CS4501: Introduction to Computer Vision

Introduction

Various slides from previous courses by:
D.A. Forsyth (Berkeley / UIUC), I. Kokkinos (Ecole Centrale / UCL), S. Lazebnik (UNC / UIUC), S. Seitz (MSR / Facebook), J. Hays (Brown / Georgia Tech), A. Berg (Stony Brook / UNC), D. Samaras (Stony Brook). J. M. Frahm (UNC), V. Ordonez (UVA).
Today’s Class

Who am I?
What is Computer Vision?
Why is Computer Vision Hard?
Cameras
Questions
About Me

Assistant Professor
2016 - Now

Visiting Professor
2019 - Now

Visiting Researcher
2015 - 2016

MS, PhD in CS,
2009-2015

... also spent time at:

Adobe Research

Allen Institute for Artificial Intelligence

The University of North Carolina at Chapel Hill

Stony Brook University

Google
Microsoft
eBay
Describing Images with Language

NEW! Obj2Text: Generating Visually Descriptive Language from Object Layouts
Xuwang Yin, Vicente Ordonez.

Im2Text: Describing Images Using 1 Million Captioned Photographs
Vicente Ordonez, Girish Kulkarni, Tamara L. Berg.

Large Scale Retrieval and Generation of Image Descriptions
International Journal of Computer Vision. IJCV 2015. [August 2016 Issue]. [pdf] [link] [bibtex]
Naming Objects

Superordinates: animal, vertebrate
Basic Level: bird
Entry Level: bird
Subordinates: American robin

Superordinates: animal, vertebrate
Basic Level: bird
Entry Level: penguin
Subordinates: Chinstrap penguin

From Large Scale Image Categorization to Entry-Level Categories
Vicente Ordonez, Jia Deng, Yejin Choi, Alexander C. Berg, Tamara L. Berg.

Predicting Entry-Level Categories
Vicente Ordonez, Wei Liu, Jia Deng, Yejin Choi, Alexander C. Berg, Tamara L. Berg.

Learning to Name Objects
Vicente Ordonez, Wei Liu, Jia Deng, Yejin Choi, Alexander C. Berg, Tamara L. Berg.
Communications of the ACM. March 2016 (Vol. 59, No. 3). (~Research Highlight)
Recognizing Commonly Uncommon Situations

Commonly Uncommon: Semantic Sparsity in Situation Recognition

http://imsitu.org/demo/
Accelerating Neural Networks: XNOR-Net

<table>
<thead>
<tr>
<th>Network Variations</th>
<th>Operations used in Convolution</th>
<th>Memory Saving (Inference)</th>
<th>Computation Saving (Inference)</th>
<th>Accuracy on ImageNet (AlexNet)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Convolutions</td>
<td>Real-Value Inputs</td>
<td>+, -, ×</td>
<td>1x</td>
<td>%56.7</td>
</tr>
<tr>
<td></td>
<td>Real-Value Weights</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Binary Weight Coefficients</td>
<td>Real-Value Inputs</td>
<td>+, -</td>
<td>~32x</td>
<td>%56.8</td>
</tr>
<tr>
<td></td>
<td>Real-Value Weights</td>
<td></td>
<td>~2x</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Binary Weight Coefficients</td>
<td>Binary Inputs</td>
<td>XNOR, bitcount</td>
<td>~32x</td>
<td>%44.2</td>
</tr>
<tr>
<td></td>
<td>Binary Weights</td>
<td></td>
<td>~58x</td>
<td></td>
</tr>
</tbody>
</table>

**XNOR-Net: ImageNet Classification Using Binary Convolutional Neural Networks**

Mohammad Rastegari, Vicente Ordonez, Joseph Redmon, Ali Farhadi.

Removing Gender Bias from Situation Recognition

NEW! Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints
Jiyeu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez, Kai-Wei Chang.

NEW! Balanced Datasets Are Not Enough: Estimating and Mitigating Gender Bias in Deep Image Representations
Tianlu Wang, Jiyeu Zhao, Mark Yatskar, Kai-Wei Chang, Vicente Ordonez.
Synthesizing Images from Text

Mike is surprised at the duck. The duck is standing on the grill. Jenny is running towards Mike and the duck.

A guy on a motorcycle with some people watching.

Several elephants walking together in a line near water.

NEW! Text2Scene: Generating Compositional Scenes from Textual Descriptions
Fuwen Tan, Song Feng, Vicente Ordonez.
Long Beach, California. June 2019. [arxiv] [bibtex] (Oral presentation + Best Paper Finalist — top 1% of submissions)
What is Computer Vision?
Make computers understand images and video

What kind of scene?

Where are the cars?

How far is the building?

