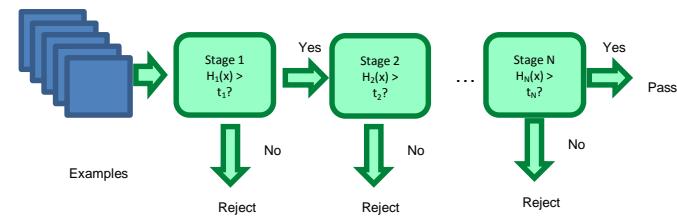
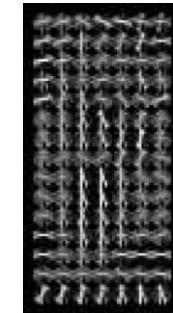


Object Detection Wrapup

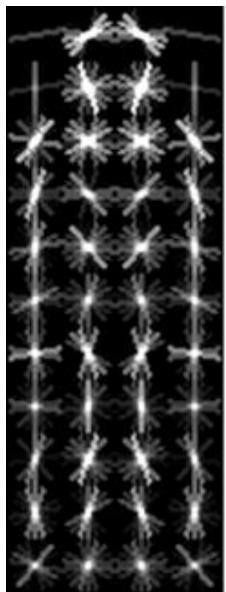
Things to remember

- Sliding window for search
- Features based on differences of intensity (gradient, wavelet, etc.)
 - Excellent results require careful feature design
- Boosting for feature selection
- Integral images, cascade for speed
- Bootstrapping to deal with many, many negative examples

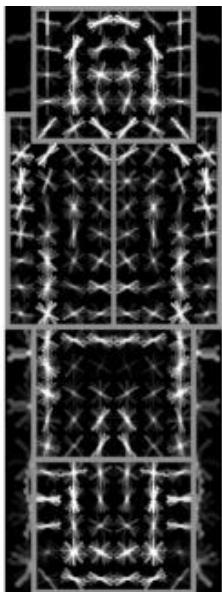


Discriminative part-based models

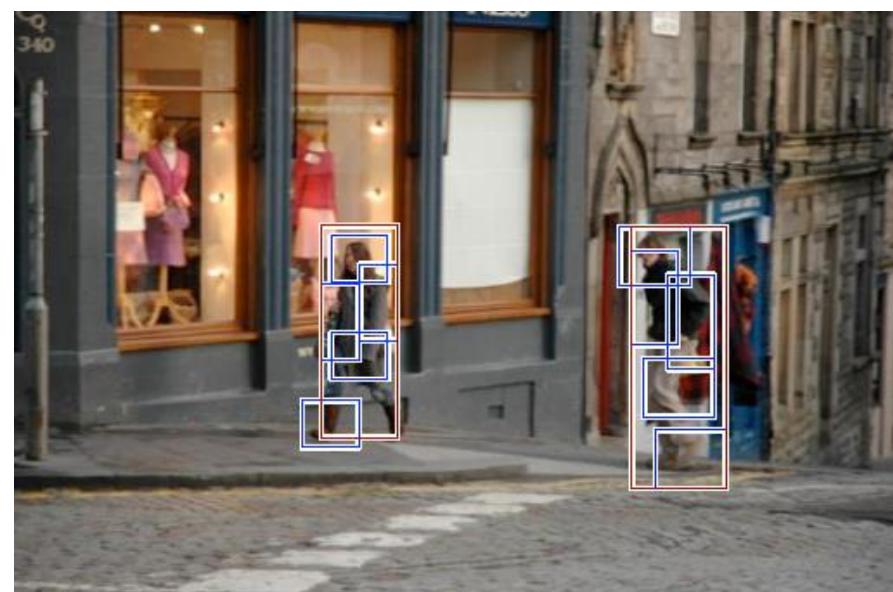
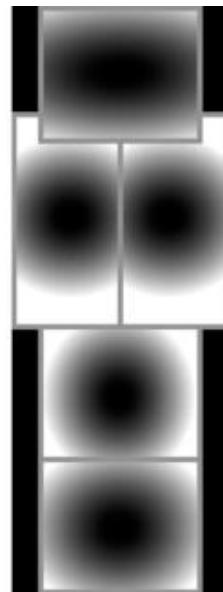
Root
filter



Part
filters



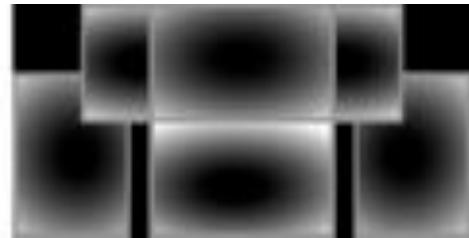
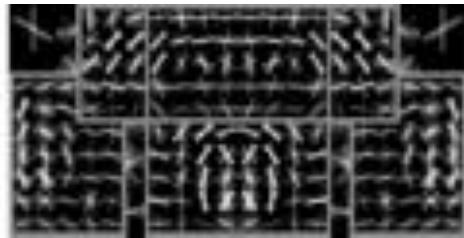
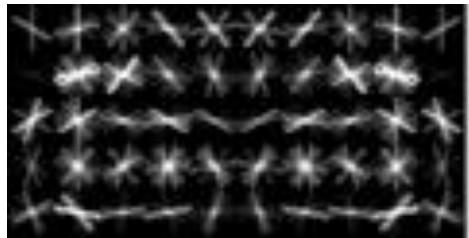
Deformation
weights



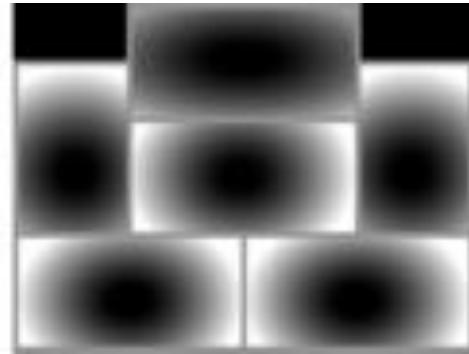
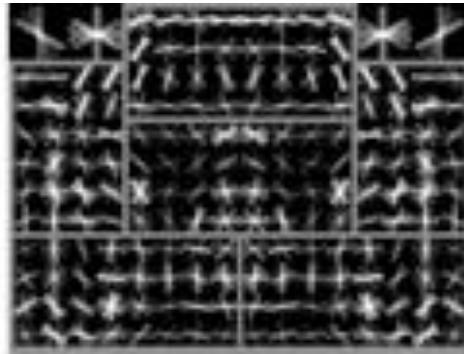
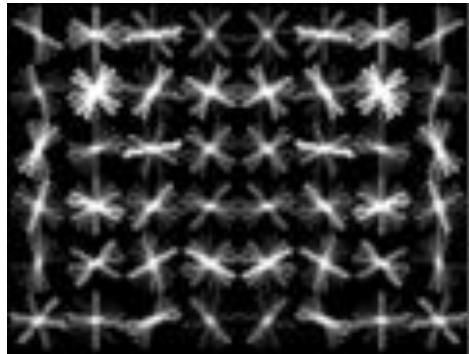
P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan, [Object Detection with Discriminatively Trained Part Based Models](#), PAMI 32(9), 2010

Car model

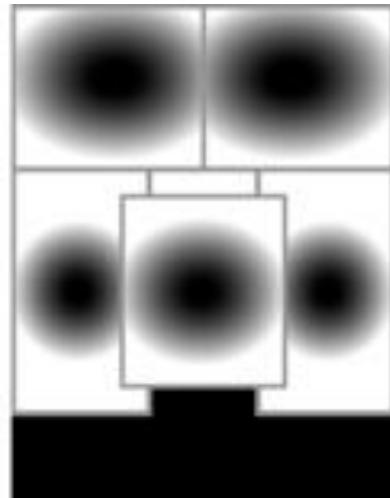
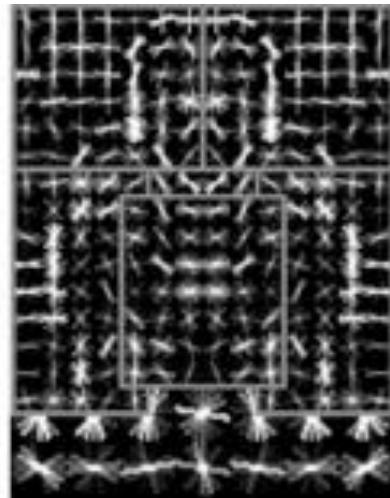
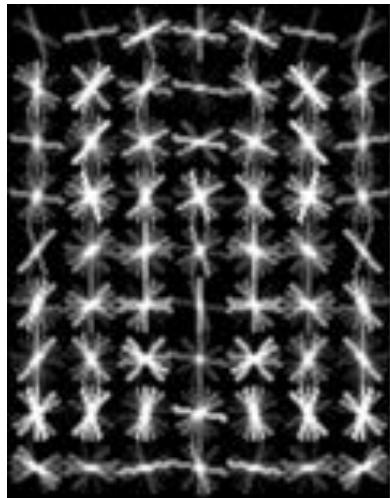
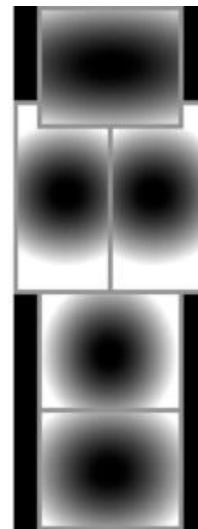
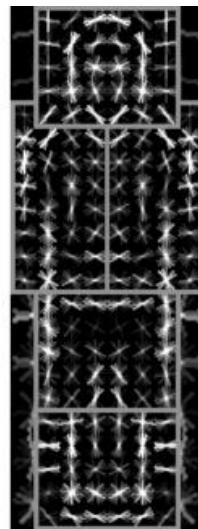
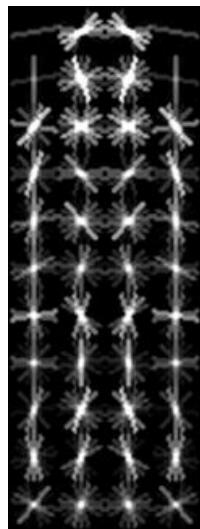
Component 1



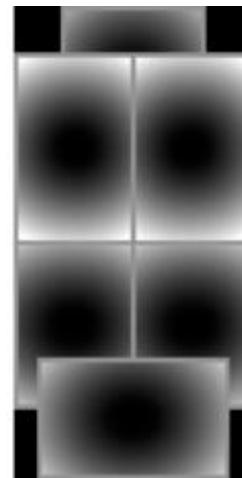
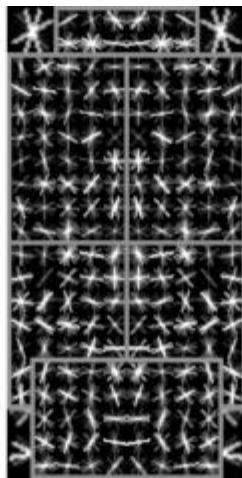
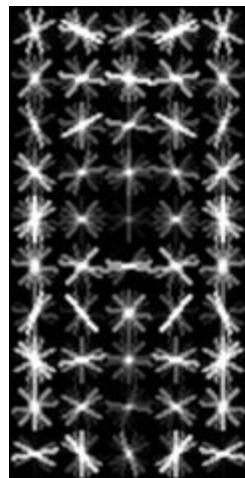
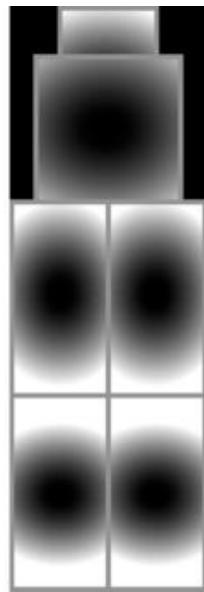
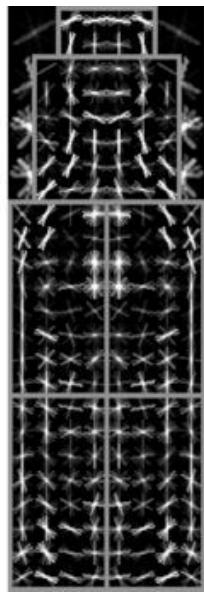
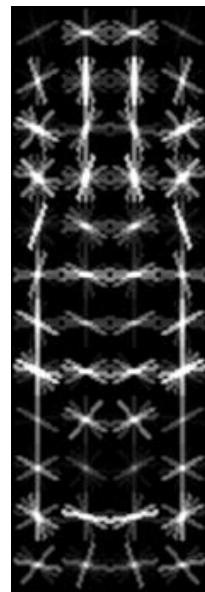
Component 2



Person model

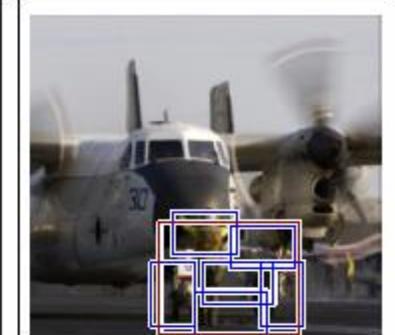
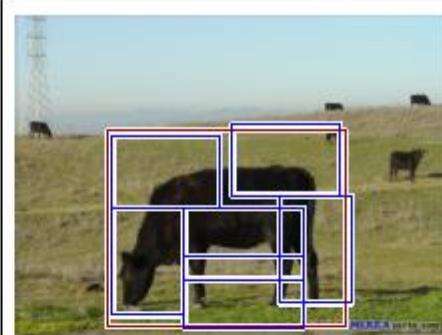
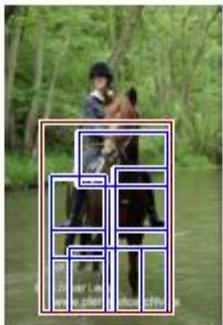
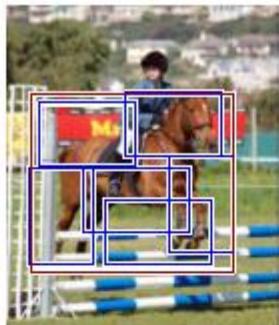
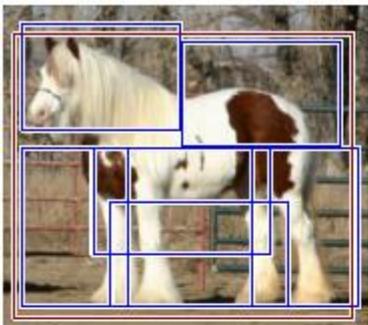


Bottle model

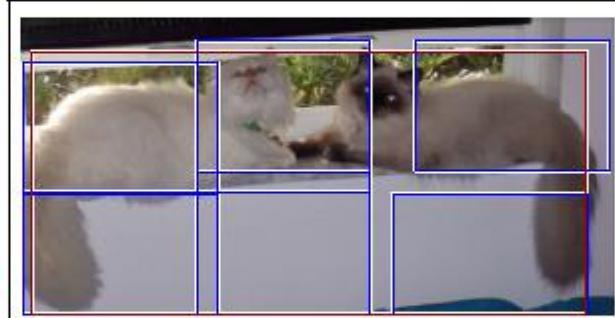
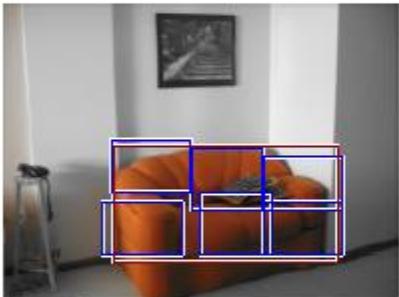


More detections

horse



sofa

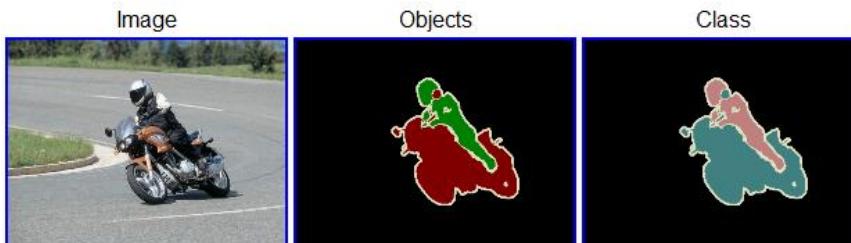


bottle



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 written guidelines
 - High quality (?)

