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

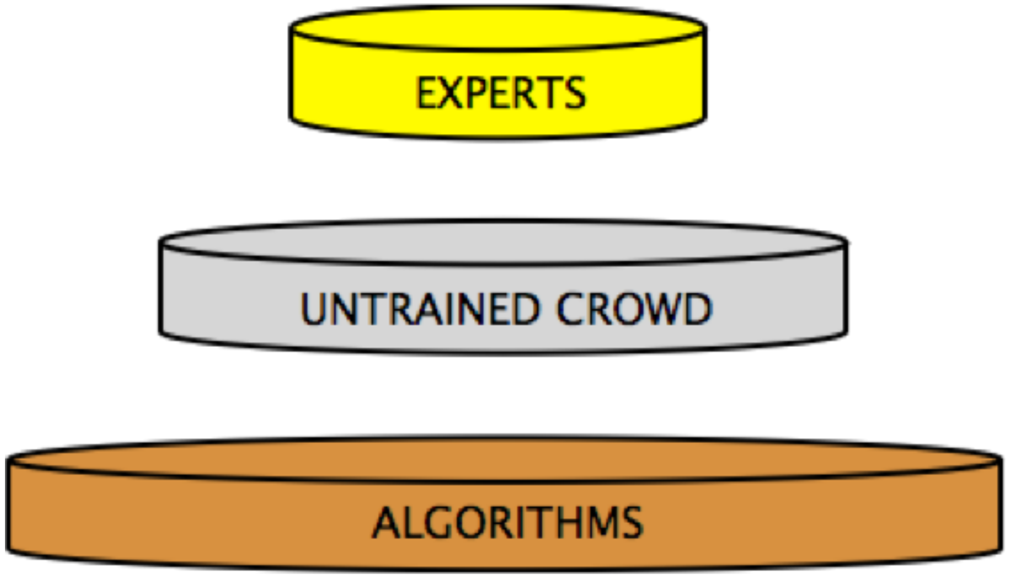
- Dataset Creation
- Getting More Out of Your Training Data
- Human-Machine Partnerships
- Understanding Machine Learning Algorithms

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

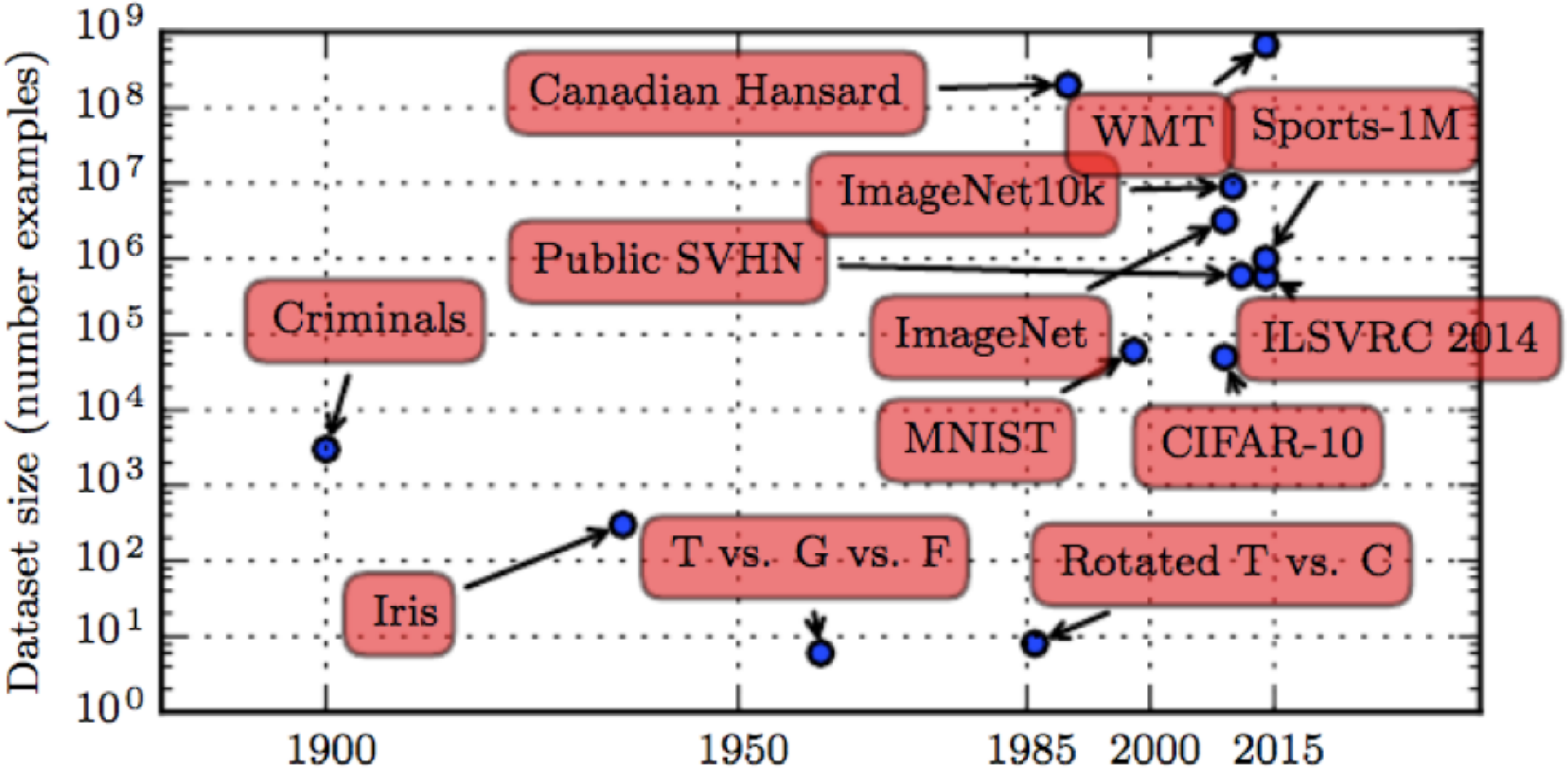


amazon mechanical turk™
Artificial Artificial Intelligence

Group Discussion

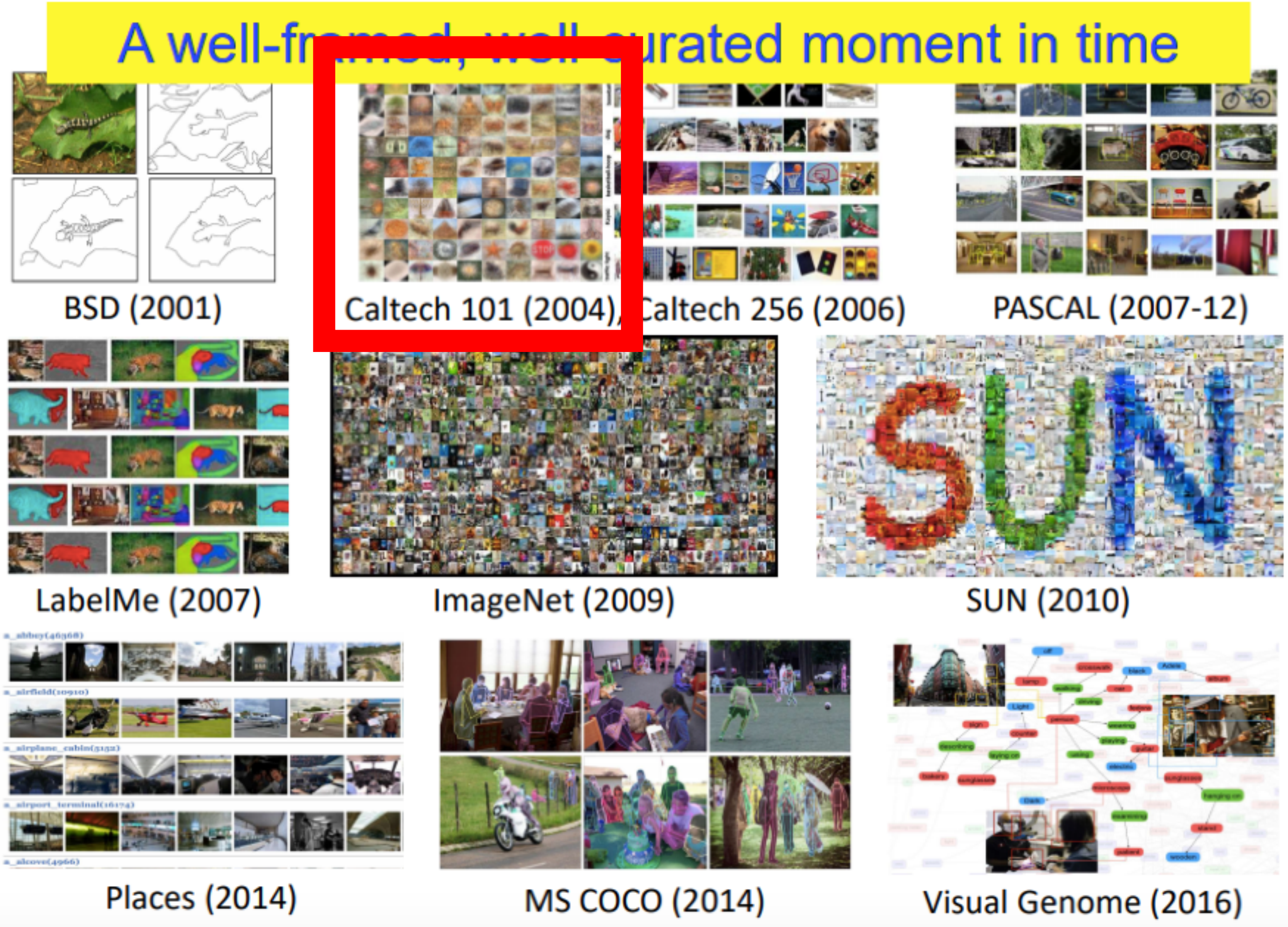
- Toy Problem: predict if a comment on a social network is toxic or not
- Create a budget and process for creating a labeled dataset in-house
- Create a budget and process for creating a labeled dataset with anonymous online crowd workers

