The Promise and Peril of Big Data

Alexei (Alyosha) Efros
CMU (school-year), INRIA (summer)
It’s hard to be a computer...
How the computer sees the world:

The Guitar Player
Pablo Picasso (1911)
Some early work…

“Data, Data, Data… Watson, I need Data!”
[Sherlock Holmes, 1886]
But If you want to publish a NIPS paper…

Data

Features

Learning Algorithm
Face Detection: Big Success Story

- Rowley, Baluja, and Kanade, 1998
- Schniderman & Kanade, 1999
- Viola & Jones, 2001
Modern Recognition is largely Data-Driven

• In non-linear SVMs:
  – In ML, people report ~10% of data are support vectors
  – In recognition, up to 2/3 of data are support vectors!!!

• In linear SVMs:
  – Typical setup: 4000 dim. HOG, only 300 “chair” examples

Figure that Francis Bach hates
Recognition Learning Spectrum

Extrapolation problem
Generalization

Interpolation problem
Correspondence

Number of training samples

Traditional datasets

1 10 10² 10³ 10⁴ 10⁵ 10⁶

∞
Everything else being equal…

… the visual world is just much richer!

- **MNIST Digits**
  - 10 digits *
  - ~1,000 variations = 10,000

- **English words**
  - ~100,000 words *
  - ~5 variations = 500,000

- **Visual world**
  - ~100,000 objects *
  - ~10,000 variations (pose, scale, lighting, intra-category)
  - = 1,000,000,000 (1 billion!)
Yet, we train on 15 examples?!
If you want to start a company…

- Data
- Features
- Algorithm

Amnon Shashua
To make research progress…

Data / Features / Algorithms
Big Message...

Keep the data -- you never know when you will need it!
Texture

- Texture depicts spatially repeating patterns
- Many natural phenomena are textures

radishes  rocks  yogurt
Texture Synthesis
Classical Texture Synthesis

Synthesis

Parametric Texture Model

Analysis

Sample texture

Novel texture

This is hard!
Throwing away too much too soon?

input texture

synthesized texture
Non-parametric Approach

Synthesis

Analysis

Sample texture

Novel texture
Motivation from Language

- [Shannon, ’48] proposed a way to generate English-looking text using N-grams:
  - Assume a generalized Markov model
  - Use a large text to compute prob. distributions of each letter given N-1 previous letters
  - Starting from a seed repeatedly sample this Markov chain to generate new letters
  - Also works for whole words

WE NEED TO EAT CAKE
Results (using alt.singles corpus):
  – “As I've commented before, really relating to someone involves standing next to impossible.”
  – “One morning I shot an elephant in my arms and kissed him.”
  – “I spent an interesting evening recently with a grain of salt”

Notice how well local structure is preserved!
  – Now, instead of letters let’s try pixels…
[Efros & Leung, ’99]

Input image

non-parametric sampling

Input image
Texture Growing
Homage to Shannon

Ingrain the unsatisfactory!
Mr. Dick Gephardt was fail-

ful riffs on the looming

ly asked, "What's your

story about the emerg-

people about continua-

Gephardt began, patiently ob-

s, that the legal system has

with this latest tangle.
Two Kinds of Things in the World

Navier-Stokes Equation

\[ \frac{\partial u}{\partial t} = - (u \cdot \nabla) u + \nu \nabla^2 u - \frac{1}{\rho} \nabla p + f \]

+ weather
+ location
+ ...
Lots of data available
“Unreasonable Effectiveness of Data”  
[Halevy, Norvig, Pereira 2009]

- Parts of our world can be explained by elegant mathematics:
  - physics, chemistry, astronomy, etc.

- But much cannot:
  - psychology, genetics, economics, etc.