Examples

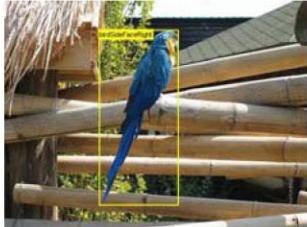
Aeroplane



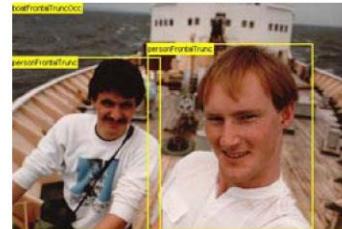
Bicycle



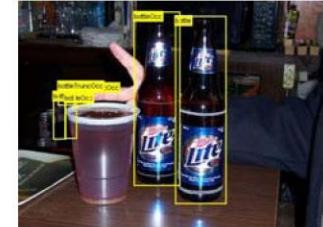
Bird



Boat



Bottle



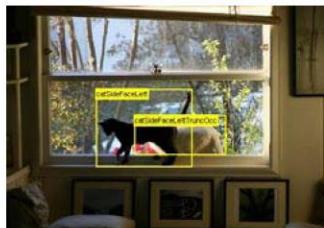
Bus



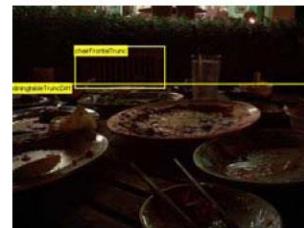
Car



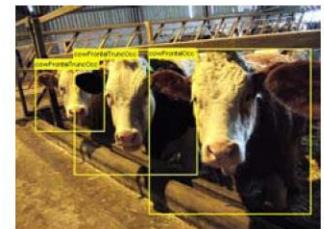
Cat



Chair



Cow



Examples

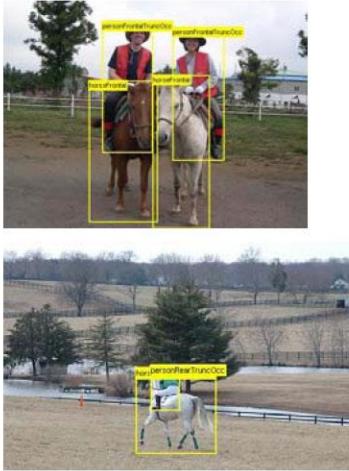
Dining Table



Dog



Horse



Motorbike



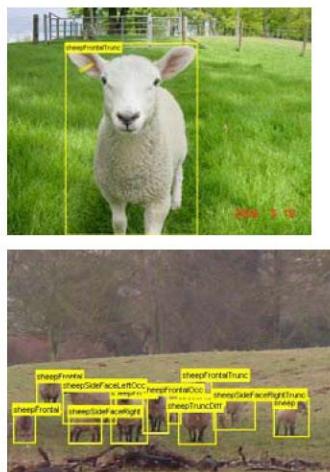
Person



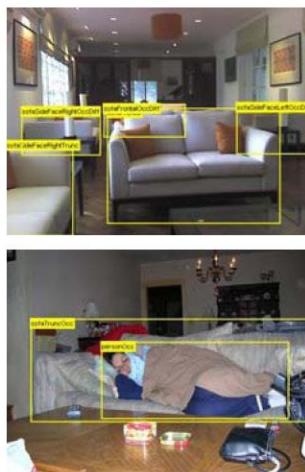
Potted Plant



Sheep



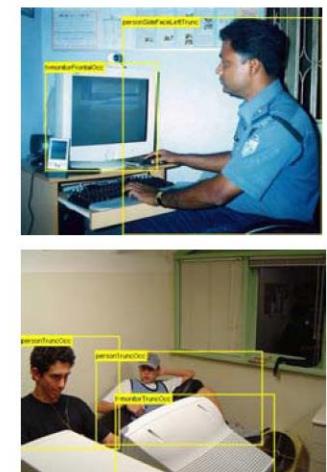
Sofa



Train



TV/Monitor



Classification Challenge

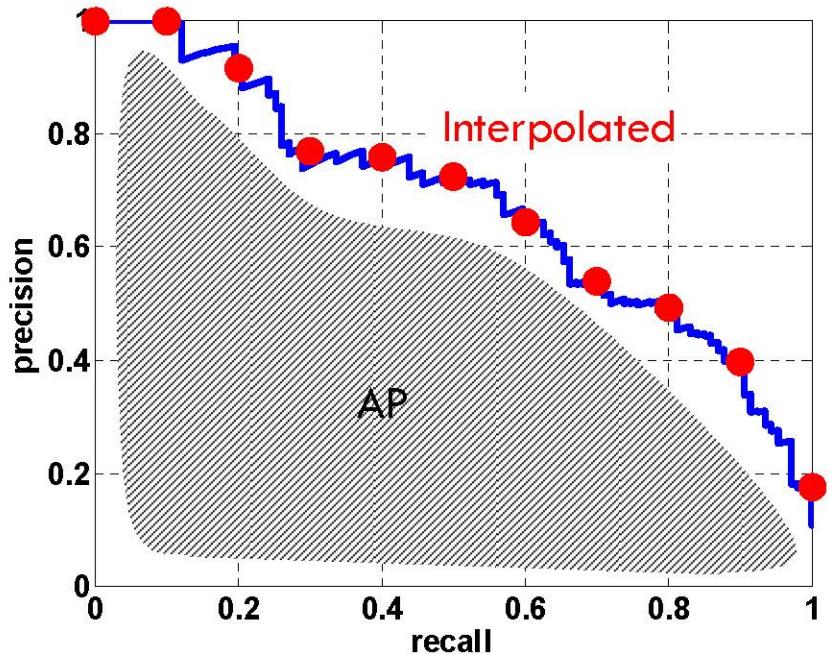
- Predict whether at least one object of a given class is present in an image



is there a cat?

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

Participation

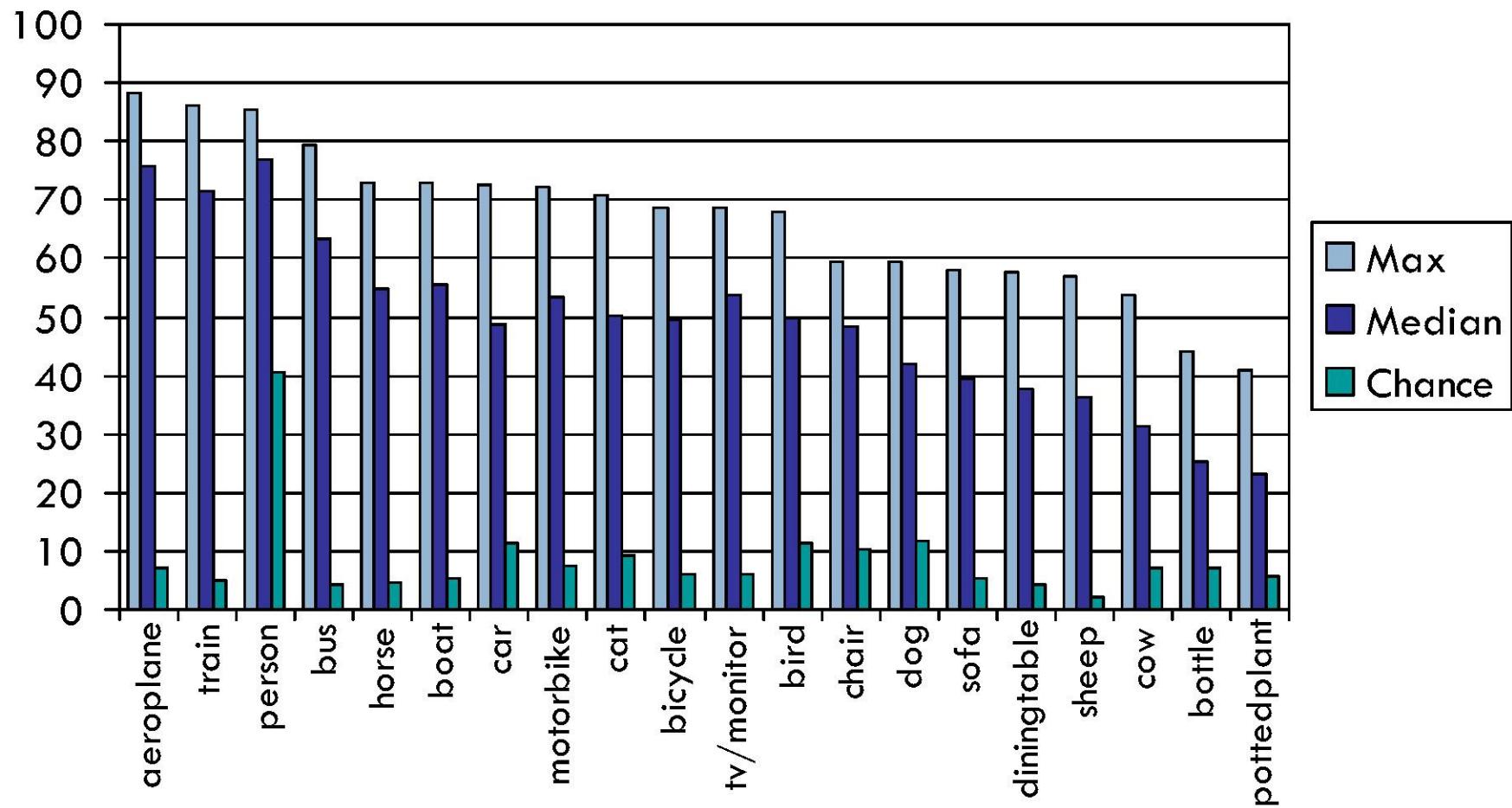
- 48 Methods, 20 Groups

Results: AP by Method and Class

	aero plane	bicycle	bird	boat	bottle	bus	car	cat	chair	cow	dining table	dog	horse	motor bike	person	potted plant	sheep	sofa	train	tv/ monitor
CVC_FLAT	85.3	57.8	66.0	66.1	36.2	70.6	60.6	63.5	55.1	44.6	53.4	49.1	64.4	66.8	84.8	37.4	44.1	47.9	81.9	67.5
CVC_FLAT-HOG-ESS	86.3	60.7	66.4	65.3	41.0	71.7	64.7	63.9	55.5	40.1	51.3	45.9	65.2	68.9	85.0	40.8	49.0	49.1	81.8	68.6
CVC_PLUS	86.6	58.4	66.7	67.3	34.8	70.4	60.0	64.2	52.5	43.0	50.8	46.5	64.1	66.8	84.4	37.5	45.1	45.4	82.1	67.0
FIRSTNIKON_AVGSRKDA	83.3	59.3	62.7	65.3	30.2	71.6	58.2	62.2	54.3	40.7	49.2	50.0	66.6	62.9	83.3	34.2	48.2	46.1	83.4	65.5
FIRSTNIKON_AVGSVM	83.8	58.2	62.6	65.2	32.0	69.8	57.7	61.1	54.5	44.0	50.3	49.6	64.6	61.7	83.2	33.4	46.5	48.0	81.6	65.3
FIRSTNIKON_BOOSTSRKDA	83.0	59.2	61.4	64.6	33.2	71.1	57.5	61.0	54.8	40.7	48.3	50.0	65.5	63.4	82.8	32.8	47.0	47.1	83.3	64.6
FIRSTNIKON_BOOSTSVMS	83.5	56.8	61.8	65.5	33.2	69.7	57.3	60.5	54.6	43.1	48.3	50.3	64.3	62.4	82.3	32.9	46.9	48.4	82.0	64.2
LEAR_CHI-SVM-MULT-LOC	79.5	55.5	54.5	63.9	43.7	70.3	66.4	56.5	54.4	38.8	44.1	46.2	58.5	64.2	82.2	39.1	41.3	39.8	73.6	66.2
NECUIUC_CDCV	88.1	68.0	68.0	72.5	41.0	78.9	70.4	70.4	58.1	53.4	55.7	59.3	73.1	71.3	84.5	32.3	53.3	56.7	86.0	66.8
NECUIUC_CLS-DTCT	88.0	68.6	67.9	72.9	44.2	79.5	72.5	70.8	59.5	53.6	57.5	59.0	72.6	72.3	85.3	36.6	56.9	57.9	85.9	68.0
NECUIUC_LL-CDCV	87.1	67.4	65.8	72.3	40.9	78.3	69.7	69.7	58.5	50.1	55.1	56.3	71.8	70.8	84.1	31.4	51.5	55.1	84.7	65.2
NECUIUC_LN-CDCV	87.7	67.8	68.1	71.1	39.1	78.5	70.6	70.7	57.4	51.7	53.3	59.2	71.6	70.6	84.0	30.9	51.7	55.9	85.9	66.7
UVASURREY_BASELINE	84.1	59.2	62.7	65.4	35.7	70.6	59.8	61.3	56.7	45.3	52.4	50.6	66.1	66.6	83.7	34.8	47.2	47.7	80.8	65.9
UVASURREY_MKFDA+BOW	84.7	63.9	66.1	67.3	37.9	74.1	63.2	64.0	57.1	46.2	54.7	53.5	68.1	70.6	85.2	38.5	47.2	49.3	83.2	68.1
UVASURREY_TUNECOLORKERNELSEL	85.0	62.8	65.1	66.5	37.6	73.5	62.1	62.0	57.4	45.1	54.5	52.5	67.7	69.8	84.8	39.1	46.8	49.9	82.9	68.1
UVASURREY_TUNECOLORSPECDA	84.6	62.4	65.6	67.2	39.4	74.0	63.4	62.8	56.7	43.8	54.7	52.7	67.3	70.6	85.0	38.8	46.9	50.0	82.2	66.2