…

Slide by James Hays.
Why computer vision matters

Safety
Health
Security
Comfort
Fun
Access
Computer Vision

Create an algorithm to distinguish dogs from cats

Face Detection in Cameras

Birdsnap
Relationship with Other Fields

- Computer Vision: Image ➔ Knowledge

[Diagram showing deer and cat]
Relationship with Other Fields

- Image Processing: Image $\rightarrow$ Image
Relationship with Other Fields

- Computer Graphics: Knowledge $\rightarrow$ Image

Vertices, Locations, Objects, Shapes, Colors, Material properties, Lighting settings, Camera settings, etc.
Ridiculously brief history of computer vision

- 1960’s: interpretation of synthetic worlds
- 1970’s: some progress on interpreting selected images
- 1980’s: Neural Networks come and go; shift toward geometry
- 1990’s: face recognition; statistical analysis in vogue
- 2000’s: broader recognition; large annotated datasets available; video processing starts
- 2010’s: Deep learning with ConvNets
- 2030’s: ?
How vision is used now

• Examples of real world applications
Optical character recognition (OCR)

Technology to convert scanned docs to text

- If you have a scanner, it probably came with OCR software

Digit recognition, AT&T labs
http://www.research.att.com/~yann/

License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition
Face detection

• Digital cameras detect faces
Smile detection

The Smile Shutter flow

Imagine a camera smart enough to catch every smile! In Smile Shutter Mode, your Cyber-shot® camera can automatically trip the shutter at just the right instant to catch the perfect expression.

Sony Cyber-shot® T70 Digital Still Camera
Vision-based biometrics

“How the Afghan Girl was Identified by Her Iris Patterns” Read the story wikipedia
Login without a password...

Fingerprint scanners on many new laptops, other devices

Face recognition systems now beginning to appear more widely
http://www.sensiblevision.com/
Object recognition (in mobile phones)

Point & Find, Nokia
Google Goggles
iNaturalist

https://www.inaturalist.org/pages/computer_vision_demo
Special effects: shape capture

*The Matrix* movies, ESC Entertainment, XYZRGB, NRC
Special effects: motion capture

Pirates of the Caribbean, Industrial Light and Magic
Sports

Sportvision first down line
Nice explanation on www.howstuffworks.com

http://www.sportvision.com/video.html
Medical Imaging

3D imaging
MRI, CT

Image guided surgery
Grimson et al., MIT
Smart cars

- Mobileye
  - Market Capitalization: 11 Billion dollars
Self-driving Cars e.g. Google’s Waymo

June 24, 2011. "Nevada state law paves the way for driverless cars". Financial Post. Christine Dobby
Aug 9, 2011, "Human error blamed after Google’s driverless car sparks five-vehicle crash". The Star (Toronto)
Ford acquires SAIPS for self-driving machine learning and computer vision tech

Posted Aug 16, 2016 by Darrell Etherington (@etherington)

Ford outlined a few of the ways it's aiming to ship driverless cars by 2021, and part of the plan involves acquisitions. CEO Mark Fields revealed at a press event in Palo Alto today that the automaker acquired SAIPS, an Israeli company focusing on machine learning and computer vision. It's also partnering exclusively with Nirenberg Neuroscience, to bring more "humanlike intelligence" to machine learning components of driverless car systems.

SAIPS' technology brings image and video processing algorithms, as well as deep learning tech focused on processing and classifying input signals, all key ingredients in the special sauce that makes up autonomous vehicle tech. This company's expertise should help with on-board interpretation of data captured by sensors on Ford's self-driving cars, and turning that data into usable info for the car's virtual driver system. SAIPS' offerings include detection of anomalies, persistent tracking of objects detected by sensors, and much more. The company's past clients include HP and Trax, but its partner group doesn't appear to have included much in the way of device-specific applications.

CrunchBase

Ford Motor Company

FOUNDED
1903

OVERVIEW
Ford is an automotive company that develops, manufactures, distributes, and services vehicles, parts, and accessories worldwide. It operates through two sectors: automotive and financial services. The automotive sector offers vehicles primarily under the Ford and Lincoln brand names. This sector markets cars, trucks, parts, and accessories through retail dealers in North America and distributors...

LOCATION
Dearborn, MI

CATEGORIES
Automotive

WEBSITE
http://www.ford.com/

Full profile for Ford Motor Company

TC NEWSLETTERS

- The Daily Crunch
  Our top headlines
  Delivered daily

- TC Week-in-Review
  Top stories of the week
  Delivered weekly

- CrunchBase Daily
  Daily tech news from CrunchBase
Interactive Games: Kinect – (Maybe)

- Object Recognition: [video](http://www.youtube.com/watch?feature=iv&v=fQ59dXOo63o)
- Mario: [video](http://www.youtube.com/watch?v=8CTJL5IUjHg)
- 3D: [video](http://www.youtube.com/watch?v=7QrnwoO1-8A)
- Robot: [video](http://www.youtube.com/watch?v=w8BmgMKFbY)
Industrial robots

