16-824: Visual Learning and Recognition

Many slides from A. Farhadi, A. Efros
Course Information

• Time:
  – Monday, Wednesday 1:30-2:50

• Location:
  – NSH 1305

• Office Hours:
  – Email me for appointments

• Contact:
  – abhinavg@cs, EDSH 213

• Website:
People - Instructor

- Abhinav Gupta
- Ph.D. 2009, University of Maryland
Pitcher pitches the ball before Batter hits. Batter hits and then simultaneously Batter runs to base and Fielder runs towards the ball. Fielder runs towards the ball and then Fielder catches the ball. Fielder catches the ball and then Fielder throws to the base. Fielder at Base catches the ball at base after Fielder throws to the base.
People

• Abhinav Gupta
• Ph.D. 2009, University of Maryland
• Postdoctoral Fellow, Carnegie Mellon University, 2009-11
blocks world revisited

Original Image

3D Parse Graph

All results and Code: http://www.cs.cmu.edu/~abhinavg/blocksworld
People

• David Fouhey
• Ph.D. Student, Robotics Institute
People

• David Fouhey
• Ph.D. Student, Robotics Institute
• Research Interests
  – 3D Scene Understanding
  – Understanding Humans
People - TA

- Xiaolong Wang
- PhD Student, Robotics Institute
- Working with me
- Research Interests:
  - Learning Visual Representation via ConvNets
  - Representing actions via ConvNets
People - TA

- Rohit Girdhar
- MS Student, Robotics Institute
- Working with me
- Research Interests:
  - 3D Understanding
  - Affordances
16-824: Learning-based Methods in Vision

What is this course about?
What is the goal of Computer Vision?

Systems that can “understand” Visual Data
understanding visual data
understanding visual data
understanding visual data
What does it mean to understand?
The Vision Story Begins…

“What does it mean, to see? The plain man's answer (and Aristotle's, too). would be, to know what is where by looking.”

Vision: a split personality

“What does it mean, to see? The plain man’s answer (and Aristotle’s, too). would be, to know what is where by looking. In other words, vision is the process of discovering from images what is present in the world, and where it is.”

Answer #1: pixel of brightness 243 at position (124,54) …and depth .7 meters

Answer #2: looks like flat sittable surface of the couch

Which do we want? Is the difference just a matter of scale or is there some fundamental difference?
Measurement vs. Perception
Brightness: Measurement vs. Perception

Edward H. Adelson
Brightness: Measurement vs. Perception

Proof!

Slide Credit: Alyosha
Measurement

Length

Müller-Lyer Illusion

http://www.michaelbach.de/ot/sze_muelue/index.html
Measurement
Capturing physical quantities like pixel brightness, depth, etc.

Perception/Understanding

• a high-level representation that captures the semantic structure of the scene and its constituent objects.

• Subjective - Depends on Task and Agent

• Intersection of what you see and what you believe (prior knowledge)
...but why do we care about perception?

The goals of computer vision (what + where) are in terms of what humans care about.
So what do humans care about?
Image Classification/ Scene Recognition

Living Room
Object Detection

Couch

Table
Object Segmentation/Categorization
3D Understanding
Functional Understanding

- Can Sit
- Can Move
- Can Push
- Can Walk
Pose Estimation:
Activity Recognition:

What is he doing?
Why are these problems hard?
Challenges 1: view point variation

Michelangelo 1475-1564

slide by Fei Fei, Fergus & Torralba
Challenges 2: illumination
Challenges 3: occlusion

Magritte, 1957

slide by Fei Fei, Fergus & Torralba
Challenges 4: scale
Challenges 5: deformation

Xu, Beihong 1943
Challenges 6: background clutter

Klimt, 1913

slide by Fei Fei, Fergus & Torralba
Challenges 7: object intra-class variation
Challenges 8: local ambiguity

slide by Fei-Fei, Fergus & Torralba
Challenges 9: the world behind the image
ill-posed

- EXAMPLE:
- Recovering 3D geometry from single 2D projection

- Infinite number of possible solutions!

from [Sinha and Adelson 1993]
How do we solve it?
Data to Rescue !!
• Data to build observation models..
• Data to build priors about the visual world.
• Use the models and prior information to infer..

Machine-Learning!
In this course, we will:

Take a few baby steps...
Technical Challenges
Challenges: scale, efficiency

- **flickr**: 6 billion images
- **Facebook**: 70 billion images
- **Imgur**: 1 billion images served daily
- **YouTube**: 100 hours uploaded per minute
- **From Cisco**: Almost 90% of web traffic is visual!
Challenges: scale, efficiency

~10,000 to 30,000 object categories

Fei Fei Li, Rob Fergus, Antonio Torralba

Biederman 1987
Challenges: learning with minimal supervision

Less:
- Unlabeled, multiple objects

More:
- Cropped to object, parts and classes

Classes labeled, some clutter

Kristen Grauman
This is a pottopod
Find the pottopod

Slide from Pietro Perona, 2004 Object Recognition workshop
What to expect in the class?
Graphical Models

Learning as a tool to exploit big data, build prior models etc.

Not formulate problem in complicated manner...
But that said...

- We will still look at the learning methods which give the state of the art performance on these tasks.

- For example, most focus this year will be on deep learning - Convolutional Neural Networks (CNN)...
Is this a research course?

• One year ago - YES!

• But times have changed: Computer Vision is a hot topic in industry now..

• 2012 - Resurgence of Deep Networks (CNNs)
2014 - Deep Learning is Everywhere

- Google, Facebook, Baidu, Apple
  - Strong deep learning groups hiring everywhere..
  - Beyond Research: Development
    - Image Search
    - Automated Driving

Startups Sold Everyday
- Vision Factory, EuVision, Flutter...

*Come Back to this in Next Class!*
Course Outline
Goals

• Read some interesting papers together
  – Learn something new: both you and us!

• Get up to speed on big chunk of vision research
  – understand 70% of CVPR papers!

• Use learning-based vision in your own work

• Learn how to speak
• Learn how think critically about papers
Course Organization

• Requirements:
  1. Class Participation (15%)
     • Keep annotated bibliography
     • Post on the Class Blog before each class
     • Ask questions / debate / flight / be involved!
  2. Presentation (20 %)
  3. Project (25%)
  4. Assignment (2x20%)
Class Participation

• Keep annotated bibliography of papers you read (always a good idea!). The format is up to you. At least, it needs to have:
  – Summary of key points
  – A few Interesting insights, “aha moments”, keen observations, etc.
  – Weaknesses of approach. Unanswered questions. Areas of further investigation, improvement.

• Submit a comment on the Class Blog
  – ask a question, answer a question, post your thoughts, praise, criticism, start a discussion, etc.
Presentation

1. Pick a topic from the list

2. Understand it as if you were the author
   - If there is code, understand the code completely

3. Prepare an amazing 15min presentation
   - Discuss with me/David before the presentation, 5 days before the presentation
Class Assignment

Two assignments to get you familiar with deep learning.

Toolboxes
- CAFFE
- TORCH

Fine-tuning and Learning-from-scratch
Class Project

Opportunity to work on the crazy idea which your advisor would not let you do! (Group of 2-3)

Merit Criteria
1. Crazy (the more different it sounds the better it is)
2. Amount of Work/Results.
3. Report/Presentation

Failure/Success has no points! An idea with interesting failure results is a successful project!
End of Semester Awards

• We will vote for:
  – Best Project
  – Best Presentation
Logistics

• Waitlist - Class size restricted to 51 students

• Talk to me after class!