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


2 year old kid


15 scenes database (2006)

8 scenes database (2001)
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

The evolution of scene and object centered databases

COIL-20

Caltech 101

MNIST (1998)

IMAGENET (2009)

2 year old kid


PASCAL (2005)

15 scenes database (2006)

8 scenes database (2001)

SUN dataset (2010)
SUN dataset
900 Scene categories & 130,000 images
The evolution of scene and object centered databases

COIL-20

Caltech 101

MNIST (1998)

2 year old kid

IMAGENET (2009)

10^3 10^4 10^5 10^6 10^7 10^8 10^9 # images


15 scenes database (2006)

SUN database (2010)

10^3 10^4 10^5 10^6 10^7 10^8 10^9 # images

8 scenes database (2001)

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

Google Image Search

flickr

bing
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, cosy, common, comparable, comparative, competitive, complementary, complete, complex, complicated, concealed, conceivable, confined, considerable, contemporary, cool, coolest, cosmopolitan, cost-effective, cozy, cozier, cream-white, creative, crowded, cultivated, cultural, current, damp, dangerous, dark, darkened, darker, darkest, decorative, delightful, designated, designed, desirable, deserved, 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, dull, 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, fashionable, favourite, fictional, fictitious, filmed, filthy, fine, foggy, foreign, formal, fractured, friendly, frightening, frightful, frosty, frozen, frustrating, full, funny, furnished, fuzzy, gaudy, ghostly, ghastly, glorious, glassy, glazed, glittering, gloomy, glorious, glossy, gold-like, gold-plated, good, gorgeous, graceful, gracious, grand, gray, great, greatest, green, greener, grey, grisly, gruesome, habitable, 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, impertinent, important, impossible, impressive, improbable, improper, insidious, 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, isolated, joint, joyful, key, known, large, large-scale, largest, less-favored, lesser, licensed, limitless, 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, marine, marvellous, mysterious-looking, mystic, mythical, 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, 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, possible, 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, probable, 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, reusable, required, reserved, residential, respectable, respected, restful, 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, shocking, shoddy, short-term, significant, silent, silly, similar, 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, strong, stressful, striking, stunning, stupendous, stupid, stylish, successful, sufficient, sunny, super, superb, superior, surrealistic, suspicious, symbolic, typical, terrible, terrific, theoretical, thrilling, thriving, tiptop, tip, 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, wood, 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

Is this a cliff scene?

Definition: a high, steep or overhanging face of rock.

Task

For each of the 810 images, answer yes or no to the above question. Only answer Yes to real photos. Always answer No to cartoon, drawing, CG rendering, or real photos with a large text overlay on the photo. Here are some examples:

No Single Object  No Text Overlay  No Drawing  No Screenshot  No Graphics  No Bad Photo

Not Only Logo  No Magazine/Newspaper  No  No  Yes  Yes

Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes
Annotation Interface
1st round

**Is this a cliff scene?**

*Definition*: a high, steep or overhanging face of rock.

No
Annotation Interface:

2nd round

Is this a living room scene?

Definition: a room in a private residence intended for general social and leisure activities.

Yes
cleaner bedroom:205
messy bedroom:808
Places Database for Scene Recognition

http://places.csail.mit.edu

10 million images from 476 scene categories

More than one year of time!

Bolei

Grad School:

+Ten thousands of anonymous AMT workers
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.

Zhou, Lapedriza, Xiao, Torralba & Oliva (NIPS 2014)
Places-CNN Demo

2675 anonymous users report 77% top-5 recognition accuracy

Demo, data, and Places-CNNs could be downloaded at http://places.csail.mit.edu
Analyzing the CNNs
What are all those units doing?

what is being learned inside the network?
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),

**Deformable Part model**


**Bag-of-word model**


**Class-specific graph model**

Kumar, Torr and Zisserman (2005), Felzenszwalb & Huttenlocher (2005)
Learning to Represent Objects

Possible internal representations:

- Object parts
- Textures
- Attributes
How Objects are Represented in CNN?