- Enter: The Magic of **Big Data**
  - Great advances in several fields:
    - e.g. speech recognition, machine translation, Google
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 Good News

Really stupid algorithms + Lots of Data

= “Unreasonable Effectiveness”
The Bad News

Visual Data is much more difficult

- **text:**
  - clean, segmented, compact, 1D

- **Visual data:**
  - Noisy, unsegmented, high entropy, 2D/3D
Distance Metrics

\[ CLIME - CRIME \] = hamming distance of 1 letter

\[ \begin{array}{c|c}
\text{Y} & \text{x} \\
\hline
\end{array} \begin{array}{c|c}
\text{Y} & \text{x} \\
\hline
\end{array} \] = Euclidian distance of 5 units

\[ \begin{array}{c|c}
\text{Y} & \text{x} \\
\hline
\end{array} \begin{array}{c|c}
\text{Y} & \text{x} \\
\hline
\end{array} \] = Grayvalue distance of 50 values

\[ \begin{array}{c|c}
\text{Y} & \text{x} \\
\hline
\end{array} \begin{array}{c|c}
\text{Y} & \text{x} \\
\hline
\end{array} \] = ?
L2 norm says these are not similar
Make them tiny!
Lots of Tiny Images

c) Segmentation of 32x32 images
Human Scene Recognition

![Graph showing correct recognition rate and true positive rate as a function of image resolution for color and grayscale images. The graph is labeled as (a) Scene recognition.]
Lots Of Images

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

Grayscale input High resolution

Colorization of input using average

A. Torralba, R. Fergus, W.T.Freeman. 2008
Not a pixel lover? No problem!

Let’s match gradients
[Hays and Efros. Scene Completion Using Millions of Photographs. SIGGRAPH 2007 and CACM October 2008.]
Efros and Leung result
Scene Matching for Image Completion
Scene Completion Result
The Algorithm
Scene Matching
Scene Descriptor

Scene Gist Descriptor

(Oliva and Torralba 2001)
Scene Descriptor

Scene Gist Descriptor
(Oliva and Torralba 2001)
Scene Descriptor

+ 

[Diagram showing a color distribution matrix with axes labeled as Frequency and Edge Orientation, and color intensity indicating high and low edge energy]
2 Million Flickr Images
Context Matching
Graph cut + Poisson blending
... 200 scene matches
Why does it work?
Nearest neighbors from a collection of 20 thousand images
Nearest neighbors from a collection of 2 million images
Scene matching with camera transformations

Sivic, Kaneva, Torralba, Avidan, Freeman, Internet Vision Workshop, 2008
Creating and exploring a large, photorealistic virtual space

Josef Sivic (INRIA/ENS), Biliana Kaneva (MIT), Antonio Torralba (MIT), Shai Avidan (Adobe) and Bill Freeman (MIT)

IEEE Workshop on Internet Vision, 2008
Cross-Domain Matching

“Local – bad
Global – good!”
(after Orwell)
Medici Fountain, Paris
Search by image

Drop image here

Watch a short video to learn more.
Medici Fountain, Paris (winter)
Google search results for Luxembourg Gardens.

- Image size: 713 × 600
- No other sizes of this image found.

Visually similar images related to Luxembourg Gardens.
Our Goal [SIGGRAPH Asia’11]
WHY IS THIS SO HARD?
EXAMPLE: SIFT MATCHING

[SIFT: Lowe, 2004]
Example: SIFT Matching

[SIFT: Lowe, 2004]
Input Query

Top Matches
“Data-driven Uniqueness”
Possible Explanation

- If the space of images was uniform, nearest neighbor would work perfectly well.
- But the space is very non-linear, non-Euclidean.
- The Exemplar-SVM is trying to make a small, linear rescaling of visual space, near the query point.
  - i.e. capturing the natural image statistics near the exemplar.
  - Or maybe global is good enough (see Deva).
FEATURE REPRESENTATION
HISTOGRAM OF ORIENTED GRADIENTS (HOG)

[Dalal and Triggs, CVPR, 2005]
Input Query

Learnt Weights

HOG

Top Match

Top Match
SEARCH USING IMAGES

Input Query

Top Matches
SEARCH USING IMAGES

Input Query

Top Matches
SEARCH USING IMAGES

Input Query

Top Matches
SEARCH USING PAINTINGS

Input Painting

Our Approach

GIST

Bag-of-Words

Tiny Images

HOG
SEARCH USING PAINTINGS
SEARCH USING PAINTINGS

Input Painting

Top Matches
SEARCH USING SKETCHES

Input Sketch

Tiny Images

GIST

Bag-of-Words

HOG

Our Approach
SEARCH USING SKETCHES
SALIENCY

[Yarbus, 1962]
PROXY FOR S Aliency?
PREDICTING Saliency

Saliency Dataset [Judd et al., 2009]
APPLICATIONS
WHERE WAS THE PAINTER STANDING?

