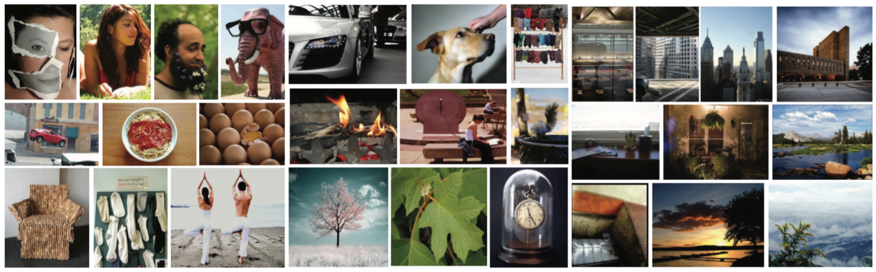
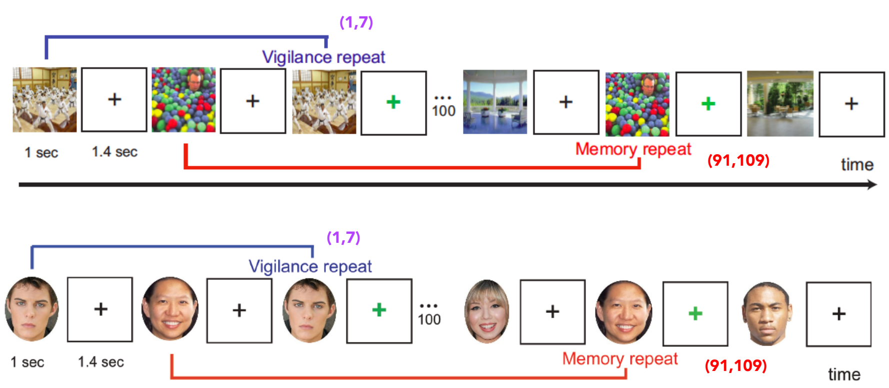
The memorability of visual images

Khosla, Raju, Torralba, Oliva (2015)



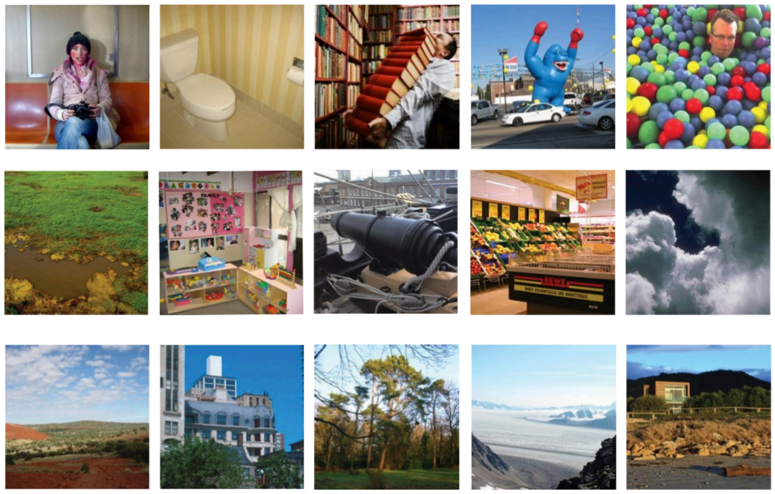
Visual memory experiments



- Continuous repeat detection task
- ~**10,000** unique images sampled from 900 scene categories (Standing, 1973; Brady et al., 2008)
- ~**10,000** unique face sampled from US demographic adult distribution (gender, race, age)
- ~ 10,000 unique words
- **2222** target images (memory repeats) whose repeats occurred ~ 91-109 after the first presentation

- Vigilance repeats every 1-7 images
- Each game level has 120 images
- ~ 80 scores per target images

Large difference in image memorability & high consistency between observers' groups



