Applications: Deep Learning for Vision I
6.869 Advances in Computer Vision

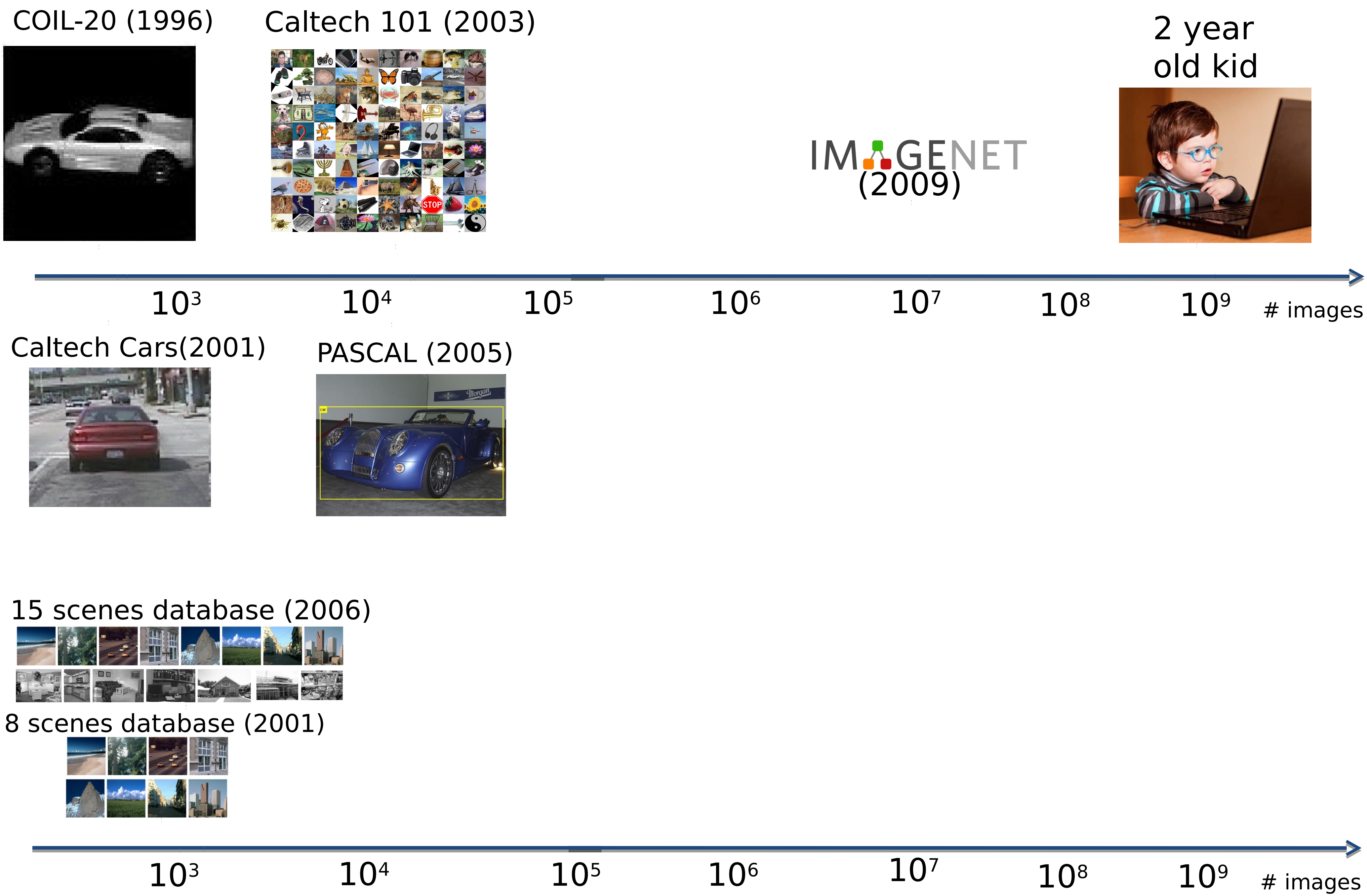
Places Database and Places-CNNs for Large-scale Scene Recognition

Bolei Zhou
Nov.9, 2015

Lecture Outline

- Building the Places Database
- Training CNN on Places Database
- Analyzing the Places-CNNs

The evolution of object and scene databases



The 15-scenes benchmark

15 scene categories and 5, 000 images


Skyscrapers


Suburb


Building facade


Coast


Forest


Bedroom


Living room


Office


Kitchen


Store


Industrial


Street

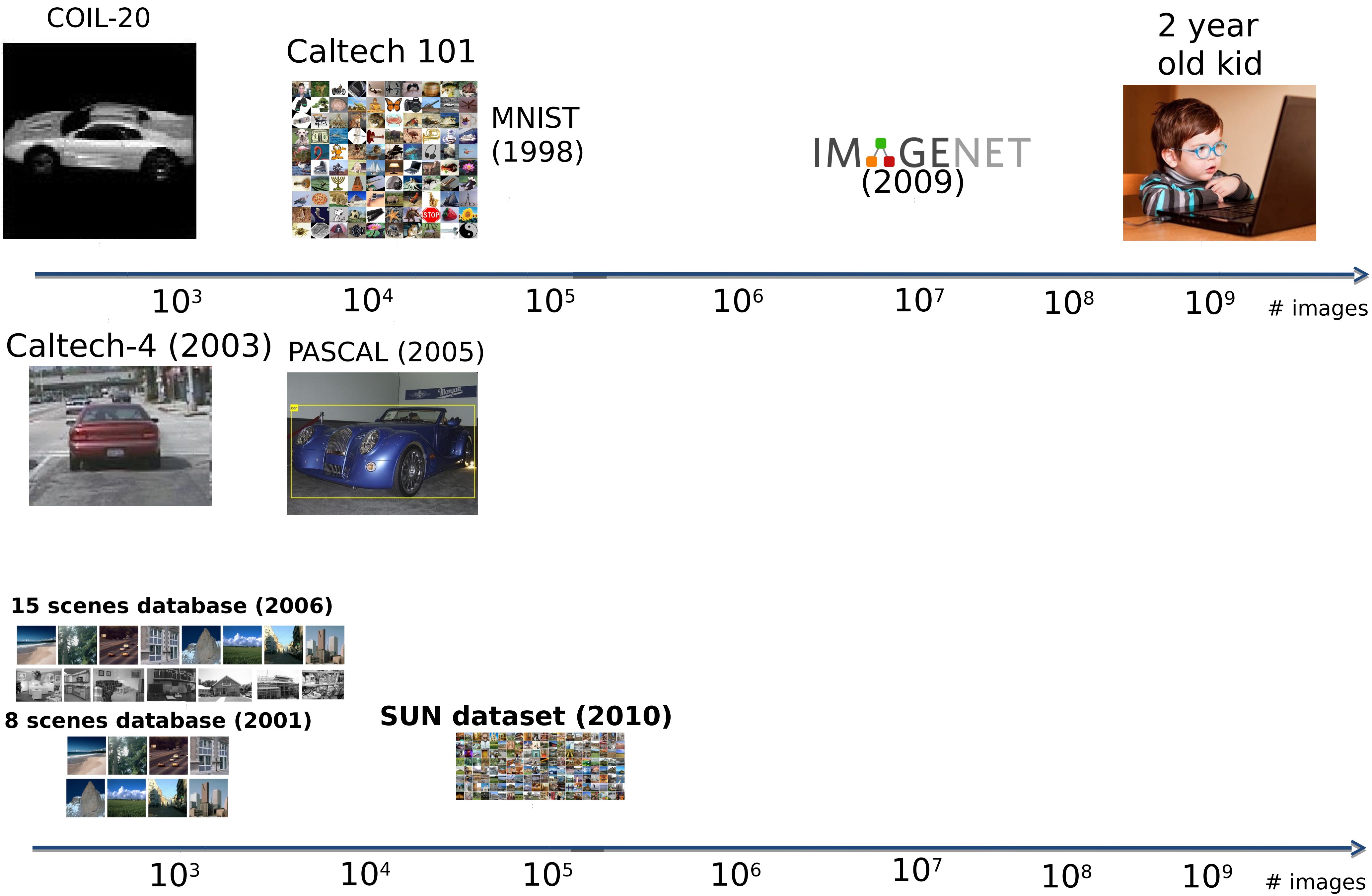

Highway


Mountain

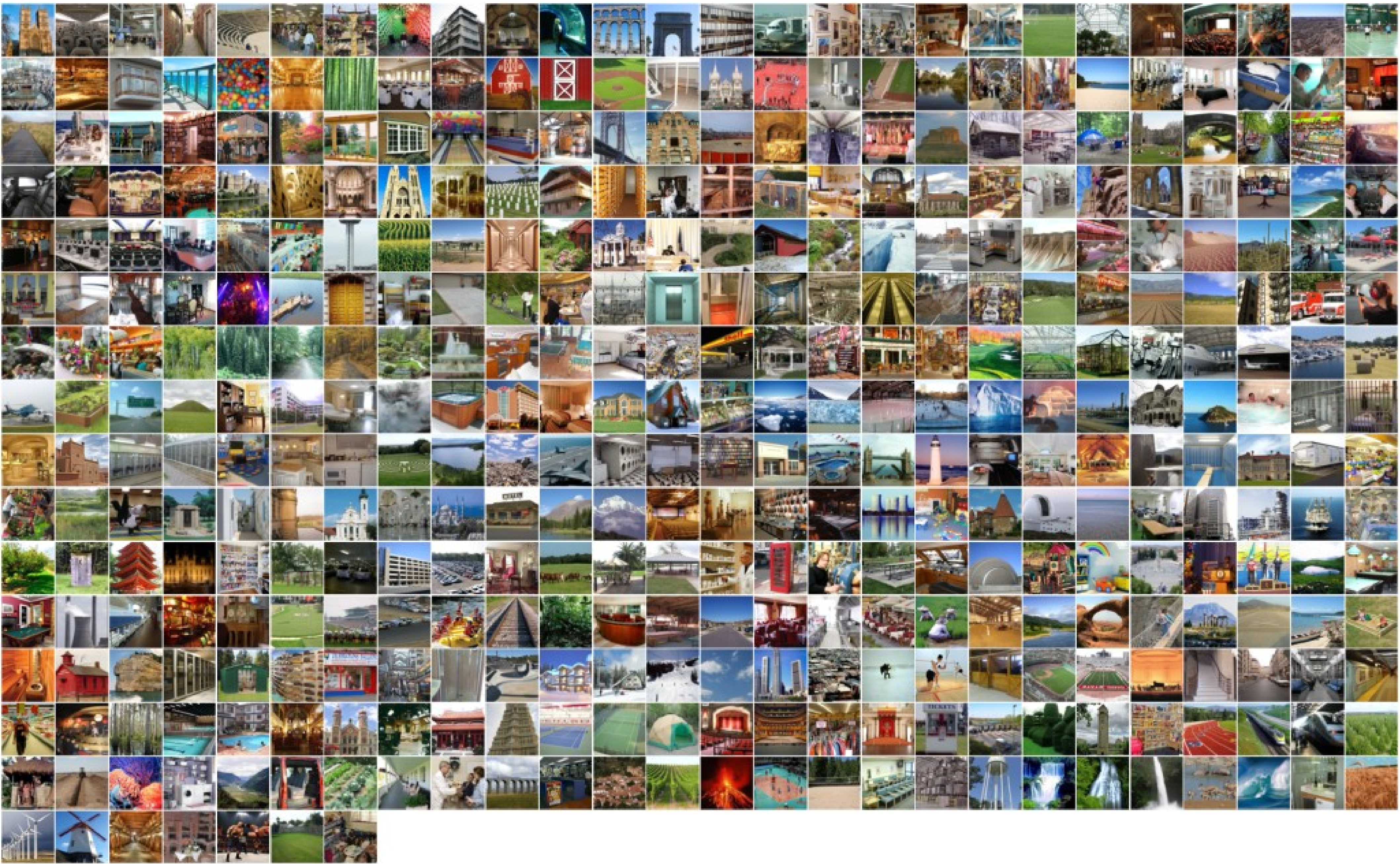

Open country

Oliva & Torralba, 2001, Fei Fei & Perona, 2005, Lazebnik, et al 2006

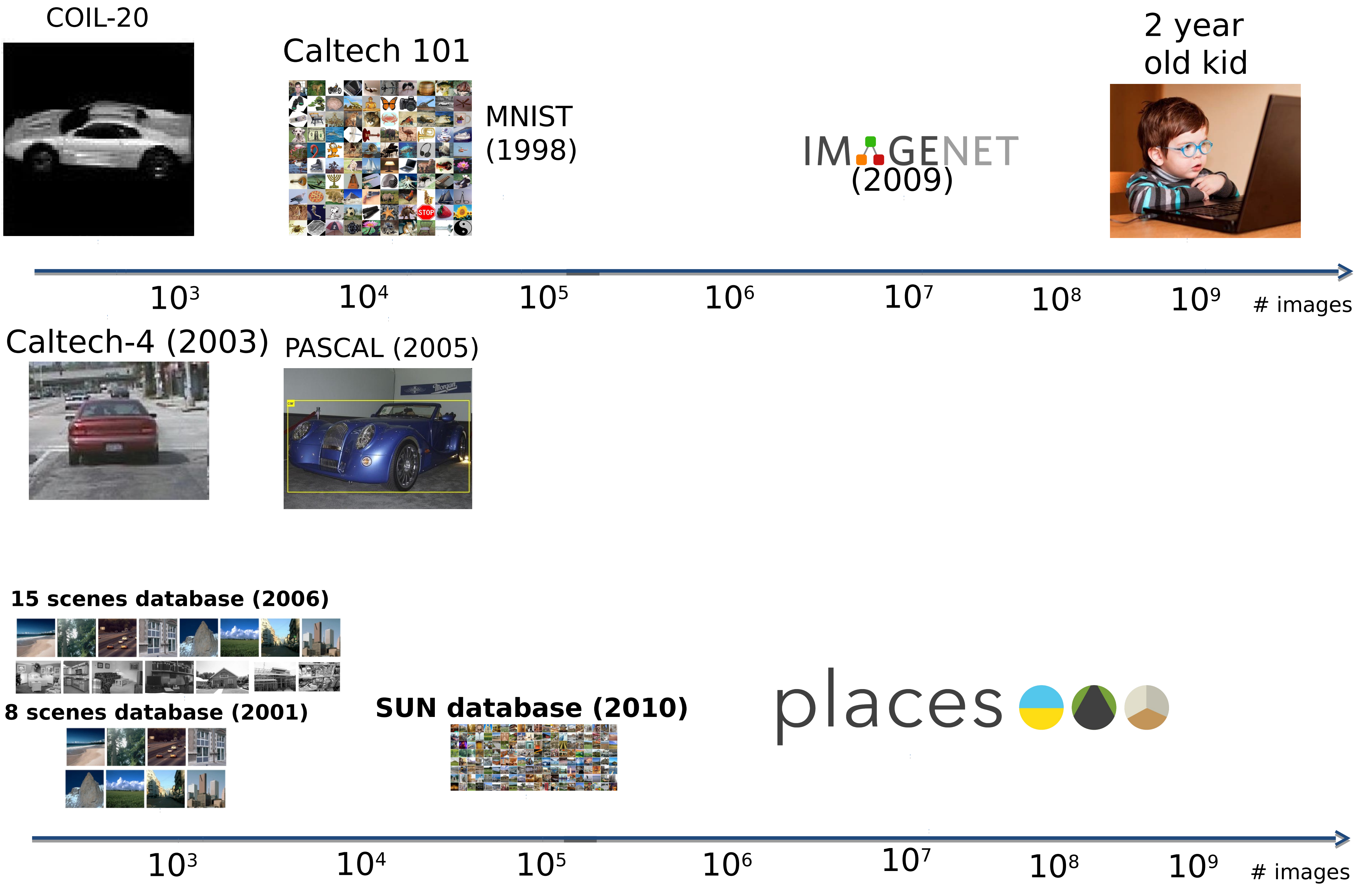
The evolution of scene and object centered databases



SUN dataset
900 Scene categories & 130,000 images



The evolution of scene and object centered databases

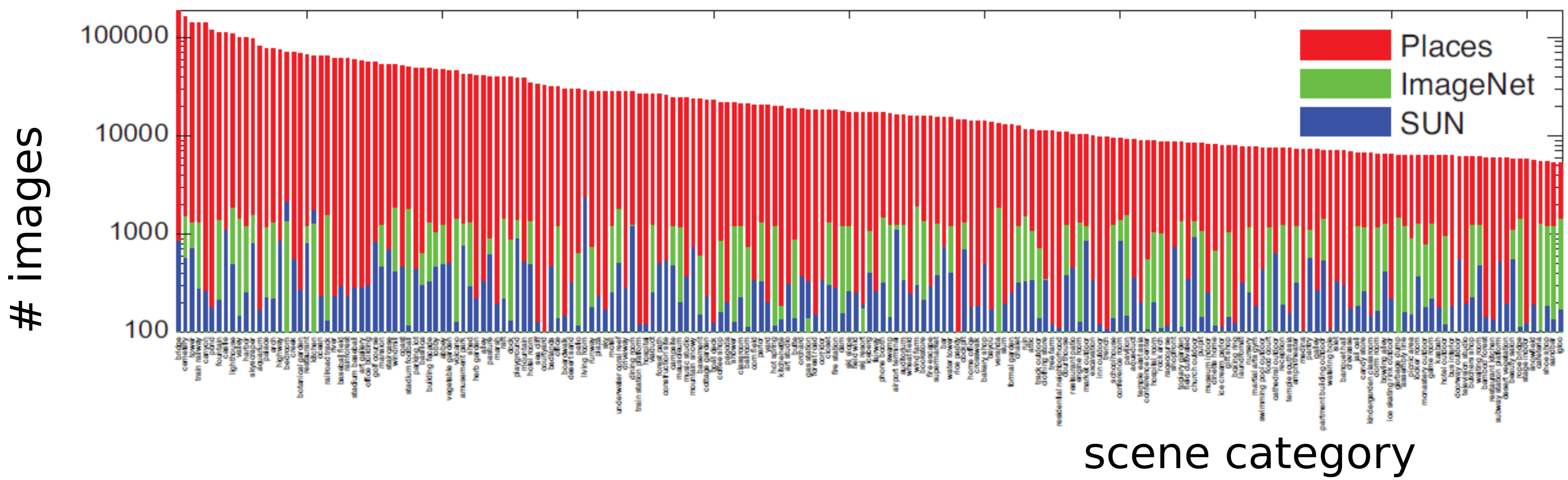


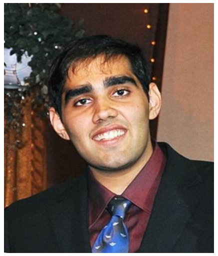
places

Places Database for Scene Recognition

http://places.csail.mit.edu

10 million images from 476 scene categories



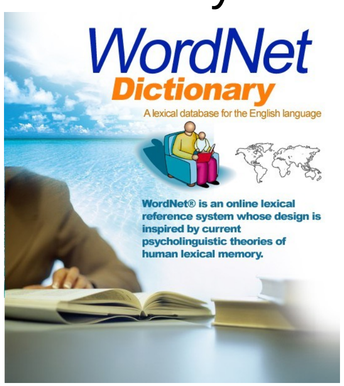



Aditya Khosla, Agata Lapedriza, Aude Oliva, Antonio Torralba

Places Database for Scene Recognition

Dataset building process:

1. Scene words are collected from a dictionary



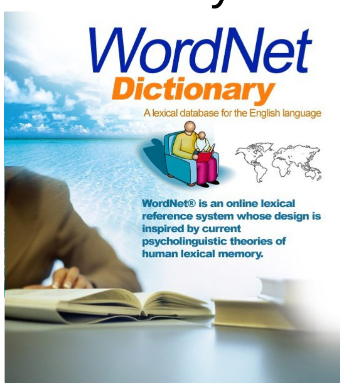
2. Images are queried and downloaded

3. Crowd sourcing annotation

Places Database for Scene Recognition

Dataset building process:

1. Scene words are collected from a dictionary

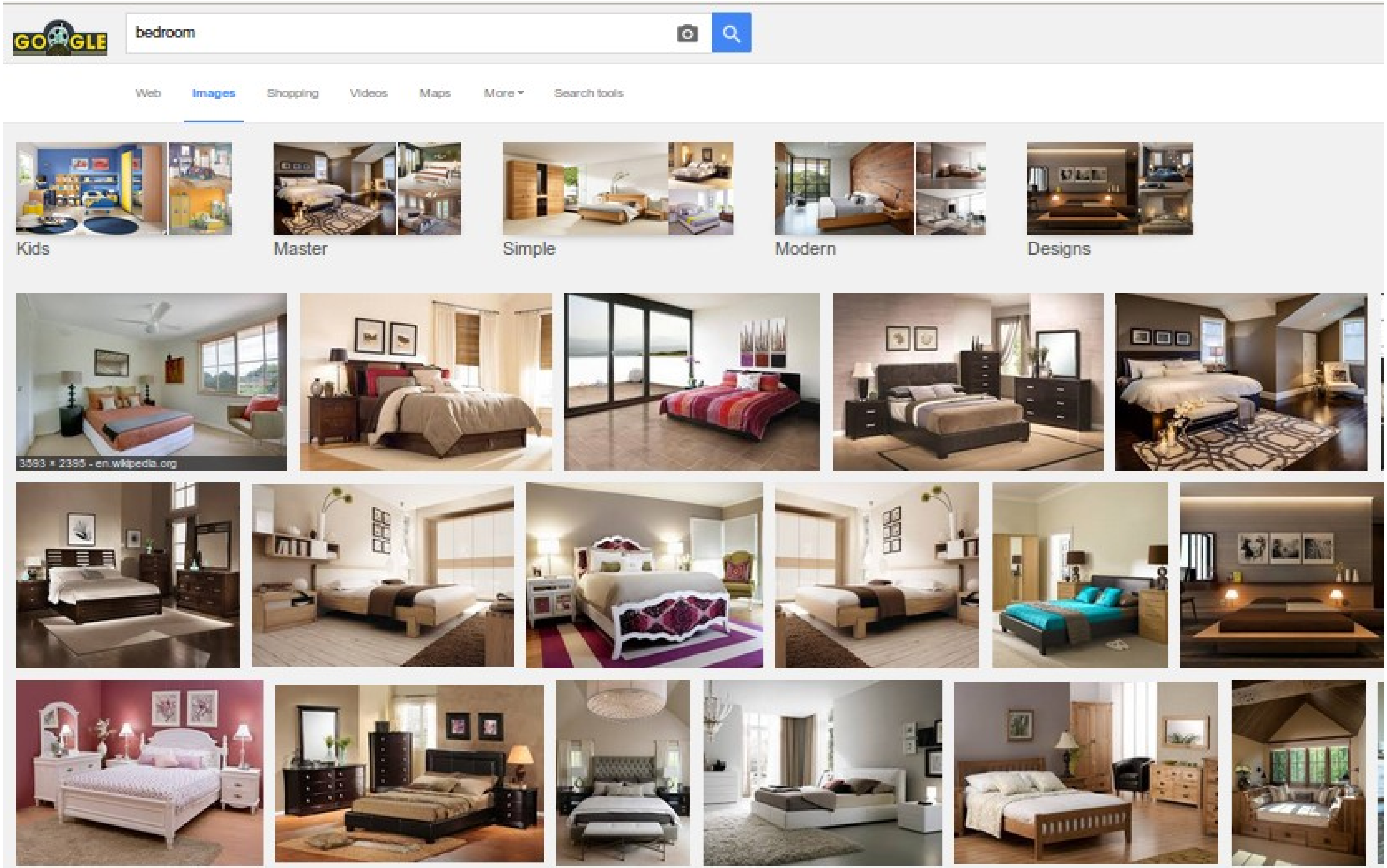


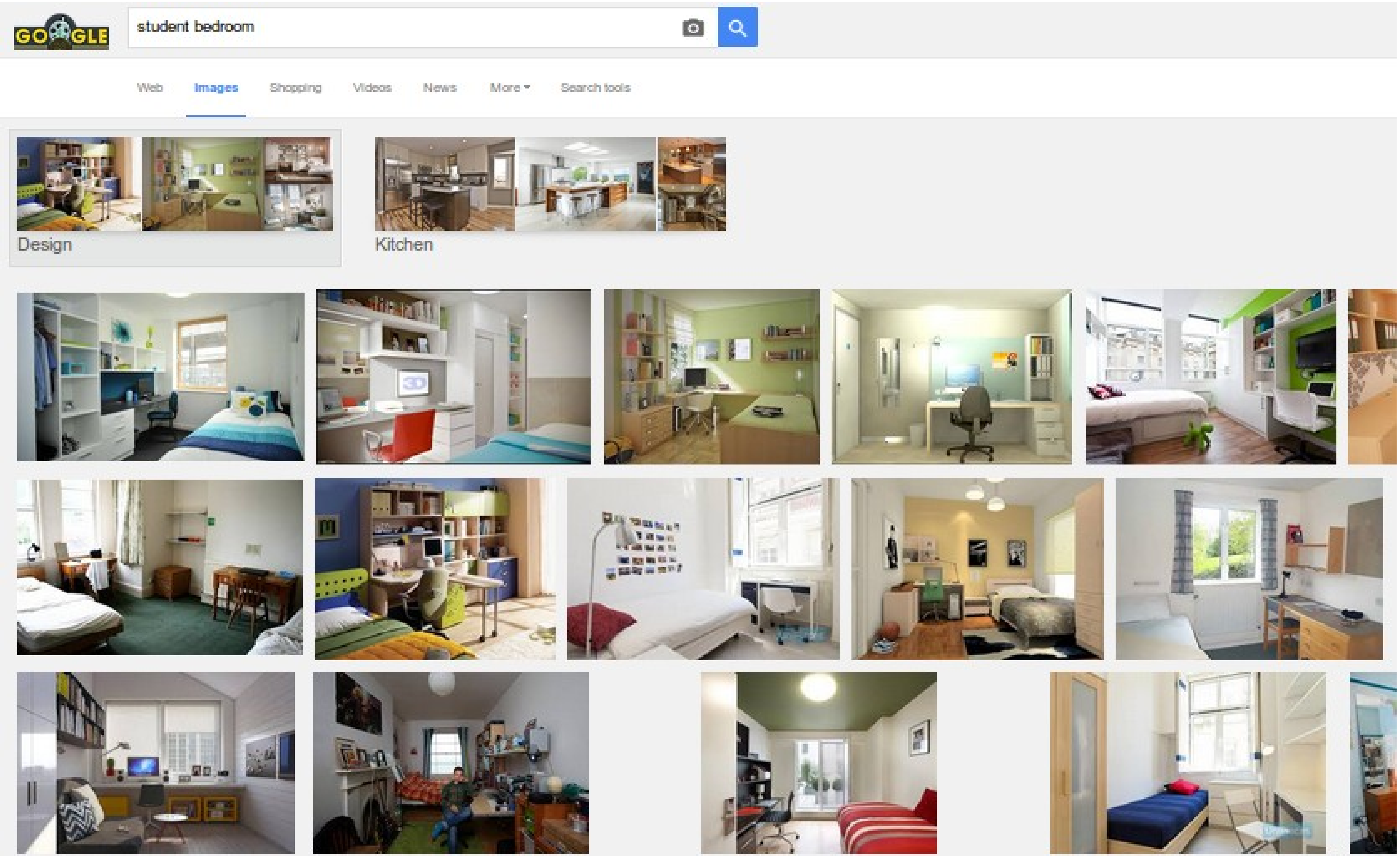
any concrete noun which could reasonably complete the phrase “I am in a place ”, or “Let’s go to the place ”

Places Database for Scene Recognition

Dataset building process:

2. Images are queried and downloaded


bedroom
Web
Images
Shopping
Videos
Maps
More
Search tools
Kids
Master
Simple
Modern
Designs
3593 × 2395 - en.wikipedia.org


student bedroom
Web
Images
Shopping
Videos
News
More
Search tools
Design
Kitchen

In reality, student bedroom should be like this:

900 adjective to improve diversity

abandoned, acceptable, accessible, additional, adjacent, advertised, affordable, air-conditioned, alternative, american, amusing, ancient, antique, appealing, appropriate, architectural, asian, astonishing, astounding, attractive, austere, authentic, available, average, awesome, beautiful, beguiling, beloved, best, better, better-known, big, bigger, biggest, bizarre, black, black-and-white, bland, boring, breezy, brick-built, bright, brighter, brightest, brilliant, broken, busiest, business-like, bustling, busy, central, centralized, certain, changed, changing, charming, cheap, cheaper, cheapest, cheerful, cheerless, cheery, cherished, chilling, chilly, civilized, classic, classical, clean, cleaner, clear, clearer, clinical, closer, closest, closing, cloudy, coastal, cold, coldest, colourful, comfortable, comforting, comfortless, comfy, common, comparable, comparative, competitive, complementary, complete, complex, complicated, concealed, conceivable, confined, considerable, contemporary, cool, coolest, cosmopolitan, cost-effective, cosy, cozy, cream-white, creative, crowded, cultivated, cultural, current, damp, dangerous, dark, darkened, darker, darkest, decorative, delightful, designated, designed, desirable, desired, desolate, desolated, different, difficult, dilapidated, dim, dimly-lit, dingy, dirty, disadvantageous, disorderly, do-it-yourself, domestic, double, double-fronted, double-length, downtown, drab, dreadful, driest, dry, dual, dull, duller, dullest, dusty, early, economic, economical, elegant, embarrassing, empty, enormous, especial, european, everyday, exciting, exemplary, exotic, exterior, external, extraordinary, extravagant, familiar, famous, fancy, fantastic, far-away, fascinating, fashionable, fashioned, favourable, fictional, fictitious, filmed, filthy, fine, foggy, foreign, formal, fractured, friendly, frightening, frightful, frosty, frozen, frustrating, full, funny, furnished, fuzzy, gaudy, ghastly, ghostly, glamorous, glassy, glazed, glittering, gloomy, glorious, glossy, godlike, gold-plated, good, gorgeous, graceful, gracious, grand, gray, great, greatest, green, greener, grey, grisly, gruesome, habitable, habitual, handy, happy, harmonious, harrowing, harsh, hazardous, healthful, healthy, heart-breaking, heart-rending, heavy, hideous, hiding, higgledy-piggledy, high, hilarious, historic, historical, holiest, home, horizontal, hospitable, hostile, hot, huge, humid, idyllic, illegal, imaginary, immaculate, immense, imminent, immortal, impassable, impassioned, impersonal, important, impossible, impressive, improbable, improper, inauspicious, inconceivable, inconvenient, incredible, independent, individual, indoor, industrial, ineffable, inexpensive, informal, inhabited, inhospitable, initial, innovatory, innumerable, insecure, insignificant, inspiring, integrated, intentional, interesting, intermediate, internal, international, intimidating, intriguing, inviting, irrational, irregular, isolated, joint, joyful, key, known, large, large-scale, largest, less-favored, lesser, licensed, lifeless, light, limited, little, little-frequented, little-known, lively, living, local, lofty, logical, lone, long, long-awaited, long-forgotten, long-inhabited, long-netting, long-stays, long-term, lost, lousy, lovely, low, low-ceilinged, low-cost, low-energy, lower, lucky, luxury, magical, magnificent, main, majestic, major, marginal, marine, marvellous, massive, masterful, maximum, mean, meaningless, mechanised, medieval, mediocre, medium-sized, melancholy, memorable, messy, middle, middle-order, mighty, miniature, minor, miserable, missing, misty, mixed, modern, moist, mouldy, mountainous, moving, muddy, multi-functional, multiple, mundane, murky, musty, muted, mysterious, mysterious-looking, mystic, mystical, mythic, naff, named, nameless, narrow, national, native, natural, naturalistic, nearby, neat, necessary, neglected, neighboring, new, nice, night-time, nineteenth-century, noisy, nondescript, normal, northern, notable, notorious, numerous, odd, odorous, official, old, only, open, open-air, operatic, orderly, ordinary, organic, original, ornamental, out-of-homes, out-of-the-way, outdoor, outlying, outside, outstanding, over-crowded, overgrown, overwhelming, paid, painful, painted, palatial, pastoral, peaceful, peculiar, perfect, periodic, peripheral, permanent, permitted, personal, petty, pictorial, picturesque, pitiful, placid, plain, planted, pleasant, pleasing, poisonous, poor, popular, populated, populous, positive, possible, post-war, posterior, postmodern, potential, powerful, practical, pre-arranged, pre-eminent, precise, predictable, present, present-day, preserved, pretty, previous, pricey, primal, prior, private, privileged, probable, professional, profitable, promising, proven, public, pure, queer, quiet, rainy, rare, real, realistic, reasonable, rebuilt, recent, recognized, recommended, reconstructed, recreated, recurring, red, red-brick, redundant, refused, regional, regular, related, relative, relaxing, relevant, reliable, religious, remaining, remarkable, remote, rented, representative, reputable, required, reserved, residential, respectable, respected, restful, restless, restricted, retail, rich, ridiculous, right, rigid, river-crossing, rocky, romantic, rural, sacred, sad, safe, salubrious, satisfying, scary, scattered, scenic, scientific, secondary, secret, secured, selected, senior, separated, serious, sexy, shiny, shocking, shoddy, short-term, significant, silent, silly, similar, simple, single, sizable, slack, small, smelly, smoke-free, smoking, snowy, sobering, soft, solid, sombre, soothing, sophisticated, sorrowful, sound-filled, southern, spare, spatial, special, specialized, spectacular, sporting, stable, standard, static, steady, stifling, strange, stressful, striking, stunning, stupendous, stupid, stylish, successful, sufficient, sunny, super, superb, superior, surrealistic, suspicious, symbolic, teenage, terrible, terrific, theoretical, thrilling, thriving, tidier, tight, tiny, tough, tragic, unattractive, unbelievable, uncertain, unchanging, uncharted, uncivilized, uncomfortable, unconventional, underground, underwater, undisturbed, uneven, unexpected, unfamiliar, unforgettable, unfriendly, unhappy, unhealthy, unimportant, unknown, unnatural, unnecessary, unparalleled, unpleasant, unsafe, unseemly, unsuitable, unusual, upmarket, urban, vague, valuable, varied, various, vertical, very, vibrant, virtual, visual, vital, vivid, voluntary, vulgar, vulnerable, wacky, waiting, warm, wealthy, weeping, weird, weird-looking, well-assured, well-defended, well-designed, well-hidden, well-insulated, well-known, well-lit, well-loved, well-ordered, well-organized, well-secured, well-sheltered, well-used, wet, white, whole, wicked, wide, widespread, wild, windy, wintering, wonderful, wondrous, wooded, wordless, working, worldly, worldwide, worst, worthwhile, worthy, wretched, wrong, young, yucky,

Places Database for Scene Recognition

Dataset building process:

3. Crowd sourcing annotation

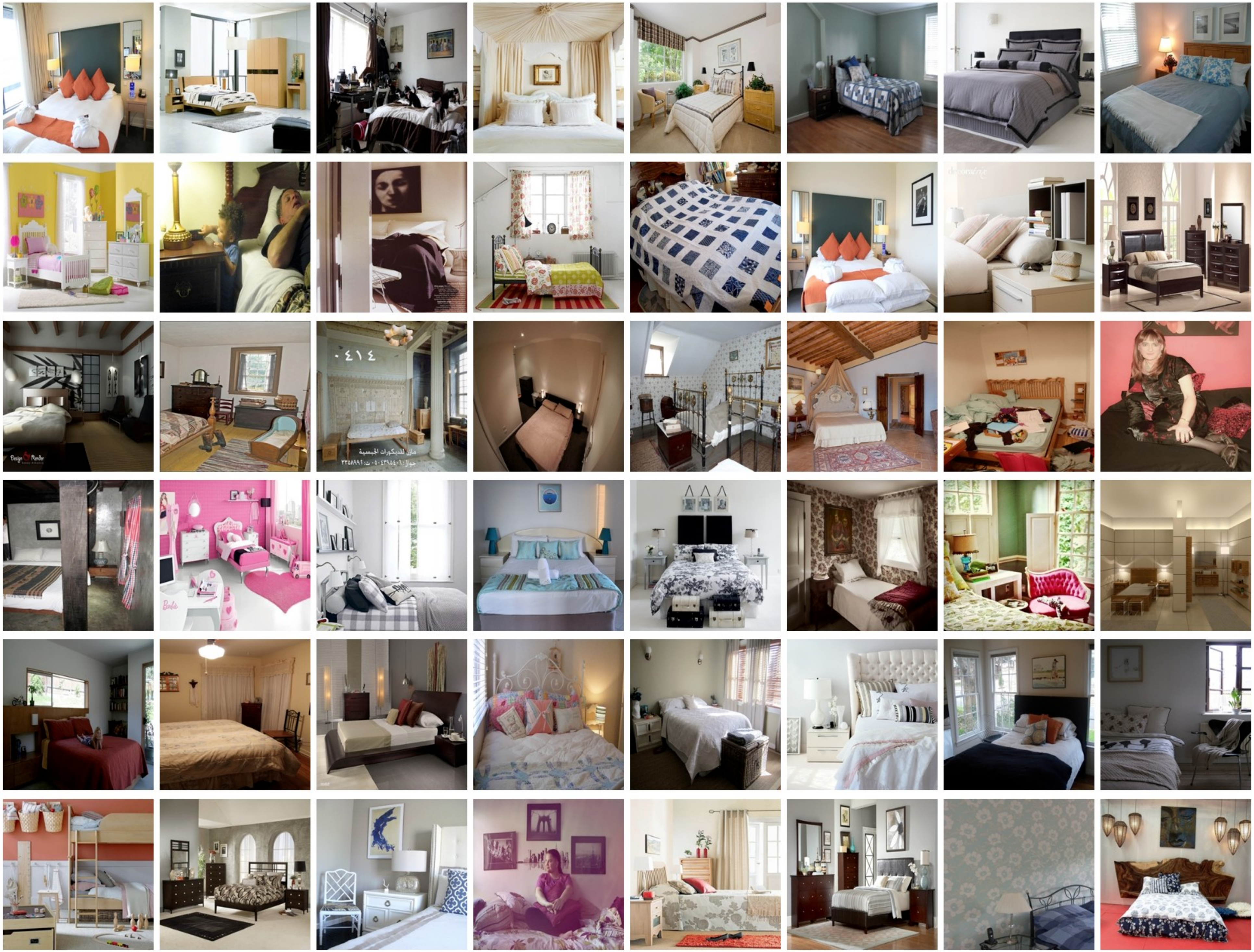
AMT workers get paid to annotate the images

