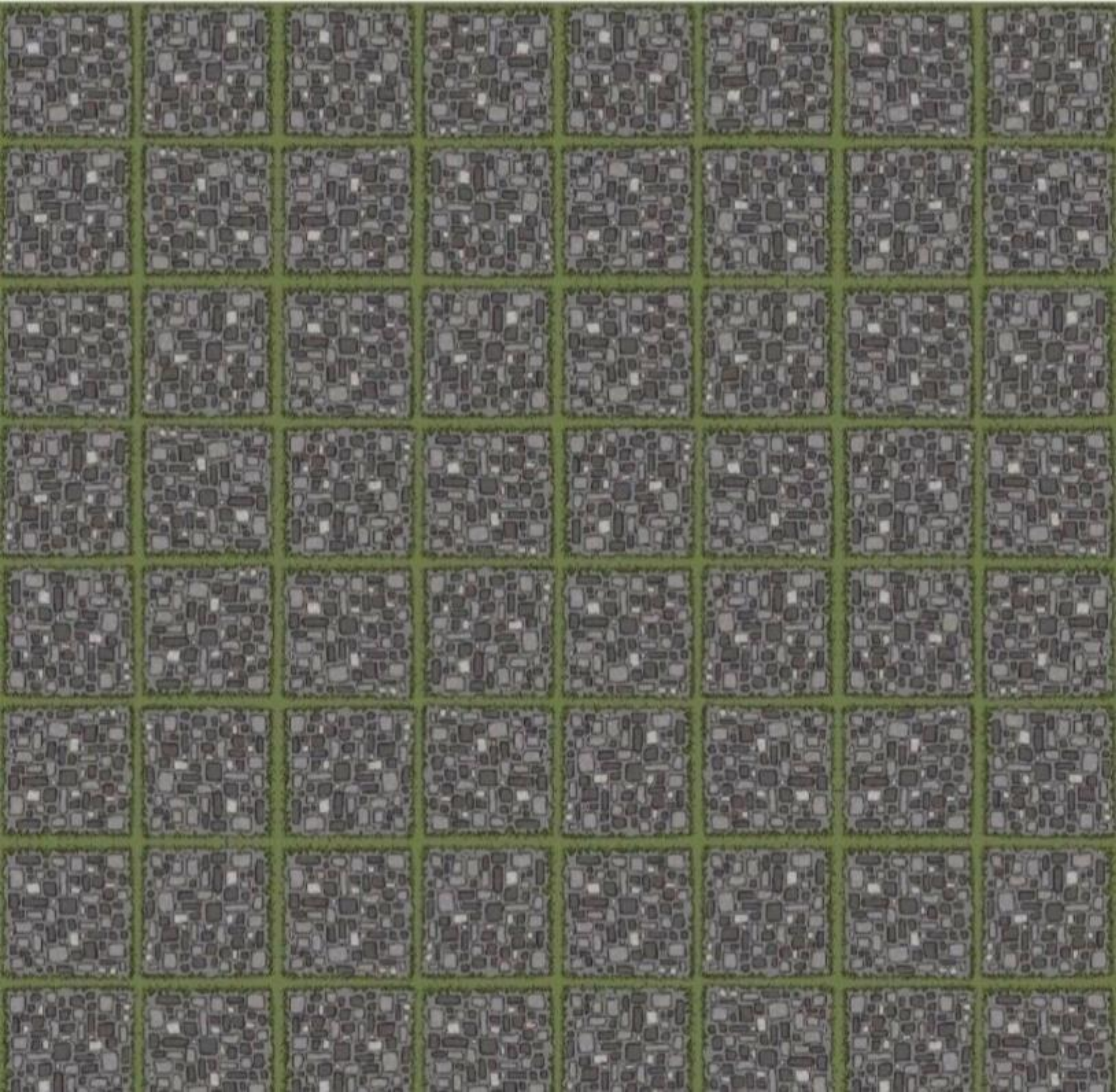
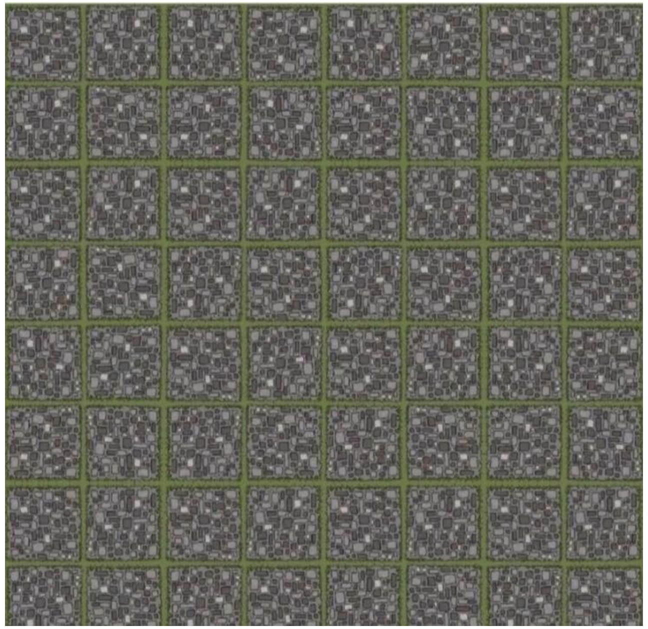
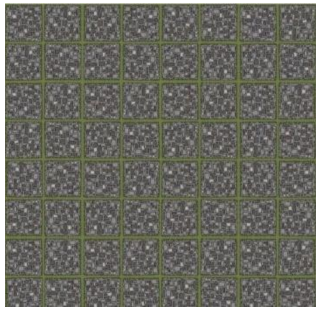
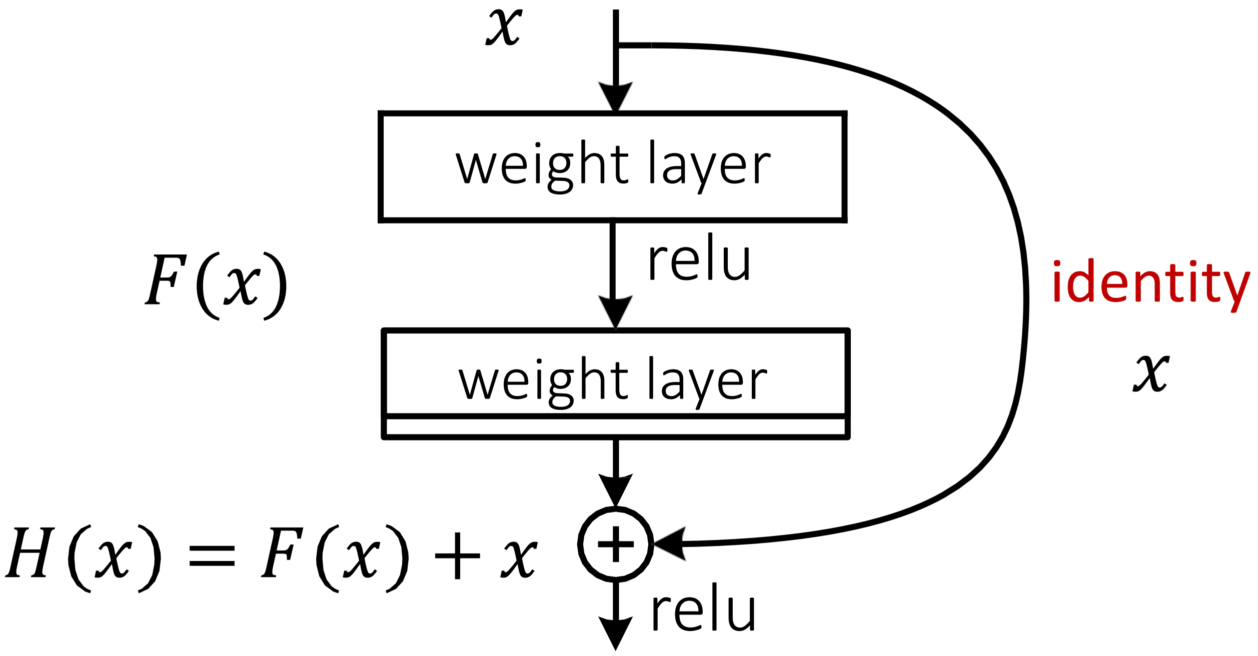
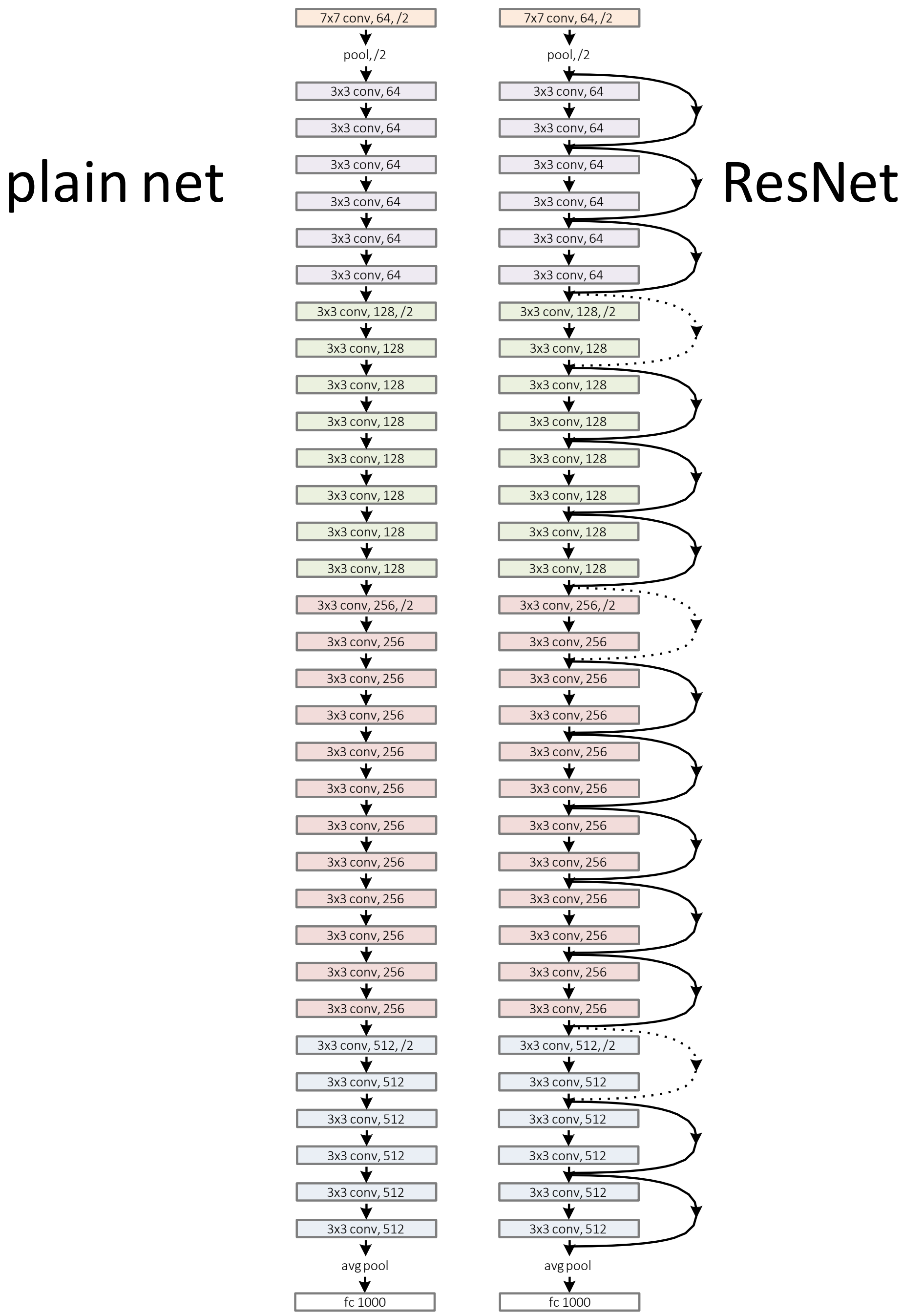




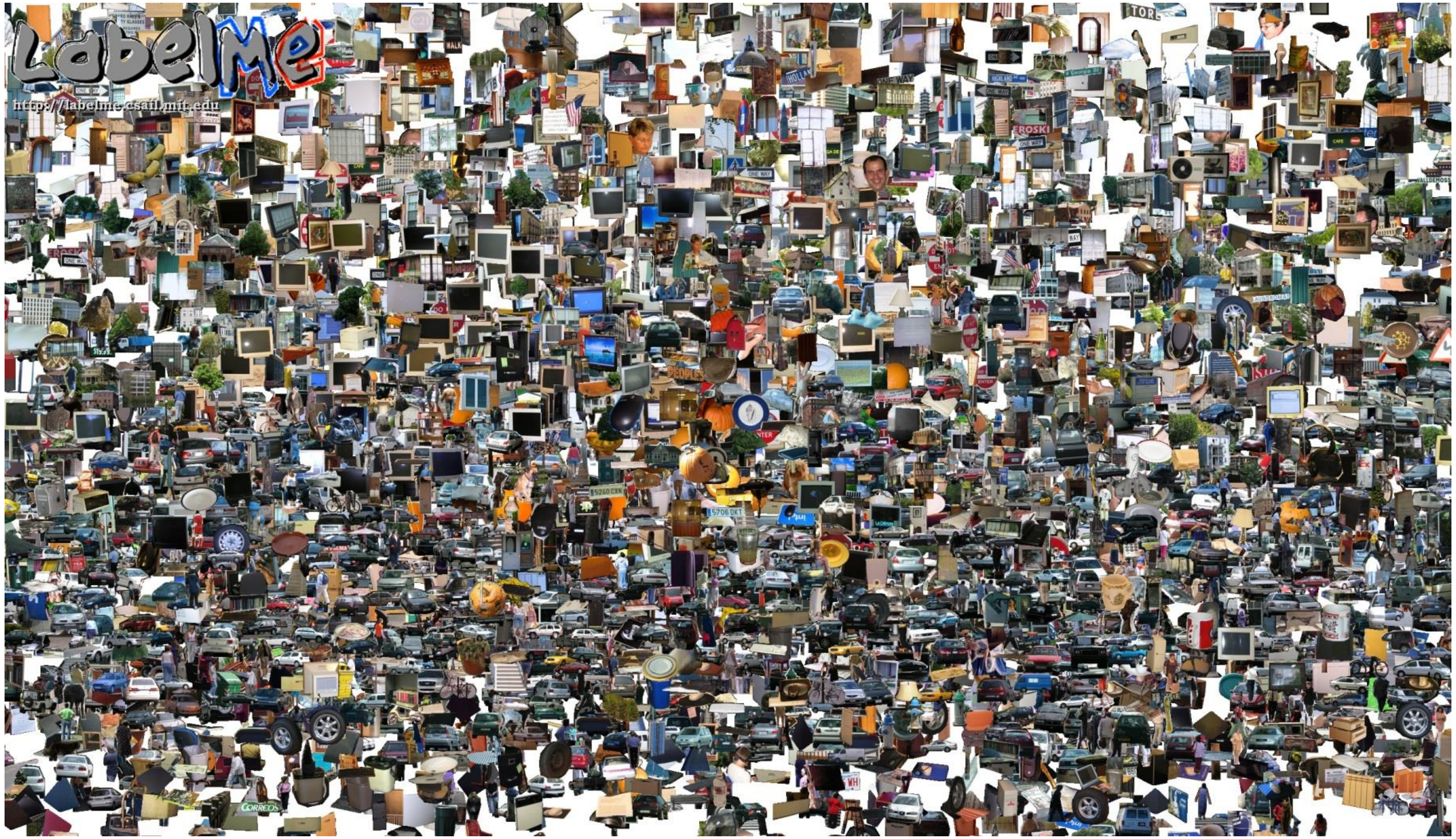
Recap: Resnet

- $F(x)$ is a residual mapping w.r.t. identity





Opportunities of Scale



Computer Vision

James Hays

Many slides from James Hays, Alyosha Efros, and Derek Hoiem

Graphic from Antonio Torralba

Outline

Opportunities of Scale: Data-driven methods

- The Unreasonable Effectiveness of Data
- Scene Completion
- Im2gps
- Recognition via Tiny Images

Computer Vision Class so far

- The geometry of image formation
  – Ancient / Renaissance
- Signal processing / Convolution
  – 1800, but really the 50’s and 60’s
- Hand-designed Features for recognition, either instance-level or categorical
  – 1999 (SIFT), 2003 (Video Google), 2005 (Dalal-Triggs), 2006 (spatial pyramid bag of words)
- Learning from Data
  – 1991 (EigenFaces) but late 90’s to now especially

What has changed in the last 15 years?