Historical Overview of Datasets



Deep Learning, Ian Goodfellow, Yoshua Bengio, and Aaron Courville

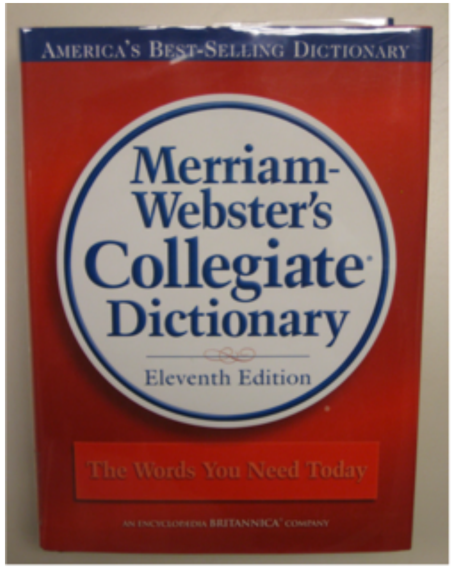
Historical Overview of Datasets



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

→

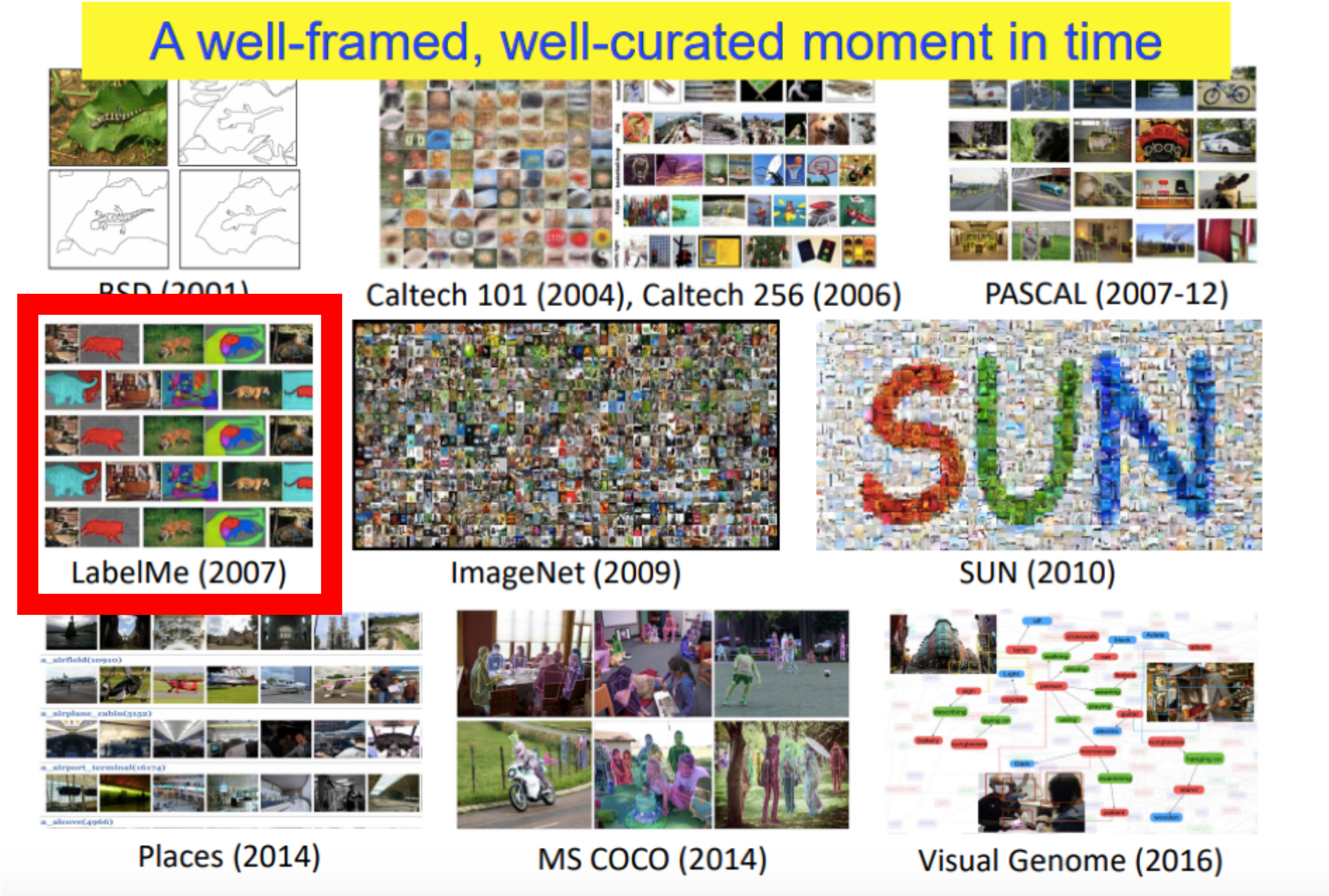
3. Image Verification

2 graduate students reviewed all images and discarded irrelevant images

 # Images: 9,144

Li Fei Fei, Rob Fergus, & Pietro Perona, CVPR 2004 & CVIU 2007; *1,060 citations in 1/17*

Historical Overview of Datasets



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

LabelMe

1. Image Collection

- Most images taken by authors using a variety of hand-held digital cameras.

- Also includes video sequences taken with a head-mounted web camera.

2. Object Segmentation

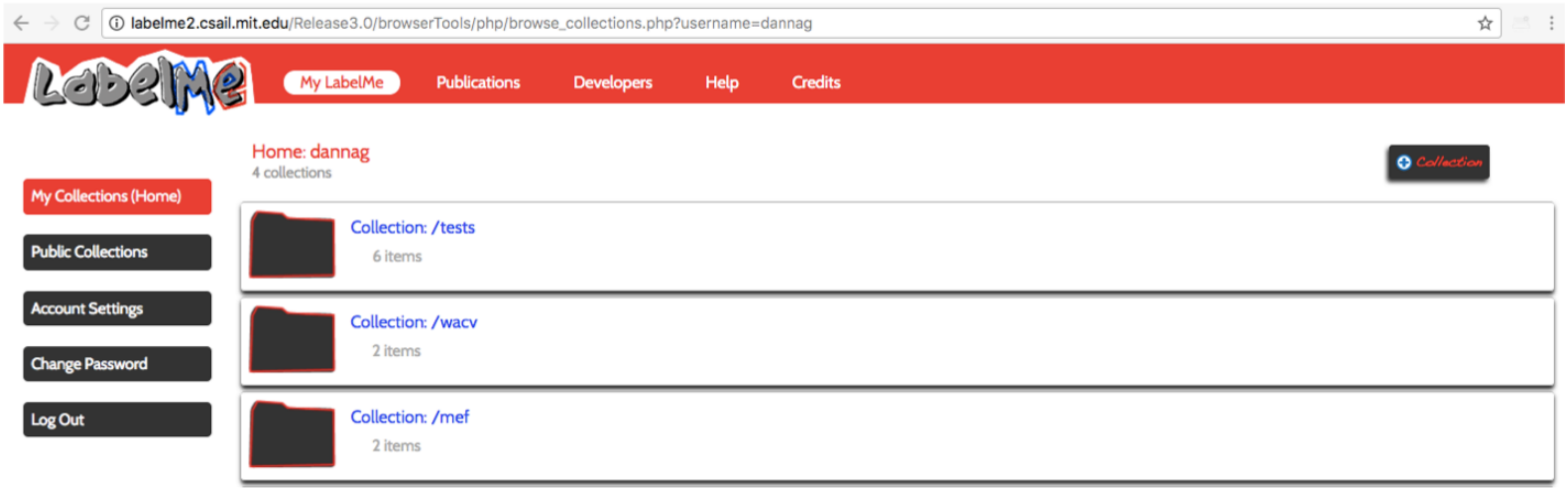
- Any user that logs in uses online tool

