict the bounding boxes of all objects of a given class in an image (if any)

Evaluation

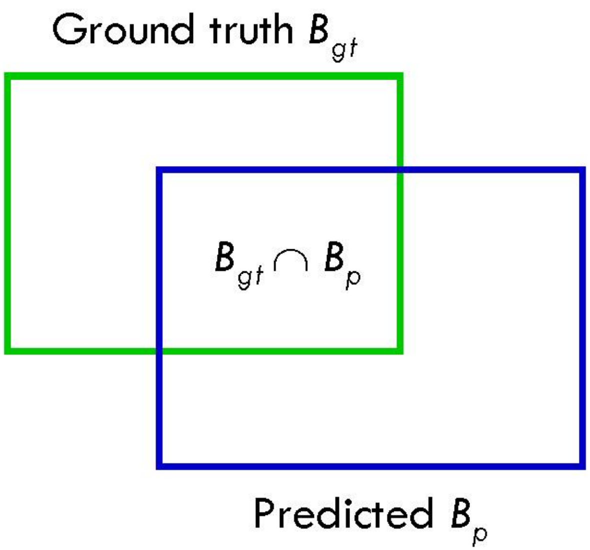
- Average Precision [TREC] averages precision over the entire range of recall
  - Curve interpolated to reduce influence of "outliers"



- A good score requires both high recall and high precision
- Application-independent
- Penalizes methods giving high precision but low recall

Evaluating Bounding Boxes

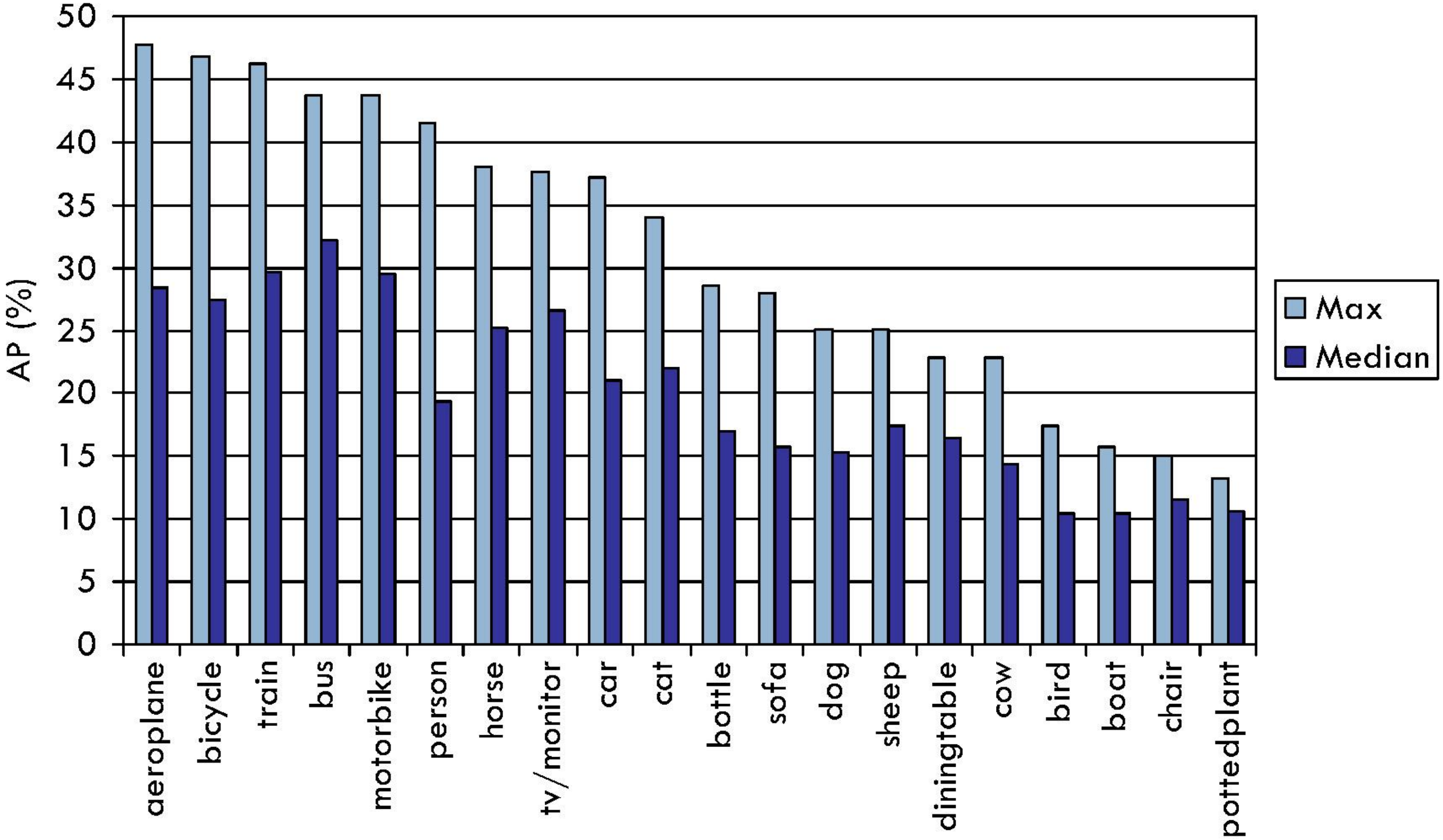
- Area of Overlap (AO) Measure



$$AO(B_{gt}, B_p) = \frac{|B_{gt} \bigcap B_p|}{|B_{gt} \bigcup B_p|}$$

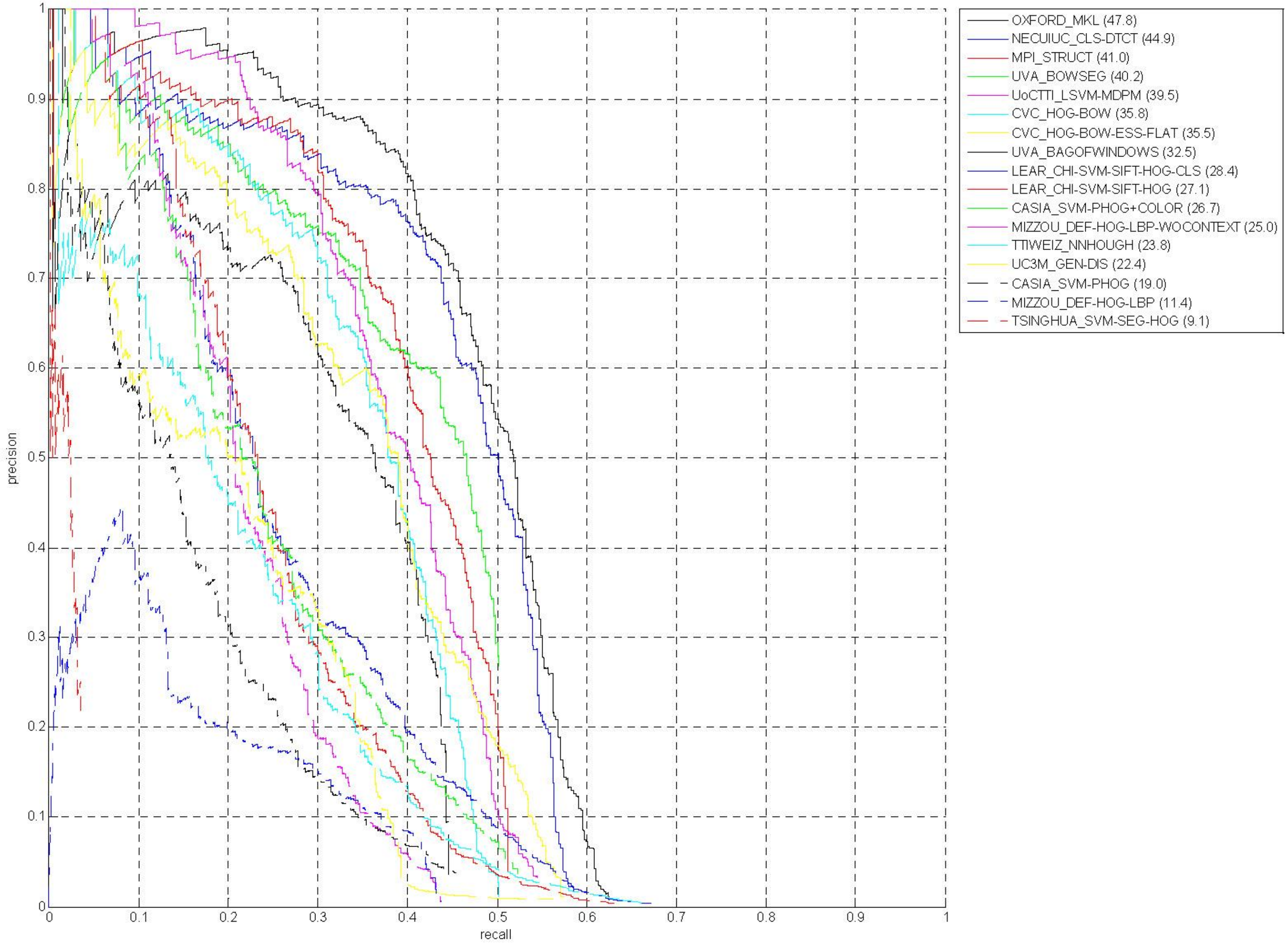
- Need to define a threshold t such that $AO(B_{gt},B_p)$ implies a correct detection: 50%

