

Human-Machine Partnerships

Spring 2018

Review

- Last week:
 - Autoencoders
 - Fine-tuning
 - Software for deep learning
- Assignments (Canvas):
 - Project outline and prototype due yesterday
 - Final project presentation due in two weeks
 - Final project report due in three weeks
- Next Week: Guest Lecturers
 - Kolina "Koko" Koltai: PhD candidate on Ethics
 - Beth Hallmark: Career Development Director
 - Lab: video creation training
- Questions?

Final Project: Resources and Next Steps

- Writing center: <http://uwc.utexas.edu/>
 - can schedule four individual 45-minutes consultation per month
- Tutoring
 - <https://utdirect.utexas.edu/apps/ugs/my/tutoring/student/tutoring-agreement/>
- Final project presentation, peer review, and paper

Today's Topics

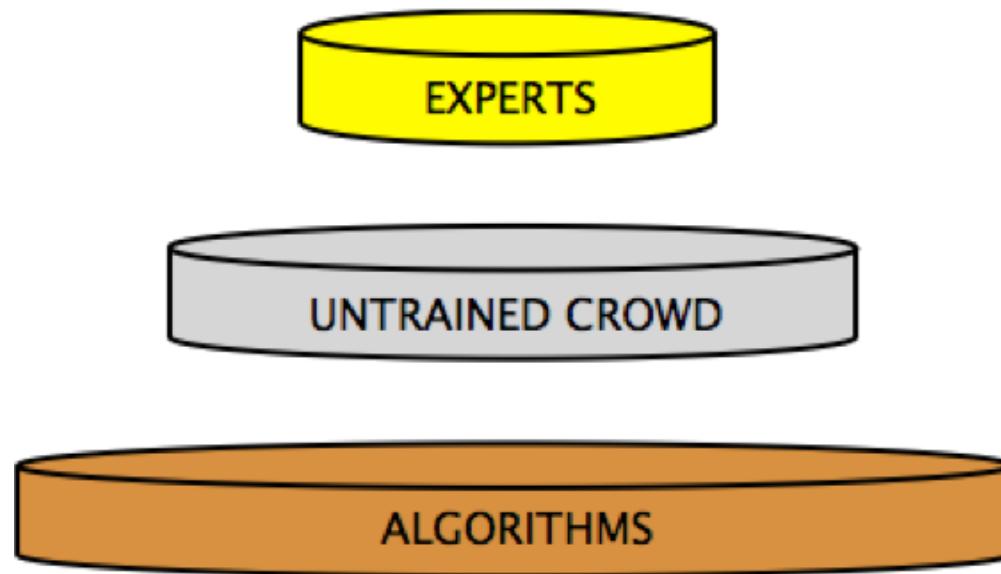
- Dataset Creation
- Getting More Out of Your Training Data
- Human-Machine Partnerships
- Understanding Machine Learning Algorithms

Today's Topics

- Dataset Creation
- Getting More Out of Your Training Data
- Human-Machine Partnerships
- Understanding Machine Learning Algorithms

Crowdsourcing

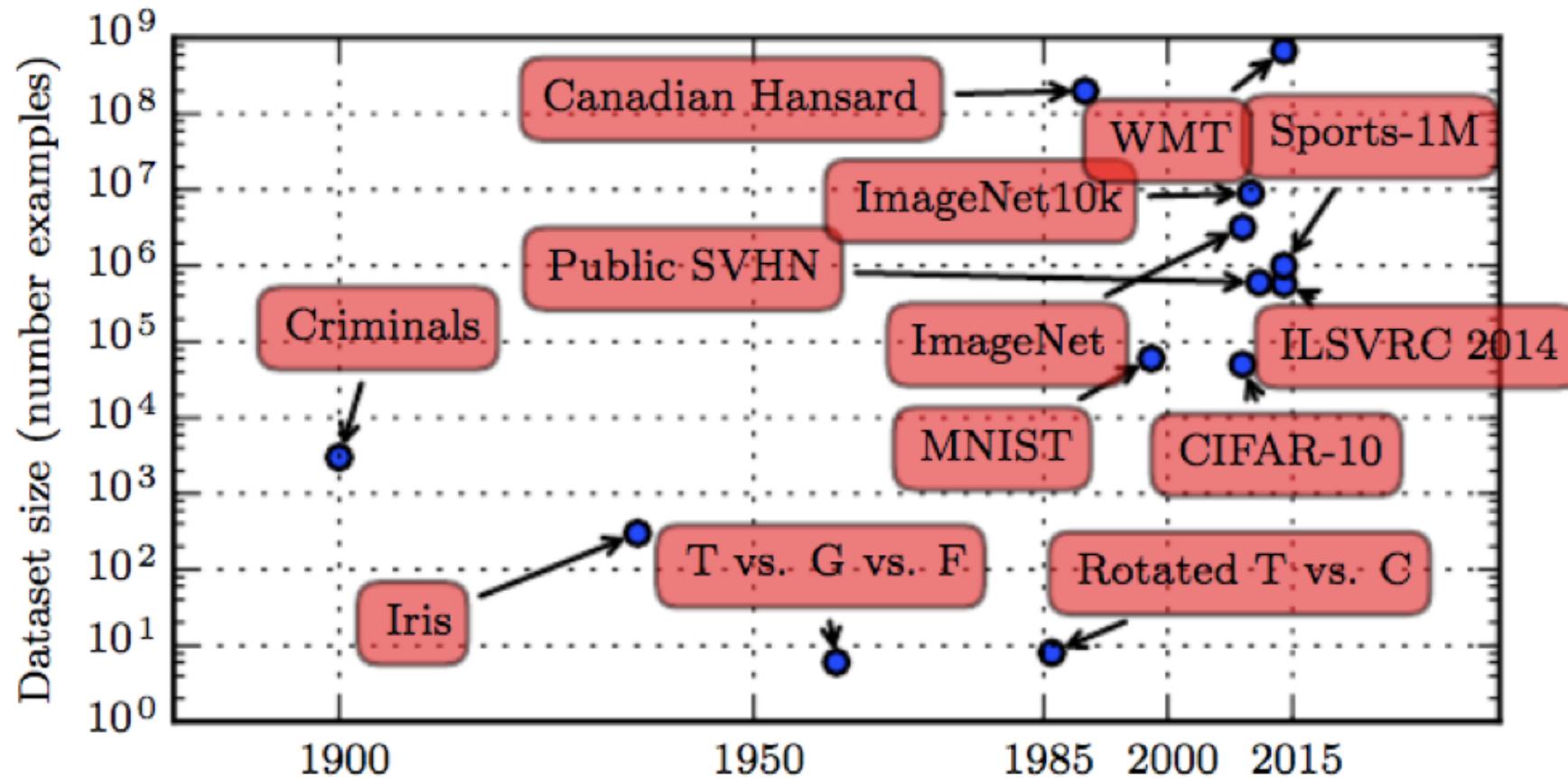
Crowds offer a middle-ground between expensive “experts” and scalable algorithms



Group Discussion

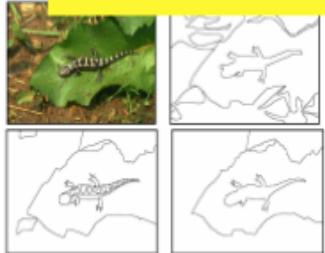
- Toy Problem: predict if a comment on a social network is toxic or not
- Create a budget and process for creating a labeled dataset in-house
- Create a budget and process for creating a labeled dataset with anonymous online crowd workers

Historical Overview of Datasets

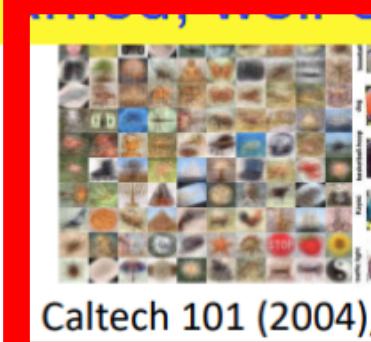


Historical Overview of Datasets

A well-framed, well-curated moment in time



BSD (2001)



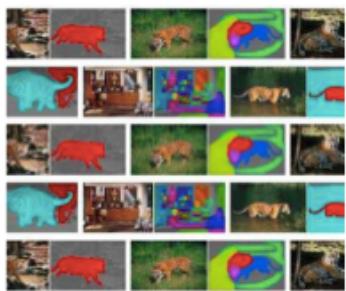
Caltech 101 (2004)



Caltech 256 (2006)



PASCAL (2007-12)



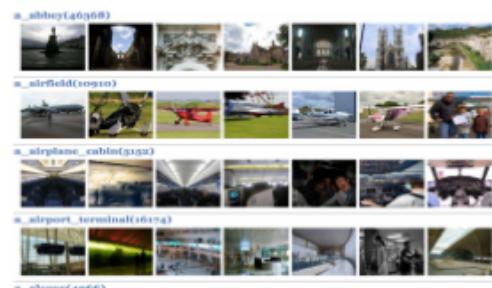
LabelMe (2007)



ImageNet (2009)



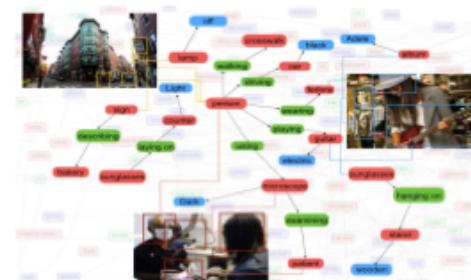
SUN (2010)



Places (2014)



MS COCO (2014)