Input Painting
PAINTING2GPS

Input Painting

Retrieval set
10,000 Geo-tagged Flickr Images
100 top matches used to estimation
PAINTING2GPS

Input Painting

Estimated Geo-location

Estimated using 100 top matches
Input Painting

Sydney Opera House
VISUAL SCENE EXPLORATION
VISUAL SCENE EXPLORATION
Organizing Visual “Garbage Heap”

“It irritated him that the “dog” of 3:14 in the afternoon, seen in profile, should be indicated by the same noun as the dog of 3:15, seen frontally…”

“My memory, sir, is like a garbage heap.”

-- Jorge Luis Borges, *Funes the Memorious*
PHOTOSYNTH
[Snavely et al., 2006]

Dataset size: 136 photos (from flickr)
# of discovered synths: 14
82 photos not part of any synth
FINDING SIMILAR IMAGES

Query image
PAIRWISE SIMILARITY MATRIX
TRAIVERSING THE GRAPH
Ways to use Big Data

1. See what different subsets of data think of you
2. Use kNN to make a smaller, cozier sub-problem
3. Visual Data Mining: find needles in a haystack
4. Ditch Categories – keep all instances and connect them
Priors for Large Photo Collections

&

What they Reveal about Cameras

Sujit Kuthirummal
Aseem Agarwala
Dan B Goldman
Shree K. Nayar

Columbia University
Adobe Systems, Inc.
Adobe Systems, Inc.
Columbia University
Ways to use Big Data

1. See what different subsets of data think of you
2. Use kNN to make a smaller, cozier sub-problem
3. Visual Data Mining: find needles in a haystack
4. Ditch Categories – keep all instances and connect them
1. kNN + Label Transfer

Sky, Water, Hills, Beach, Sunny, mid-day
80 Million Tiny Images [PAMI’08]

a) Input image

b) Neighbors
c) Ground truth
d) Wordnet voted branches

Torralba, Fergus, Freeman, PAMI 2008
Non-parametric Scene Parsing
[CVPR’09]
im2gps [CVPR’08]

Query Photograph

Hays & Efros, CVPR 2008
Ways to use Big Data

1. See what different subsets of data think of you
2. Use kNN to make a smaller, cozier sub-problem
3. Visual Data Mining: find needles in a haystack
4. Ditch Categories – keep all instances and connect them
Visual Words or Letters?
Spectrum of Visual Features

Low-Level
- Pixel
- Filter-Banks
- Sparse-SIFT

High-Level
- Parts, Segments
- Objects
- Image

Visual Words

Our Approach
Discriminative Patches

Two key requirements

1. Need to occur frequently (representative)

2. …but not **too** frequently
   
   Discriminative: Need to be different enough from the rest of the visual world.
First some examples
Finding needles in a haystack
K-Means Clusters
K-Means
Discriminative K-means
Discriminative Clustering+
Discriminative Clustering+
Discriminative Clustering++

KMeans

Iter 1

Iter 2

Iter 3

Iter 4
Discriminative Clustering++

KMeans

Iter 1

Iter 2

Iter 3

Iter 4
More Discovered Patches
What makes Paris look like Paris?

Carl Doersch, Saurabh Singh, Abhinav Gupta, Josef Sivic and Alexei Efros, [SIGGRAPH ’12]
How well can people do?

- [http://baikal.graphics.cs.cmu.edu/cdoersch/im2gps2/corr/test2.html](http://baikal.graphics.cs.cmu.edu/cdoersch/im2gps2/corr/test2.html)
How well can people do?

- [Link](http://baikal.graphics.cs.cmu.edu/cdoersch/im2gps2/corr/test2.html)
- Mean performance = 79%
  - Subjects who have been to Paris
    (up to 90% if allowed to scrutinize the images)
What makes Paris look like Paris?
What makes Paris look like Paris?
Goal

To automatically discover geo-informative visual elements, that (hopefuly) capture the "look and feel" of a place
Goal

To automatically discover geo-informative visual elements, that are:

• **Representative**: frequently occur in Paris.