- The Internet
- Crowdsourcing
- Learning representations from the data these sources provide (deep learning)

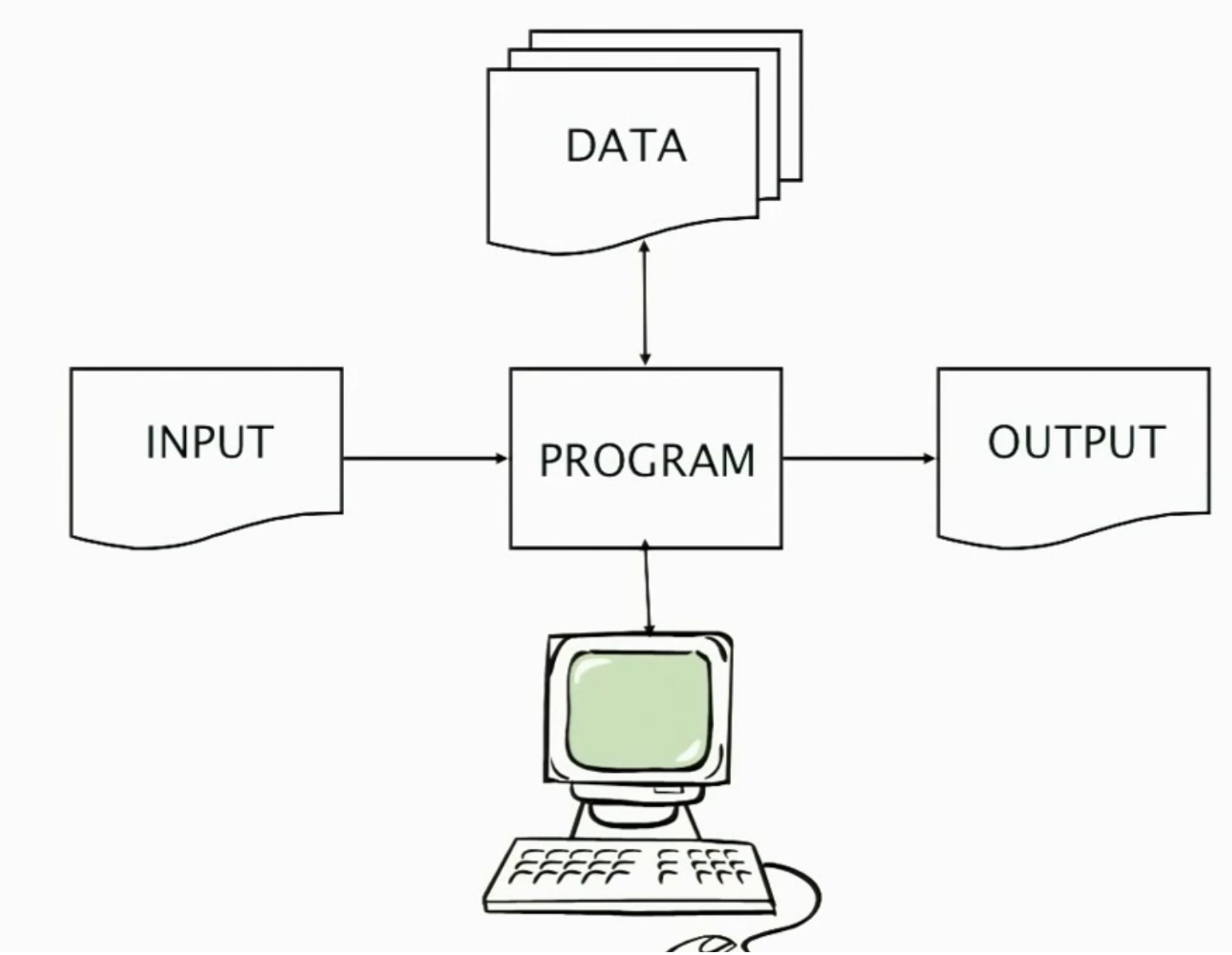
Google and massive data-driven algorithms

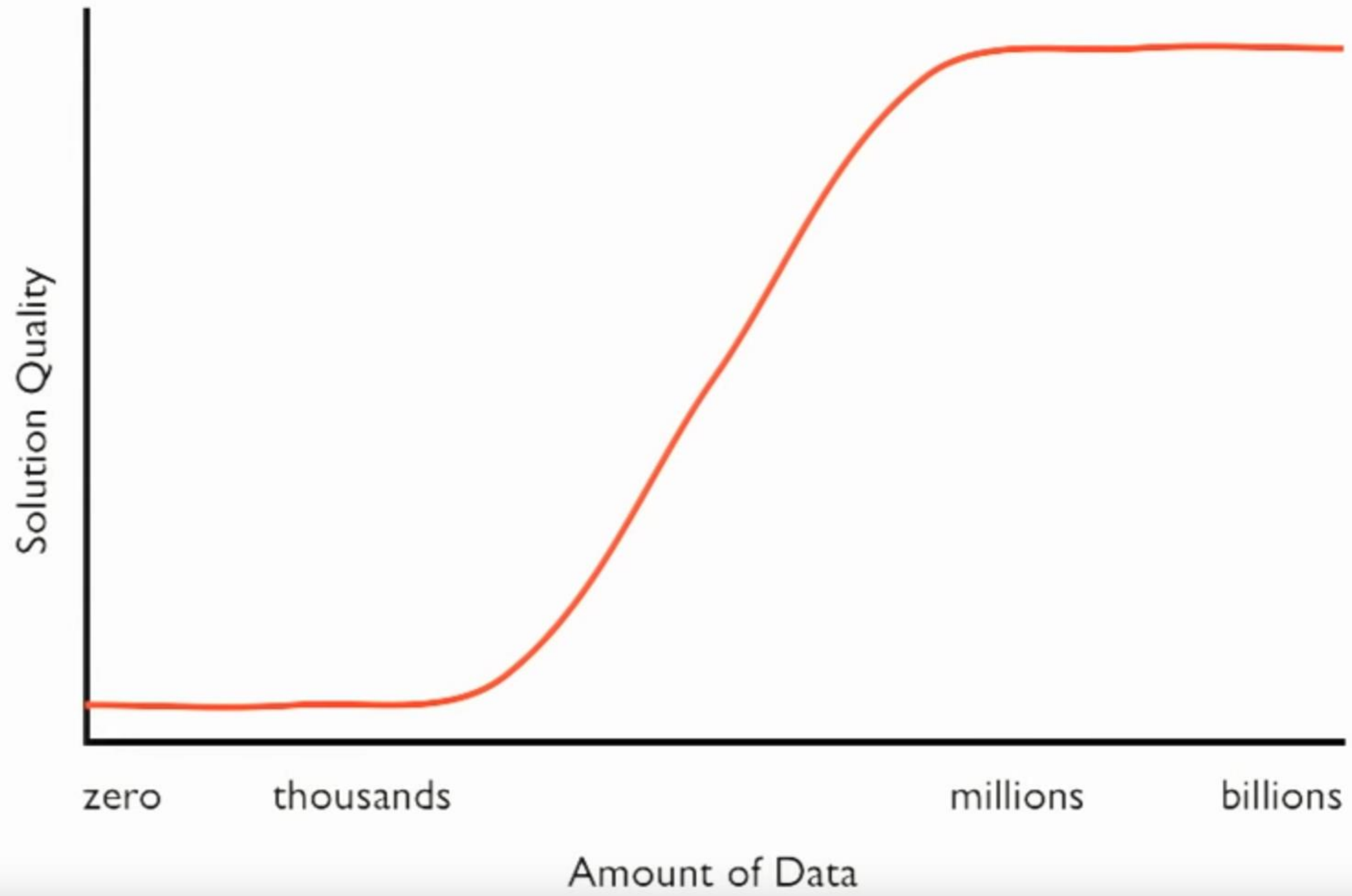
A.I. for the postmodern world:

– all questions have already been answered…many times, in many ways

– Google is dumb, the “intelligence” is <u>in the data</u>

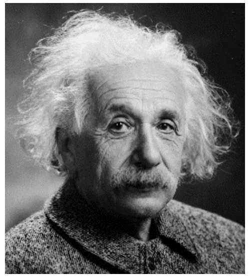
https://youtu.be/yvDCzhbjYWs?t=24
Watch until 9:42


DATA
INPUT
PROGRAM
OUTPUT


Solution Quality
zero
thousands
millions
billions
Amount of Data

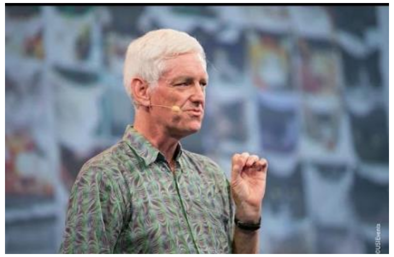
The Unreasonable Effectiveness of Math

- **“**The miracle of the appropriateness of the language of mathematics…” **Eugene Wigner**



- “The most incomprehensible thing about the universe is that it is comprehensible.” **Albert Einstein**

- “There is only one thing which is more unreasonable than the unreasonable effectiveness of mathematics in physics, and this is the unreasonable ineffectiveness of mathematics in biology.” **Israel Gelfand**

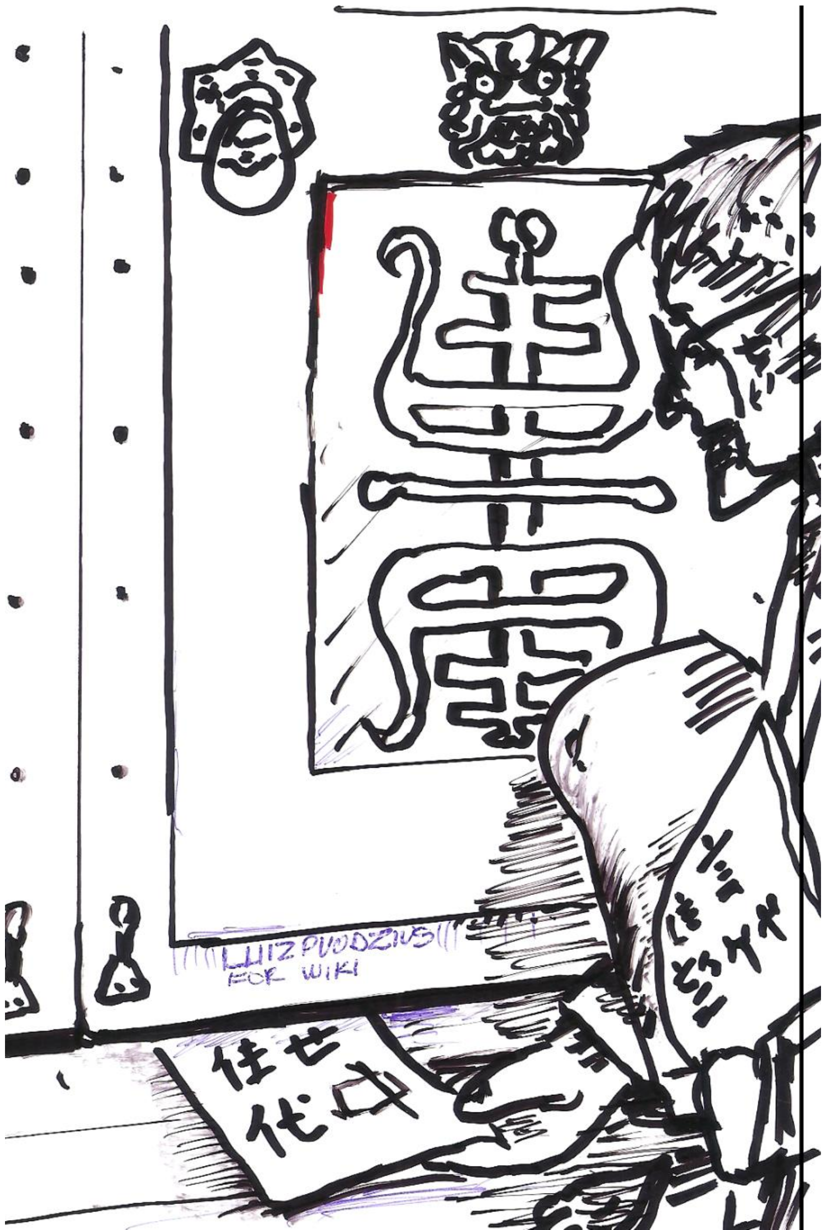


- “We should stop acting as if our goal is to author extremely elegant theories, and instead embrace complexity and make use of the best ally we have: the unreasonable effectiveness of data.” **Peter Norvig**

Chinese Room, John Searle (1980)

If a machine can convincingly simulate an intelligent conversation, does it necessarily understand? In the experiment, Searle imagines himself in a room, acting as a computer by manually executing a program that convincingly simulates the behavior of a native Chinese speaker.

Most of the discussion consists of attempts to refute it. "The overwhelming majority," notes *BBS* editor Stevan Harnad,“ still think that the Chinese Room Argument is dead wrong." The sheer volume of the literature that has grown up around it inspired Pat Hayes to quip that the field of cognitive science ought to be redefined as "the ongoing research program of showing Searle's Chinese Room Argument to be false.



Yann LeCun

October 23 at 9:58pm ·

Questions from the piece:

Q1. Does the Chinese Room argument prove the impossibility of machine consciousness?
A1: Hell no. ... See More

Can Machines Become Moral?

The question is heard more and more often, both from those who think that machines cannot become moral, and who think that to believe otherwise is a dangerous illusion, and from those who think that machines must become moral,...

BIGQUESTIONSONLINE.COM | BY DON HOWARD

You and 156 others   30 Comments 20 Shares

Like   Comment   Share

Big Idea

- Do we need computer vision systems to have strong AI-like reasoning about our world?
- What if invariance / generalization isn't actually the core difficulty of computer vision?
- What if we can perform high level reasoning with brute-force, data-driven algorithms?

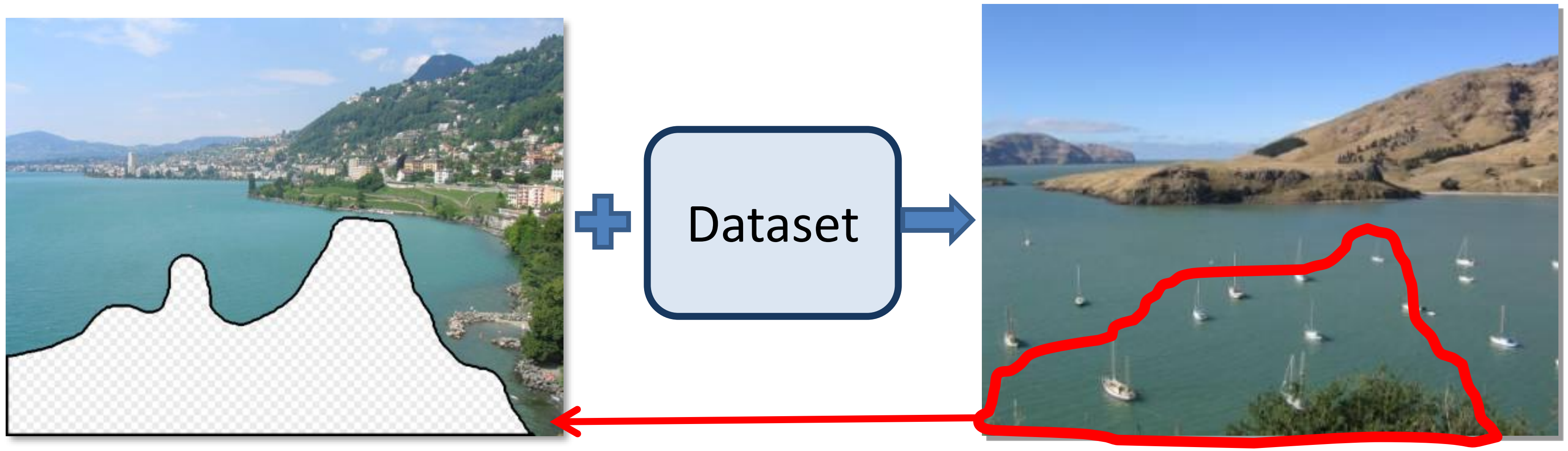
Scene Completion

[Hays and Efros. Scene Completion Using Millions of Photographs. SIGGRAPH 2007 and CACM October 2008.]