3. Category Assignment

- User that segments the object also labels the object

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

LabelMe: Online Database and Tool



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

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

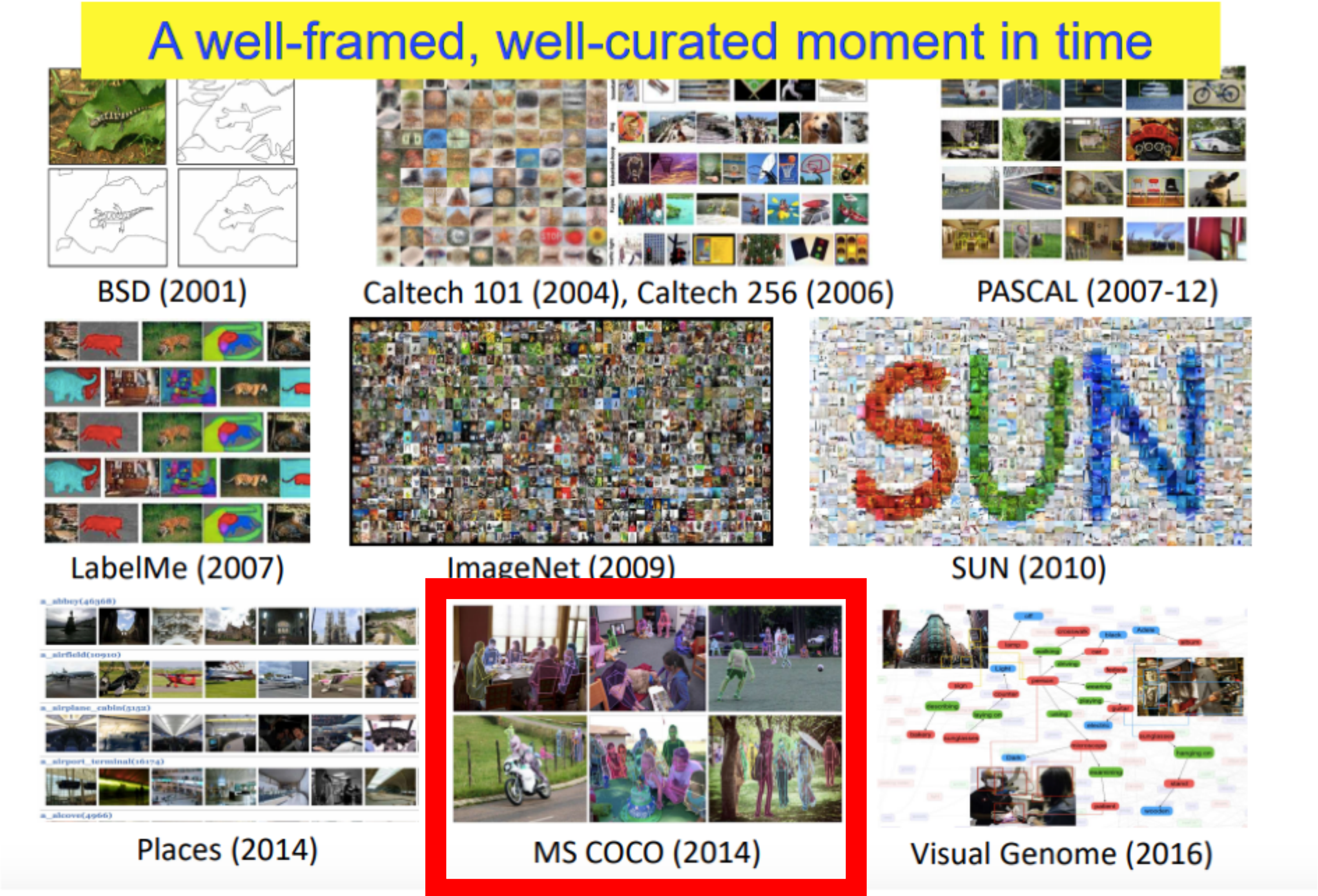
"she has labeled more than 250,000 objects"

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*

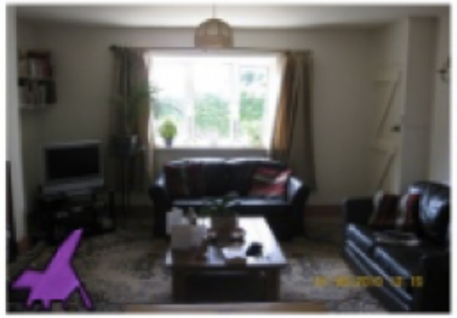
Historical Overview of Datasets



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,

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

Rationale: primary interest is in precise localization of object instances

Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, Piotr Dollár; ECCV 2014; *809 citations in 2/17*

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

Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, Piotr Dolla ́r; ECCV 2014; *809 citations in 2/17*

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	hyppo	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
jetpack	robots								

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

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

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

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:

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

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:

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

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

→

3. Object presence labeling

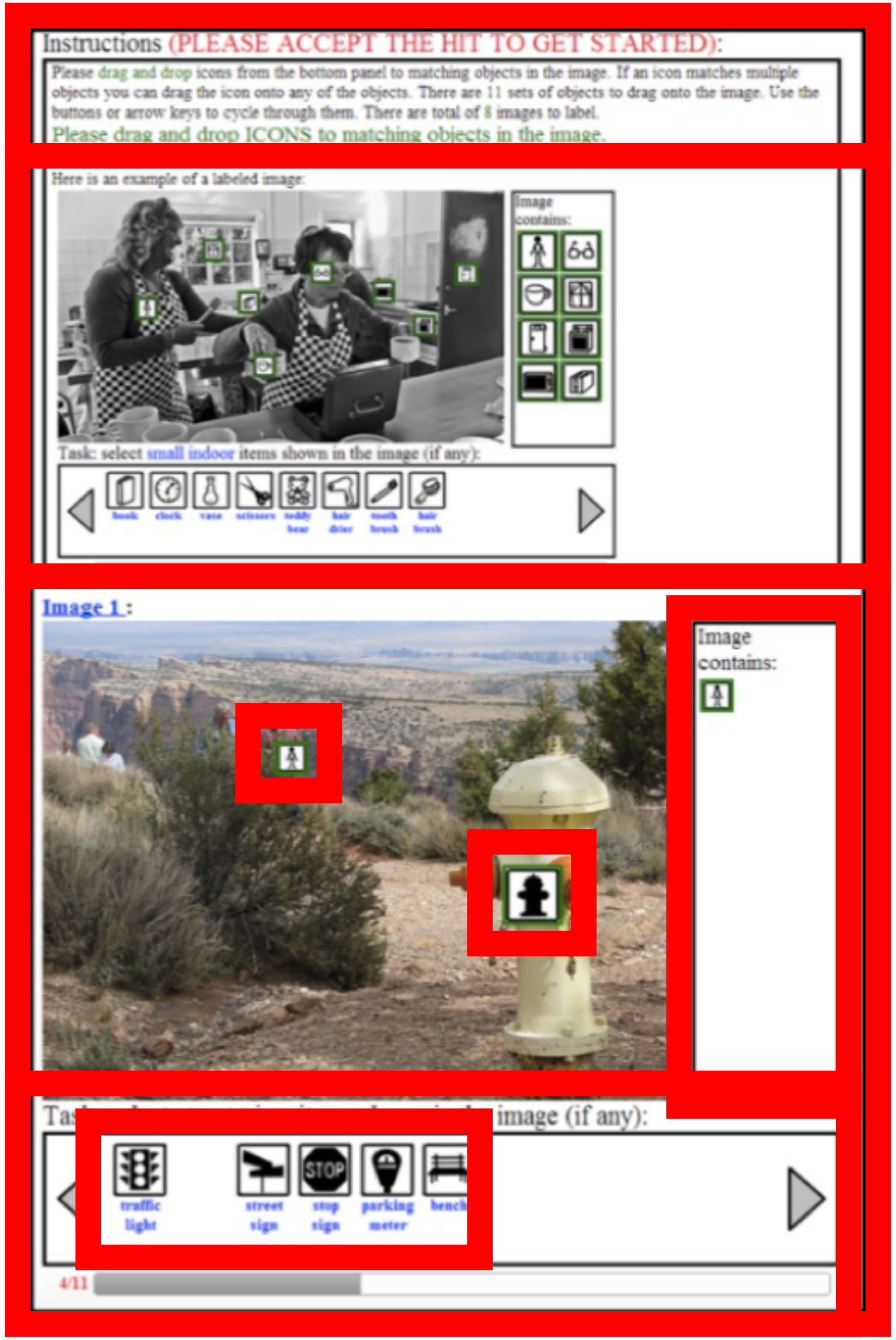
- Assign all relevant categories to each image by locating one instance

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 – object presence

11 Groupings

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



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

3. Object presence labels

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

4. Instance Spotting

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

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

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

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

3. Object presence labels
   - Assign all relevant categories to each image by locating one instance

4. Instance Spotting

5. Object segmentation
   - Adaptation of OpenSurfaces user interface

~20,000 worker hours

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

Training task per object category required.