• **Discriminative**: occur in Paris but not at other places.
Need both conditions

Discriminative only:
Need both conditions

Frequently occurring only:
Extracted Elements from Barcelona
Correspondence

Paris, France

Prague, Czech Republic

London, England
Many Elements Capture Context
Mapping architectural features

Figure 6: Examples of geographic patterns in Paris (shown as red dots on the maps) for three discovered visual elements (shown below each map). Balconies with cast-iron railings are concentrated on the main boulevards (left). Windows with railings mostly occur on smaller streets (middle). Arch supporting columns are concentrated on Place des Vosges and the St. Germain market (right).
Ways to use Big Data

1. See what different subsets of data think of you

2. Use kNN to make a smaller, cozier sub-problem

3. Visual Data Mining: find needles in a haystack

4. Ditch Categories – keep all instances and connect them
Down with Categories!!!

Ceci n'est pas une pipe.

Alexei (Alyosha) Efros
CMU
Acknowledgements

Murphy
Big Book of Concepts

Weinberger
Everything is Miscellaneous

Talks by Moshe Bar; writings of Shimon Edelman

Many great discussions with many colleagues, especially Tomasz Malisiewicz, James Hays, and Derek Hoiem
Understanding an Image
Object naming -> Object categorization

- sky
- building
- flag
- banner
- face
- street lamp
- wall
- bus
- cars

Slide by Fei Fei, Fergus & Torralba.
Object categorization

sky
flag
banner
bus
building
street lamp
wall
face
cars
Why Categorize?

1. Knowledge Transfer
2. Communication
Classical View of Categories

• Dates back to Plato & Aristotle
  1. Categories are defined by a list of properties shared by all elements in a category
  2. Category membership is binary
  3. Every member in the category is equal
Problems with Classical View

• Humans don’t do this!
  – People don’t rely on abstract definitions / lists of shared properties (Wittgenstein 1953, Rosch 1973)
    • e.g. define the properties shared by all “games”
    • e.g. are curtains furniture? Are olives fruit?
  – Typicality
    • e.g. Chicken -> bird, but bird -> eagle, pigeon, etc.
  – Language-dependent
    • e.g. “Women, Fire, and Dangerous Things” category is Australian aboriginal language (Lakoff 1987)
  – Doesn’t work even in human-defined domains
    • e.g. Is Pluto a planet?
Problems with **Visual Categories**

- A lot of categories are functional
- World is too varied
- Categories are 3D, but images are 2D
Typical HOG car detector

Felzenszwalb et al, PASCAL 2007
Why not?
Solution: hierarchy?

Ontologies, hierarchies, levels of categories (Rosch), etc.
WordNet, ImageNet, etc etc
Still Problematic!

– Intransitivity
  • e.g. car seat is chair, chair is furniture, but …
– Multiple category membership
  • it’s not a tree, it’s a forest!

Clay Shirky, “Ontologies are Overrated”
Fundamental Problem with Categorization

Making decisions too early!
We should only categorize at run-time, once we know the task!
The Dictatorship of Librarians
categories are losing…

vs.

YAHOO!

Google
On-the-fly Categorization?

1. Knowledge Transfer
2. Communication
Association instead of categorization

Ask not “what is this?”, ask “what is this like”

– Moshe Bar

• Exemplar Theory (Medin & Schaffer 1978, Nosofsky 1986, Krushke 1992)
  – categories represented in terms of remembered objects (exemplars)
  – Similarity is measured between input and all exemplars
  – think non-parametric density estimation

• Vanevar Bush (1945), Memex (MEMory EXtender)
  – Inspired hypertext, WWW, Google…
Bush’s Memex (1945)

- Store publications, correspondence, personal work, on microfilm
- Items retrieved rapidly using index codes
  - Builds on “rapid selector”
- Can annotate text with margin notes, comments
- Can construct a *trail* through the material and save it
  - Roots of hypertext
- Acts as an external memory
Visual Memex, a proposal
[Malisiewicz & Efros]

Nodes = instances
Edges = associations

types of edges:
• visual similarity
• spatial, temporal co-occurrence
• geometric structure
• language
• geography
• ..
Object Detection

“bus”
Exemplar Object Detection
Ensemble of Exemplar-SVMs
“What is this?”

