6.819 / 6.869: Advances in Computer Vision

Website:

http://6.869.csail.mit.edu/fa15/

Aude Oliva

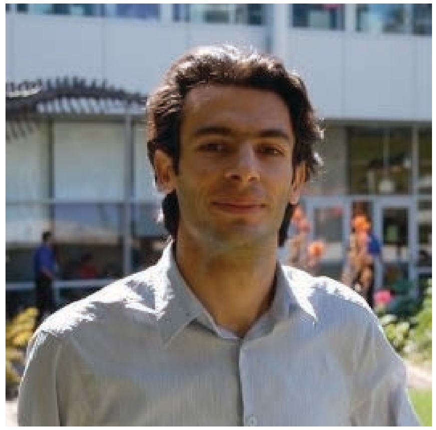
Lecture TR 9:30AM – 11:00AM
(Room 34-101)

Instructors & TAs

Aude Oliva
Principal Research Scientist

oliva@mit.edu

Office hours:
32-D432
Tuesday 11-12 pm
or
by appointment T/Th



Yusuf Aytar
Postdoctoral Researcher

yusuf@csail.mit.edu

Office hours:
Room TBD
Tuesday 5-6 pm

Aditya Khosla
EECS graduate student

khosla@csail. mit.edu

Office hours:
Room TBD
Monday 5-6 pm

Bolei Zhou
EECS graduate student

bolei@mit.edu

Office hours:
Room TBD
Wednesday 5-6 pm

Contacting us by email

- Put **course number in Subject line**:
- **6.819** (or **6.869**)
- Put **topic in subject line**:

Ex: 6.819 Missing Tuesday lecture
6.869 Meeting request: Tuesday 2 pm?
6.819 Late for PS2: Interview
6.869 Project Option 2: Suggestion
6.819 Project Option 3: Literature survey

Schedule at a glance

#	Day	Date	Topic	Assignments
1	Thursday	9/10/2015	Introduction	
	Monday	9/14/2015	Tutorial: Matlab ← **32-D463- Monday 5-6 pm**	
2	Tuesday	9/15/2015	Early vision: Basic of Image processing I	
3	Thursday	9/17/2015	Early vision: Basic of Image processing II	
4	Tuesday	9/22/2015	Early vision: Basic of Image processing II	
5	Thursday	9/24/2015	Early vision: Image Statistics	PS1given
6	Tuesday	9/29/2015	Early vision: Segmentation	
7	Thursday	10/1/2015	Early vision: Texture synthesis	PS2 given - PS1 due
8	Tuesday	10/6/2015	Mid level: Image formation, lenses	
9	Thursday	10/8/2015	Mid level: Basics 3D, Shape from X	
10	Thursday	10/15/2015	Mid level: Motion: continuous	
11	Tuesday	10/20/2015	Mid level: Motion: discrete	PS3 given - PS2 due
12	Thursday	10/22/2015	Learning: Intro. Machine Learning for Vision	
13	Tuesday	10/27/2015	Learning: Intro. to Deep Learning (CNN)	
14	Thursday	10/29/2015	Session on Project Topics	
15	Tuesday	11/3/2015	High level : Object and Scene Recognition I	Project summary due
16	Thursday	11/5/2015	High level : Object and Scene Recognition II	
17	Tuesday	11/10/2015	Applications: Deep Learning for Vision I	PS4 given - PS3 due
		11/11/2015	Tutorial for CNN	
18	Thursday	11/12/2015	Applications: Deep Learning for Vision II	
19	Tuesday	11/17/2015	Applications: Deep Learning for Vision III	
20	Thursday	11/19/2015	Applications: Image Retrieval	
21	Tuesday	11/24/2015	Tutorial: How to give a short talk	PS4 due
22	Tuesday	12/1/2015	Applications: Human and Artificial Visual Brains	
23	Thursday	12/3/2015	Project Presentation: Mini Places Challenge	
24	Tuesday	12/8/2015	Project presentation	
25	Thursday	12/10/2015	Project presentation	Final report due

Assignments

- Problem sets (60%)
- Final project (40%)
  - Summary project (5%)
  - Final presentation (5%)
  - Research component of final project (30%)
- No exams or quizzes

Materials

- **Piazza**: for student collaboration and finding project groups. Instructors and TAs will not be active on Piazza.
- **Stellar**: for turning in late assignments and receiving grades

- Readings: see class website
- http://6.869.csail.mit.edu/fa15/

Problem sets (60%)

- Four problem sets: 15 % each
- Collaboration policy
  - Psets are due individually
  - Done individually but you can talk to people
  - Writing always individually
- **Turn a printed version in class.** Late due on Stellar.
- Up to 4 days late total, for the 4 Psets altogether

  (i.e. if you use all the 4 days on PS1 for instance, you have none left for the other PSets).

  If you are late after that, the grade of the late PS will be zero.

Projects (40%)

Three Project Options

1) Summary of final project proposal (5%): 1 page (template)
   - Individually
   - Due the first week of November (earlier, better!)
2) Research component of final project (30%, template) and final presentation (5%).
   - Presentation (2-5 minutes each): Dec 3, 8, 10
   - Everybody presents.

You are welcome to come to our office hours to brainstorm and suggest your project ideas.

Summary of Project Proposal

- The project proposal should be one page maximum following this template:
- **What is the problem/question** that you will be investigating?
- **What are the most relevant readings**? (2-4 papers)
- **What data will you use?**
- **What method or algorithm will you use?**

- **How will you evaluate your results?**

  Qualitatively, what kind of results do you expect (e.g. plots or figures)

  Quantitatively, what kind of analysis will you use to evaluate and/or compare your results (e.g. what performance metrics or statistical tests)?

Project : A survey (individual)

- Select a topic (to be discussed with one of us)
- Select 10-12 papers: send the list
- Read the papers
- Write a 2500 words survey article (a survey template will be given).
- You can opt for that option, change from a coding project to the survey, at any moment before Thanksgiving

Project: Your own project

2-4 people

- **Applications/Models.** If you have access to a specific large image dataset (e.g. biology, engineering, physics, neuroscience) and a categorization task, you can apply models to this problem.
- From what you learn in class, you can choose a topic/question and propose an approach/model (including questions related to neuroscience).

Project: Mini Places Challenge

2-4 people – Challenge *to be announced*

Introduction

The goal of this challenge is to identify the scene category depicted in a photograph. The data for this task comes from the Places2 dataset which contains 10+ million images belonging to 400+ unique scene categories. Specifically, the challenge data will be divided into 8.1M images for training, 20k images for validation and 381k images for testing coming from 401 scene categories. Note that there is a non-uniform distribution of images per category for training, ranging from 4,000 to 30,000, mimicking a more natural frequency of occurrence of the scene.

For each image, algorithms will produce a list of at most 5 scene categories in descending order of confidence. The quality of a labeling will be evaluated based on the label that best matches the ground truth label for the image. The idea is to allow an algorithm to identify multiple scene categories in an image given that many environments have multi-labels (e.g. a bar can also be a restaurant) and that humans often describe a place using different words (e.g. forest path, forest, woods).

