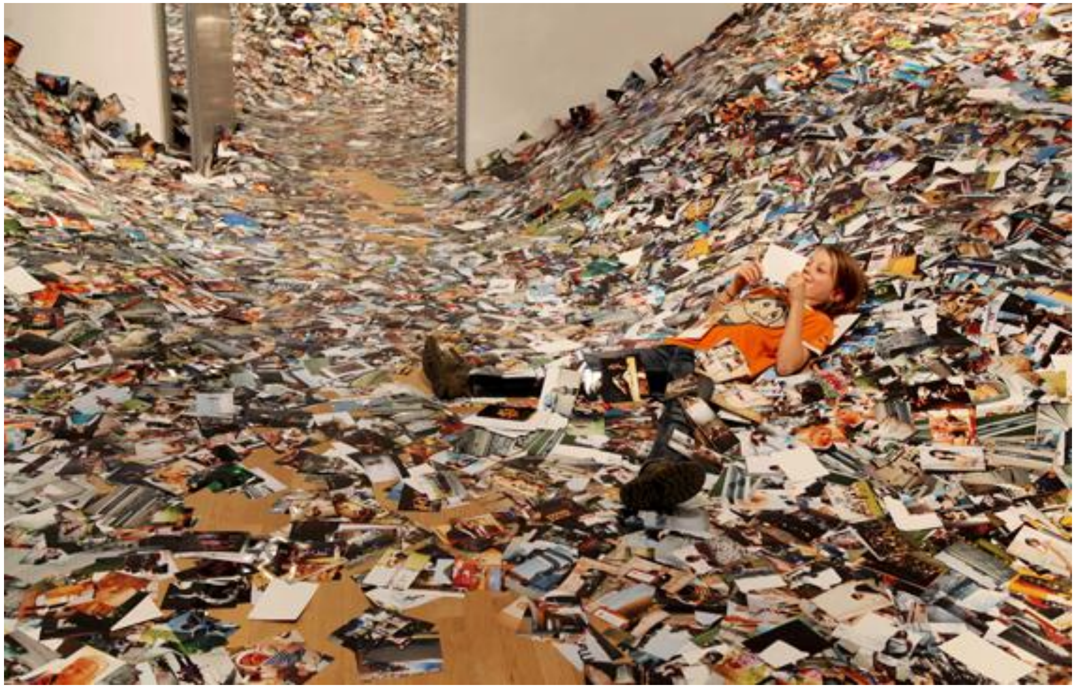
Human Computation and Computer Vision

CS143 Computer Vision

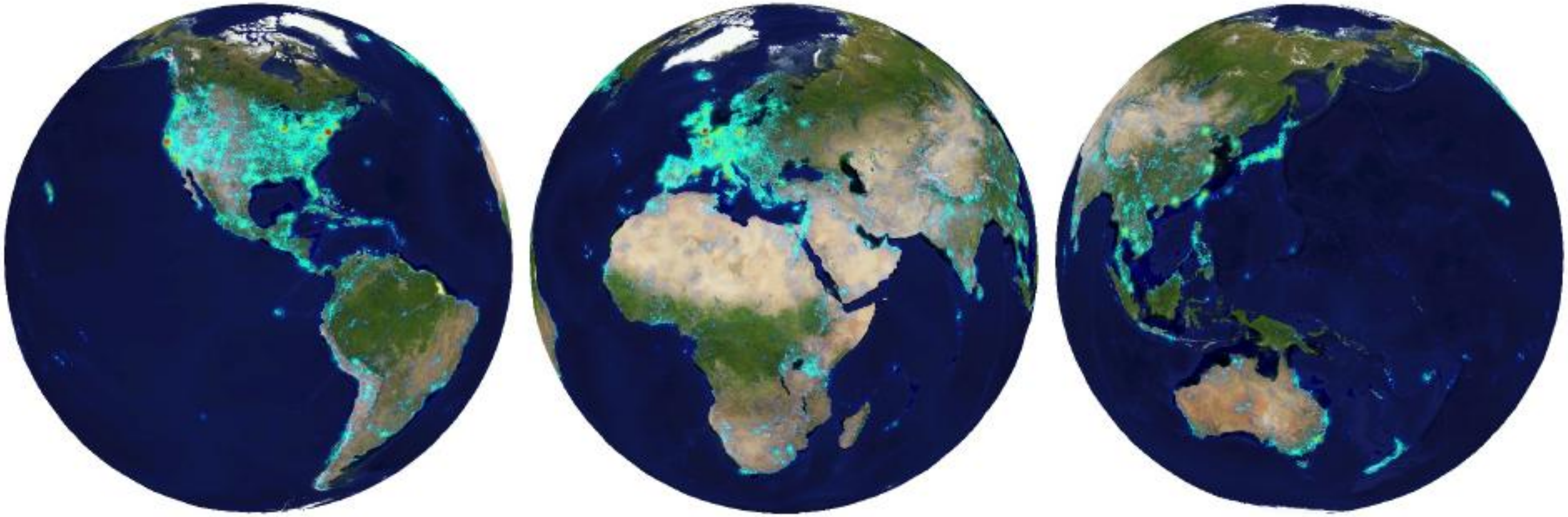
James Hays, Brown University

24 hours of Photo Sharing



installation by Erik Kessels

And sometimes Internet photos have useful labels

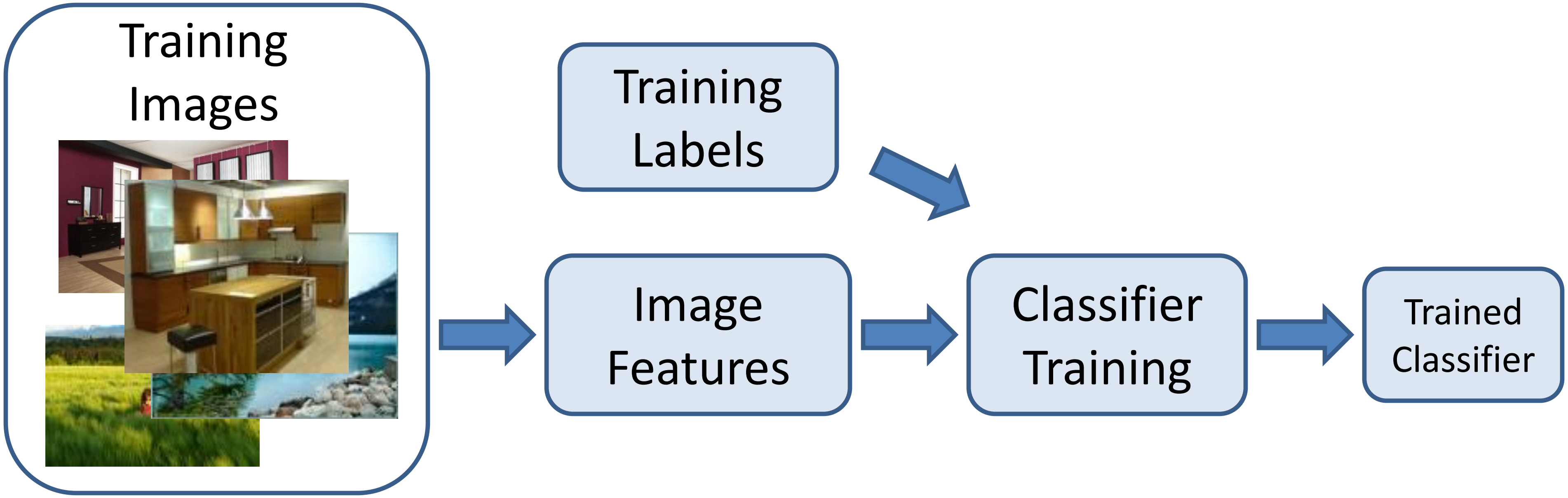


Im2gps. Hays and Efros. CVPR 2008

But what if we want more?