Annotation Interface

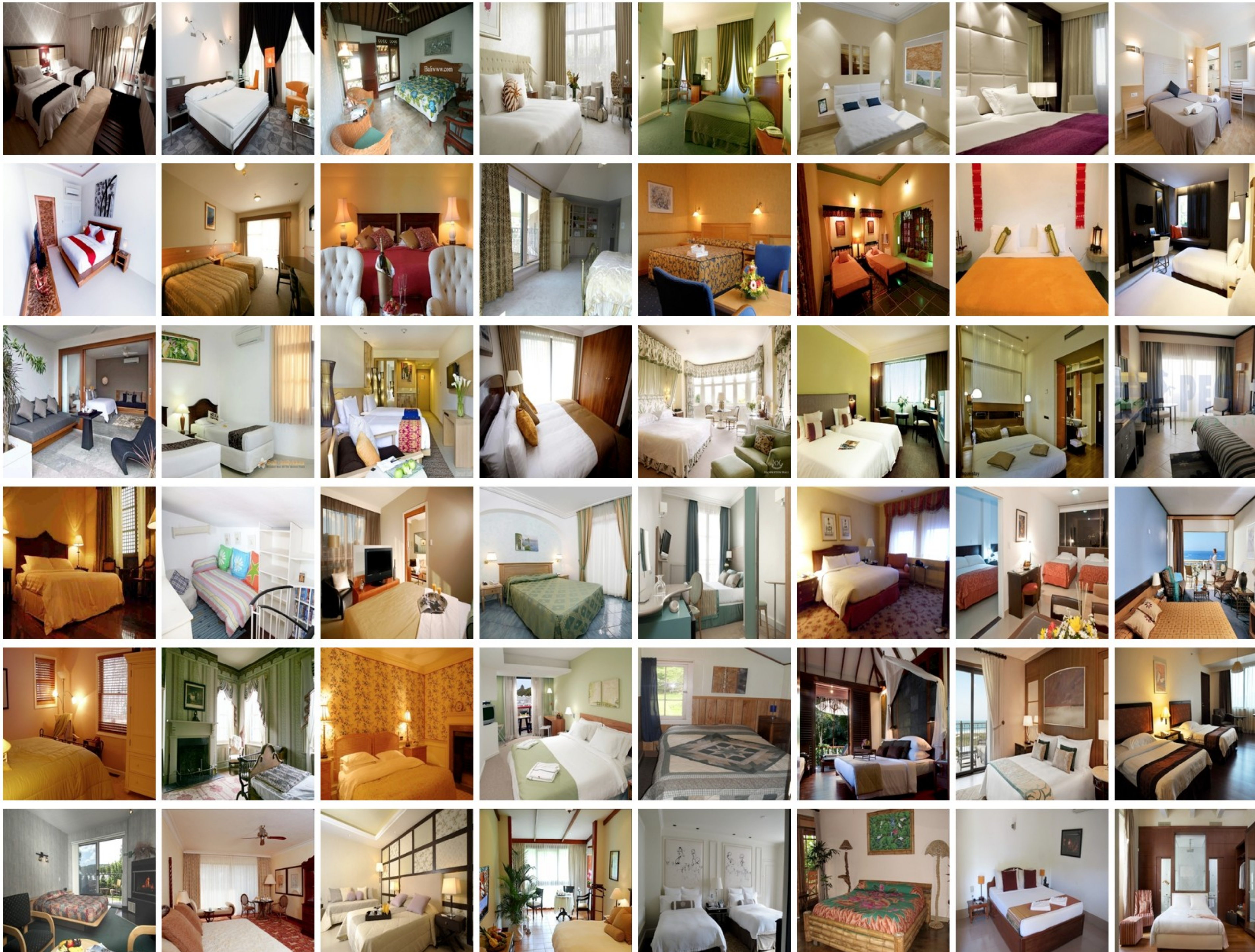
Annotation Interface
1st round

Annotation Interface: 2nd round

simple bedroom:476

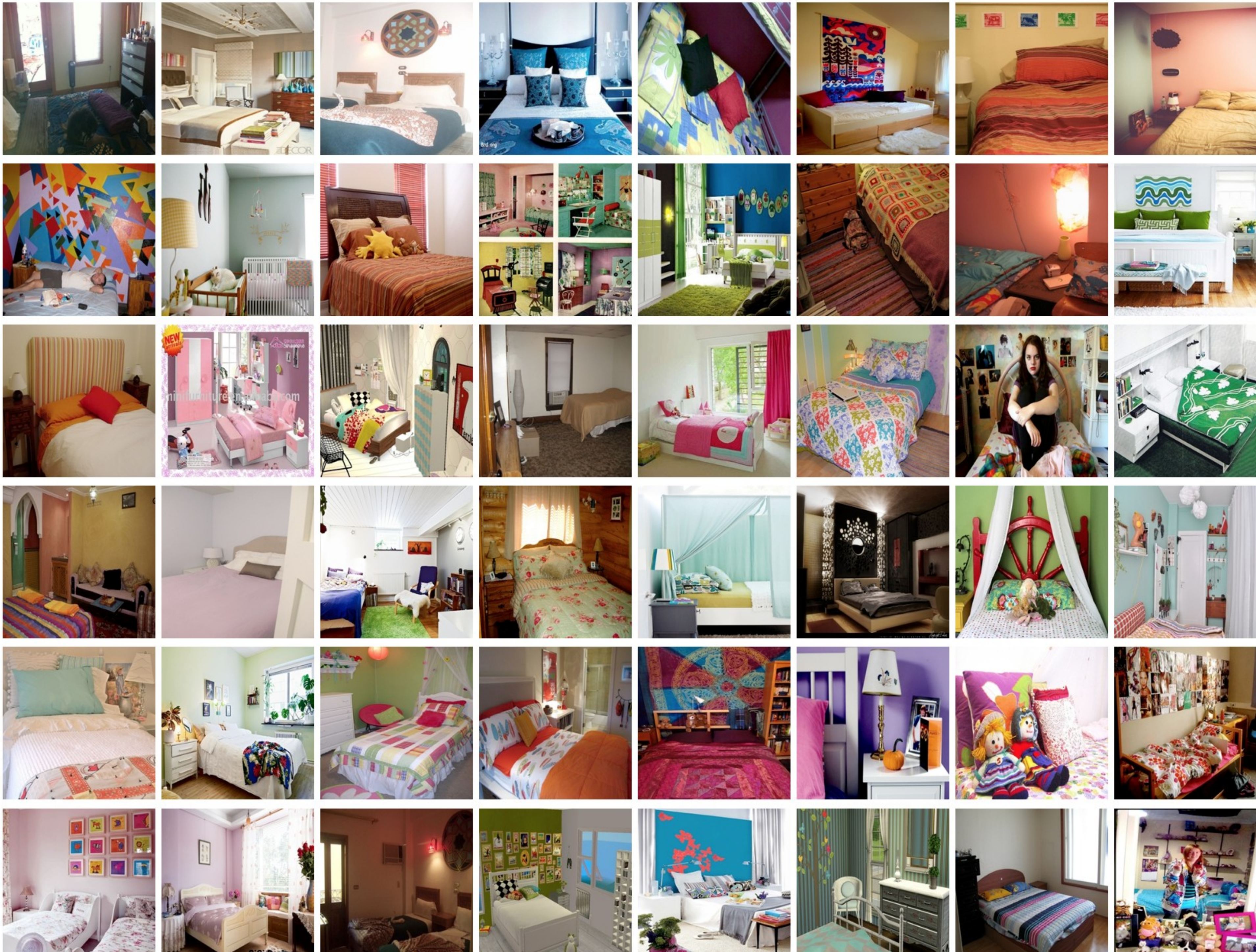


superior bedroom:423



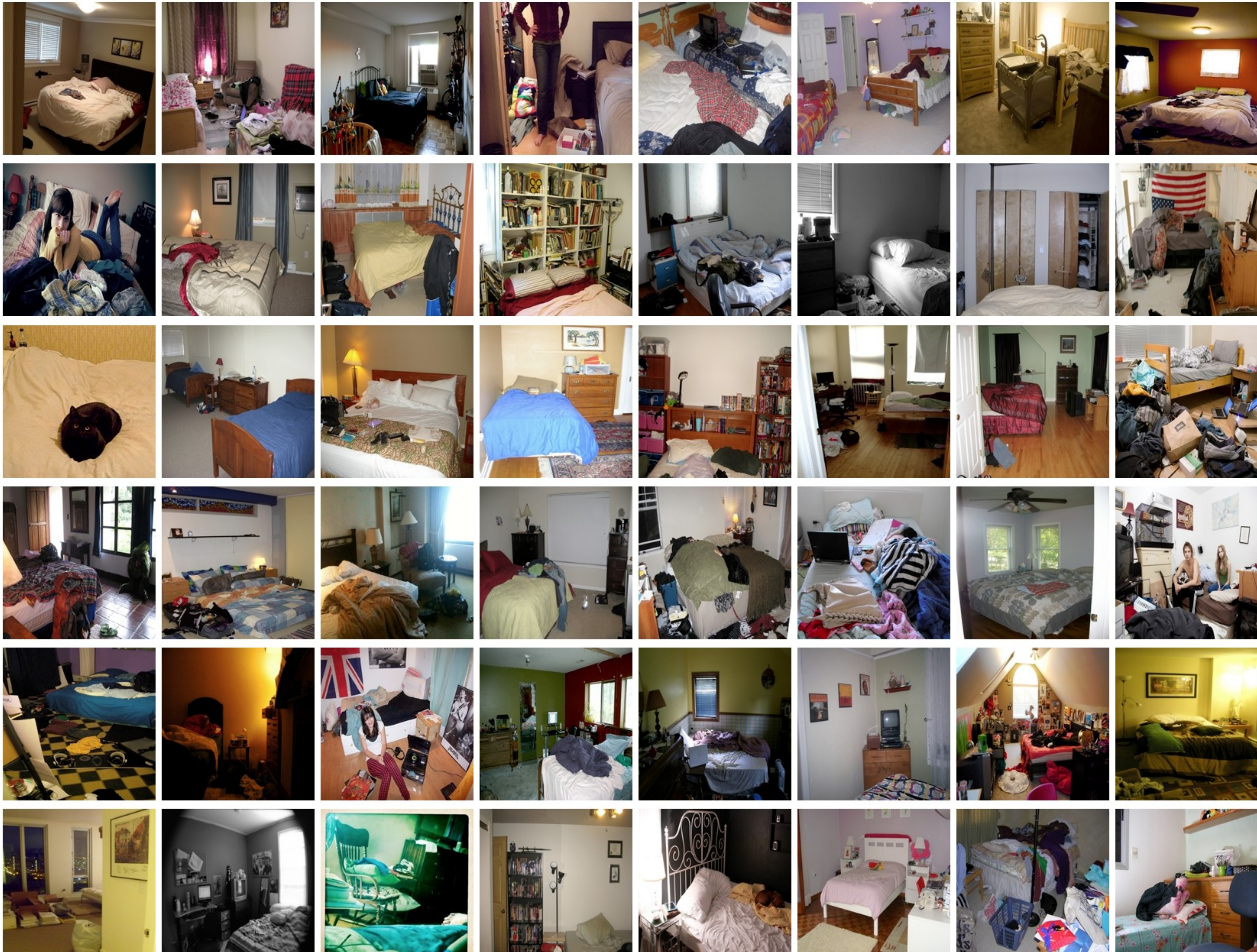
senior bedroom:319

colourful bedroom:209



cleaner bedroom:205

messy bedroom:808

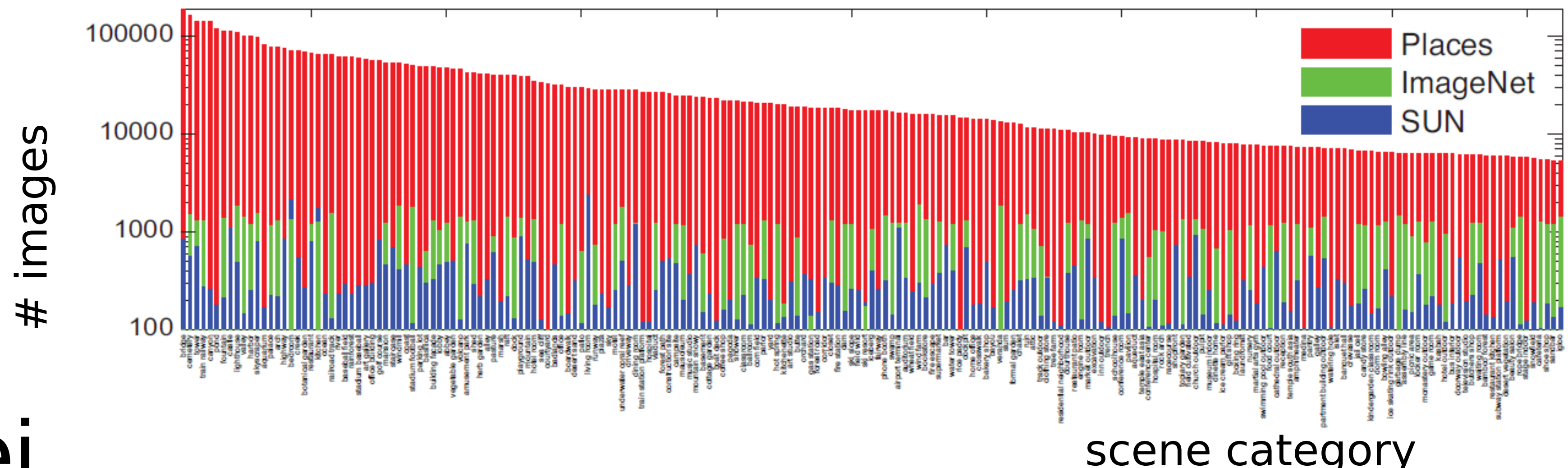


places

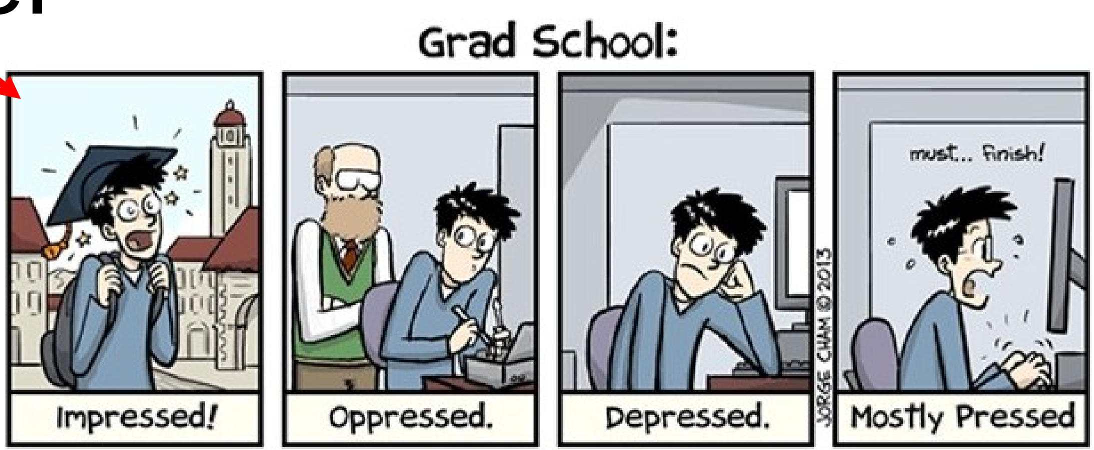
Places Database for Scene Recognition

http://places.csail.mit.edu

10 million images from 476 scene categories



Bolei



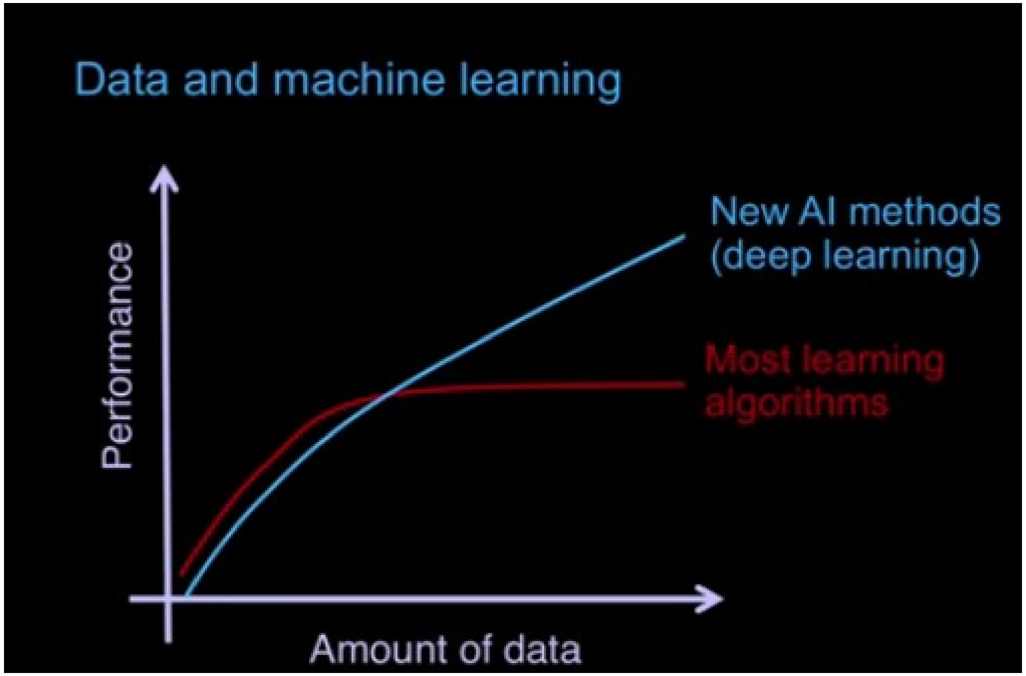
+Ten thousands of anonymous AMT workers

amazon mechanical turk

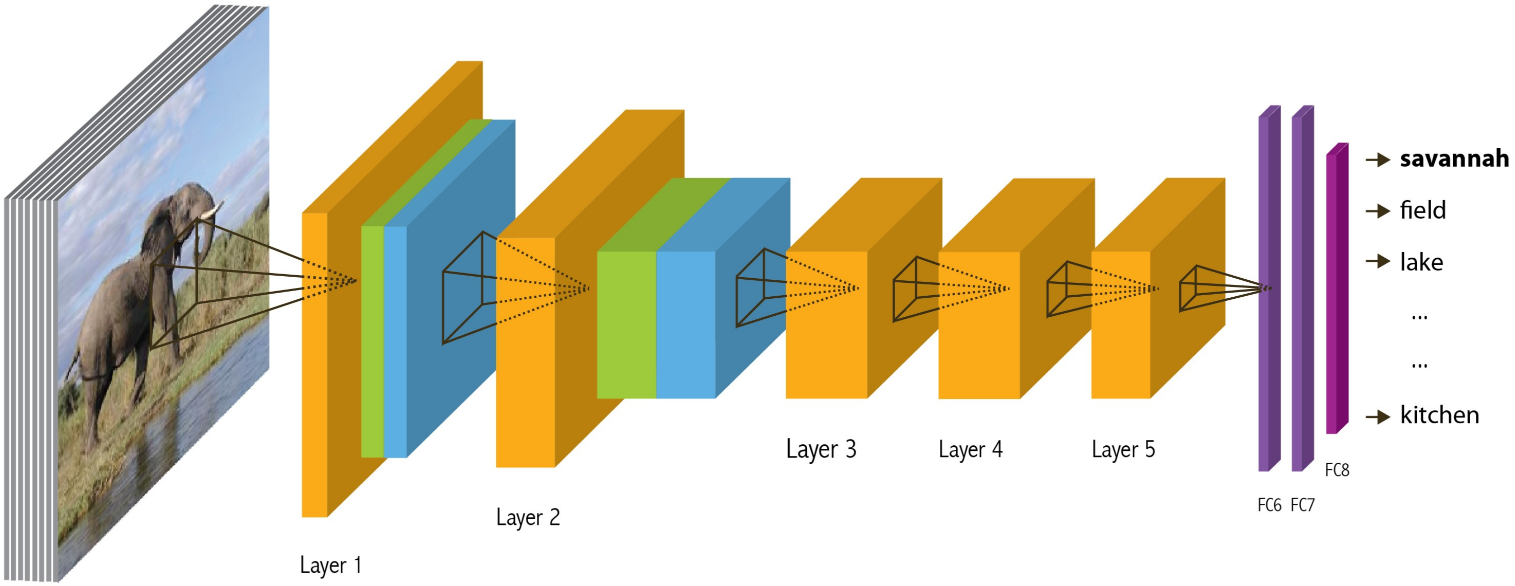
More than one year of time!