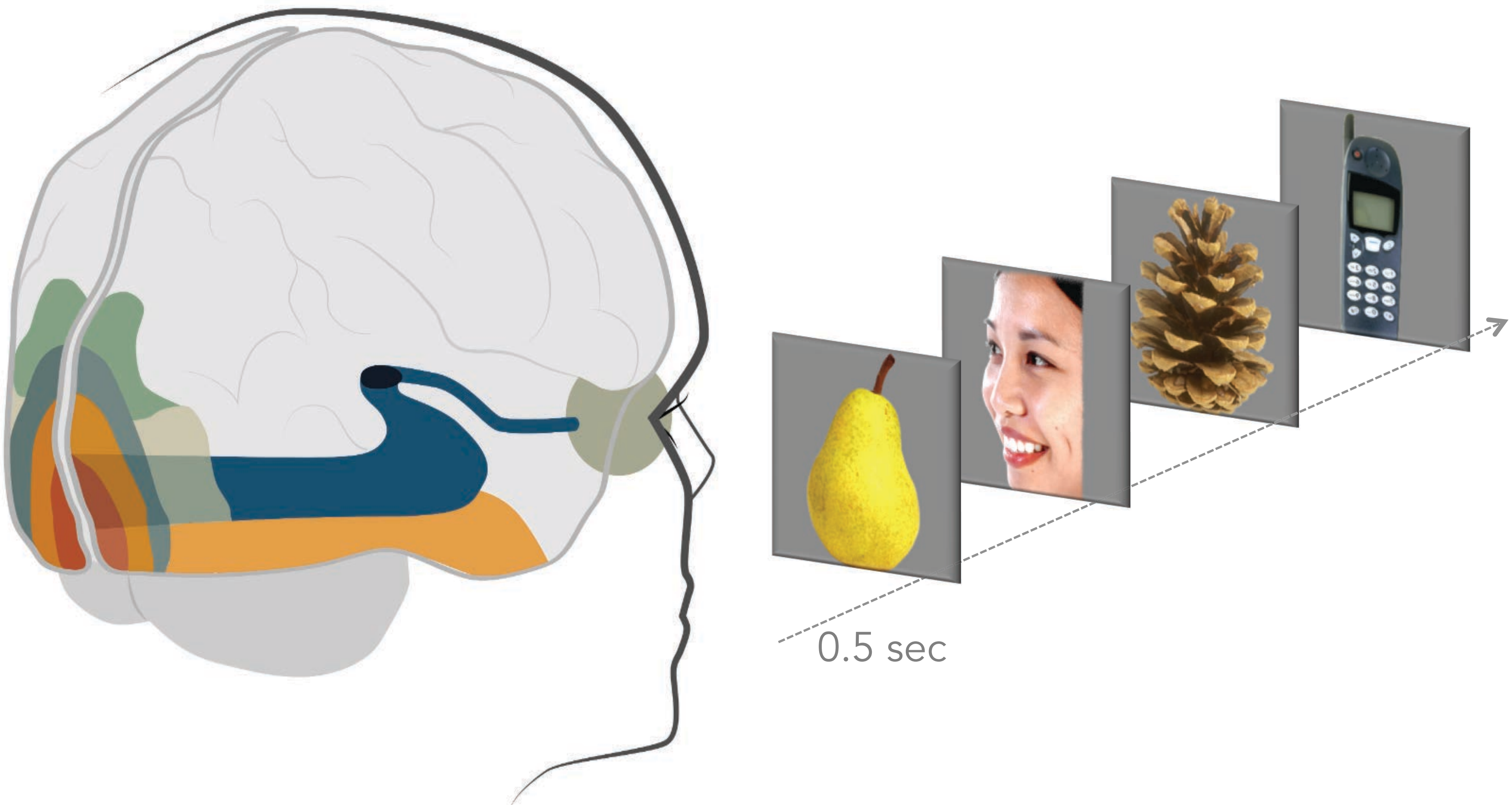
Dates

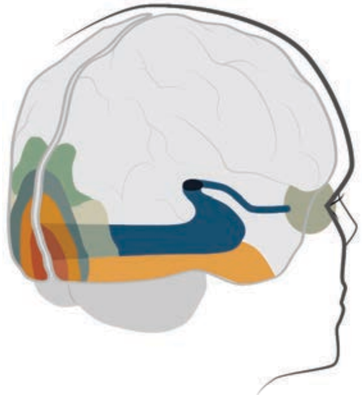
- **August 15, 2015**: Development kit, data, and evaluation software made available
- **November 13, 2015, 5pm PST**: Submission deadline
- **December 10, 2015**: Challenge results released
- **December 17, 2015**: Winner(s) presents at ICCV 2015 Workshop

Organizers

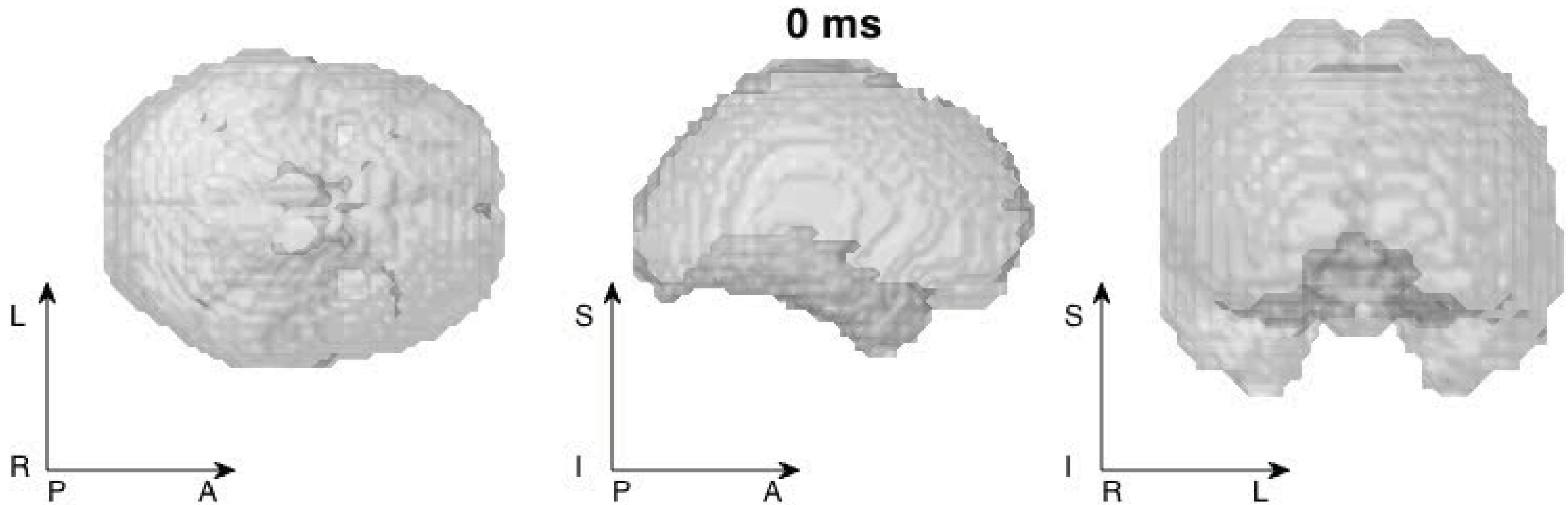
- Aditya Khosla
- Bolei Zhou
- Agata Lapedriza
- Antonio Torralba
- Aude Oliva