Image Categorization

Training



Training Images

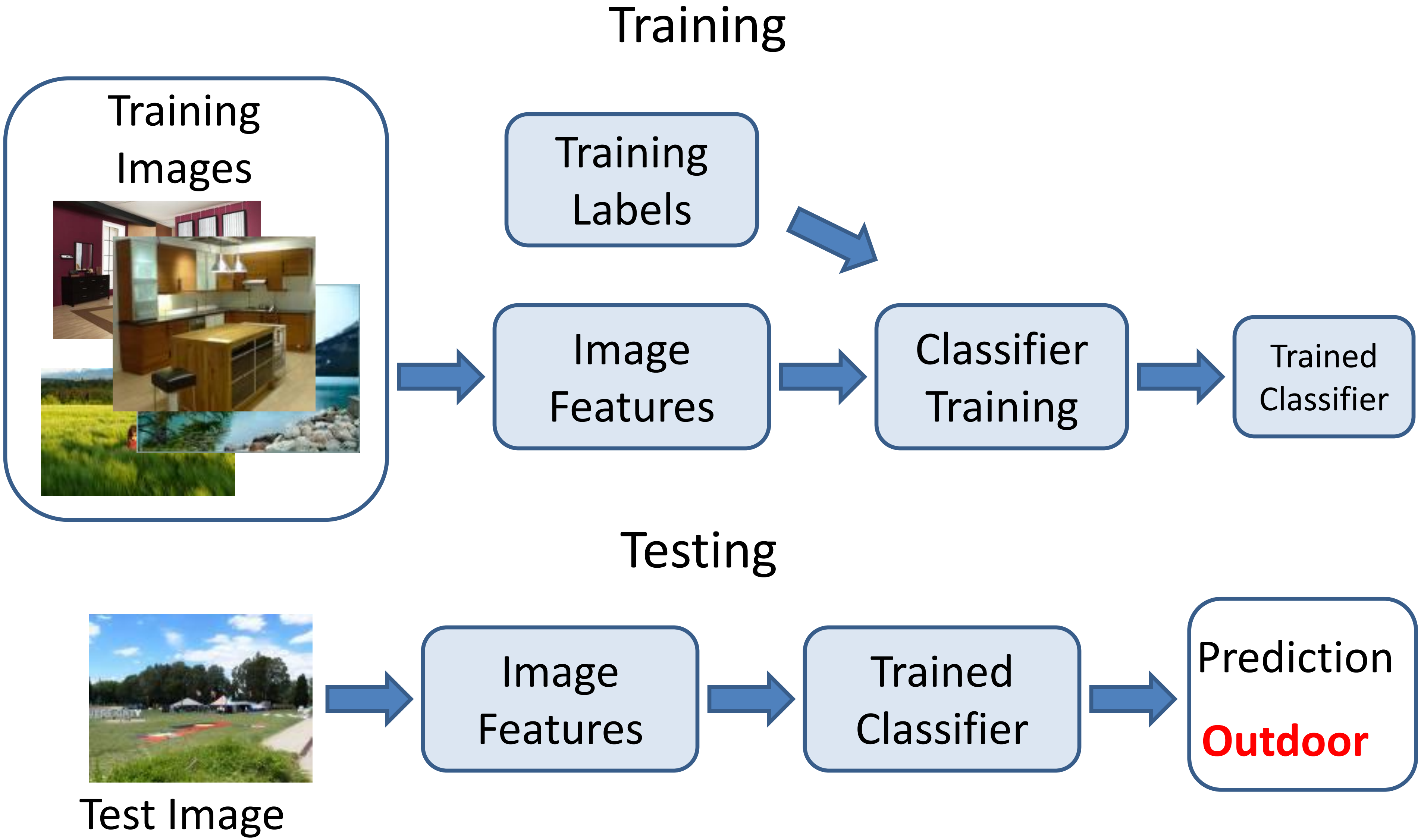
Training Labels

Image Features

Classifier Training

Trained Classifier

Image Categorization



Training

Training Images

Training Labels

Image Features

Classifier Training

Trained Classifier

Testing

Test Image

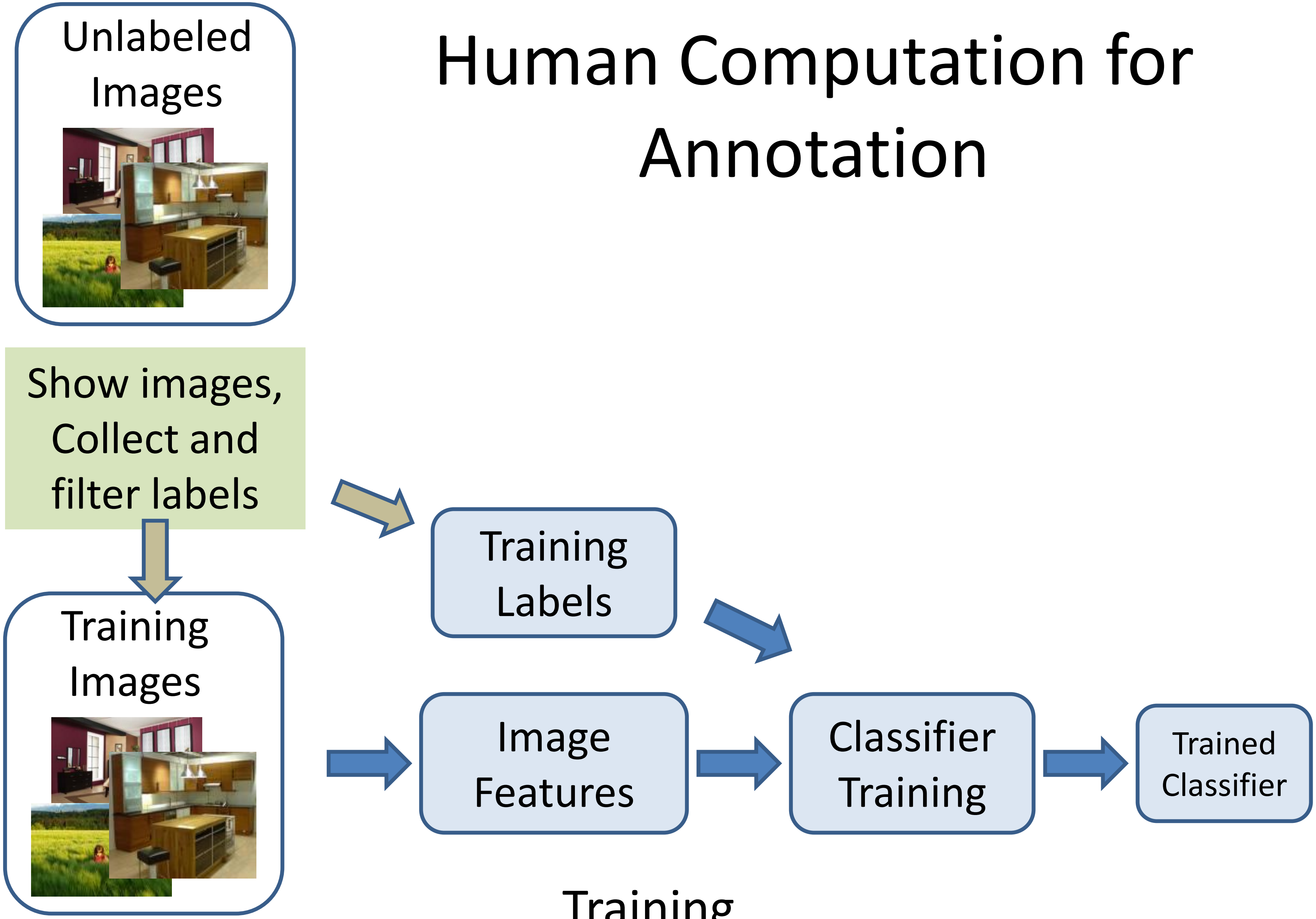
Image Features

Trained Classifier

Prediction

Outdoor

Human Computation for Annotation



Unlabeled Images

Show images, Collect and filter labels

Training Images

Training Labels

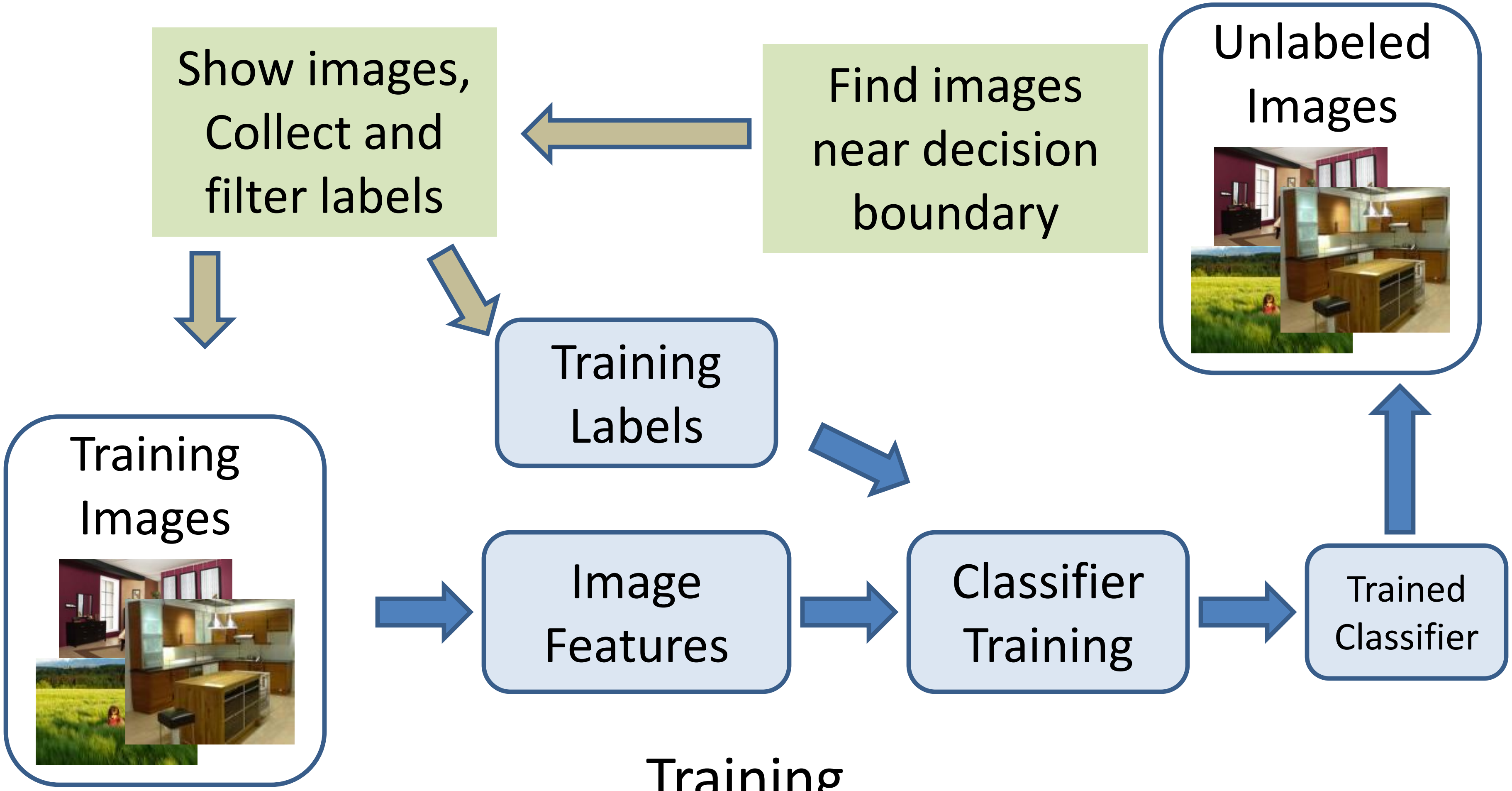
Image Features

Classifier Training

Trained Classifier

Training

Active Learning



Show images, Collect and filter labels

Find images near decision boundary

Unlabeled Images

Training Labels

Training Images

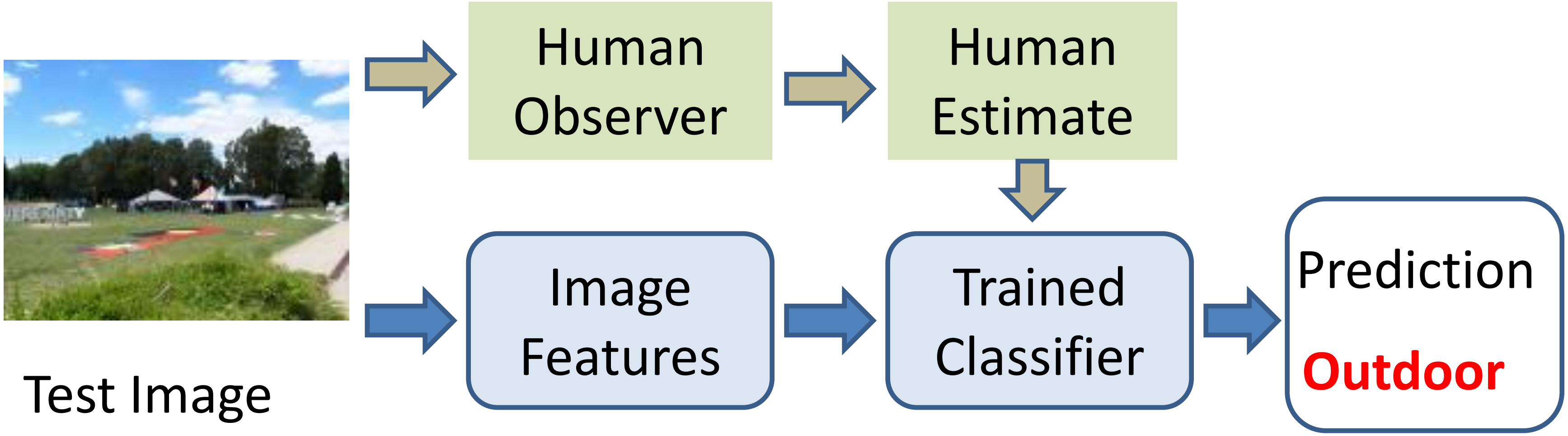
Image Features

Classifier Training

Trained Classifier

Training

Human-in-the-loop Recognition



Test Image

Human Observer

Human Estimate

Image Features

Trained Classifier

Prediction

Outdoor

Testing

Outline

- Human Computation for Annotation
  - ESP Game
  - Mechanical Turk
- Human-in-the-loop Recognition
  - Visipedia