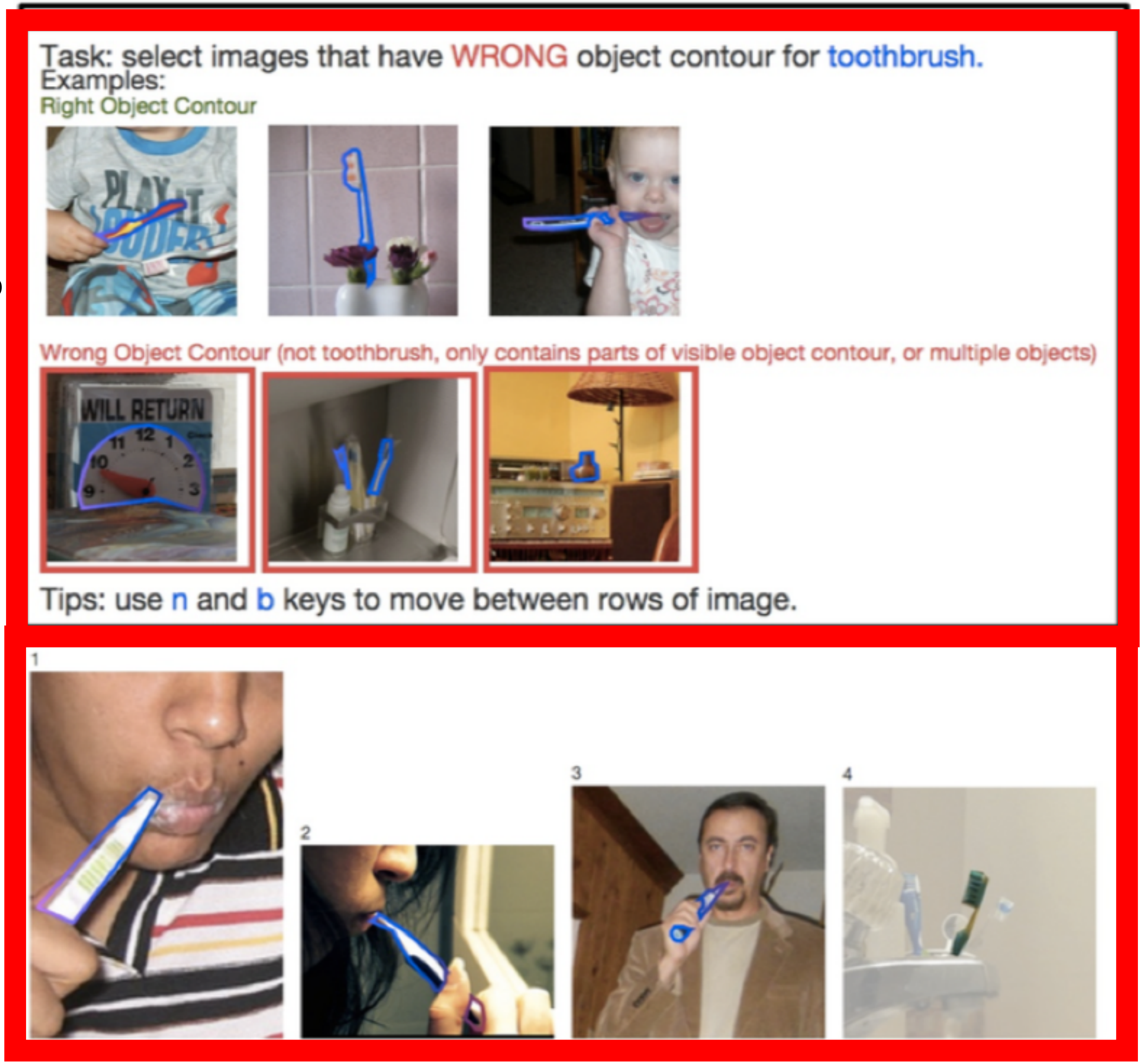
MSCOCO

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.

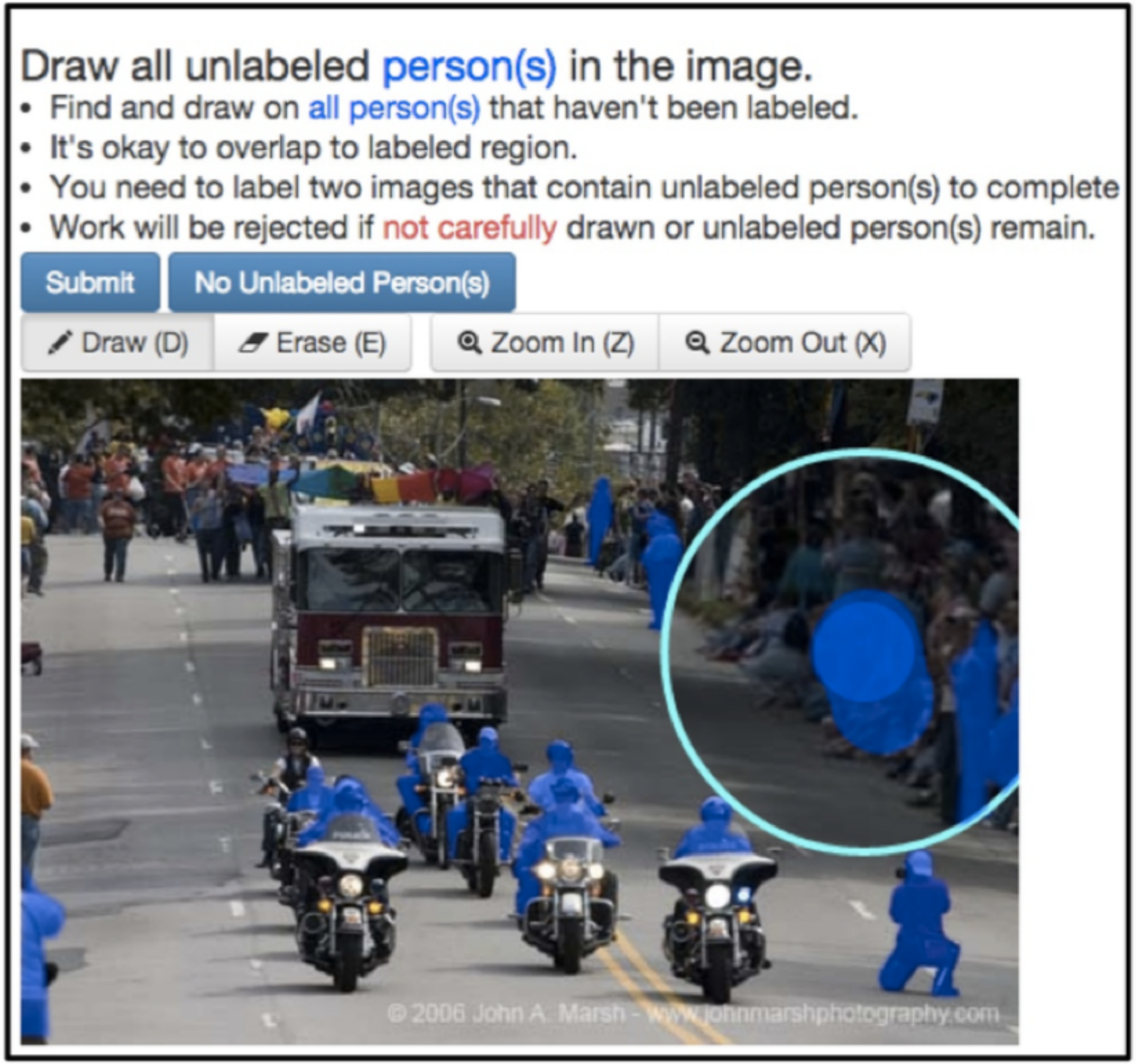
Blocked workers: those who often produced poor segmentations were blocked and their work not used

64 examples



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.



For More, You Can Take….

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

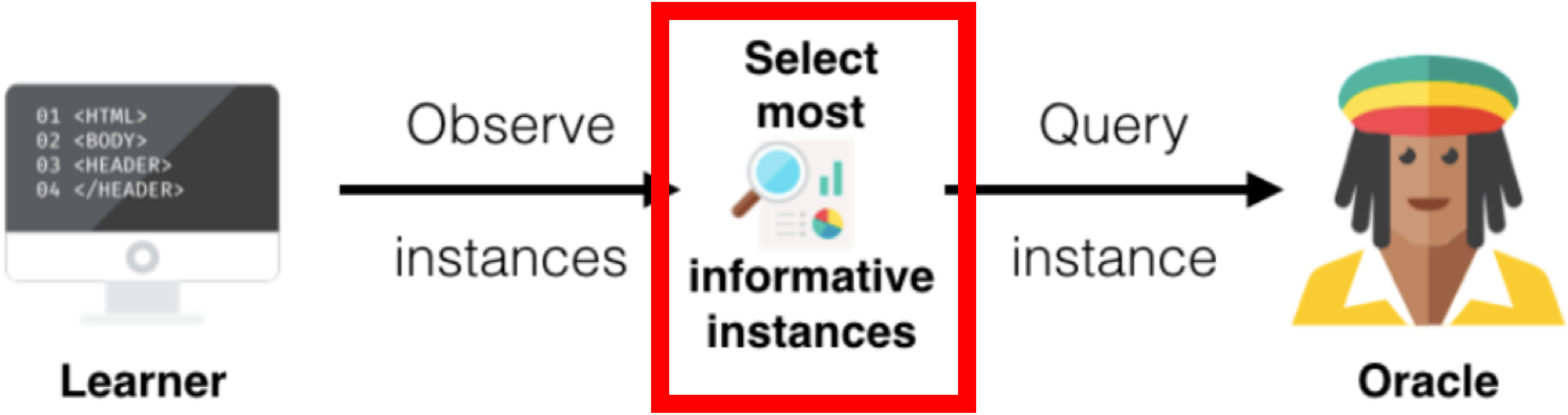
Today’s Topics

- 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



https://www.datacamp.com/community/tutorials/active-learning

Curriculum Learning

Decide order in which you provide data.