Memorable ~ 90%

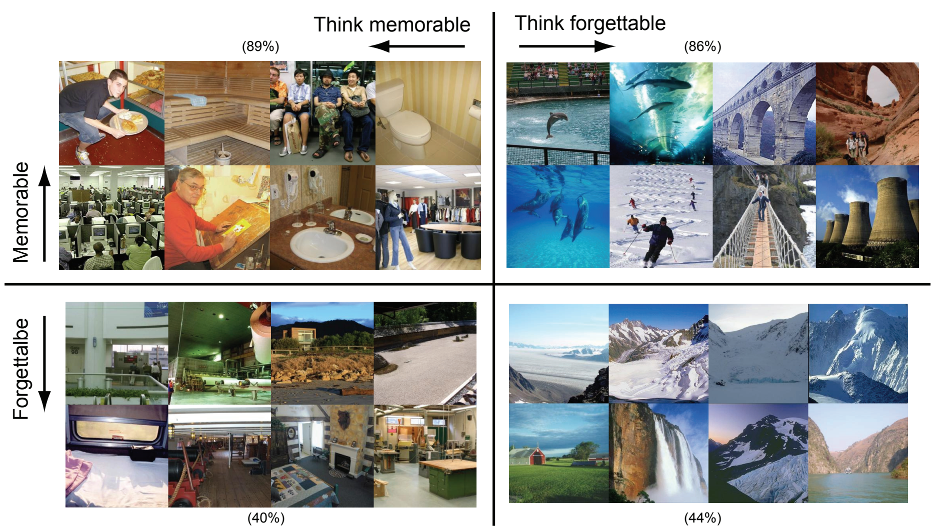
Average ~ 67%

Forgettable ~40%

Mean HIT rate: 67.5% • SD: 13.5%
Mean False alarm rate: 10.7% • SD: 7.6%

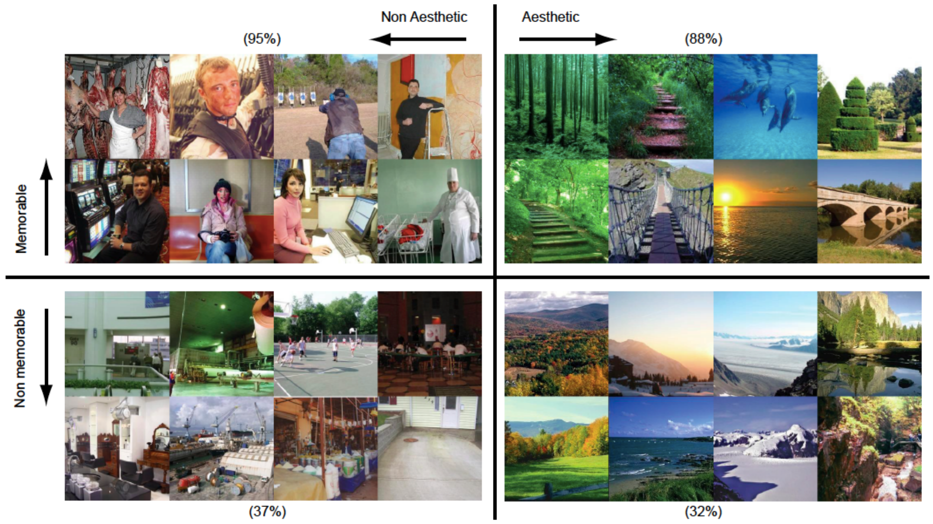
Isola et al (2011). *IEEE CVPR*

Subjective judgments <u>do not predict</u> image memorability



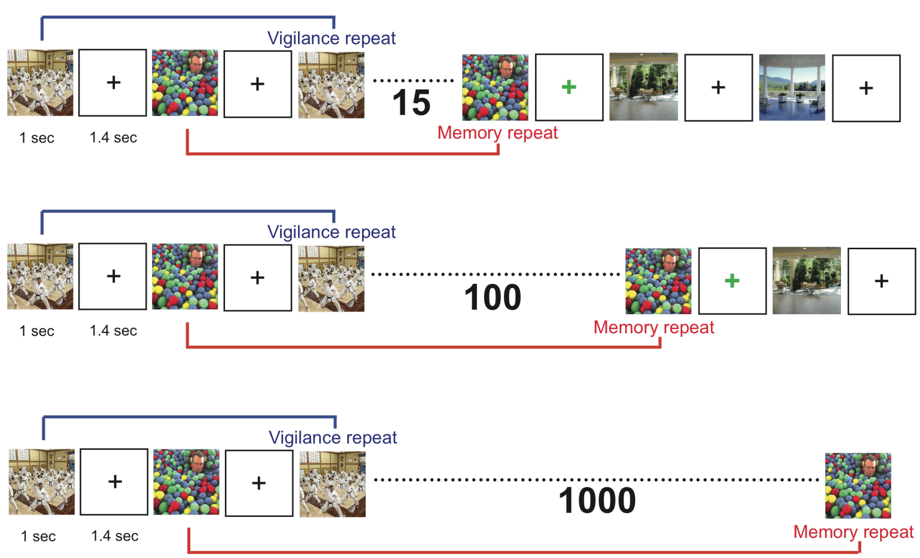
Isola et al (2011). *Neural Information Processing Systems (NIPS). PAMI (2014)*