AP by Class

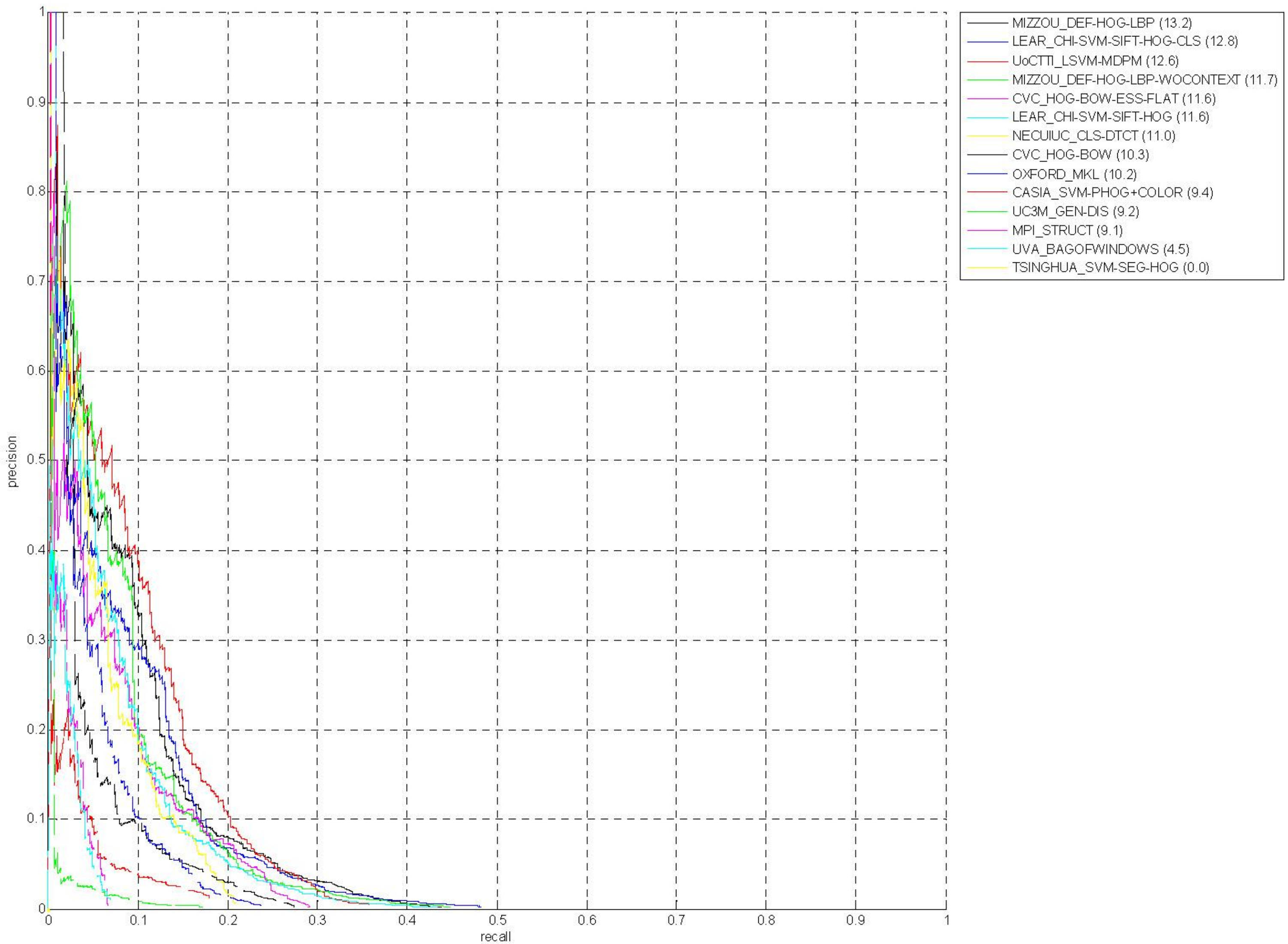


Chance essentially 0

Precision/Recall - Aeroplane



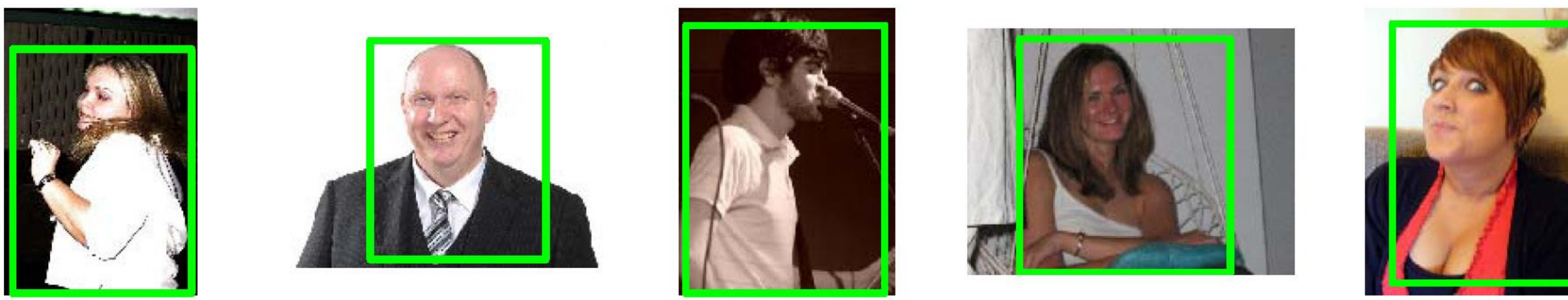
Precision/Recall – Potted plant



True Positives - Person

UoCTTI_LSVM-MDPM

MIZZOU_DEF-HOG-LBP

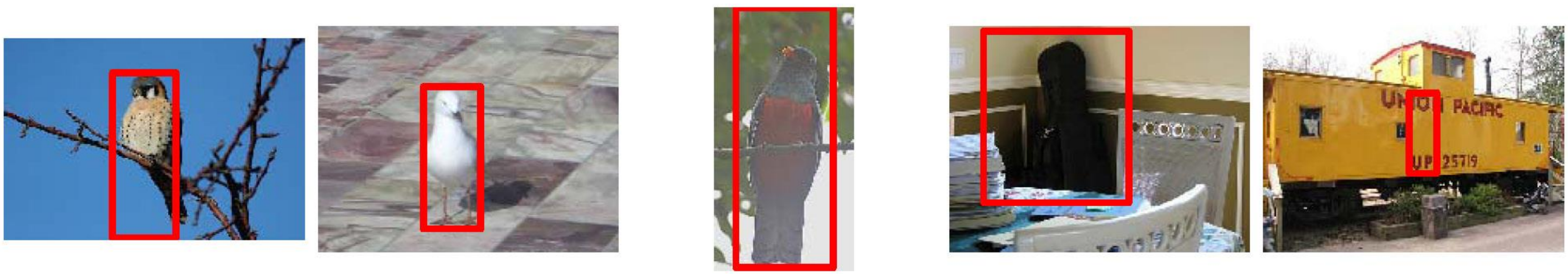


NECUIUC_CLS-DTCT

False Positives - Person

UoCTTI_LSVM-MDPM

MIZZOU_DEF-HOG-LBP



NECUIUC_CLS-DTCT

“Near Misses” - Person

UoCTTI_LSVM-MDPM

MIZZOU_DEF-HOG-LBP

NECUIUC_CLS-DTCT

True Positives - Bicycle

UoCTTI_LSVM-MDPM

OXFORD_MKL