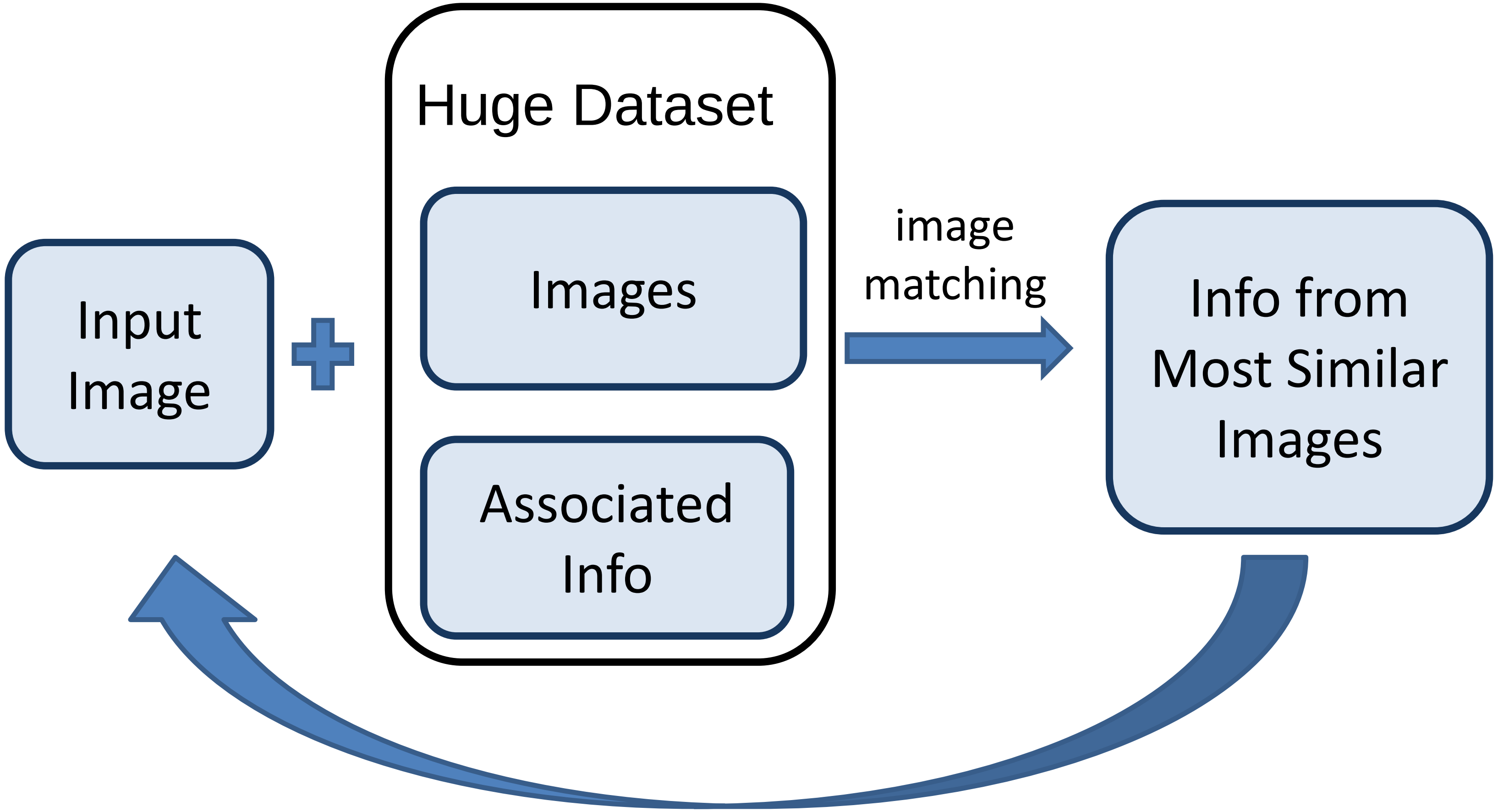
http://graphics.cs.cmu.edu/projects/scene-completion/

How it works

- Find a similar image from a large dataset
- Blend a region from that image into the hole

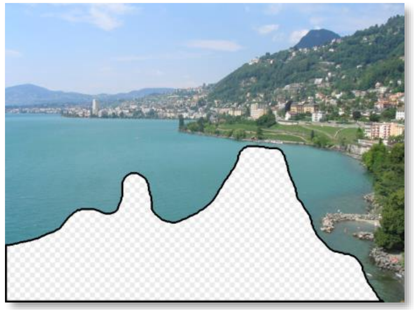


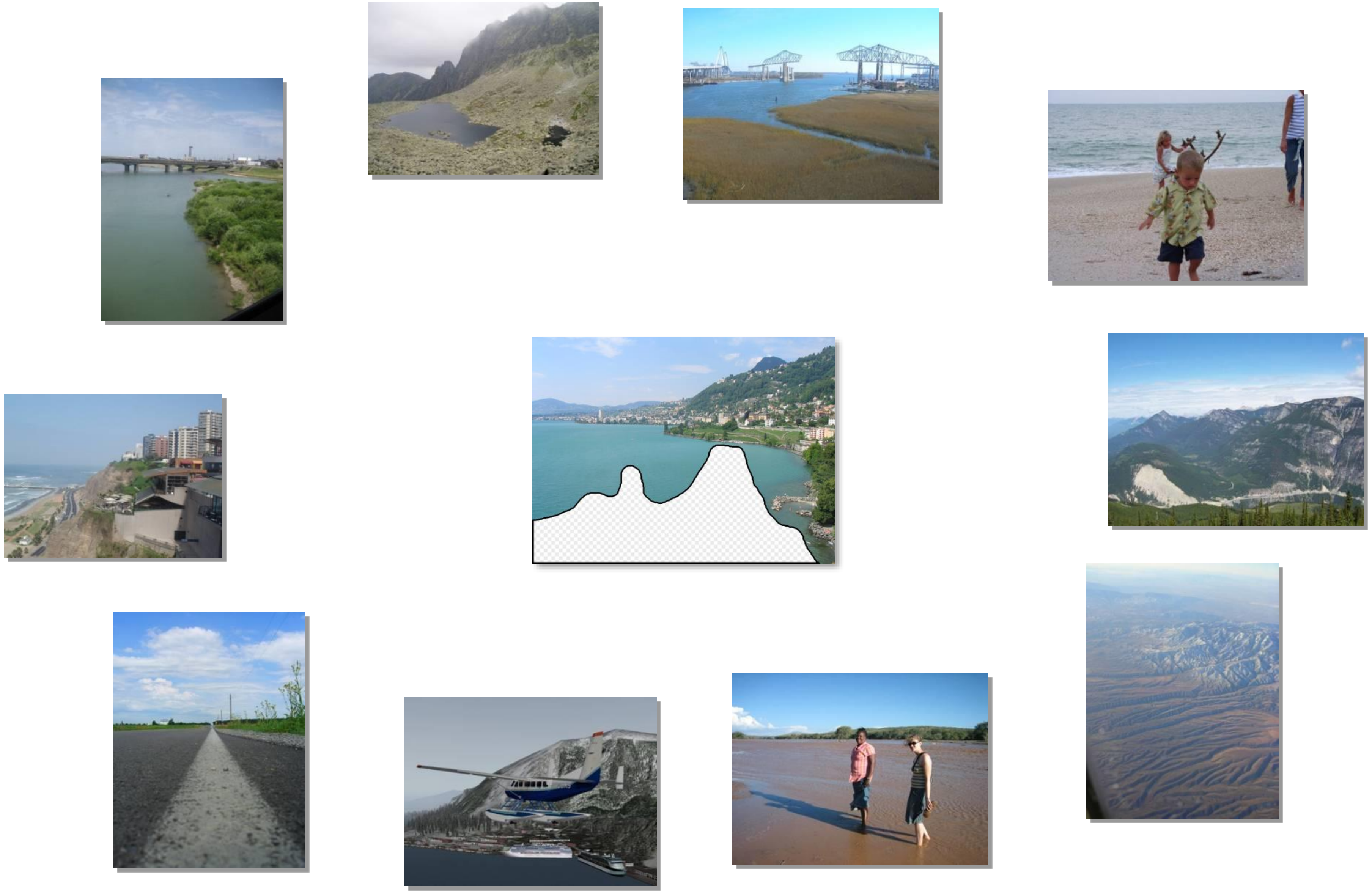
General Principal



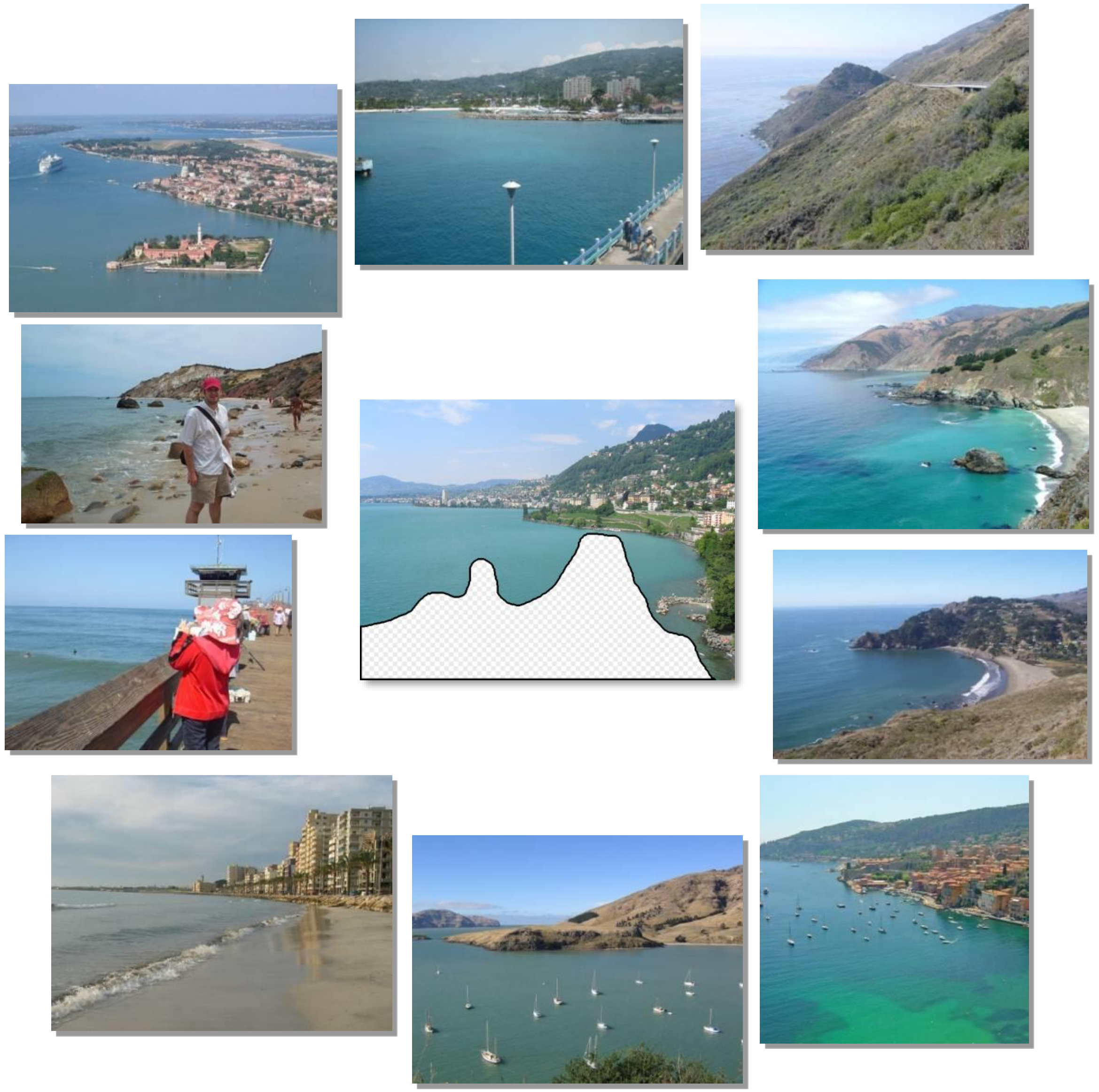
Hopefully, If you have enough images, the dataset will contain very similar images that you can find with simple matching methods.

How many images is enough?





Nearest neighbors from a
collection of 20 thousand images



Nearest neighbors from a
collection of 2 million images

Image Data on the Internet

- Flickr (as of Sept. 19th, 2010)
  - 5 billion photographs
  - 100+ million geotagged images
- Facebook (as of 2009)
  - 15 billion

http://royal.pingdom.com/2010/01/22/internet-2009-in-numbers/

Image Data on the Internet

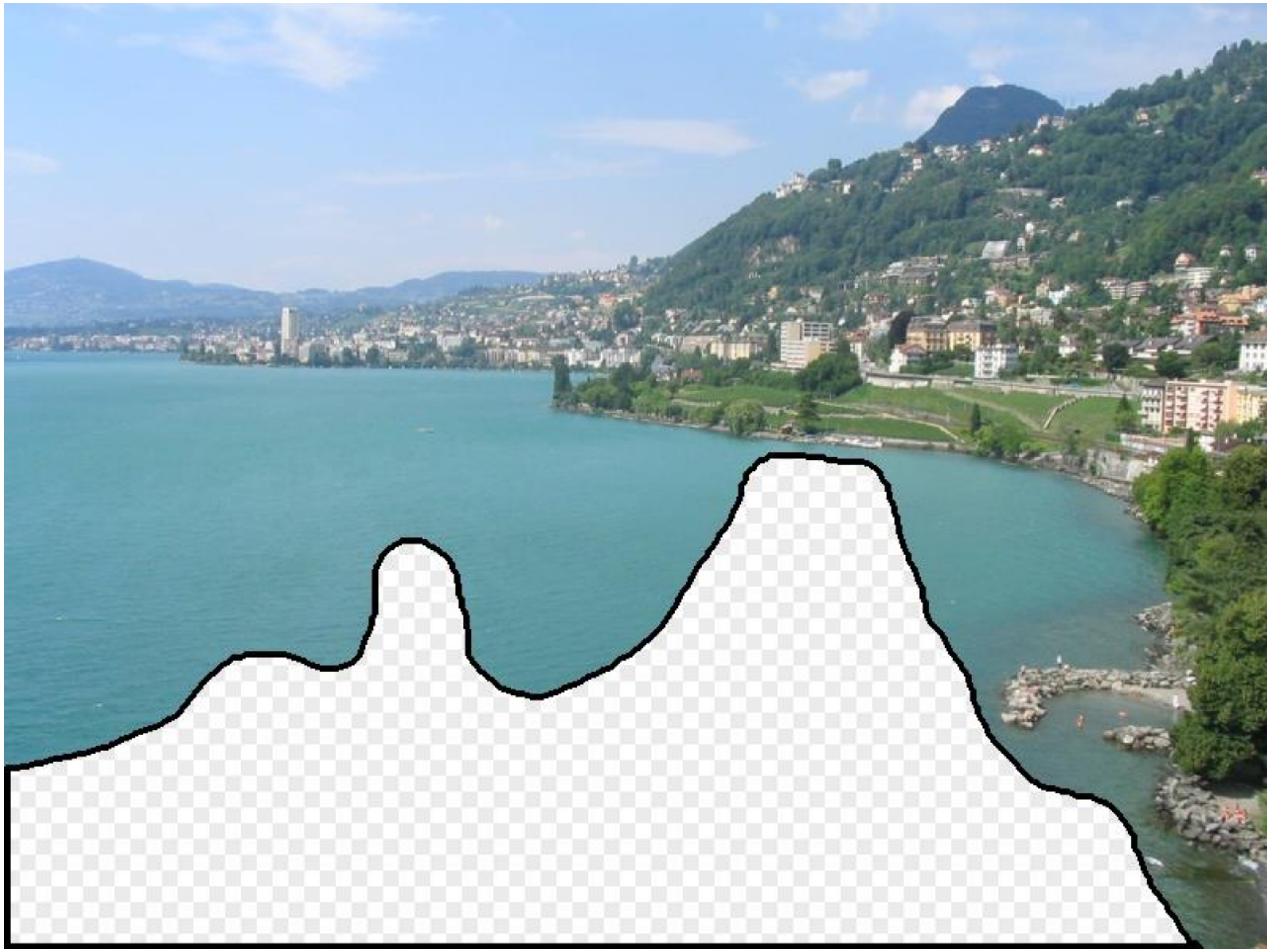
- Flickr (as of Nov 2013)
  - 10 billion photographs
  - 100+ million geotagged images
  - 3.5 million a day
- Facebook (as of Sept 2013)
  - 250 billion+
  - 300 million a day
- Instagram
  - 55 million a day

Scene Completion: how it works

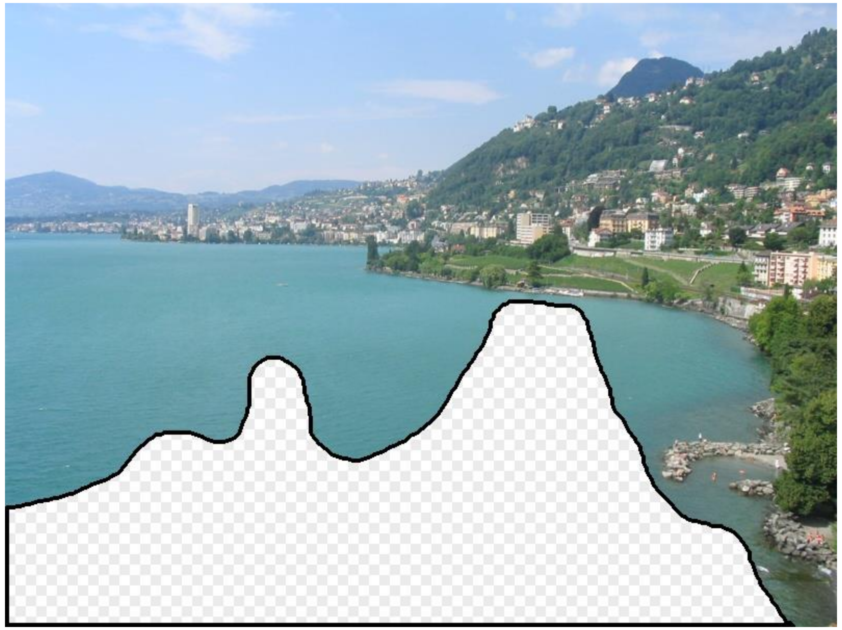
[Hays and Efros. Scene Completion Using Millions of Photographs. SIGGRAPH 2007 and CACM October 2008.]