Vision: High-Powered Engine

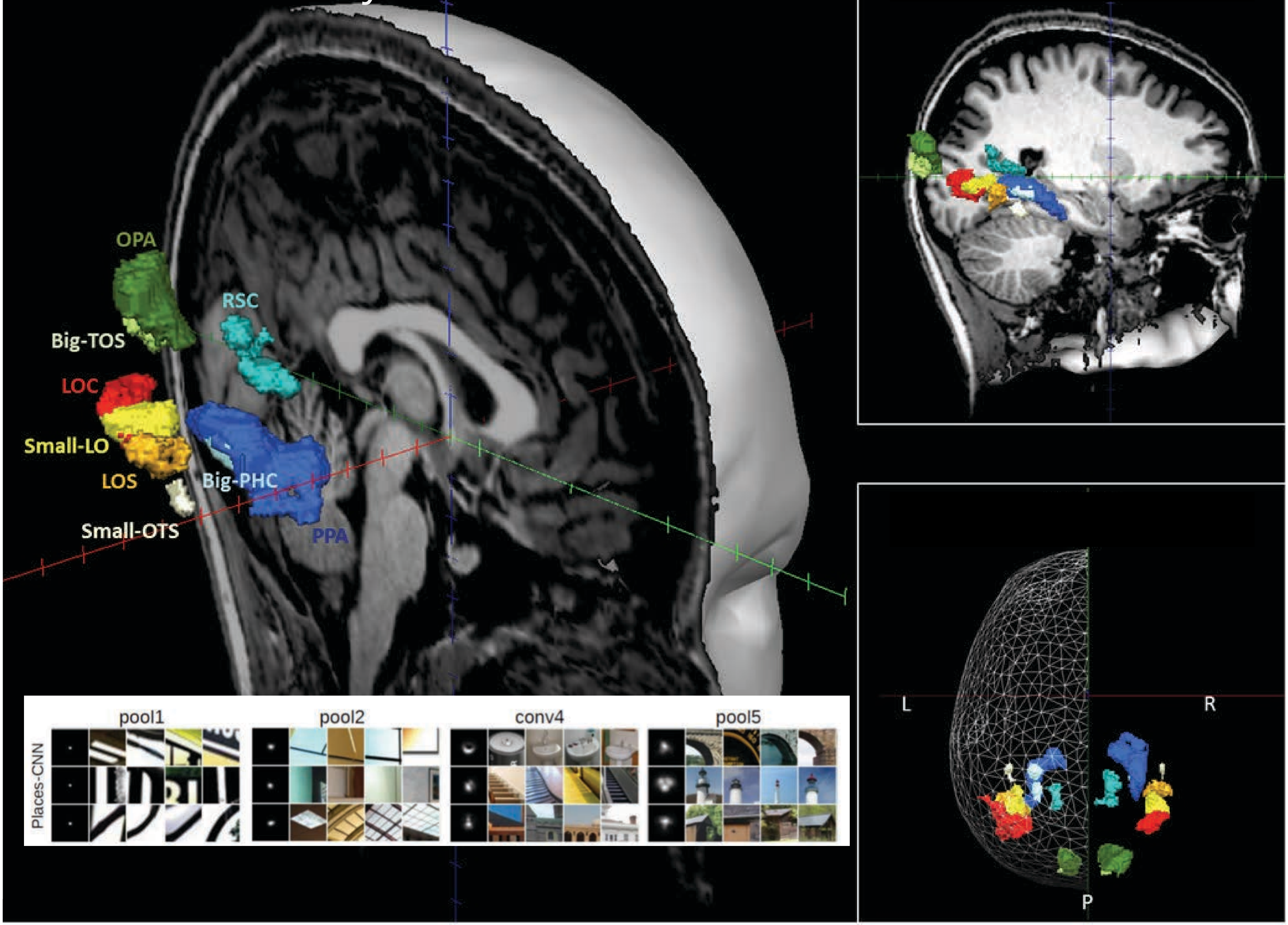


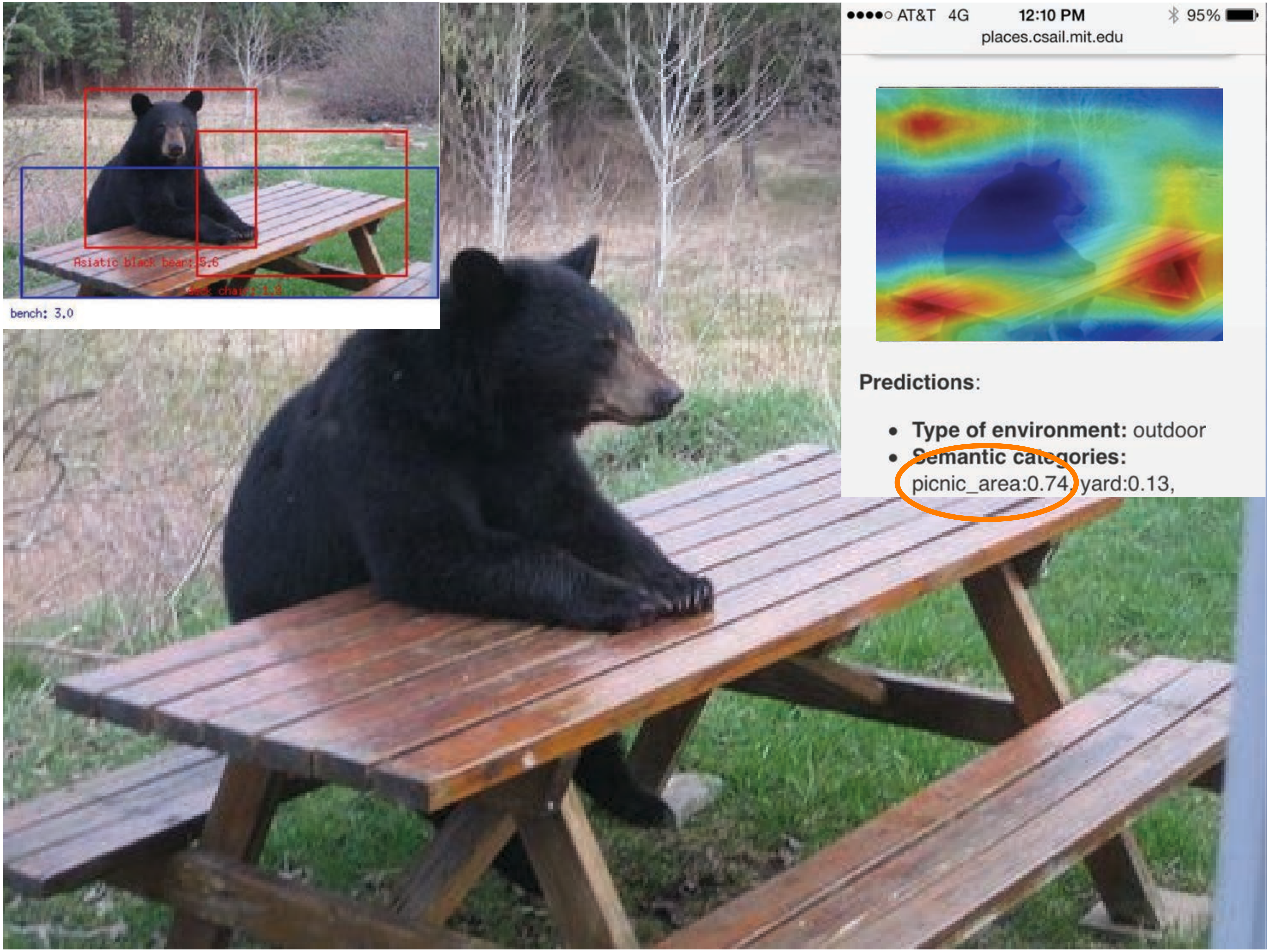


Brain dynamics of seeing



Human Visual System as a Model




Asiatic black bear: 5.6
bench: 3.0
●●●●○ AT&T 4G
12:10 PM
95%
places.csail.mit.edu
Predictions:
• Type of environment: outdoor
• Semantic categories:
picnic_area:0.74, yard:0.13,


Computation: learning from millions of instances

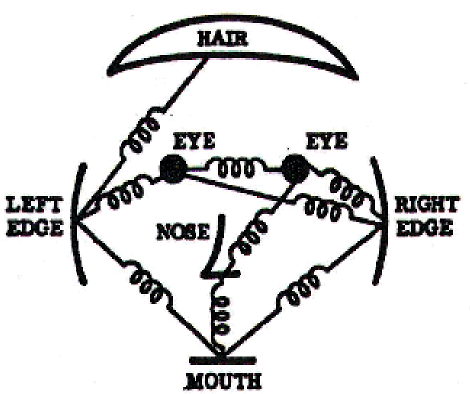
Two extremes of visual learning

Extrapolation problem
Generalization
Diagnostic features

Interpolation problem
Correspondence
Finding the differences

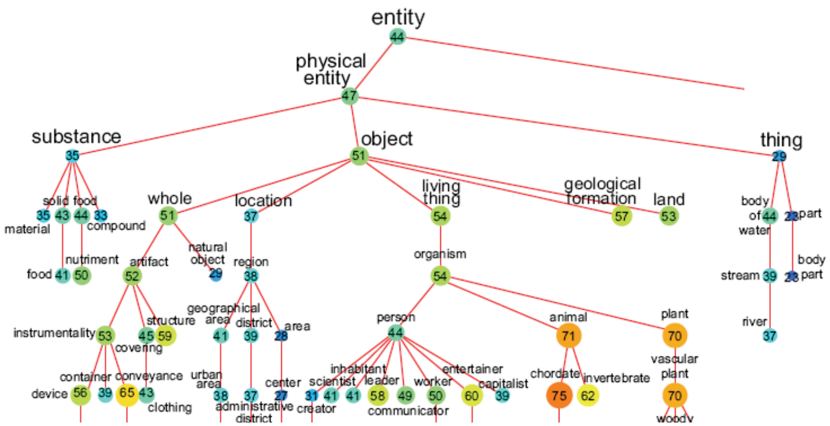
1 — 10 — 10^2 — 10^3 — 10^4 — 10^5 — 10^6 — ∞