Image memorability is distinct from image aesthetic



Isola et al (2011). *Neural Information Processing Systems (NIPS)*

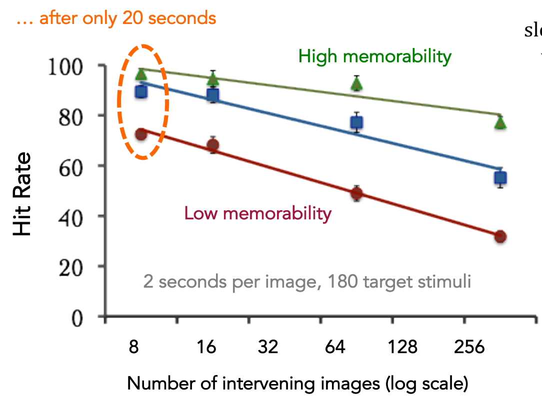
Is memorability stable across time? Yes



Isola et al (2014). IEEE PAMI (Pattern Analysis and Machine Intelligence)

When do memorability differences arise?

At stage of encoding: data suggest that some images (features) are encoded in less sufficient detail than others



In-lab experiment:
slope of forgetting increases with memorability levels

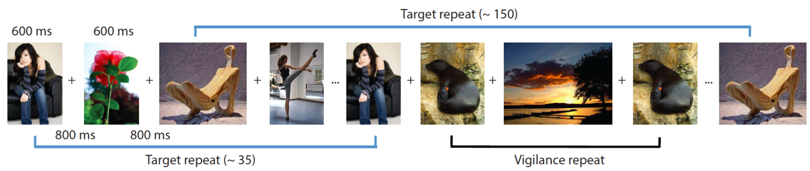
What about faces?

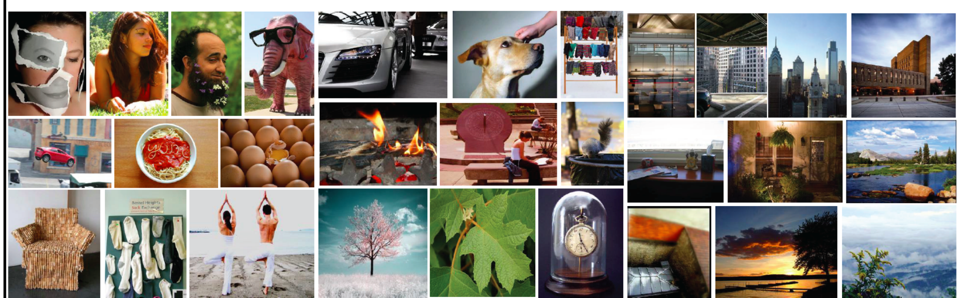
large differences in face memorability, high consistency between observers' groups

Bainbridge, Isola, Oliva, (2013). J. Exp. Psychology: General.