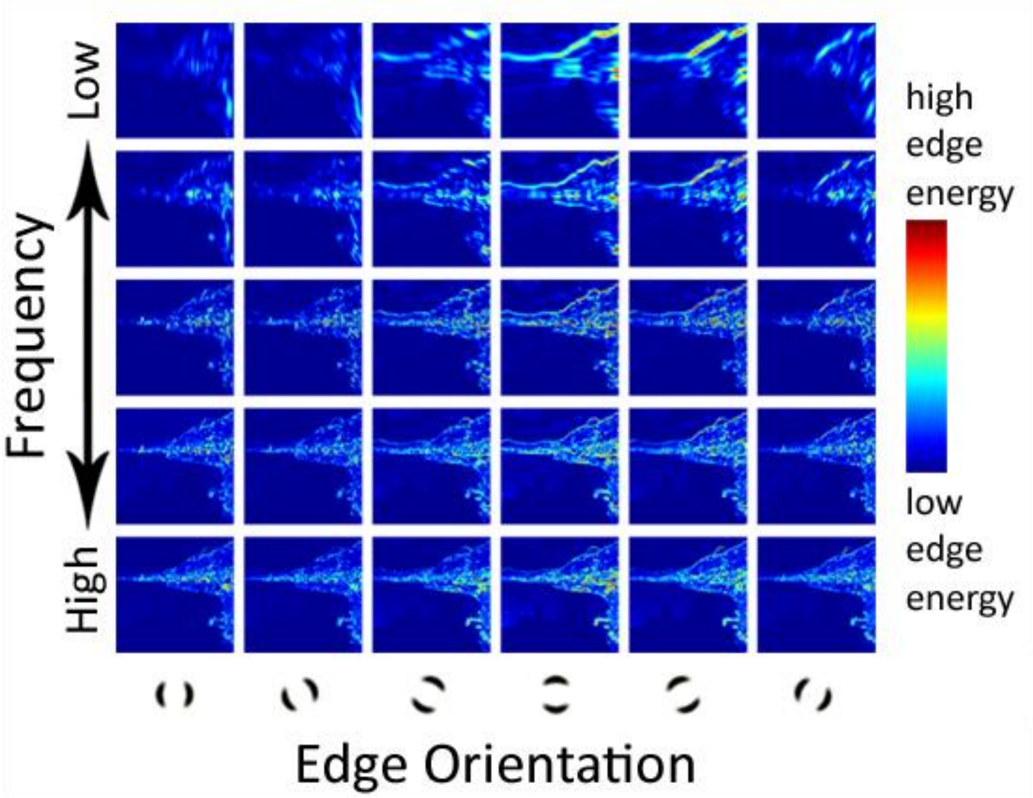
The Algorithm

Scene Matching

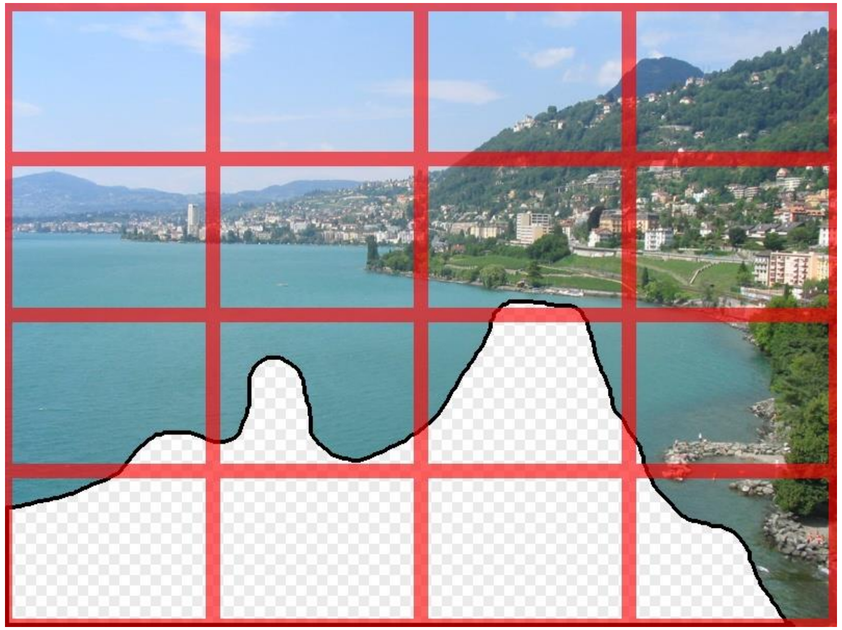


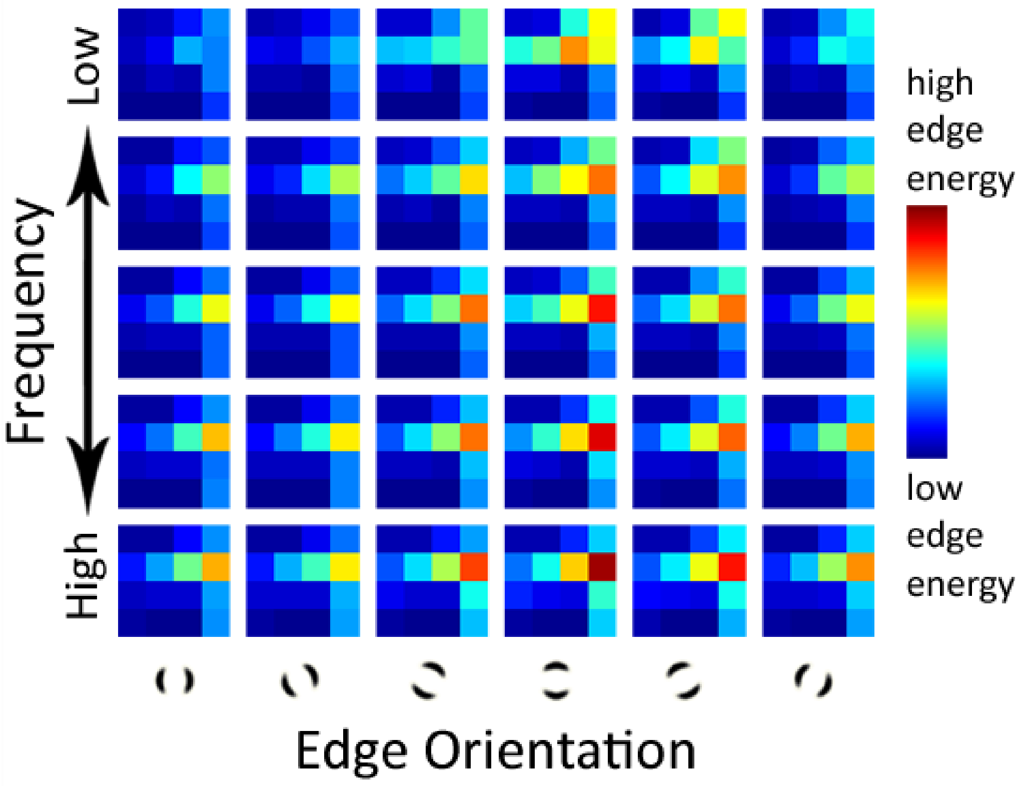
Scene Descriptor





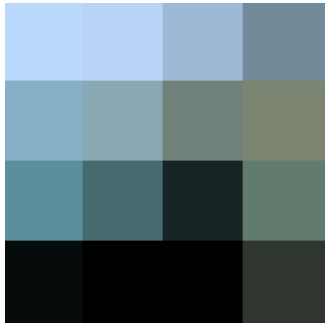
Scene Descriptor



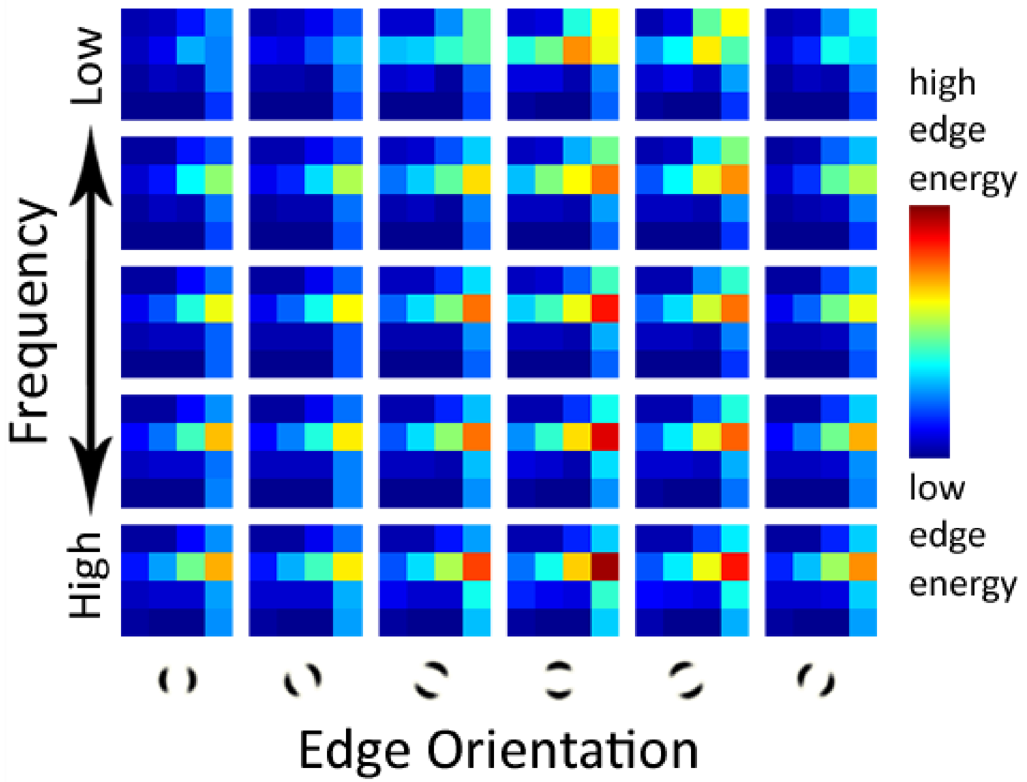


Scene Gist Descriptor
(Oliva and Torralba 2001)

Scene Descriptor

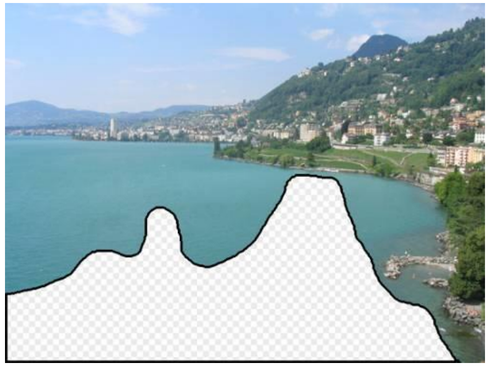


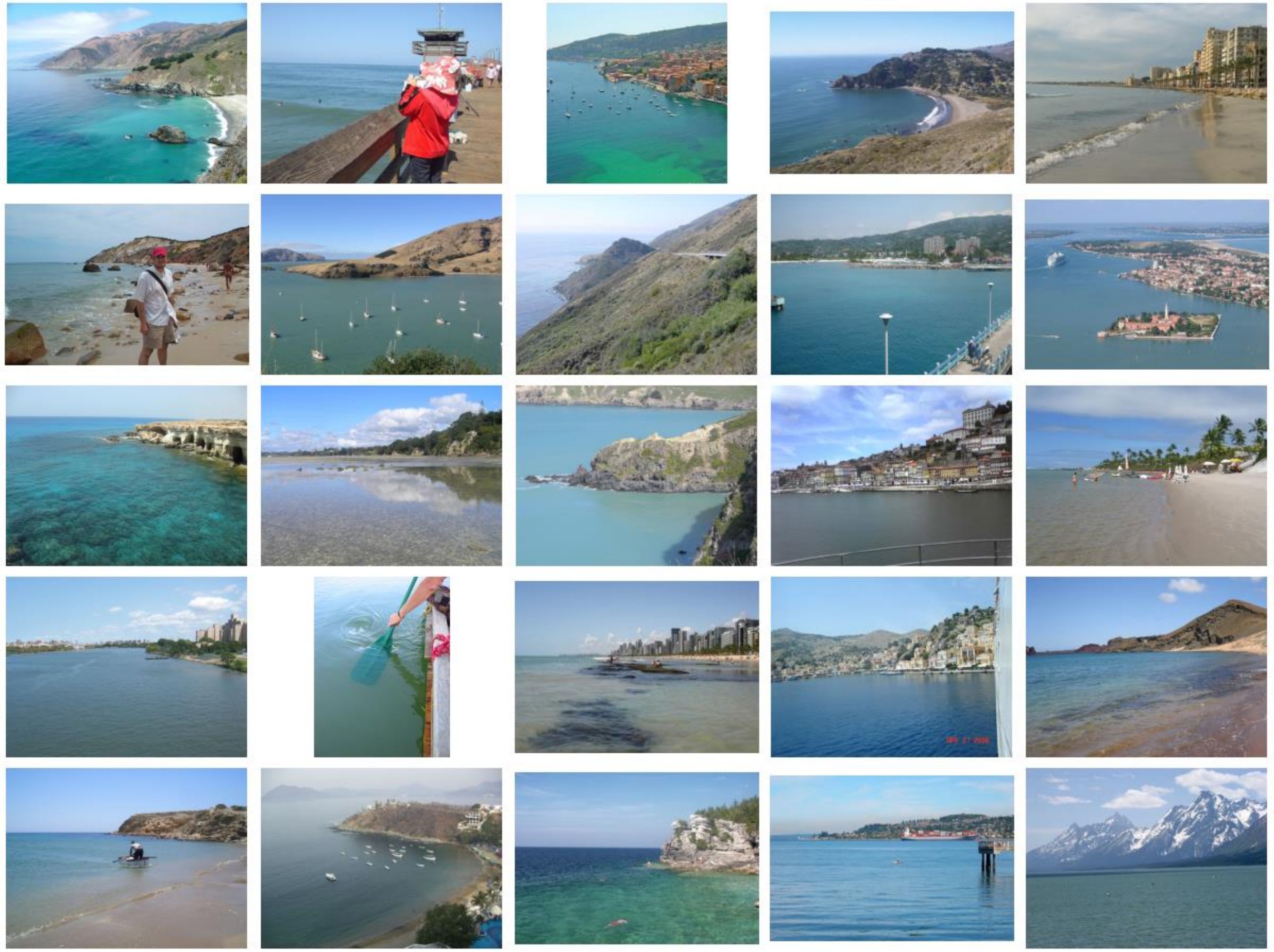
+



Scene Gist Descriptor
(Oliva and Torralba 2001)

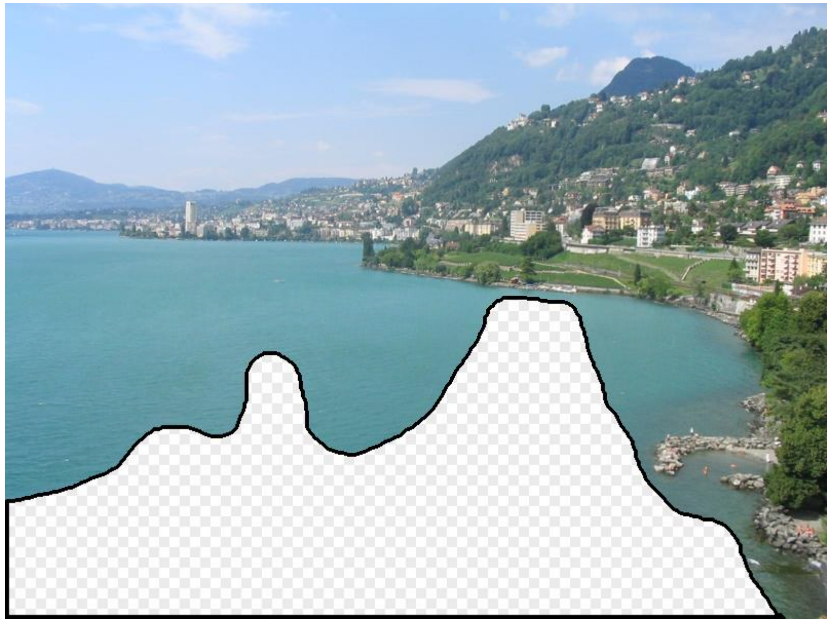
2 Million Flickr Images





… 200 total

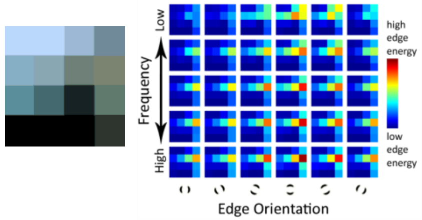
Context Matching



Graph cut + Poisson blending

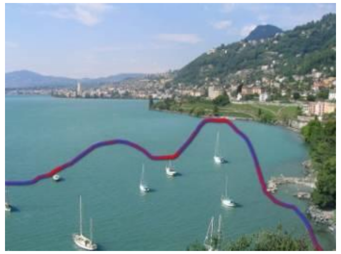
Result Ranking

We assign each of the 200 results a score which is the sum of:

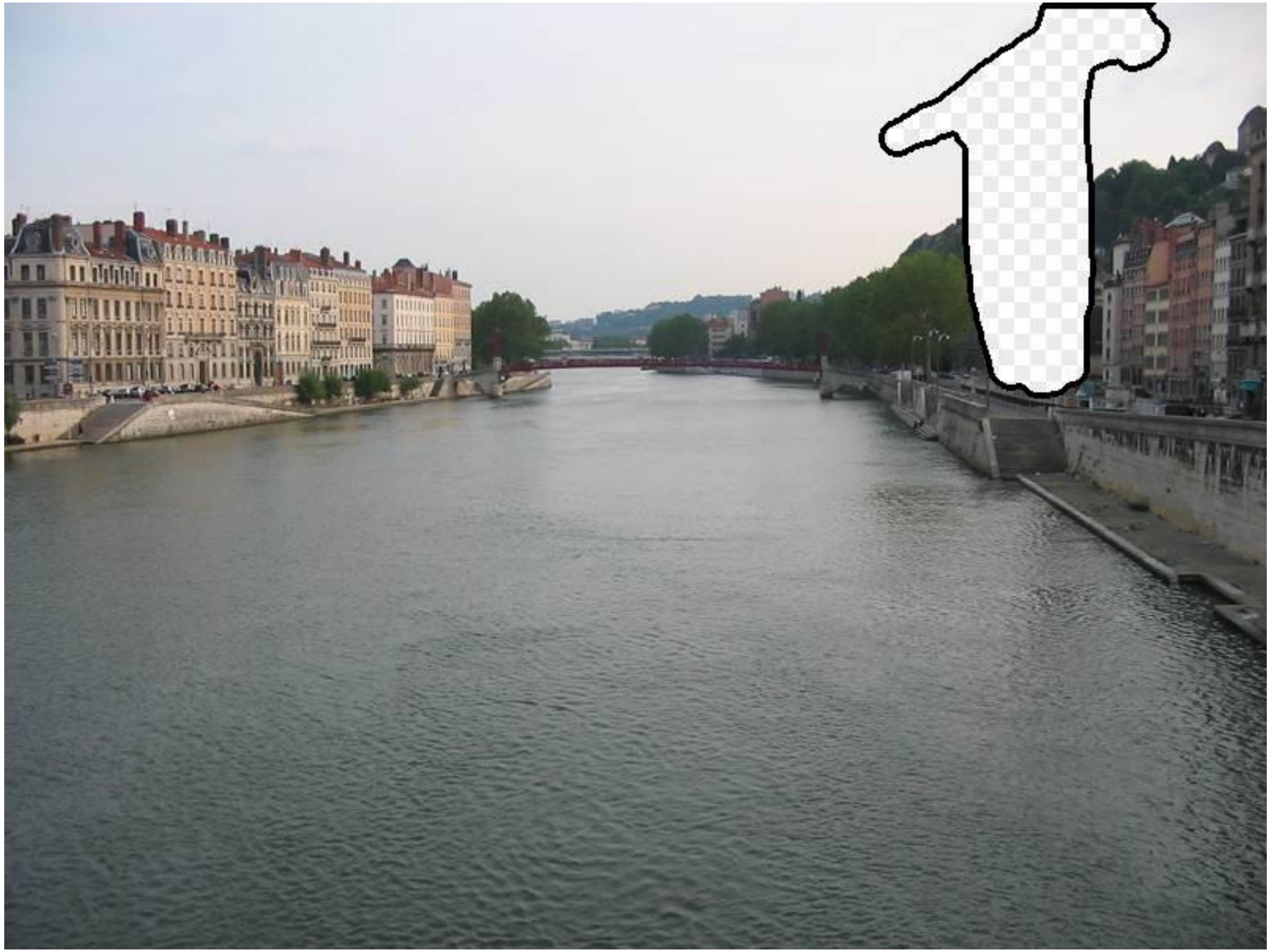


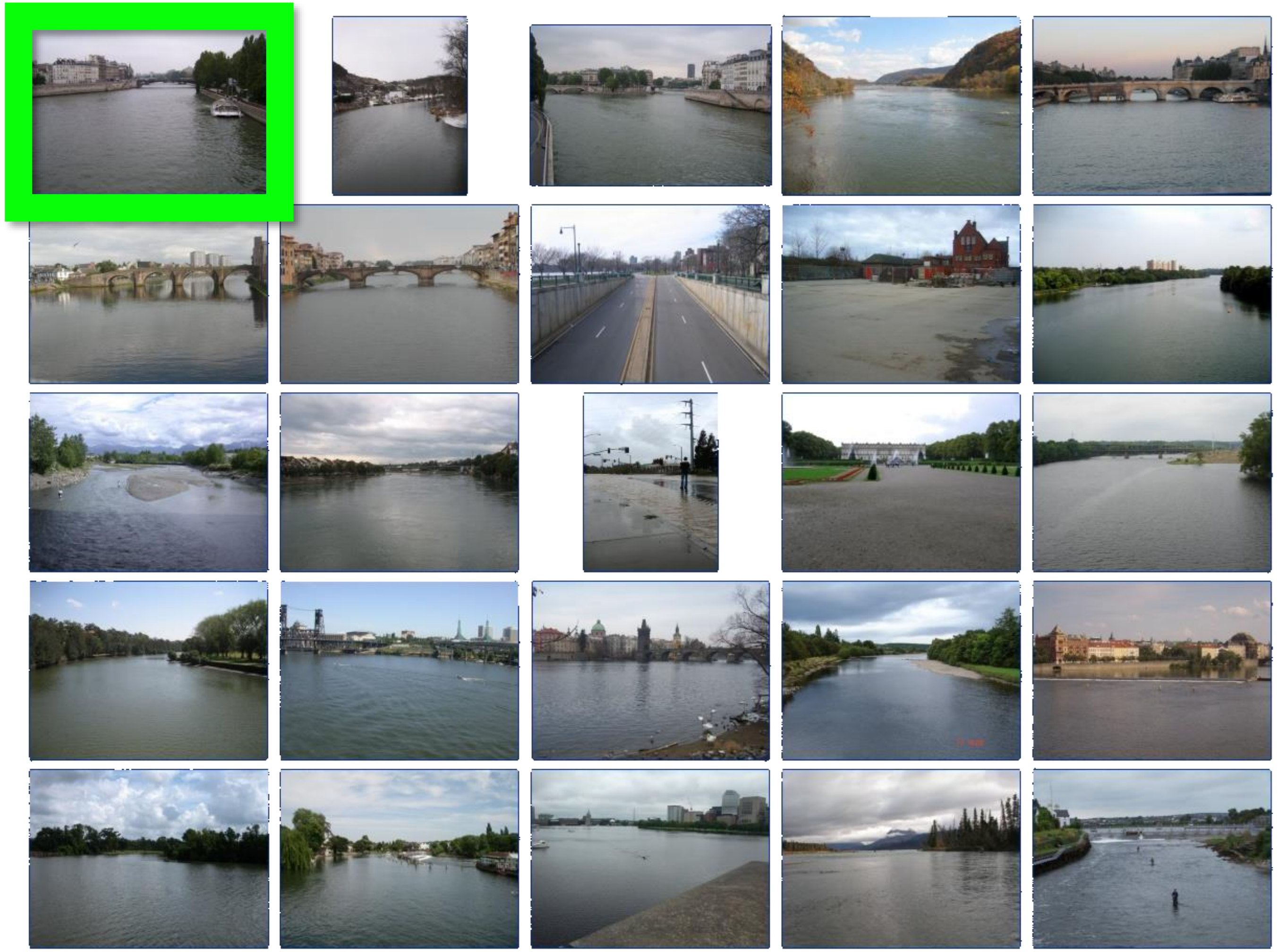
The scene matching distance

The context matching distance
(color + texture)

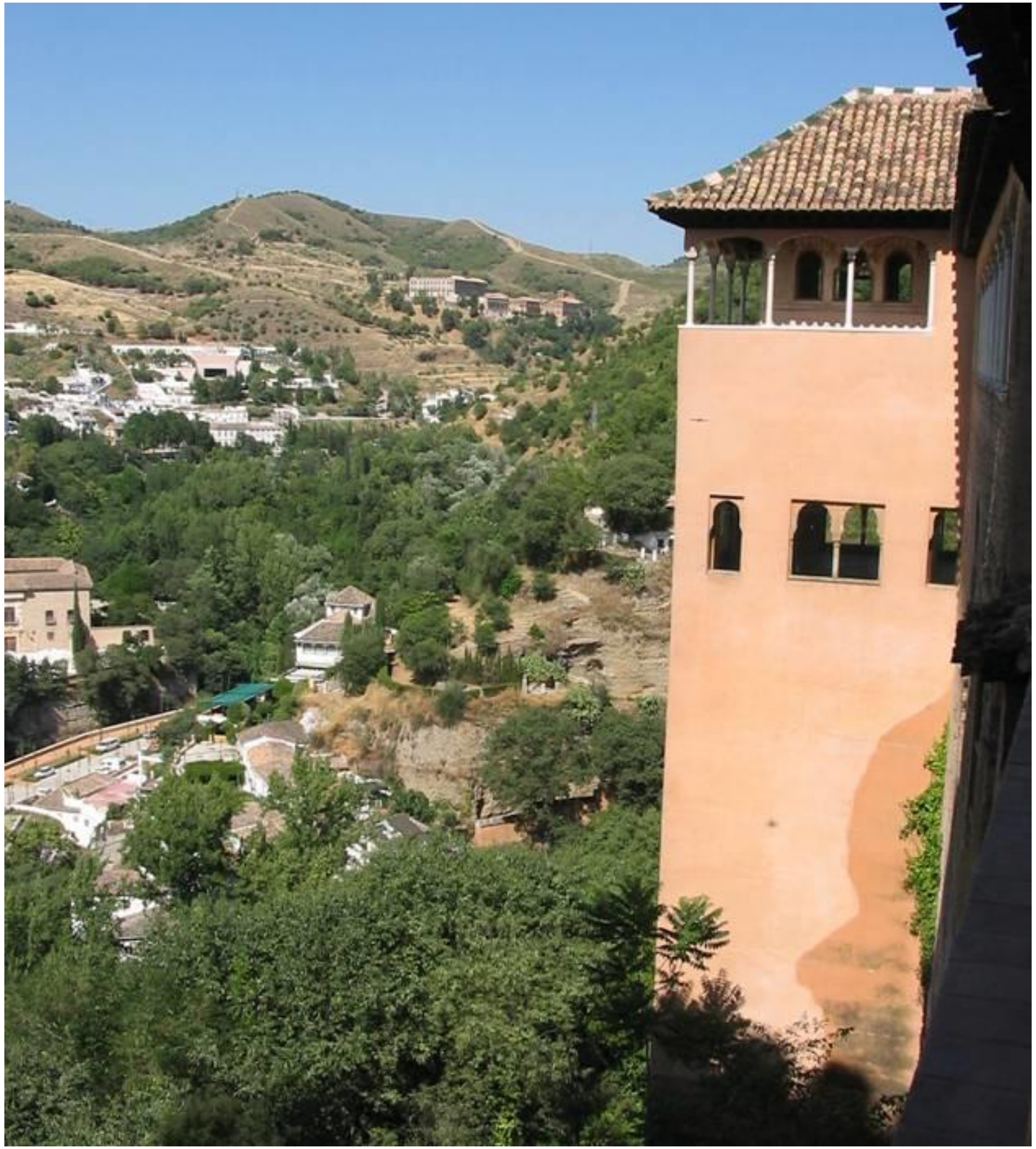


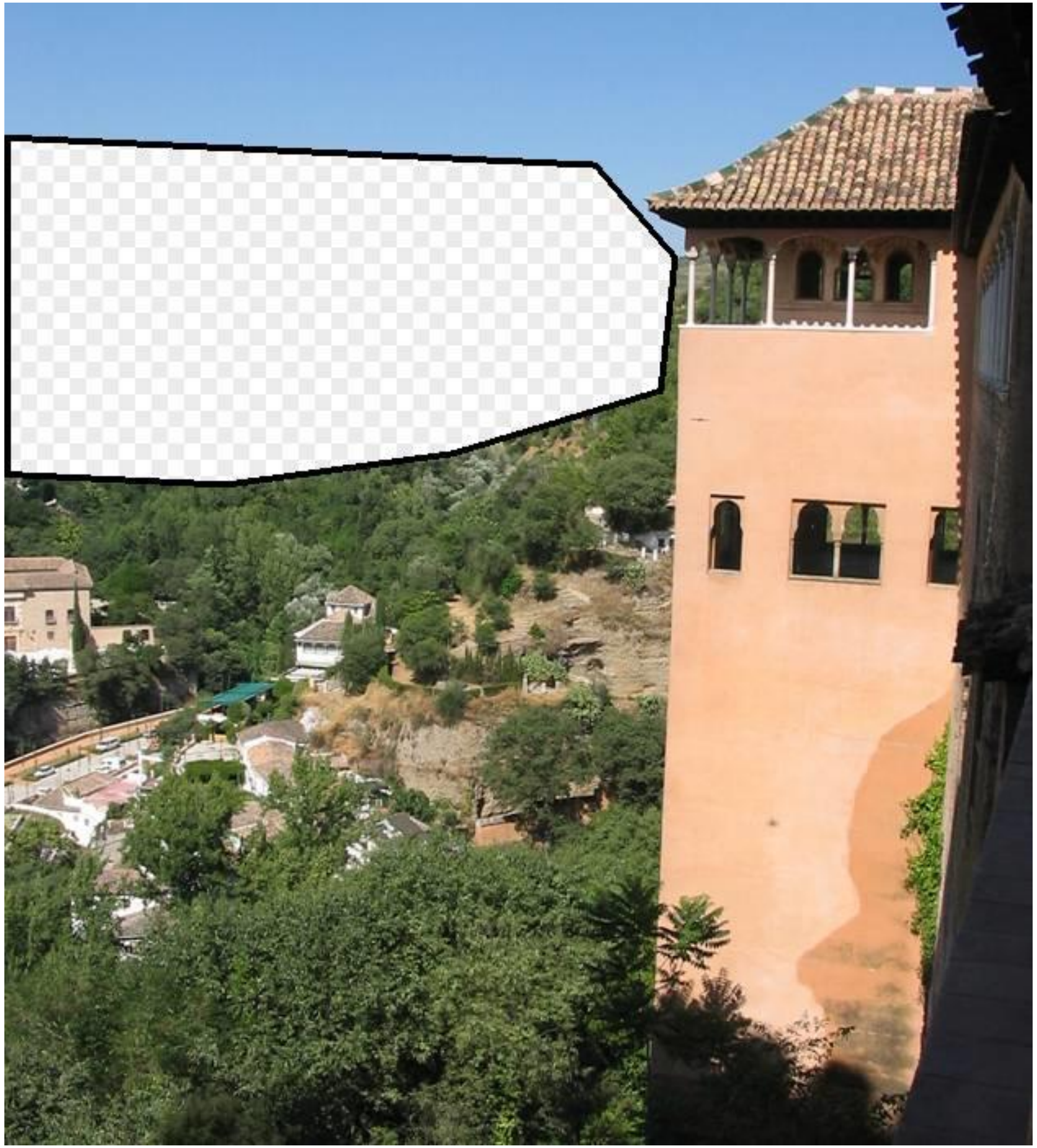
The graph cut cost



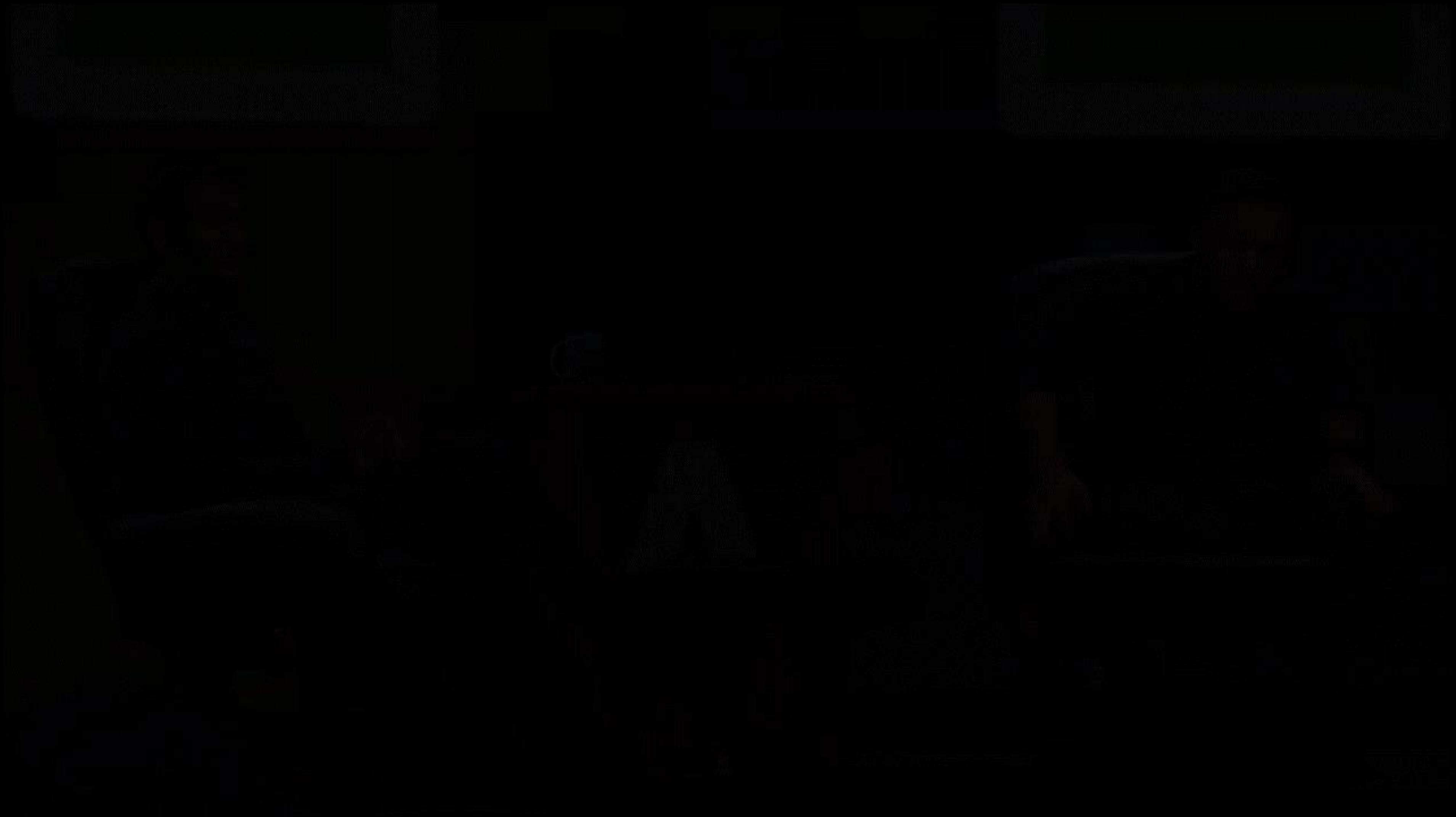


… 200 scene matches





Which is the original?