Number of training samples



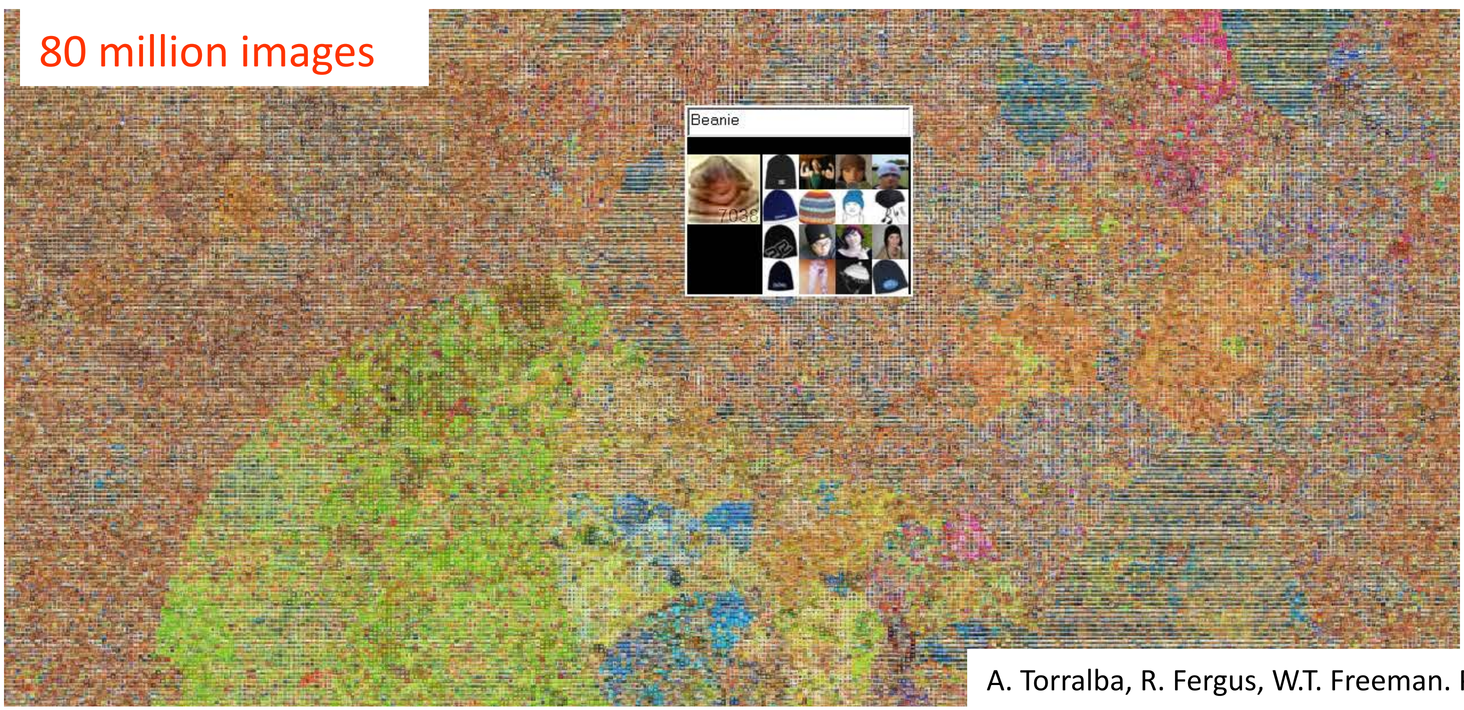
80.000.000 images

75.000 non-abstract nouns from WordNet

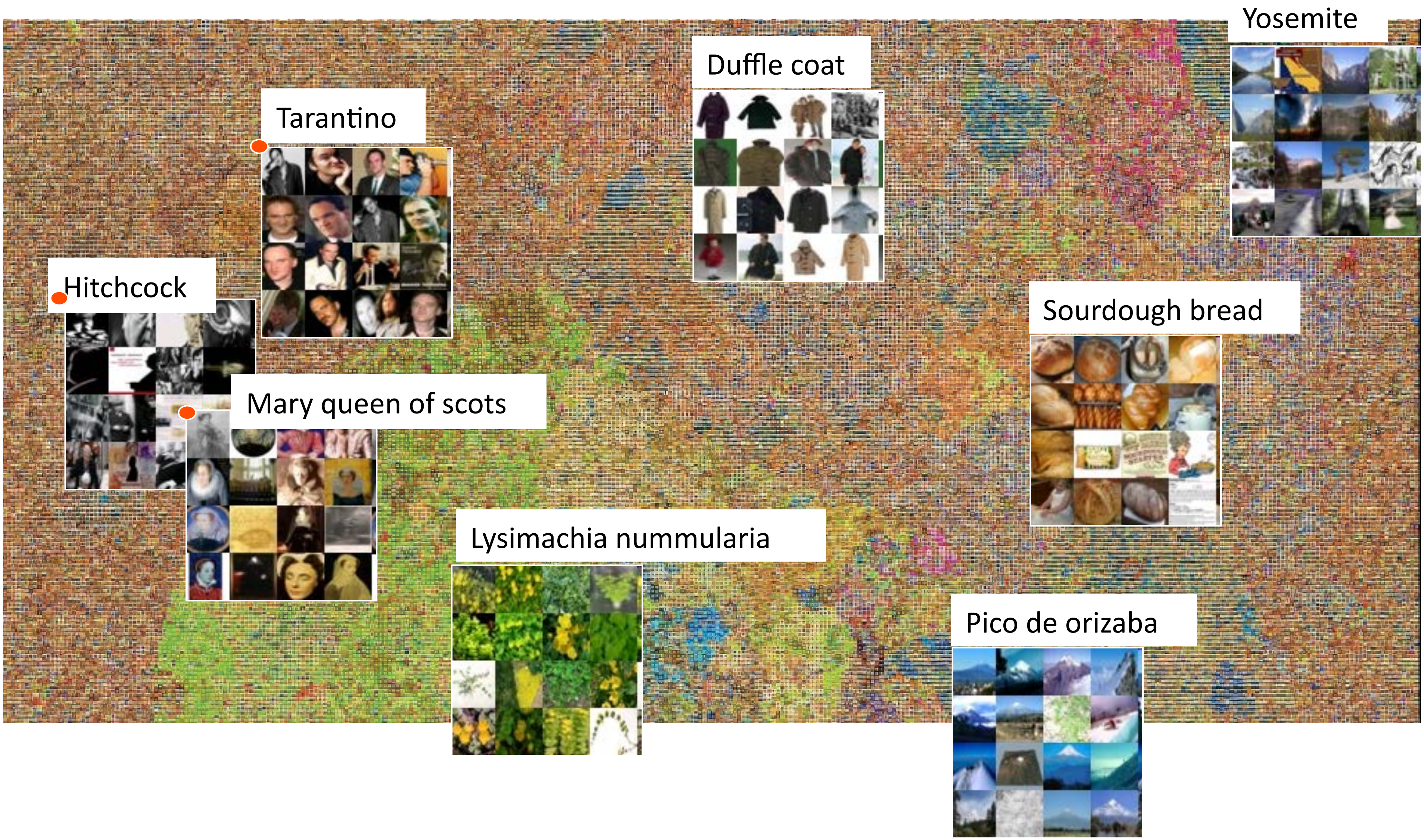


Online image search engines

And after 1 year downloading images

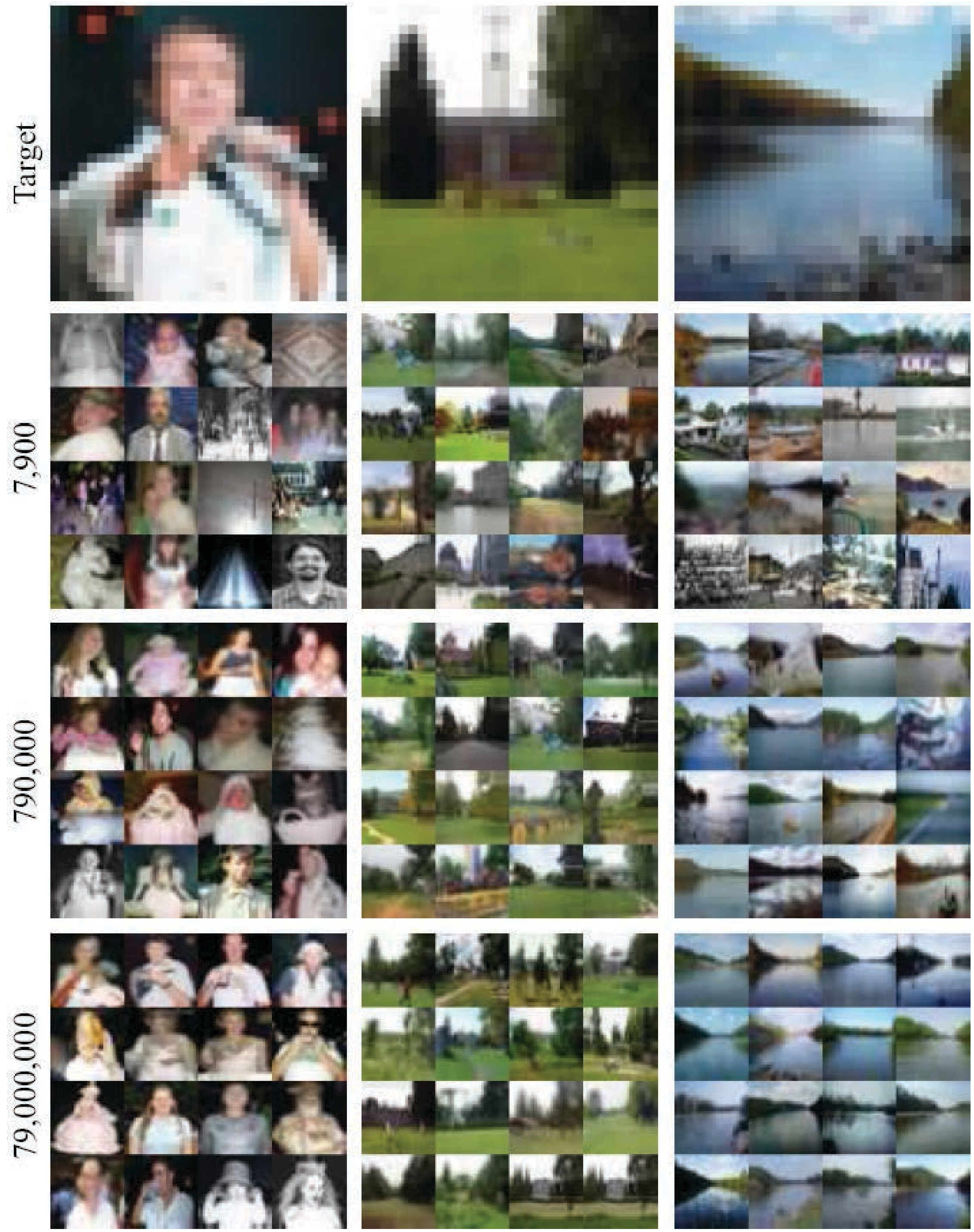


A. Torralba, R. Fergus, W.T. Freeman. PAMI 2008



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