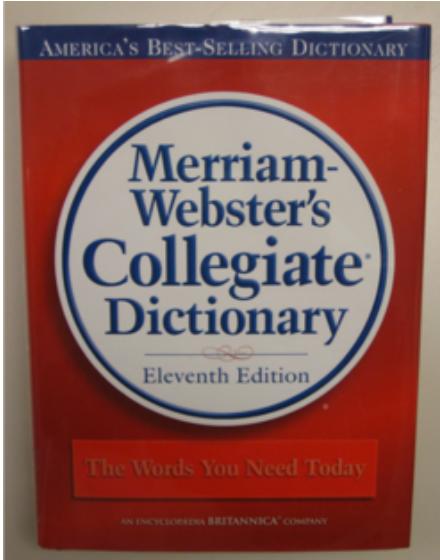


Visual Genome (2016)

Slide Credit:
<http://www.cs.utexas.edu/~grauman/slides/grauman-epic-iccv2017.pdf>

Caltech-101: In-House Annotators

1. Category Selection



Flipped through a dictionary and chose 101 categories associated with a drawing

2. Image Collection



3. Image Verification

2 graduate students reviewed all images and discarded irrelevant images

Images: 9,144

Historical Overview of Datasets

A well-framed, well-curated moment in time



BSD (2001)



Caltech 101 (2004), Caltech 256 (2006)



PASCAL (2007-12)



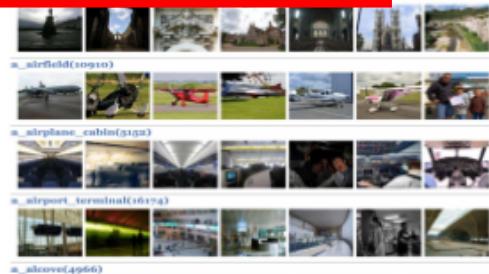
LabelMe (2007)



ImageNet (2009)



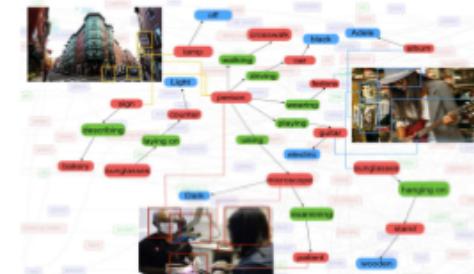
SUN (2010)



Places (2014)



MS COCO (2014)



Visual Genome (2016)

Slide Credit:
<http://www.cs.utexas.edu/~grauman/slides/grauman-epic-iccv2017.pdf>

LabelMe

1. Image Collection 2. Object Segmentation 3. Category Assignment

- Most images taken by authors using a variety of hand-held digital cameras.
- Also includes video sequences taken with a head-mounted web camera.

- Any user that logs in uses online tool

- User that segments the object also labels the object

LabelMe: Online Database and Tool

The screenshot shows the LabelMe web application interface. At the top, there is a navigation bar with links for 'My LabelMe', 'Publications', 'Developers', 'Help', and 'Credits'. Below the navigation bar, the main content area displays the user's collections. A red sidebar on the left contains links for 'My Collections (Home)', 'Public Collections', 'Account Settings', 'Change Password', and 'Log Out'. The main content area shows four collections under 'Home: dannag' (4 collections): 'Collection: /tests' (6 items), 'Collection: /wacv' (2 items), and 'Collection: /mef' (2 items). A 'Collection' button is located in the top right corner of the main content area.

labelme2.csail.mit.edu/Release3.0/browserTools/php/browse_collections.php?username=dannag

My LabelMe Publications Developers Help Credits

Home: dannag
4 collections

Collection: /tests
6 items

Collection: /wacv
2 items

Collection: /mef
2 items

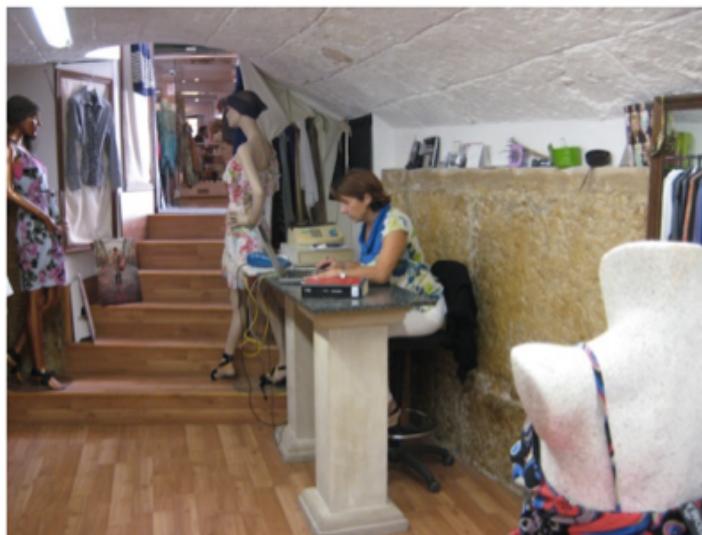
Collection

LabelMe: Online Database

"I work in a small clothing shop. The shop is open from 10am to 8pm with only a short break at 2pm. Despite the long working hours I have a lot of free time. As I am the owner of the shop, I can do whatever I want during that time. I am always ready for the clients, however, in such a long day there are many hours of inactivity. I used to read a lot and books passed by my hands a great speed. I was starting to lose the pleasure that one feels when reading a good book. For this reason, when I started working with LabelMe it was very satisfying to know that I was doing something that had some scientific value and that it could be of use for somebody in the future."



Antonio



"she has labeled more than 250,000 objects"



Antonio's Mom

Adela Barriuso and **Antonio Torralba**; arXiv 2012; *16 citations in 2/17*

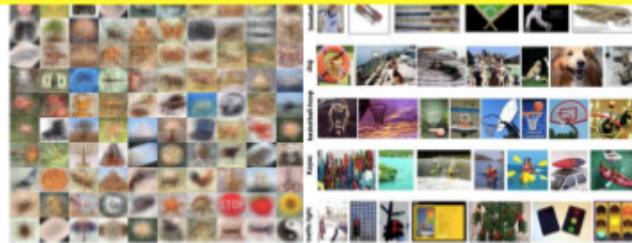
Bryan. C. Russell, **Antonio Torralba**, Kevin P. Murphy, and William T. Freeman; IJCV 2008; *1950 citations in 2/17*

Historical Overview of Datasets

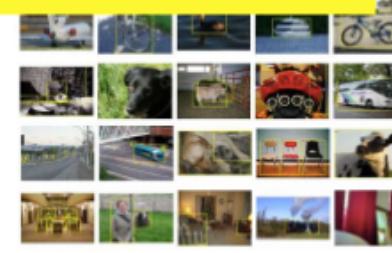
A well-framed, well-curated moment in time



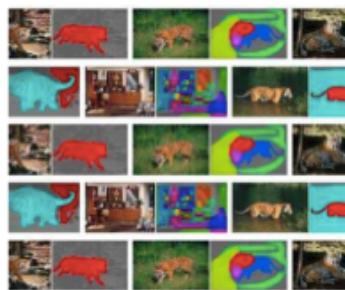
BSD (2001)



Caltech 101 (2004), Caltech 256 (2006)



PASCAL (2007-12)



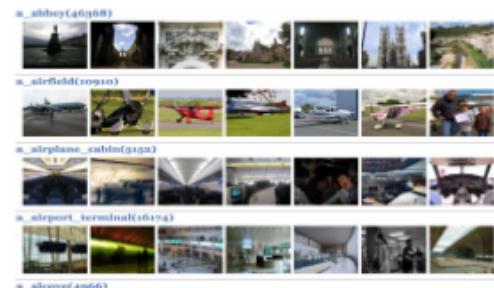
LabelMe (2007)



ImageNet (2009)



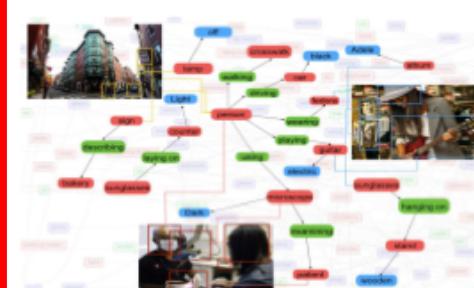
SUN (2010)



Places (2014)



MS COCO (2014)



Visual Genome (2016)

Slide Credit:
<http://www.cs.utexas.edu/~grauman/slides/grauman-epic-iccv2017.pdf>

MSCOCO

Include “things”: objects that can easily be labeled; e.g., person, chair



Exclude “stuff”: objects with no clear boundaries; e.g., sky, grass,



Rationale: primary interest is in precise localization of object instances

MSCOCO

“What is the difference between labeling ‘things’ and labeling ‘stuff’?”