Vision-guided robots position nut runners on wheels
Vision in space

Vision systems (JPL) used for several tasks

• Panorama stitching
• 3D terrain modeling
• Obstacle detection, position tracking
• For more, read “Computer Vision on Mars” by Matthies et al.
Augmented Reality and Virtual Reality

Magic Leap, Oculus, Hololens, etc.
Vision is really hard

- Vision is an amazing feat of natural intelligence
- Visual cortex occupies about 50% of Macaque brain
- One third of human brain devoted to vision (more than anything else)
Is seeing trivial?
Is seeing trivial?

http://web.mit.edu/persci/people/adelson/checkershadow_illusion.html
Is seeing trivial?
Is seeing trivial?
Is seeing trivial?
Face or non-face?
Face or non-face?
Why is vision so hard?
This is an image to us:
This is an image to a computer:
Vision is Hard
Illumination
Occlusion
Deformation

Xu, Beihong 1943
Intra-class variation

slide by Fei-Fei, Fergus & Torralba
What is the state of the art today?

• Given enough training data Computer Vision systems are surprisingly robust to the previously outlined challenges e.g. illumination changes, intra-class variation.

• Still not at the same level as humans, despite the hype.

• Still many open challenges, such as few-shot learning, transfer learning, and unsupervised learning.
Deep Learning and Vision

- Deep Learning has been a great disruption into the field of Computer Vision. Has made a lot of new things work!

- Many deep learning methods being applied to vision these days.

- This is not a deep learning course. We will study the important pre-deep learning methods, and then some deep learning.
About the Course

CS4501-008: Introduction to Computer Vision

• Instructor: Vicente Ordóñez
• Email: vicente@virginia.edu
• Website: http://vicenteordonez.com/vision/
• Class Location: Thornton Hall E316
• Class Times: Monday-Wednesday 5pm - 6:15pm
• Piazza?
• Office hours: TBD
Teaching Assistants

Paola Cascante-Bonilla (pc9za at virginia.edu)
Office Hours: Tuesdays from 1pm to 2pm (Rice Hall 442)

Ziyan Yang (zy3cx at virginia.edu)
Office Hours: Thursdays from 12:30pm to 2:30pm (Rice Hall 442).
Objectives

• Develop Intuitions Between Human Vision and Computer Vision

• Understanding the Basics of 2D and 3D Computer Vision

• Become familiar with the technical approaches in computer vision such as registration, matching, and recognition

• Obtain practical experience in the implementation of computer vision applications.
Pre-requisites

• Python programming skills
• Calculus / Linear Algebra / Probability
Grading

• Assignments: 50% (5 assignments)
  (10% + 10% + 10% + 10% + 10%)

• Quiz: 20% (2 quizzes)
  (10% + 10%)

• Final project: 30% (group project – 3 people max)
Textbook


Cameras
Cameras

Polaroid
$100

EOS Rebel T6i
$900

Canon EOS C300
$40,000
What do you need to make a camera from scratch?
Accidental Pinhole and Pinspeck Cameras
Revealing the scene outside the picture.
Antonio Torralba, William T. Freeman
Accidental Cameras

a) Input (occluder present)  

b) Reference (occluder absent)

c) Difference image (b-a)  
d) Crop upside down  
e) True view
Let’s design a camera
- Idea 1: put a piece of film in front of an object
- Do we get a reasonable image?
Idea 2: add a barrier to block off most of the rays
  • This reduces blurring
  • The opening known as the aperture
Camera obscura
Known by the Greeks and the Chinese 470BC-322BC

Recorded in writings by Chinese Philosopher Mozi (墨子)
470-390BCE

Recorded in writings by Aristotle or one of his disciples
384-322BCE
Camera obscura: the pre-camera

Freestanding camera obscura at UNC Chapel Hill

Photo by Seth Ilys
Camera Obscura used for Tracing

Lens Based Camera Obscura, 1568
First Photograph

Oldest surviving photograph
- Took 8 hours on pewter plate

Joseph Niepce, 1826

Photograph of the first photograph

Stored at UT Austin

Bitumen of Judea: Naturally occurring asphalt that is photo-sensitive
From Joseph Niepce to Louis Daguerre

• Act 1: Joseph Niepce tells his nice idea to pal Louis Daguerre

• Act 2: Good friend Louis Daguerre improves idea and names it Daguerrotypes

• Act 3: Louis Daguerre makes history (and money).
1846 Daguerrotype of a young Abraham Lincoln

Daguerreotype camera built by La Maison Susse Frères in 1839, with a lens by Charles Chevalier
Hercules Florence’s *Photographie*

- Brazilian painter and inventor
- Before Daguerre but after Niepce
- Included the idea of negatives
George Eastman 1885 (Rochester, NY)

• Founder of pioneering Eastman Kodak Company (Kodak)

• Popularization of film photography (nitrate film)

Should look familiar if you were born in the 80’s or earlier or if you are a true modern-day hipster!
• Photo negatives
So, who invented cameras?

Maybe the wrong question to ask?
Questions?