NECUIUC_CLS-DTCT

False Positives - Bicycle

UoCTTI_LSVM-MDPM

OXFORD_MKL

NECUIUC_CLS-DTCT

Outline

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

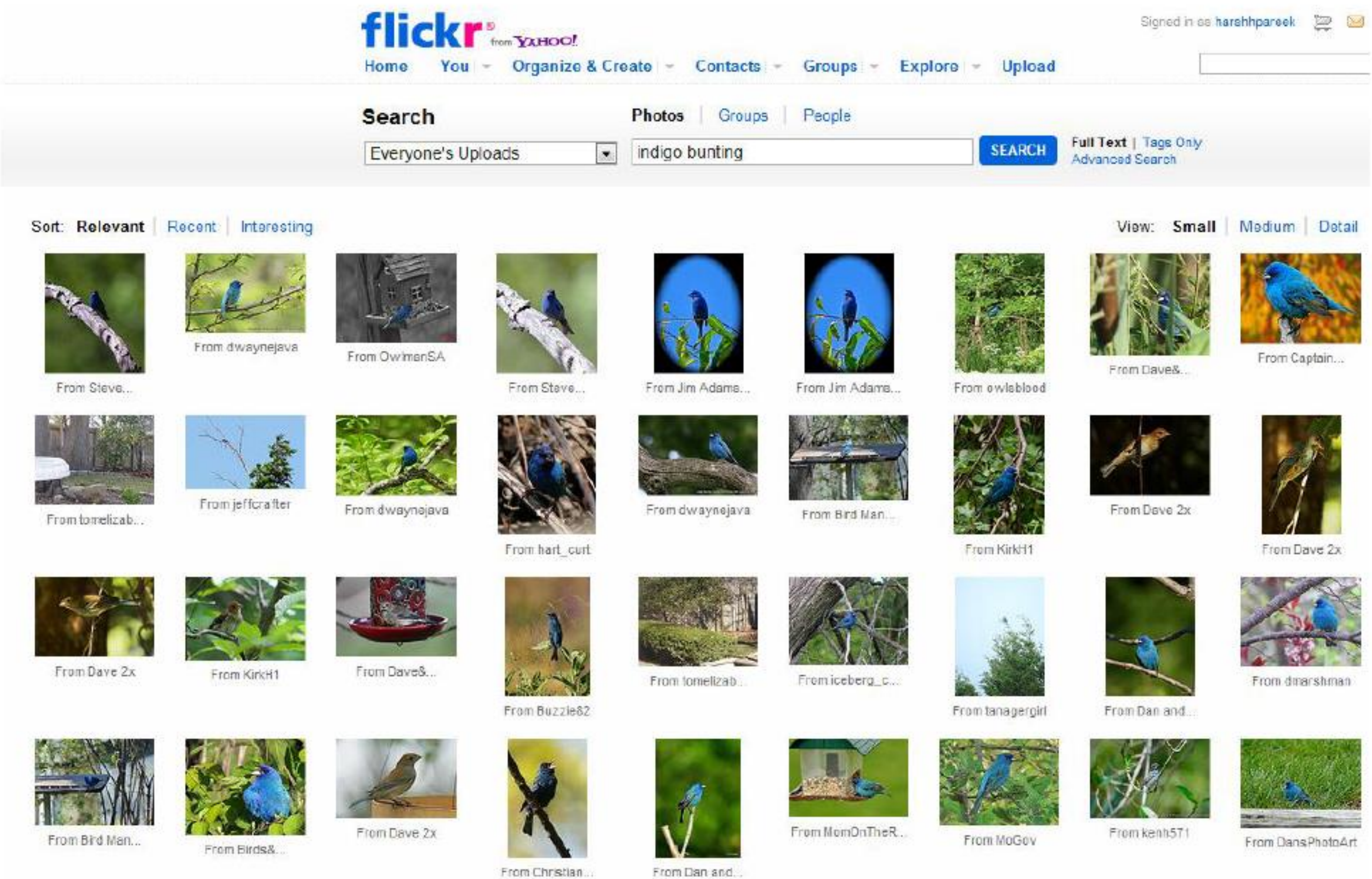


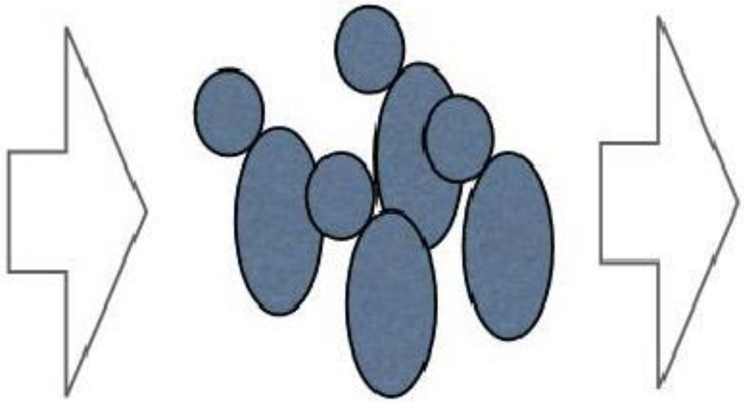
Image credit: Flickr.com

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

Building datasets

6000 images
from flickr.com

Annotators



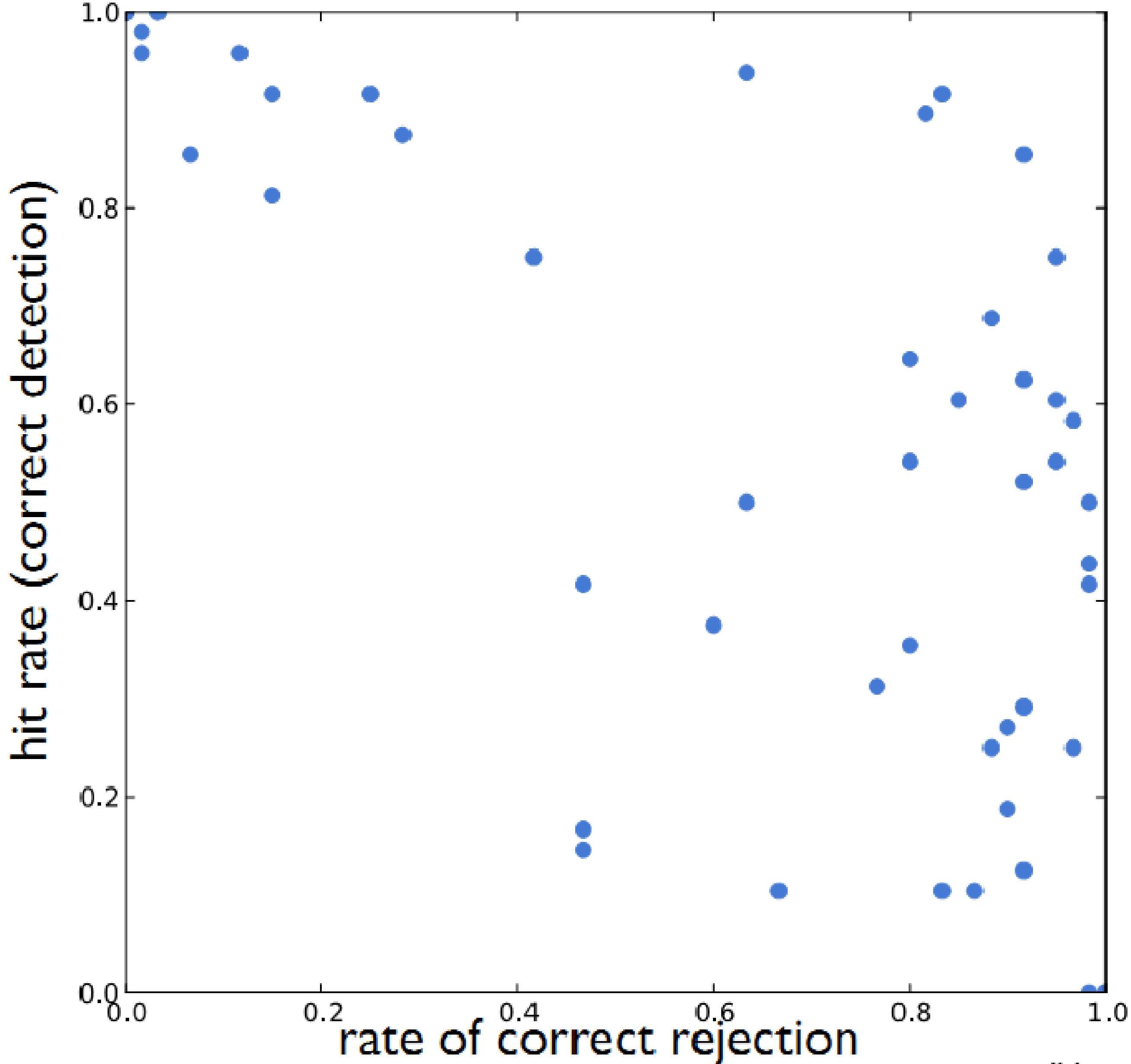
amazonmechanical turk
Artificial Artificial Intelligence

Is there an Indigo bunting in the image?