The importance of having
Lots Of Images



What can we do with a good similarity metric and **a lot of data**?

Input image

- Labels
- Motion
- Depth
- …

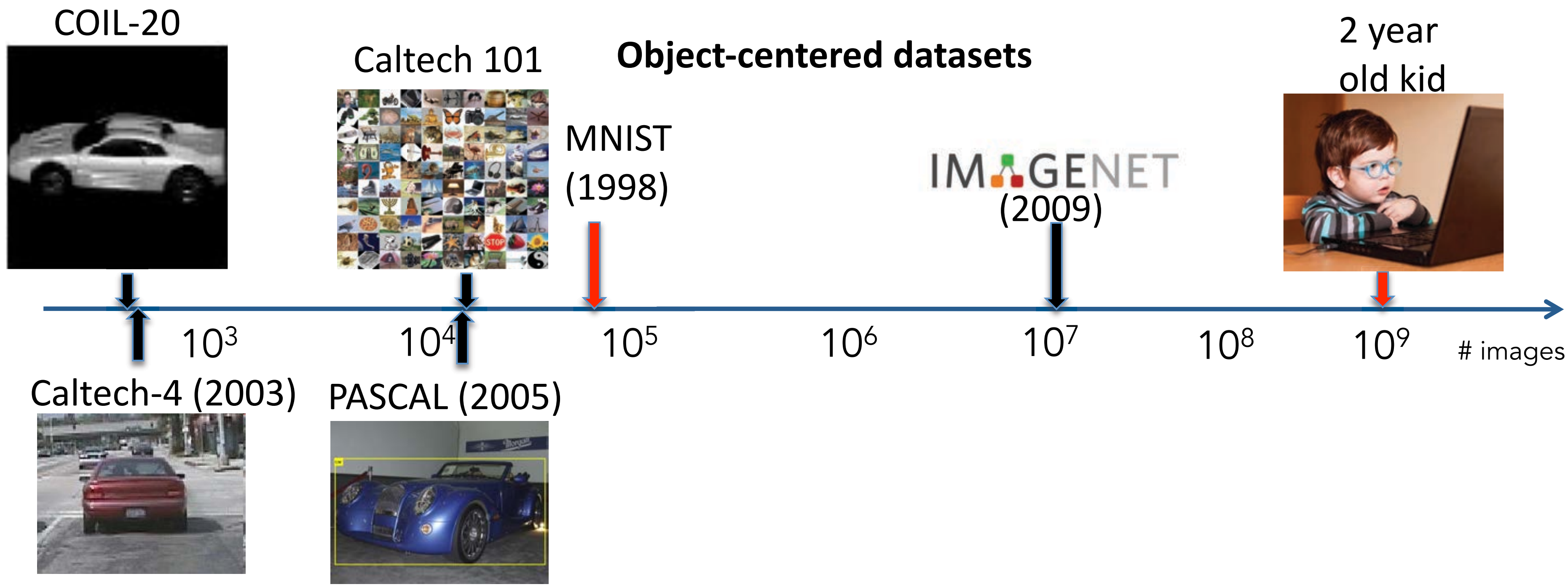
Nearest neighbors

- Labels
- Motion
- Depth
- …

The space of world images: As large databases become available, this opens the door to effective data driven methods.