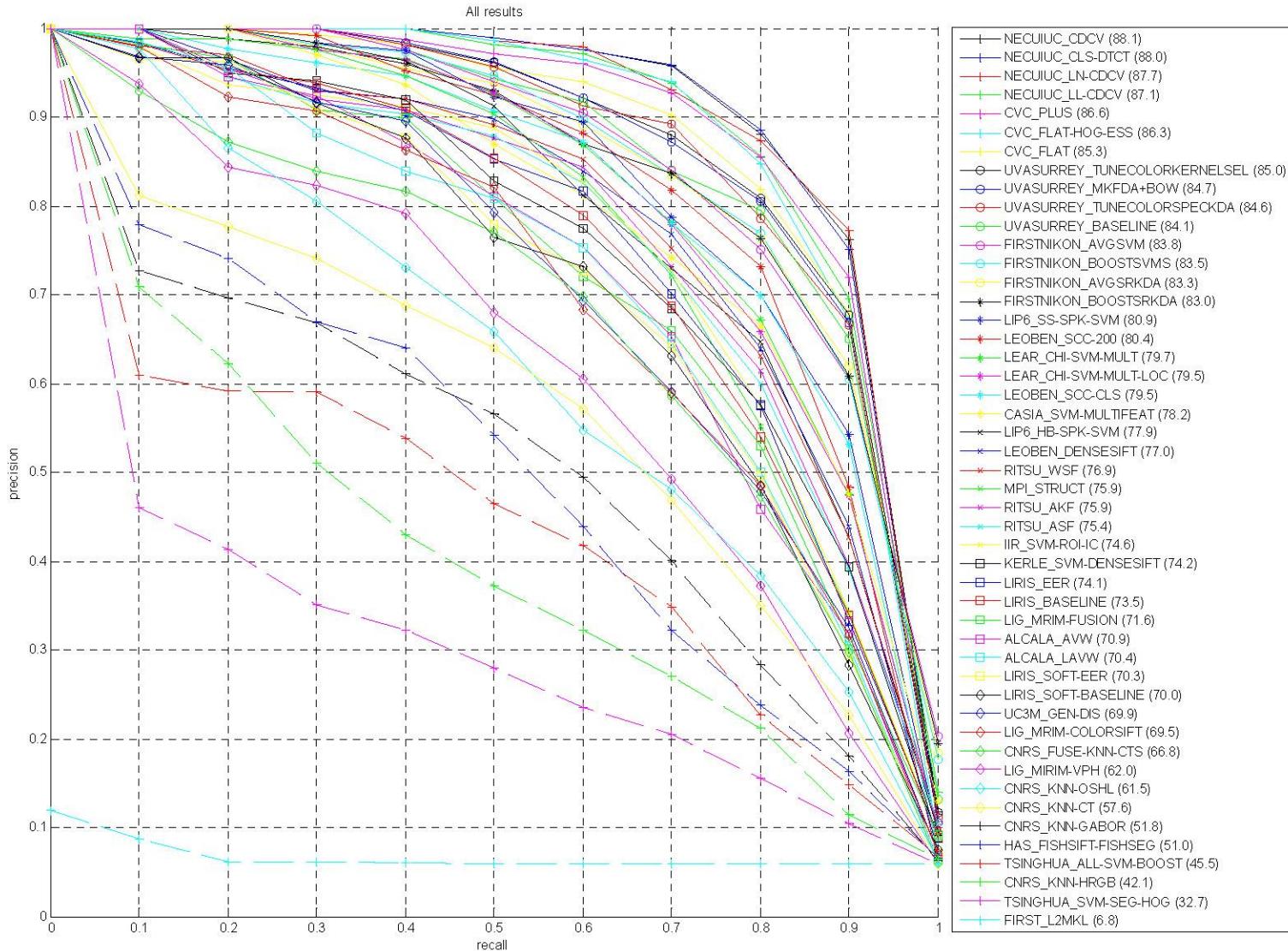
- Only methods in 1st, 2nd or 3rd place by group shown
- Groups: CVC, FIRST/Nikon, NEC/UIUC, UVA/Surrey

AP by Class

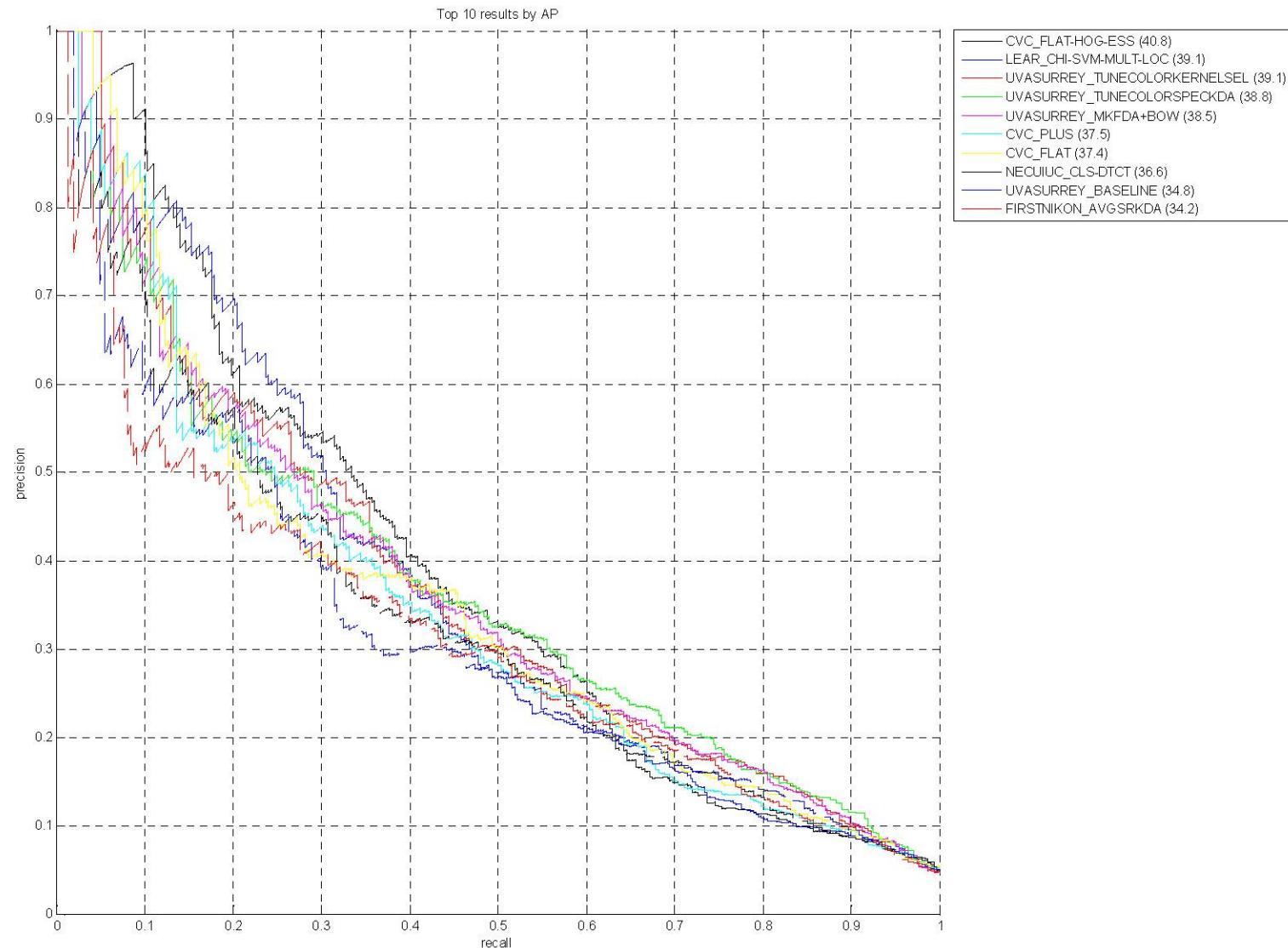


- Max AP: 88.1% (aeroplane) ... 40.8% (potted plant)

Precision/Recall: Aeroplane (All)



Precision/Recall: Potted plant (Top 10 by AP)