How to train with million of images

Traditional machine learning algorithm cannot handle large-scale data



Training CNN on Places Database



AlexNet CNN: 5 conv layers + 2 FC layers + 1 softmax layer

Imagenet classification with deep convolutional neural networks. NIPS'12

Training CNN on Places Database

We train AlexNet CNN on 2.5 million images from 205 categories of Places.

- trained on GPU NVIDIA Titan Black for 7 days using Caffe Package.
- 60,000,000 parameters and 630,000,000 connections.

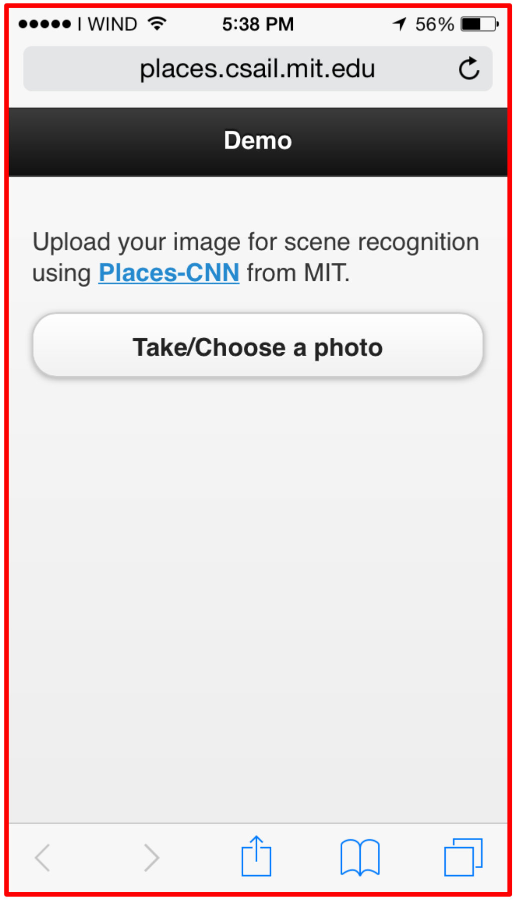
Classification accuracy on the test set of Places 205 and the test set of SUN 205.

	Places 205	SUN 205
Places-CNN	**50.0%**	**66.2%**
ImageNet CNN feature+SVM	40.8%	49.6%

Zhou, Lapedriza, Xiao, Torralba & Oliva (NIPS 2014)

Places-CNN Demo

2675 anonymous users report 77% top-5 recognition accuracy



Predictions:

- **type**: indoor
- **semantic categories:** coffee_shop:0.47, restaurant:0.17, cafeteria:0.08, food_court:0.06,

Predictions:

- **type**: indoor
- **semantic categories:** supermarket:0.96,

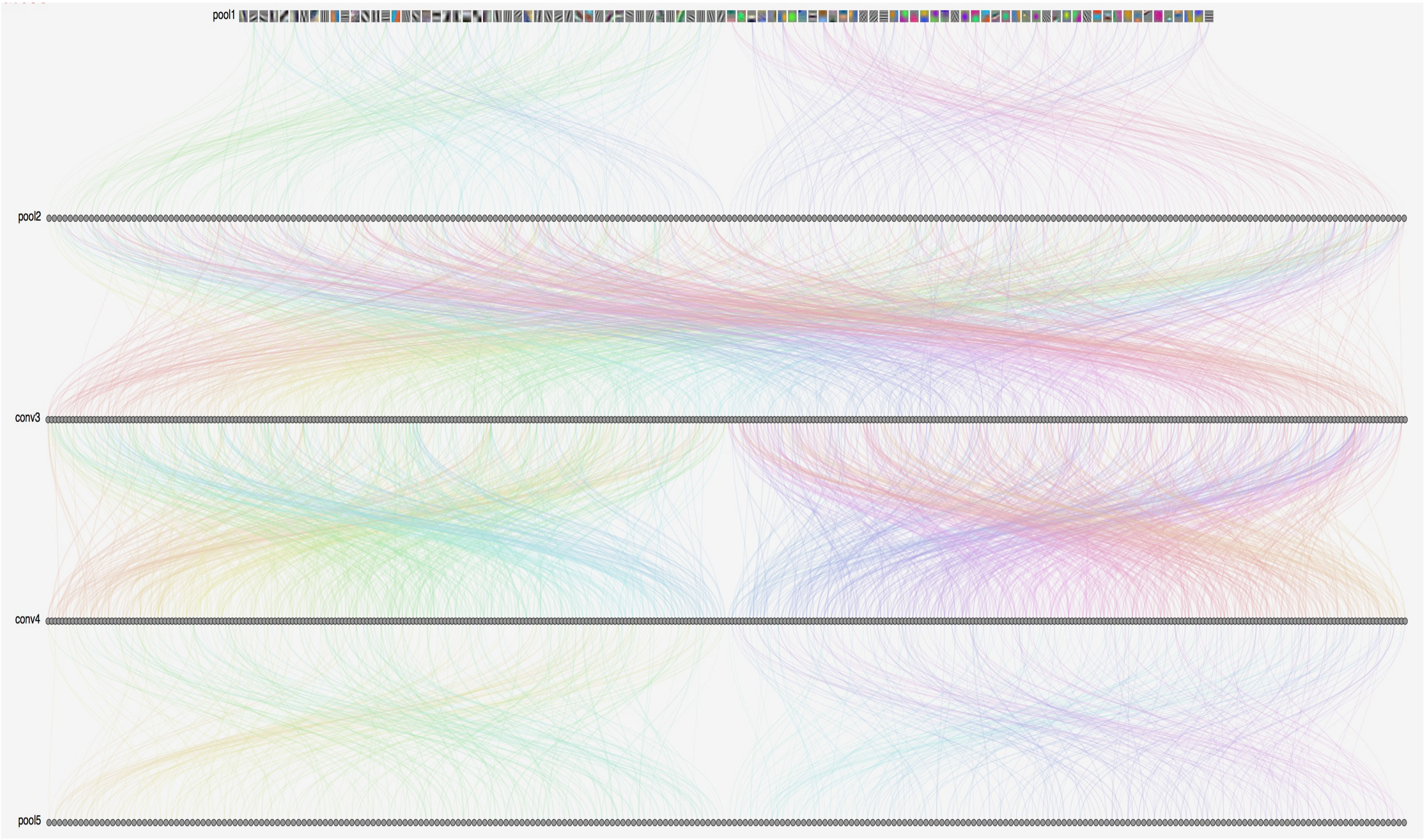
Predictions:

- **type**: indoor
- **semantic categories:** conference_center:0.51, auditorium:0.12, office:0.08,

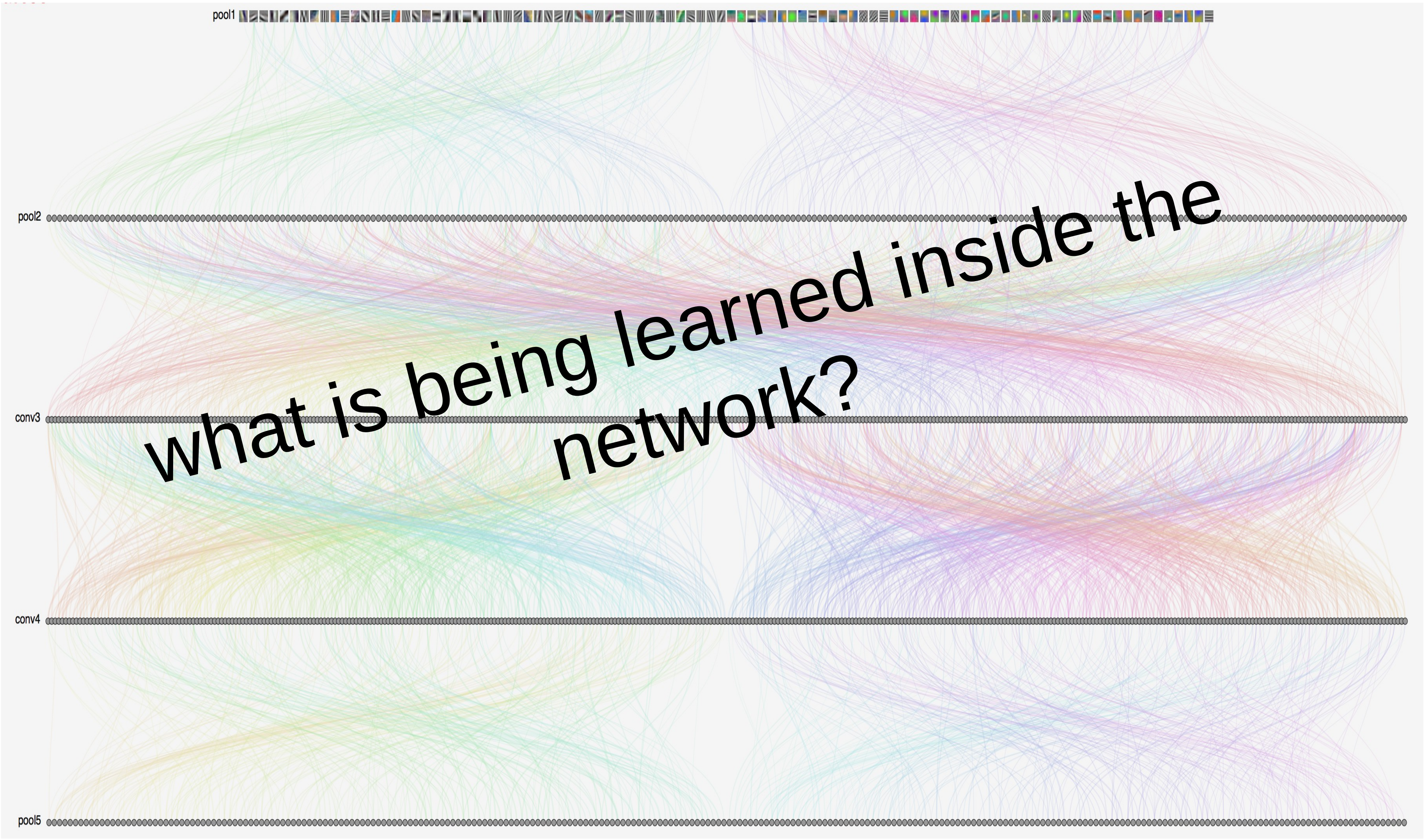
Demo, data, and Places-CNNs could be downloaded at http://places.csail.mit.edu

places
THE SCENE RECOGNITION DATABASE

Analyzing the CNNs



What are all those units doing?



Object Representations in Computer Vision

Part-based models are used to represent objects and visual patterns.

-Object as a set of parts

-Relative locations between parts

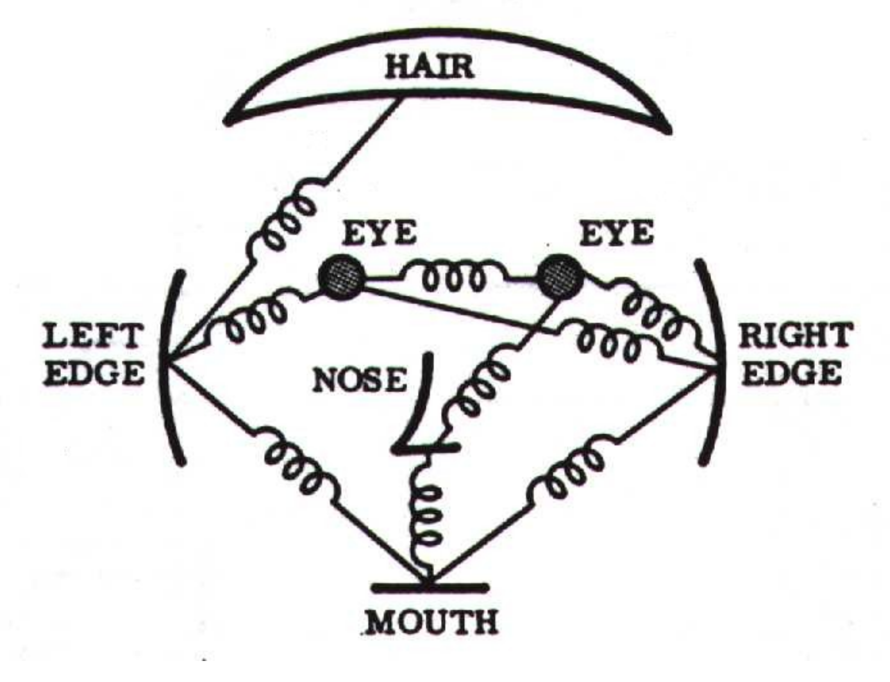


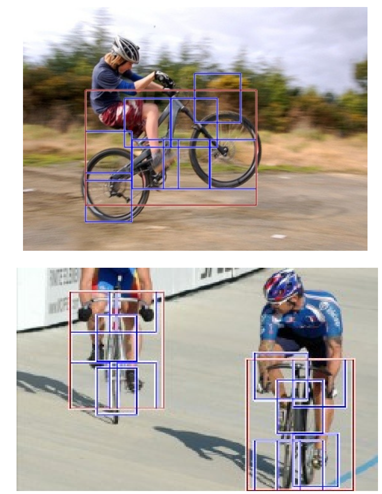
Figure from Fischler & Elschlager (1973)

Object Representations in Computer Vision

Constellation model

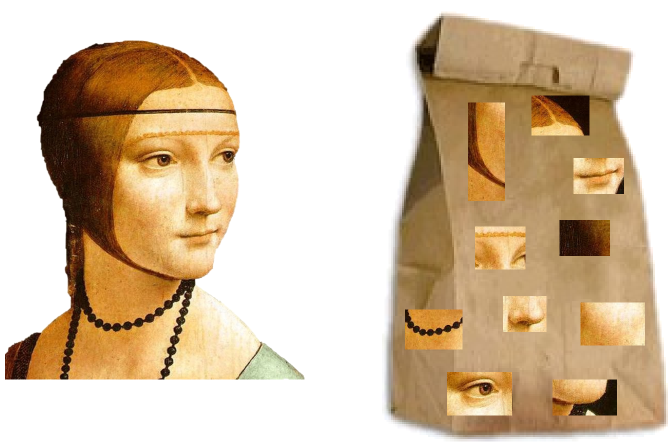
Weber, Welling & Perona (2000),
Fergus, Perona & Zisserman (2003)

Deformable Part model



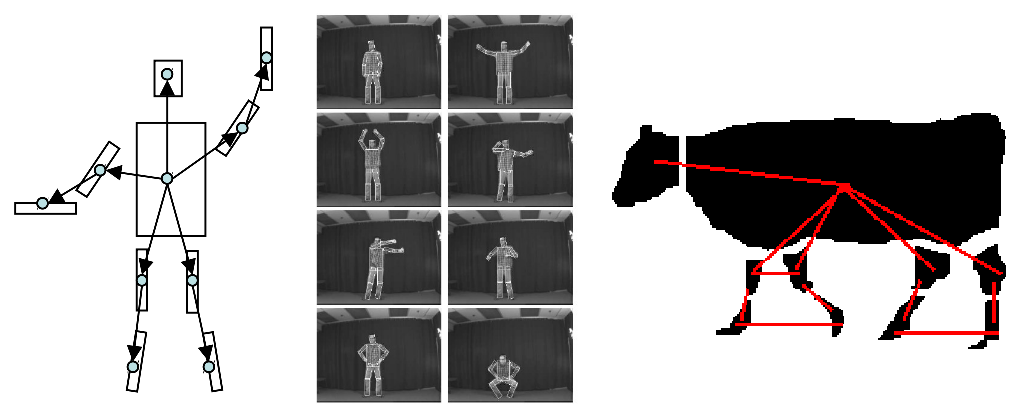
P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan (2010)

Bag-of-word model



Lazebnik, Schmid & Ponce(2003), Fei-Fei Perona (2005)

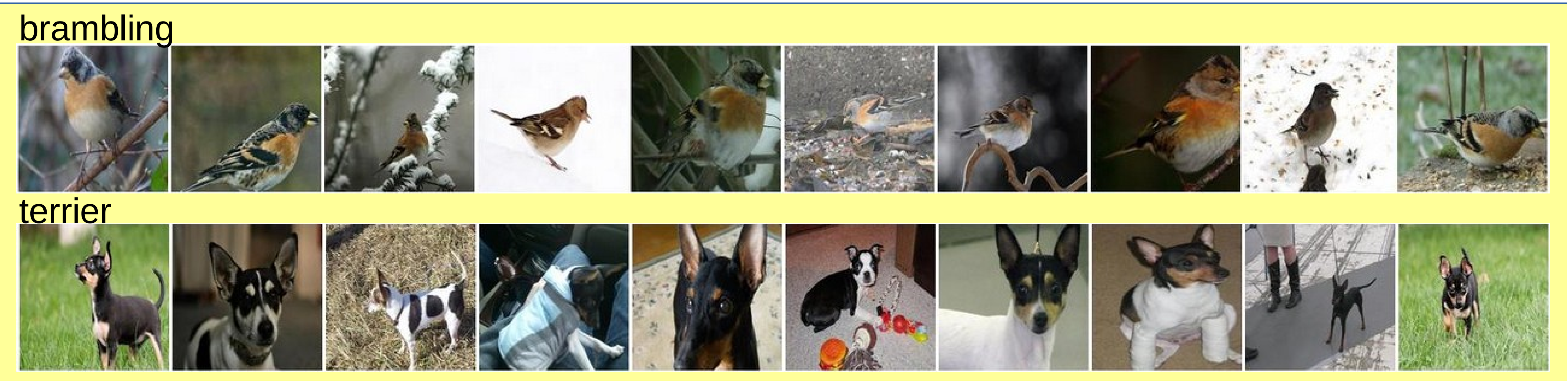
Class-specific graph model



Kumar, Torr and Zisserman (2005), Felzenszwalb & Huttenlocher (2005)