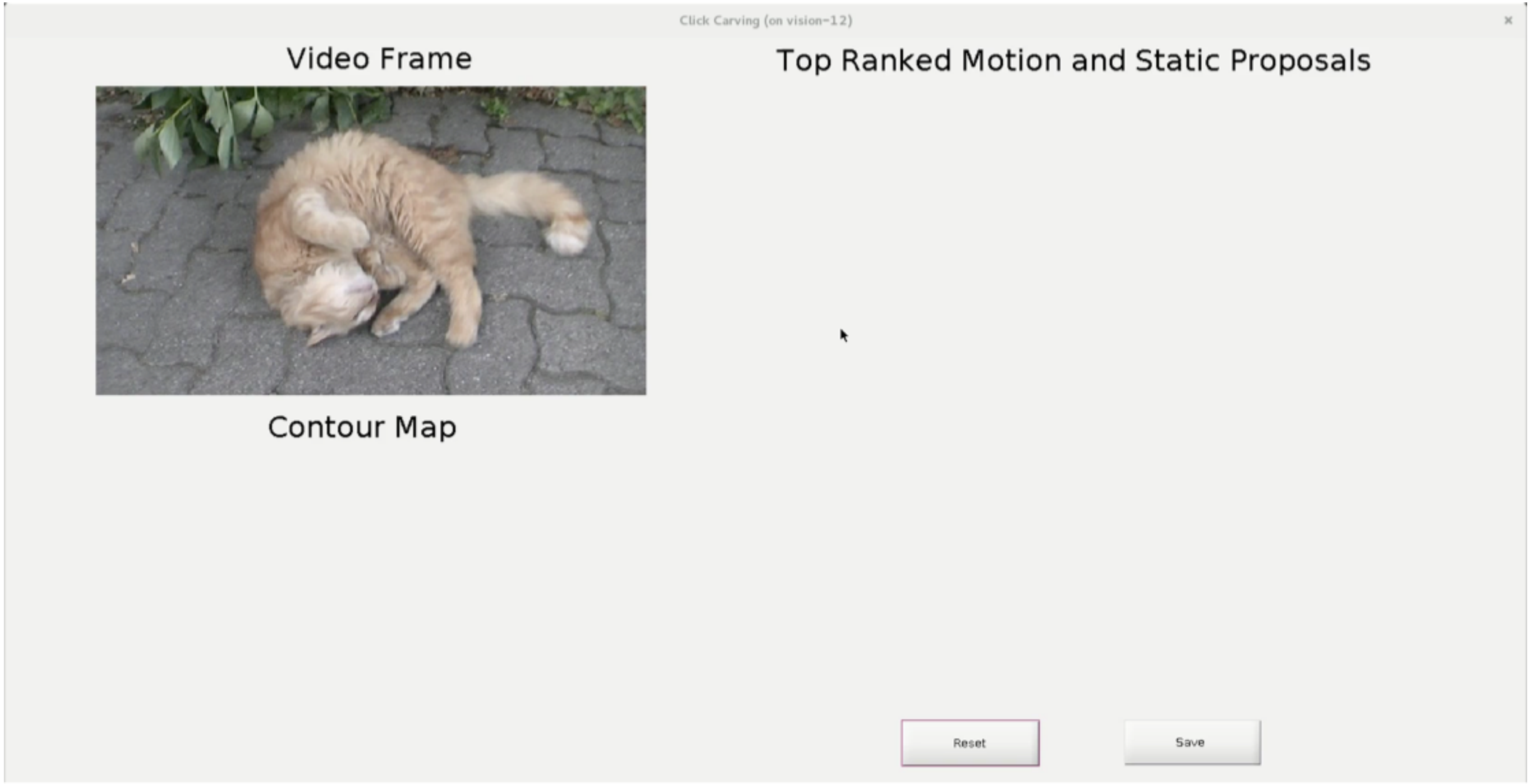
Today’s Topics

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

Interactive Vision Applications

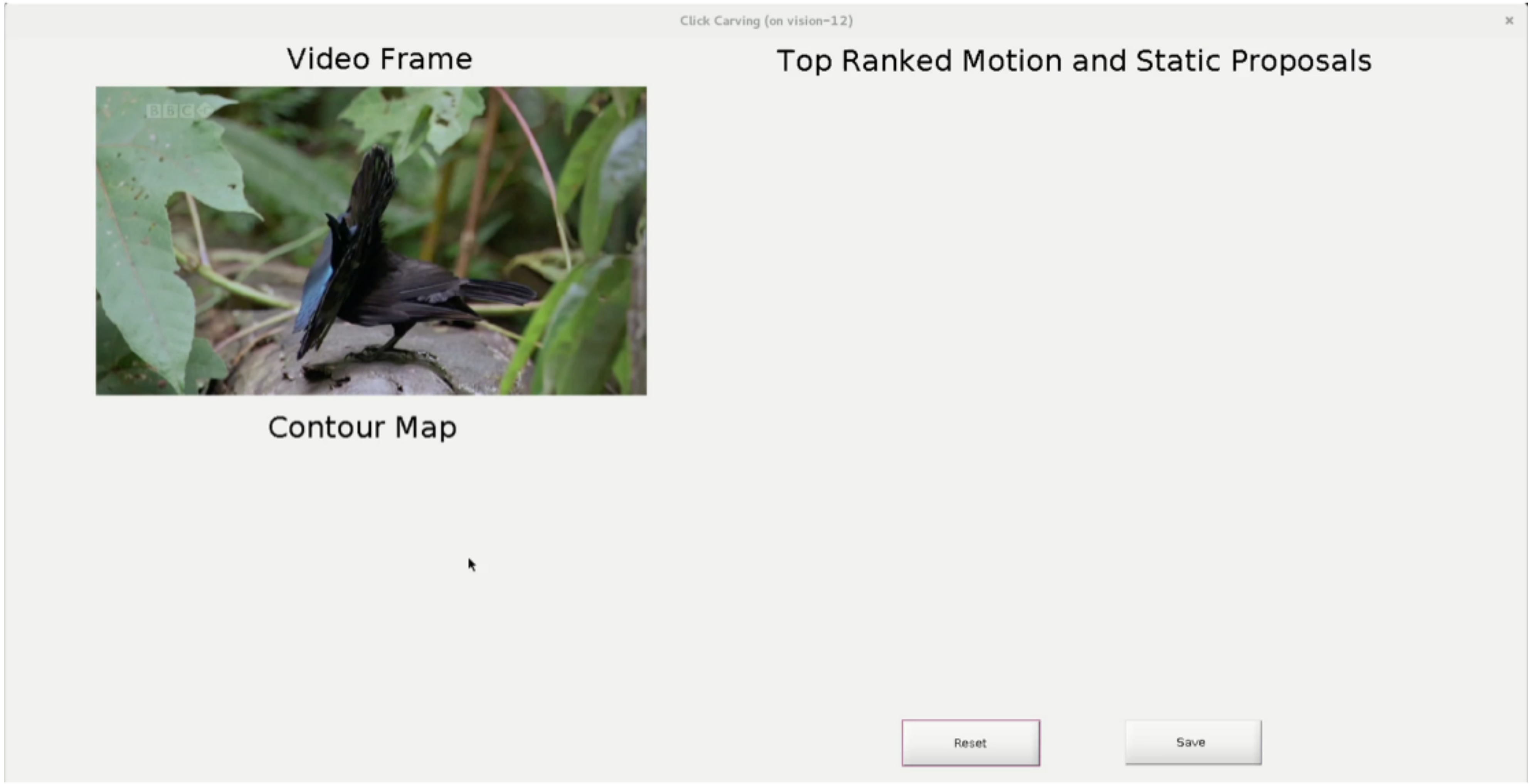
http://ibird.com/

Object Segmentation: Click Carving



Suyog Jain and Kristen Grauman; HCOMP 2016

Object Segmentation: Click Carving

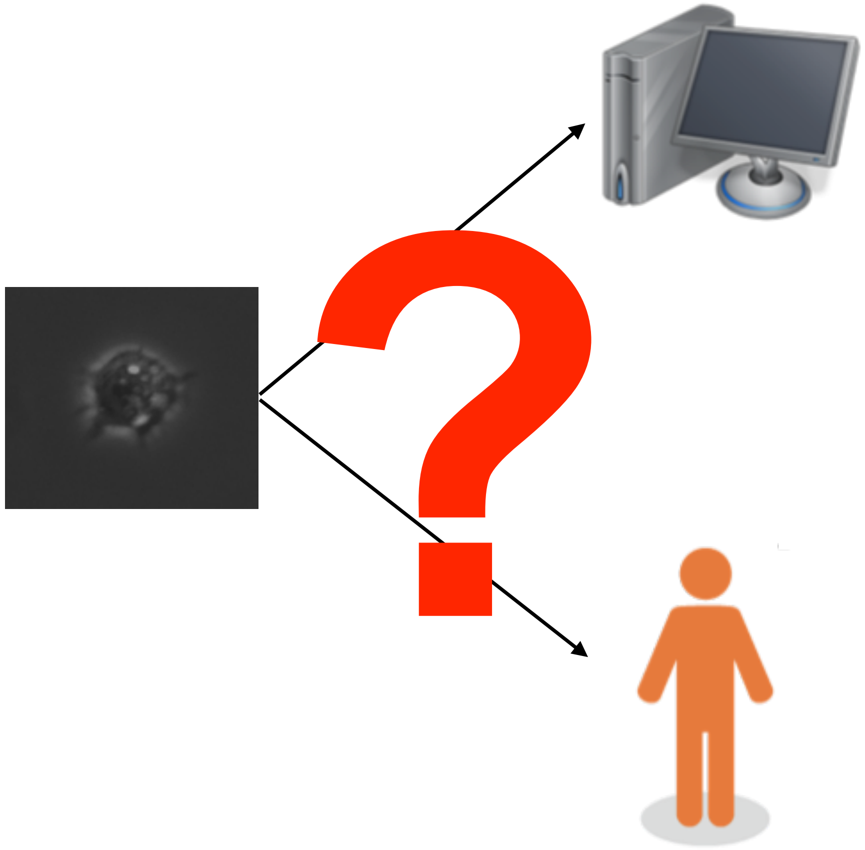


Suyog Jain and Kristen Grauman; HCOMP 2016

Why Interactive Vision?