Include “things”: objects that can easily be labeled; e.g., person, chair



Exclude “stuff”: objects with no clear boundaries; e.g., sky, grass,



Rationale: primary interest is in precise localization of object instances

Selected 91 from 272 categories in bold

(without *)

person	bicycle	car	motorcycle	bird	cat	dog	horse	sheep	bottle
chair	couch	potted plant	tv	cow	airplane	hat*	license plate	bed	laptop
fridge	microwave	sink	oven	toaster	bus	train	mirror*	dining table	elephant
banana	bread	toilet	book	boat	plate*	cell phone	mouse	remote	clock
face	hand	apple	keyboard	backpack	steering wheel	wine glass	chicken	zebra	shoe*
eye	mouth	scissors	truck	traffic light	eyeglasses*	cup	blender*	hair drier	wheel
street sign*	umbrella	door*	fire hydrant	bowl	teapot	fork	knife	spoon	bear
headlights	window*	desk*	computer	refrigerator	pizza	squirrel	duck	frisbee	guitar
nose	teddy bear	tie	stop sign	surfboard	sandwich	pen/pencil	kite	orange	toothbrush
printer	pans	head	sports ball	broccoli	suitcase	carrot	chandelier	parking meter	fish
handbag	hot dog	stapler	basketball hoop	donut	vase	baseball bat	baseball glove	giraffe	jacket
skis	snowboard	table lamp	egg	door handle	power outlet	hair	tiger	table	coffee table
skateboard	helicopter	tomato	tree	bunny	pillow	tennis racket	cake	feet	bench
chopping board	washer	lion	monkey	hair brush*	light switch	arms	legs	house	cheese
goat	magazine	key	picture frame	cupcake	fan (ceil/floor)	frogs	rabbit	owl	scarf
ears	home phone	pig	strawberries	pumpkin	van	kangaroo	rhinoceros	sailboat	deer
playing cards	towel	hippo	can	dollar bill	doll	soup	meat	window	muffins
tire	necklace	tablet	corn	ladder	pineapple	candle	desktop	carpet	cookie
toy cars	bracelet	bat	balloon	gloves	milk	pants	wheelchair	building	bacon
box	platypus	pancake	cabinet	whale	dryer	torso	lizard	shirt	shorts
pasta	grapes	shark	swan	fingers	towel	side table	gate	beans	flip flops
moon	road/street	fountain	fax machine	bat	hot air balloon	cereal	seahorse	rocket	cabinets
basketball	telephone	movie (disc)	football	goose	long sleeve shirt	short sleeve shirt	raft	rooster	copier
radio	fences	goal net	toys	engine	soccer ball	field goal posts	socks	tennis net	seats
elbows	aardvark	dinosaur	unicycle	honey	legos	fly	roof	baseball	mat
ipad	iphone	hoop	hen	back	table cloth	soccer nets	turkey	pajamas	underpants
goldfish	robot	crusher	animal crackers	basketball court	horn	firefly	armpits	nectar	super hero costume

MSCOCO

1. Category Selection

- 272 categories:
 - 1) WordNet, SUN, VOC, ...
 - 2) Most frequent words describing visual objects
 - 3) 4-8 yr olds listing objects in indoors/outdoors
 - 91 by author votes + coverage

2. Image Collection

- Search on Flickr because it often has non-iconic images
 - Query: object + object or scene + scene
 - Query: unusual categories



Commonly retrieved with Google, Bing, etc

Goal for this paper
images with **contextual** information and taken from **non-canonical** viewpoints

MSCOCO

1. Category Selection

- 272 categories:
 - 1) WordNet, SUN, VOC, ...
 - 2) Most frequent words describing visual objects
 - 3) 4-8 yr olds listing objects in indoors/outdoors
- 91 by author votes + coverage

2. Image Collection

- Search on Flickr because it often has non-iconic images
 - Query: object + object or scene + scene
 - Query. unusual categories
 - Quality control with crowd

Users remove invalid or iconic images from grid of 128 images:

Task: select images that contain **BOTH** a person AND a bicycle

Instructions:
Please click and select images that contain **BOTH** a person(s) **AND** a bicycle(s).
Do **NOT** select an image that contains **ONLY** a person(s) or **ONLY** a bicycle(s).
(It is right to not select any image if none of image contains both categories.)

cartoons paintings

MSCOCO

1. Category Selection

- 272 categories:
 - 1) WordNet, SUN, VOC, ...
 - 2) Most frequent words describing visual objects
 - 3) 4-8 yr olds listing objects in indoors/outdoors
 - 91 by author votes + coverage

2. Image Collection

- Search on Flickr because it often has non-iconic images
 - Query: object + object or scene + scene
 - Query: unusual categories
 - Quality control with crowd

Users remove invalid or iconic images from grid of 128 images:

Task: select images that contain a bear(s)

Instructions:
Please click and select images that contain **MULTIPLE** objects AND at least one bear(s).

Do NOT select an image that contains **ONLY** a bear(s).

Do NOT select an image that contains **NO** bear(s).

You can de-select the image by clicking on it again.
Please do not select cartoons or paintings.

cartoons paintings

MSCOCO

1. Category Selection 2. Image Collection 3. Object presence labeling

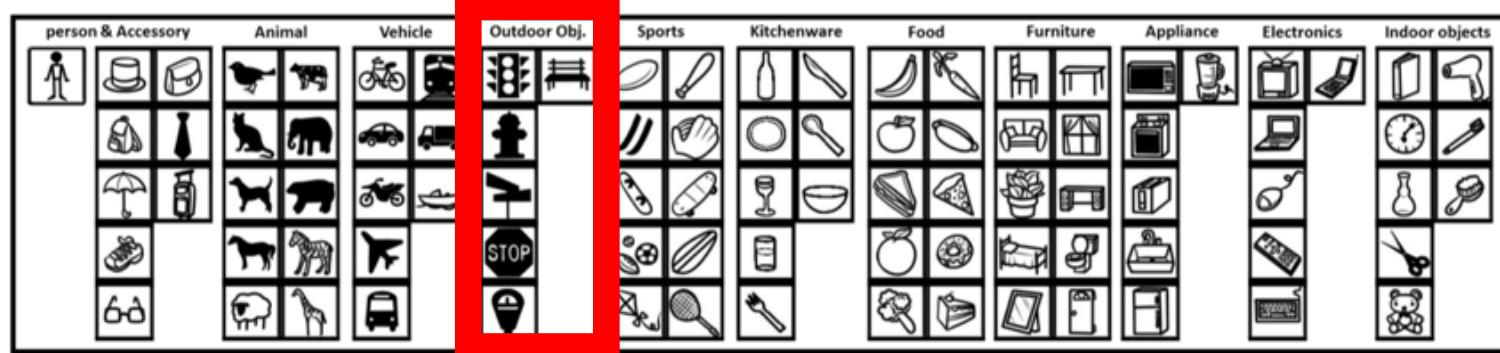
- 272 categories:
1) WordNet,
SUN, VOC, ...
2) Most frequent
words describing
visual objects
3) 4-8 yr olds
listing objects in
indoors/outdoors
- 91 by author
votes + coverage

- Search on
Flickr because it
often has non-
iconic images
- Query: object
+ object or
scene + scene
- Query: unusual
categories
- Quality control
with crowd

- Assign all
relevant
categories to
each image by
locating one
instance

MSCOCO – object presence

11 Groupings



For high recall, 8 people per image were solicited to do this task

Instructions (PLEASE ACCEPT THE HIT TO GET STARTED):
Please drag and drop icons from the bottom panel to matching objects in the image. If an icon matches multiple objects you can drag the icon onto any of the objects. There are 11 sets of objects to drag onto the image. Use the buttons or arrow keys to cycle through them. There are total of 8 images to label.
Please drag and drop ICONS to matching objects in the image.

Here is an example of a labeled image:

Task: select small indoor items shown in the image (if any):

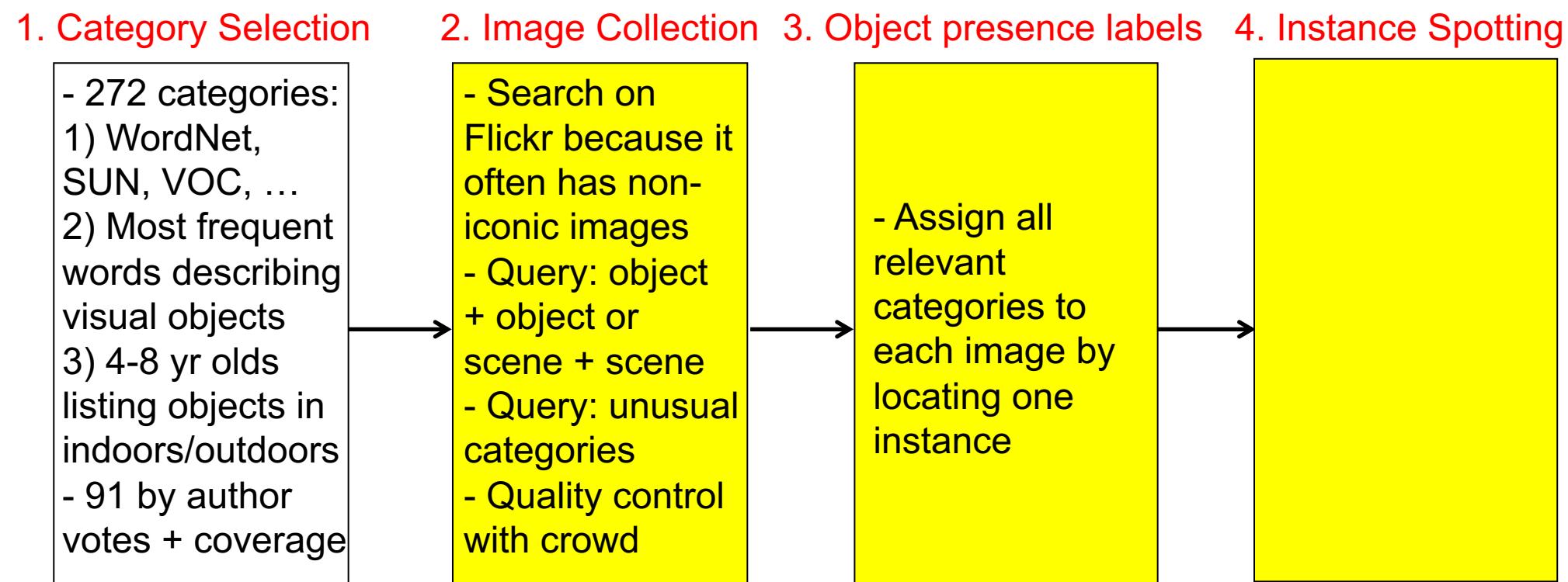
Image 1:

Image contains:

Task: select outdoor objects shown in the image (if any):

4/11

MSCOCO



MSCOCO

“magnifying glass” feature: doubles resolution of currently selected region to assist with small objects.