100s of
training images

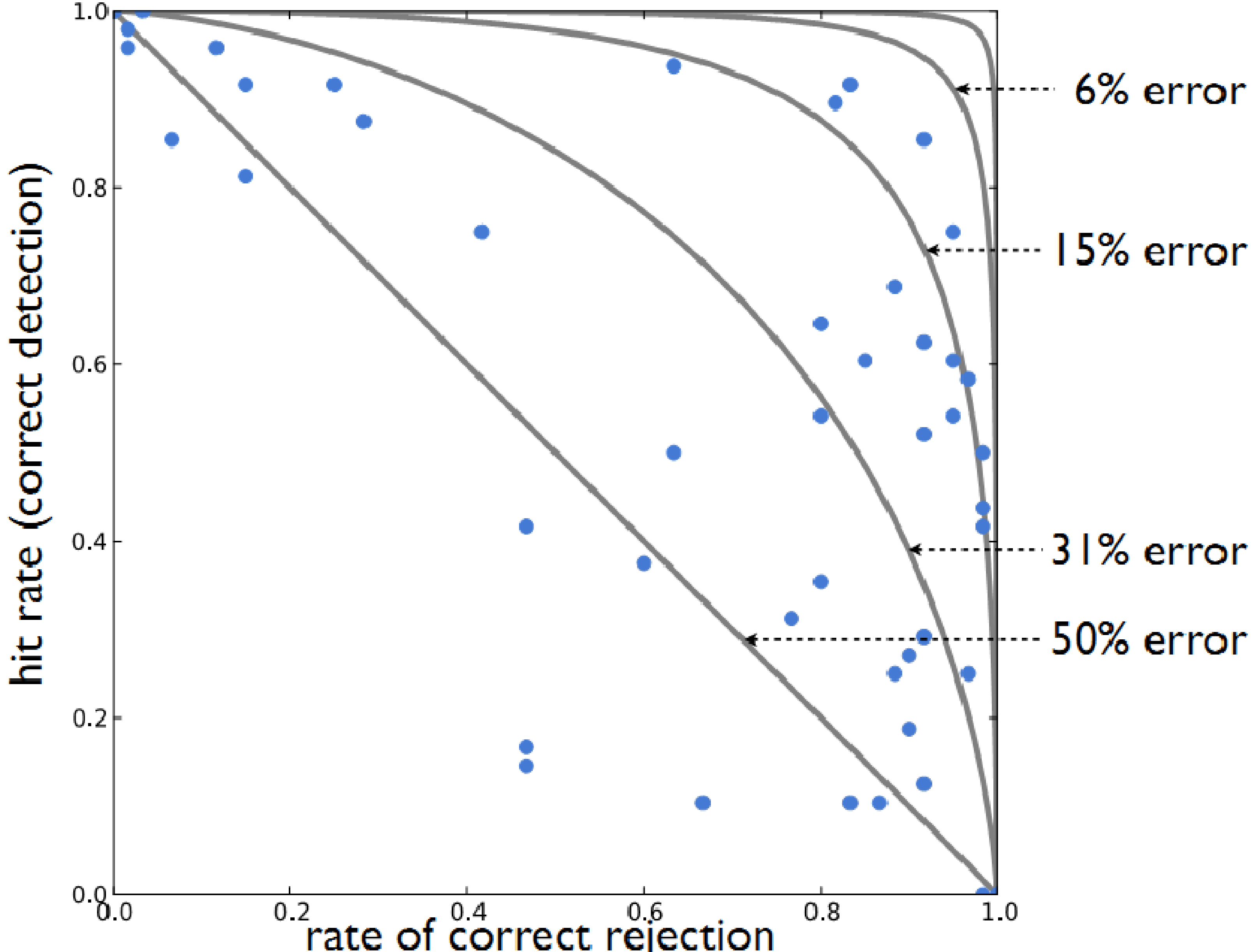
Slide credit: Welinder et al

Task: Find the Indigo Bunting



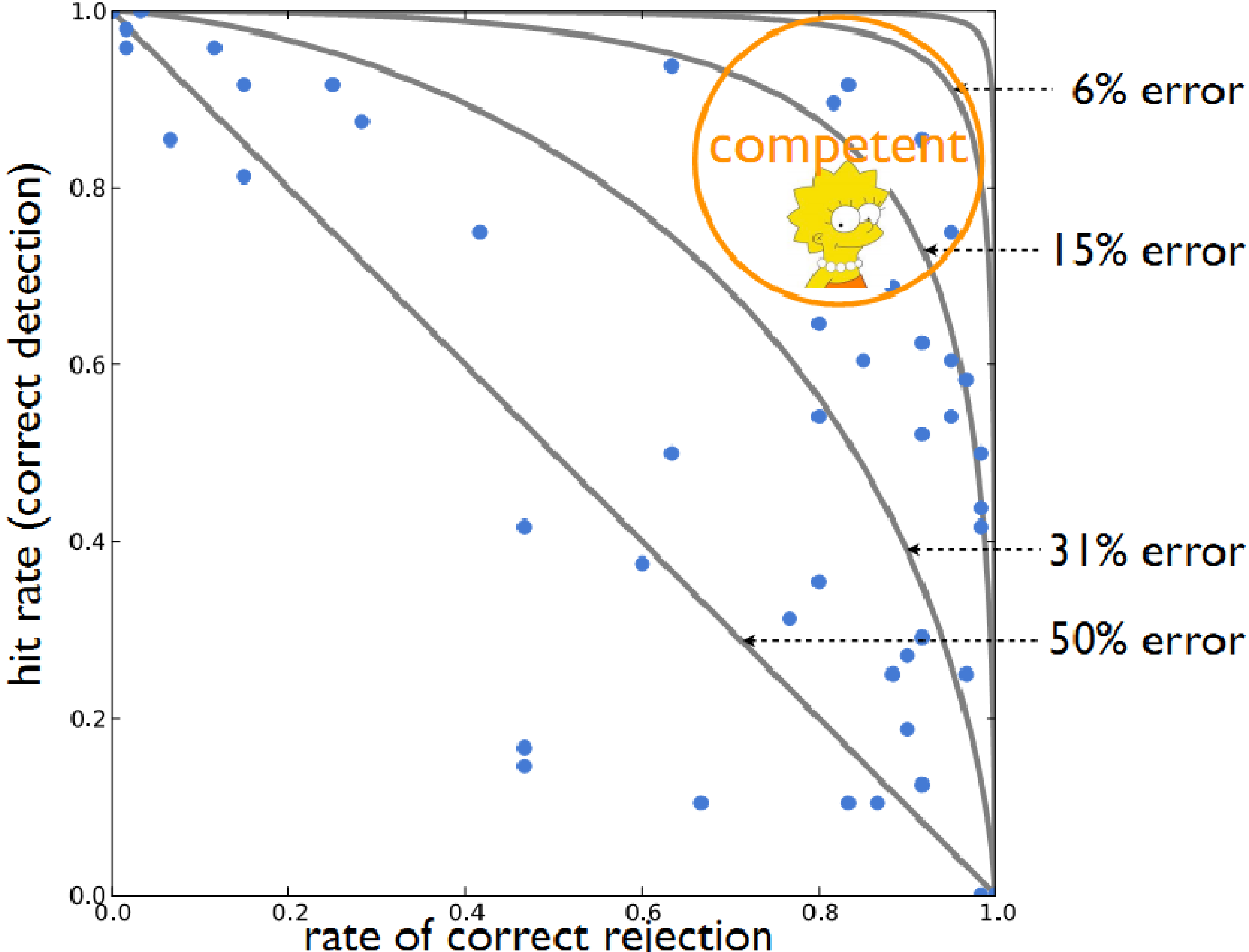
Slide credit: Welinder et al

Task: Find the Indigo Bunting



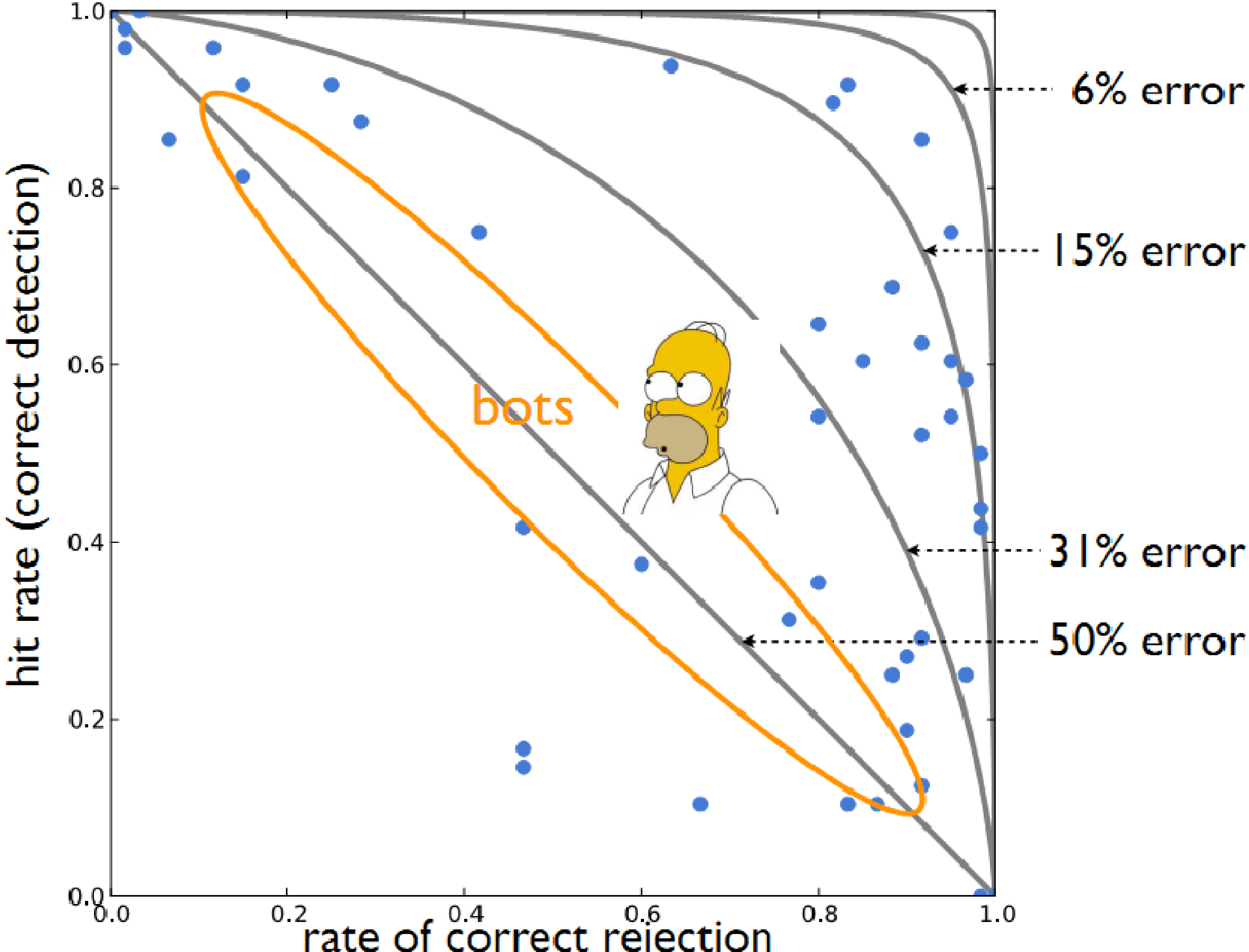
Slide credit: Welinder et al

Task: Find the Indigo Bunting



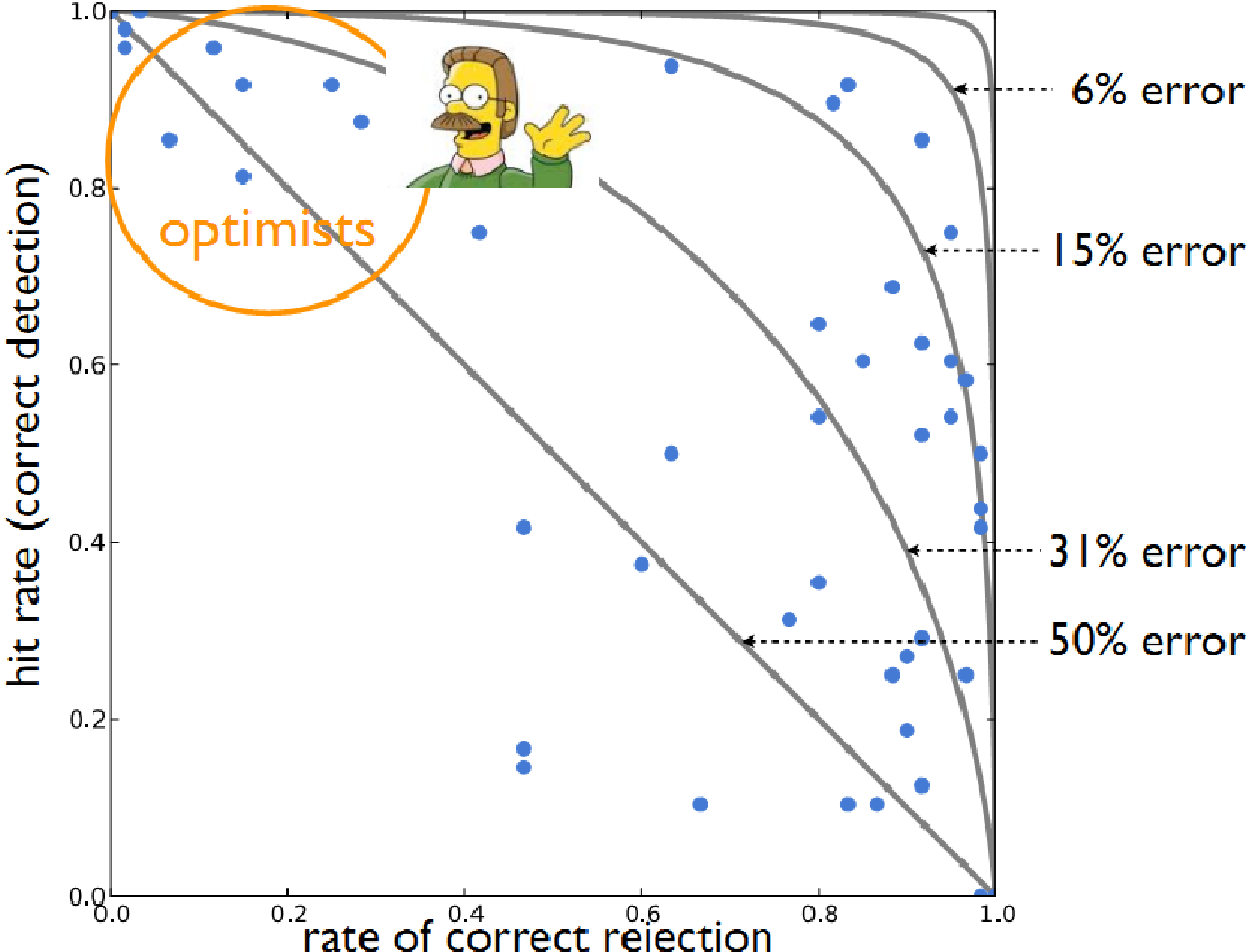
Slide credit: Welinder et al

Task: Find the Indigo Bunting



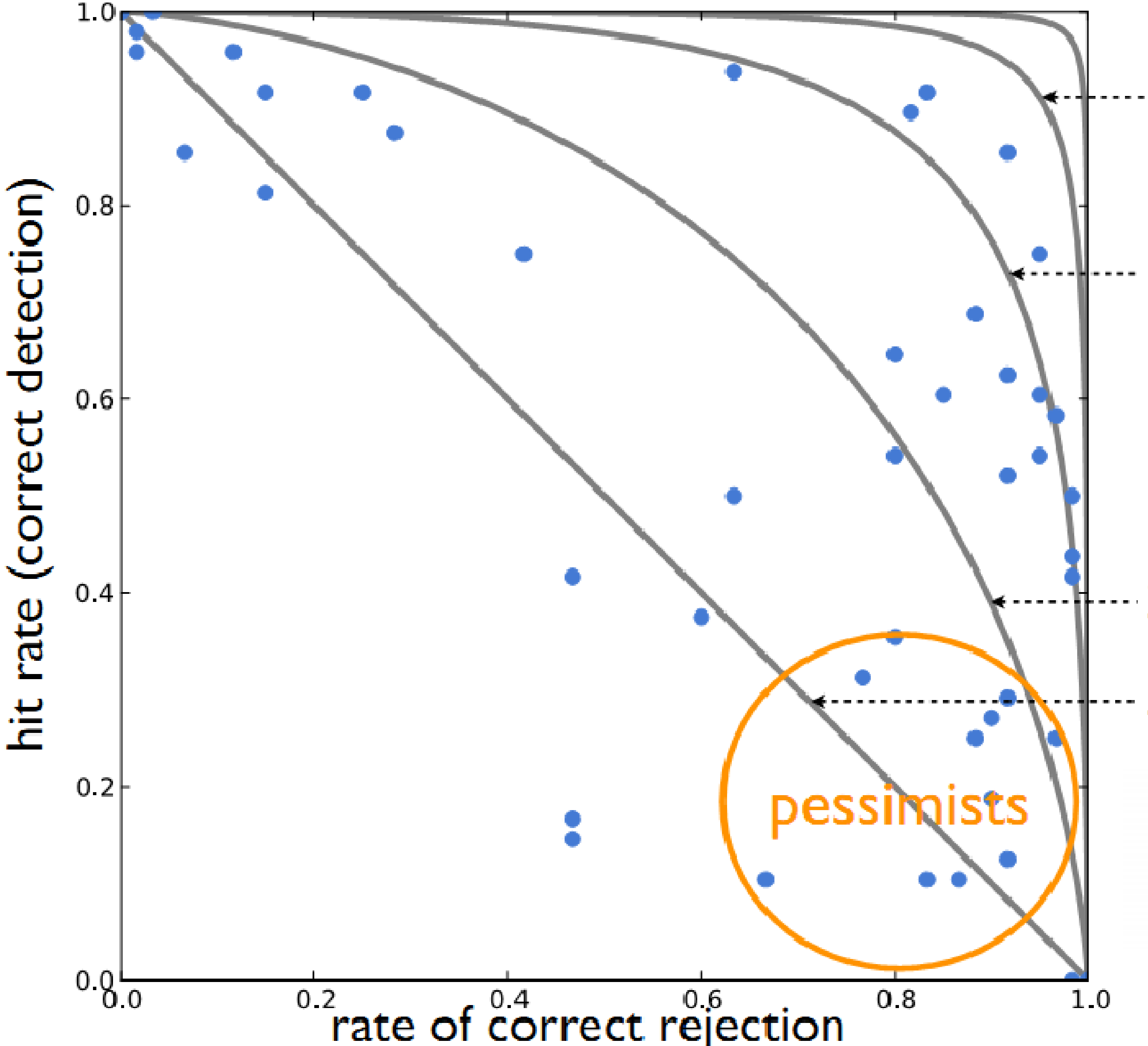