Hays, Efros, Siggraph 2006
Russell, Liu, Torralba, Fergus, Freeman. NIPS 2007`

The evolution of vision databases



IMAGENET

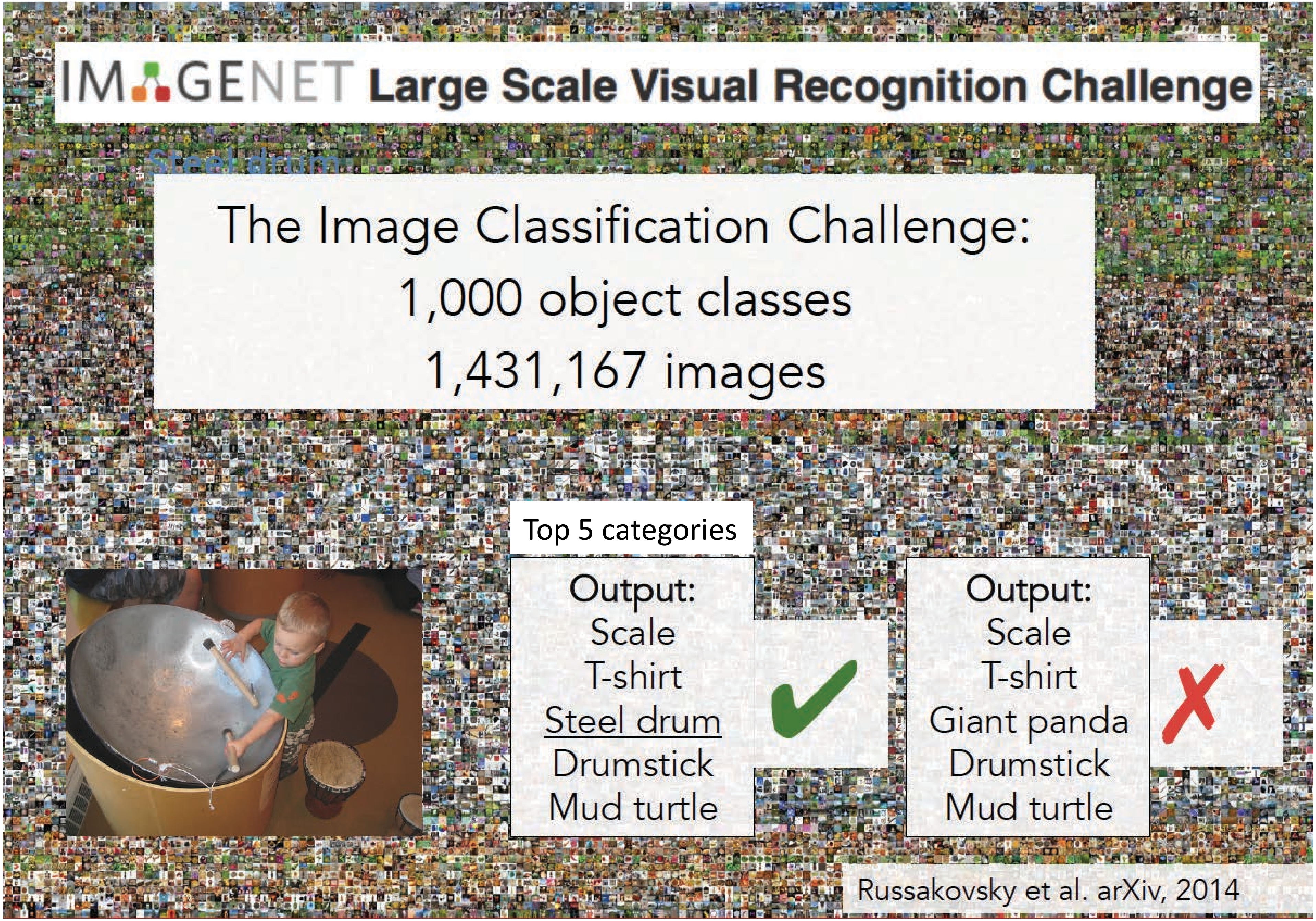
www.image-net.org

22K categories and **14M** images

- Animals
  - Bird
  - Fish
  - Mammal
  - Invertebrate
- Plants
  - Tree
  - Flower
- Food
- Materials
- Structures
- Artifact
  - Tools
  - Appliances
  - Structures
- Person
- Scenes
  - Indoor
  - Geological Formations
- Sport Activities

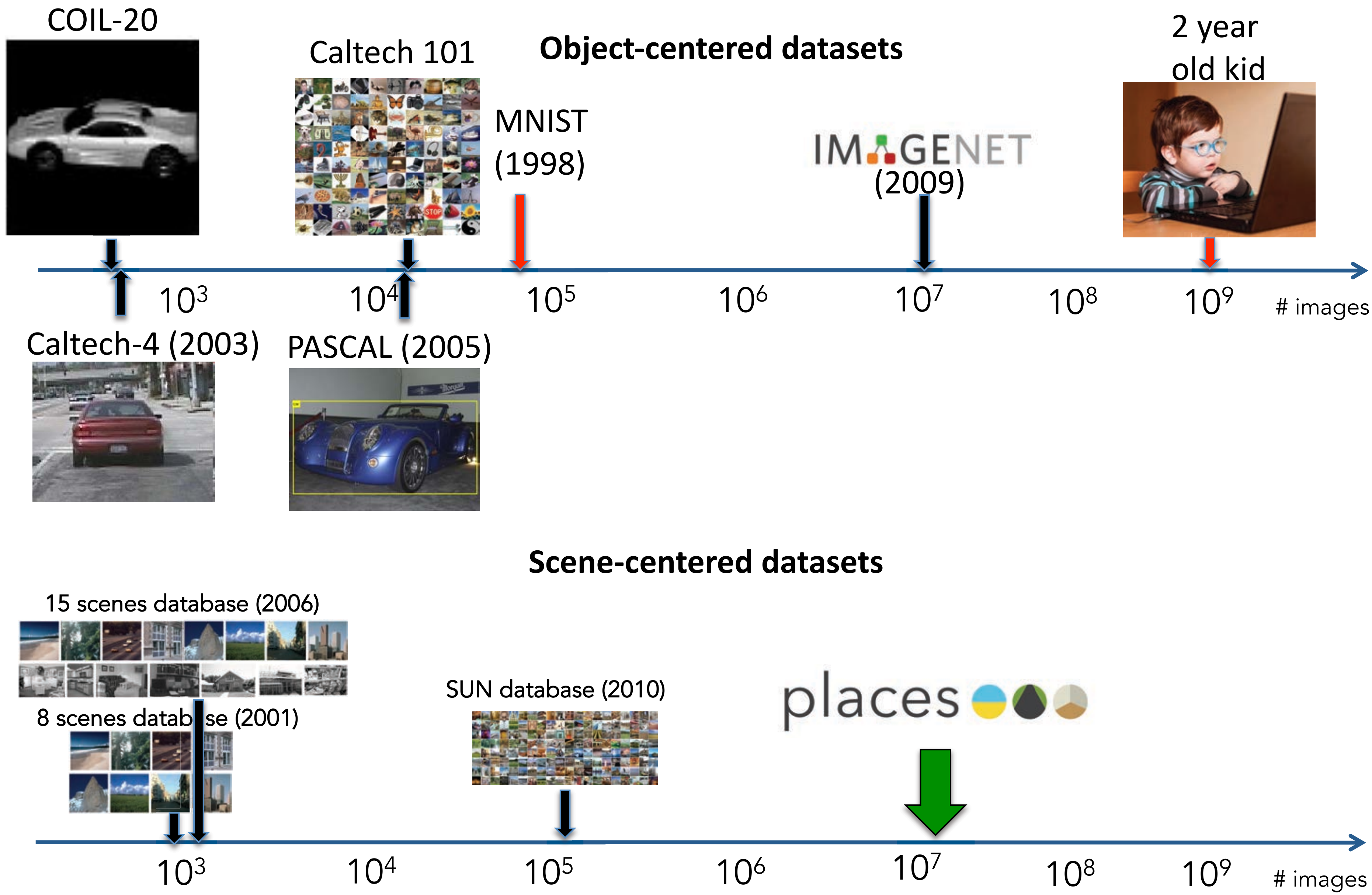
Deng, Dong, Socher, Li, Li, & Fei-Fei, 2009