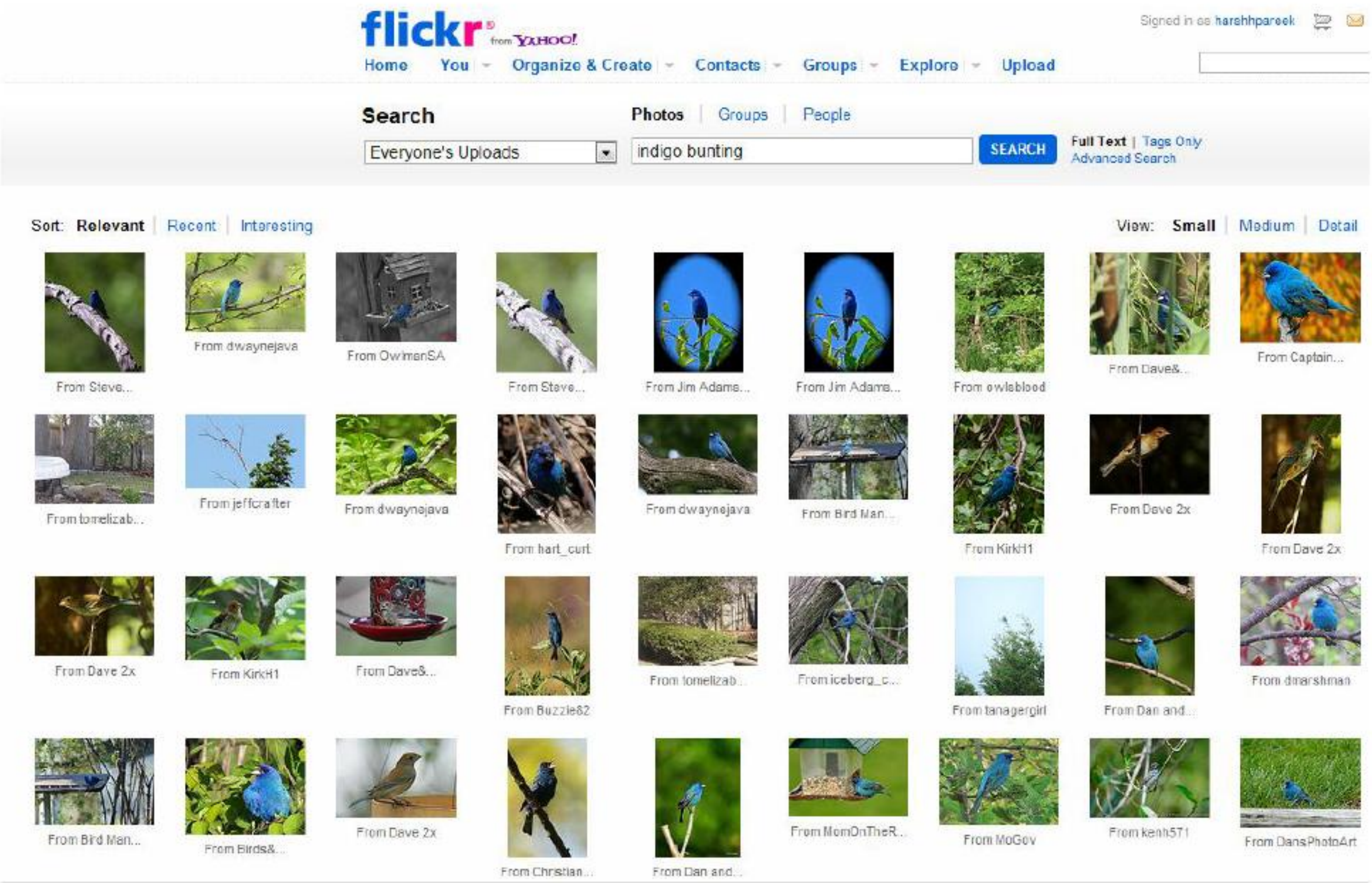
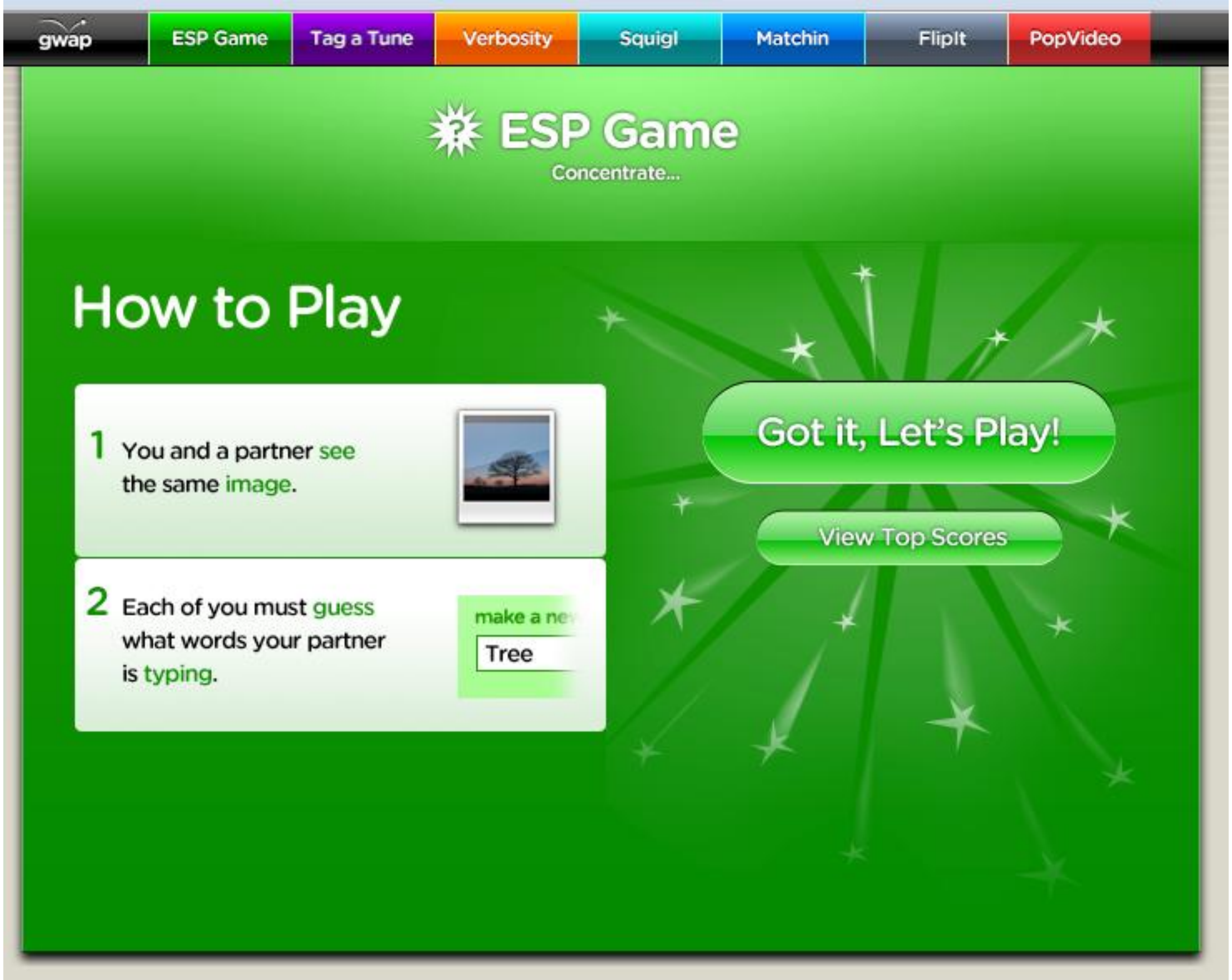


Image credit: Flickr.com



ESP Game

Concentrate...

How to Play

1 You and a partner see the same image.

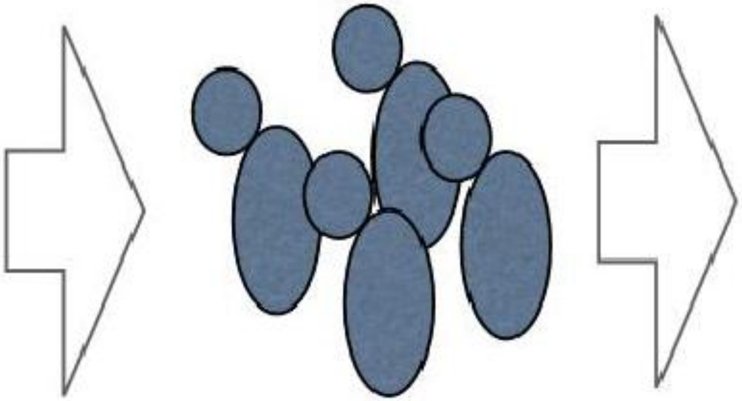
2 Each of you must guess what words your partner is typing.

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

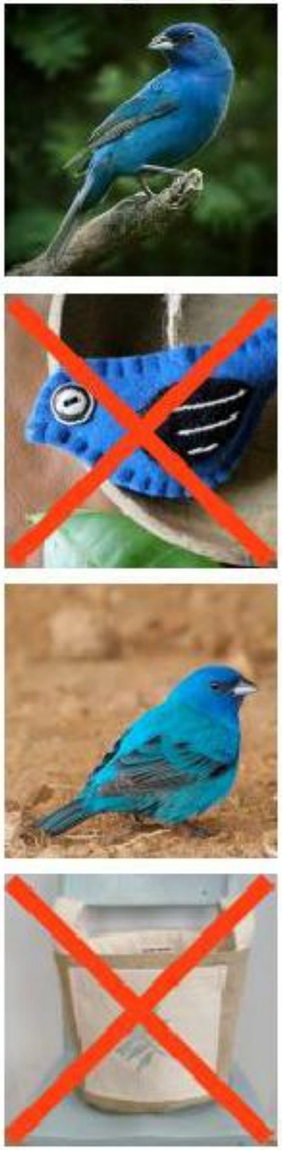
Annotators



amazonmechanical turk
Artificial Artificial Intelligence

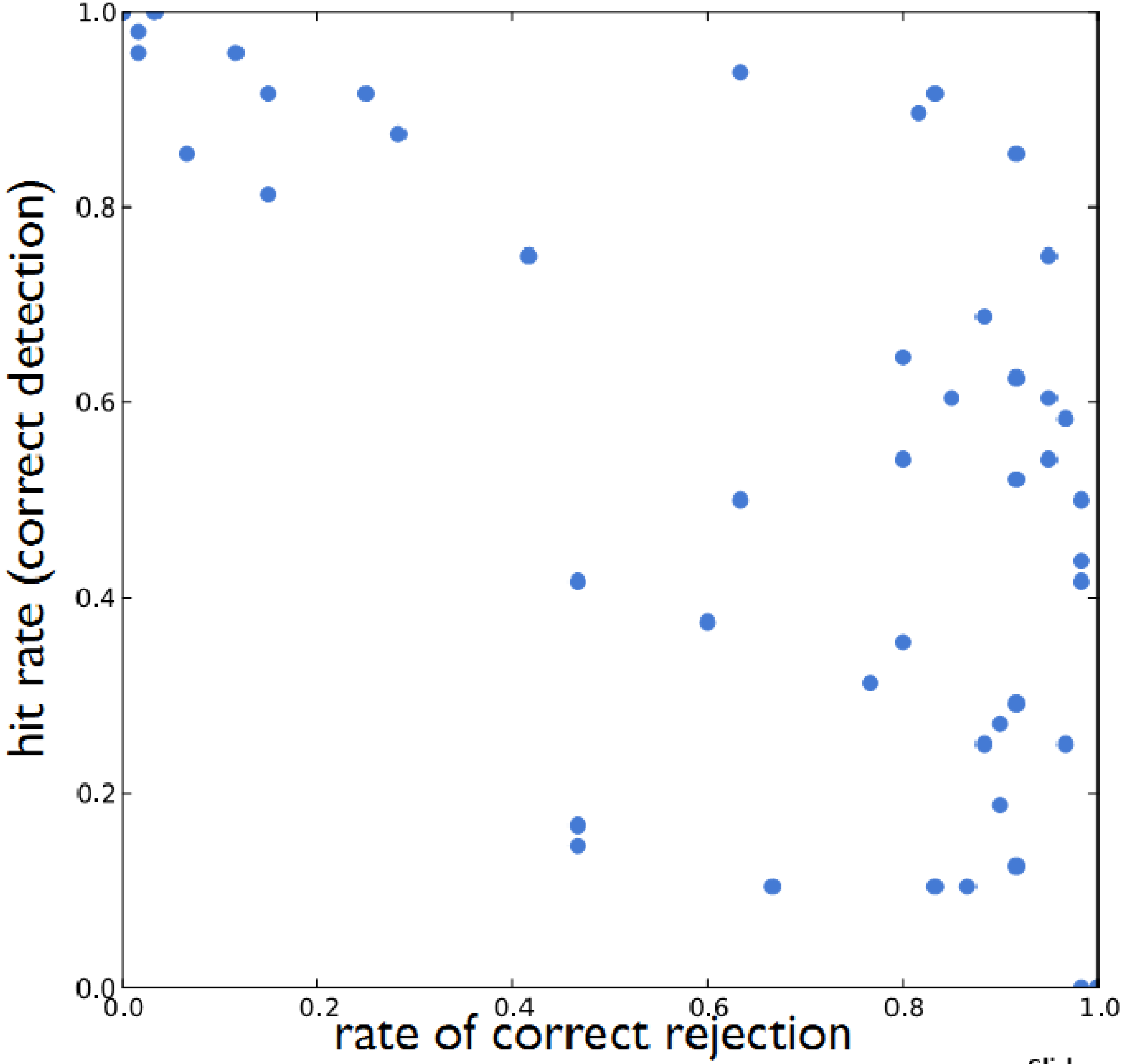
Is there an Indigo bunting in the image?