Slide credit: Welinder et al

Task: Find the Indigo Bunting



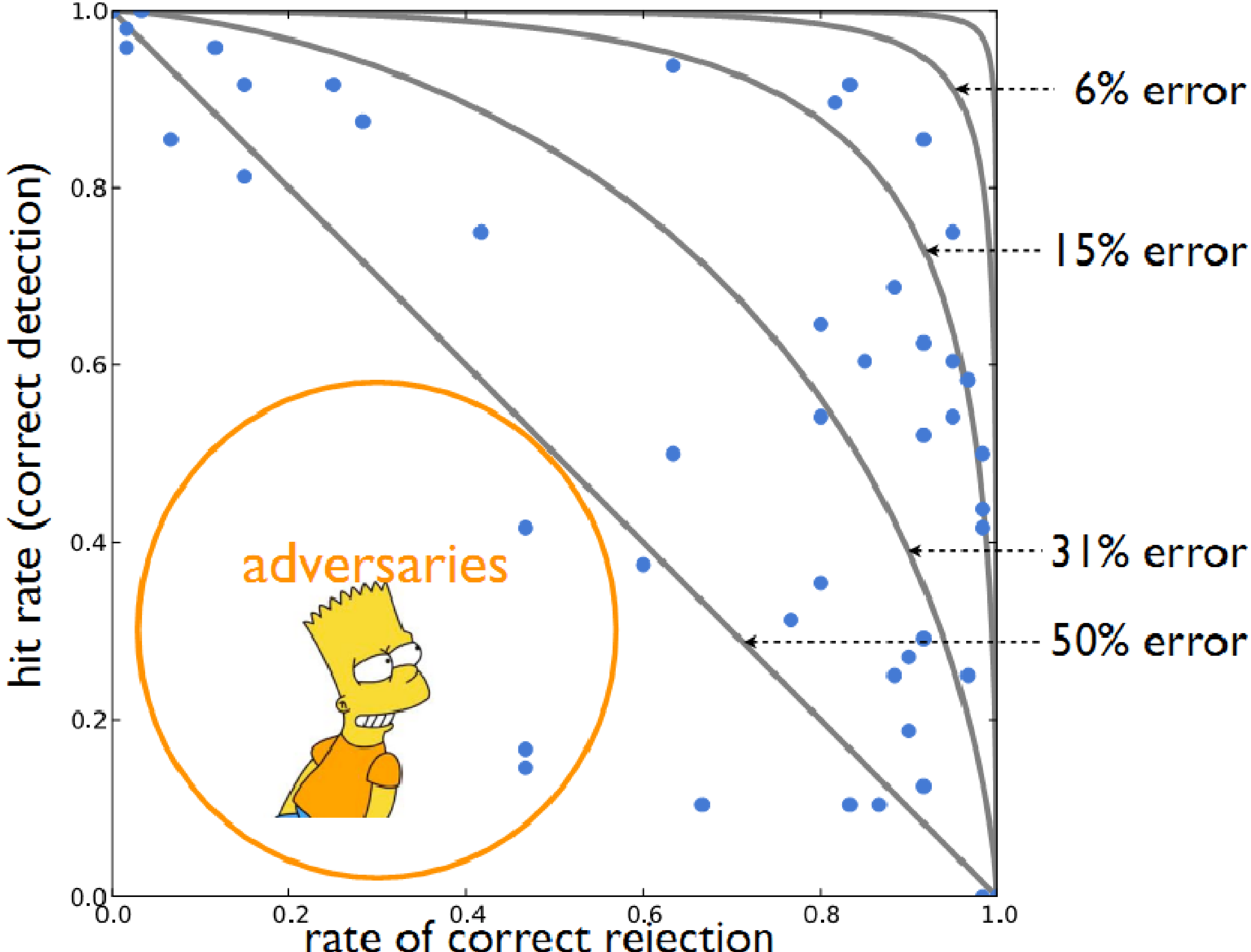
Slide credit: Welinder et al

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

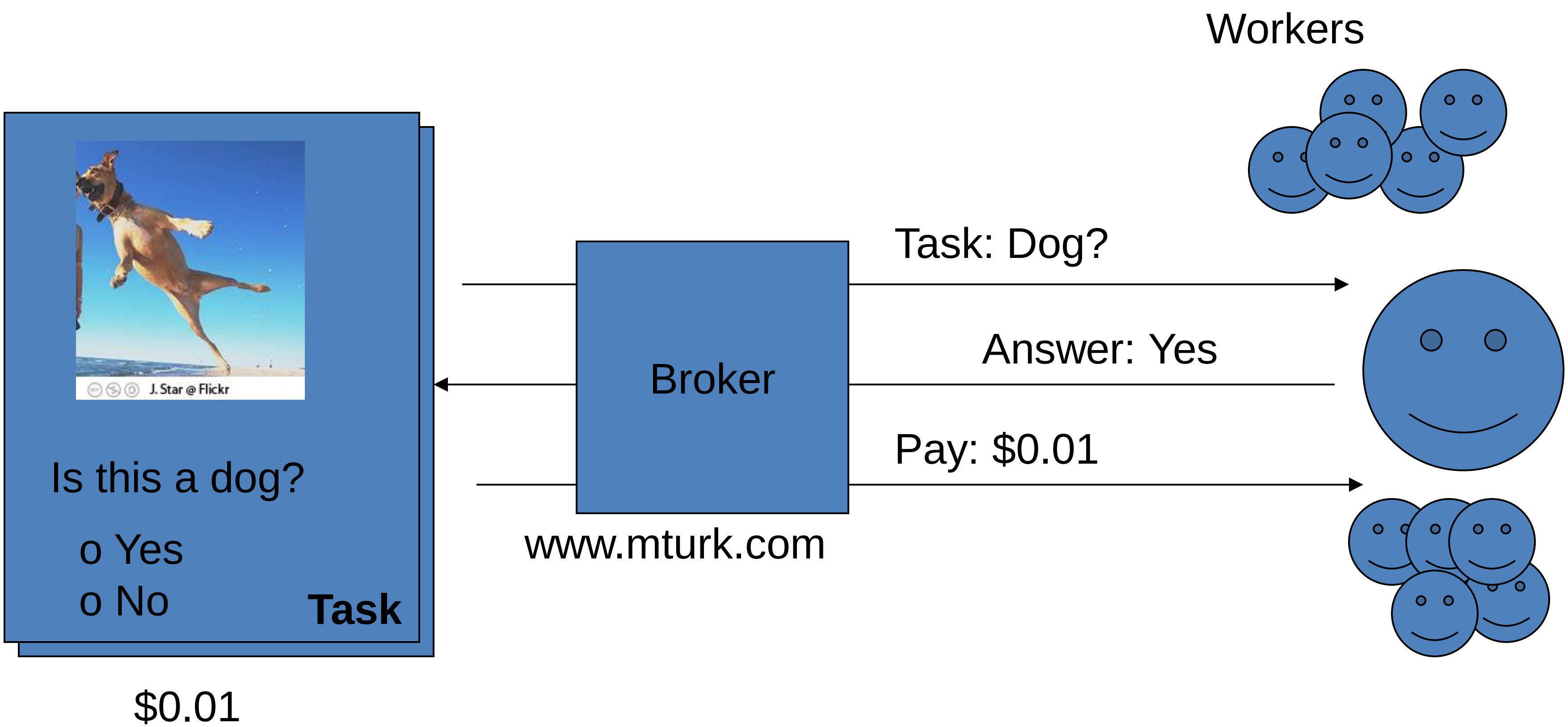
Alexander Sorokin

David Forsyth

University of Illinois at Urbana-Champaign

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

Mechanical Turk Project

If you're using the turk, Be sure to copy the text back into the HIT page so that you can be credited.