Instructions (PLEASE ACCEPT THE HIT TO GET STARTED):

- Mark [each occurrence](#) (if any) of the following object: [cow](#).
- You only need to mark up to 10 instances if multiple cow(s) exist in the image. It is possible for some images that this object does not appear.
- The blinking icon (Hint) shows where one instance of the object could be. The Hint is [NOT ALWAYS](#) correct.
- Type [N](#) to go to the next image and [B](#) to go back.
- There are 50 images in this HIT.

Good Example



Left Click: Add marker

Bad Example
(Do not click)



Right Click: Delete marker

Drag & Drop Move marker

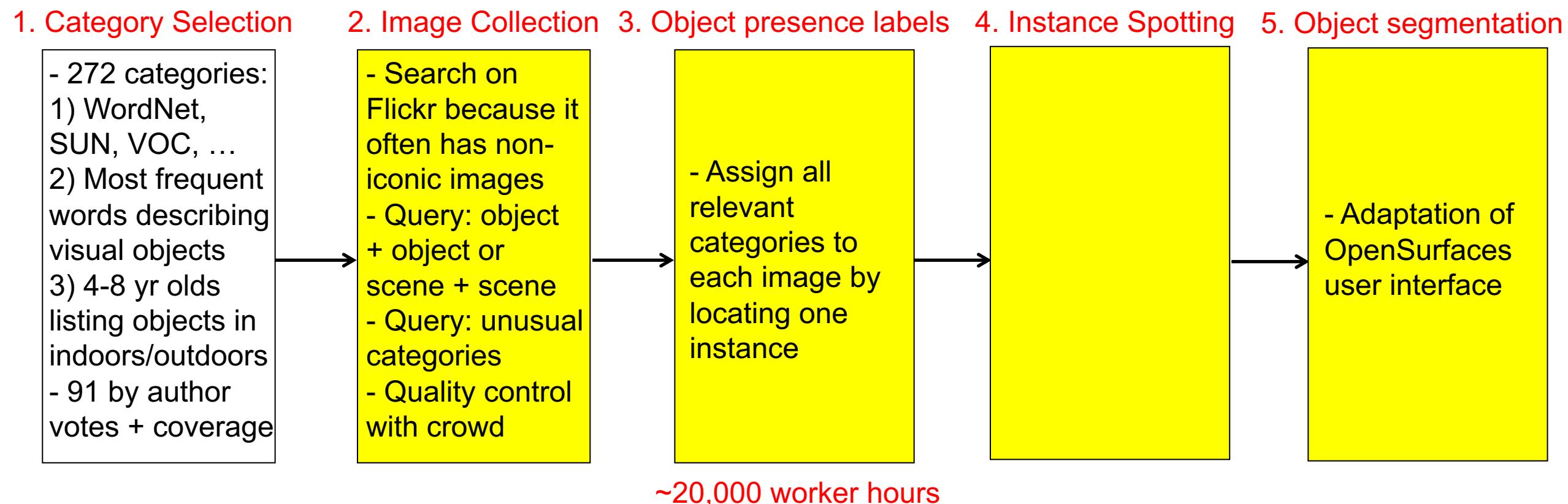
 7 cow(s) found in this image.

COW

Back [B] Next [N] Hint [H]



MSCOCO



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; 1,955 citations in 2/17

MSCOCO

Instructions: carefully trace around regions that have a **single sports ball** indicated by the icon. (1/3)

Not sure what object sports ball is? Click on [here](#) to see examples!

Draw (D) Adjust (A) Undo (Ctrl-Z) Redo (Ctrl-Y) Close (Right-click) Delete (Delete)
Move to Target (M) Zoom In (I) Zoom Out (O) Reset Zoom (ESC)

Please Accept HIT to get started! [Examples](#) [Instructions](#)

Tips: Using "[Move to taget](#)" (M) and "[Zoom In](#)" (I) for the small object!
Please pay attentions to trace boundary carefully. Work will be rejected if not follow the instruction.

Training task per object category required.

MSCOCO

Seeded gold standards: 4 of 64 segmentation known to be bad; a worker must identify 3 of the 4 known bad segmentations to complete the task.

Verification step: 3-5 workers judged each segmentation and indicate whether it matched the instance well or not.

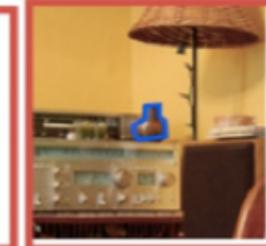
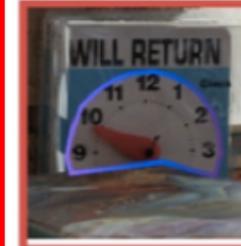
Blocked workers: those who often produced poor segmentations were blocked and their work not used

64 examples

Task: select images that have **WRONG** object contour for toothbrush.
Examples:
Right Object Contour



Wrong Object Contour (not toothbrush, only contains parts of visible object contour, or multiple objects)



Tips: use **n** and **b** keys to move between rows of image.



MSCOCO

- Crowd labeling is similar to semantic segmentation as object instances are not individually identified.
- Crowd labeling is employed for images containing 10+ instances of an object category.

Draw all unlabeled **person(s)** in the image.

- Find and draw on **all person(s)** that haven't been labeled.
- It's okay to overlap to labeled region.
- You need to label two images that contain unlabeled person(s) to complete
- Work will be rejected if **not carefully** drawn or unlabeled person(s) remain.

Submit No Unlabeled Person(s)

Draw (D) Erase (E) Zoom In (Z) Zoom Out (X)

© 2006 John A. Marsh - www.johnmarshphotography.com

For More, You Can Take....

- iSchool Courses:
 - Crowdsourcing for Computer Vision
 - Human Computation&Crowdsourcing

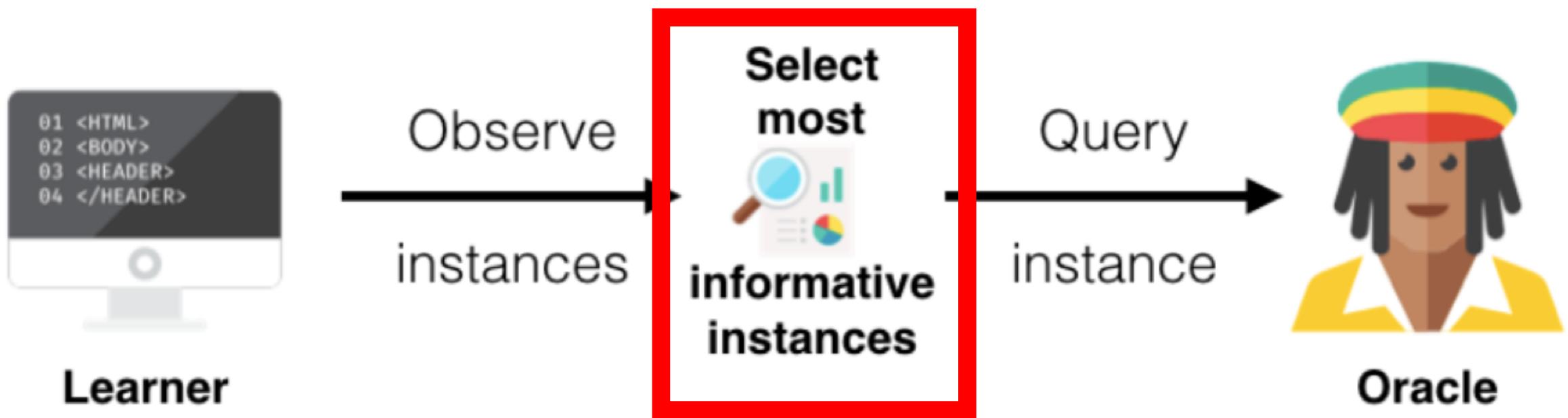
Today's Topics

- Dataset Creation
- Getting More Out of Your Training Data
- Human-Machine Partnerships
- Understanding Machine Learning Algorithms

Group Discussion

Given that training data is expensive, how would you design a machine learning algorithm to learn more with its data?

Active Learning



Curriculum Learning

Decide order in which you provide data.