“What is this *like*?”

Malisiewicz & Efros, CVPR’08
Image Parsing with Context

Figure 1: The **Visual Memex** graph encodes object similarity (solid black edge) and spatial context (dotted red edge) between pairs of object exemplars. A spatial context feature is stored for each context edge. The Memex graph can be used to interpret a new image (left) by associating image segments with exemplars in the graph (orange edges) and propagating the information.
Torralba’s Context Challenge
Torralba’s Context Challenge
Torralba’s Context Challenge
Our Challenge Setup

Figure 2: Torralba’s Context Challenge: “How far can you go without running a local object detector?” The task is to reason about the identity of the hidden object (denoted by a “?”) without local information. In our category-free Visual Memex model, object predictions are generated in the form of exemplar associations for the hidden object. In a category-based model, the category of the hidden object is directly estimated.

Malisiewicz & Efros, NIPS’09
3 models

Visual Memex: exemplars, non-parametric object-object relationships
  • Recurse through the graph

Baseline: CoLA: categories, parametric object-object relationships

Reduced Memex: categories, non-parametric relationships
Qual. results

<table>
<thead>
<tr>
<th>Input Image + Hidden Region</th>
<th>Visual Memex Exemplar Predictions</th>
<th>Categorization Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image 1" /></td>
<td><img src="predictions1.png" alt="Predictions 1" /></td>
<td><img src="results1.png" alt="Results 1" /></td>
</tr>
<tr>
<td><img src="image2.png" alt="Image 2" /></td>
<td><img src="predictions2.png" alt="Predictions 2" /></td>
<td><img src="results2.png" alt="Results 2" /></td>
</tr>
<tr>
<td><img src="image3.png" alt="Image 3" /></td>
<td><img src="predictions3.png" alt="Predictions 3" /></td>
<td><img src="results3.png" alt="Results 3" /></td>
</tr>
</tbody>
</table>
Figure 3: a.) Context Challenge confusion matrices for the 3 methods: Visual Memex, KDE, and CoLA. b.) Recognition Precision versus Recall when thresholding output based on confidence. c) Side by side comparison of the 3 methods’ accuracies for 30 categories.
will Big Data solve all your problems?
1. Data is Biased

- Internet is a tremendous repository of visual data (Flickr, YouTube, Picassa, etc)
- But it’s not random samples of visual world
My Paris
Real Notre Dame
Sampling Bias

- People like to take pictures on vacation
Photographer Bias

- People want their pictures to be recognizable and/or interesting
Social Bias

“100 Special Moments” by Jason Salavon
Social Bias

Mildred and Lisa

Source: U.S. Social Security Administration

Gallagher et al CVPR 2008
Social Bias

Gallagher et al CVPR 2008

Gallagher et al, CVPR 2009
2. We will never have enough data
Long Tails -- Unfamiliar is Common

10% of the objects account for 90% of the data

~Zipf’s law

Slide by Antonio Torralba
Dealing with sparse data (rare scenes)

Quick Fixes:
better alignment
  • e.g. reduce resolution, sifting, warping, etc.

Understand the simple stuff first
Recognize when it’s easy!

People take on a variety of poses, aspects, scales

- self-occlusion
- rare pose
- motion blur

- non-distinctive pose
- too small
- just right
detect this

Ramanan, Forsyth, Zisserman, 2004
“Poping out” foreground objects

Hoiem et al, ICCV 2007

Figure 10. Object popout. We show five out of the fifteen most “solid” regions in the Geometric Context dataset. Our algorithm often finds foreground objects, which would be helpful for unsupervised object discovery [21].
Guess structure

David C. Lee, Martial Hebert, Takeo Kanade, CVPR’09
Guess structure

David C. Lee, Martial Hebert, Takeo Kanade, CVPR’09
Subtracting away structure

Structure

Objects

Wall appearance modeling

David C. Lee, Martial Hebert, Takeo Kanade, CVPR’09
Dealing with sparse data (rare scenes)

Long-term Fixes:

segment into chunks
  • e.g. segmentation for recognition approaches

Attributes – densifying the labels
From categorization to association
  • Ask not “what is this?”, ask “what is this like?”
Conclusion…

“If you torture data long enough, it might confess”
- Ronald Coase