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

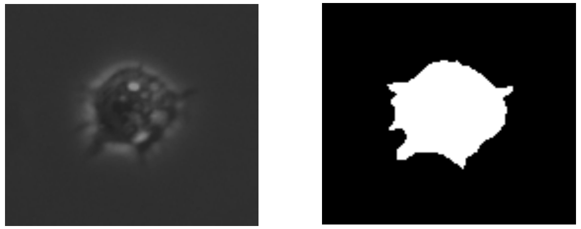
Divide Labor Between Humans & Machines



[Gurari et al; CVPR 2016]

Predict Quality of Algorithm Results

Ground Truth

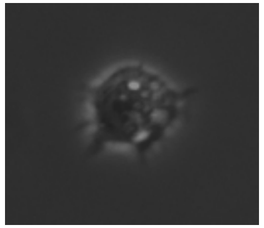


Algorithm-Drawn Segmentations

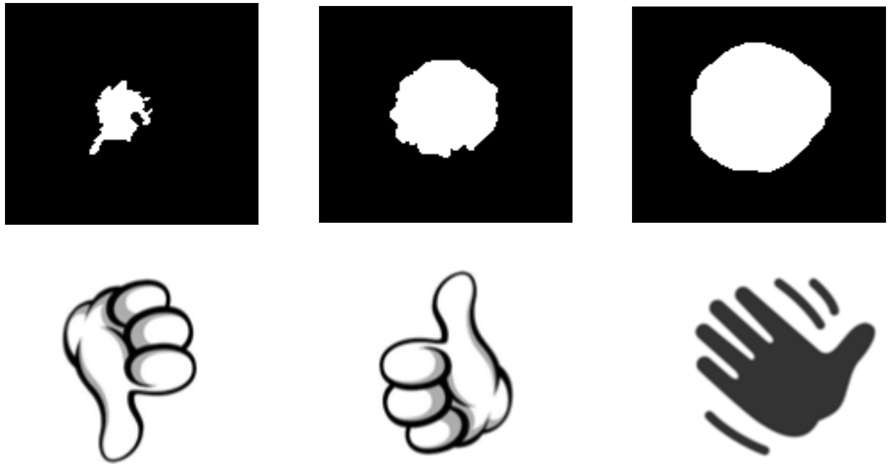
[Gurari et al; CVPR 2016]

Predict Quality of Algorithm Results

Ground Truth

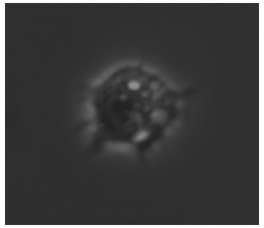


Algorithm-Drawn Segmentations

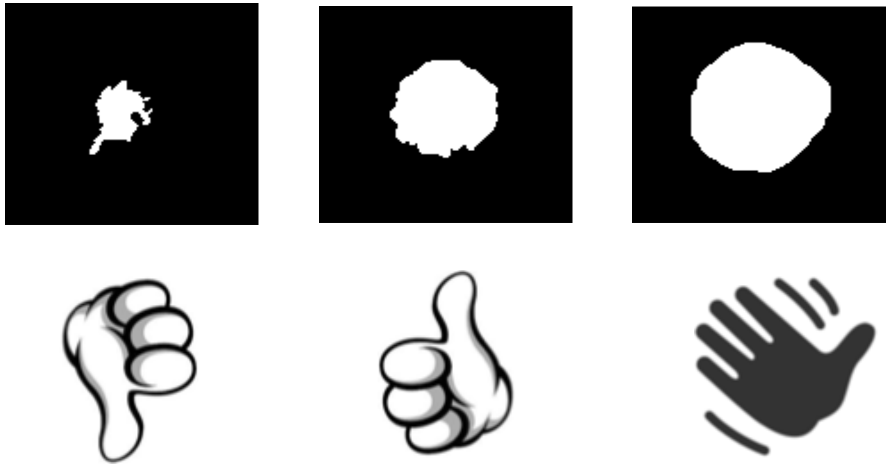


[Gurari et al; CVPR 2016]

Predict Quality of Algorithm Results

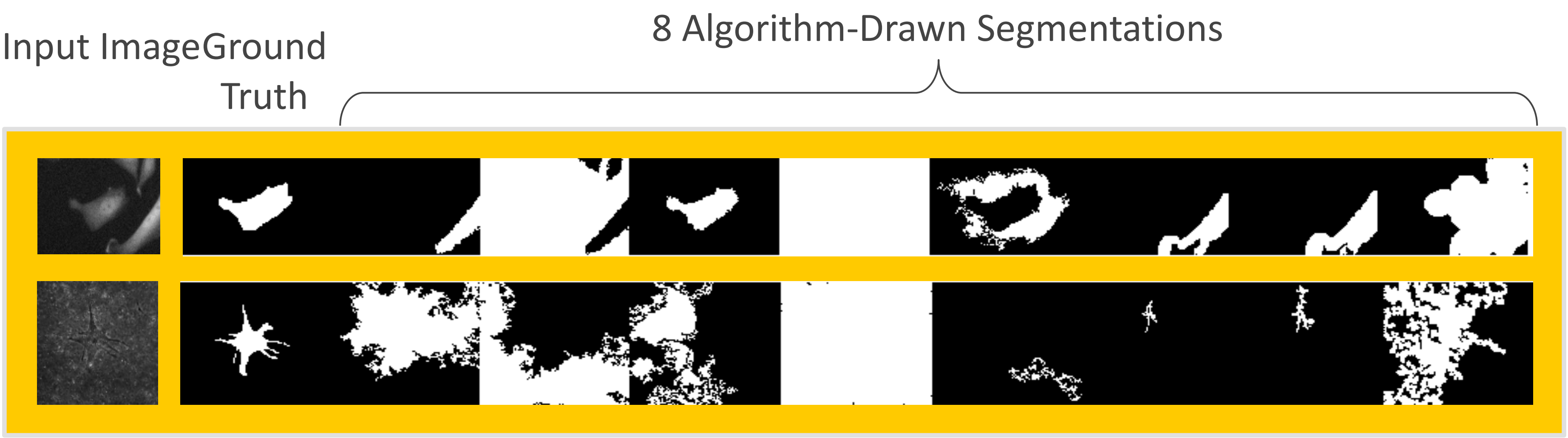


Algorithm-Drawn Segmentations



[Gurari et al; CVPR 2016]

What Features Show Algorithms Fail?

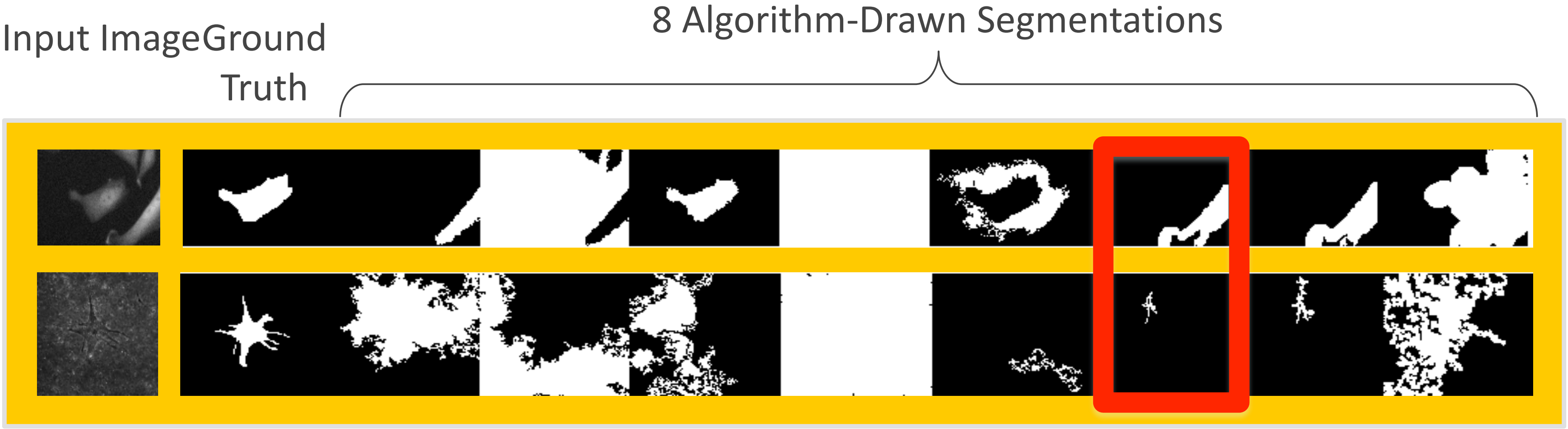


[Gurari et al; CVPR 2016]

What Features Show Algorithms Fail?

Predictive Cues

Object Location - Edges



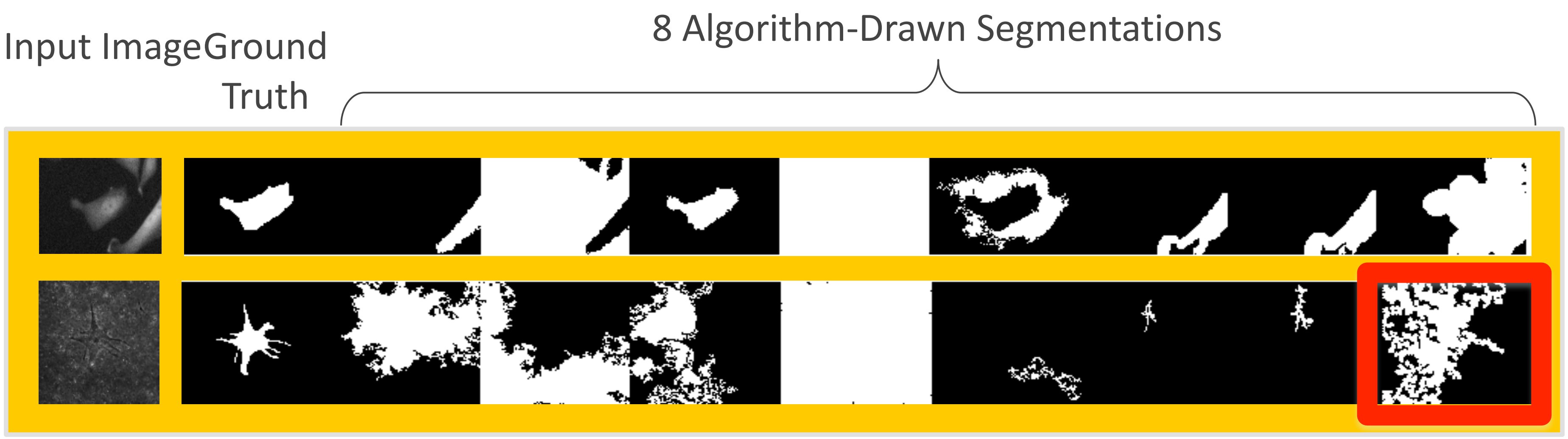
[Gurari et al; CVPR 2016]

What Features Show Algorithms Fail?