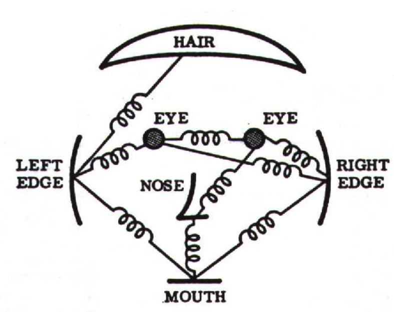
Learning to Represent Objects

IMAGENET



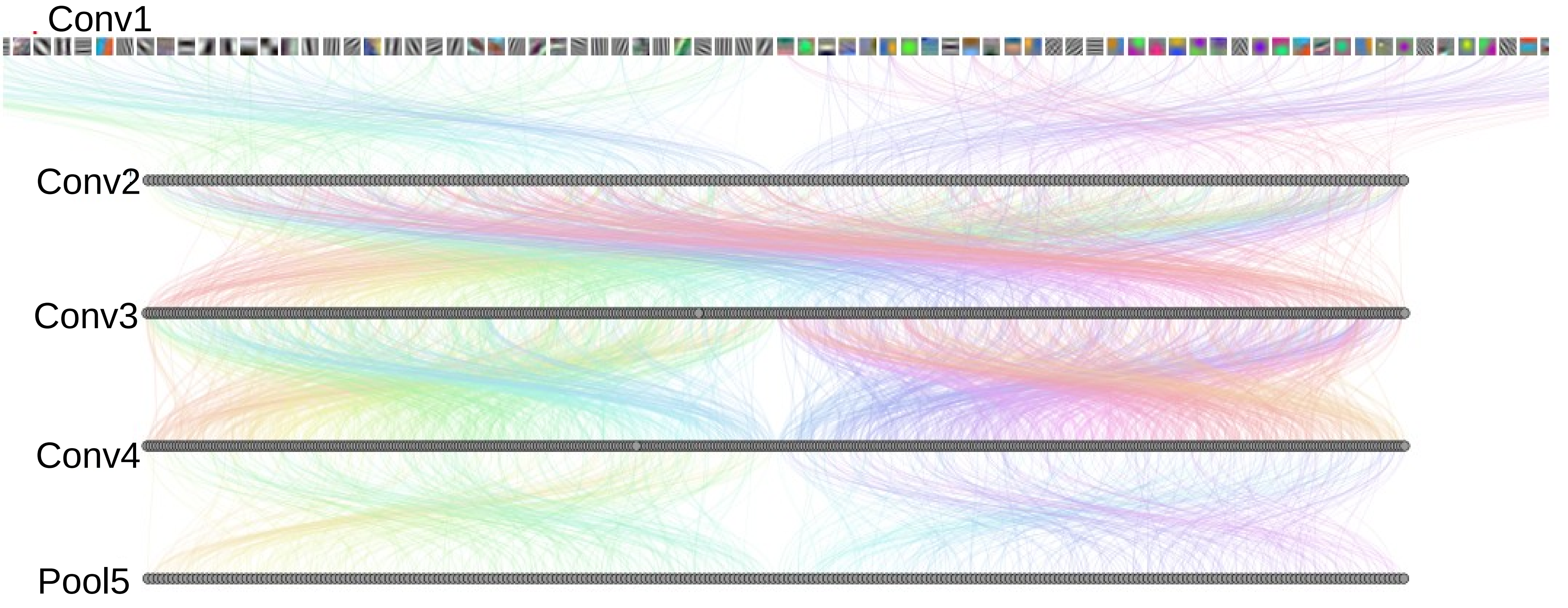
Possible internal representations:

- Object parts
- Textures
- Attributes



How Objects are Represented in CNN?

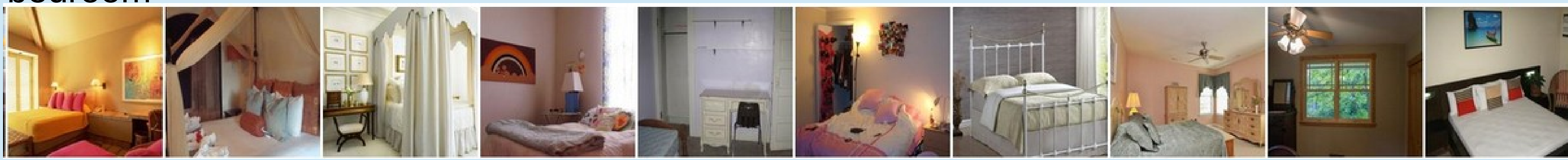
CNN uses **distributed code** to represent objects.



Agrawal, et al. Analyzing the performance of multilayer neural networks for object recognition. ECCV, 2014
Szegedy, et al. Intriguing properties of neural networks.arXiv preprint arXiv:1312.6199, 2013.
Zeiler, M. et al. Visualizing and Understanding Convolutional Networks, ECCV 2014.

Learning to Represent Scenes

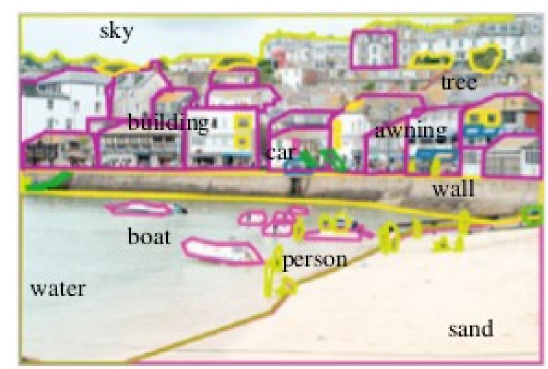
bedroom



mountain

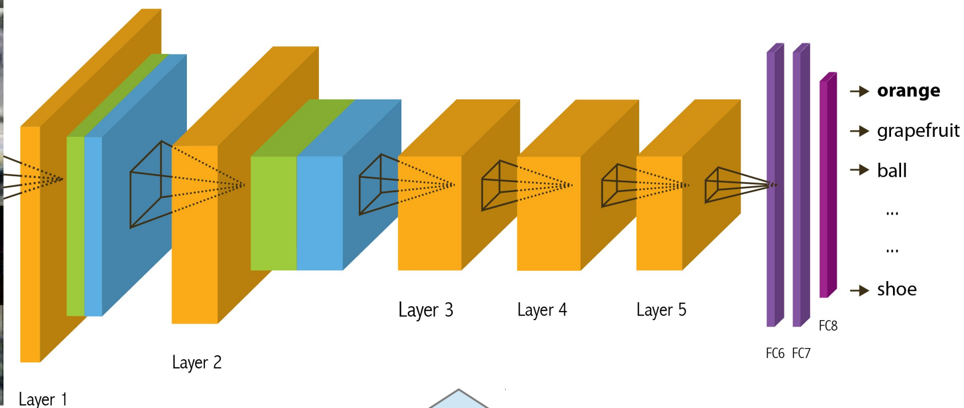
Possible internal representations:

- Objects (scene parts?)
- Scene attributes
- Object parts
- Textures



ImageNet CNN and Places CNN

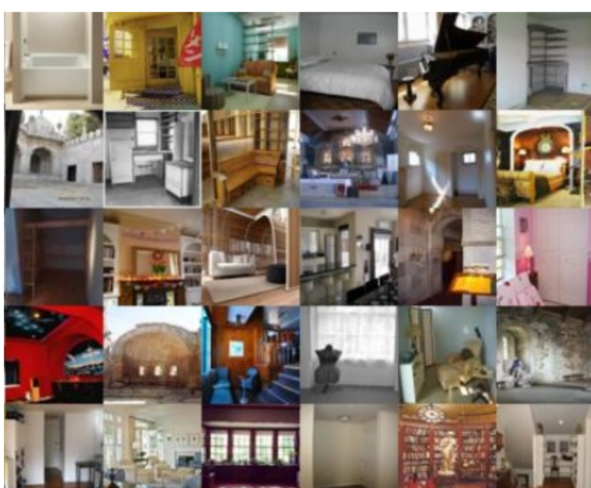
ImageNet CNN for Object Classification

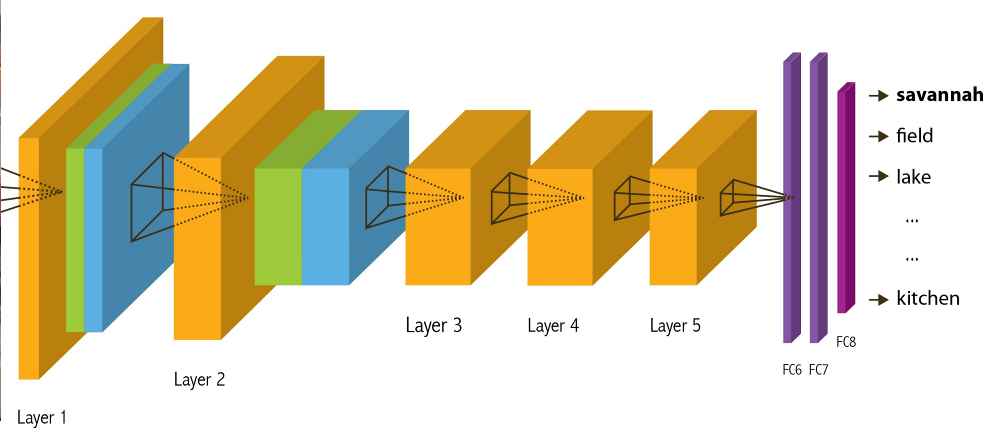


IMAGENET

Same architecture: AlexNet

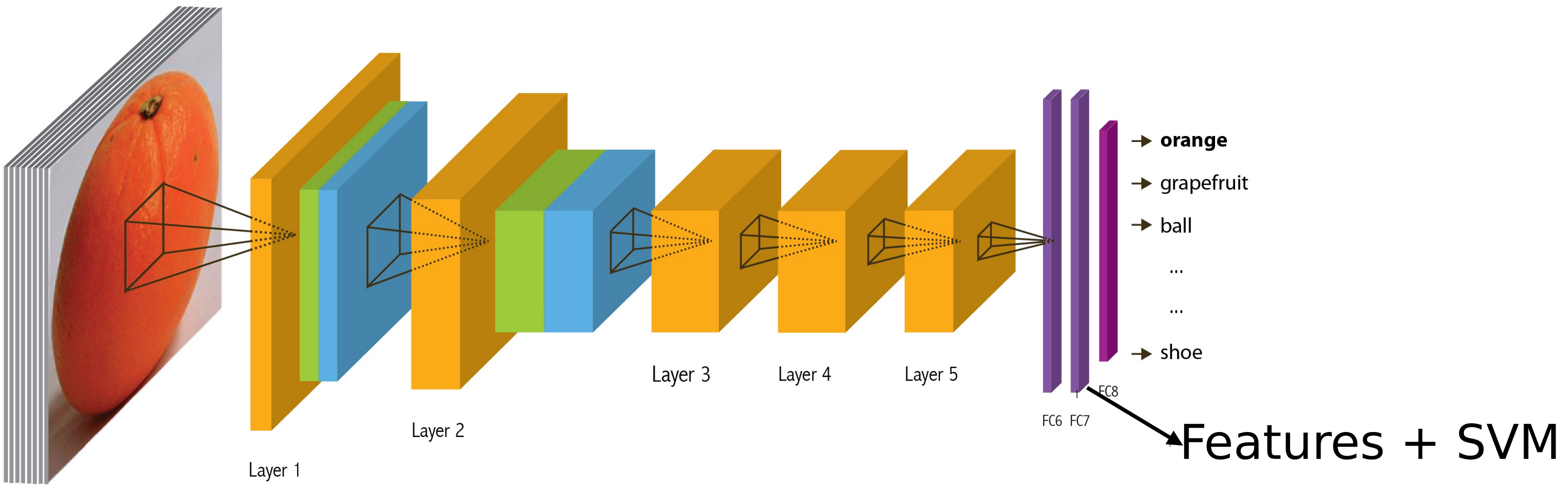
Places CNN for Scene Classification





places
THE SCENE RECOGNITION DATABASE

Generic Visual Feature



Scene datasets

	SUN397	MIT Indoor67	Scene15	SUN Attribute
Places-CNN feature	**54.32±0.14**	**68.24**	**90.19±0.34**	**91.29**
ImageNet-CNN feature	42.61±0.16	56.79	84.23±0.37	89.85

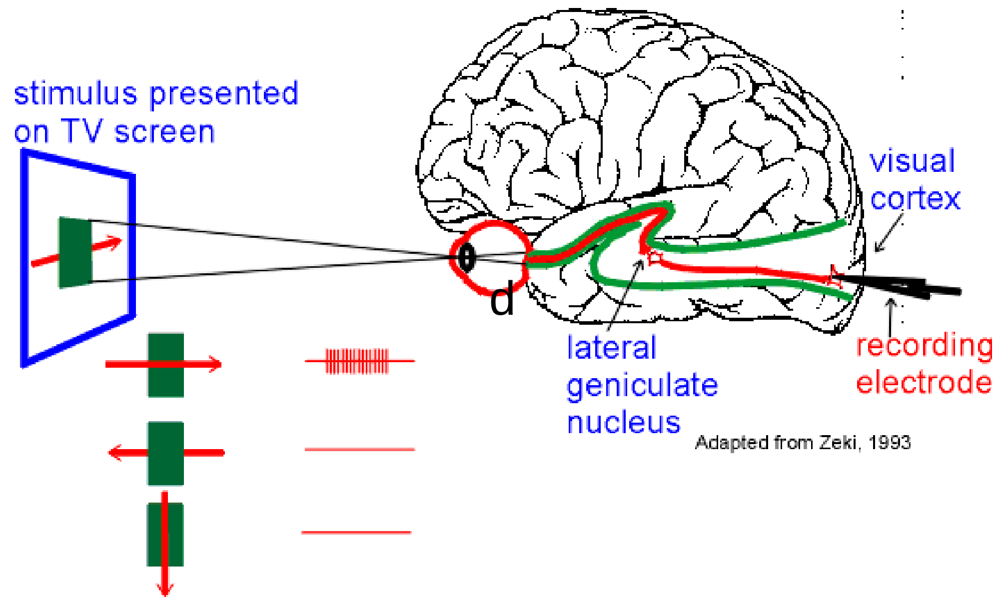
Object datasets

	Caltech101	Caltech256	Action40	Event8
Places-CNN feature	65.18±0.88	45.59±0.31	42.86±0.25	94.12±0.99
ImageNet-CNN feature	**87.22±0.92**	**67.23±0.27**	**54.92±0.33**	**94.42±0.76**

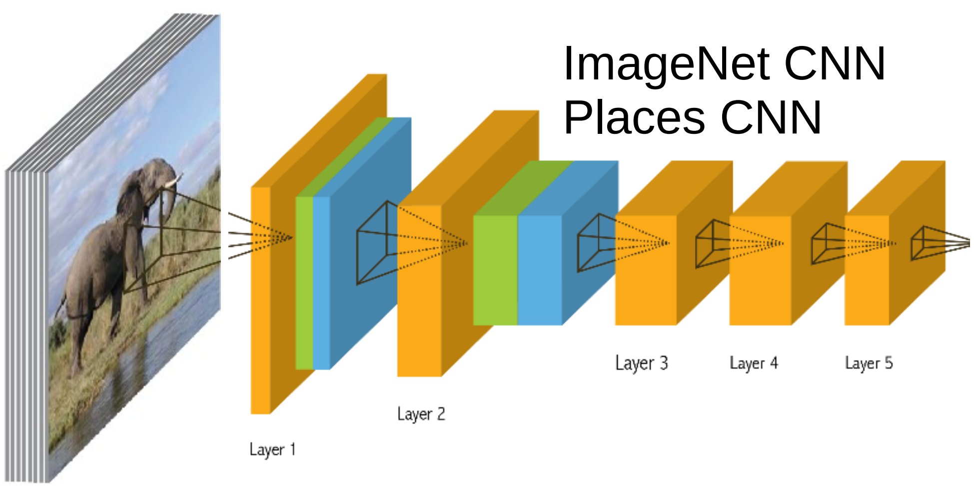
Zhou, Lapedriza, Xiao, Torralba & Oliva (NIPS 2014)