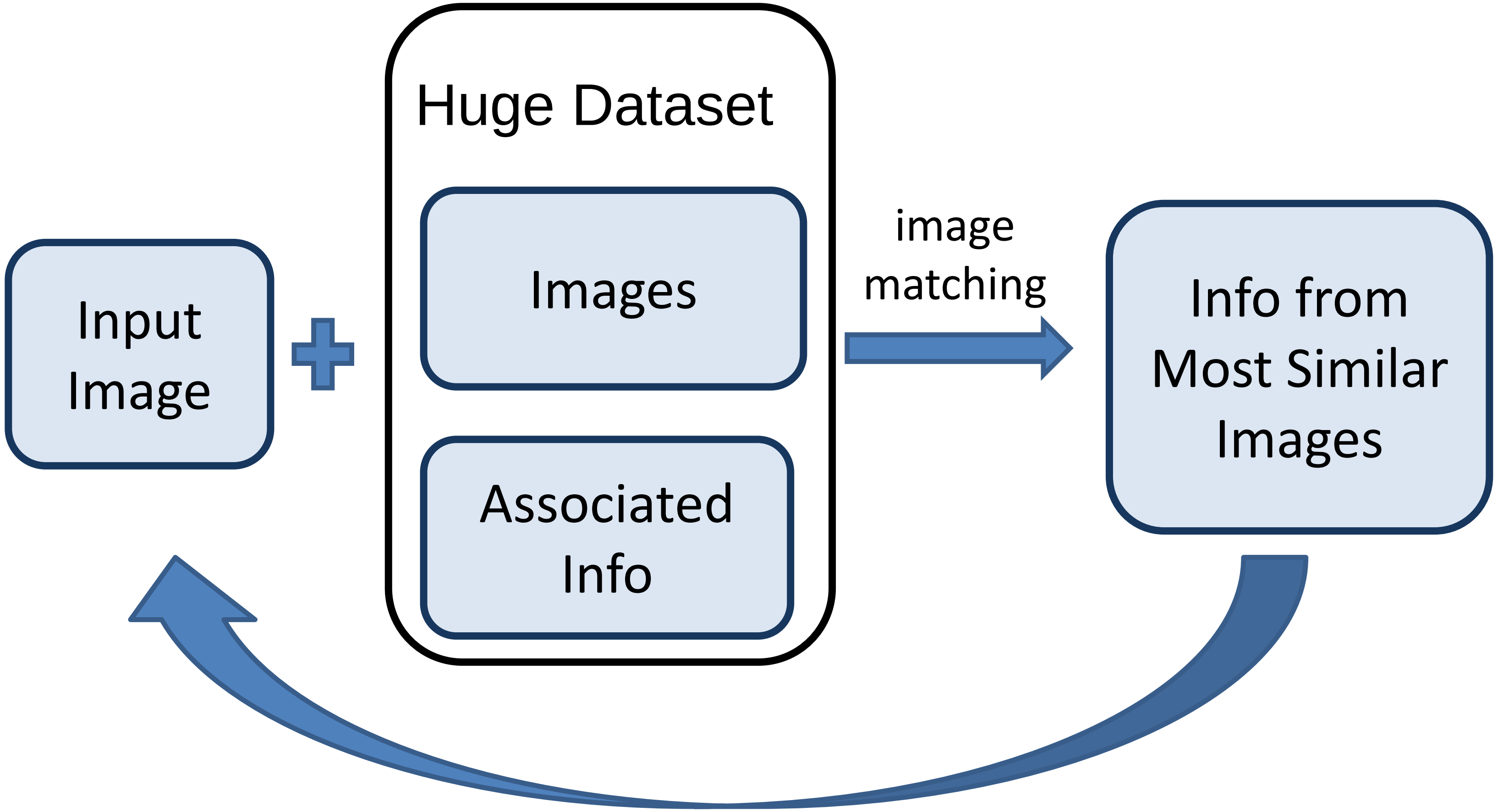


Outline

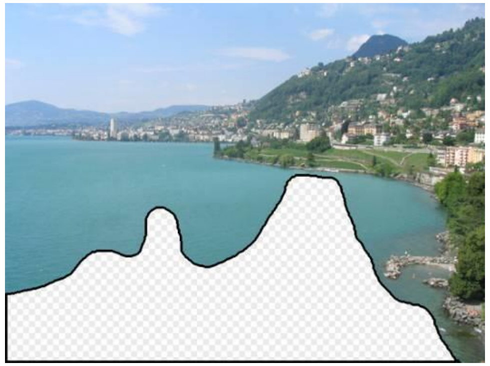
Opportunities of Scale: Data-driven methods

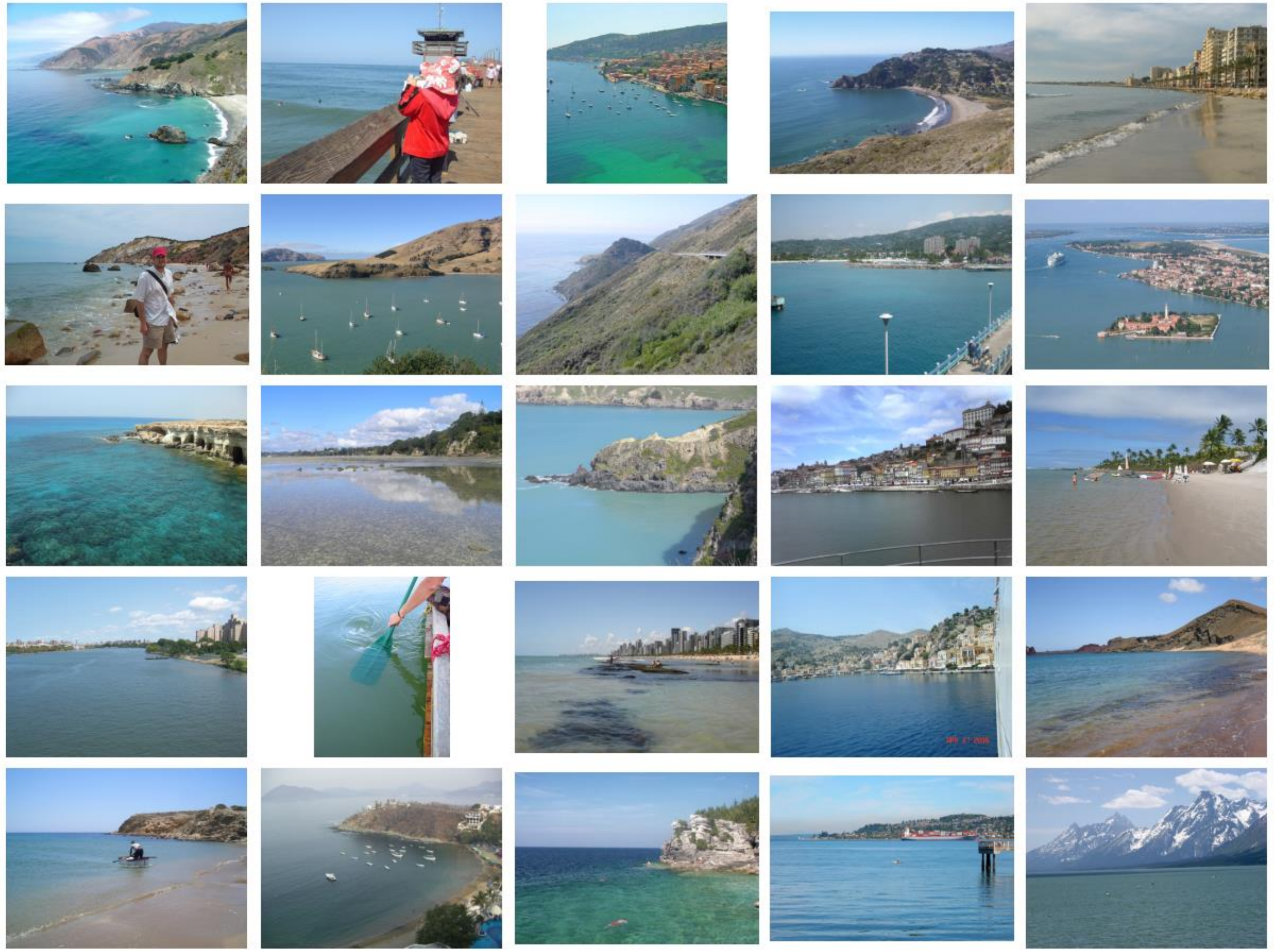
- The Unreasonable Effectiveness of Data
- Scene Completion
- Im2gps
- Recognition via Tiny Images

General Principal



Hopefully, If you have enough images, the dataset will contain very similar images that you can find with simple matching methods.





... 200 total

Graph cut + Poisson blending

Kosta Derpanis
@CSProfKGD

This reminded me of @jhhays and Efros' large-scale image geolocalization work

This Geography Genius Can Figure Out Exactly Where a Photo Was Shot
Tom Davies (AKA GeoWizard) is a human photo geotagger. He can figure out exactly where an outdoor photo was shot by studying it carefully.
petapixel.com

11:08 PM · Mar 4, 2021 from Toronto, Ontario · Twitter for iPhone

3 Likes

https://www.geoguessr.com/

https://www.youtube.com/c/GeoWizard/videos

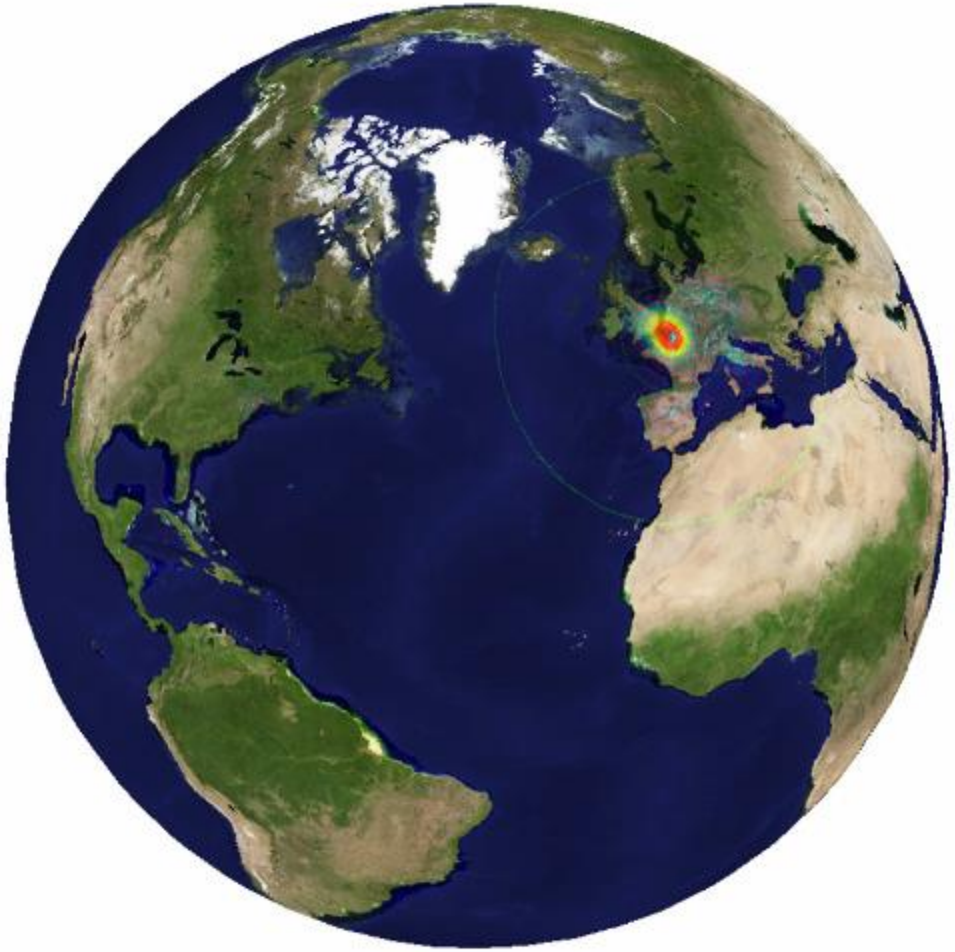
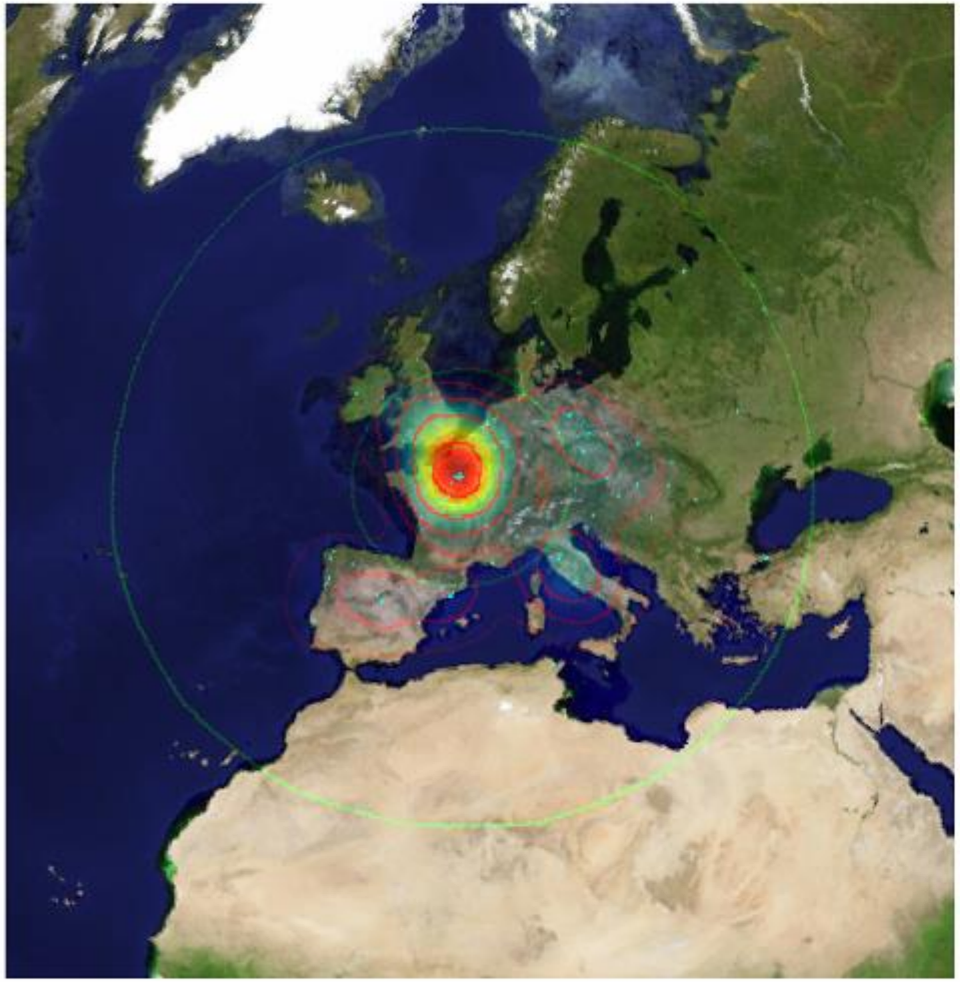
im2gps (Hays & Efros, CVPR 2008)

6 million geo-tagged Flickr images

http://graphics.cs.cmu.edu/projects/im2gps/

How much can an image tell about its geographic location?