Predictive Cues

Object Location - Edges

Object Boundary - Very Jagged



[Gurari et al; CVPR 2016]

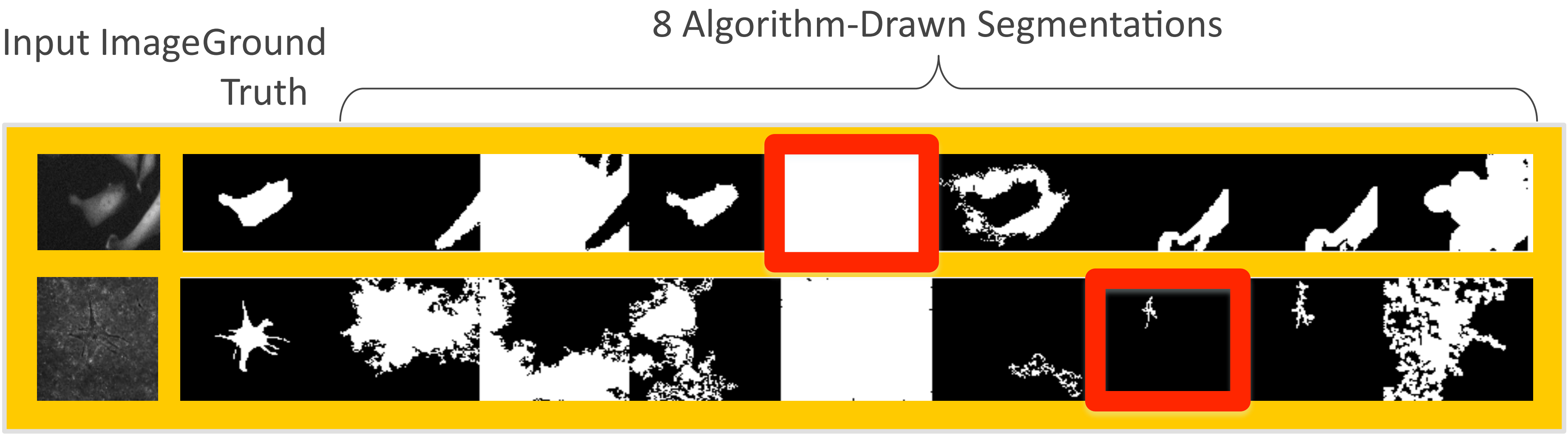
What Features Show Algorithms Fail?

Predictive Cues

Object Location - Edges

Object Boundary - Very Jagged

Image Coverage - Abnormally Large & Small



[Gurari et al; CVPR 2016]

System Training

Training Instances

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

[Gurari et al; CVPR 2016]

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	*F1*	*F2*	● ● ●	*F9*
Seg 1:	0.9	0.7	0.2	● ● ●	0.81
⋮	⋮				
Seg 5742:	0.13	0.5	0.2	● ● ●	0.3

[Gurari et al; CVPR 2016]

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

$$\begin{array}{r|c|ccccc|} & \textit{Score} & \textit{F1} & \textit{F2} & & \cdots & \textit{F9} \\ \hline \textit{Seg 1:} & 0.9 & 0.7 & 0.2 & & \cdots & 0.81 \\ \vdots & \vdots & & & & & \\ \textit{Seg 5742:} & 0.13 & 0.5 & 0.2 & & \cdots & 0.3 \\ \hline \end{array}$$

$$\mathrm{y} = X\underbrace{\boldsymbol{\beta}}_{\text{Model}} + \underbrace{\boldsymbol{\alpha}}_{\text{Bias Term}}$$

Training Model

- Linear Regression

[Gurari et al; CVPR 2016]

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	*F1*	*F2*	●●●	*F9*
Seg 1:	0.9	0.7	0.2	●●●	0.81
⋮	⋮				
Seg 5742:	0.13	0.5	0.2	●●●	0.3

$$y = X\boxed{\beta} + \alpha$$

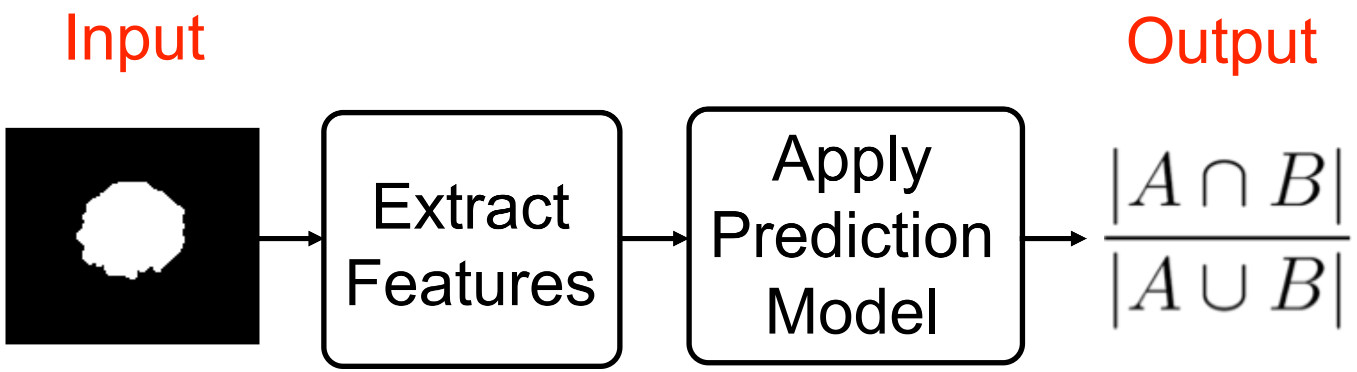
Model Bias Term

Training Model

- Linear Regression

[Gurari et al; CVPR 2016]

System Use for Novel Image



[Gurari et al; CVPR 2016]

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

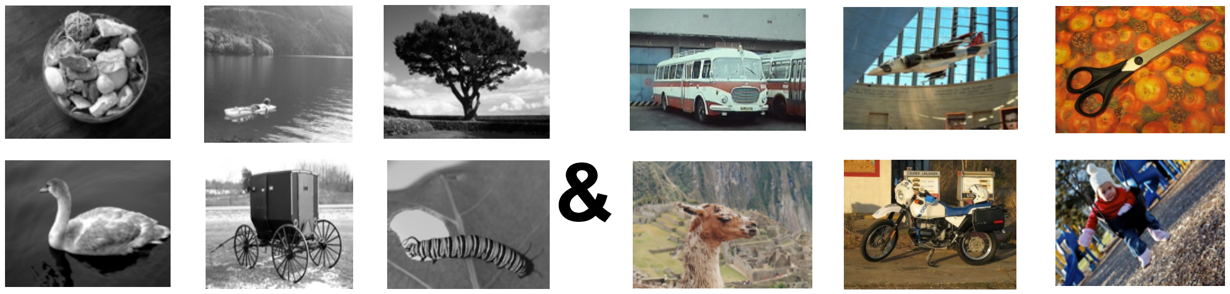
[Gurari et al; CVPR 2016]

Evaluation

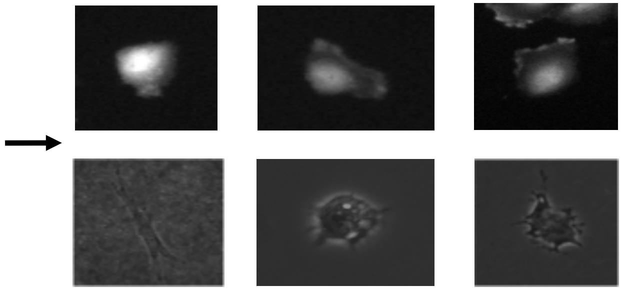
	Weizmann 1,100 Segs		IIS 1,661 Segs		BU-BIL 2,981 Segs	
	CC	**MAE**	**CC**	**MAE**	**CC**	**MAE**
Ours:	**0.64**	**0.24**	**0.68**	**0.22**	**0.61**	**0.31**
CPMC [CVPR 2010]:	0.61	0.32	0.67	0.31	0.36	0.33
AlexNet features [NIPS 2012]:	-0.1	26.7	-0.01	45	-0.01	3.22