Today's Topics

- Dataset Creation
- Getting More Out of Your Training Data
- Human-Machine Partnerships
- Understanding Machine Learning Algorithms

Interactive Vision Applications

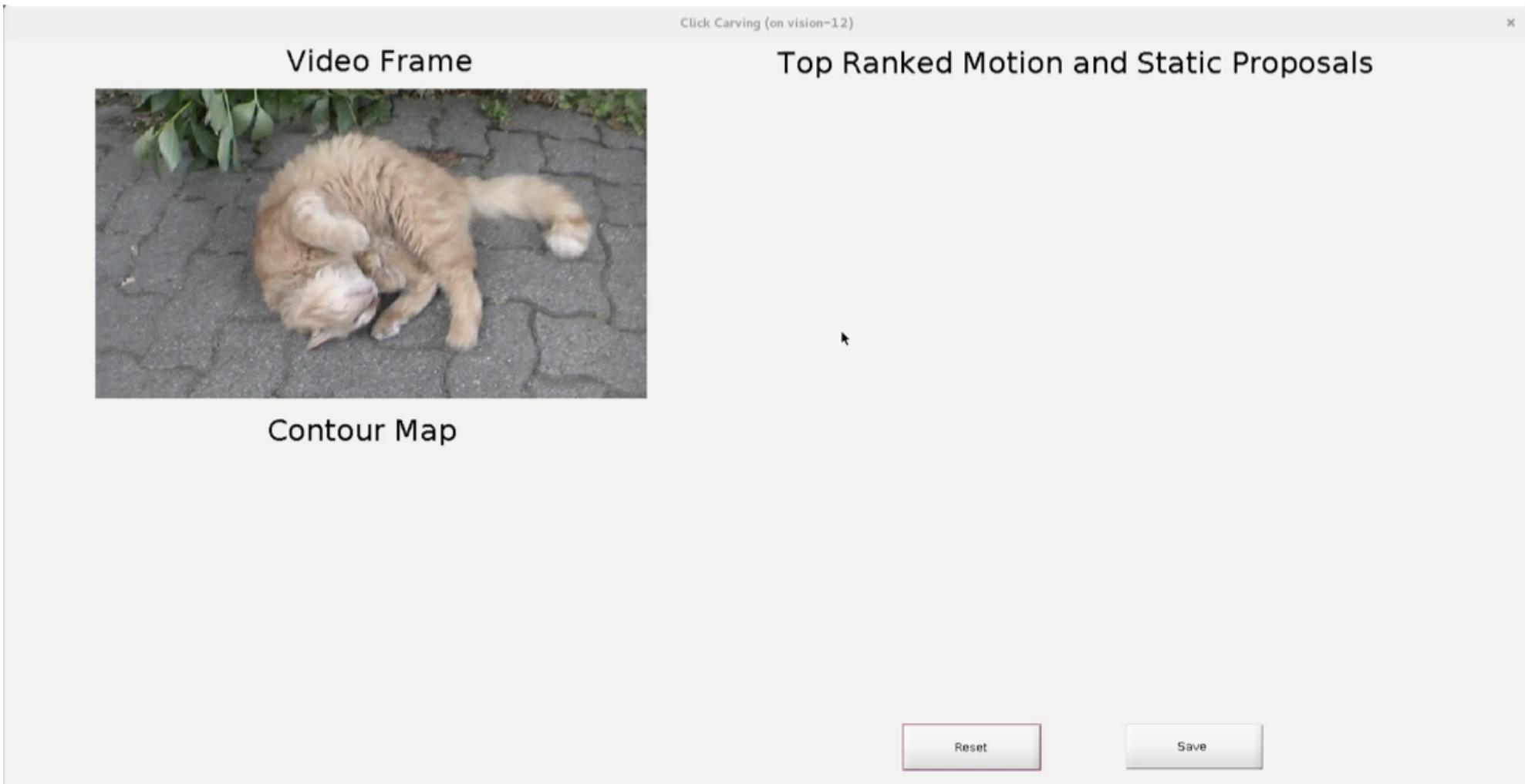


Pinterest

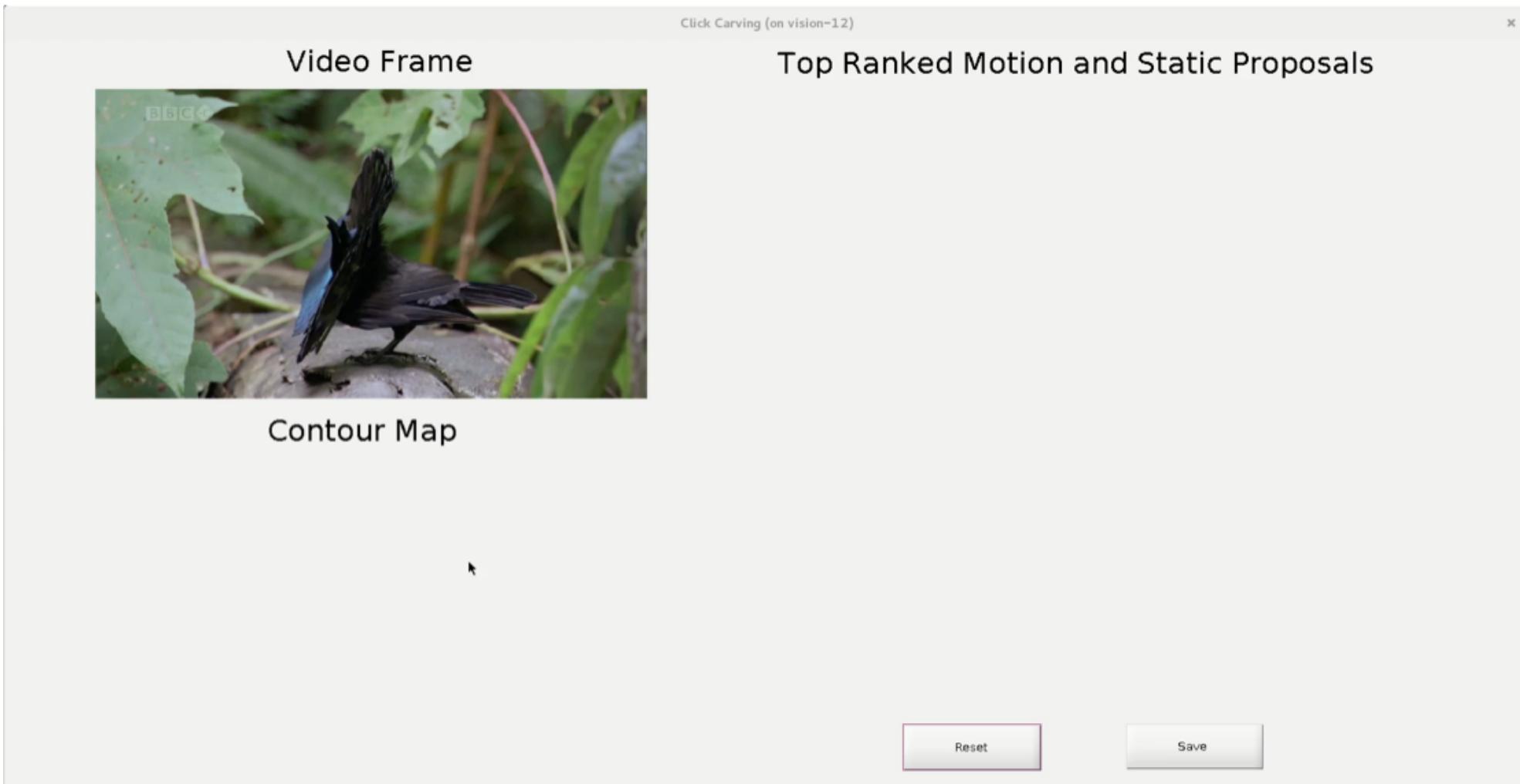


<http://ibird.com/>

Object Segmentation: Click Carving



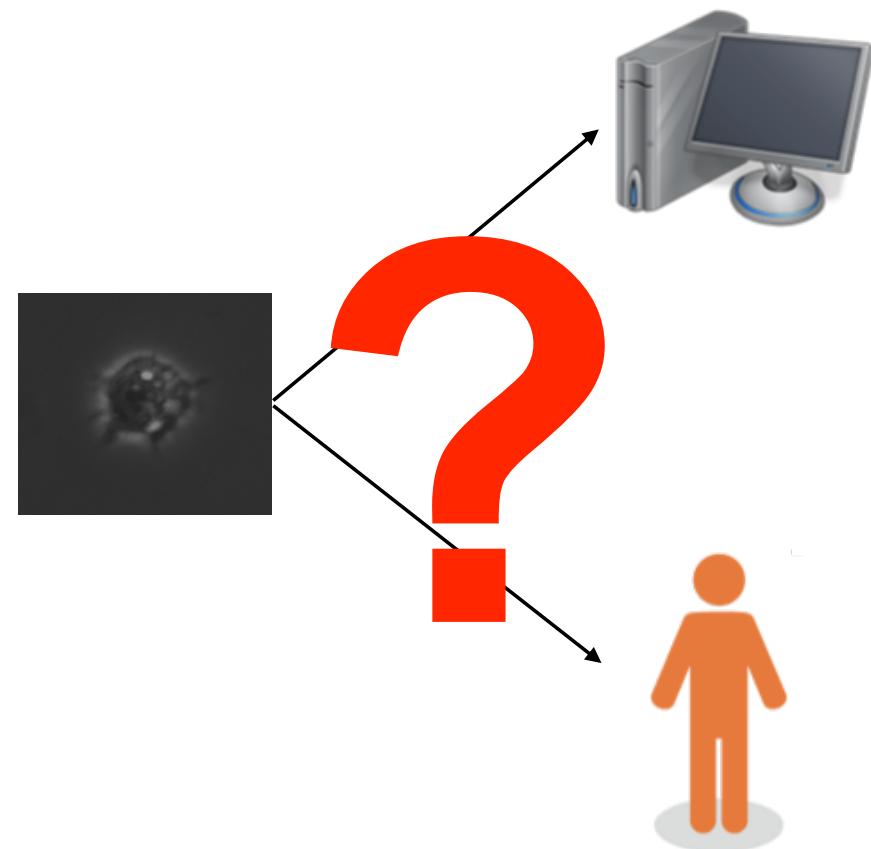
Object Segmentation: Click Carving



Why Interactive Vision?