Ranked Images: Aeroplane

- Class images:
Highest ranked



Ranked Images: Chair

- Class images:
Highest ranked



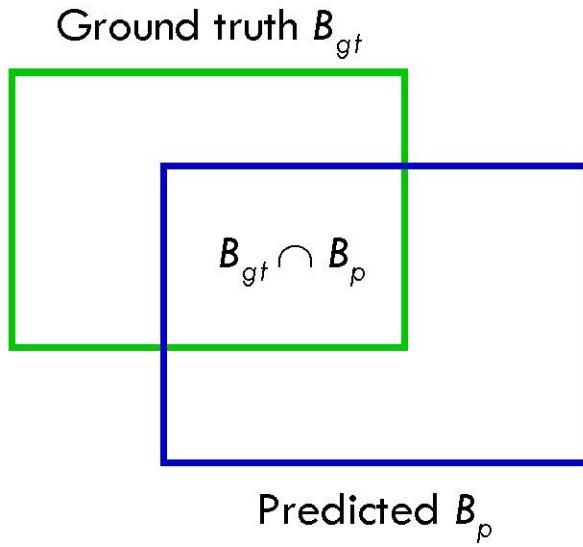
Detection Challenge

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



Evaluating Bounding Boxes

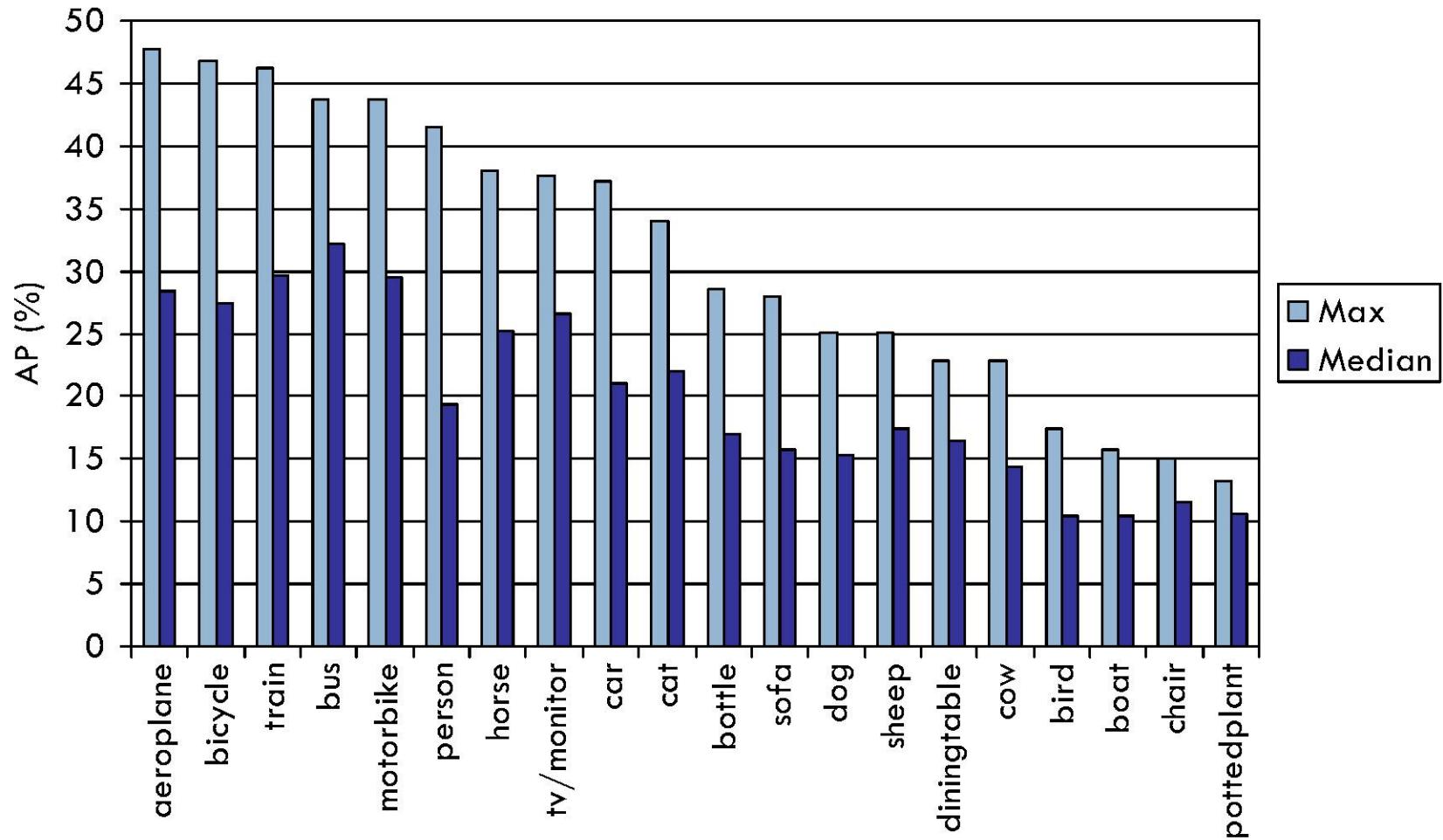
- Area of Overlap (AO) Measure



$$AO(B_{gt}, B_p) = \frac{|B_{gt} \cap B_p|}{|B_{gt} \cup B_p|}$$

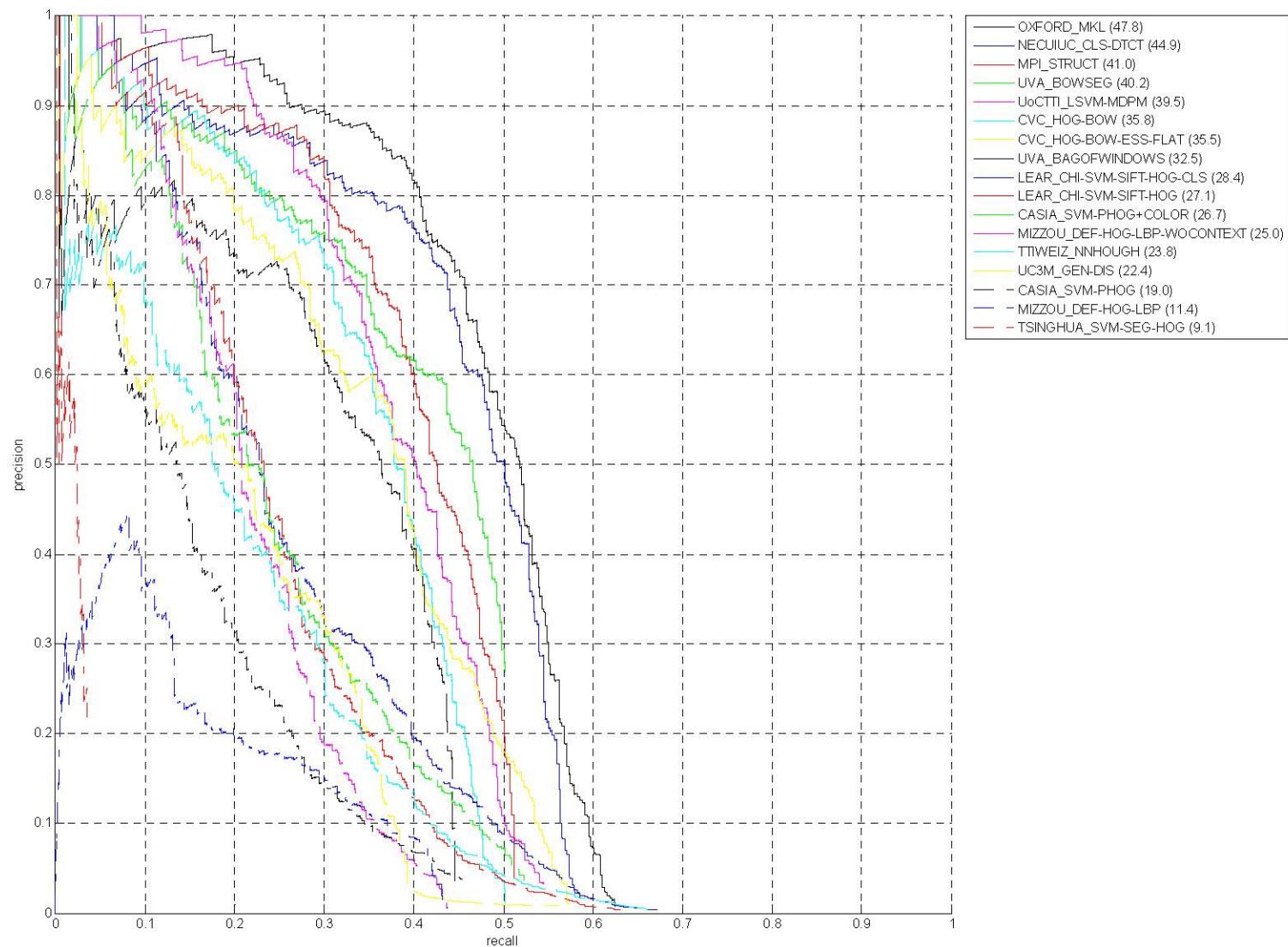
- Need to define a threshold t such that $AO(B_{gt}, B_p)$ implies a correct detection: 50%