Slide from Fei-Fei Li & Andrej Karpathy

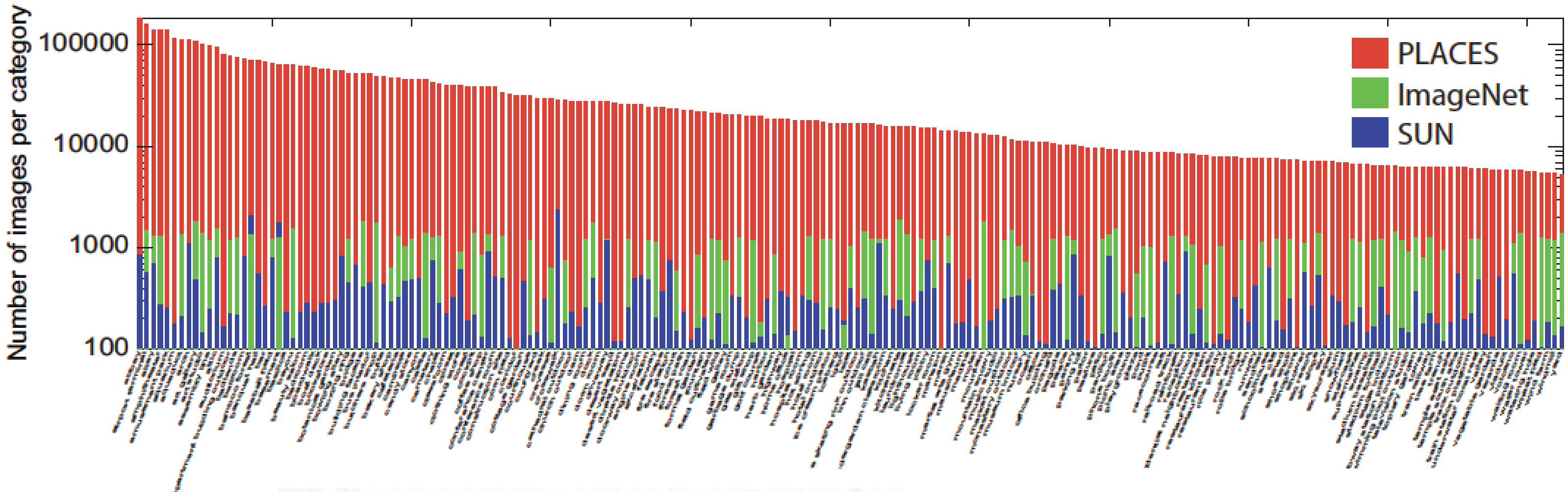


Slide from Fei-Fei Li & Andrej Karpathy

The evolution of vision databases



places



High coverage across categories

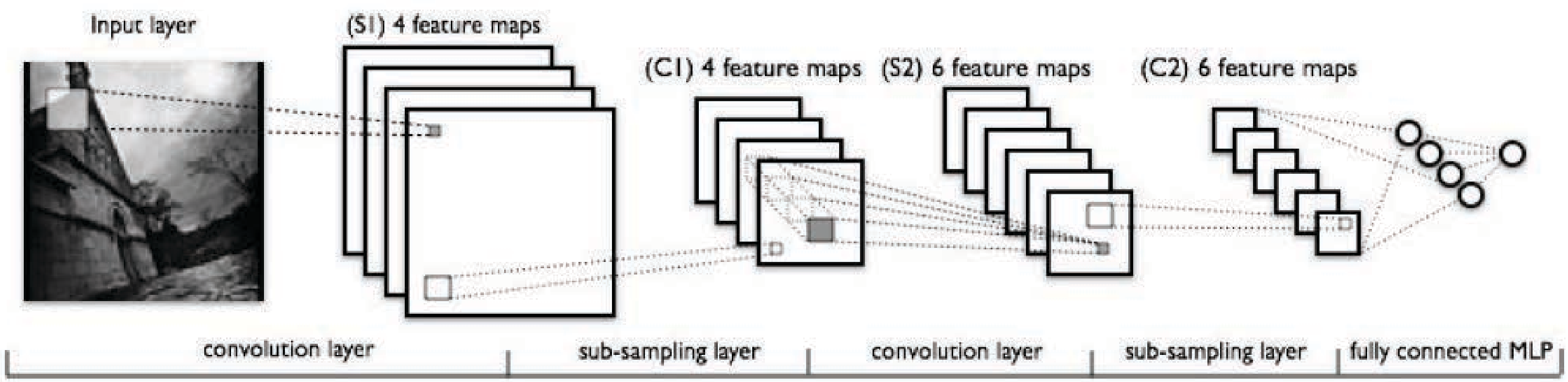
High diversity within category

spare bedroom | messy bedroom | teenage bedroom | romantic bedroom | crazy bedroom

Deep Learning in Computer Vision

Deep Convolutional Neural Network (CNN)

- stack of several layers of computations.
- computations include convolution, max-pooling, fully connecting,etc.



- First proposed in 1989, rediscovered to revolutionize computer vision in 2012.

Deep Learning for Object Recognition

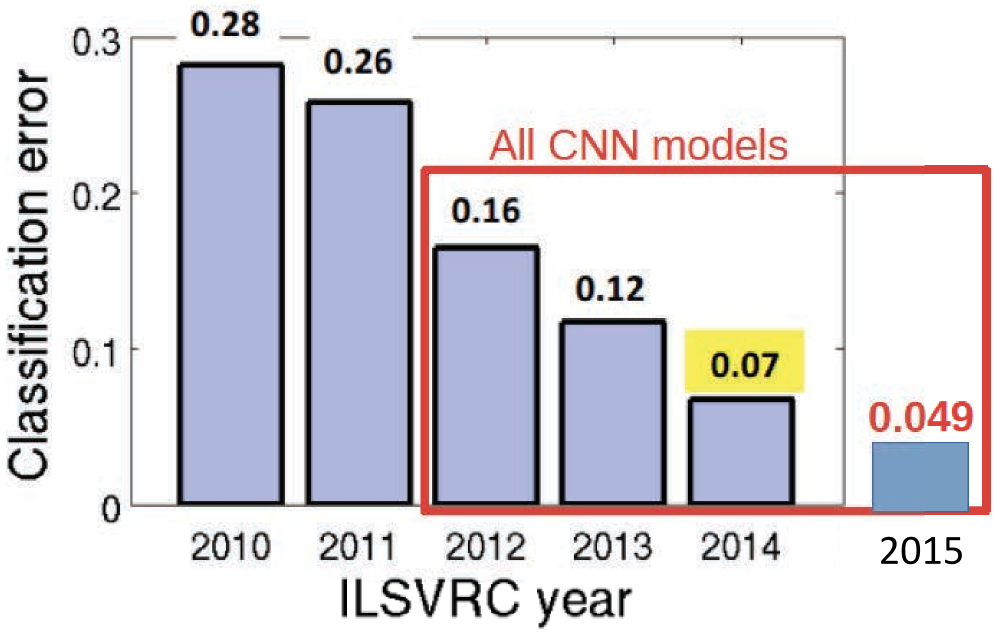
- ILSVRC: large-scale recognition on ImageNet