100s of
training images



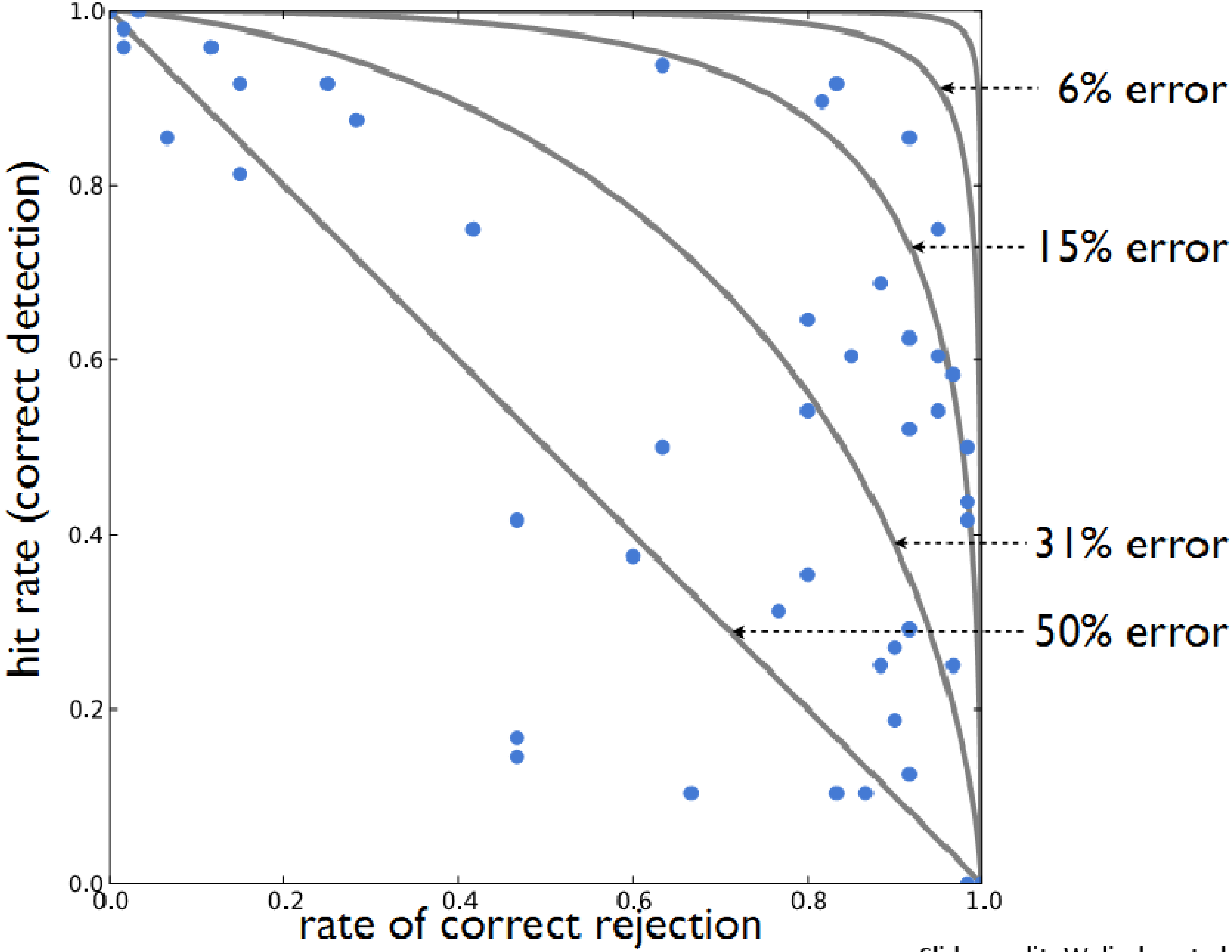
Slide credit: Welinder et al

Task: Find the Indigo Bunting



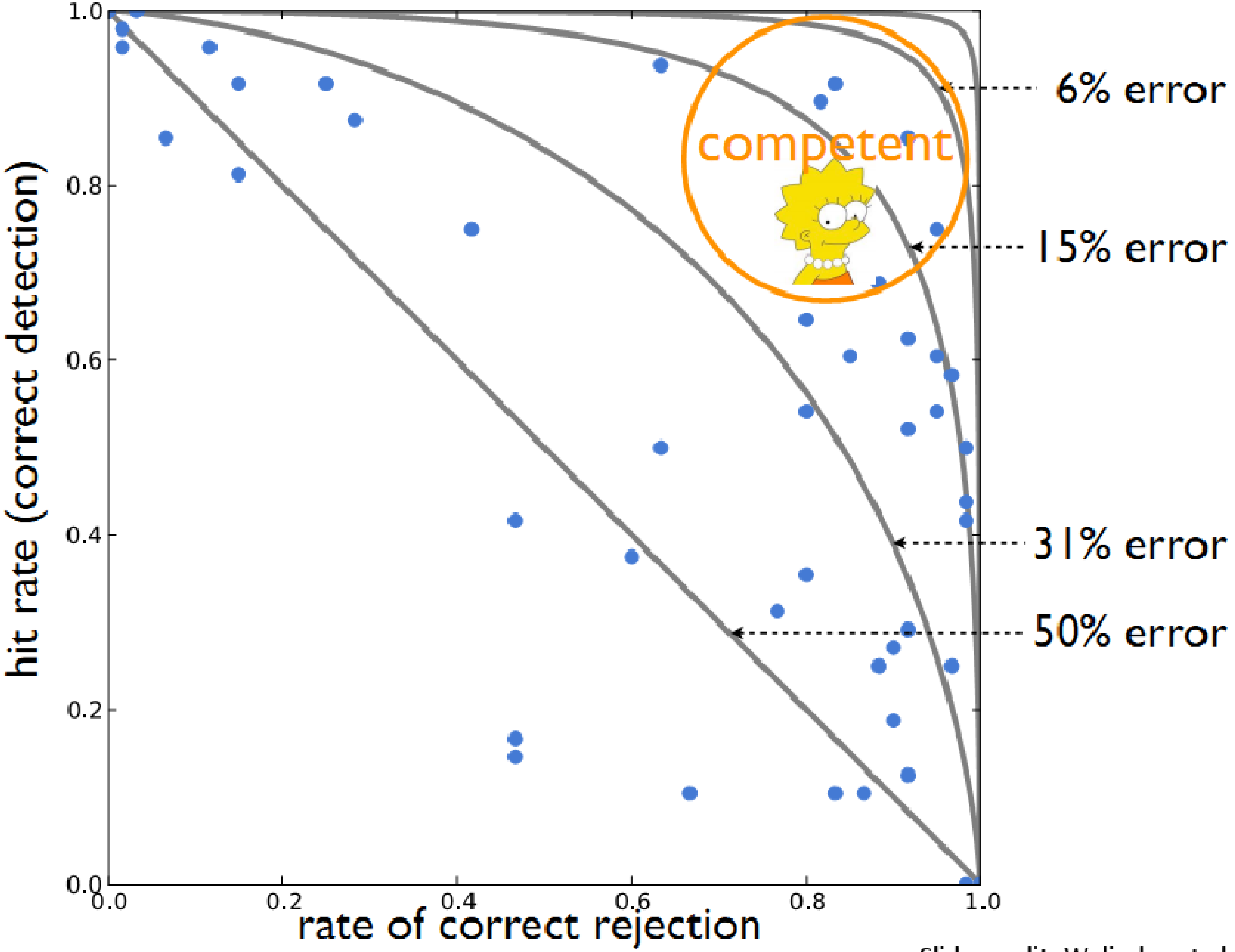
Slide credit: Welinder et al

Task: Find the Indigo Bunting



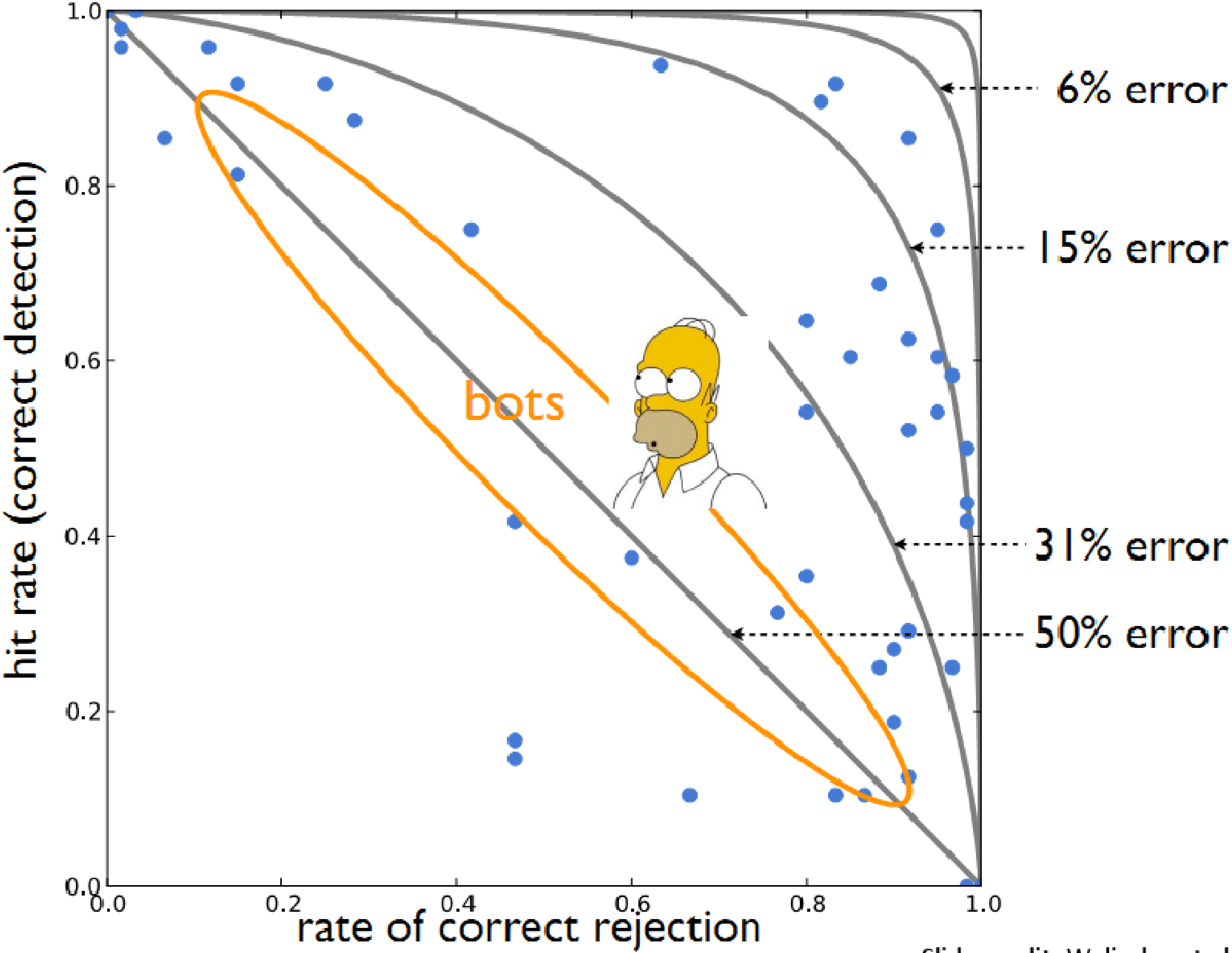
Slide credit: Welinder et al

Task: Find the Indigo Bunting



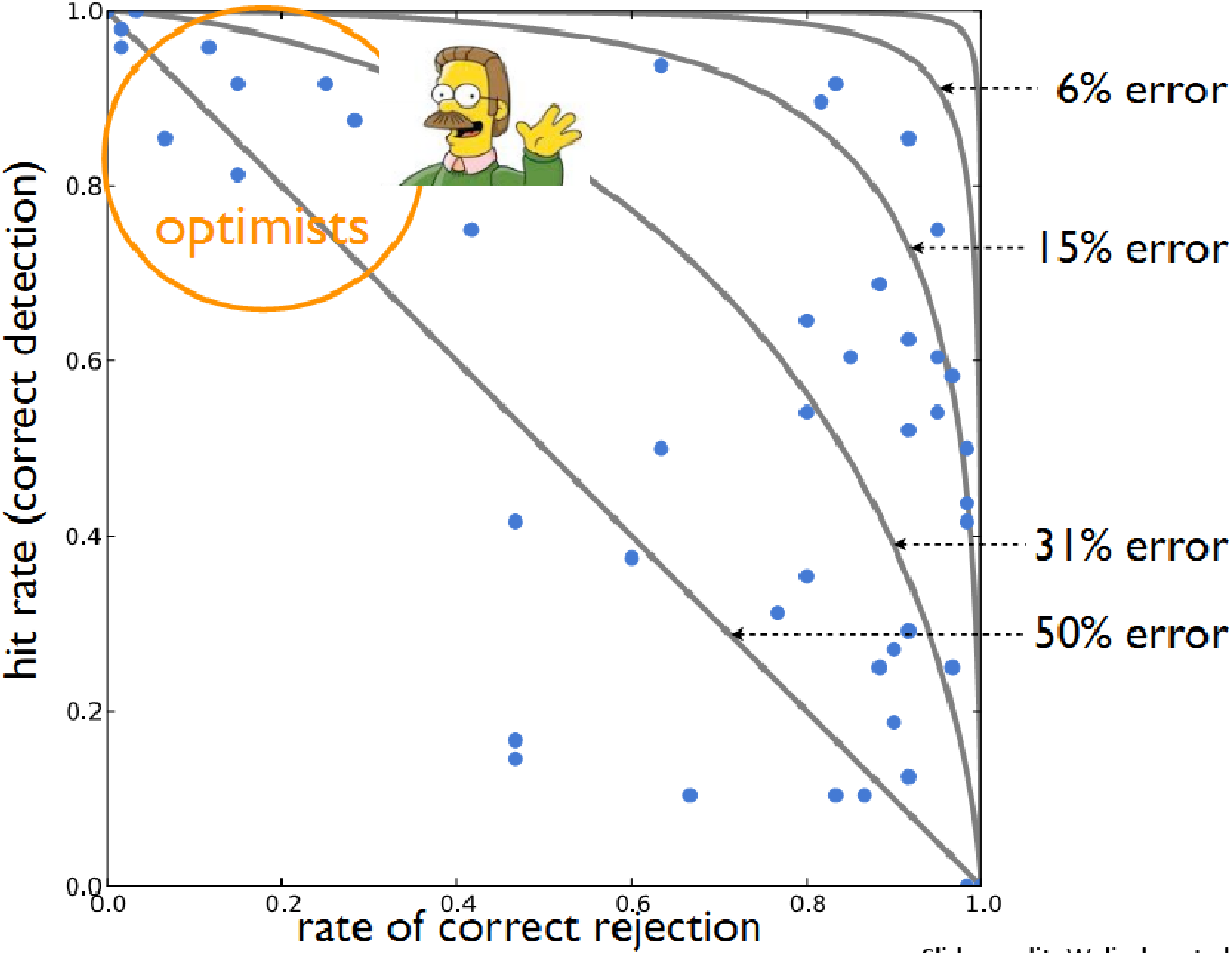
Slide credit: Welinder et al

Task: Find the Indigo Bunting



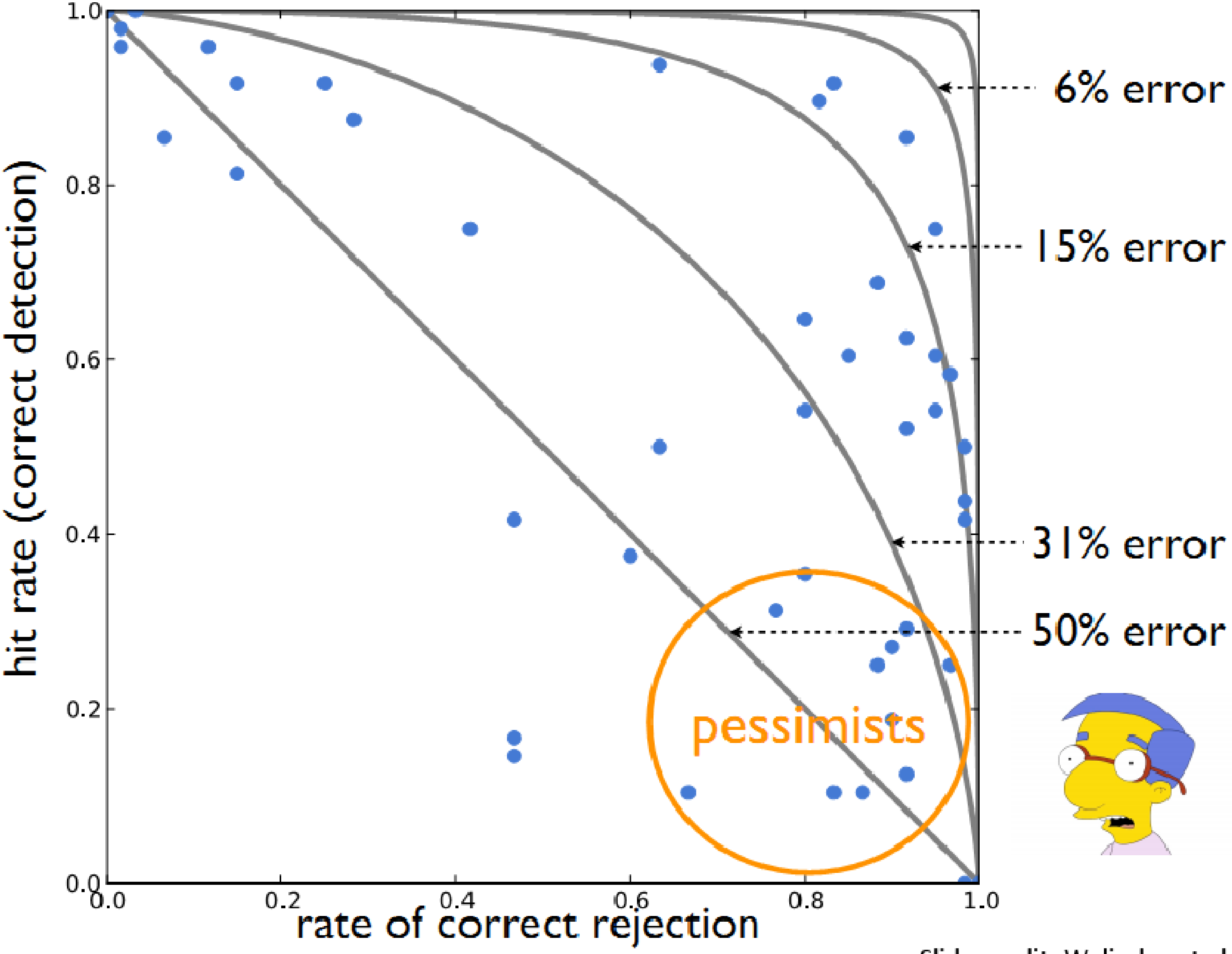
Slide credit: Welinder et al

