Recap: Resnet

- $F(x)$ is a residual mapping w.r.t. identity

\[ H(x) = F(x) + x \]
Opportunities of Scale

Computer Vision

James Hays

Many slides from James Hays, Alyosha Efros, and Derek Hoiem

Graphic from Antonio Torralba
Outline

Opportunities of Scale: Data-driven methods

– The Unreasonable Effectiveness of Data
– Scene Completion
– Im2gps
– Recognition via Tiny Images
Computer Vision Class 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 bag of words)

• Learning from Data
  – 1991 (EigenFaces) but late 90’s to now especially
What has changed in the last 15 years?

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

https://youtu.be/yvDCzhbjYWt?t=24
Watch until 9:42
The "Unreasonable Effectiveness of Math"

• “The miracle of the appropriateness of the language of mathematics…” Eugene Wigner

• “The most incomprehensible thing about the universe is that it is comprehensible.” Albert Einstein

• “There is only one thing which is more unreasonable than the unreasonable effectiveness of mathematics in physics, and this is the unreasonable ineffectiveness of mathematics in biology.” Israel Gelfand

• “We should stop acting as if our goal is to author extremely elegant theories, and instead embrace complexity and make use of the best ally we have: the unreasonable effectiveness of data.” Peter Norvig
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."
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
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?
Scene Completion

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

http://graphics.cs.cmu.edu/projects/scene-completion/
How it works

- Find a similar image from a large dataset
- Blend a region from that image into the hole
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. 19\textsuperscript{th}, 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
Scene 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
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?
Outline

Opportunities of Scale: Data-driven methods

- The Unreasonable Effectiveness of Data
- Scene Completion
- Im2gps
- Recognition via Tiny Images
Hopefully, if you have enough images, the dataset will contain very similar images that you can find with simple matching methods.
Graph cut + Poisson blending
This reminded me of @jhhays and Efros' large-scale image geolocation work
**im2gps** (Hays & Efros, CVPR 2008)

6 million geo-tagged Flickr images

http://graphics.cs.cmu.edu/projects/im2gps/
How much can an image tell about its geographic location?
Nearest Neighbors according to gist + bag of SIFT + color histogram + a few others
Example Scene Matches
Voting Scheme
Effect of Dataset Size

![Graph showing the effect of dataset size on the percentage of geolocations within 200km. The graph compares the first nearest neighbor scene match with the chance of random scenes. The x-axis represents the database size in thousands of images on a log scale, while the y-axis shows the percentage of geolocations within 200km. The green line represents the first nearest neighbor scene match, which increases significantly as the dataset size grows, while the red dashed line represents the chance of random scenes, remaining constant.]
Follow up works

- Revisiting IM2GPS in the Deep Learning Era. Nam Vo, Nathan Jacobs, James Hays. ICCV 2017

<table>
<thead>
<tr>
<th>Threshold (km)</th>
<th>Street 1</th>
<th>City 25</th>
<th>Region 200</th>
<th>Country 750</th>
<th>Cont. 2500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Im2GPS [9]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Im2GPS [10]</td>
<td>02.5</td>
<td>21.9</td>
<td>32.1</td>
<td>35.4</td>
<td>51.9</td>
</tr>
<tr>
<td>PlaNet [36]</td>
<td>08.4</td>
<td>24.5</td>
<td>37.6</td>
<td>53.6</td>
<td><strong>71.3</strong></td>
</tr>
<tr>
<td>[L] 7011C</td>
<td>06.8</td>
<td>21.9</td>
<td>34.6</td>
<td>49.4</td>
<td>63.7</td>
</tr>
<tr>
<td>[L] kNN, σ=4</td>
<td><strong>12.2</strong></td>
<td><strong>33.3</strong></td>
<td><strong>44.3</strong></td>
<td><strong>57.4</strong></td>
<td><strong>71.3</strong></td>
</tr>
<tr>
<td>... 28m database</td>
<td><strong>14.4</strong></td>
<td><strong>33.3</strong></td>
<td><strong>47.7</strong></td>
<td><strong>61.6</strong></td>
<td><strong>73.4</strong></td>
</tr>
</tbody>
</table>
Tiny Images

80 million tiny images: a large dataset for non-parametric object and scene recognition

http://groups.csail.mit.edu/vision/TinyImages/
c) Segmentation of 32x32 images
c) Segmentation of 32x32 images
Human Scene Recognition

![Graph showing Correct recognition rate vs Image resolution for Color image and Grayscale.](image-url)
Humans vs. Computers: Car-Image Classification

Humans for 32 pixel tall images
Various computer vision algorithms for full resolution images
Powers of 10

Number of images on my hard drive: \(10^4\)

Number of images seen during my first 10 years: \(10^8\)
(3 images/second * 60 * 60 * 16 * 365 * 10 = 630720000)

Number of images seen by all humanity: \(10^{20}\)
106,456,367,669 humans \(^1\) * 60 years * 3 images/second * 60 * 60 * 16 * 365 = 1 from http://www.prb.org/Articles/2002/HowManyPeopleHaveEverLivedonEarth.aspx

Number of photons in the universe: \(10^{88}\)

Number of all 32x32 images: \(10^{7373}\)
\(256^{32 \times 32} \approx 10^{7373}\)
Scenes are unique
But not all scenes are so original
Lots

Of

Images

A. Torralba, R. Fergus, W.T.Freeman. PAMI 2008
Lots
Of
Images
Lots
Of
Images
Application: Automatic Colorization

Input

Color Transfer

Color Transfer

Matches (gray)

Matches (w/ color)

Avg Color of Match
Application: Automatic Colorization

Input

Color Transfer

Color Transfer

Matches (gray)

Matches (w/ color)

Avg Color of Match
Automatic Orientation Examples

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

• With billions of images on the web, it’s often possible to find a close nearest neighbor.

• In such cases, we can shortcut hard problems by “looking up” the answer, stealing the labels from our nearest neighbor. For example, simple (or learned) associations can be used to synthesize background regions, colorize, or recognize objects.

• But we can’t really “brute force” computer vision. Still, it’s nice to get an intuition for the size of “image space”.