  1000 object classes

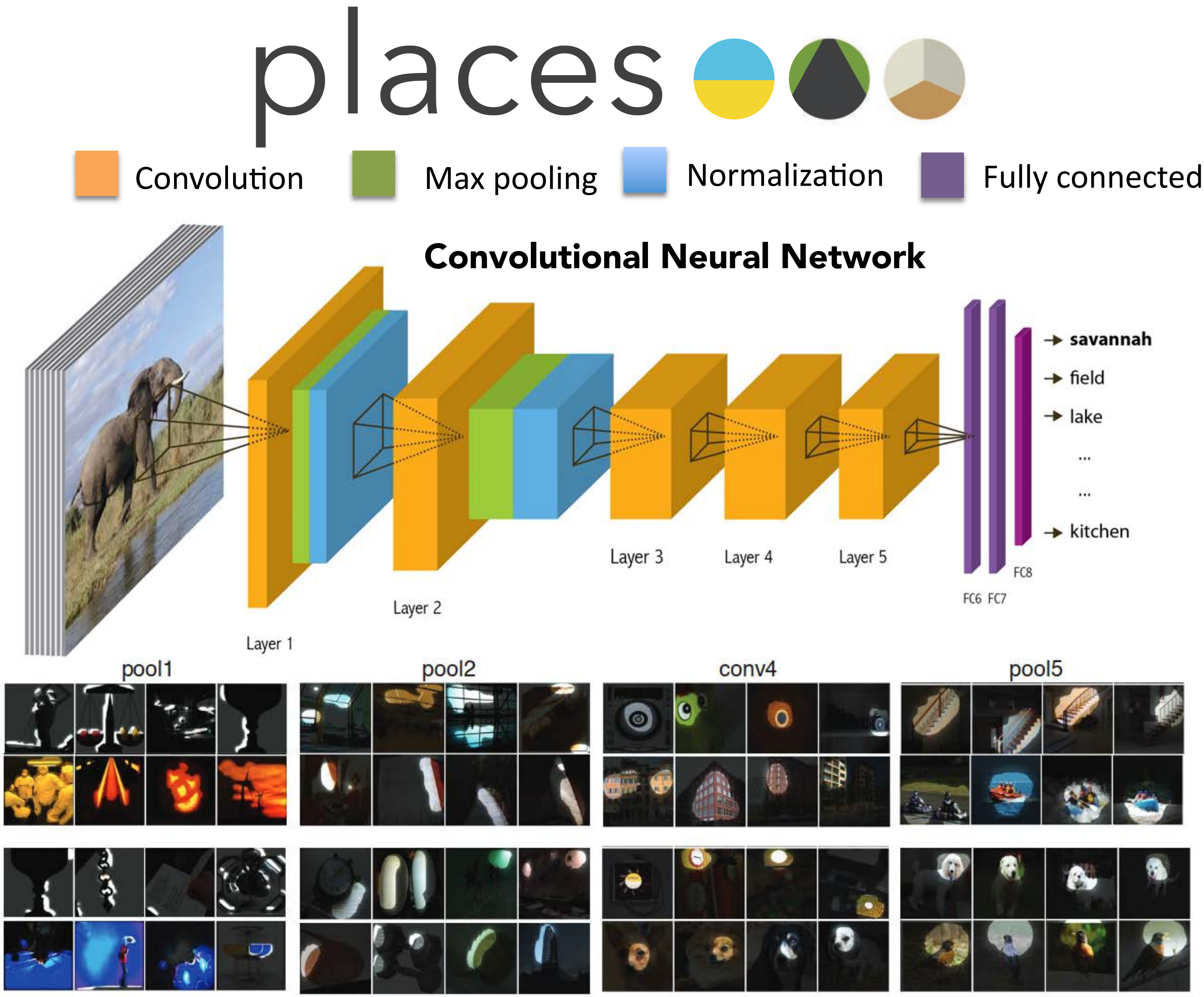
  1.2 million training data (~1000 images per class)

  100 test images per class

Top-5 error rate



Krizhevsky, Alex et al. Imagenet classification with deep convolutional neural networks. NIPS 2012

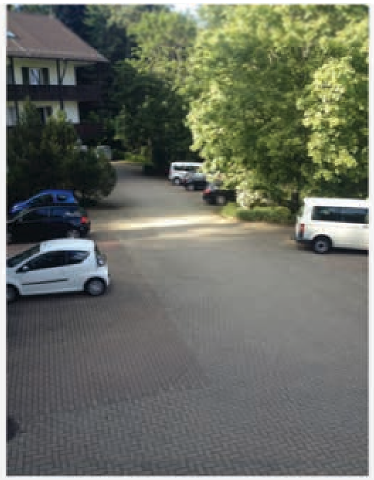

places
Convolution
Max pooling
Normalization
Fully connected
Convolutional Neural Network
savannah
field
lake
...
...
kitchen
Layer 1
Layer 2
Layer 3
Layer 4
Layer 5
FC6 FC7
FC8
pool1
pool2
conv4
pool5

places

Smart phone: places.csail.mit.edu

Predictions:

- **Type of environment:** indoor
- **Semantic categories:** restaurant:0.27, coffee_shop:0.23, cafeteria:0.21, food_court:0.12, restaurant_patio:0.09

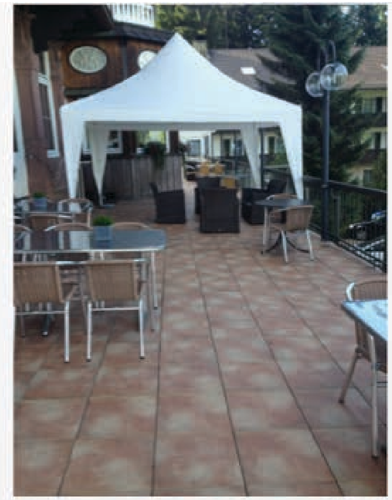


Predictions:

- **Type of environment:** outdoor
- **Semantic categories:** parking_lot:0.46, driveway:0.44,

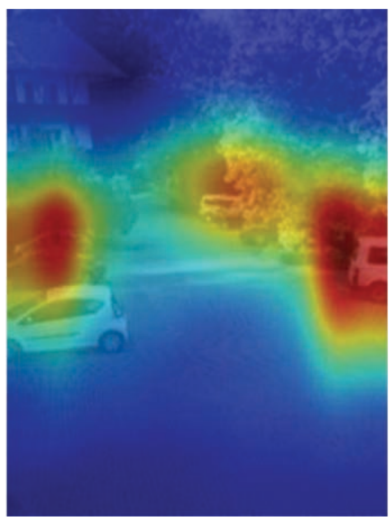
Predictions:

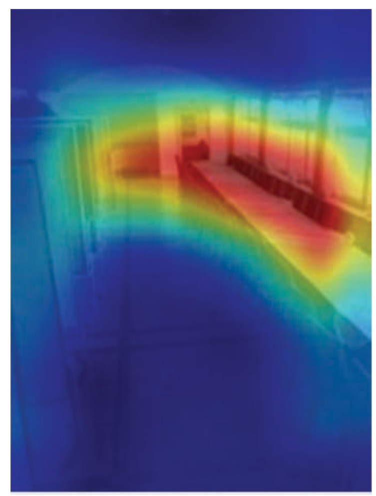
- **Type of environment:** indoor
- **Semantic categories:** conference_room:0.29, dining_room:0.27, banquet_hall:0.08, classroom:0.06,

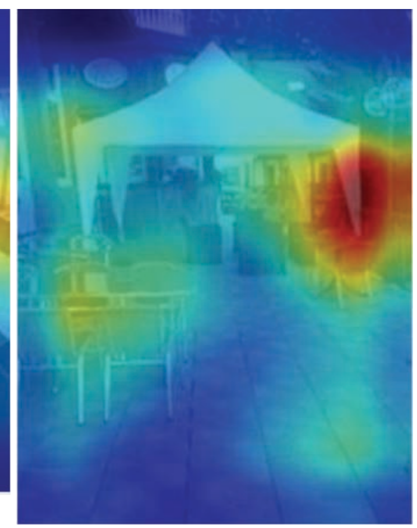


Predictions:

- **Type of environment:** outdoor
- **Semantic categories:** patio:0.38, restaurant_patio:0.35, restaurant:0.06,







Predictions:

- **Type of environment:** outdoor
- **Semantic categories:** swimming_pool/outdoor:0.80,
- **SUN scene attributes:** man-made, nohorizon, naturallight, bathing, warm, directsunsunny, swimming, diving, stillwater, openarea
- **Informative region for the category *swimming_pool_outdoor* is:**