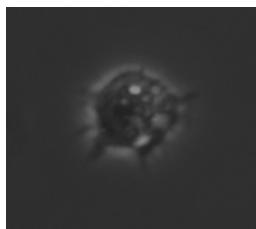
- Tasks that are
 - Time-consuming
 - Difficult for people

Divide Labor Between Humans & Machines

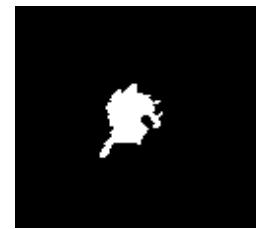


Predict Quality of Algorithm Results

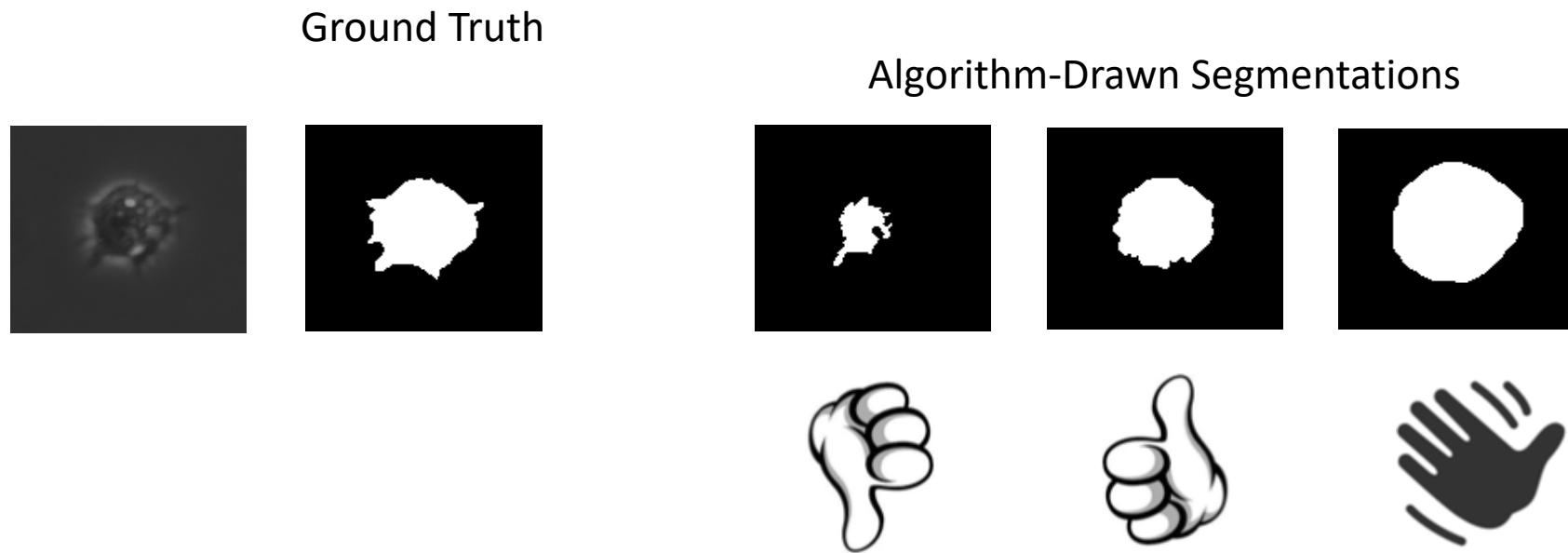
Ground Truth



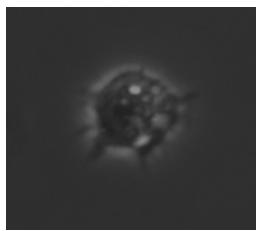
Algorithm-Drawn Segmentations



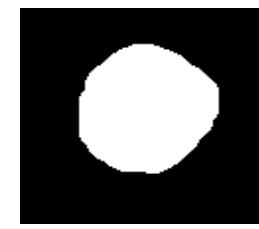
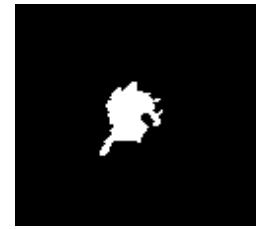
Predict Quality of Algorithm Results



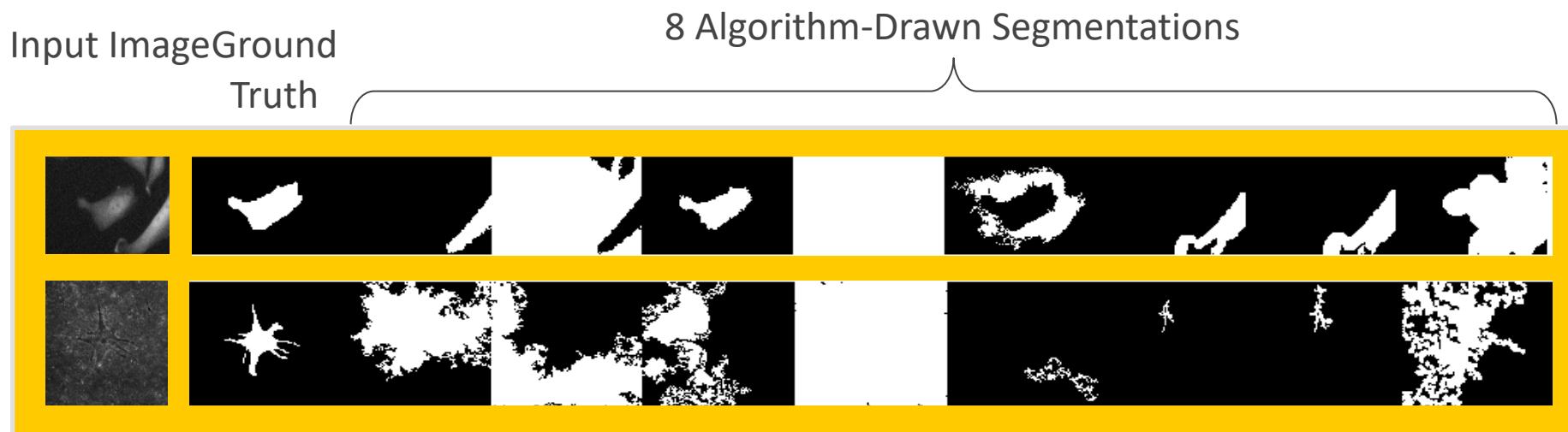
Predict Quality of Algorithm Results



Algorithm-Drawn Segmentations

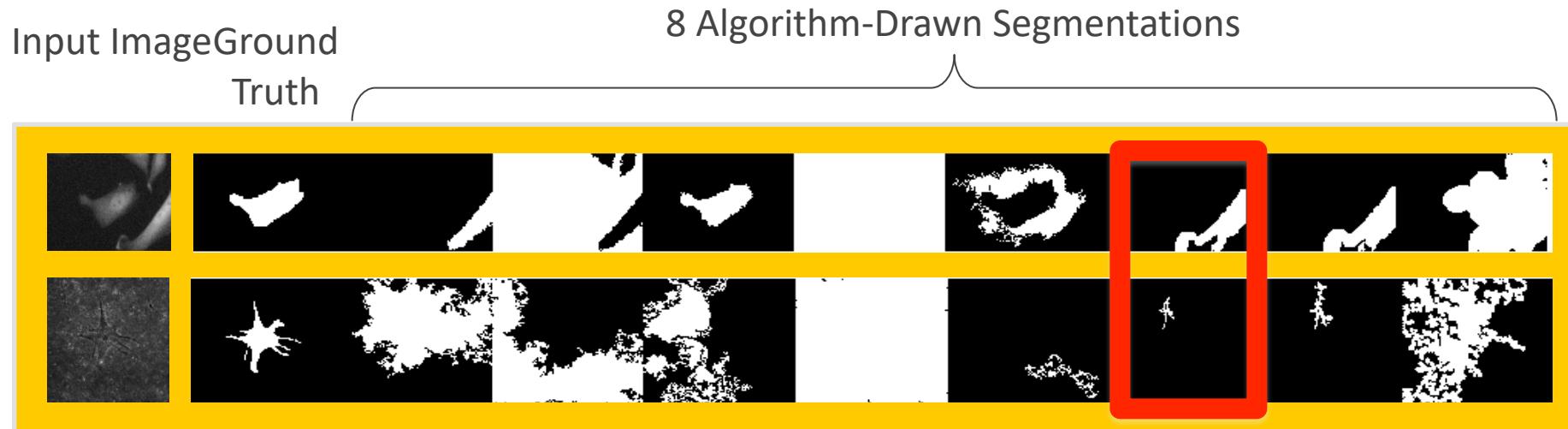


What Features Show Algorithms Fail?



What Features Show Algorithms Fail?