Task: Find the Indigo Bunting



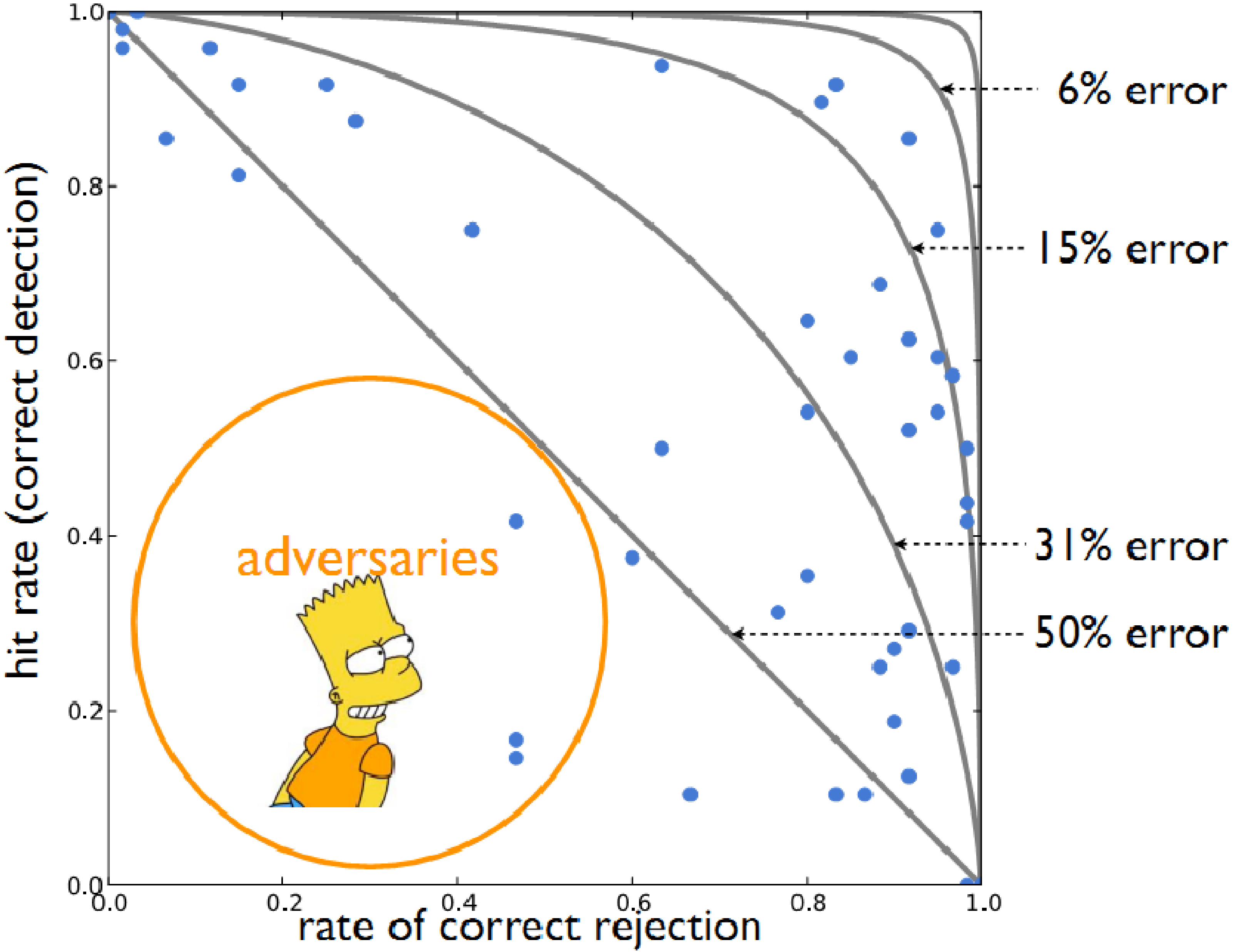
Slide credit: Welinder et al

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

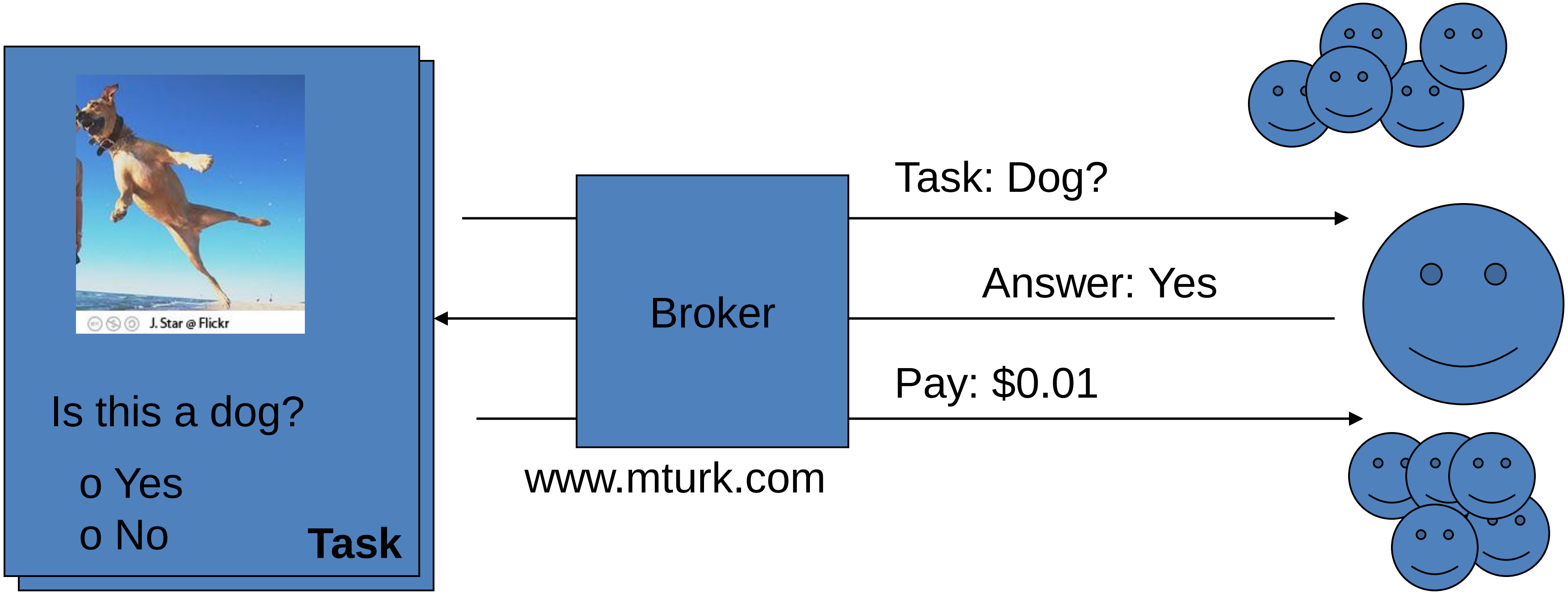
David Forsyth

University of Illinois at Urbana-Champaign

Slides by Alexander Sorokin

Amazon Mechanical Turk

Workers



J. Star @ Flickr

Is this a dog?

o Yes
o No

Task

Broker

www.mturk.com

Task: Dog?

Answer: Yes

Pay: $0.01

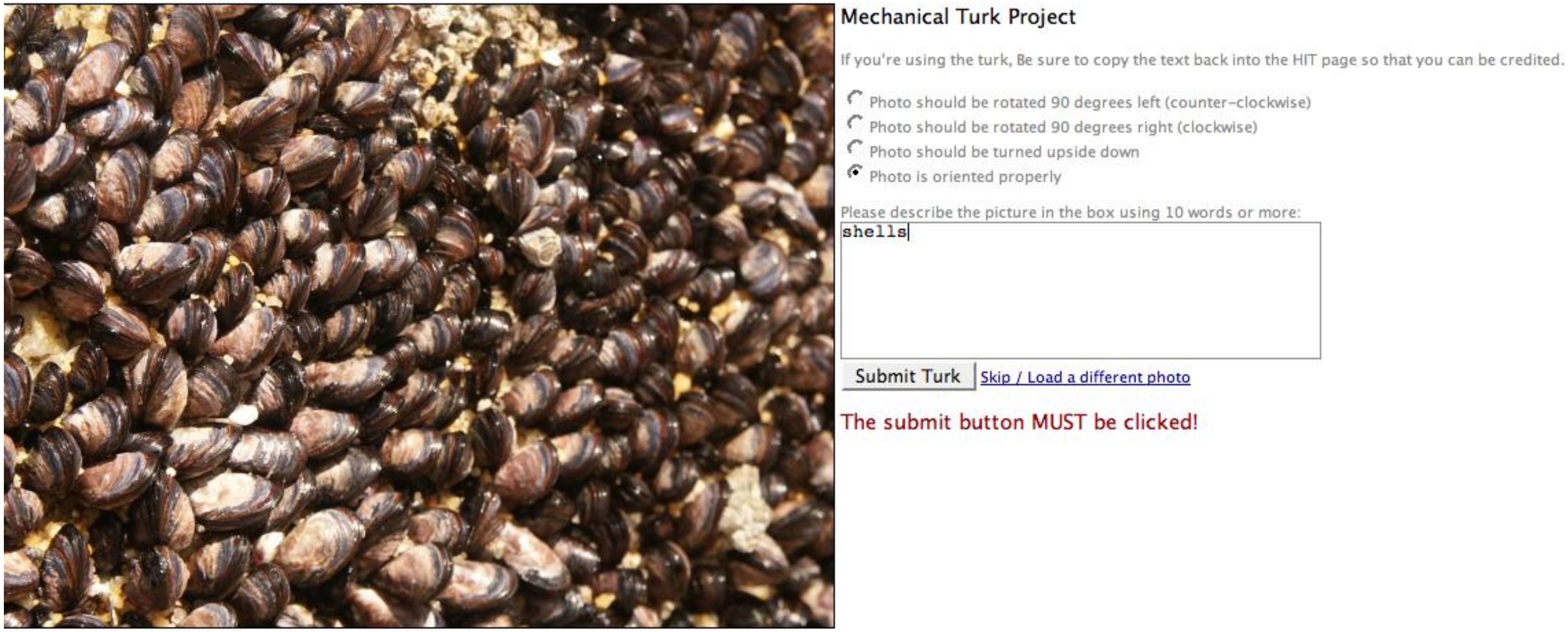
$0.01

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:

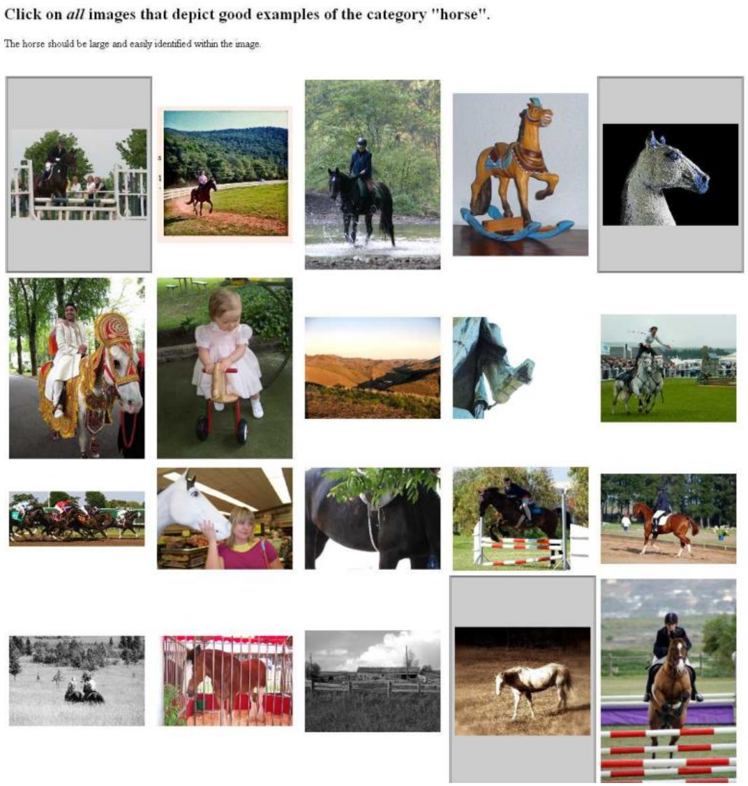
shells

The submit button MUST be clicked!

$0.01

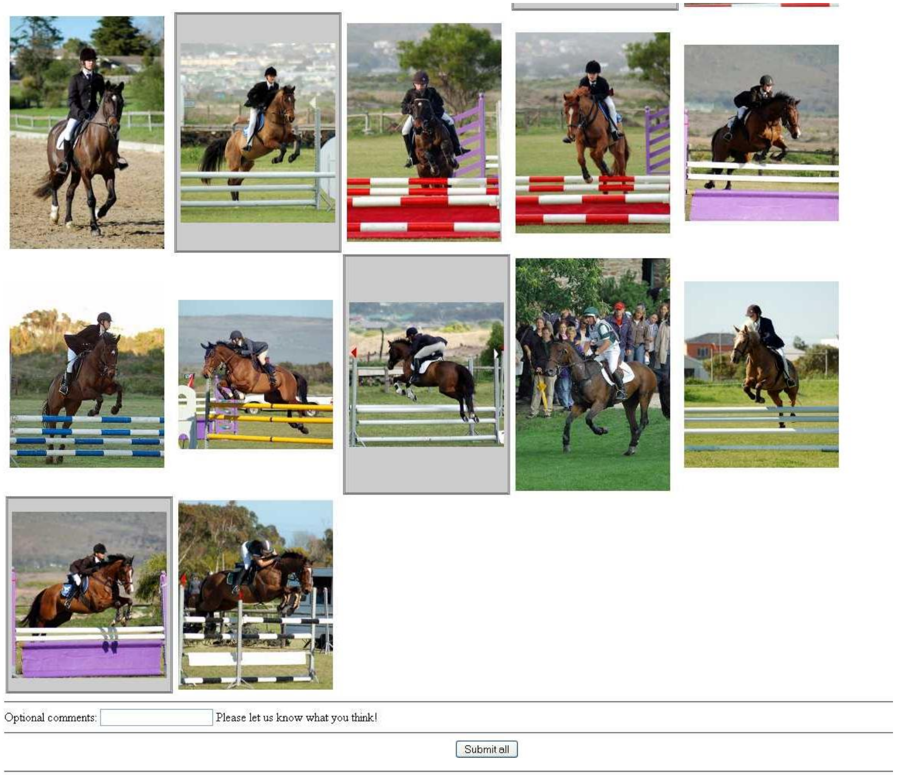
http://austinsmoke.com/turk/.

Select examples



Click on *all* images that depict good examples of the category "horse".

The horse should be large and easily identified within the image.

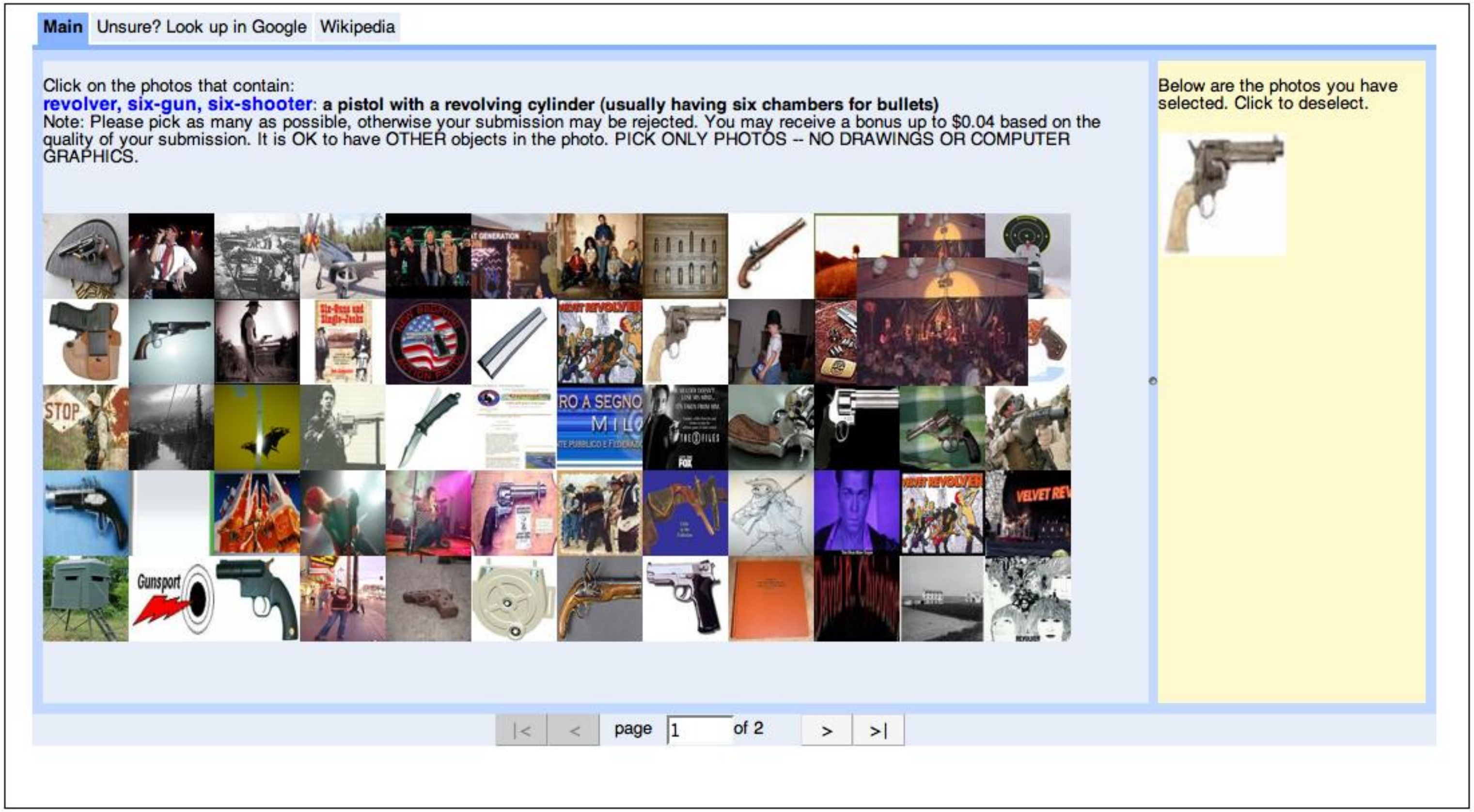


Optional comments: Please let us know what you think!

Joint work with Tamara and Alex Berg

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

Select examples



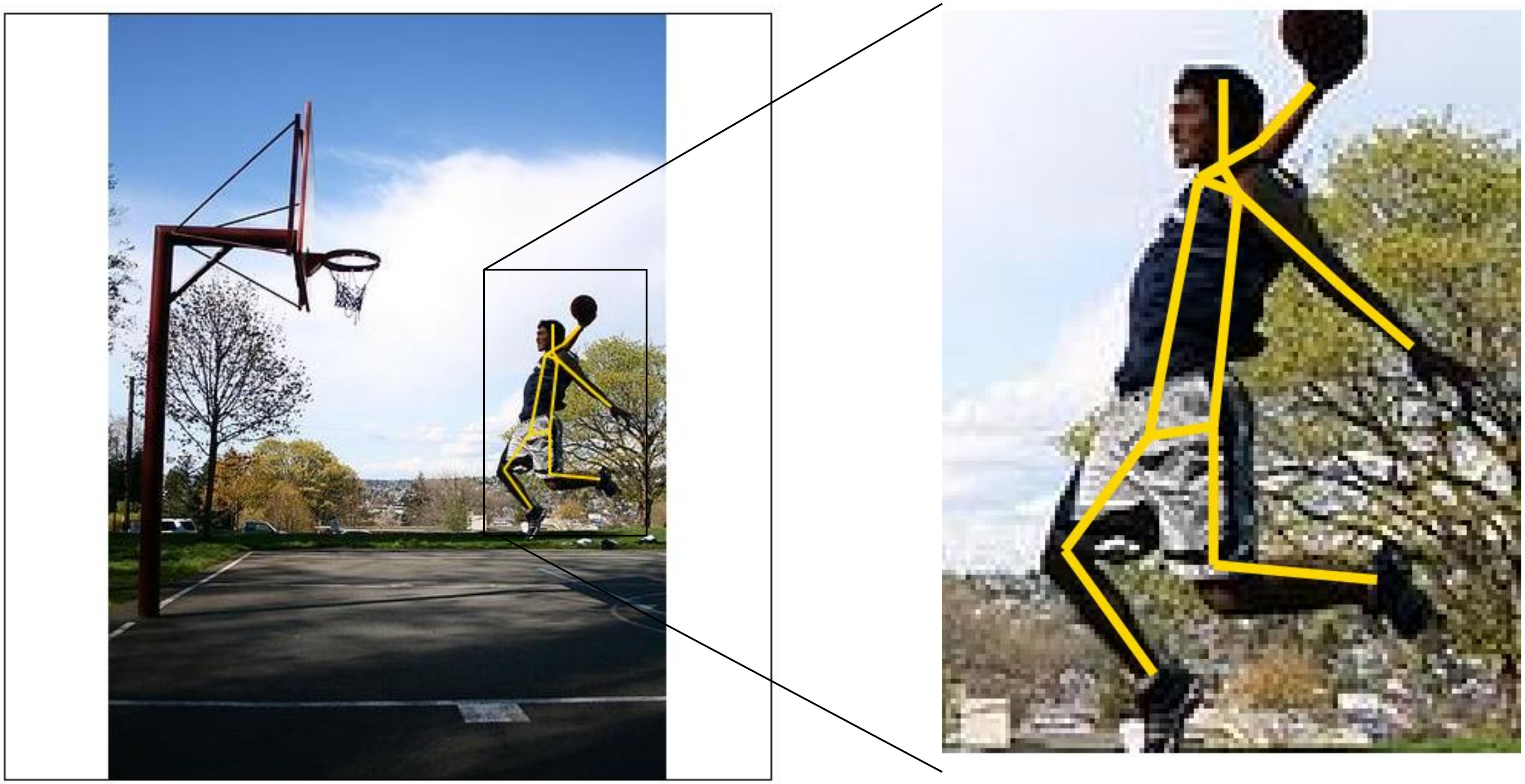
Click on the photos that contain:
revolver, six-gun, six-shooter: a pistol with a revolving cylinder (usually having six chambers for bullets)
Note: Please pick as many as possible, otherwise your submission may be rejected. You may receive a bonus up to $0.04 based on the quality of your submission. It is OK to have OTHER objects in the photo. PICK ONLY PHOTOS -- NO DRAWINGS OR COMPUTER GRAPHICS.