AP by Class

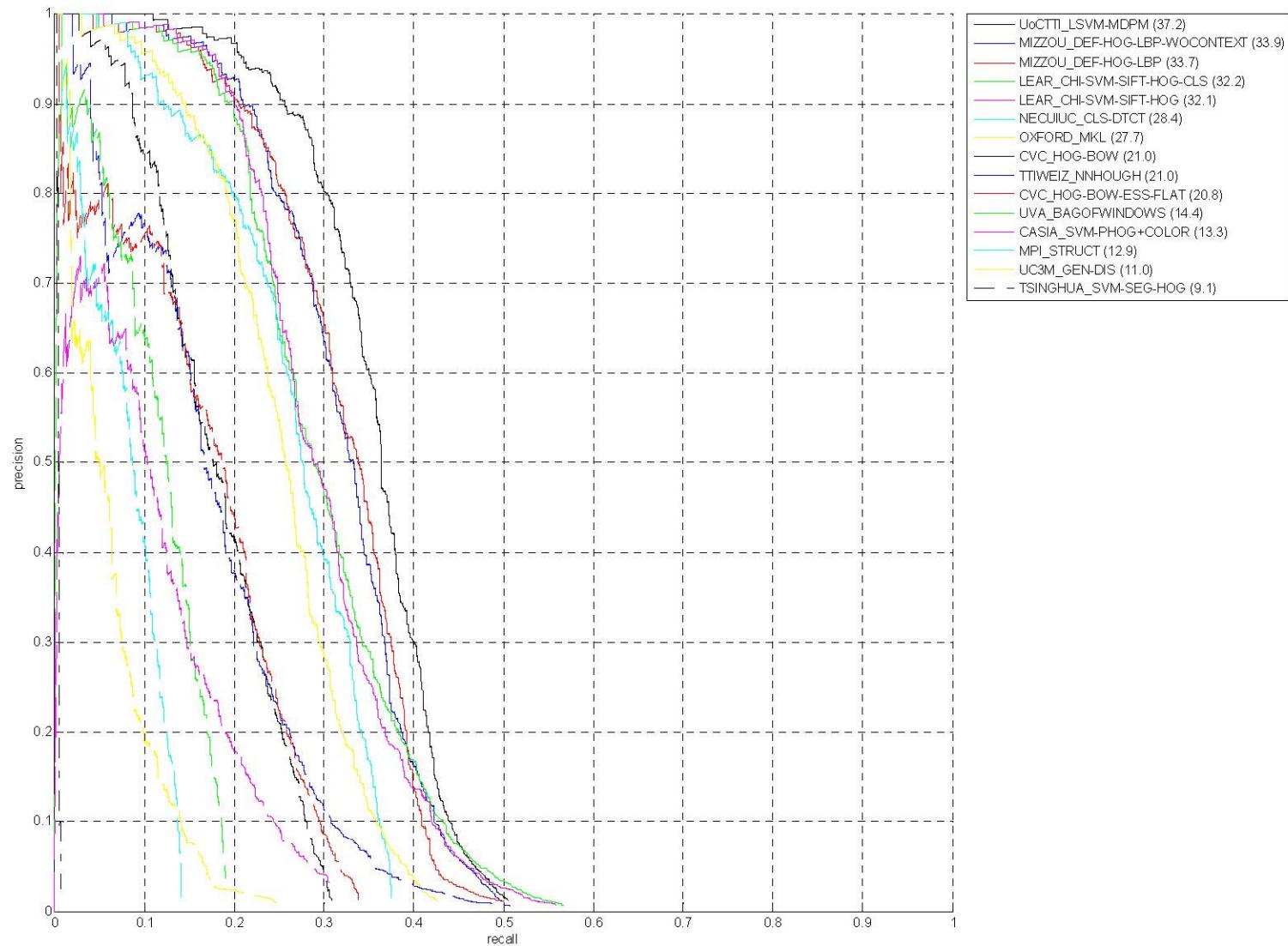


Chance essentially 0

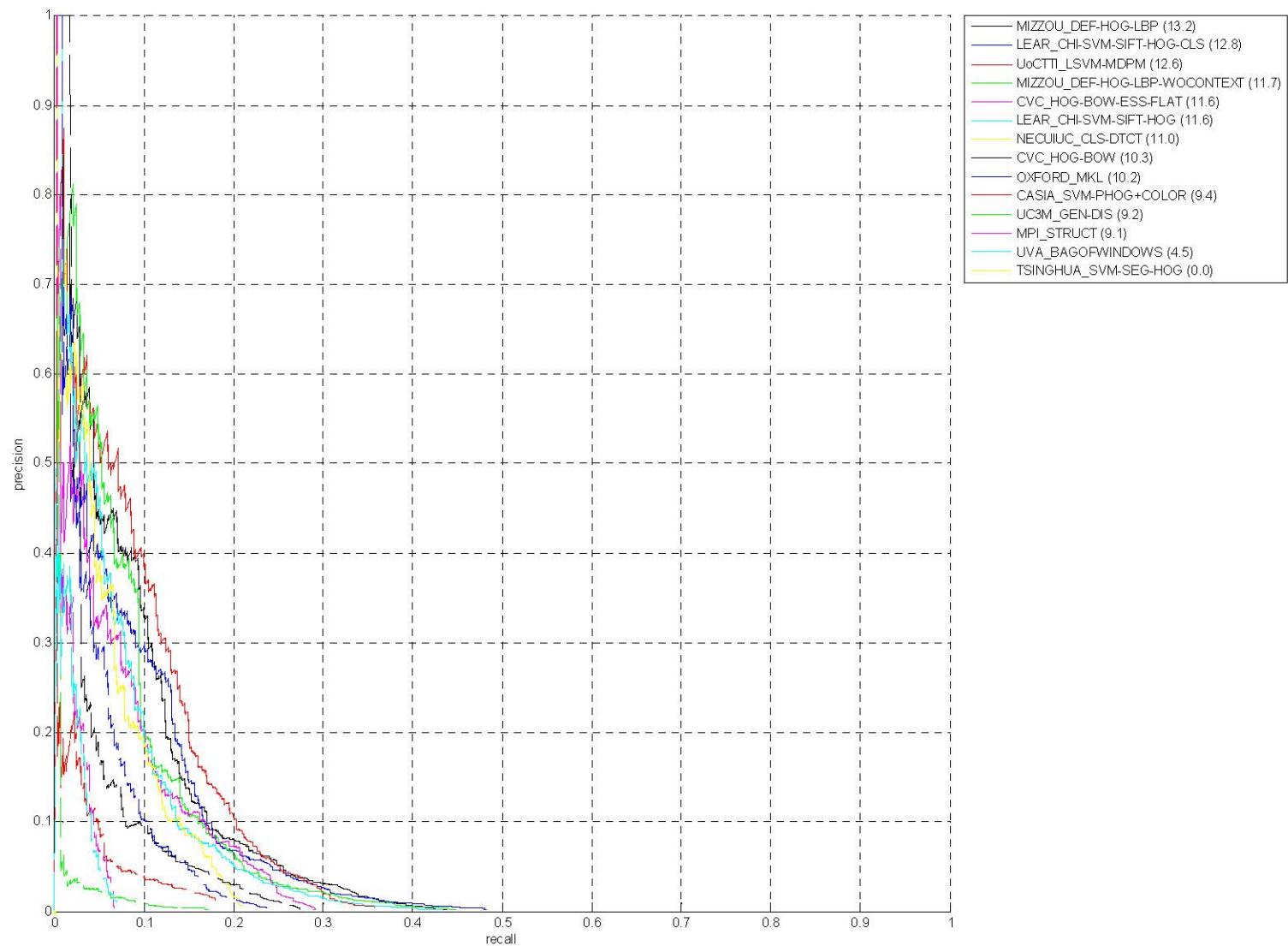
Precision/Recall - Aeroplane



Precision/Recall - Car

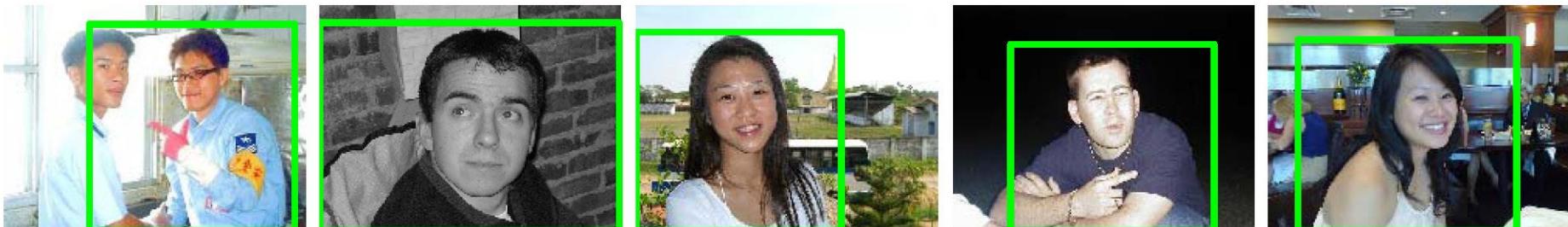


Precision/Recall – Potted plant



True Positives - Person

UoCTTI_LSVM-MDPM



MIZZOU_DEF-HOG-LBP



NECUIUC_CLS-DTCT



False Positives - Person

UoCTTI_LSVM-MDPM



MIZZOU_DEF-HOG-LBP

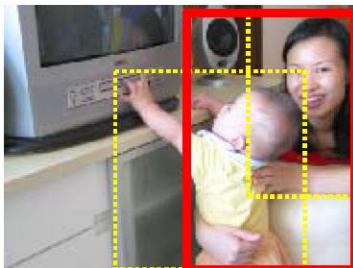


NECUIUC_CLS-DTCT

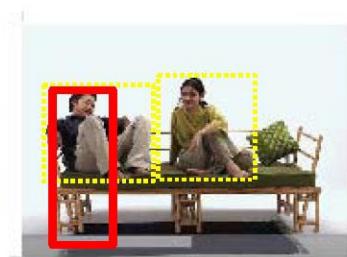


“Near Misses” - Person

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



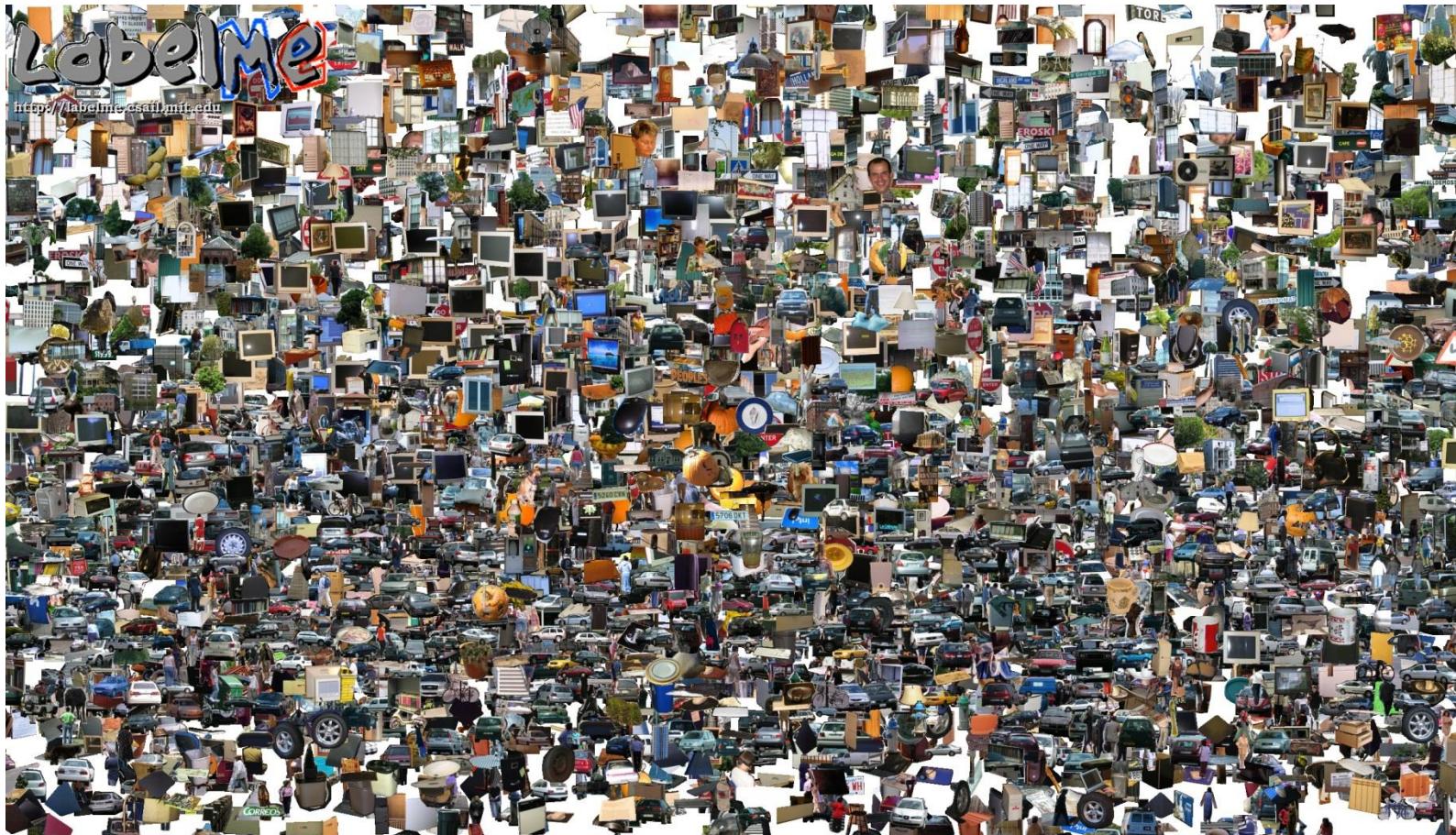
OXFORD_MKL



NECUIUC_CLS-DTCT



Opportunities of Scale



Computer Vision
James Hays

Computer Vision 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)
- Learning from Data
 - 1991 (EigenFaces) but late 90's to now especially

What has changed in the last decade?

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

Opportunities of Scale: Data-driven methods

- Scene completion (Today's class)
- Im2gps

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 in the data



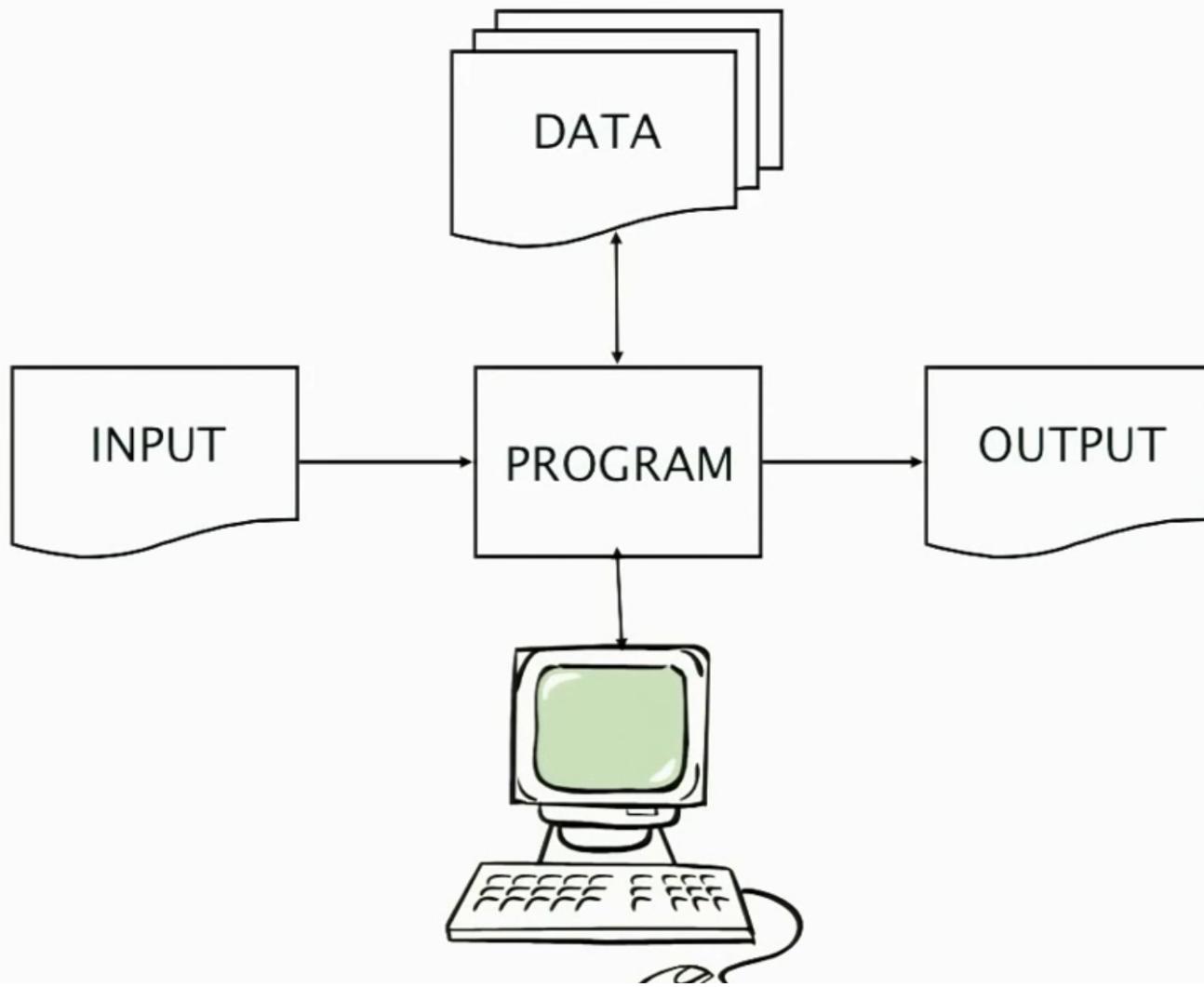
The Unreasonable Effectiveness of Data

Peter Norvig
Google



Peter Norvig

The Unreasonable Effectiveness of Data



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

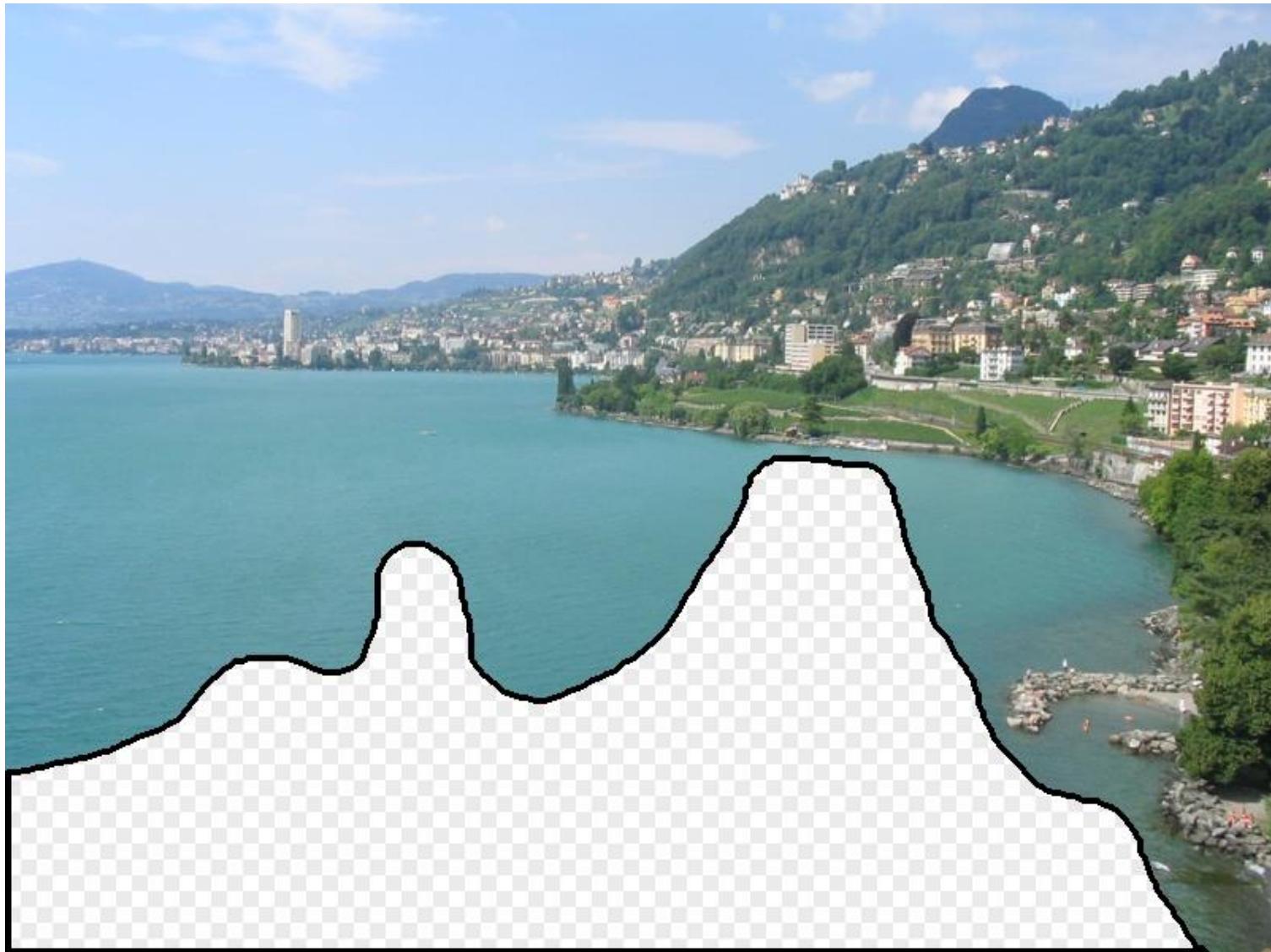
- Do we need computer vision systems to have strong AI-like reasoning about our world?
- What if invariance / generalization isn't actually the core difficulty of computer vision?
- What if we can perform high level reasoning with brute-force, data-driven algorithms?

Image Completion Example

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

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

What should the missing region contain?









Which is the original?



(a)



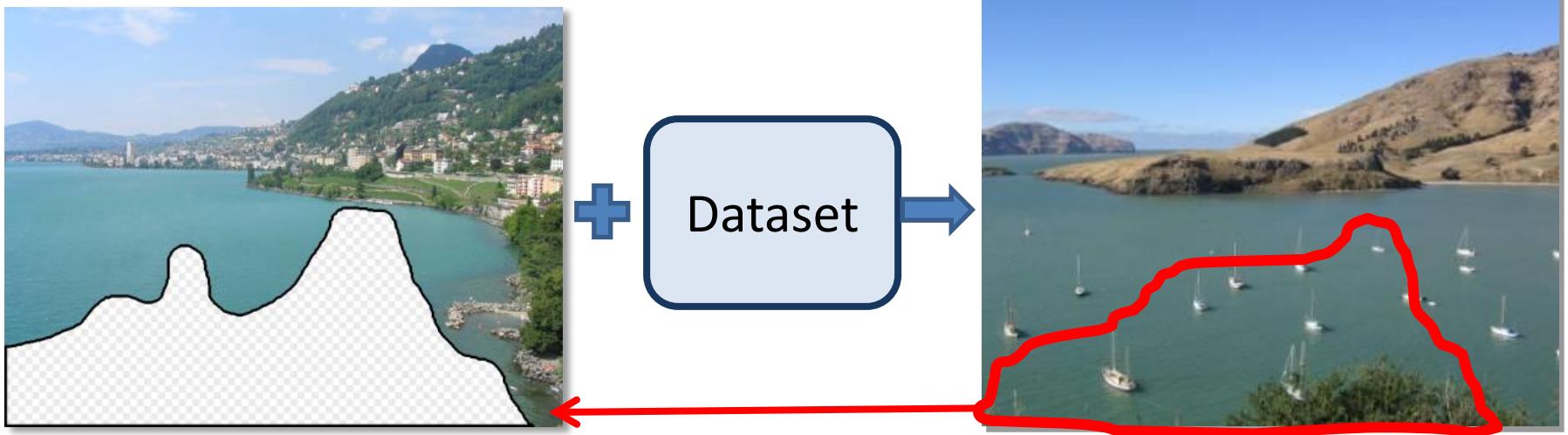
(b)