Predictive Cues
Object Location - Edges

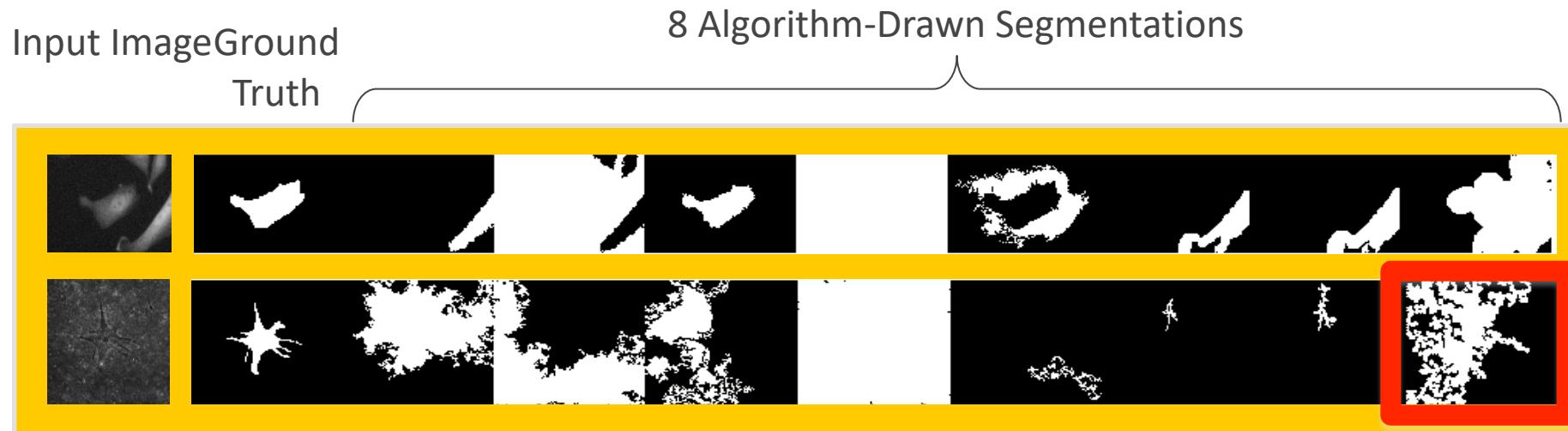


What Features Show Algorithms Fail?

Predictive Cues

Object Location - Edges

Object Boundary - Very Jagged



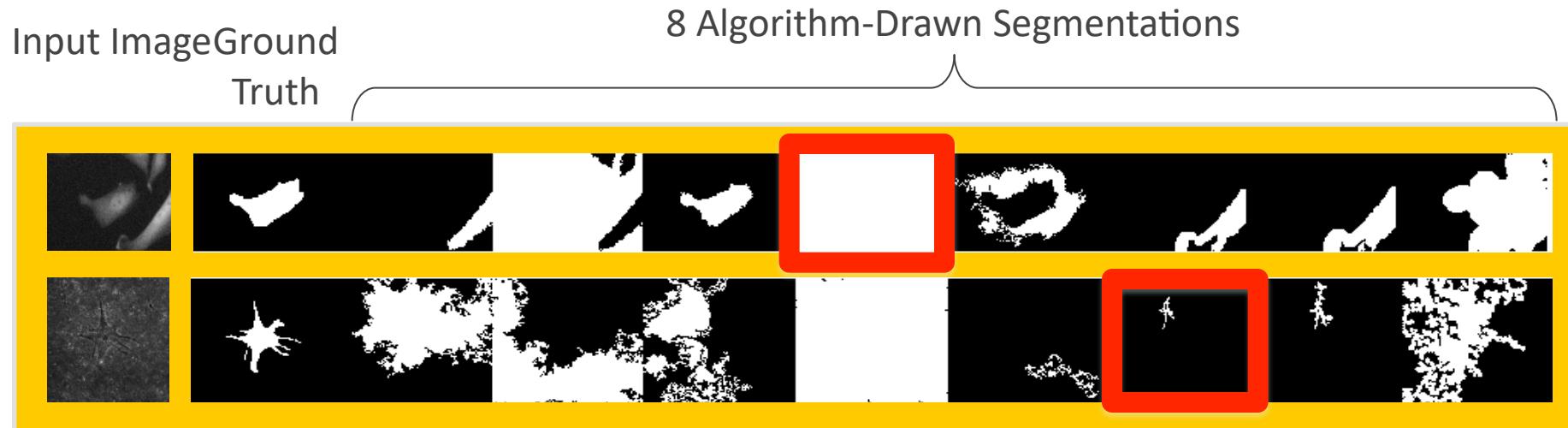
What Features Show Algorithms Fail?

Predictive Cues

Object Location - Edges

Object Boundary - Very Jagged

Image Coverage - Abnormally Large & Small



System Training

Training Instances

- 522 images
- 5,742 segmentations
 - for each image, 8 algorithms + 3 from ground truth

System Training

Training Instances

- 522 images
- 5,742 segmentations
 - for each image, 8 algorithms + 3 from ground truth

Training Features and Scores

For each training instance,

1. Evaluate segmentation similarity to ground truth

2. Compute 9 features per segmentation

Score	F_1	F_2	• • •	F_9
Seg 1:	0.9	0.7	0.2	• • • 0.81
Seg 5742:	0.13	0.5	0.2	• • • 0.3

System Training

Training Instances

- 522 images
- 5,742 segmentations
 - for each image, 8 algorithms + 3 from ground truth

Training Features and Scores

For each training instance,

1. Evaluate segmentation similarity to ground truth

2. Compute 9 features per segmentation

	Score	F_1	F_2	• • •	F_9
Seg 1:	0.9	0.7	0.2	• • •	0.81
Seg 5742:	0.13	0.5	0.2	• • •	0.3

$$y = X\beta + \alpha$$

Model

Bias Term

Training Model

- Linear Regression

System Training

Training Instances

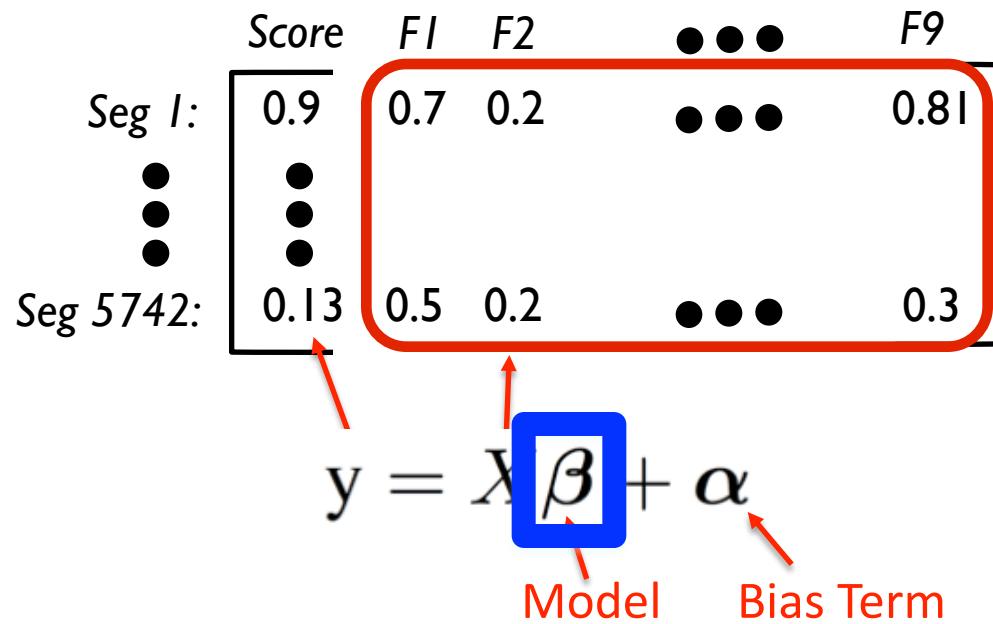
- 522 images
- 5,742 segmentations
 - for each image, 8 algorithms + 3 from ground truth