- ○ Photo should be rotated 90 degrees left (counter-clockwise)
- ○ Photo should be rotated 90 degrees right (clockwise)
- ○ Photo should be turned upside down
- ● Photo is oriented properly

Please describe the picture in the box using 10 words or more:

shells

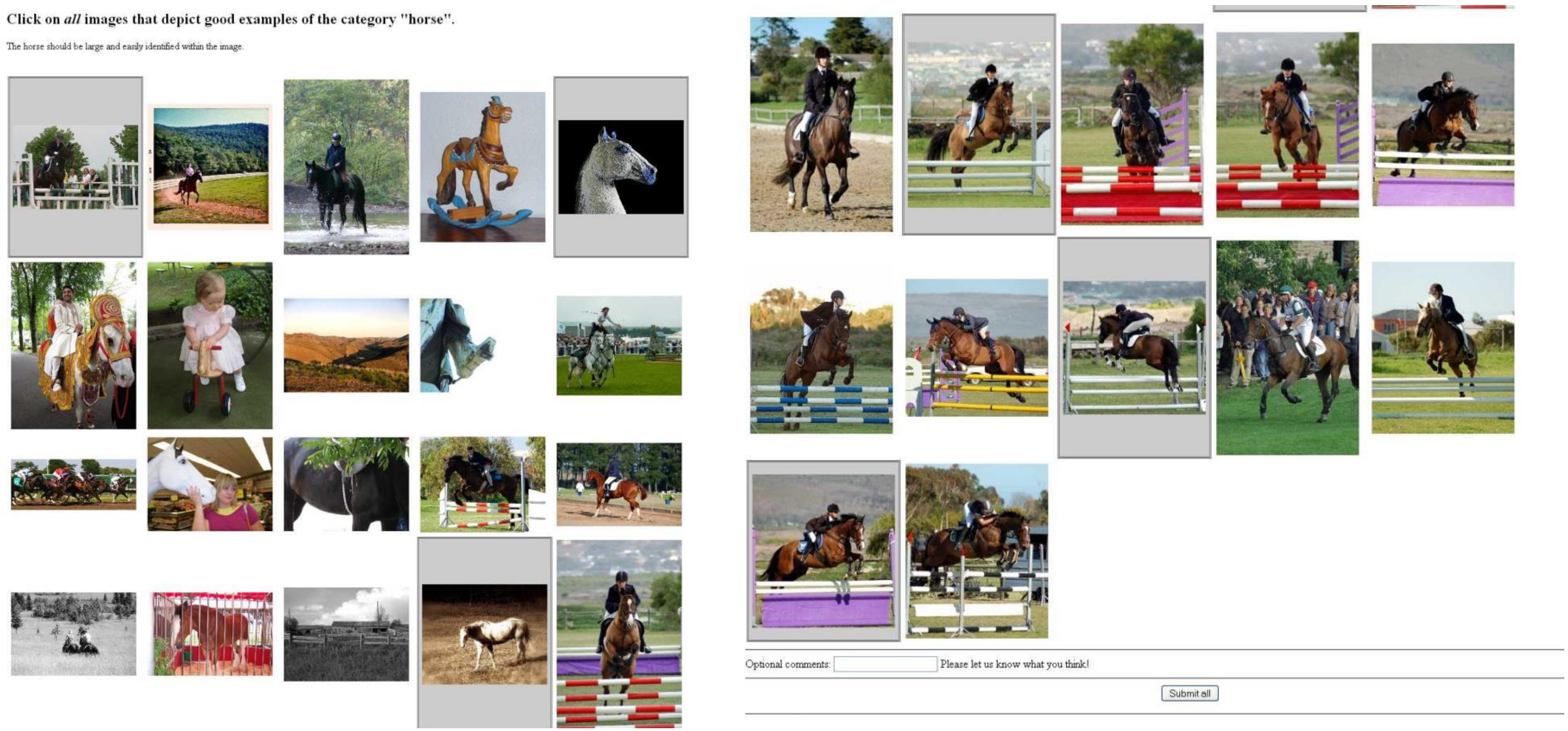
Submit Turk | Skip / Load a different photo

The submit button MUST be clicked!

$0.01

http://austinsmoke.com/turk/.