Large scale visual memorability

60,000 photographs with memorability scores



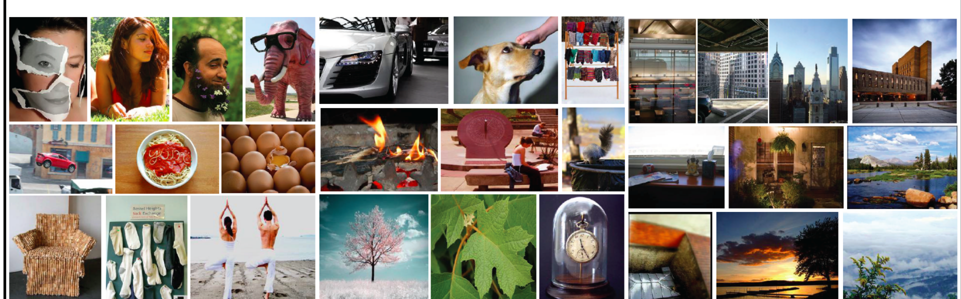


most memorable

http://memorability.csail.mit.edu/

less memorable

Most memorable / Least memorable



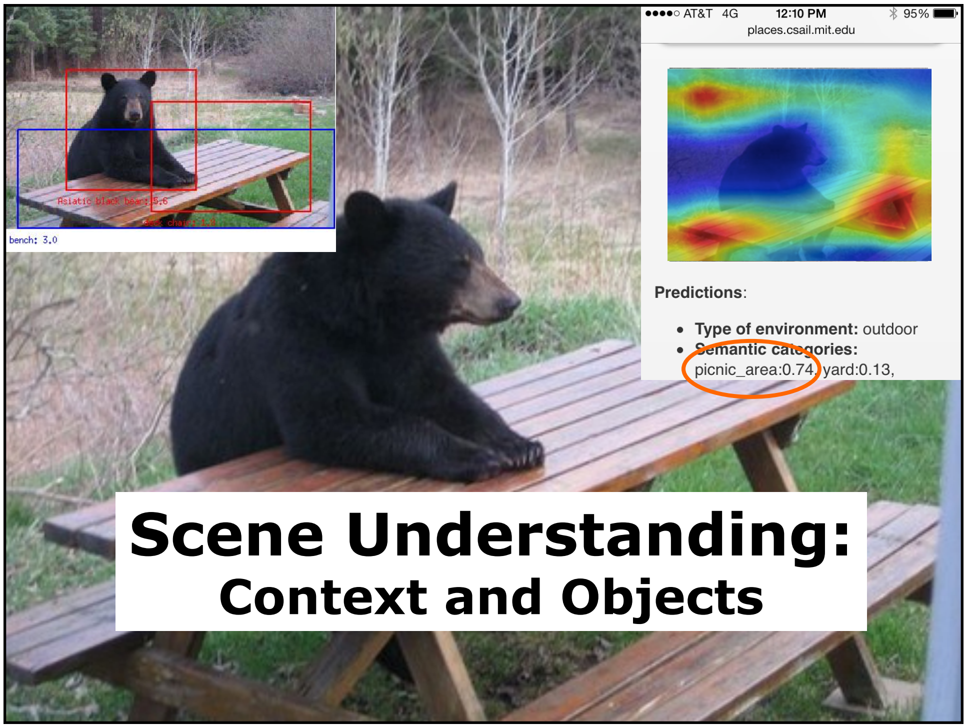
Most memorable:

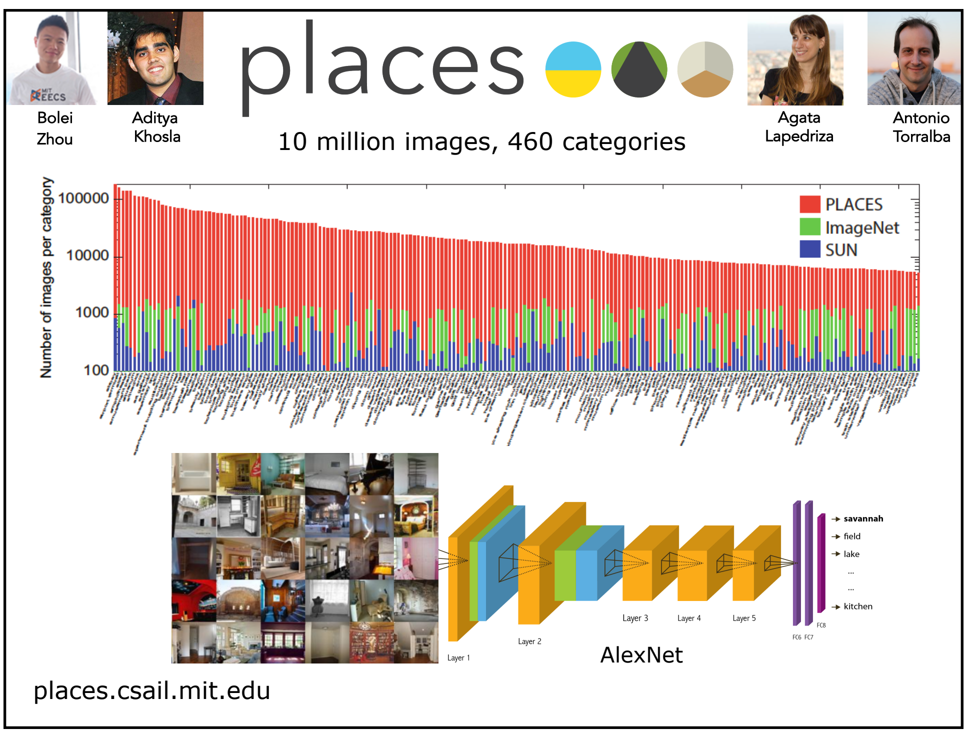
- Focus
- Enclosed Setting
- Dynamics
- Unusual