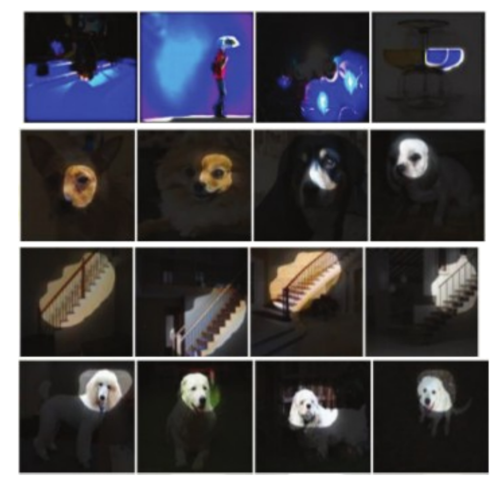
Data-Driven Approach to Visualize CNN

Neuroscientists study brain

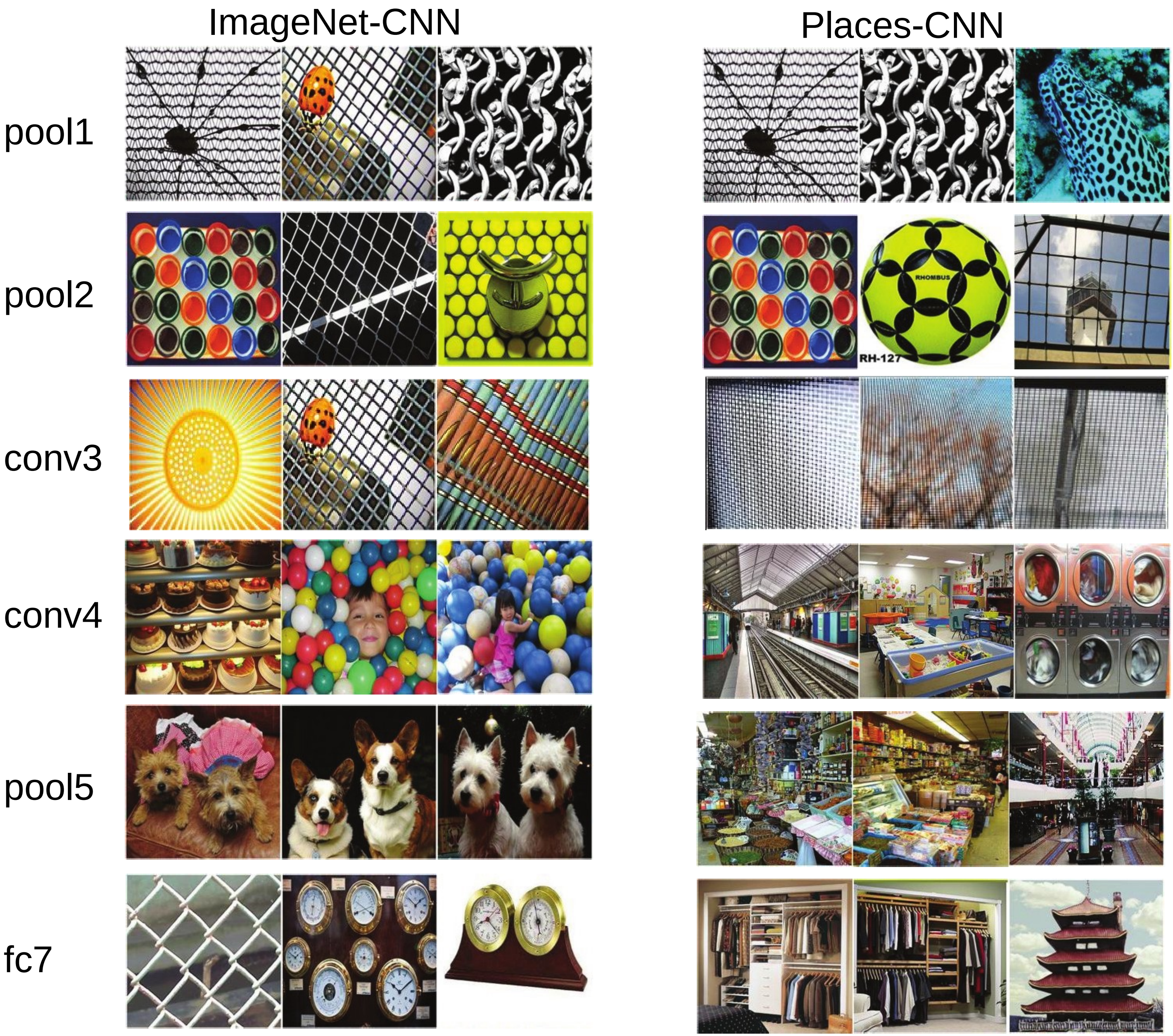


200,000 image stimuli of objects and scene categories (ImageNet TestSet+SUN database)





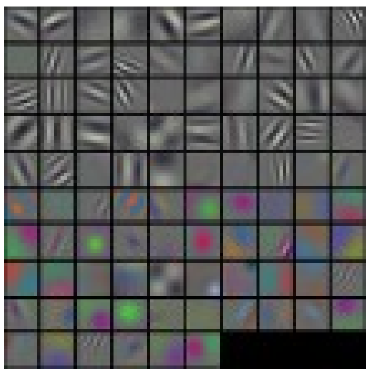
Preferred Images of Different Layers



Mean Activation Images of Internal Units

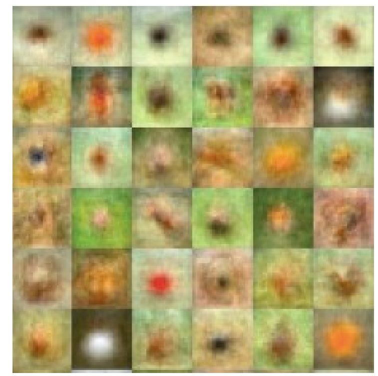
ImageNet CNN

Conv1 units

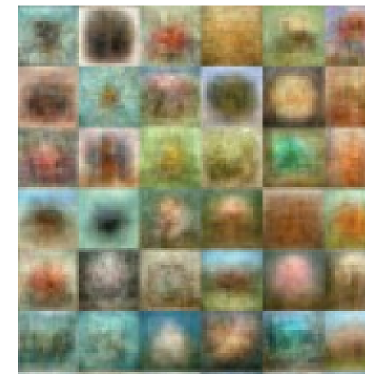


Conv2 units

Conv5 units



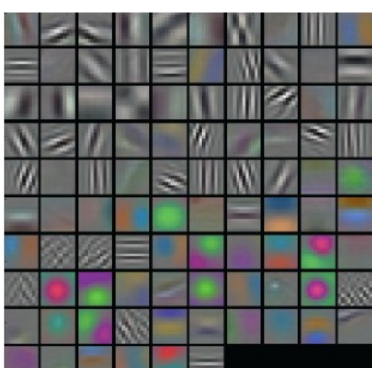
FC7 units



Object shapes

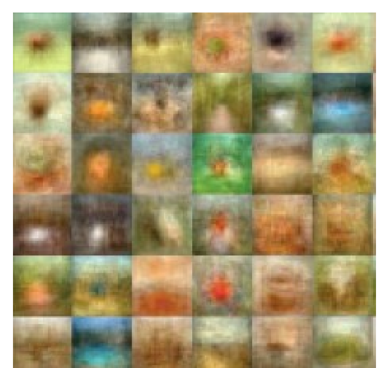
Places CNN

Conv1 units

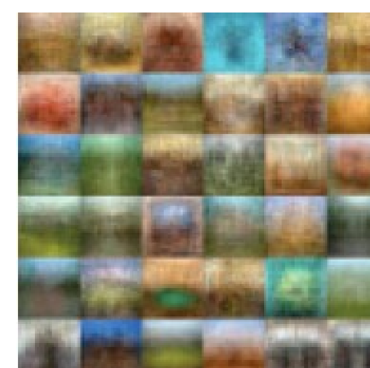


Conv2 units

Conv5 units



FC7 units



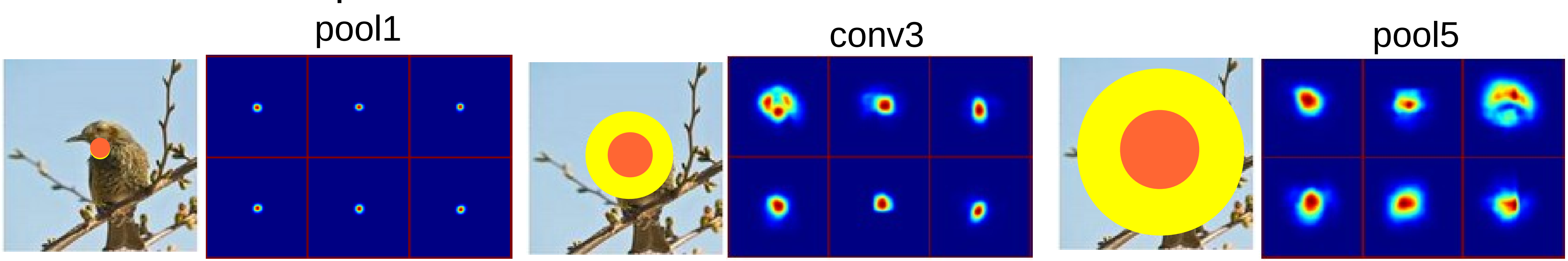
Space shapes

Zhou, Lapedriza, Xiao, Oliva & Torralba (NIPS 2014)

Estimating the Receptive Fields

Estimated receptive fields

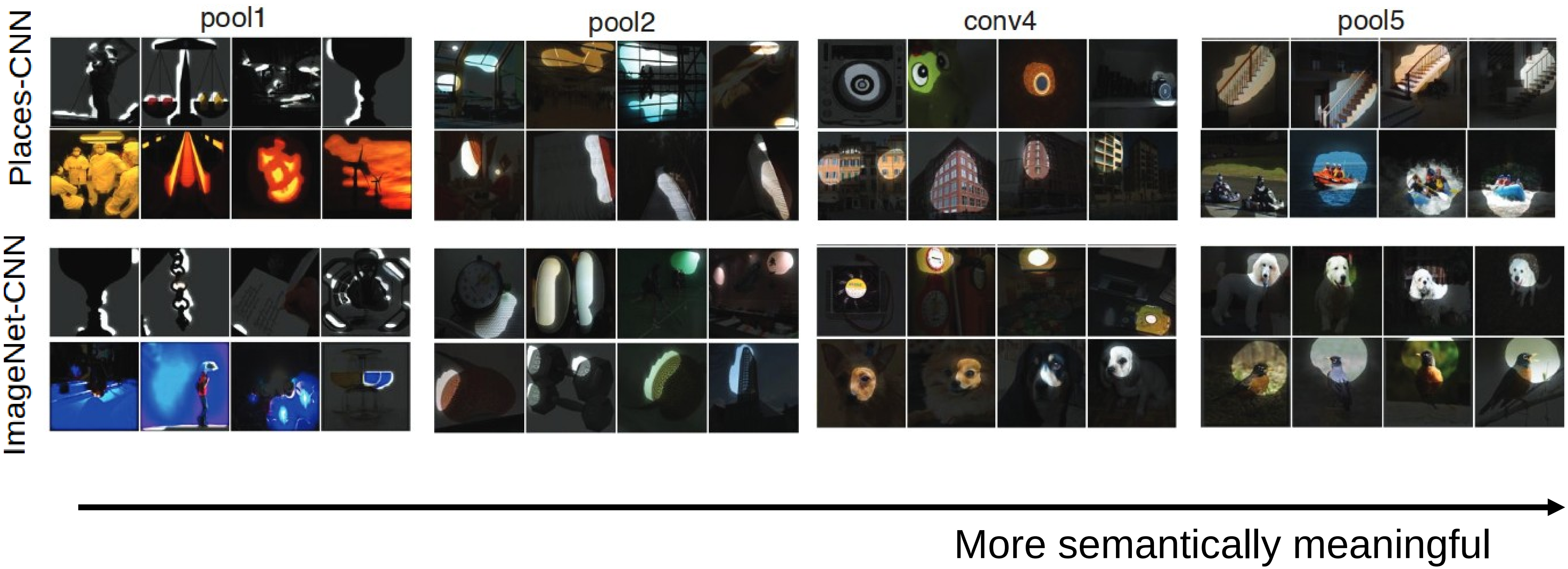
Actual size of RF is much smaller than the theoretic size



Zhou et al, ICLR'15

Image segmentation using RF of Units

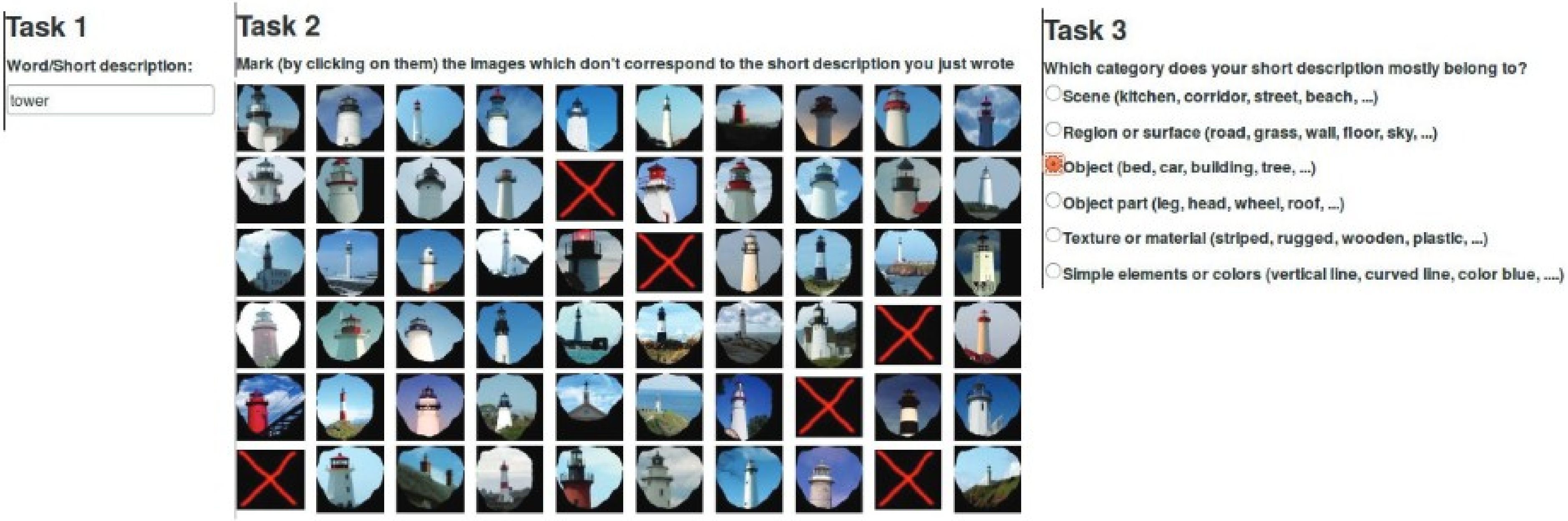
Image segmentation results for units at different layers:



Zhou et al, ICLR'15

Annotating the Semantics of Units

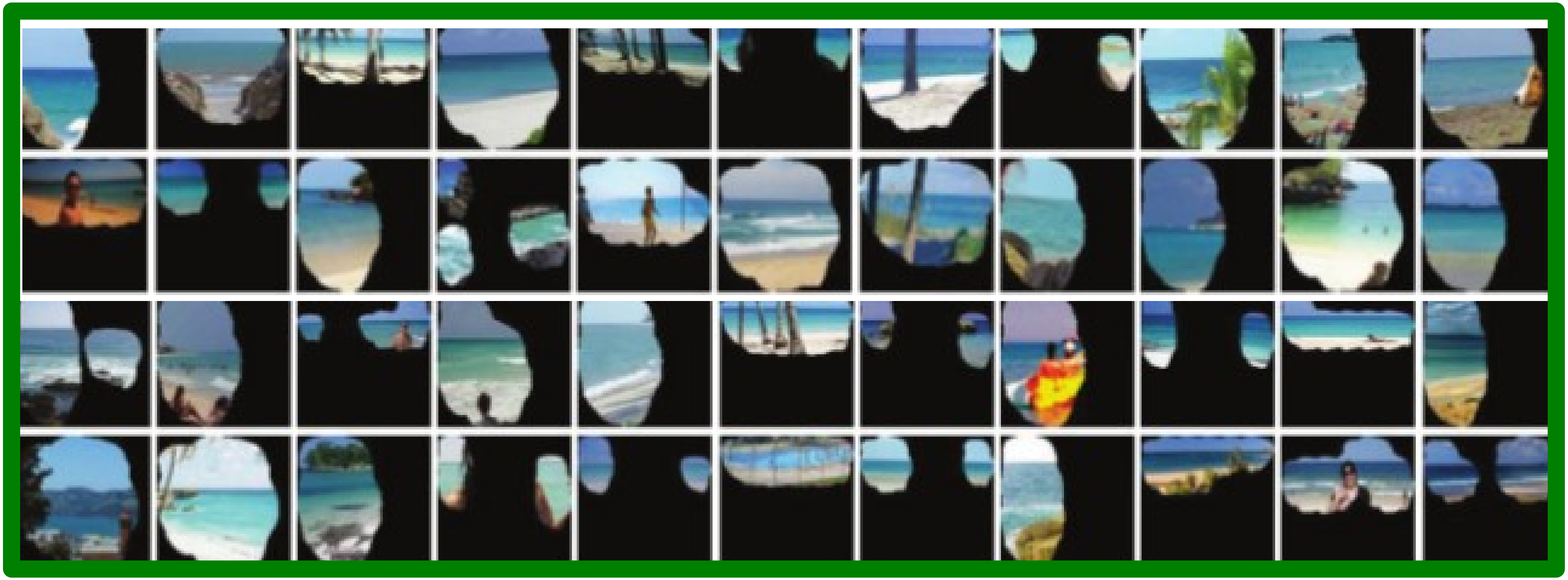
Top ranked segmented images are cropped and sent to Amazon Turk for annotation.

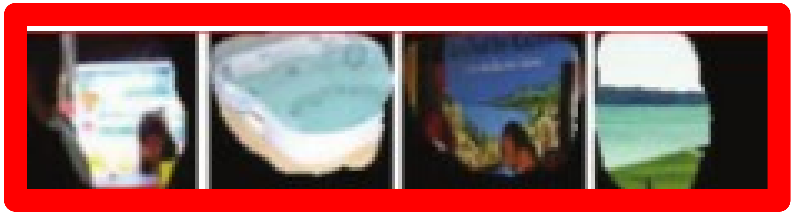


Zhou et al, ICLR'15

Annotating the Semantics of Units

Pool5, unit 76; Label: ocean; Type: scene; Precision: 93%

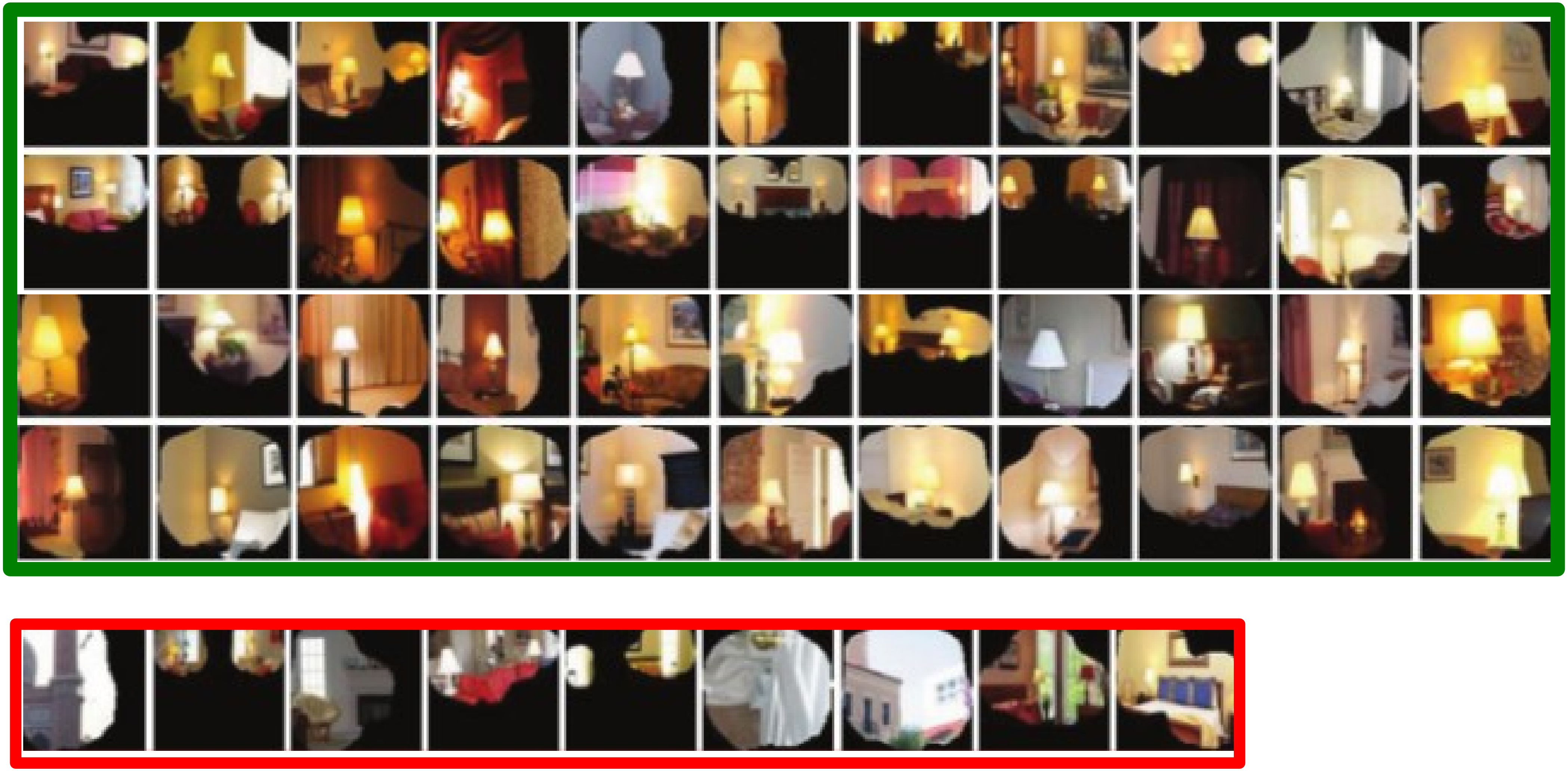




Zhou et al, ICLR'15

Annotating the Semantics of Units

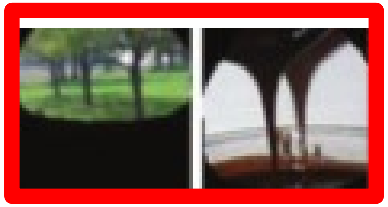
Pool5, unit 13; Label: Lamps; Type: object; Precision: 84%