Below are the photos you have selected. Click to deselect.

$0.02

requester mtlabel

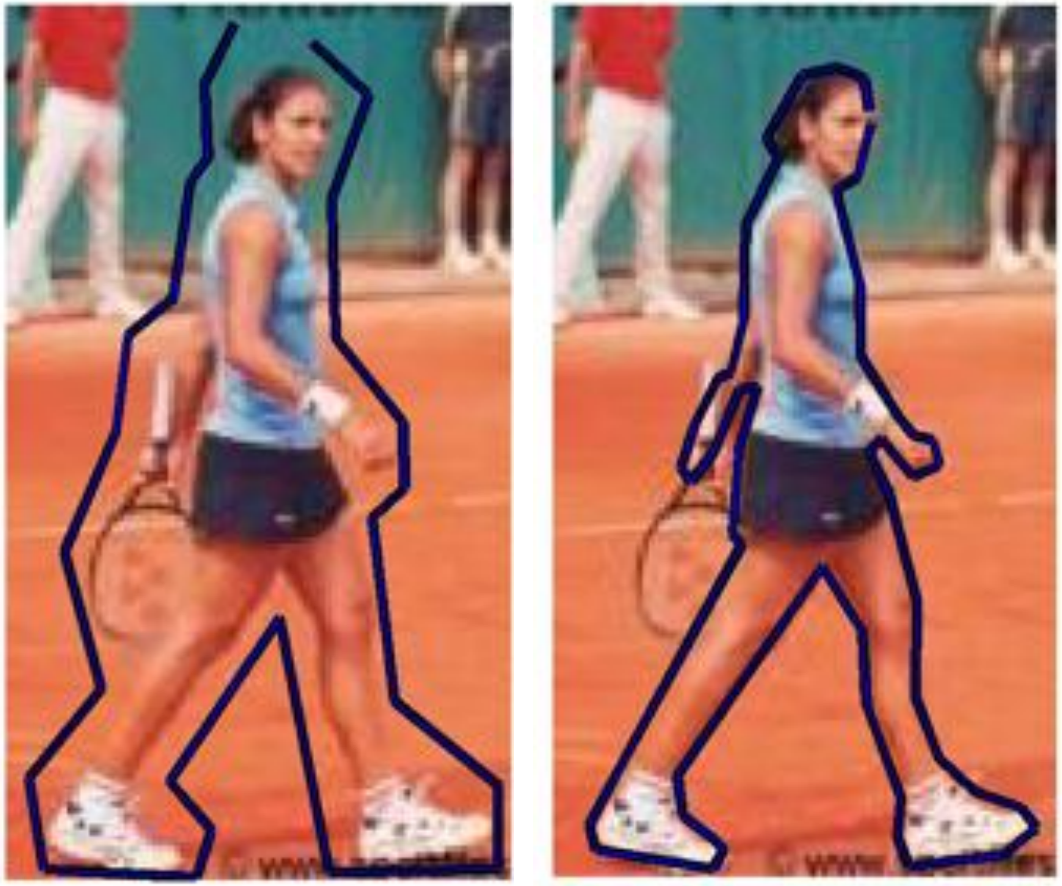
Click on landmarks



$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

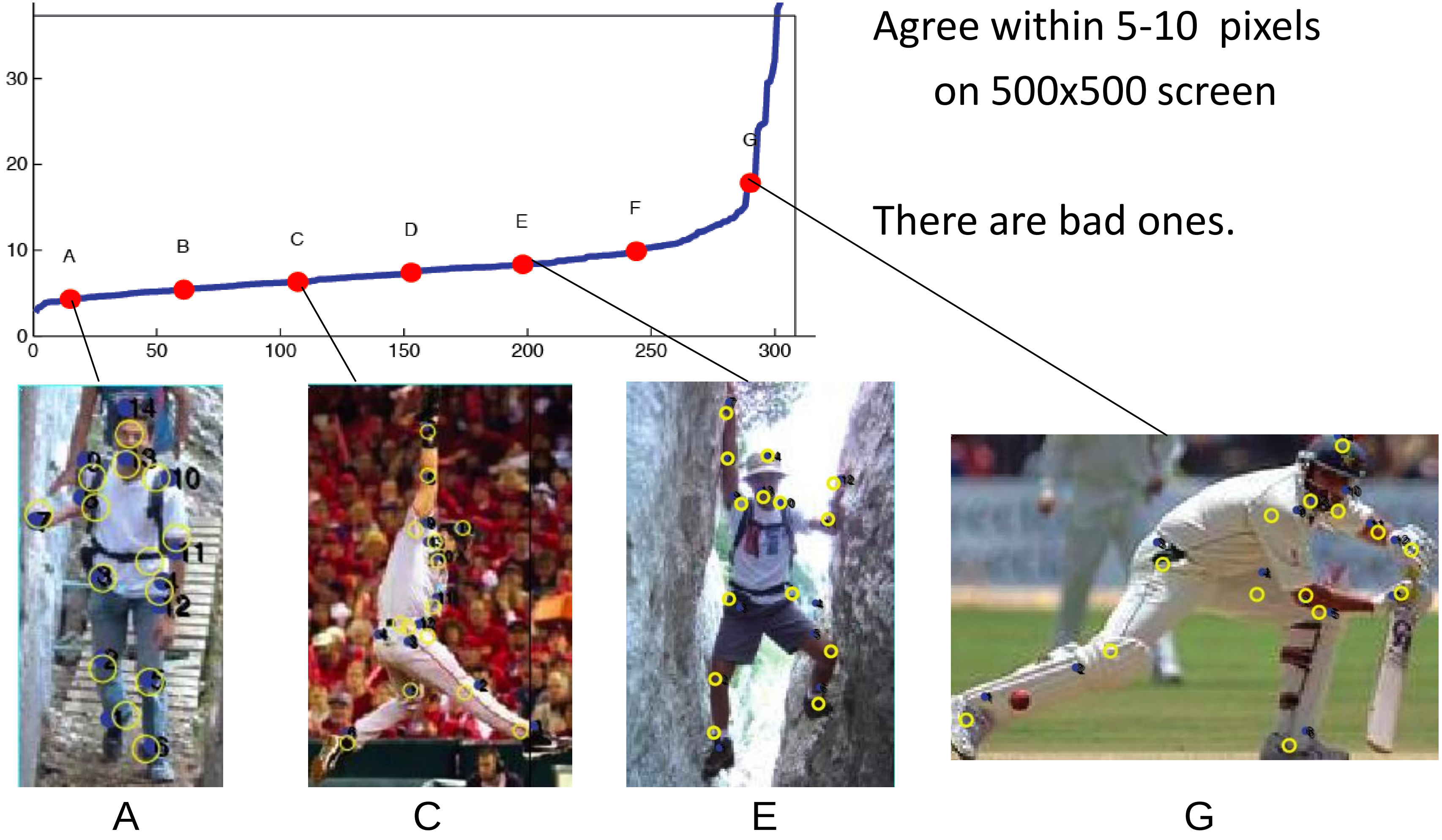
Large scale

Low price

Issues

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

Annotation quality



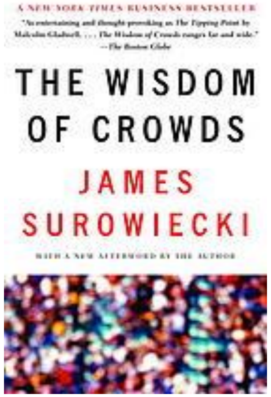
Agree within 5-10 pixels
on 500x500 screen

There are bad ones.

How do we get quality annotations?

Ensuring Annotation Quality

- Consensus / Multiple Annotation / “Wisdom of the Crowds”


A NEW YORK TIMES BUSINESS BESTSELLER
THE WISDOM OF CROWDS
JAMES SUROWIECKI
WITH A NEW AFTERWORD BY THE AUTHOR

- 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

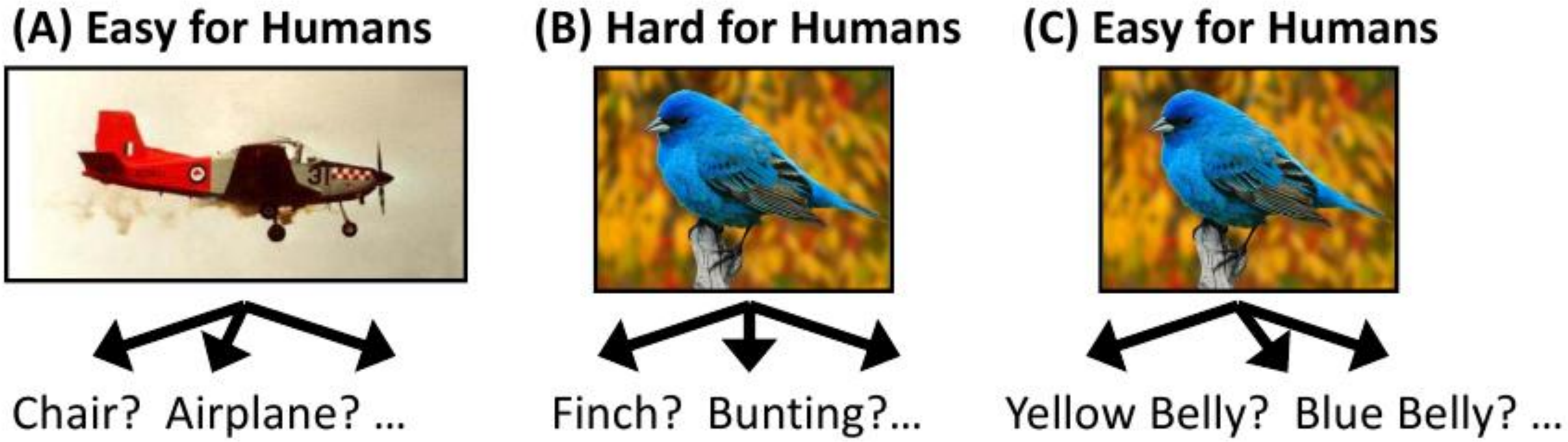
Visual Recognition with Humans in the Loop

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

Part of the Visipedia project

Slides from Brian O'Neil

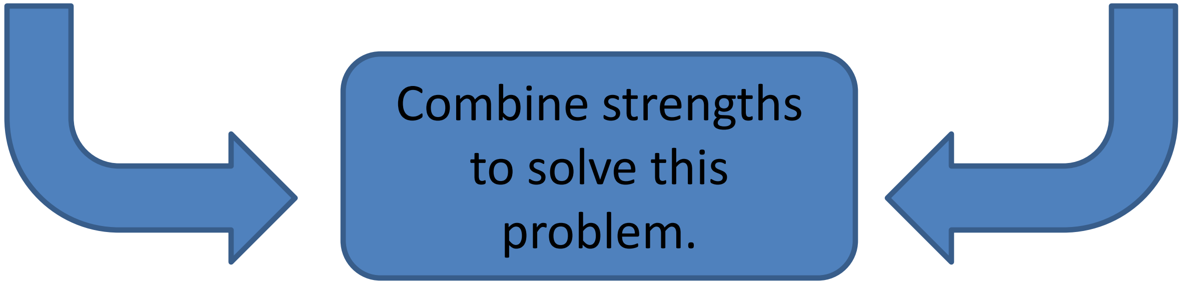
Introduction:



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.



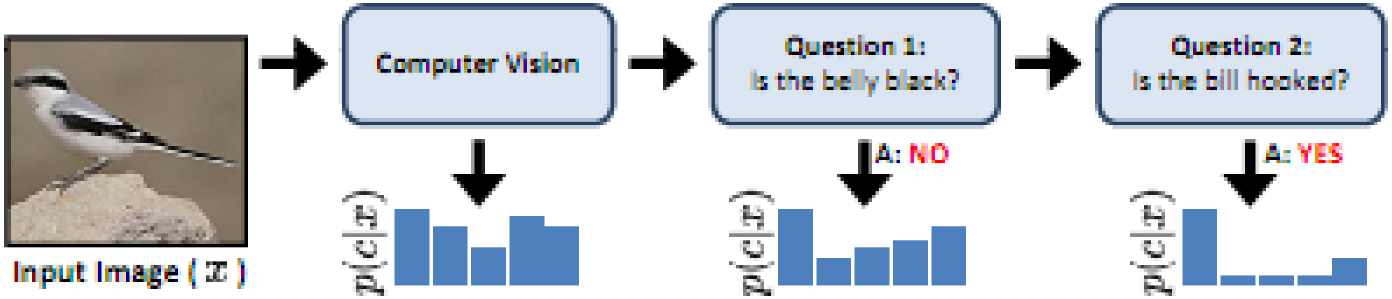
Combine strengths to solve this problem.