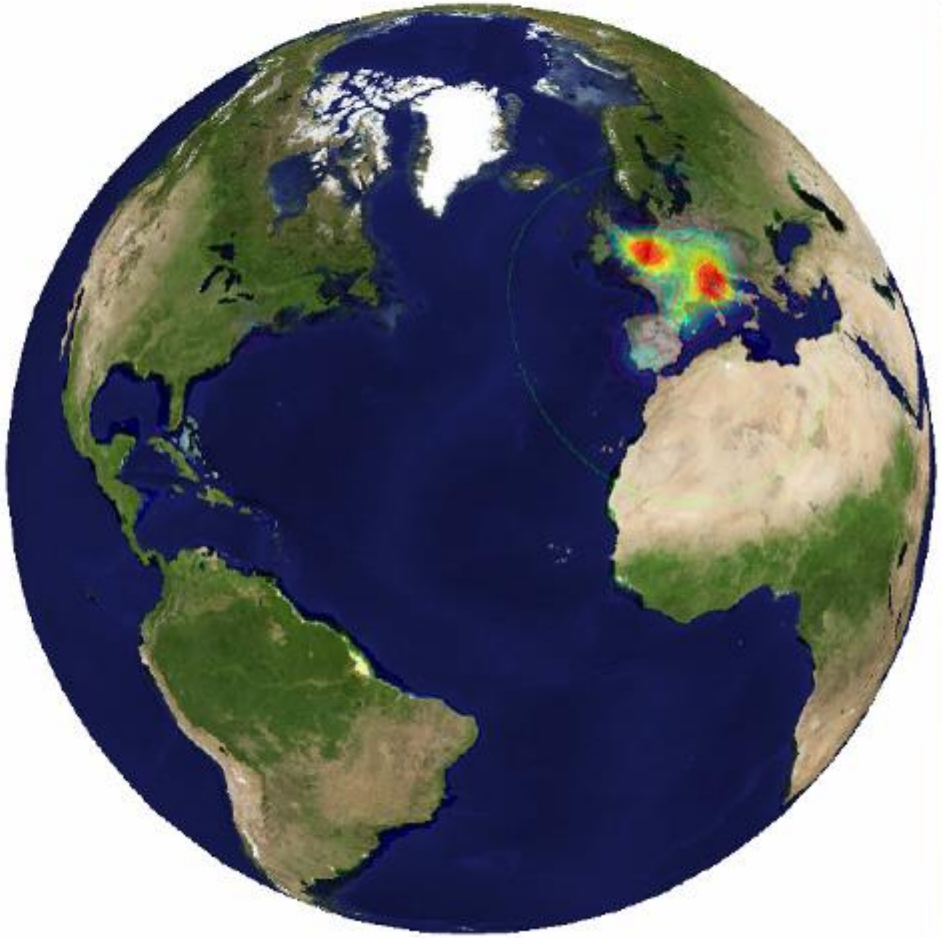
Nearest Neighbors according to gist + bag of SIFT + color histogram + a few others

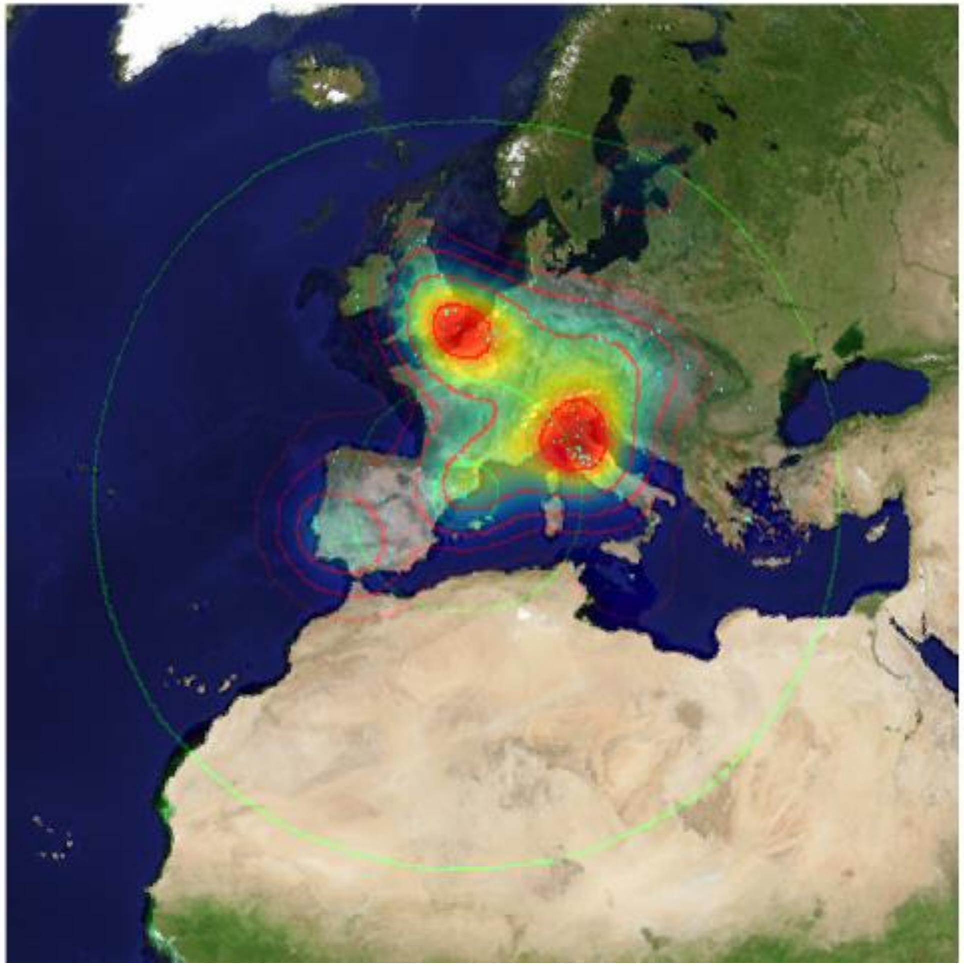



Im2gps

Example Scene Matches

Voting Scheme





im2gps

Philippines

Houston

Thailand

Houston

Maldives

Philippines

NewZealand

Bermuda

Palau

Mexico2

Brazil

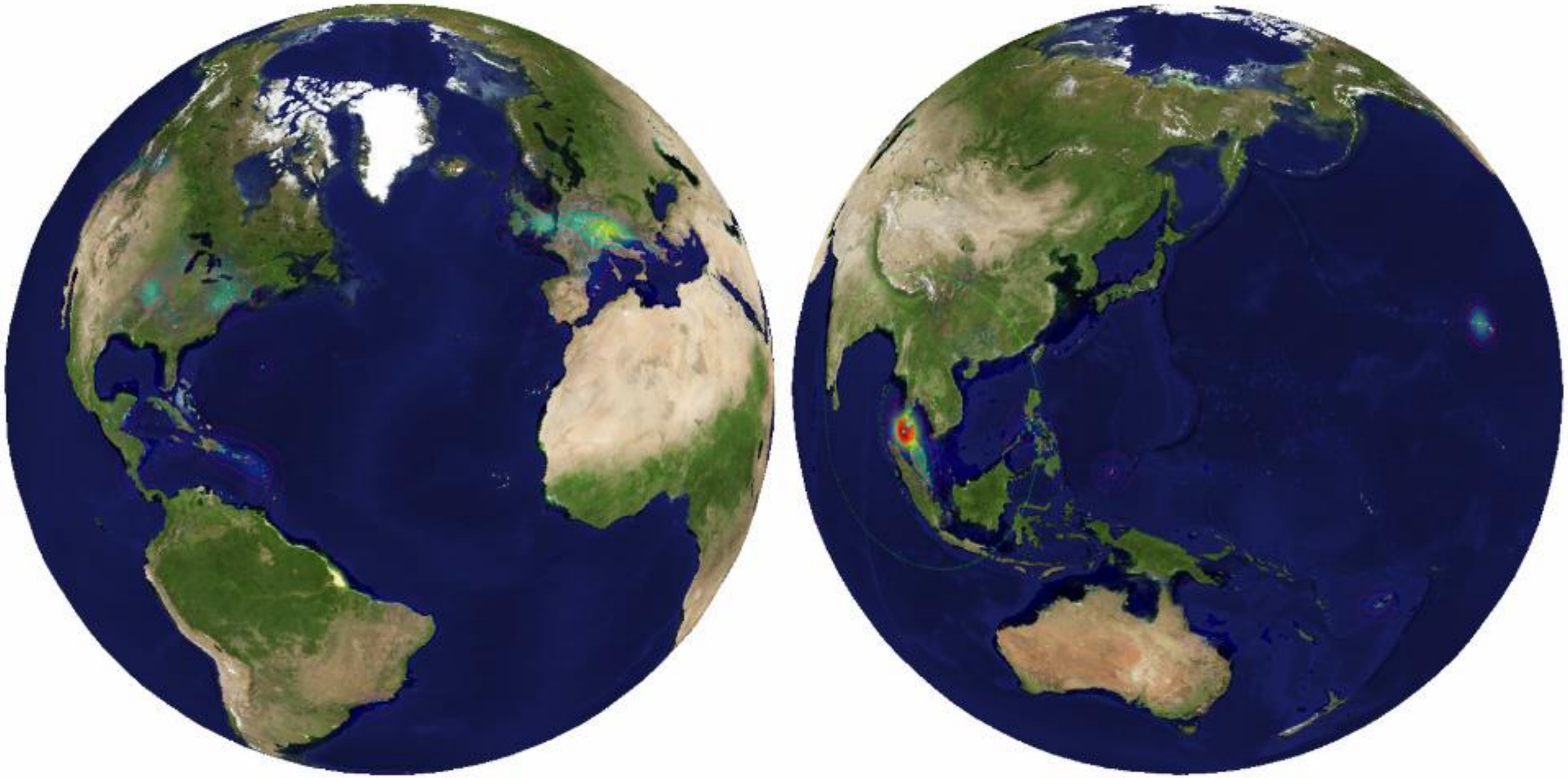
Mendoza

Brazil

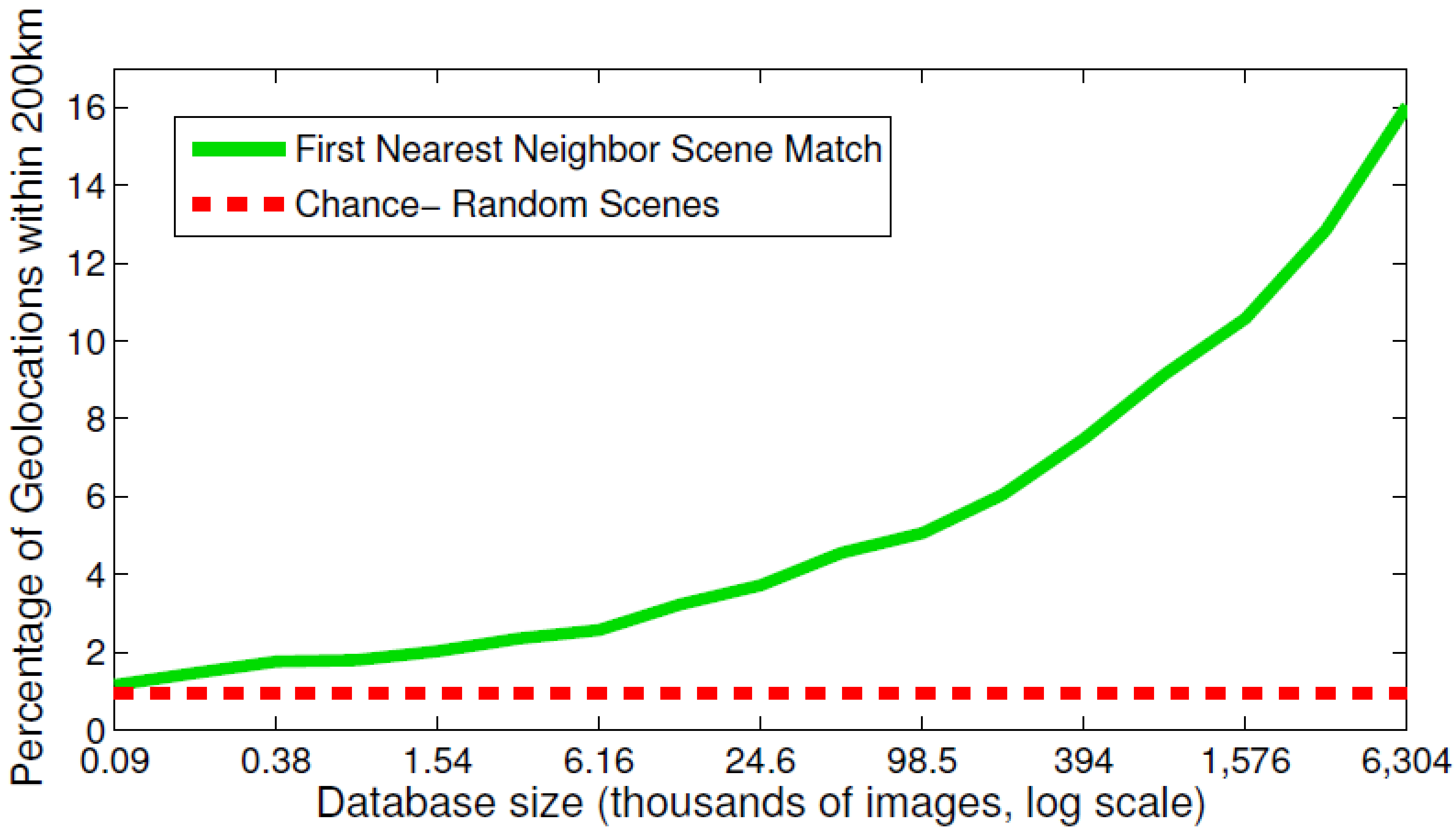
Thailand

Arkansas

Hawaii



Effect of Dataset Size



Follow up works

- PlaNet - photo geolocation with convolutional neural networks. T. Weyand, I. Kostrikov, and J. Philbin. ECCV 2016
- Revisiting IM2GPS in the Deep Learning Era.
Nam Vo, Nathan Jacobs, James Hays. ICCV 2017

Threshold (km)	Street 1	City 25	Region 200	Country 750	Cont. 2500
Human*			3.8	13.9	39.3
Im2GPS [9]		12.0	15.0	23.0	47.0
Im2GPS [10]	02.5	21.9	32.1	35.4	51.9
PlaNet [36]	08.4	24.5	37.6	53.6	**71.3**
[L] 7011C	06.8	21.9	34.6	49.4	63.7
[L] kNN, σ=4	**12.2**	**33.3**	**44.3**	**57.4**	**71.3**
... 28m database	**14.4**	**33.3**	**47.7**	**61.6**	**73.4**

Tiny Images

80 million tiny images: a large dataset for non-parametric object and scene recognition
Antonio Torralba, Rob Fergus and William T. Freeman. PAMI 2008.

http://groups.csail.mit.edu/vision/TinyImages/

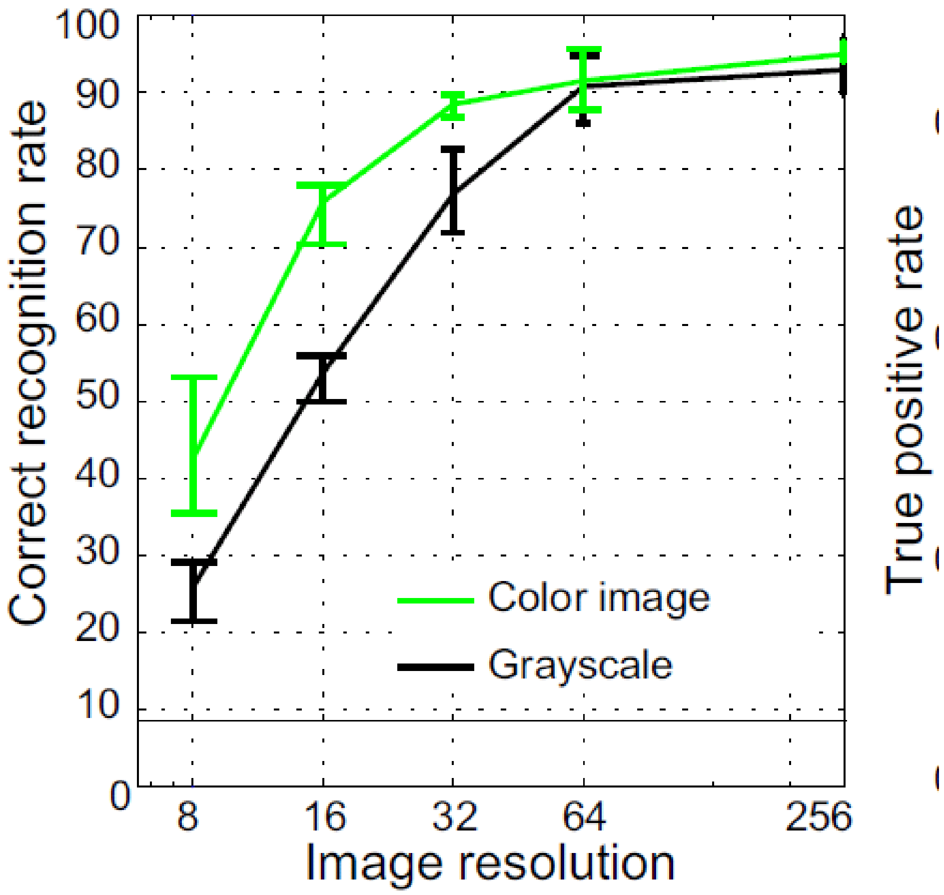
256x256


256x256
32x32
office
waiting area
dining room
dining room

c) Segmentation of 32x32 images

c) Segmentation of 32x32 images

Human Scene Recognition

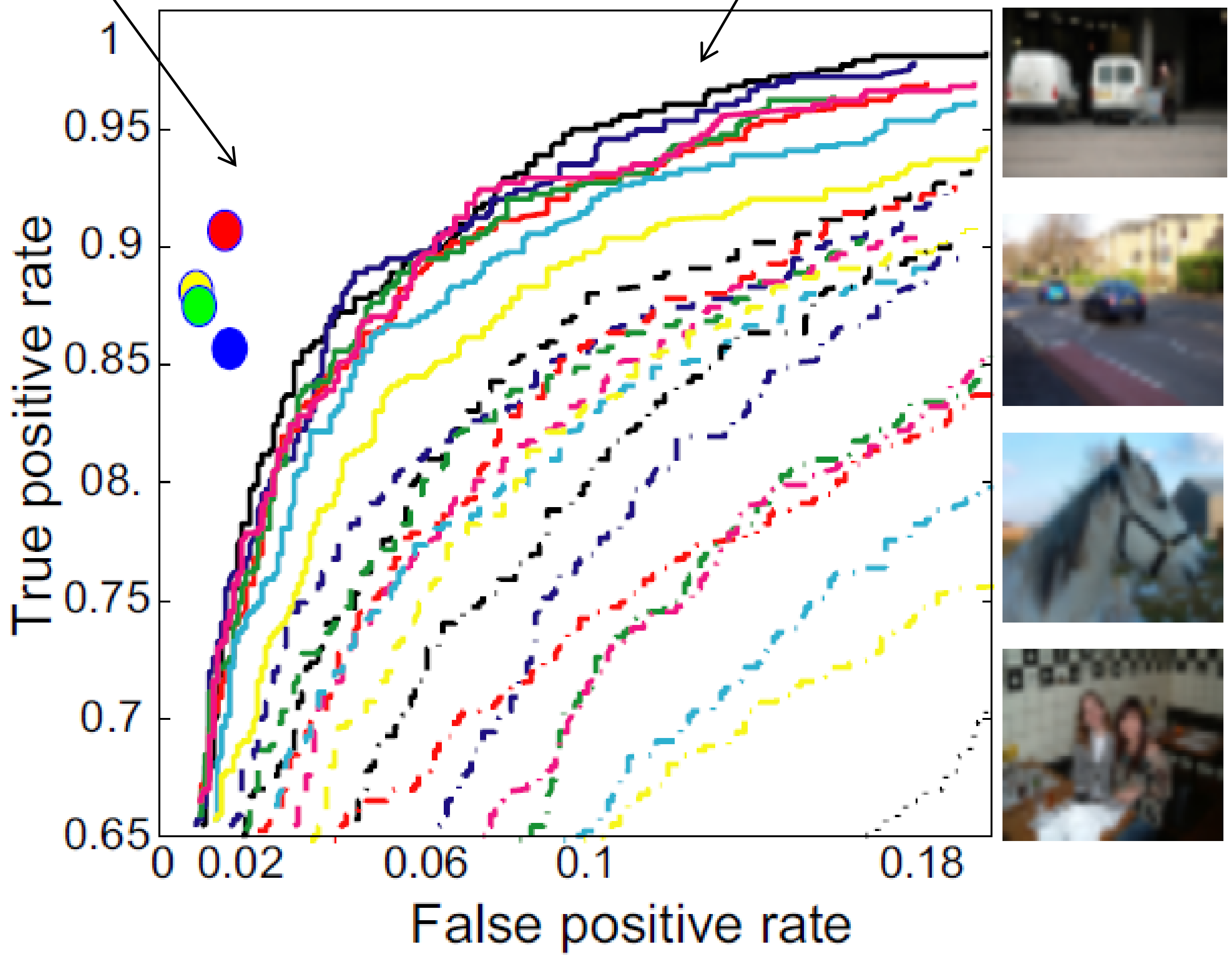


a) Scene recognition

Humans vs. Computers: Car-Image Classification

Humans for 32 pixel tall images

Various computer vision algorithms for full resolution images



Powers of 10

Number of images on my hard drive: 10^4

Number of images seen during my first 10 years: 10^8

(3 images/second * 60 * 60 * 16 * 365 * 10 = 630720000)

