CS231n: Convolutional Neural Network for Visual Recognition

Justin Johnson, Serena Yeung, Fei-Fei Li

Lecture 1: Introduction
Welcome to CS231n
Computer Vision

Biology

Neuroscience

Psychology

Cognitive sciences

graphics, algorithms, theory, ...

systems, architecture, ...

Computer Science

Mathematics

Information retrieval

Machine learning

Robotics

Image processing

Speech, NLP

Physics

optics

Engineering
Related Courses @ Stanford

• **CS131: Computer Vision: Foundations and Applications**
  - Fall 2018, Juan Carlos Niebles and Ranjay Krishna
  - Undergraduate introductory class

• **CS231a: Computer Vision, from 3D Reconstruction to Recognition**
  - Professor Silvio Savarese
  - Core computer vision class for seniors, masters, and PhDs
  - Image processing, cameras, 3D reconstruction, segmentation, object recognition, scene understanding; not just deep learning

• **CS 224n: Natural Language Processing with Deep Learning**
  - Winter 2019, Chris Manning

• **CS 230: Deep Learning**
  - Spring 2019, Prof. Andrew Ng and Kian Katanforoosh

• **CS231n: Convolutional Neural Networks for Visual Recognition**
  - This course, Justin Johnson & Serena Yeung & Fei-Fei Li
  - Focusing on applications of deep learning to computer vision
Today’s agenda

• A brief history of computer vision

• CS231n overview
Evolution’s Big Bang

543 million years, B.C.
Camera Obscura

Gemma Frisius, 1545

Leonardo da Vinci, 16th Century AD

Encyclopedie, 18th Century

This work is in the public domain
Simple cells: Response to light orientation

Complex cells: Response to light orientation and movement

Hypercomplex cells: response to movement with an end point

Hubel & Wiesel, 1959

Electrical signal from brain

- No response
- Response (end point)

Stimulus

Response

Cat image by CNX OpenStax is licensed under CC BY 4.0; changes made
Block world

Larry Roberts, 1963

(a) Original picture  (b) Differentiated picture  (c) Feature points selected
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".
<table>
<thead>
<tr>
<th>Stage</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input Image</td>
<td>Perceived intensities</td>
</tr>
<tr>
<td>Primal Sketch</td>
<td>Zero crossings, blobs, edges, bars, ends, virtual lines, groups, curves, boundaries</td>
</tr>
<tr>
<td>2 ½-D Sketch</td>
<td>Local surface orientation and discontinuities in depth and in surface orientation</td>
</tr>
<tr>
<td>3-D Model Representation</td>
<td>3-D models hierarchically organized in terms of surface and volumetric primitives</td>
</tr>
</tbody>
</table>

Stages of Visual Representation, David Marr, 1970s
• Generalized Cylinder
  Brooks & Binford, 1979

• Pictorial Structure
  Fischler and Elschlager, 1973
Normalized Cut (Shi & Malik, 1997)
“SIFT” & Object Recognition, David Lowe, 1999
Spatial Pyramid Matching, Lazebnik, Schmid & Ponce, 2006
Histogram of Gradients (HoG)
Dalal & Triggs, 2005

Deformable Part Model
Felzenswalb, McAllester, Ramanan, 2009
PASCAL Visual Object Challenge
(20 object categories)
[Everingham et al. 2006-2012]
22K categories and 15M images

- Animals
  - Bird
  - Fish
  - Mammal
  - Invertebrate
- Plants
  - Tree
  - Flower
  - Food
  - Materials
- Structures
  - Artifact
  - Tools
  - Appliances
  - Structures
- Person
  - Scenes
    - Indoor
    - Geological Formations
  - Sport Activities

Deng, Dong, Socher, Li, Li, & Fei-Fei, 2009
The Image Classification Challenge:

1,000 object classes
1,431,167 images
The Image Classification Challenge:
1,000 object classes
1,431,167 images

Lin et al
Sanchez & Perronnin
Krizhevsky et al (AlexNet)
Zeiler & Fergus
Simonyan & Zisserman (VGG)
Szegedy et al (GoogLeNet)
He et al (ResNet)
Shao et al
Hu et al (SENet)
Russakovsky et al

Russakovsky et al. IJCV 2015
Today’s agenda

• A brief history of computer vision

• CS231n overview
CS231n focuses on one of the most fundamental problems of visual recognition – *image classification*
There are many visual recognition problems that are related to image classification, such as object detection, image captioning.
- Object detection
- Action classification
- Image captioning
- …
Convolutional Neural Networks (CNN) have become an important tool for object recognition.
**IMAGENET**

**Large Scale Visual Recognition Challenge**

---

**Year 2010**

NEC-UIUC

- Dense descriptor grid: HOG, LBP
- Coding: local coordinate, super-vector
- Pooling, SPM
- Linear SVM

[Lin CVPR 2011]

---

**Year 2012**

SuperVision

---

**Year 2014**

GoogLeNet

- Pooling
- Convolution
- Softmax
- Other

[VGG](Szegedy arxiv 2014)

[He ICCV 2015]

---

**Year 2015**

MSRA

---

Figure copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.
Convolutional Neural Networks (CNN) were not invented overnight
1998
LeCun et al.