Training Features and Scores

For each training instance,

1. Evaluate segmentation similarity to ground truth

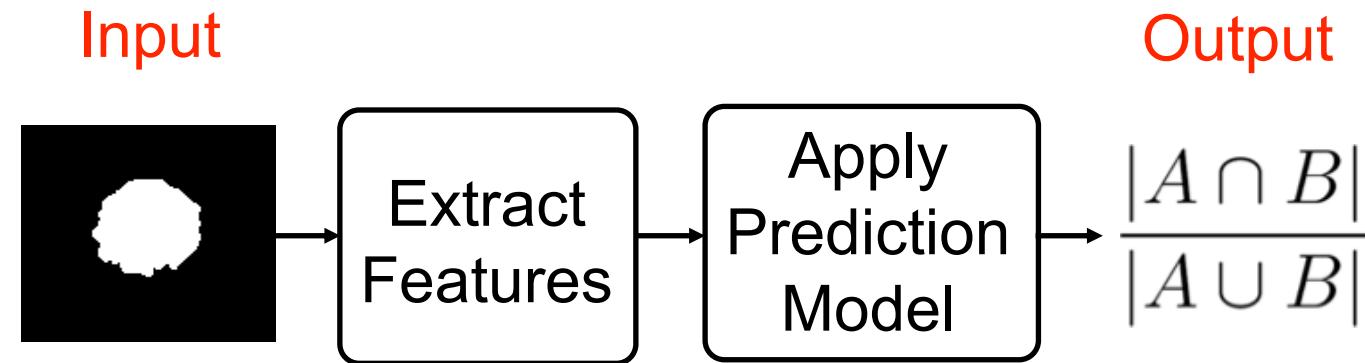
2. Compute 9 features per segmentation



Training Model

- Linear Regression

System Use for Novel Image



Evaluation

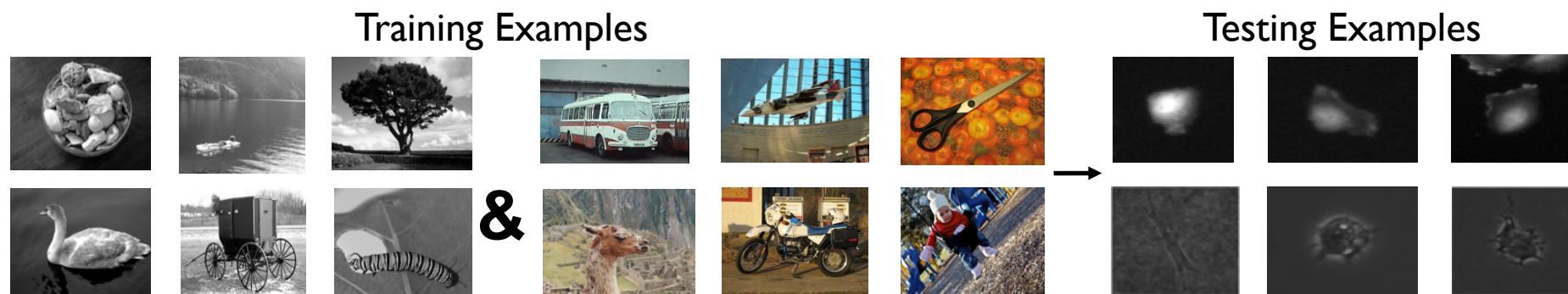
	Weizmann 1,100 Segs		IIS 1,661 Segs		BU-BIL 2,981 Segs	
	CC	MAE	CC	MAE	CC	MAE
Ours:	0.64	0.24	0.68	0.22	0.61	0.31
CPMC [CVPR 2010]:	0.61	0.32	0.67	0.31	0.36	0.33
AlexNet features [NIPS 2012]:	-0.1	26.7	-0.01	45	-0.01	3.22

CC: Correlation Coefficient, MAE: Mean Absolute Error

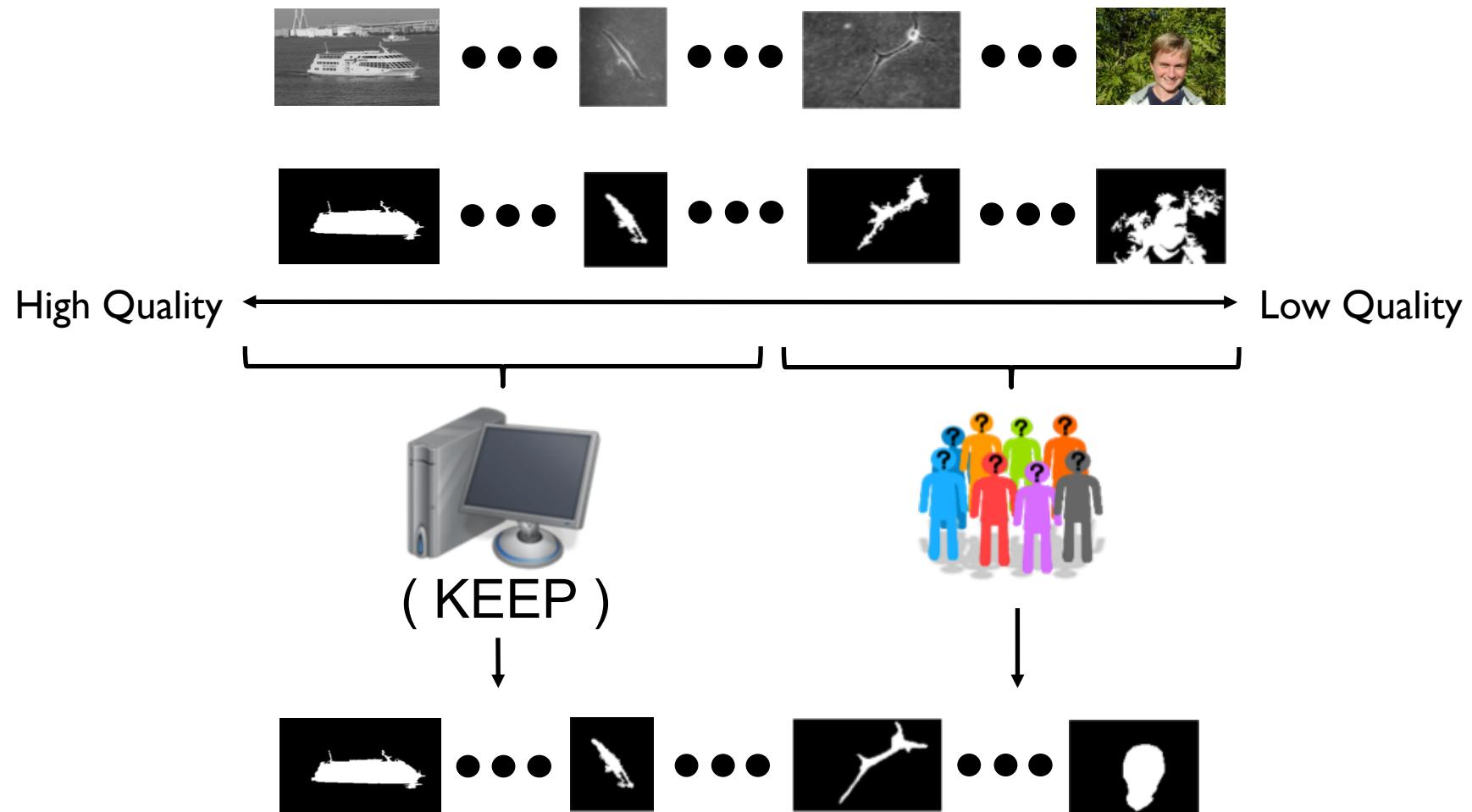
Evaluation

	Weizmann 1,100 Segs		IIS 1,661 Segs		BU-BIL 2,981 Segs	
	CC	MAE	CC	MAE	CC	MAE
Ours:	0.64	0.24	0.68	0.22	0.61	0.31
CPMC [CVPR 2010]:	0.61	0.32	0.67	0.31	0.36	0.33
AlexNet features [NIPS 2012]:	-0.1	26.7	-0.01	45	-0.01	3.22

CC: Correlation Coefficient, MAE: Mean Absolute Error



Predict When to Use Human vs Machine



Related Work

Predicting Failures of Vision Systems

Peng Zhang¹ Jiuling Wang² Ali Farhadi³ Martial Hebert⁴ Devi Parikh¹

¹Virginia Tech ²Univ. of Texas at Austin ³Univ. of Washington ⁴Carnegie Mellon University

¹{zhangp, parikh}@vt.edu ²jiuling@utexas.edu ³ali@cs.uw.edu ⁴hebert@ri.cmu.edu

Today's Topics

- Dataset Creation
- Getting More Out of Your Training Data
- Human-Machine Partnerships
- Understanding Machine Learning Algorithms

Interpretability

- Interpretable features
 - Linear regression models
 - Decision trees
- Ablation studies
 - Remove different features and report performance

Try to Break Them

who is the president

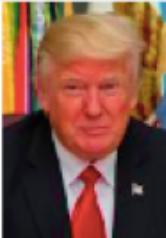
All Images News Videos Maps More Settings Tools

About 73,600,000 results (0.80 seconds)

United States of America / President

Donald Trump

Quotes and overview



how old is donald trump

All News Images Shopping Books More Settings Tools

About 9,730,000 results (0.66 seconds)

Donald Trump / Age

71 years
June 14, 1946



People also search for

 Melania Trump 47 years	 Barack Obama 56 years	 Hillary Clinton 69 years
---	--	---

how old is the president

All News Images Shopping Maps More Settings Tools

About 116,000,000 results (0.63 seconds)

Barack Obama / Age

56 years
August 4, 1961



People also search for

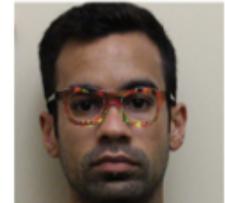
 Donald Trump 71 years	 Michelle Obama 53 years	 Hillary Clinton 69 years
--	--	---

Object Recognition Applications Gone Wrong

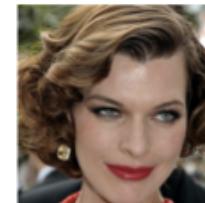
- Deadly Mistake: Self-Driving Car Crash
 - http://www.nytimes.com/interactive/2016/07/01/business/inside-tesla-accident.html?_r=0



- Ethical Mistake: Photo Tagging
 - <http://www.usatoday.com/story/tech/2015/07/01/google-apologizes-after-photos-identify-black-people-as-gorillas/29567465/>



- Security Mistake: Person Recognition
 - <https://www.cs.cmu.edu/~sbhagava/papers/face-rec-ccs16.pdf>



Object Recognition Applications Gone Wrong

- Deadly Mistake: Self-Driving Car Crash
 - http://www.nytimes.com/interactive/2016/07/01/business/inside-tesla-accident.html?_r=0

**Why do you think these
systems make such mistakes?**

- Security Mistake: Person Recognition
 - <https://www.cs.cmu.edu/~sbhagava/papers/face-rec-ccs16.pdf>

Object Recognition Applications Gone Wrong

- Deadly Mistake: Self-Driving Car Crash
 - http://www.nytimes.com/interactive/2016/07/01/business/inside-tesla-accident.html?_r=0

If you were the CEO, how

- Ethical Mistake: Photo Tagging
 - <http://www.usatoday.com/story/tech/2015/11/1/google-app-lets-after-photos-identify-black-people-as-gorillas/29567465/>

**would you change your
product in response?**

- Security Mistake: Person Recognition
 - <https://www.cs.cmu.edu/~sbhagava/papers/face-rec-ccs16.pdf>

Group Discussion

- Find ways to break different online AI systems