Least memorable:

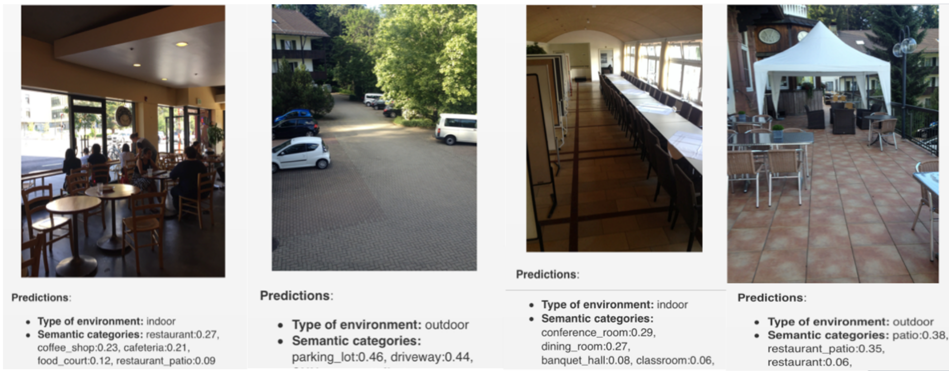
- No single focus
- Distant view
- Static
- Common

You need to recognize to remember!

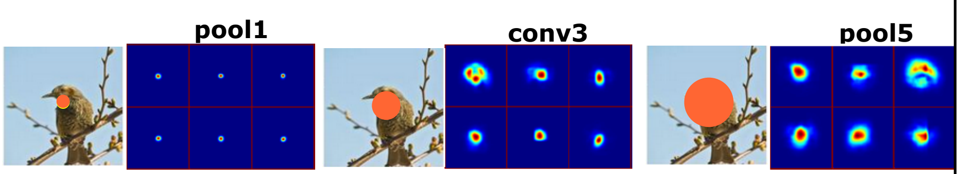

AT&T 4G
12:10 PM
95%
places.csail.mit.edu
bench: 3.0
Predictions:
• Type of environment: outdoor
• Semantic categories:
picnic_area:0.74, yard:0.13,
Scene Understanding:
Context and Objects

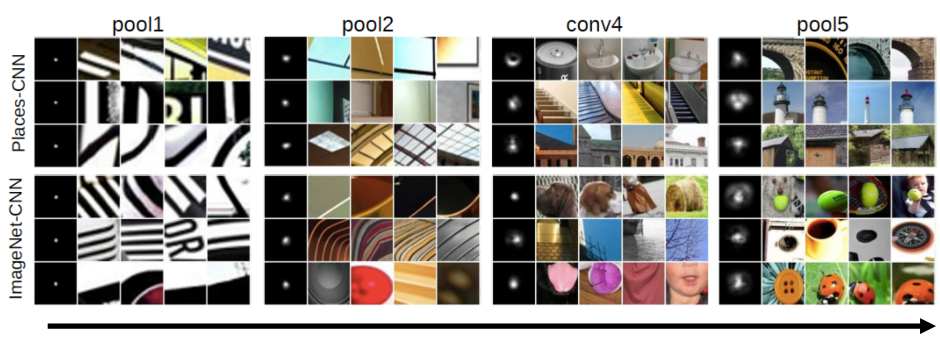

Bolei Zhou
Aditya Khosla
places
10 million images, 460 categories
Agata Lapedriza
Antonio Torralba
Number of images per category
100000
10000
1000
100
PLACES
ImageNet
SUN
savannah
field
lake
...
kitchen
Layer 1
Layer 2
Layer 3
Layer 4
Layer 5
FC6
FC7
FC8
AlexNet
places.csail.mit.edu