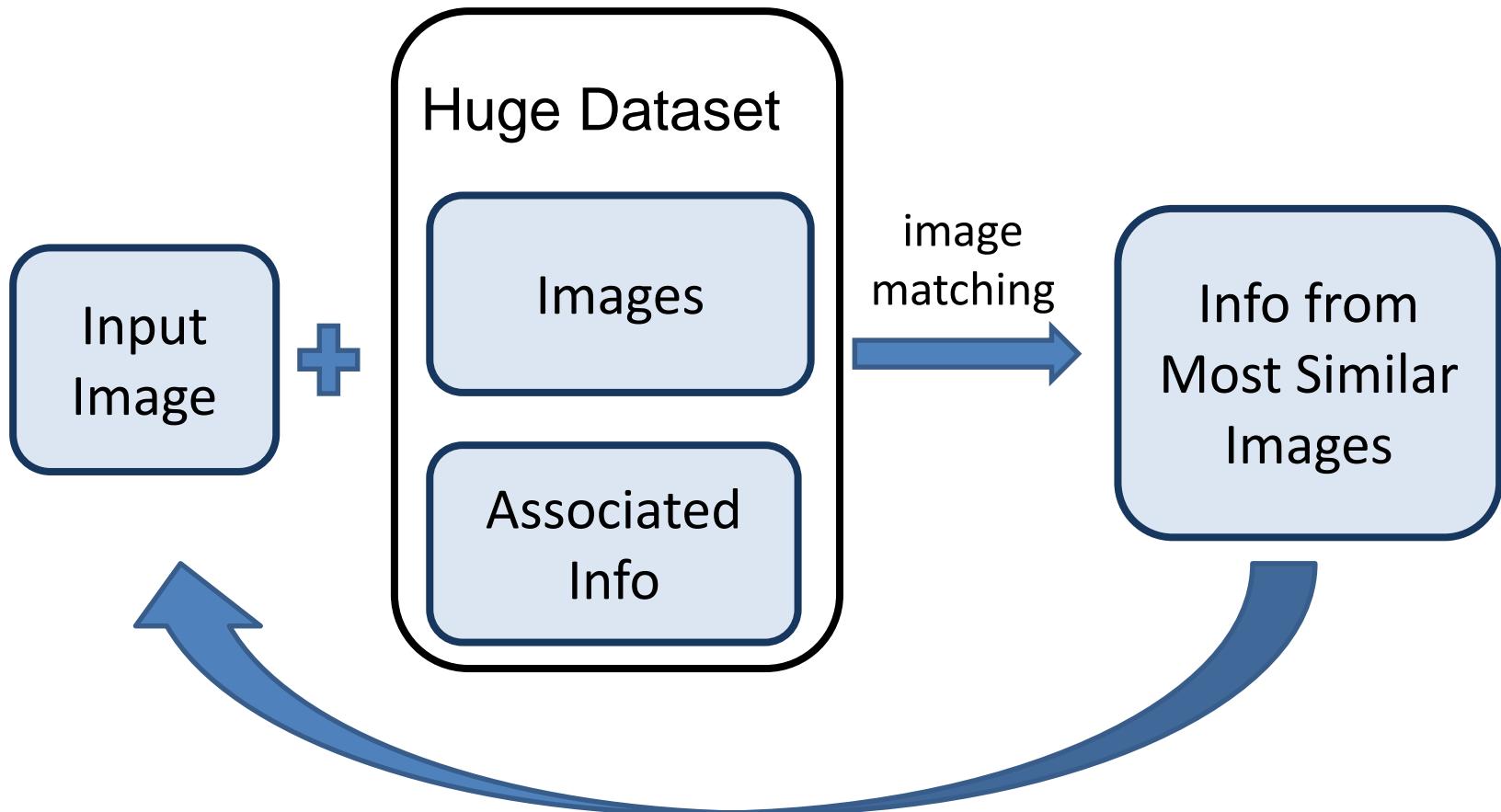
(c)

How it works

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

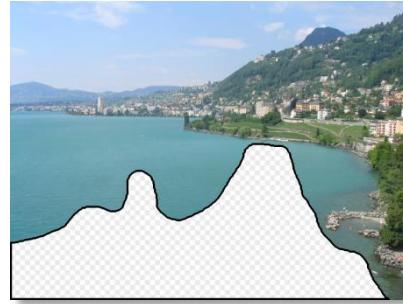


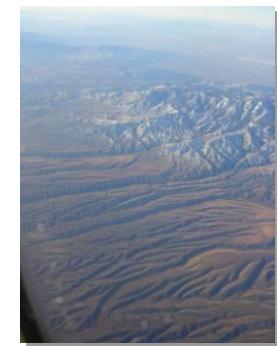
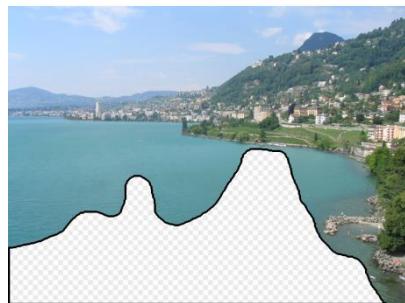
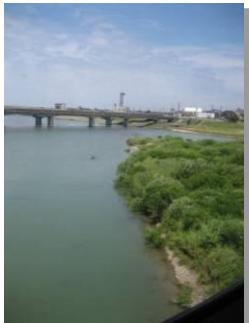
General Principal



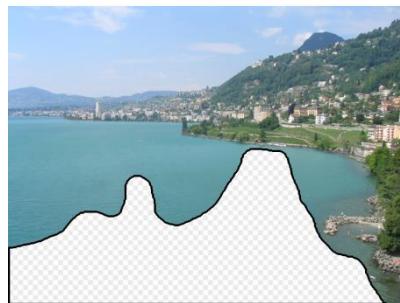
Hopefully, If you have enough images, the dataset will contain very similar images that you can find with simple matching methods.

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

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

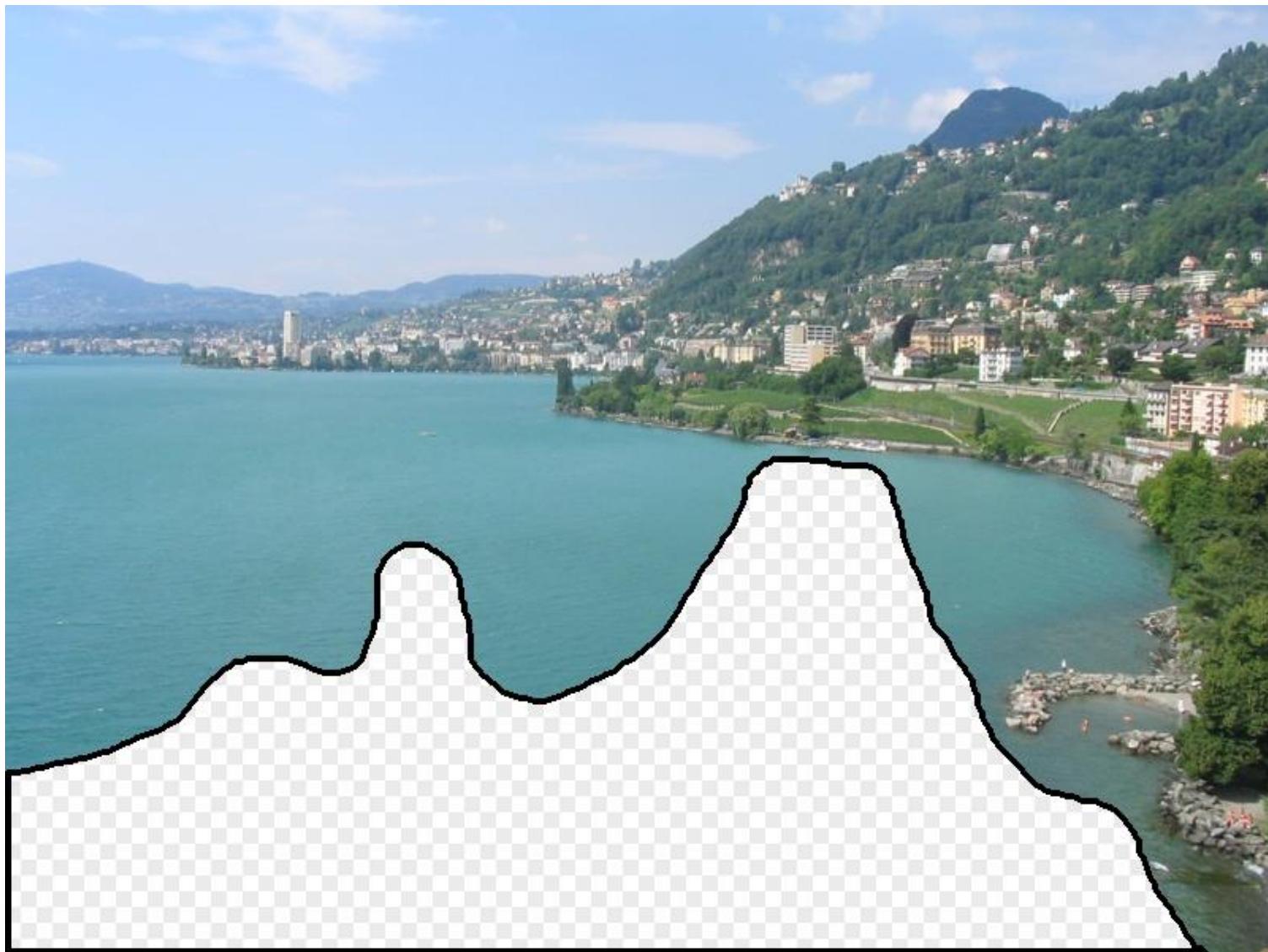
Image completion: how it works

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

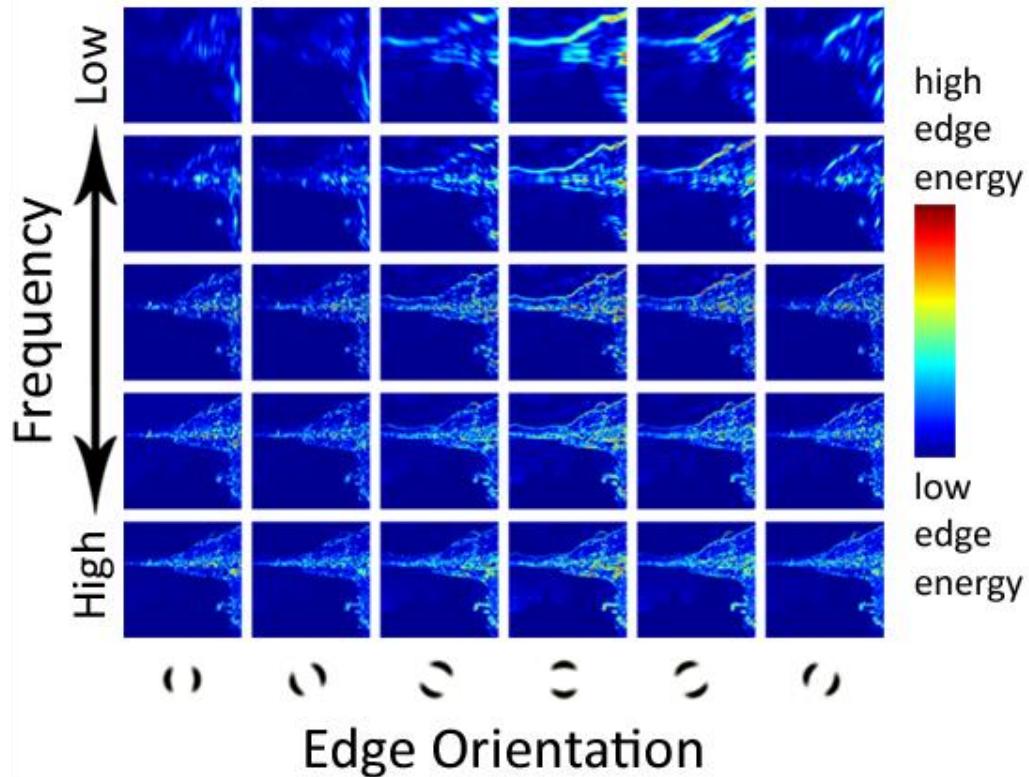
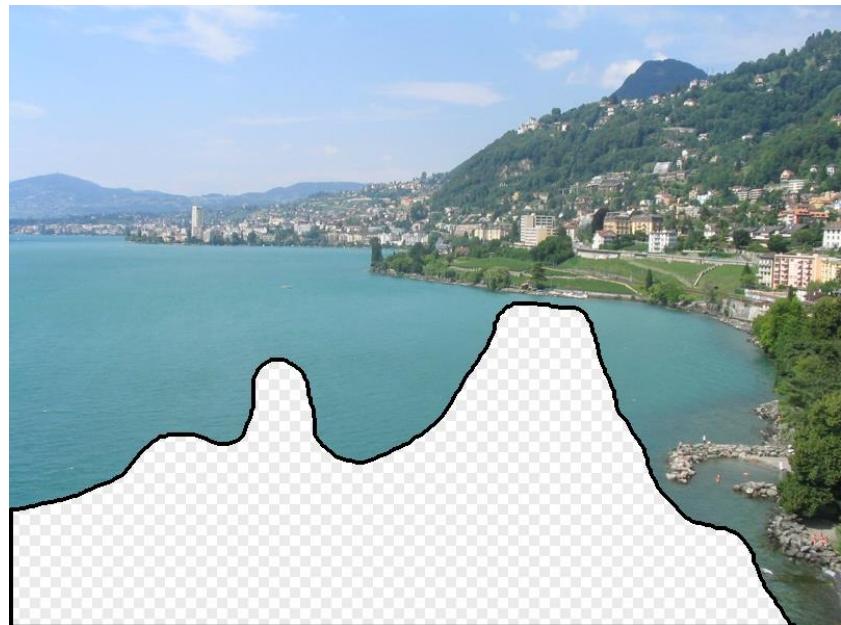
The Algorithm