CC: Correlation Coefficient, MAE: Mean Absolute Error

Training Examples

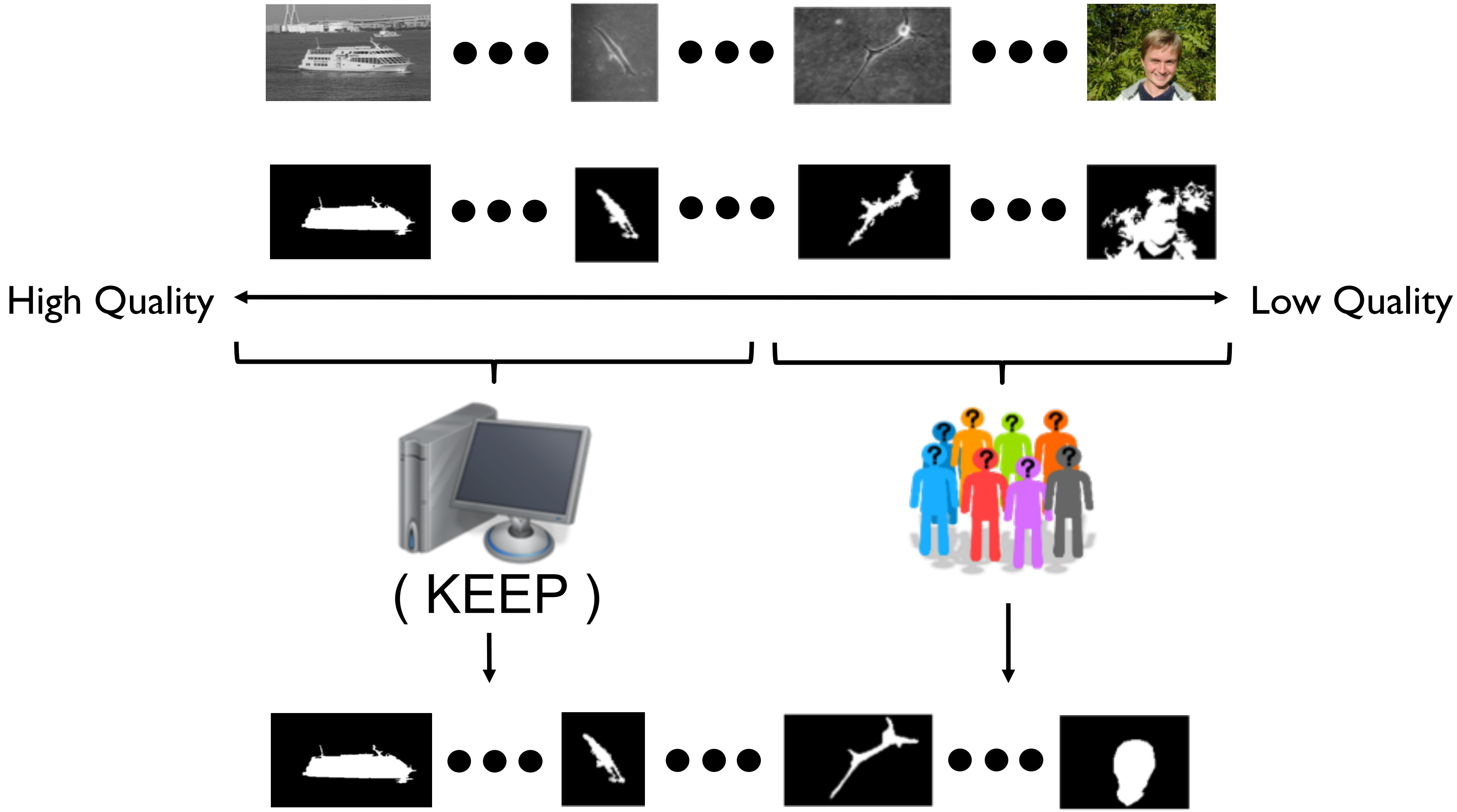


Testing Examples



[Gurari et al; CVPR 2016]

Predict When to Use Human vs Machine



[Gurari et al; CVPR 2016]

Related Work

Predicting Failures of Vision Systems

Peng Zhang[1] Jiuling Wang[2] Ali Farhadi[3] Martial Hebert[4] Devi Parikh[1]
[1]Virginia Tech [2]Univ. of Texas at Austin [3]Univ. of Washington [4]Carnegie Mellon University
[1]{zhangp, parikh}@vt.edu [2]jiuling@utexas.edu [3]ali@cs.uw.edu [4]hebert@ri.cmu.edu

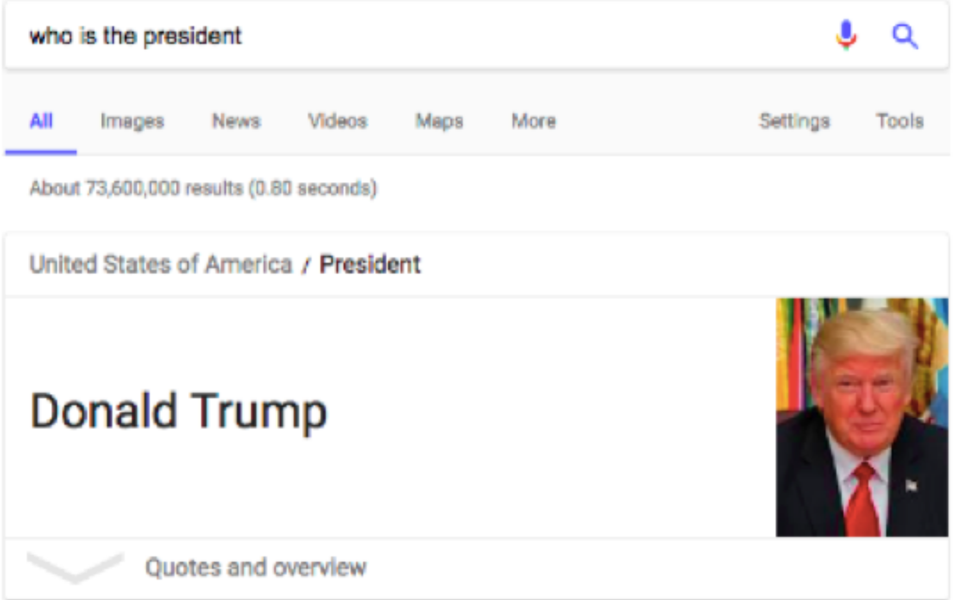
Today's Topics

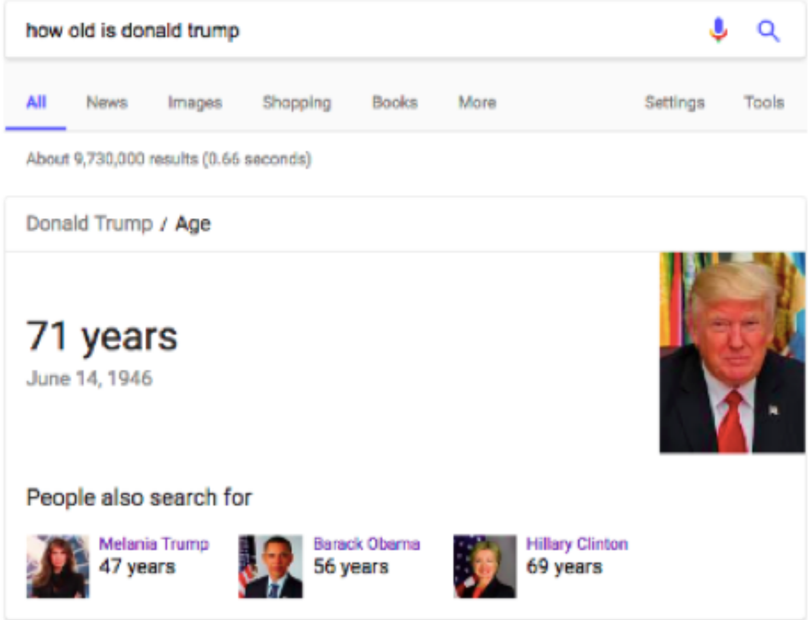
- Dataset Creation
- Getting More Out of Your Training Data
- Human-Machine Partnerships
- Understanding Machine Learning Algorithms

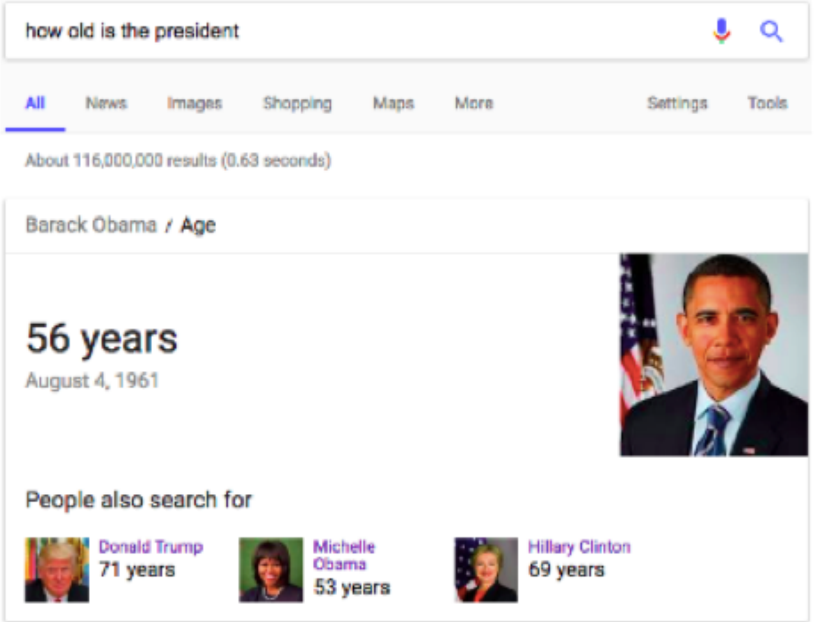
Interpretability

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

Try to Break Them



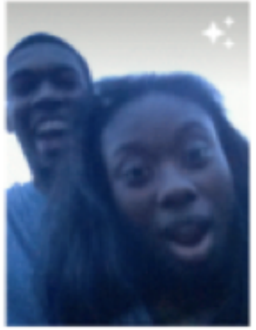




Object Recognition Applications Gone Wrong

- 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
- Ethical Mistake: Photo Tagging
  - http://www.usatoday.com/story/tech/2015/07/01/google-apologizes-after-photos-identify-black-
- Security Mistake: Person Recognition
  - https://www.cs.cmu.edu/~sbhagava/papers/face-rec-ccs16.pdf

Why do you think these systems make such mistakes?

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

If you were the CEO, how would you change your product in response?

Group Discussion

- Find ways to break different online AI systems