![Diagram of 1998 architecture]

- **# of transistors**: $10^6$
- **# of pixels used in training**: $10^7$

2012
Krizhevsky et al.

![Diagram of 2012 architecture]

- **# of transistors**: $10^9$
- **# of pixels used in training**: $10^{14}$

*Figure copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.*
Ingredients for Deep Learning

Algorithms

Data

Computation
GigaFLOPs per Dollar

- **CPU**
- **GPU**

Deep Learning Explosion

- GeForce GTX 580 (AlexNet)
- GeForce 8800 GTX
- GTX 1080 Ti

Time:
- 1/2004
- 10/2006
- 7/2009
- 4/2012
- 12/2014
- 9/2017
GigaFLOPs per Dollar

- CPU
- GPU
- TPU

Deep Learning Explosion

- TITAN V (Tensor Cores)
- GTX 1080 Ti
- GeForce GTX 580 (AlexNet)
- GeForce 8800 GTX

Time

The quest for visual intelligence goes far beyond object recognition...

Figures copyright IEEE, 2015. Reproduced for educational purposes
Some kind of game or fight. Two groups of two men? The man on the left is throwing something. Outdoors seemed like because I have an impression of grass and maybe lines on the grass? That would be why I think perhaps a game, rough game though, more like rugby than football because they pairs weren't in pads and helmets, though I did get the impression of similar clothing. maybe some trees? in the background. (Subject: SM)
Computer Vision Technology Can Better Our Lives
Who we are

Instructors

Fei-Fei Li
Justin Johnson
Serena Yeung

Teaching Assistants

Winnie Lin (Head TA)
Saahil Agrawal
Malavika Bindhi
Haoye Cai
Kaidi Cao
Apoorva Dornadula
Jim (Linxi) Fan
Pedro Pablo Garzon
Ayush Gupta
Andrew Han
Tien-Ning Hsu
Nishith Khandwala
Simon Le Cleach
Bingbin Liu
David Morales
Boxiao Pan
Ashwini Pokle
Praty Sharma
William Shen
Owen Wang
Danfei Xu
How to Contact Us

• Course Website: http://cs231n.stanford.edu/
  – Syllabus, lecture slides, links to assignment downloads, etc

• Piazza: http://piazza.com/stanford/spring2019/cs231n
  – Use this for most communication with course staff
  – Ask questions about homework, grading, logistics, etc
  – Use private questions if you want to post code

• Canvas
  – For watching lecture videos
Optional Textbook

• *Deep Learning* by Goodfellow, Bengio, and Courville
• *Free online*
Our philosophy

• Thorough and Detailed.
  – Understand how to write from scratch, debug and train convolutional neural networks.

• Practical.
  – Focus on practical techniques for training these networks at scale, and on GPUs (e.g. will touch on distributed optimization, differences between CPU vs. GPU, etc.) Also look at state of the art software tools

• State of the art.
  – Most materials are new from research world in the past 1-3 years. Very exciting stuff!
Our philosophy (cont’d)

• Fun.
  – Some fun topics such as Image Captioning (using RNN)
  – Also DeepDream, NeuralStyle, etc.
Pre-requisite

• Proficiency in Python, some high-level familiarity with C/C++
  – All class assignments will be in Python (and use numpy), but some of the deep learning libraries we may look at later in the class are written in C++.
  – A Python tutorial available on course website

• College Calculus, Linear Algebra

• Equivalent knowledge of CS229 (Machine Learning)
  – We will be formulating cost functions, taking derivatives and performing optimization with gradient descent.
Grading Policy

• 3 Problem Sets: $15\% \times 3 = 45\%$

• Midterm Exam: 20%

• Course Project: 35%
  – Project Proposal: 1%
  – Milestone: 2%
  – Poster: 2%
  – Project Report: 30%

• Late policy
  – 4 free late days – use up to 2 late days per assignment
  – Afterwards, 25% off per day late
  – No late days for project report
Collaboration Policy

- We follow the [Stanford Honor Code](#) and the [CS Department Honor Code](#) – read them!
- **Rule 1**: Don’t look at solutions or code that are not your own; everything you submit should be your own work
- **Rule 2**: Don’t share your solution code with others; however, discussing ideas or general strategies is fine and encouraged
- **Rule 3**: Indicate in your submissions anyone you worked with
- Turning in something late / incomplete is better than violating the honor code
Next Time: Image Classification

K-Nearest Neighbor

Linear Classifier
References

- Lowe, David G. "Distinctive image features from scale-invariant keypoints." International Journal of Computer Vision 60.2 (2004): 91-110. [PDF]