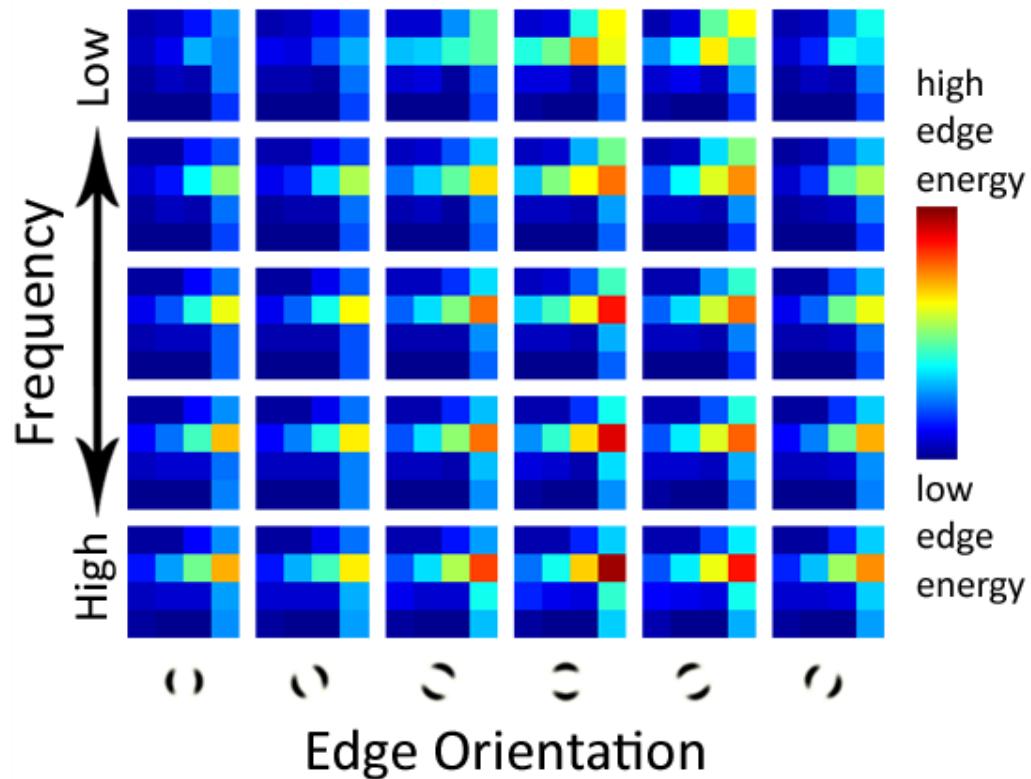
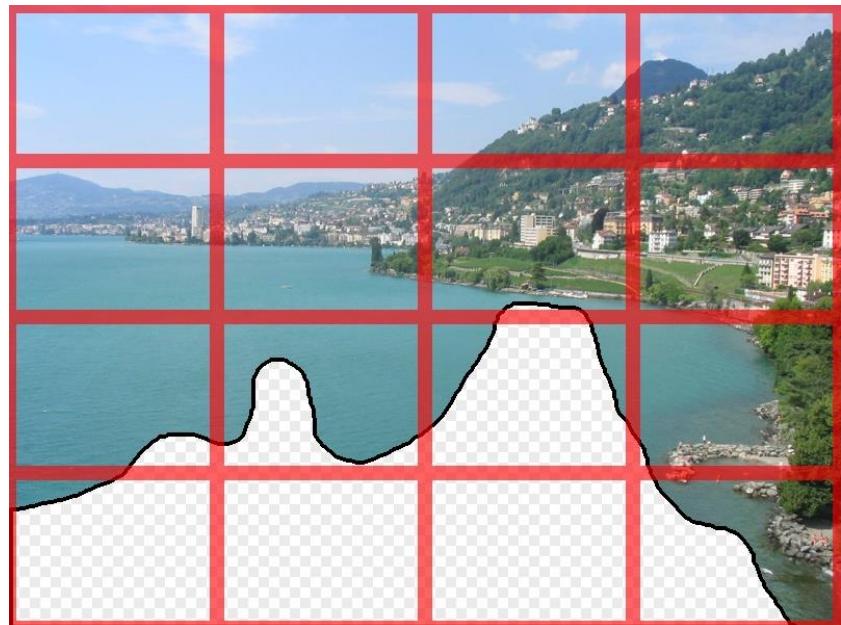
Scene Matching



Scene Descriptor

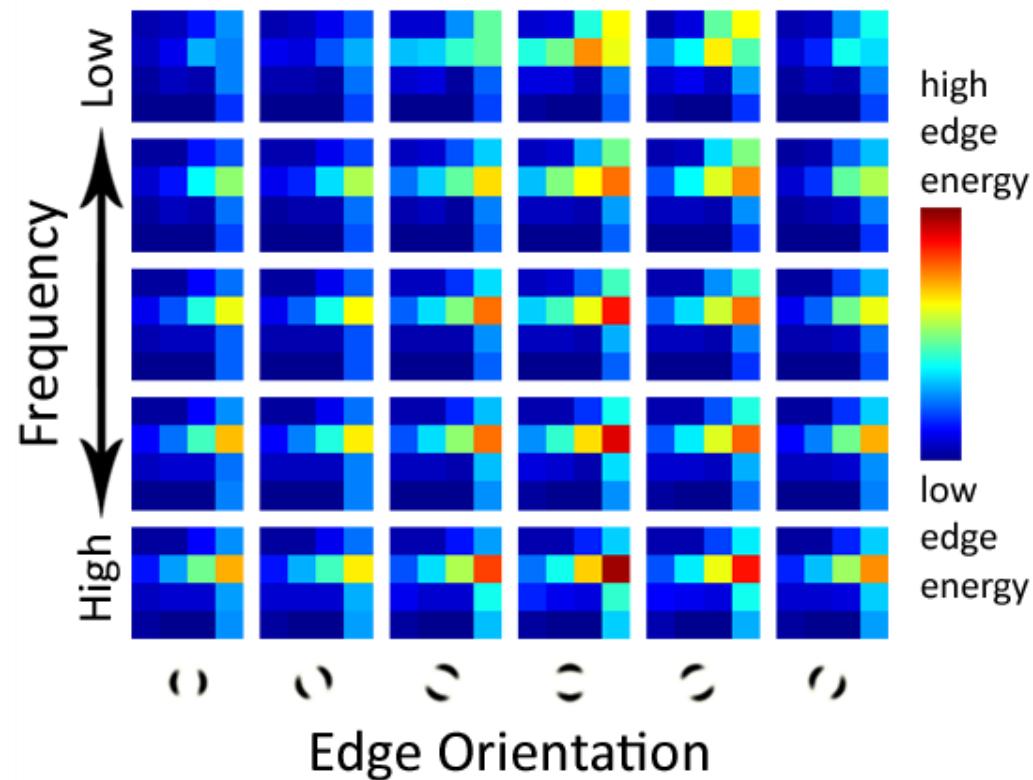
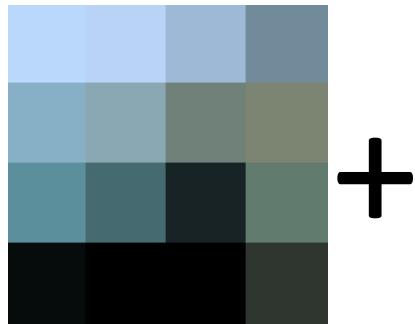


Scene Descriptor



Scene Gist Descriptor
(Oliva and Torralba 2001)

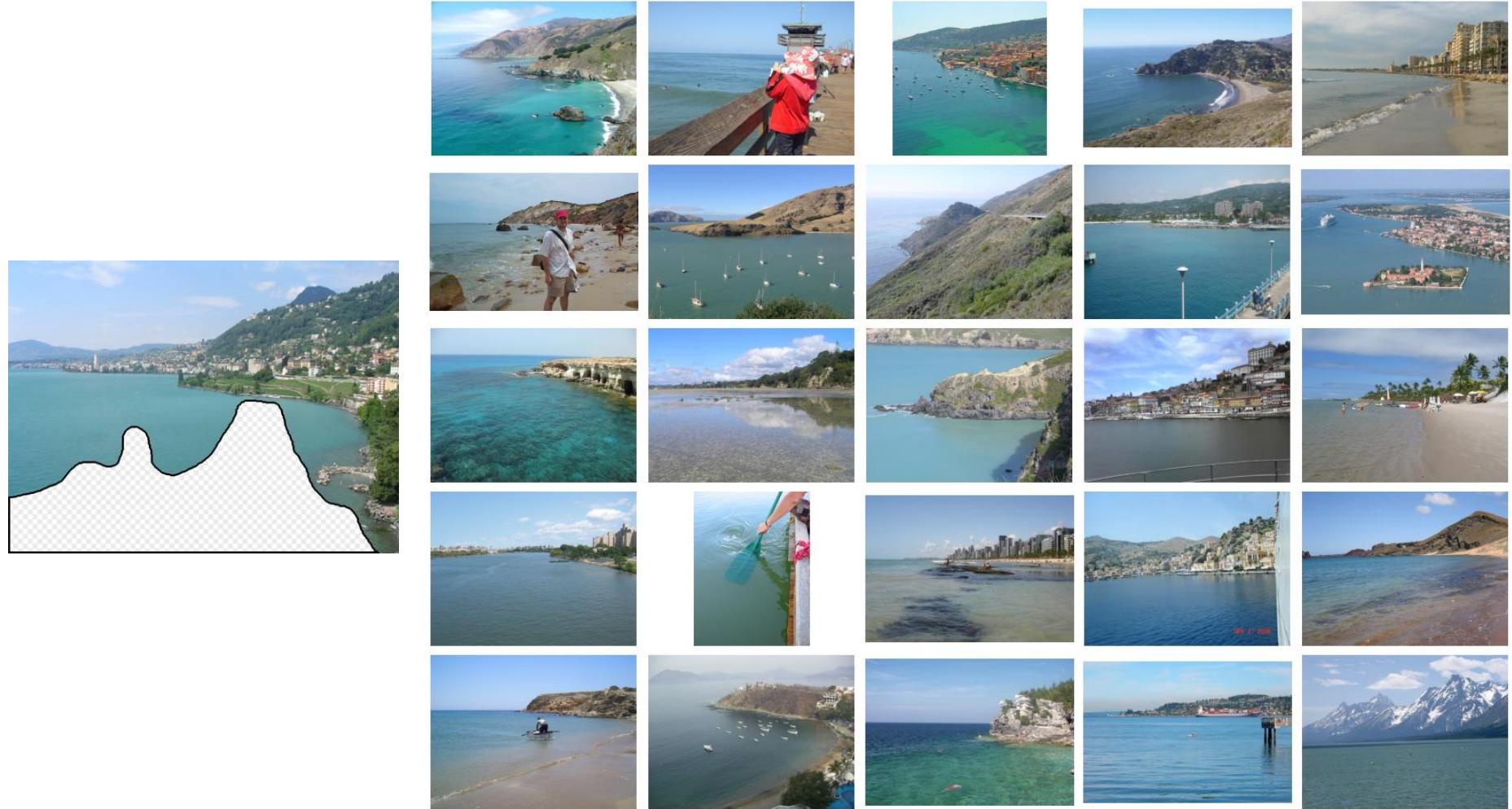
Scene Descriptor



Scene Gist Descriptor
(Oliva and Torralba 2001)

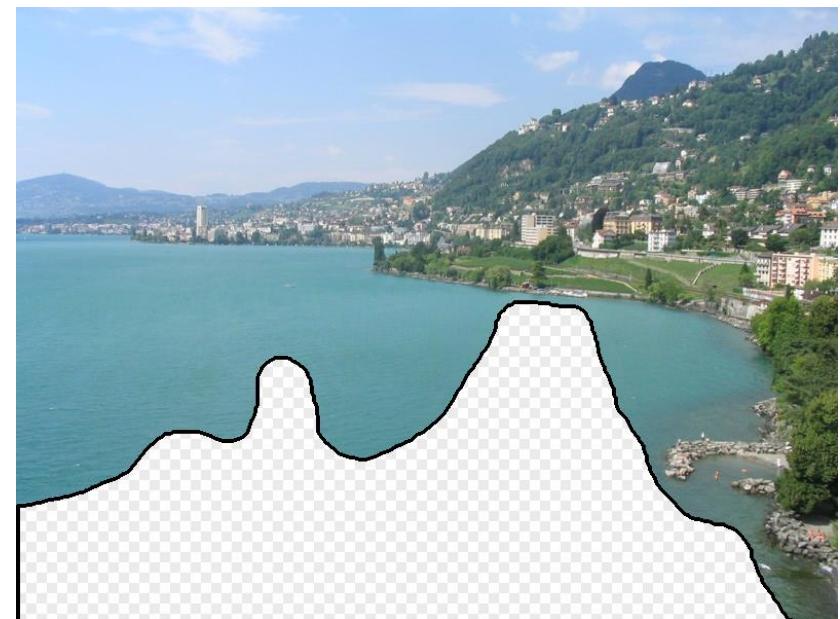
2 Million Flickr Images

The background of the image is a dense, uniform grid composed of numerous small, square thumbnail images. These thumbnails represent a vast collection of photographs from the Flickr platform, showing a wide variety of subjects and colors. The overall effect is a visual representation of the sheer volume and diversity of user-generated content available on the site.