Zhou et al, ICLR'15

Annotating the Semantics of Units

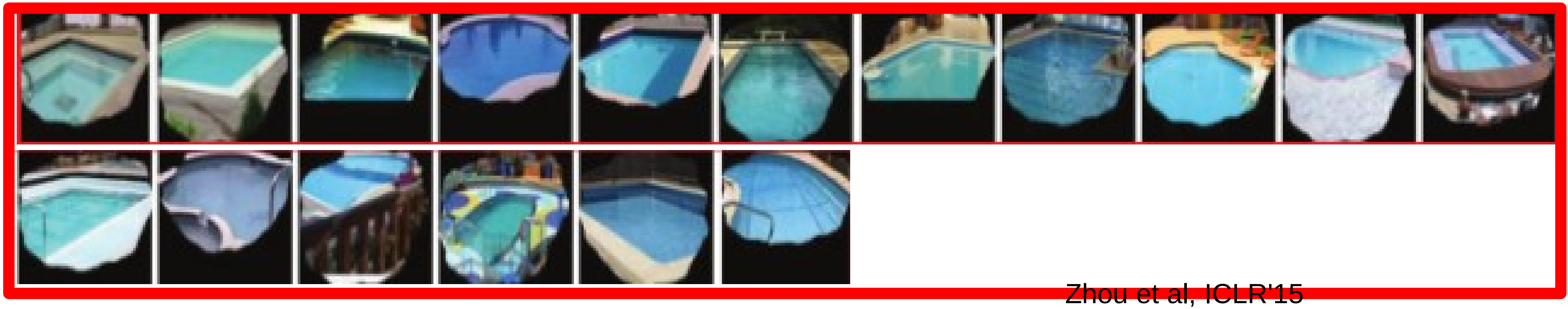
Pool5, unit 77; Label:legs; Type: object part; Precision: 96%



Zhou et al, ICLR'15

Annotating the Semantics of Units

Pool5, unit 112; Label: pool table; Type: object; Precision: 70%



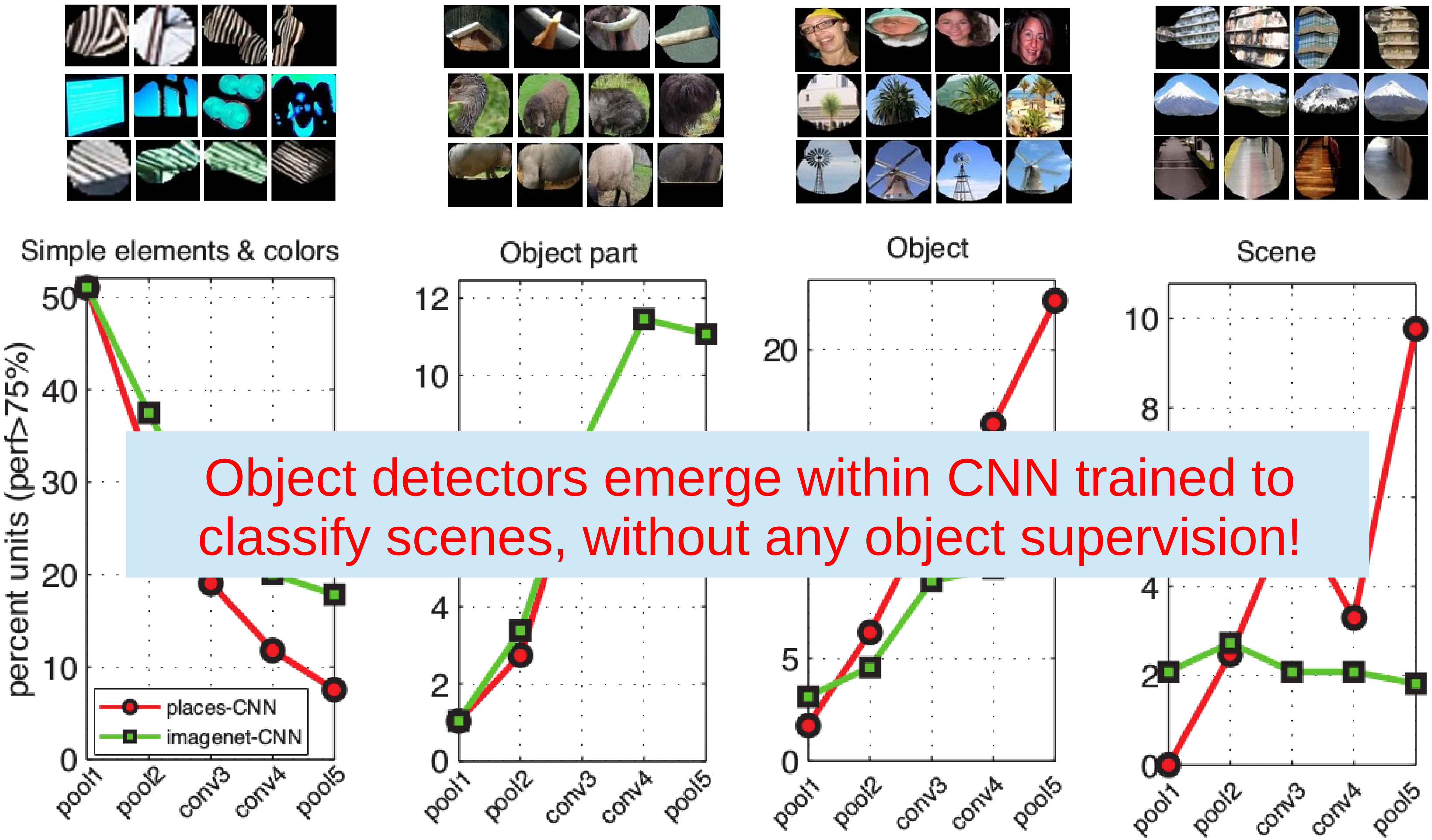
Zhou et al, ICLR'15

Annotating the Semantics of Units

Pool5, unit 22; Label: dinner table; Type: scene; Precision: 60%

Zhou et al, ICLR'15

Distribution of Semantic Types at Each Layer

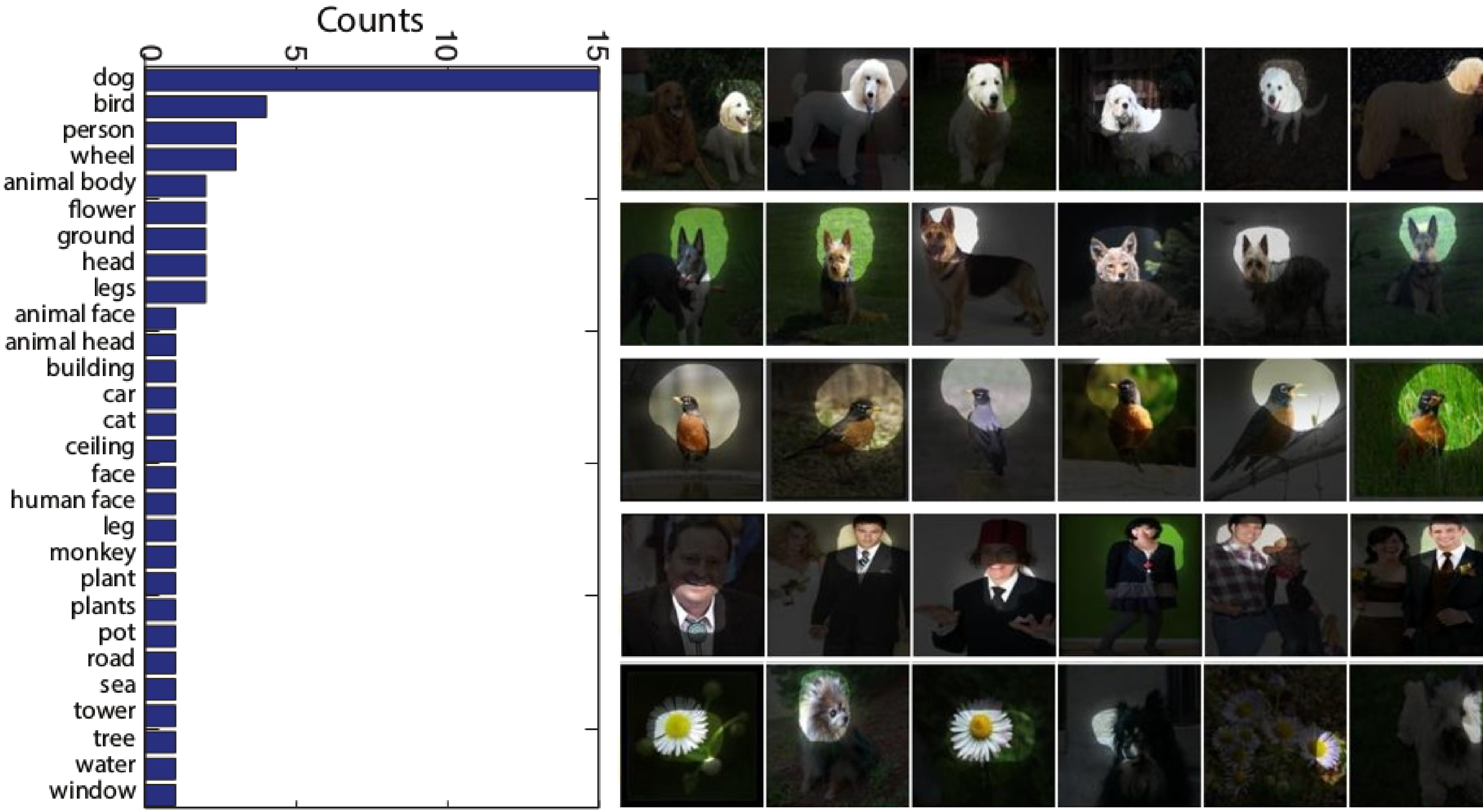


Object detectors emerge within CNN trained to classify scenes, without any object supervision!

Zhou et al, ICLR'15

Histogram of Emerged Objects in Pool5

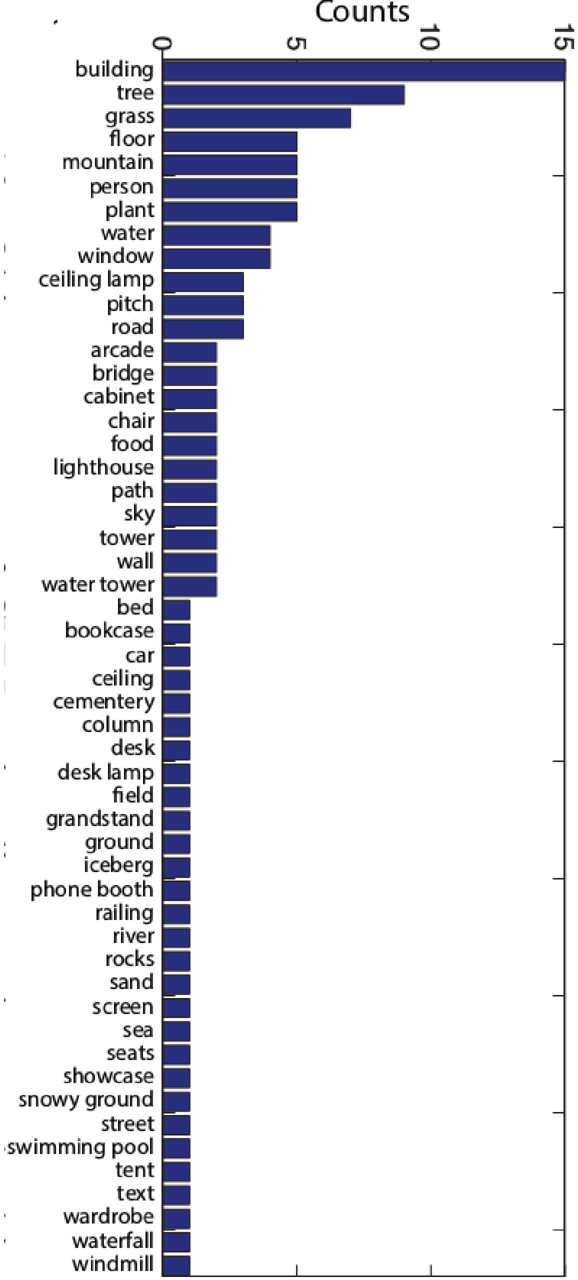
ImageNet-CNN (59/256)

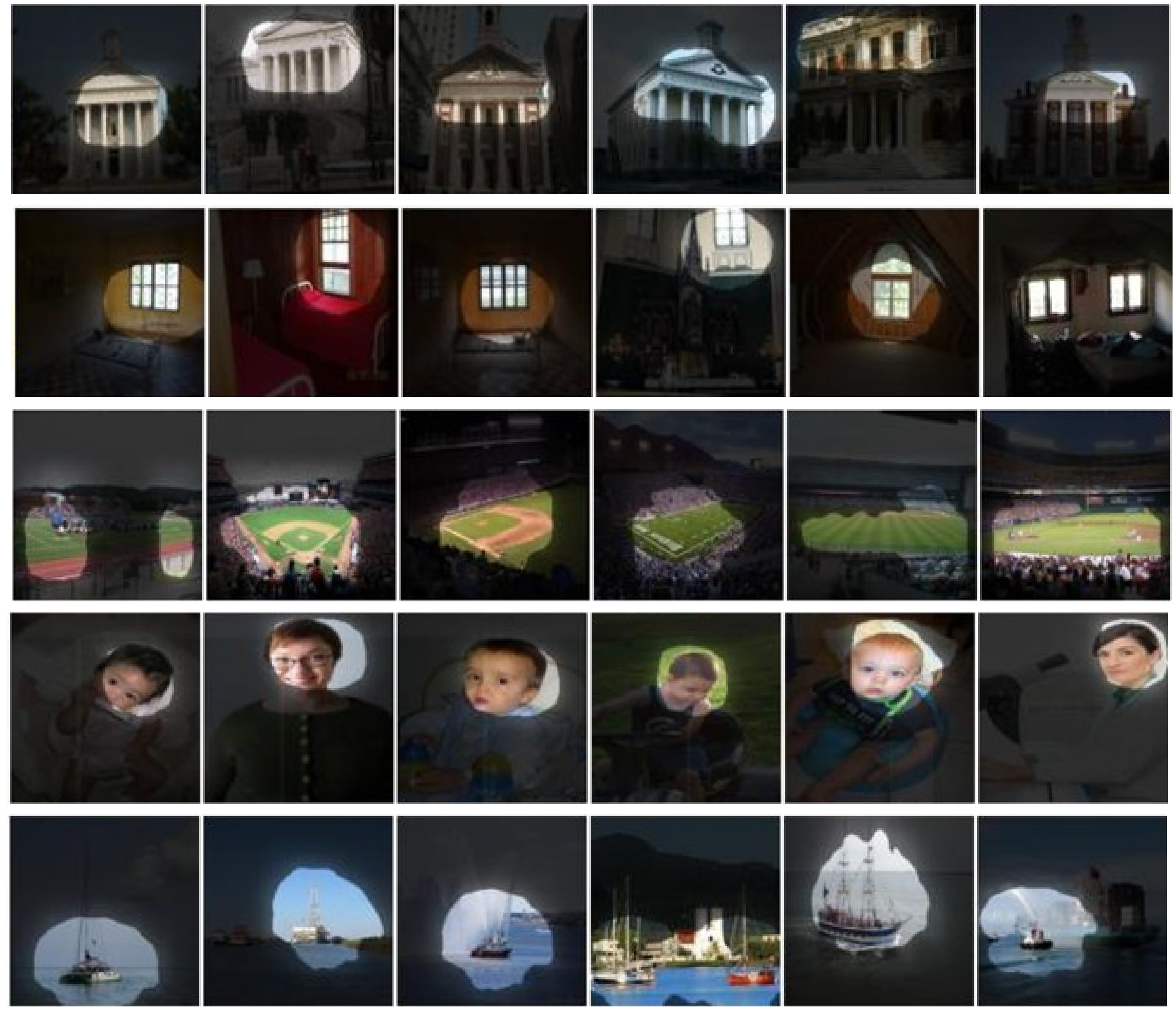


Zhou et al, ICLR'15

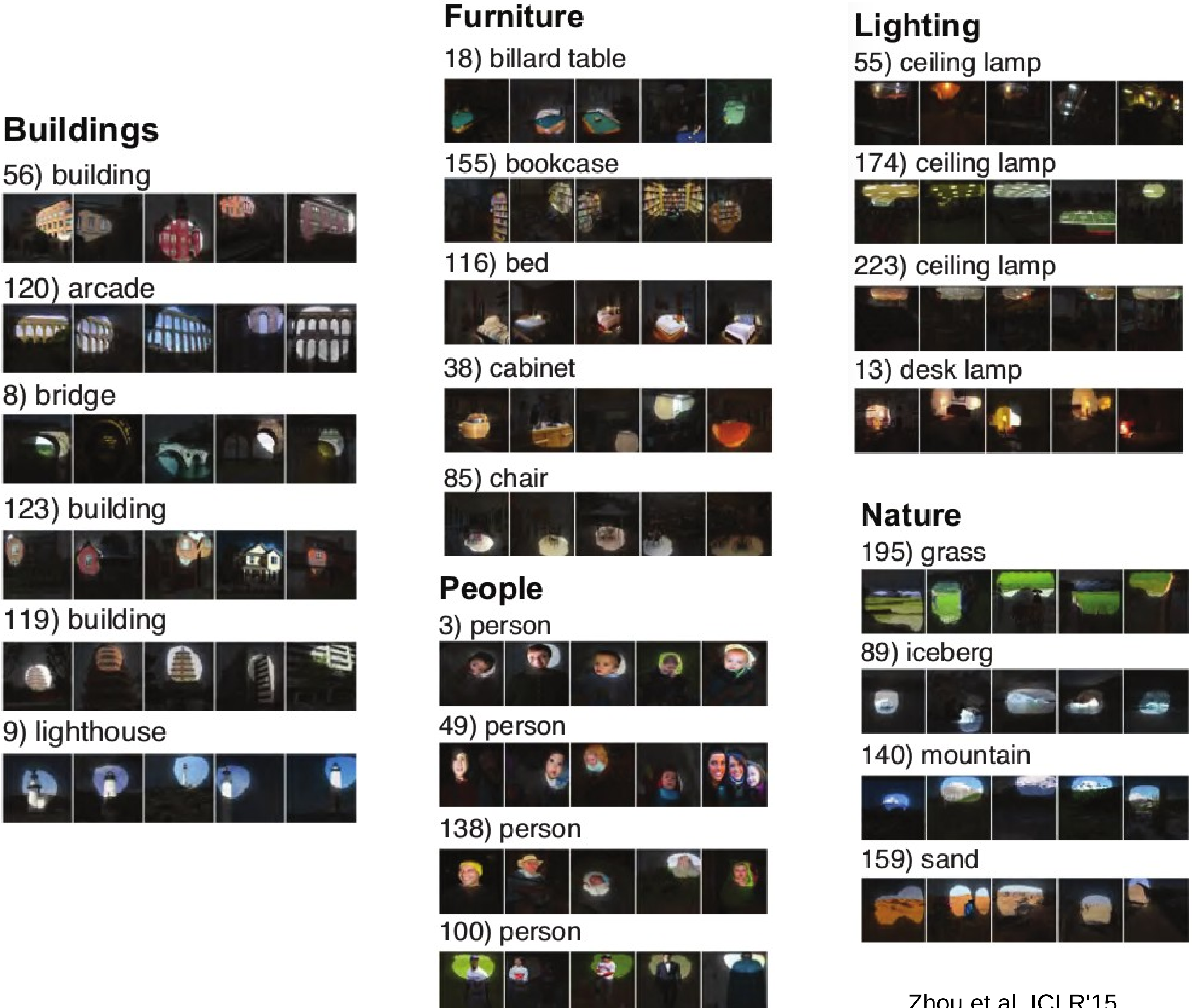
Histogram of Emerged Objects in Pool5

Places-CNN (151/256)