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.

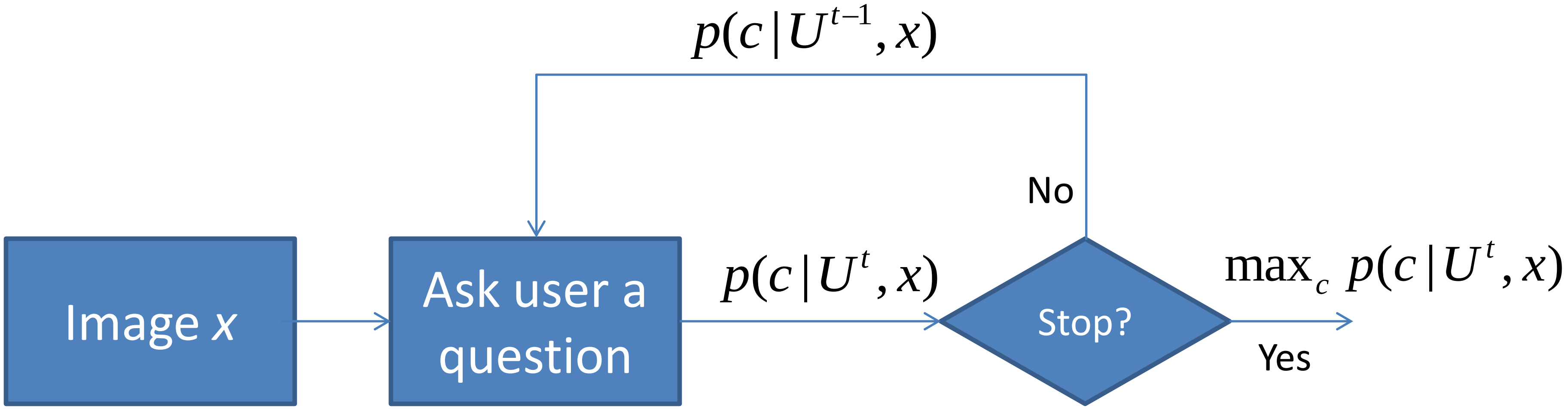
The Approach: 20 Questions

- 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.



Some definitions:

$Q = \{q_1 \ldots q_n\}$	• Set of possible questions
$a_i \in A_i$	• Possible answers to question *i*
$r_i \in V$	• Possible confidence in answer *i* (Guessing, Probably, Definitely)
$u_i = (a_i, r_i)$	• User response
U^t	• History of user responses at time *t*

Question selection

- Seek the question that gives the maximum information gain (entropy reduction) given the image and the set of previous user responses.

$$I\left(c; u_i \mid x, U^{t-1}\right) = \sum_{u_i \in A_i \times V} \boxed{p\left(u_i \mid x, U^{t-1}\right)} \; \boxed{H\left(c \mid x, u_i \cup U^{t-1}\right)} - \boxed{H\left(c \mid x, U^{t-1}\right)}$$

Probability of obtaining
Response u_i given the image
And response history

Entropy when
response is
Added to history

Entropy before response
is added.

where $$H\left(c \mid x, U^{t-1}\right) = -\sum_{c=1}^{C} p\left(c \mid x, U^{t-1}\right) \log p\left(c \mid x, U^{t-1}\right)$$

Incorporating vision

- Bayes Rule
- A visual recognition algorithm outputs a probability distribution across all classes that is used as the prior.
- A posterior probability is then computed based on the probability of obtaining a particular response history given each class.

$$p\ \ c\,|\,x,U\ \ = \eta p\ \ U\,|\,c,x\ \ p\ \ c\,|\,x\ \ = \eta p\ \ U\,|\,c\ \ p\ \ c\,|\,x$$

Modeling user responses

- Assume that the questions are answered independently.

$$p \;\; U^{t-1} \,|\, c \;\; = \prod_{i}^{t-1} p \;\; u_i \,|\, c$$

Required for posterior computation

$$p \;\; u_i \,|\, x, U^{t-1} \;\; = \sum_{c=1}^{C} p \;\; u_i \,|\, c \;\;\; p \;\; c \,|\, x, U^{t-1}$$

Required for information gain computation

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.

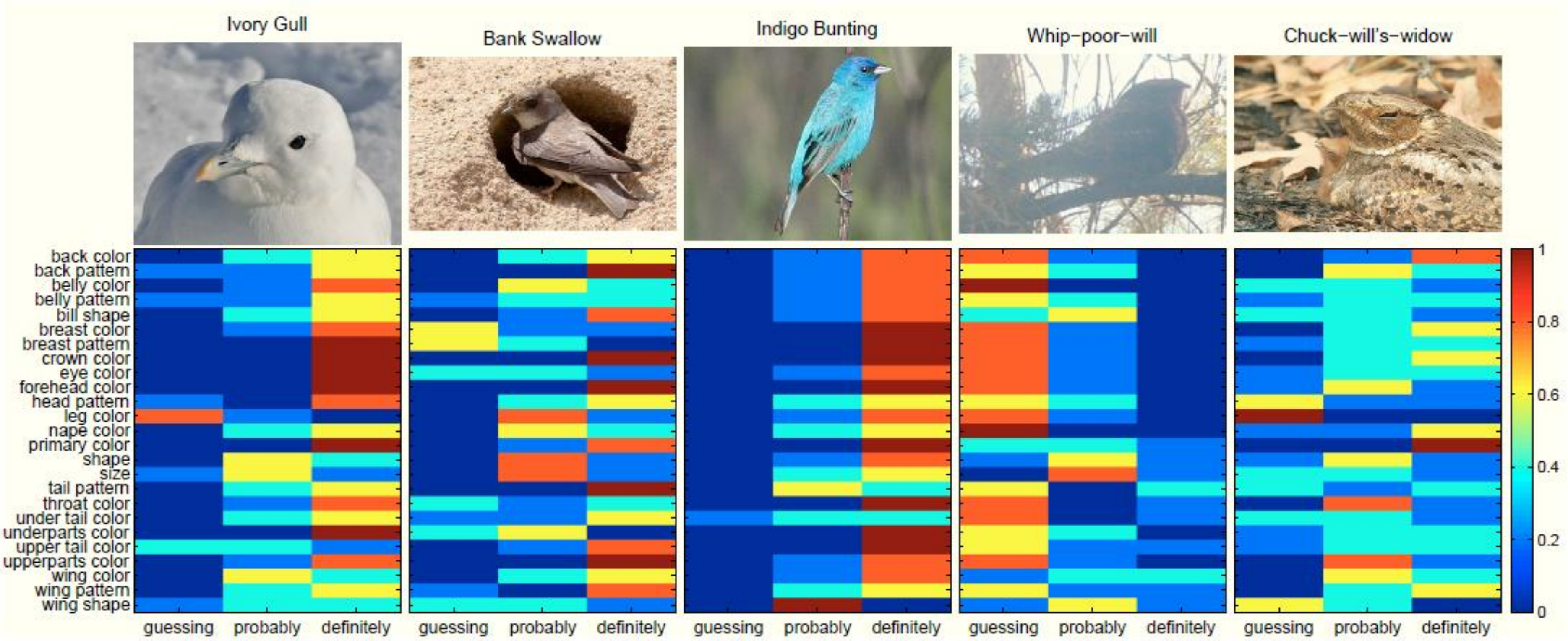
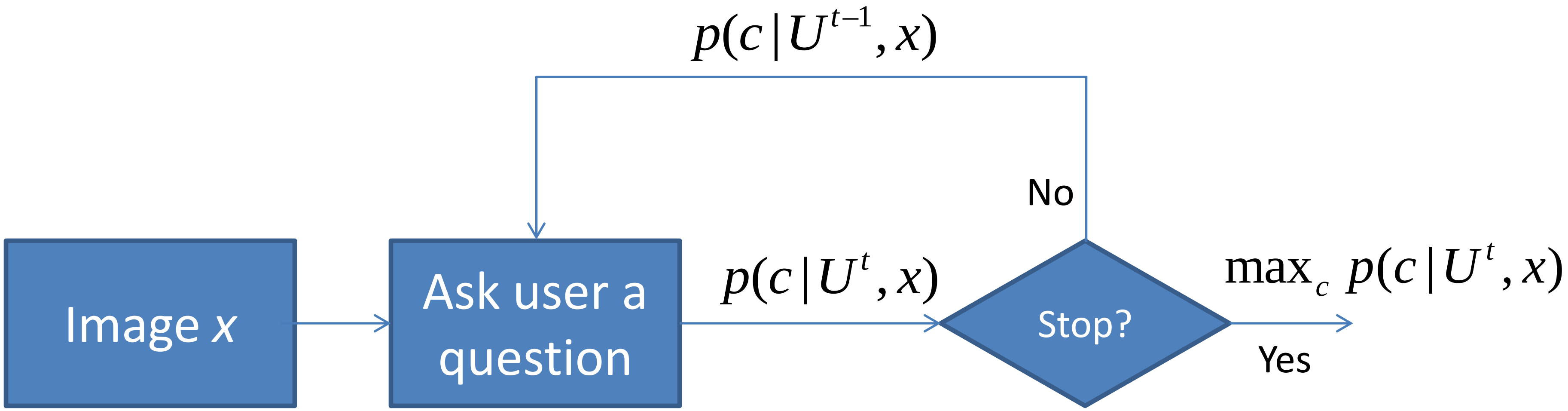


Fig. 4. Examples of user responses for each of the 25 attributes. The distribution over $\{Guessing, Probably, Definitely\}$ is color coded with blue denoting 0% and red denoting 100% of the five answers per image attribute pair.

Visual recognition

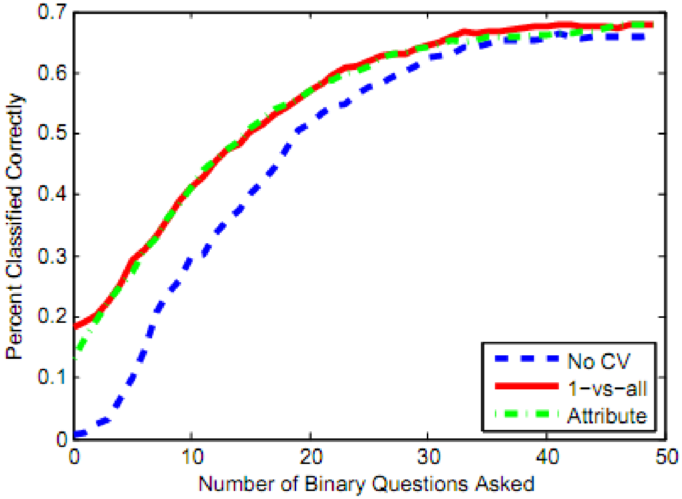
- Any vision system that can output a probability distribution across classes will work.
- Authors used Andrea Vedaldis’s code.
  - Color/gray SIFT
  - VQ geometric blur
  - 1 v All SVM
- Authors added full image color histograms and VQ color histograms

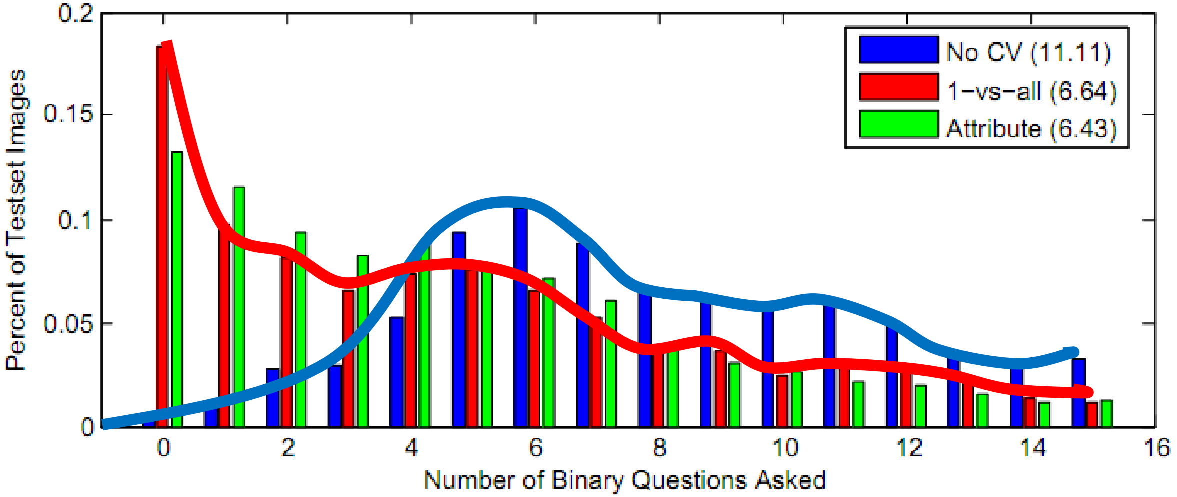
Experiments



- 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?