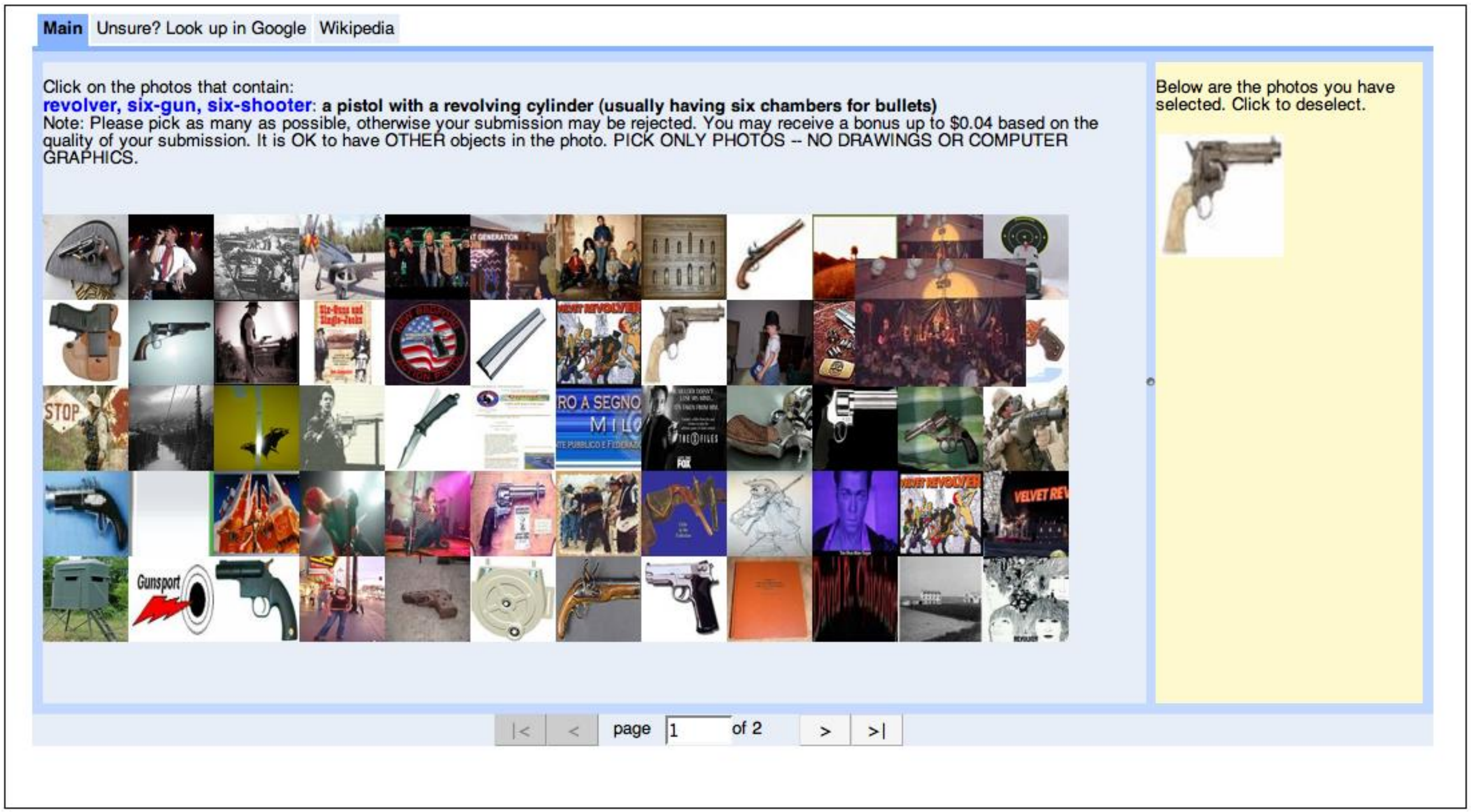
Select examples



Joint work with Tamara and Alex Berg

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

Select examples



$0.02

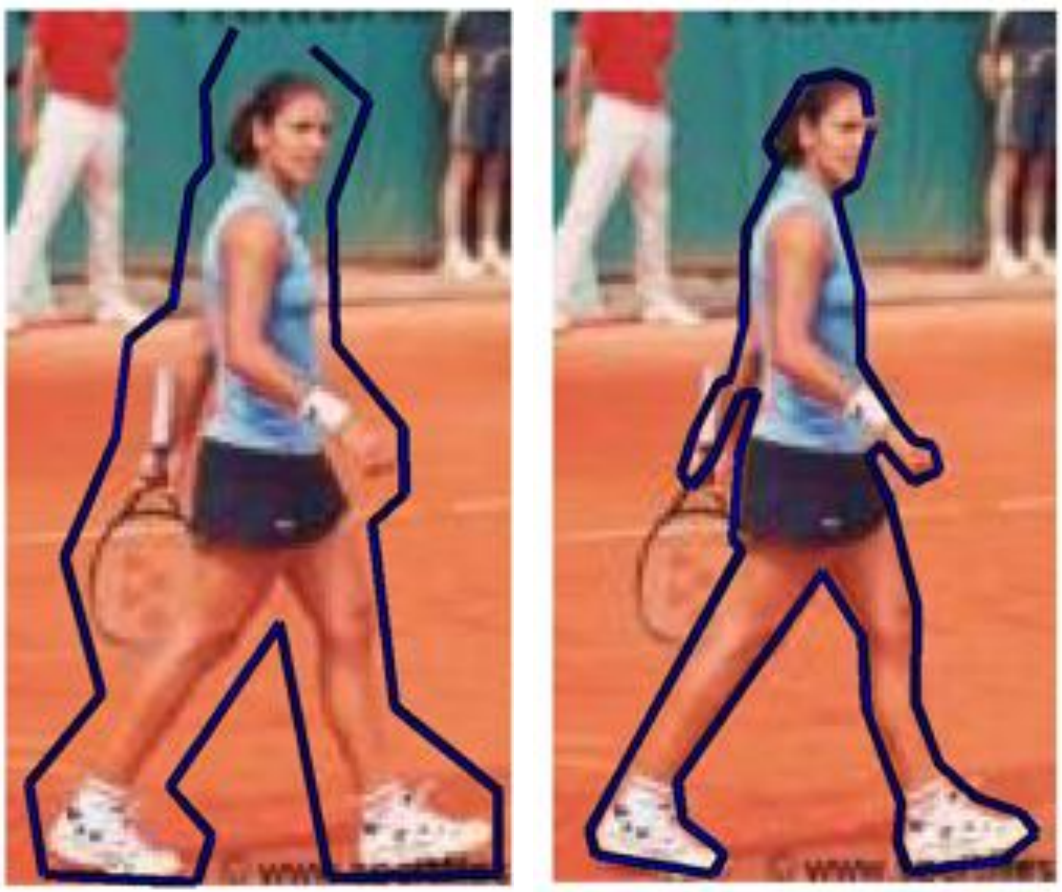
requester mtlabel

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

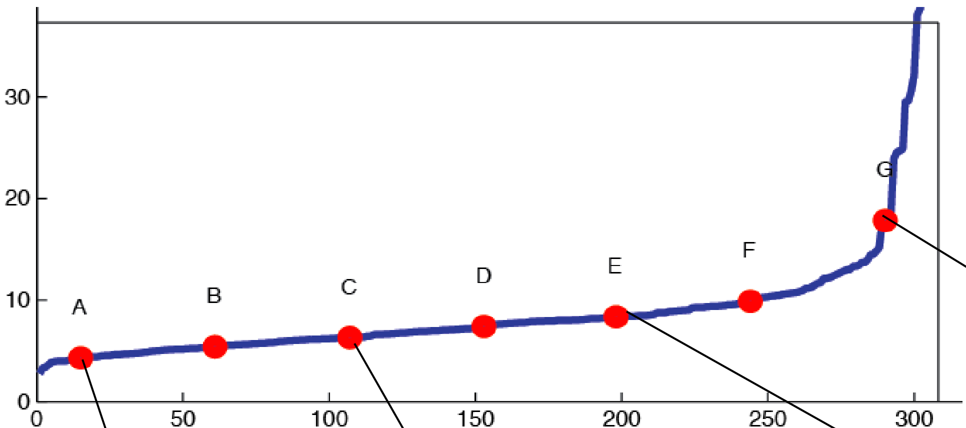
Large scale

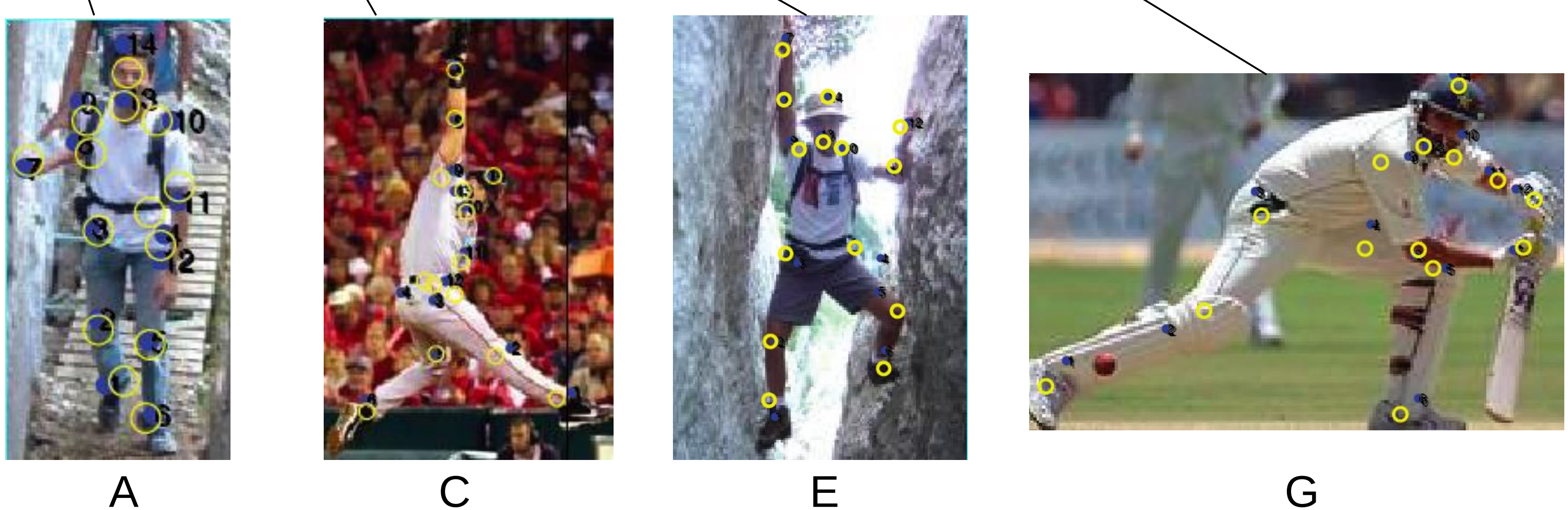
Low price

Issues

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

Annotation quality





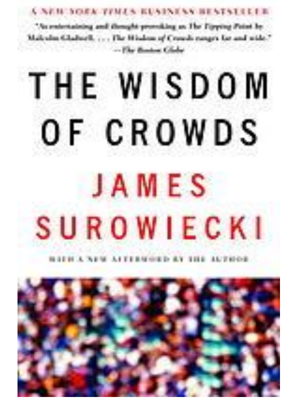
Agree within 5-10 pixels
on 500x500 screen

There are bad ones.

How do we get quality annotations?

Ensuring Annotation Quality

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



- Gold Standard / Sentinel
  - Special case: qualification exam

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

Pricing

- Trade off between throughput and cost
- Higher pay can actually attract scammers

Outline

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

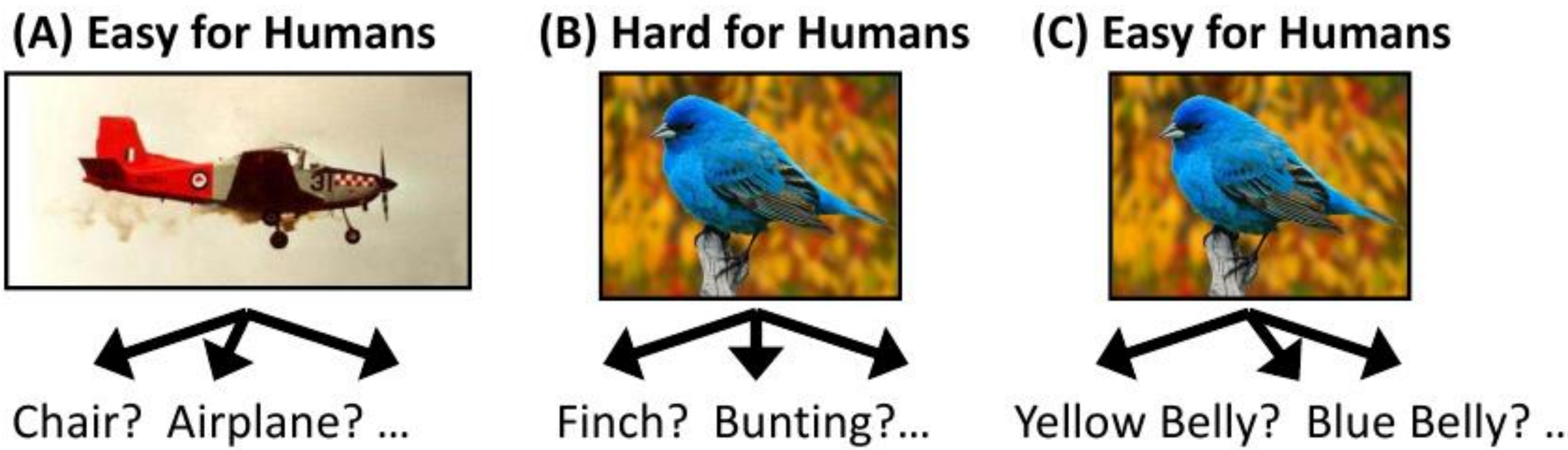
Visual Recognition with Humans in the Loop

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.