Web demo: places.csail.mit.edu



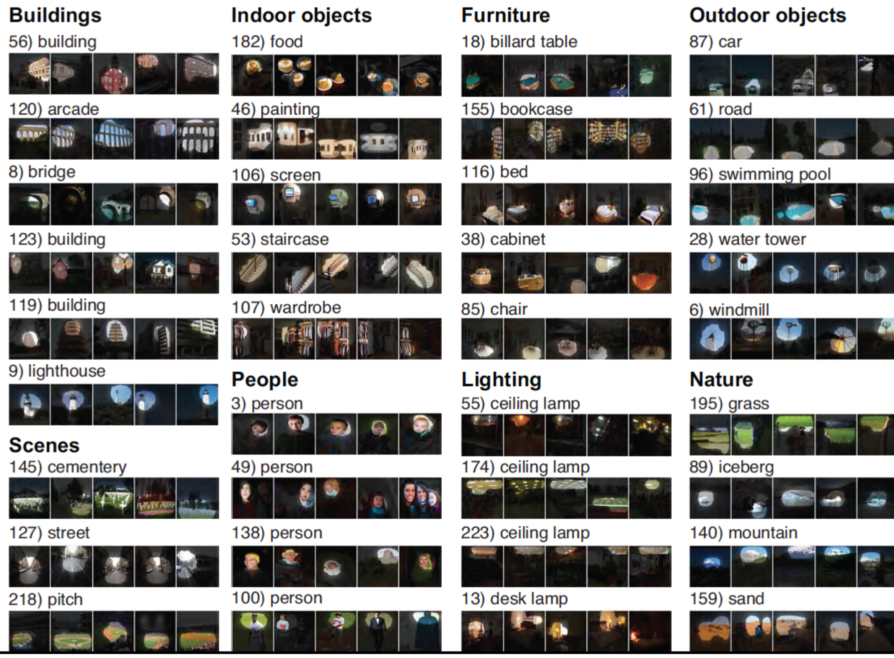
"Receptive fields" of CNN units





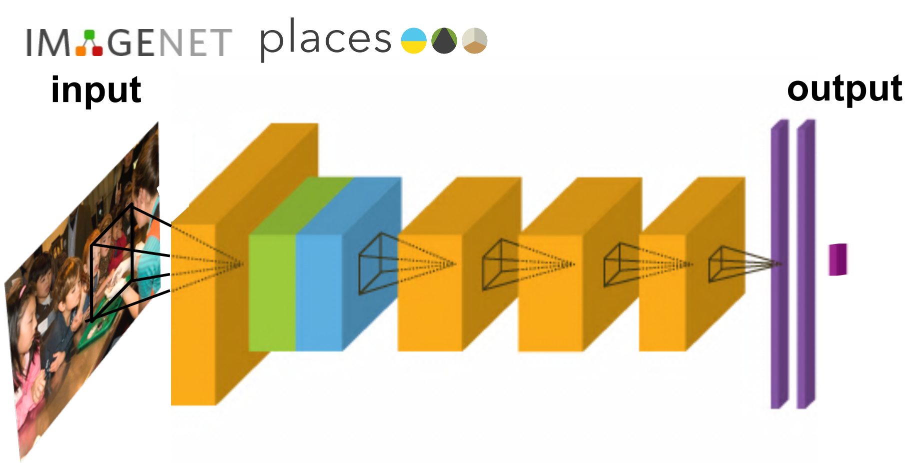
Zhou, Khosla, Lapedriza, Oliva, Torralba (2015), ICLR

Object detectors emerge inside the CNN



MemNet

CNN for Predicting Image Memorability



HybridNet from Zhou et al, NIPS 2014 (places.csail.mit.edu)