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

Learning to Represent Scenes

Possible internal representations:
- Objects (scene parts?)
- Scene attributes
- Object parts
- Textures
ImageNet CNN and Places CNN

ImageNet CNN for Object Classification

Places CNN for Scene Classification

Same architecture: AlexNet
Generic Visual Feature

Scene datasets

<table>
<thead>
<tr>
<th></th>
<th>SUN397</th>
<th>MIT Indoor67</th>
<th>Scene15</th>
<th>SUN Attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>Places-CNN feature</td>
<td>54.32±0.14</td>
<td>68.24</td>
<td>90.19±0.34</td>
<td>91.29</td>
</tr>
<tr>
<td>ImageNet-CNN feature</td>
<td>42.61±0.16</td>
<td>56.79</td>
<td>84.23±0.37</td>
<td>89.85</td>
</tr>
</tbody>
</table>

Object datasets

<table>
<thead>
<tr>
<th></th>
<th>Caltech101</th>
<th>Caltech256</th>
<th>Action40</th>
<th>Event8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Places-CNN feature</td>
<td>65.18±0.88</td>
<td>45.59±0.31</td>
<td>42.86±0.25</td>
<td>94.12±0.99</td>
</tr>
<tr>
<td>ImageNet-CNN feature</td>
<td>87.22±0.92</td>
<td>67.23±0.27</td>
<td>54.92±0.33</td>
<td>94.42±0.76</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Layer</th>
<th>ImageNet-CNN</th>
<th>Places-CNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>pool1</td>
<td><img src="image1" alt="ImageNet-CNN pool1" /></td>
<td><img src="image2" alt="Places-CNN pool1" /></td>
</tr>
<tr>
<td>pool2</td>
<td><img src="image3" alt="ImageNet-CNN pool2" /></td>
<td><img src="image4" alt="Places-CNN pool2" /></td>
</tr>
<tr>
<td>conv3</td>
<td><img src="image5" alt="ImageNet-CNN conv3" /></td>
<td><img src="image6" alt="Places-CNN conv3" /></td>
</tr>
<tr>
<td>conv4</td>
<td><img src="image7" alt="ImageNet-CNN conv4" /></td>
<td><img src="image8" alt="Places-CNN conv4" /></td>
</tr>
<tr>
<td>pool5</td>
<td><img src="image9" alt="ImageNet-CNN pool5" /></td>
<td><img src="image10" alt="Places-CNN pool5" /></td>
</tr>
<tr>
<td>fc7</td>
<td><img src="image11" alt="ImageNet-CNN fc7" /></td>
<td><img src="image12" alt="Places-CNN fc7" /></td>
</tr>
</tbody>
</table>
Mean Activation Images of Internal Units

**ImageNet CNN**

- Conv1 units
- Conv2 units
- Conv5 units
- FC7 units

**Places CNN**

- Conv1 units
- Conv2 units
- Conv5 units
- FC7 units

**Object shapes**

**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:

More semantically meaningful

Zhou et al, ICLR'15
Top ranked segmented images are cropped and sent to Amazon Turk for annotation.
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%
Annotating the Semantics of Units

Pool5, unit 77; Label:legs; Type: object part; Precision: 96%
Annotating the Semantics of Units

Pool5, unit 112; Label: pool table; Type: object; Precision: 70%
Annotating the Semantics of Units

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

Zhou et al, ICLR'15
Object detectors emerge within CNN trained to classify scenes, without any object supervision!
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
Evaluation on SUN Database

Evaluate the performance of the emerged object detectors
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?

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.

<table>
<thead>
<tr>
<th>Method</th>
<th>supervision</th>
<th>top-5 test error</th>
</tr>
</thead>
<tbody>
<tr>
<td>GoogLeNet-GAP (heuristics)</td>
<td>weakly</td>
<td><strong>37.1</strong></td>
</tr>
<tr>
<td>GoogLeNet-GAP</td>
<td>weakly</td>
<td>42.9</td>
</tr>
<tr>
<td>Backprop [22]</td>
<td>weakly</td>
<td>46.4</td>
</tr>
<tr>
<td>OverFeat [21]</td>
<td>full</td>
<td>29.9</td>
</tr>
<tr>
<td>AlexNet [24]</td>
<td>full</td>
<td>34.2</td>
</tr>
</tbody>
</table>

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