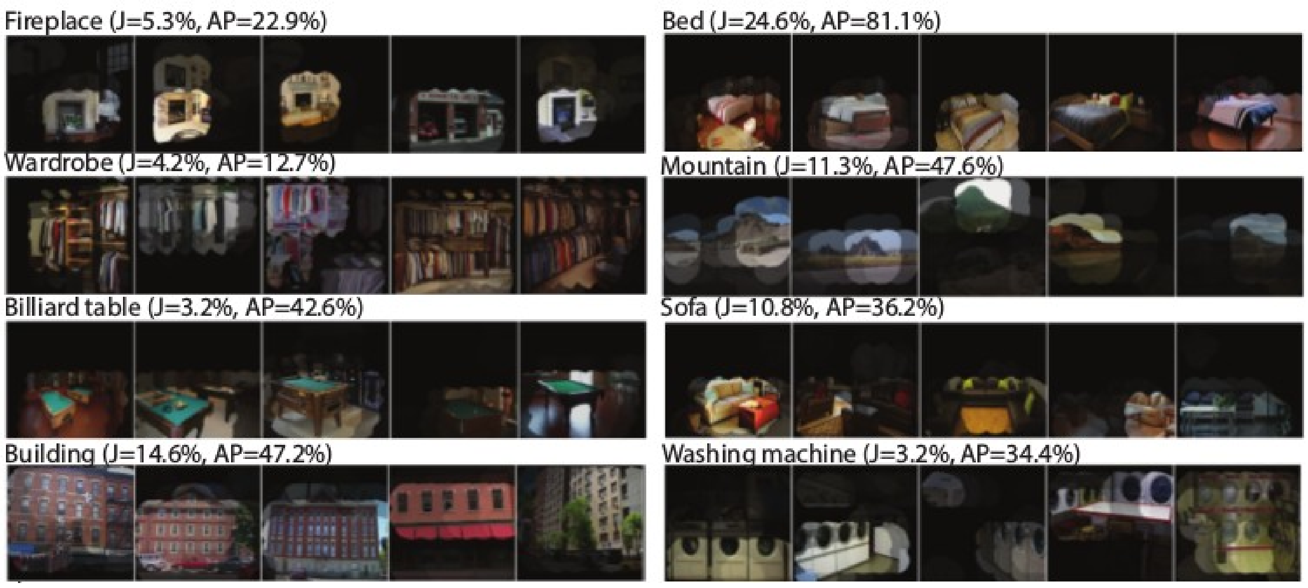
Zhou et al, ICLR'15

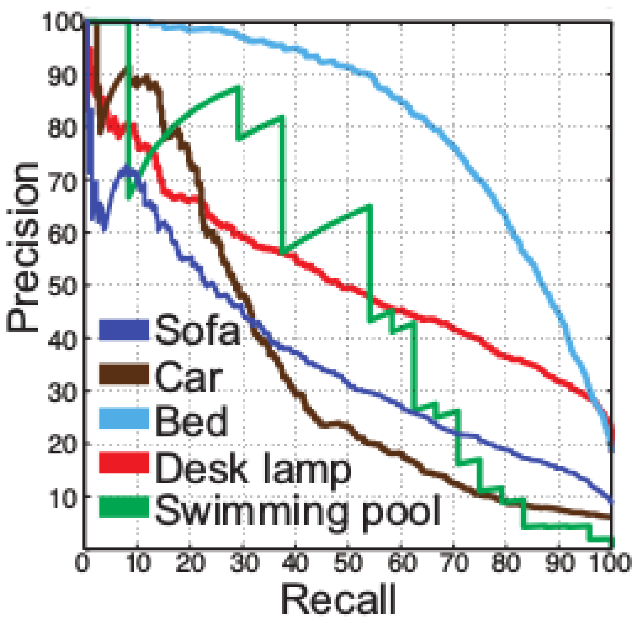


Zhou et al, ICLR'15

Evaluation on SUN Database

Evaluate the performance of the emerged object detectors





Bolei Zhou, et al, ICLR'15

Summary

We show that object detectors emerge within CNN trained for scene classification, even more than the CNN trained for object classification.

How these object detectors are relevant to the final prediction of the CNN?

Why CNN makes the prediction?

CNN Predictions:

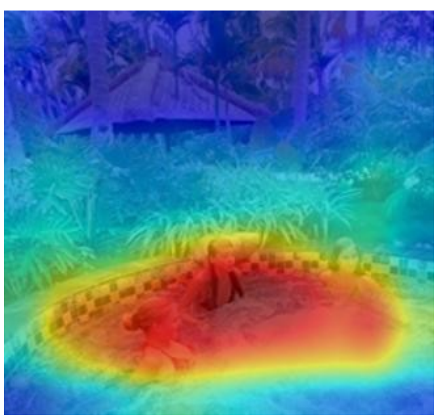
Bedroom:0.64
Dorm room:0.23

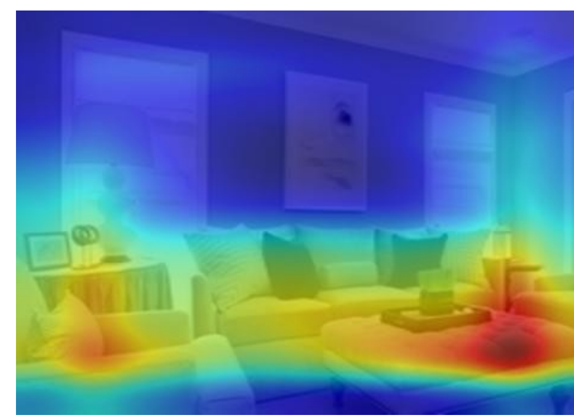
Why CNN makes the prediction?

Hot spring:0.36

Art studio:0.54

Living room:0.53





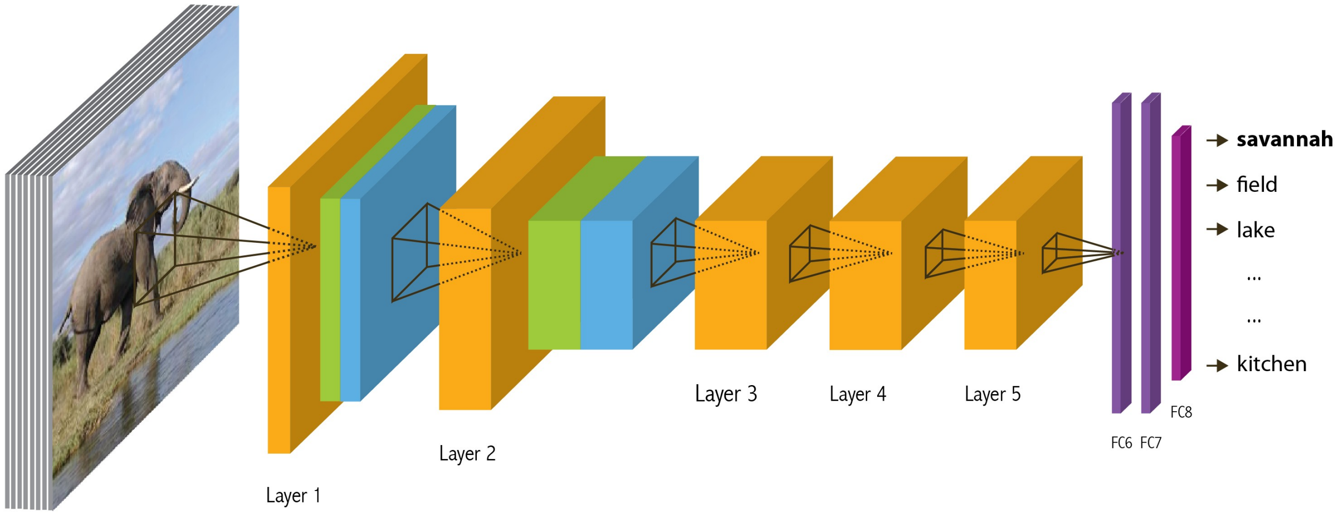
Zhou et al, Learning Deep Features for Discriminative Localization. CVPR'16 submission

Why CNN makes the prediction?

Basic idea: simplify the CNN structures

conv layers + ~~FC layers~~ + softmax layer

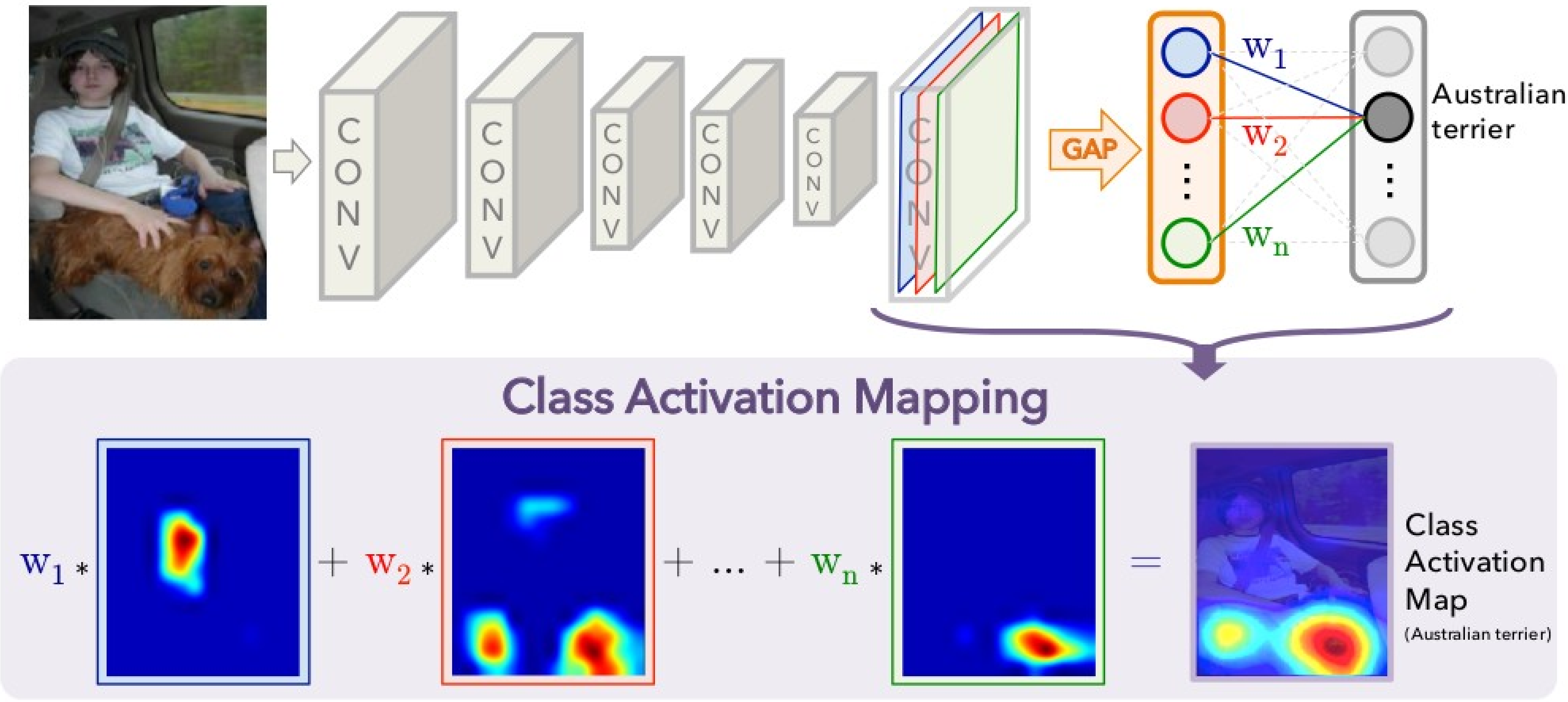
Global Average Pooling



Zhou et al, Learning Deep Features for Discriminative Localization. CVPR'16 submission

Class Activation Map

Object localization without bounding box annotation



Zhou et al, Learning Deep Features for Discriminative Localization. CVPR'16 submission

Class Activation Mapping

Different predictions leads to different class activation maps

Zhou et al, Learning Deep Features for Discriminative Localization. CVPR'16 submission

Weakly-supervised object localization

CNN trained from classification is used for object localization directly, without bounding box annotation.

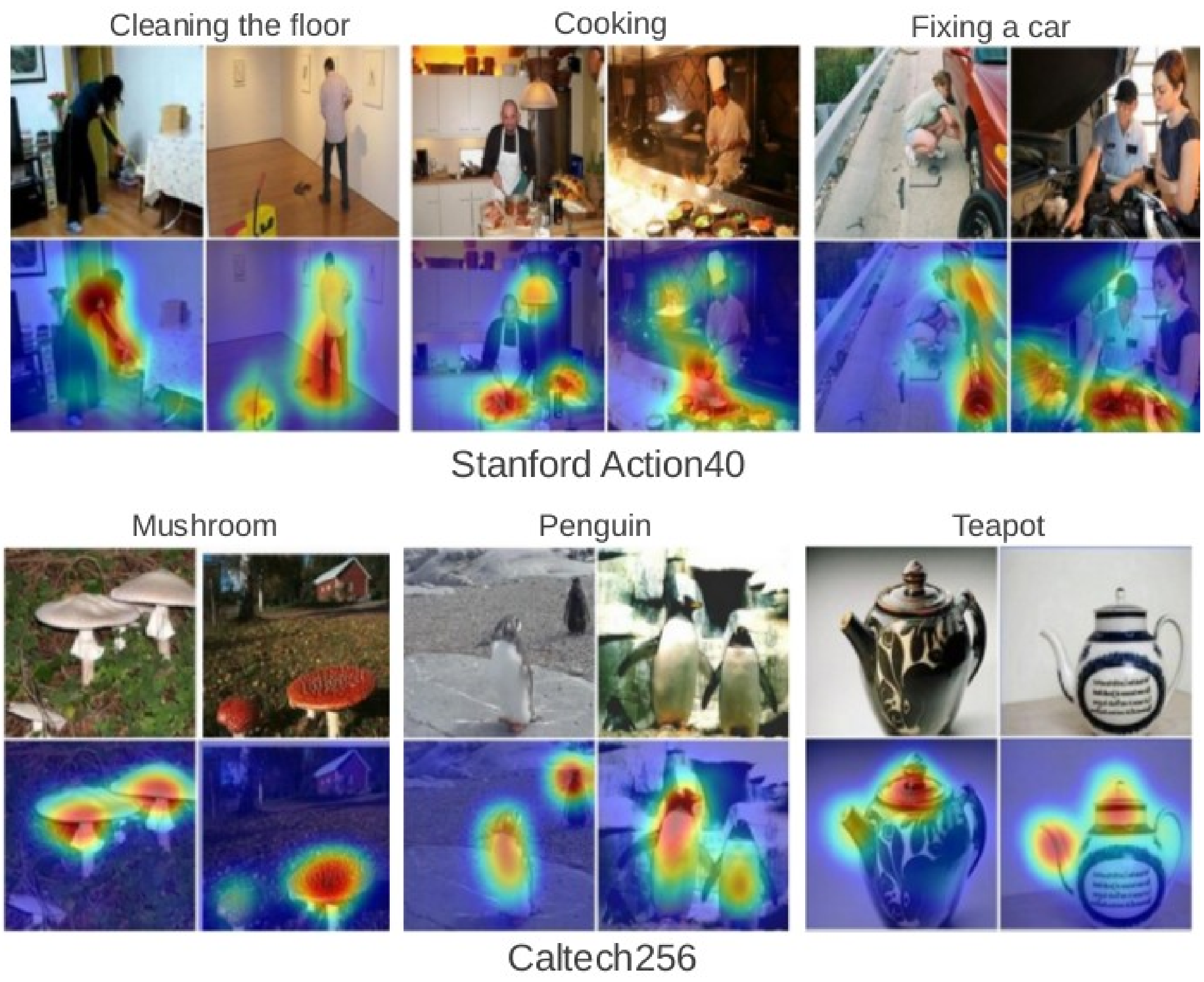
Table 3. Localization error on the ILSVRC test set for various weakly- and fully- supervised methods.

Method	supervision	top-5 test error
GoogLeNet-GAP (heuristics)	weakly	**37.1**
GoogLeNet-GAP	weakly	42.9
Backprop [22]	weakly	46.4
GoogLeNet [24]	full	26.7
OverFeat [21]	full	29.9
AlexNet [24]	full	34.2

Zhou et al, Learning Deep Features for Discriminative Localization. CVPR'16 submission

Localizable Deep Features

Deep Feature + linear SVM: localize informative regions



Zhou et al, Learning Deep Features for Discriminative Localization. CVPR'16 submission

Summary

- Places Database are built
- Places-CNNs are trained on Places Database
- Places-CNN and ImageNet-CNN are compared.

All data, demo, and pre-trained models are available at

http://places.csail.mit.edu