Number of images seen by all humanity: 10^{20}

106,456,367,669 humans[1] * 60 years * 3 images/second * 60 * 60 * 16 * 365 =

1 from http://www.prb.org/Articles/2002/HowManyPeopleHaveEverLivedonEarth.aspx

Number of photons in the universe: 10^{88}

Number of all 32x32 images: 10^{7373}

$256^{32*32*3} \sim 10^{7373}$

Scenes are unique

But not all scenes are so original





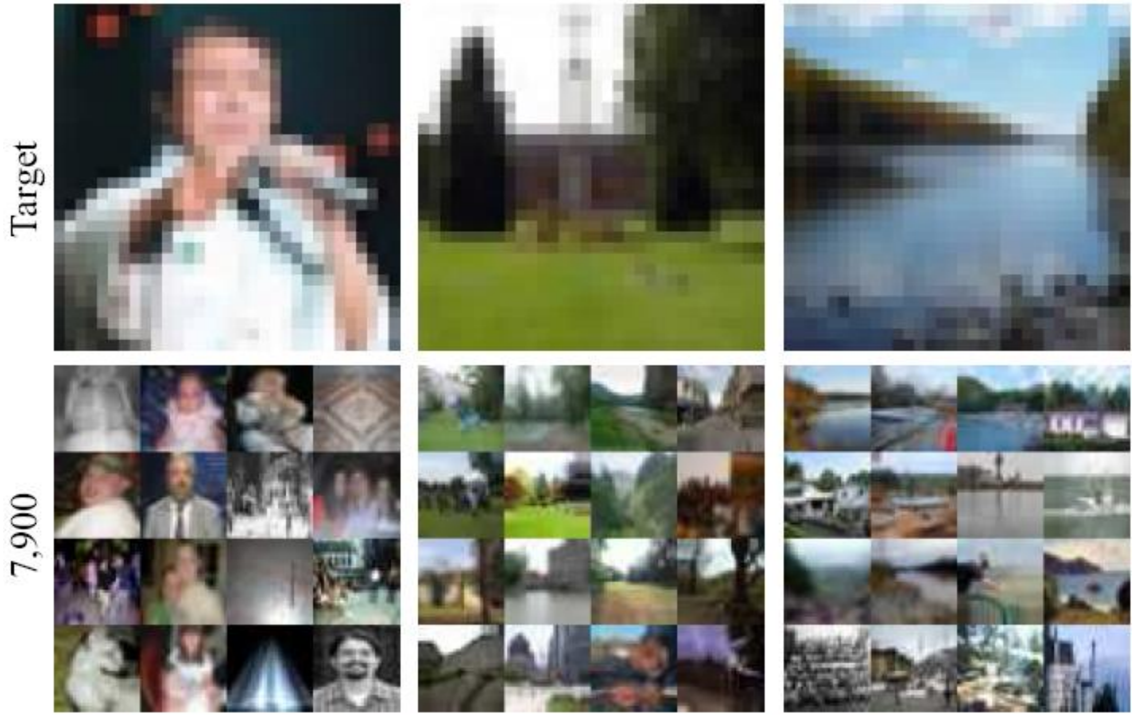




Lots

Of

Images

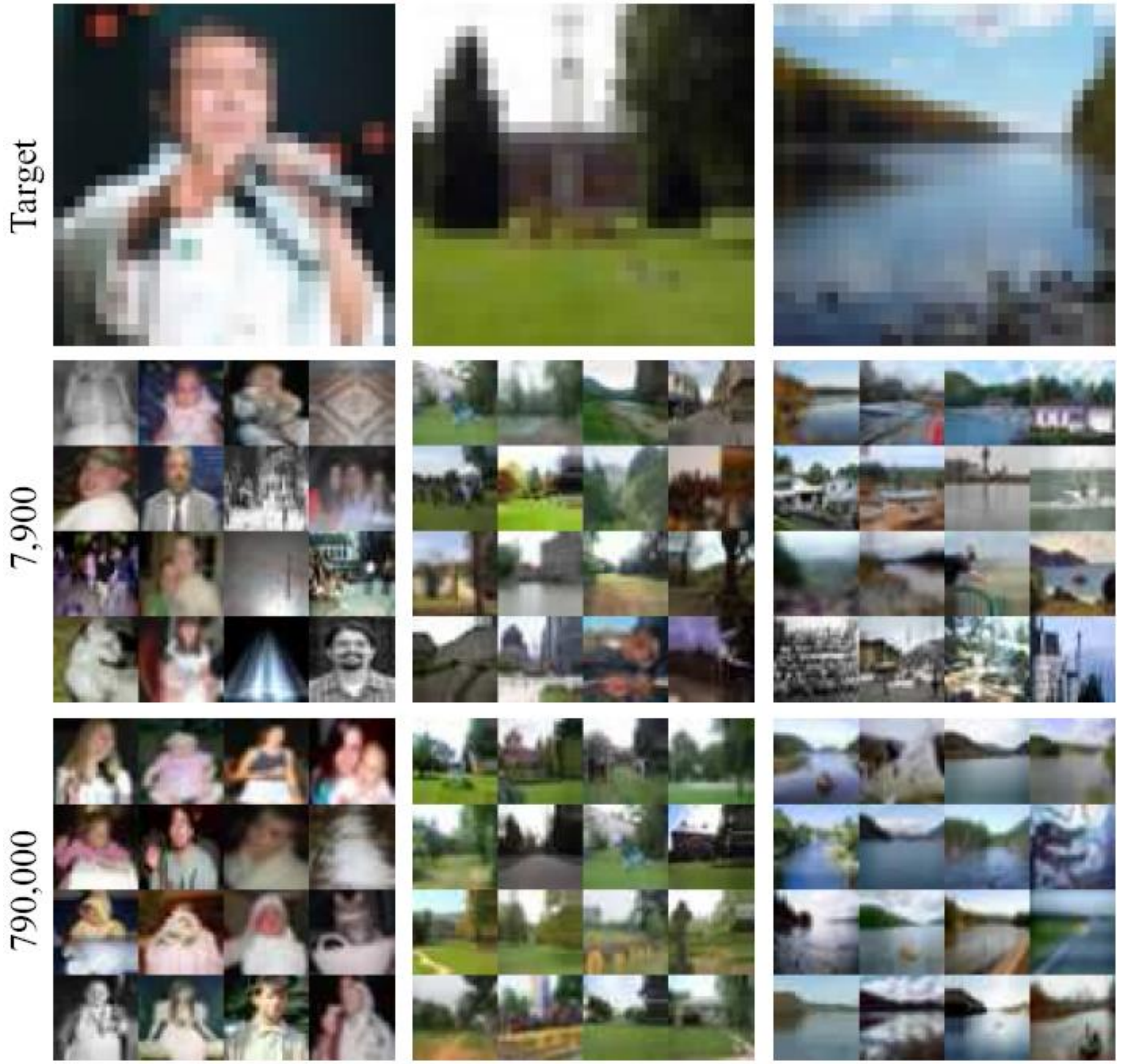


A. Torralba, R. Fergus, W.T.Freeman. PAMI 2008

Lots

Of

Images

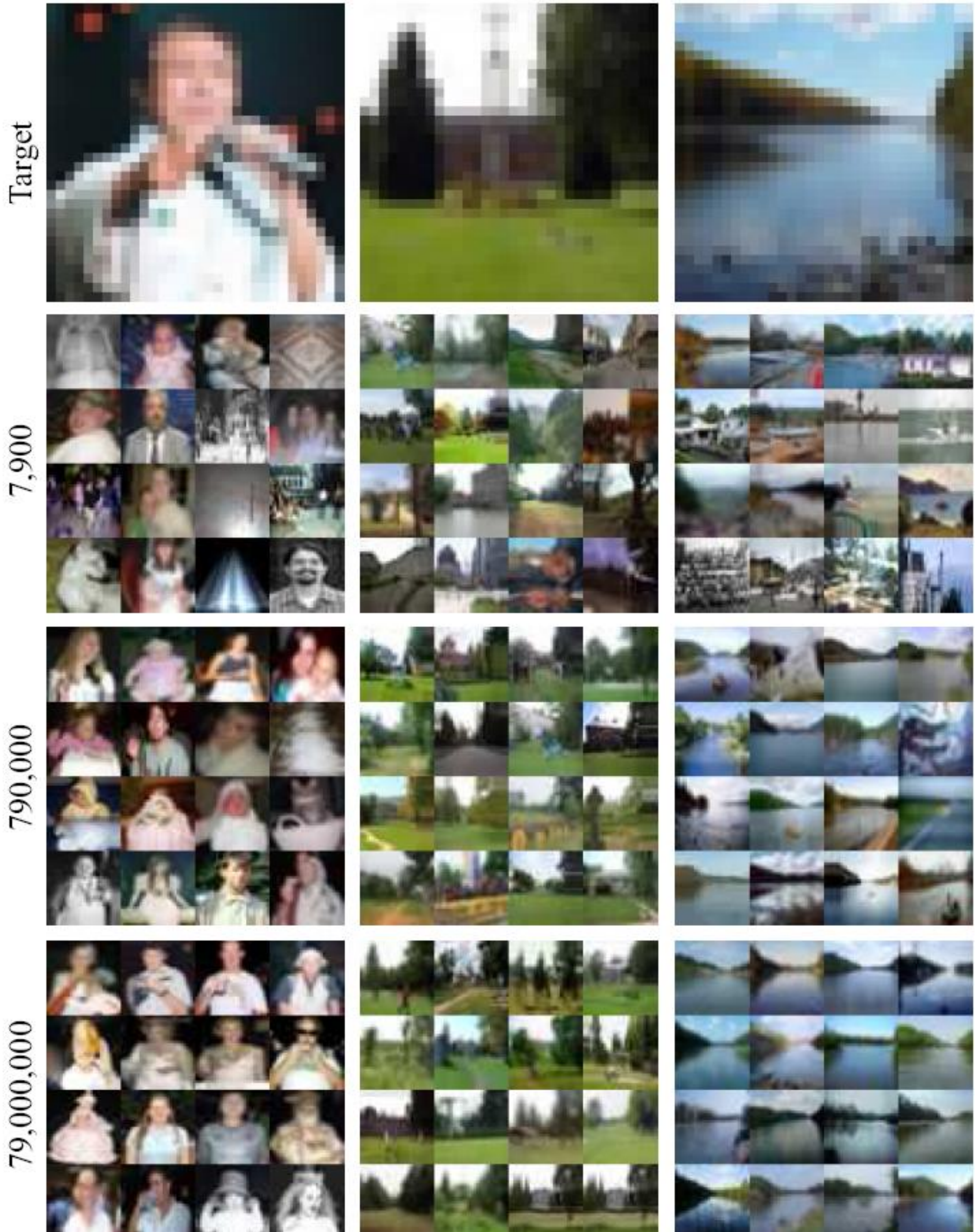


A. Torralba, R. Fergus, W.T.Freeman. PAMI 2008

Lots

Of

Images



Application: Automatic Colorization

Input

Color Transfer

Color Transfer

Matches (gray)

Matches (w/ color)

Avg Color of Match

Application: Automatic Colorization

Input

Color Transfer

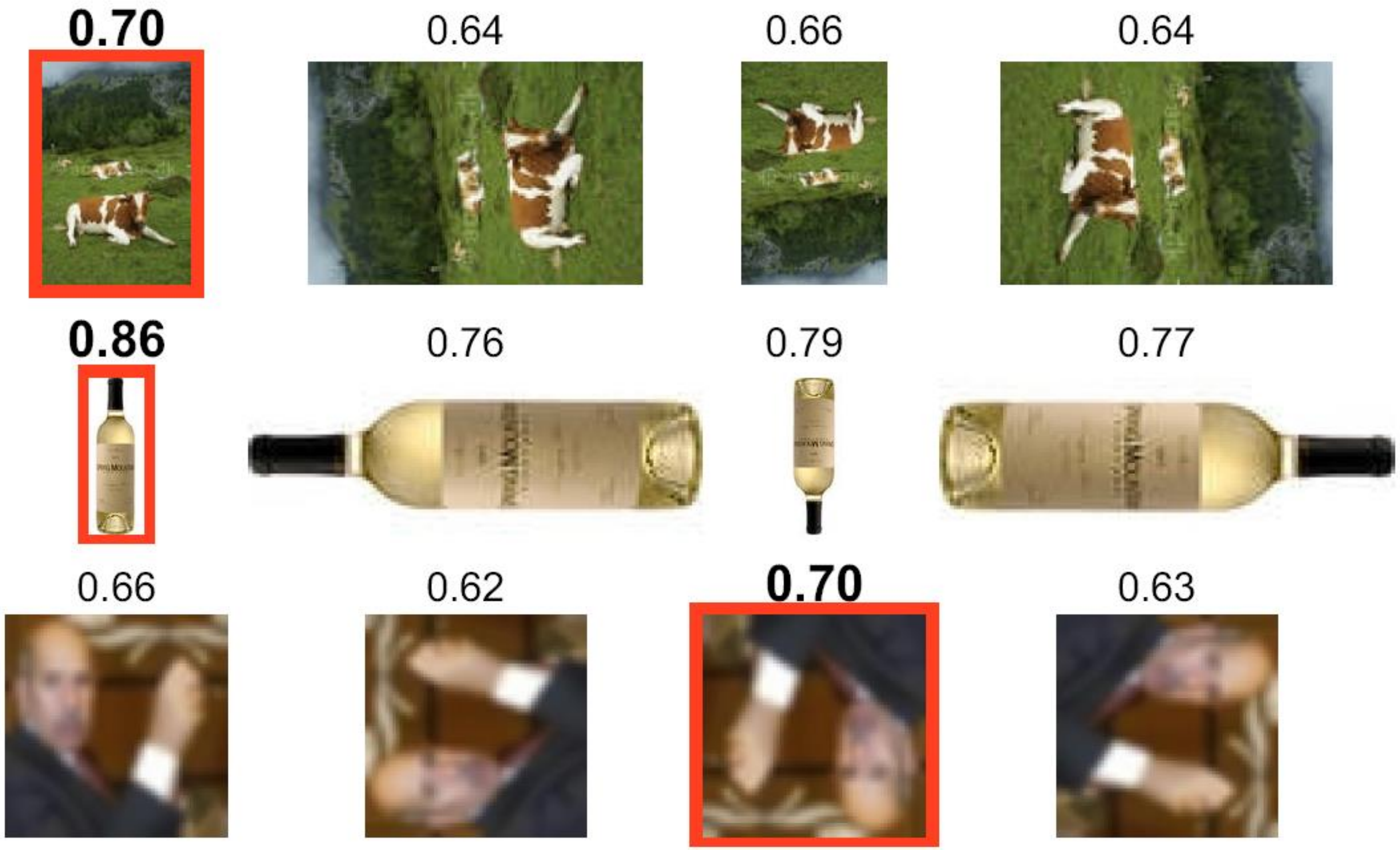
Color Transfer

Matches (gray)

Matches (w/ color)

Avg Color of Match

Automatic Orientation Examples



A. Torralba, R. Fergus, W.T.Freeman. 2008

Summary

- With billions of images on the web, it's often possible to find a close nearest neighbor
- In such cases, we can shortcut hard problems by "looking up" the answer, stealing the labels from our nearest neighbor. For example, simple (or learned) associations can be used to synthesize background regions, colorize, or recognize objects
- But we can't really "brute force" computer vision. Still, it's nice to get an intuition for the size of "image space".