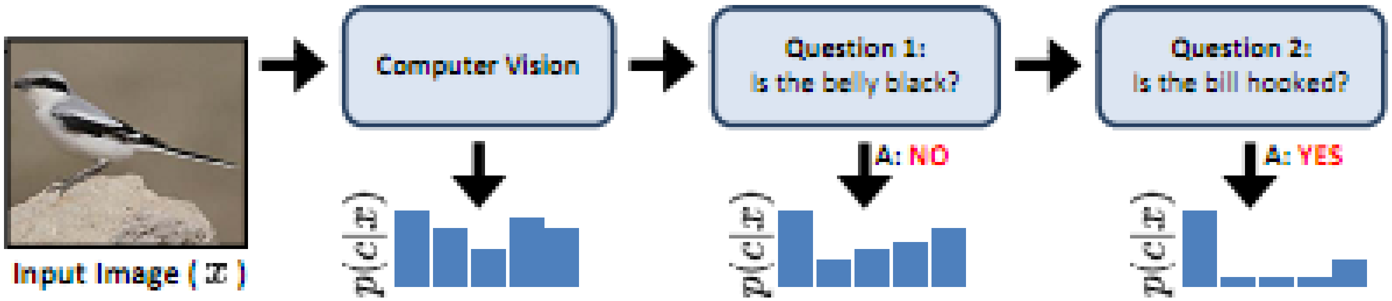
Combine strengths to solve this problem.

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.

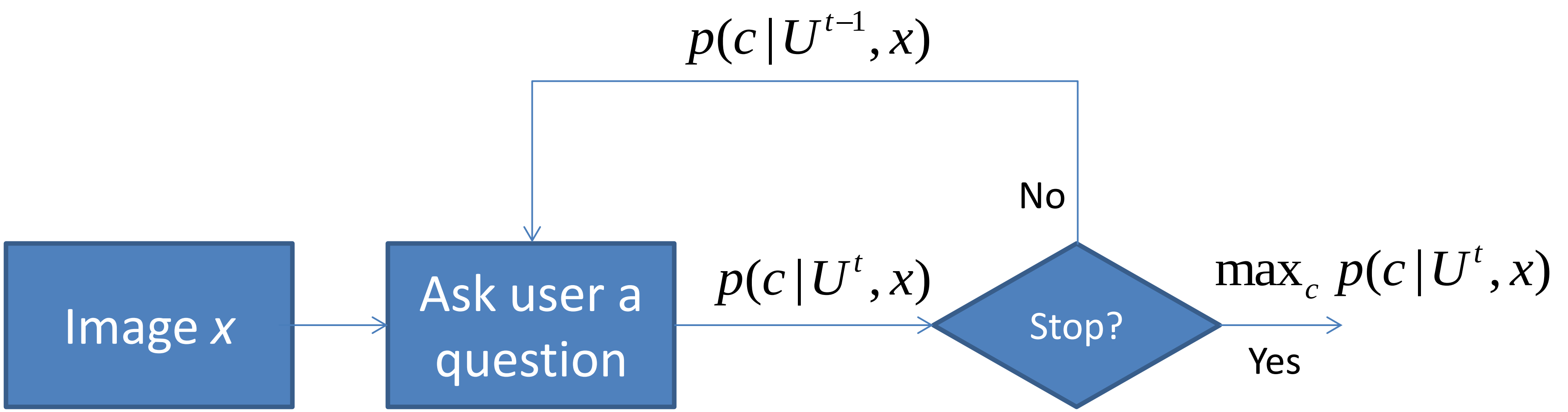
The Approach: 20 Questions

- 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.

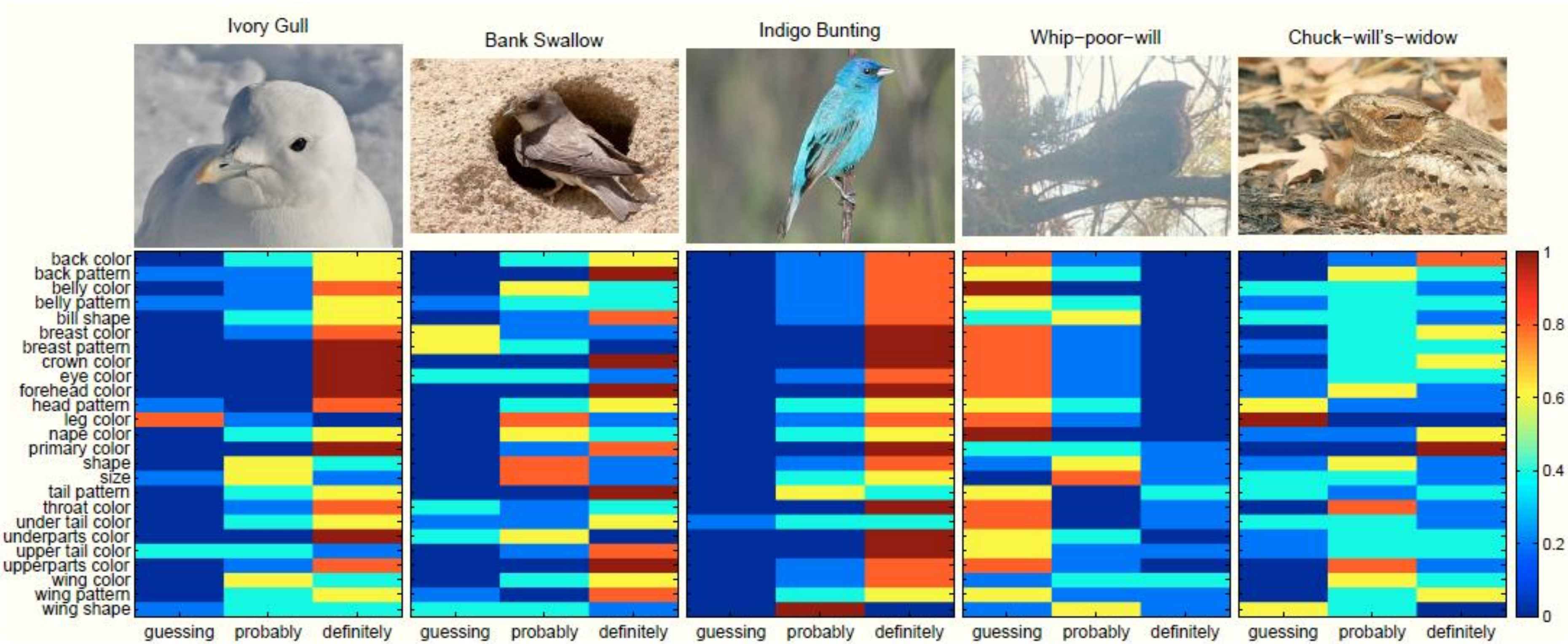
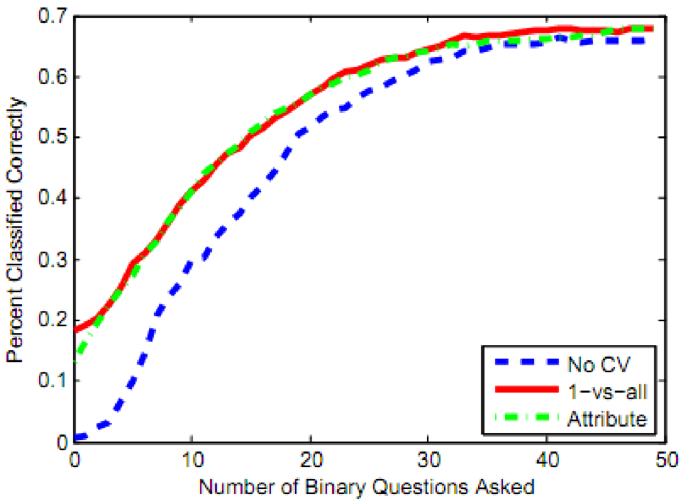
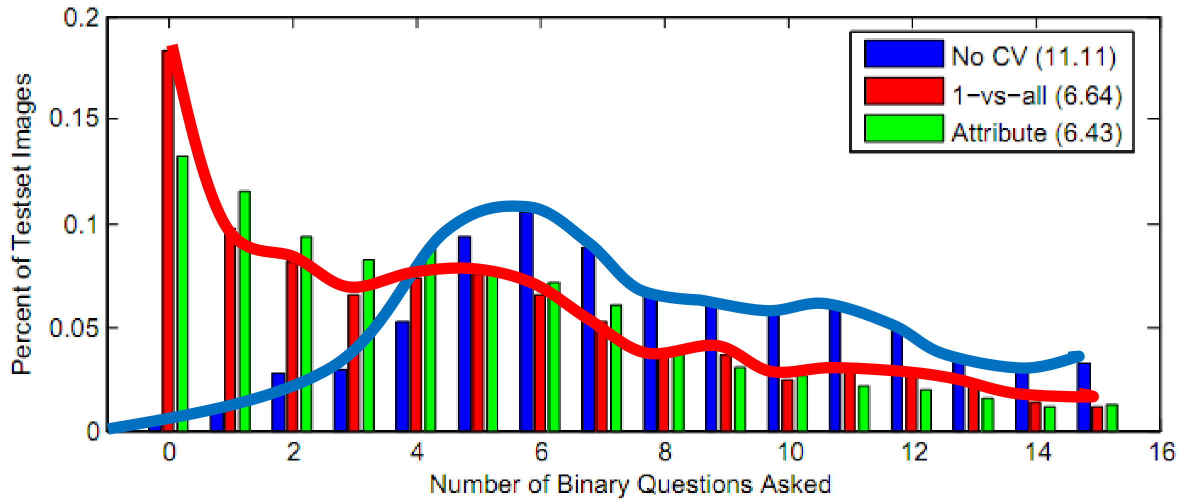


Fig. 4. Examples of user responses for each of the 25 attributes. The distribution over $\{Guessing, Probably, Definitely\}$ is color coded with blue denoting 0% and red denoting 100% of the five answers per image attribute pair.

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?