... 200 total

Context Matching

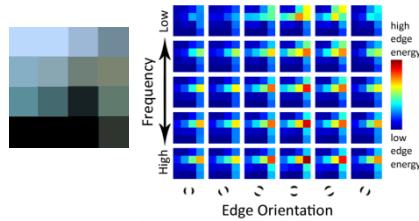




Graph cut + Poisson blending

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

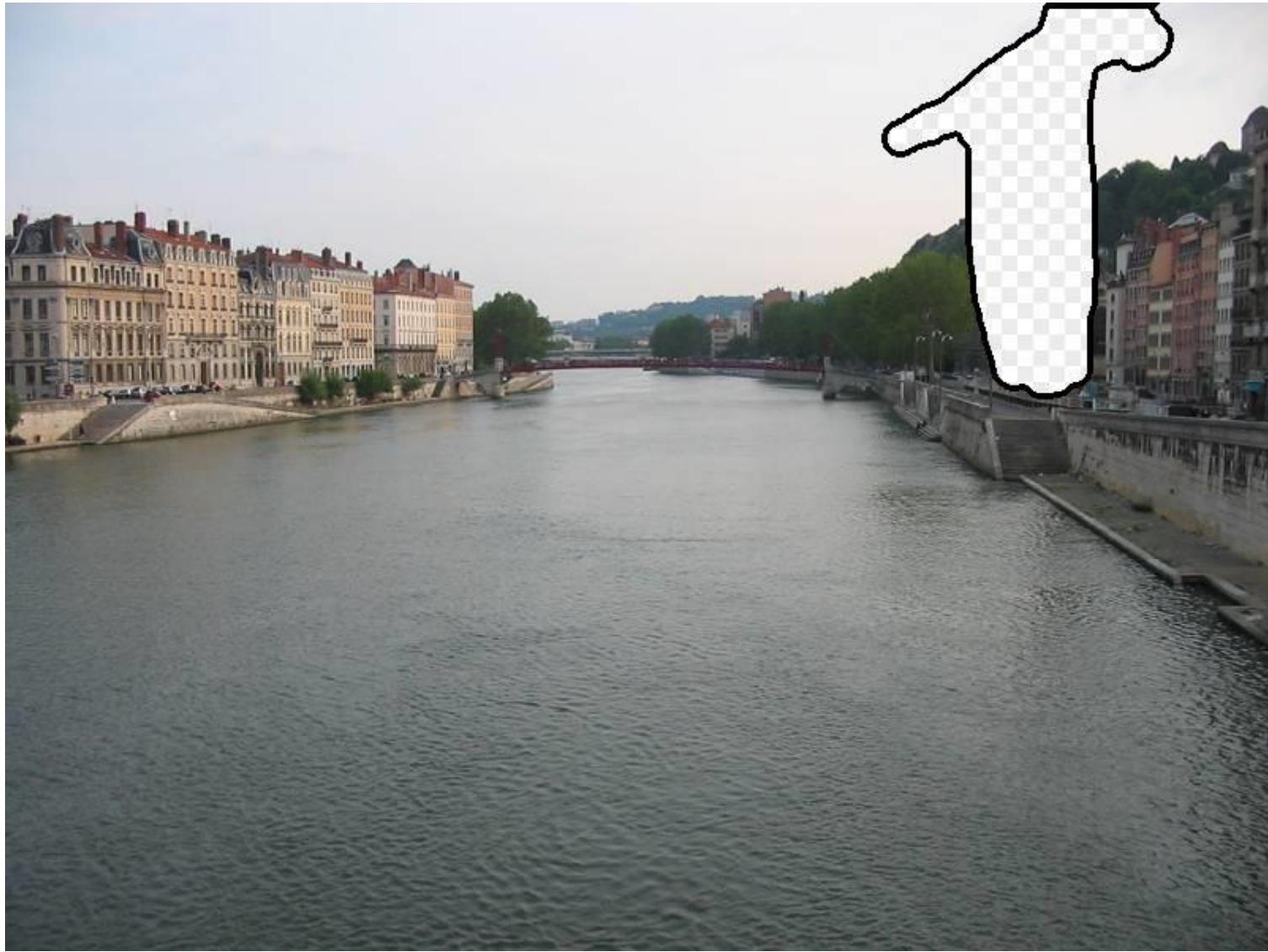




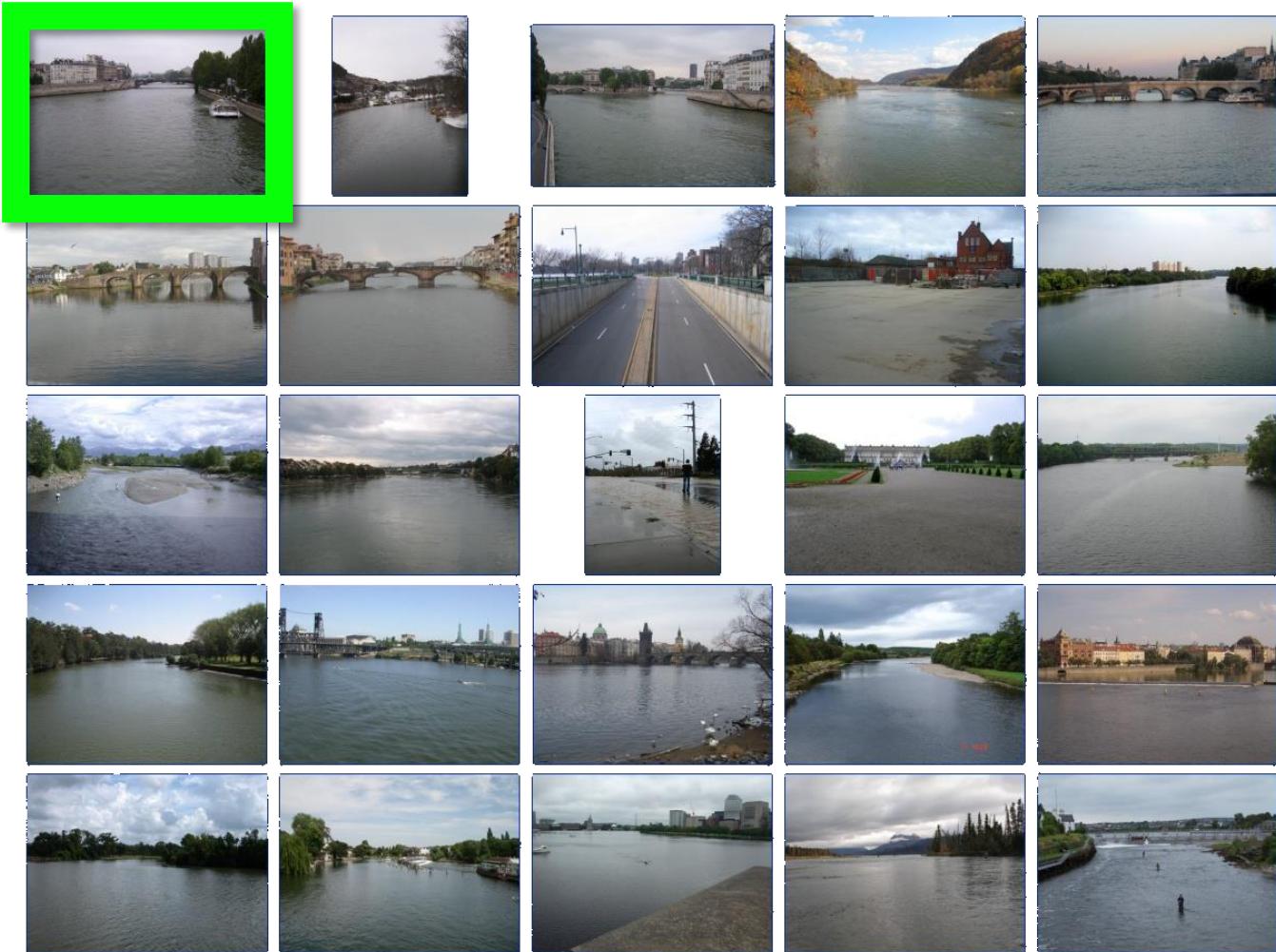








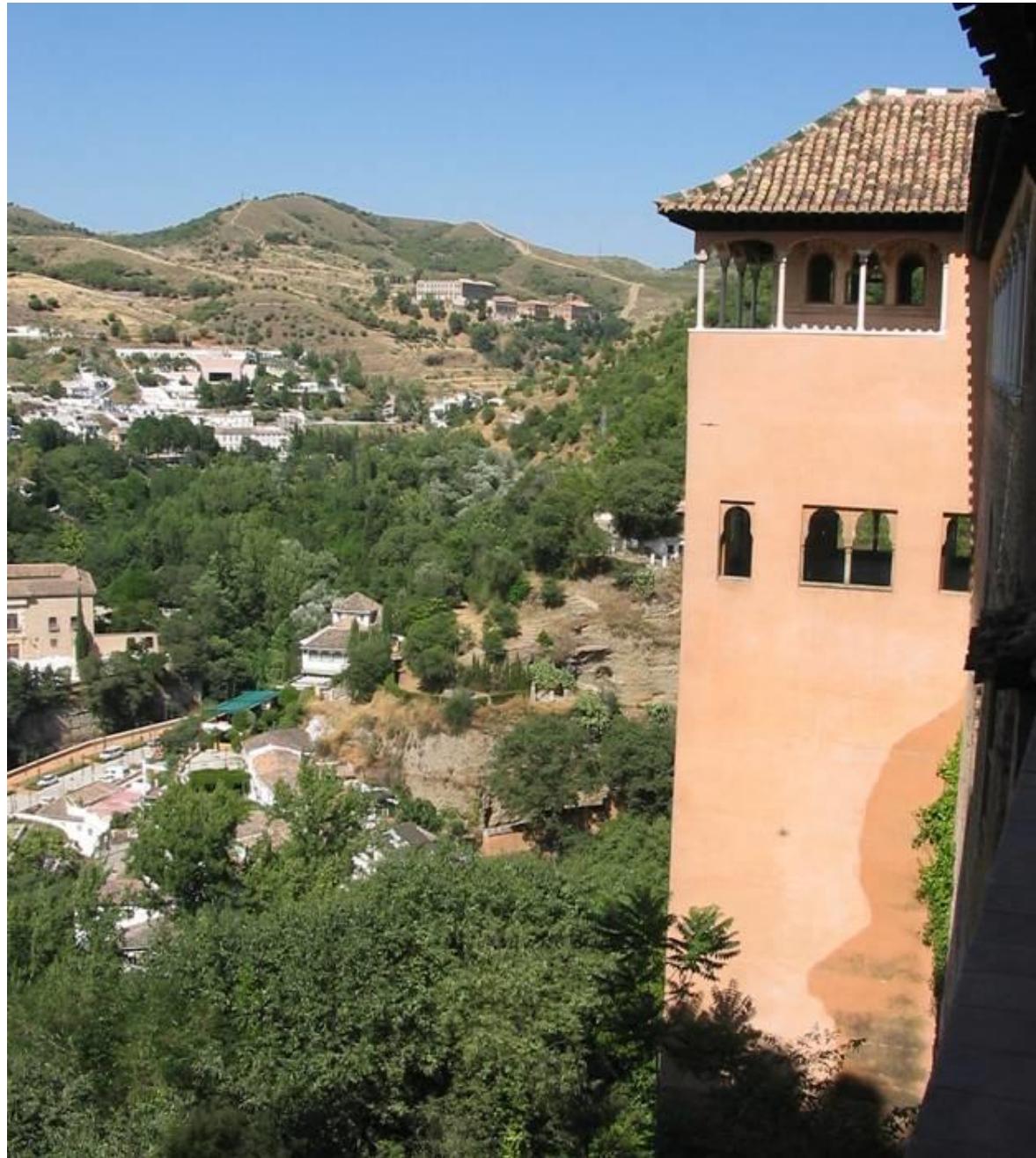


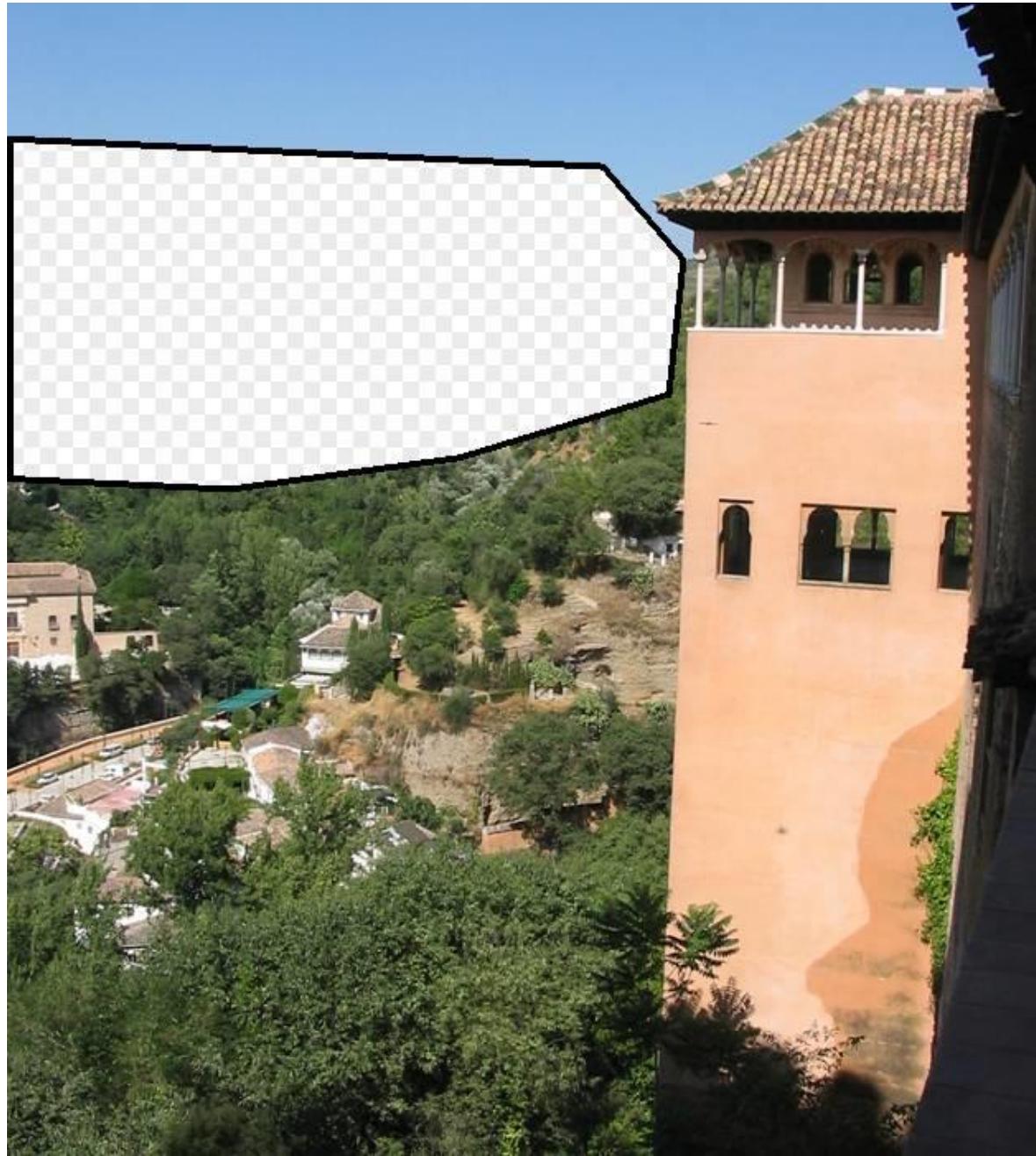


... 200 scene matches











Which is the original?





