Convolutional Neural Networks for Visual Recognition

A fundamental and general problem in Computer Vision, that has roots in Cognitive Science

Biederman, Irving. "Recognition-by-components: a theory of human image understanding." *Psychological review* 94.2 (1987): 115.

There are many visual recognition problems that are related to image classification, such as object detection, image captioning, semantic segmentation, visual question answering, visual instruction navigation, scene graph generation

Object detection
car

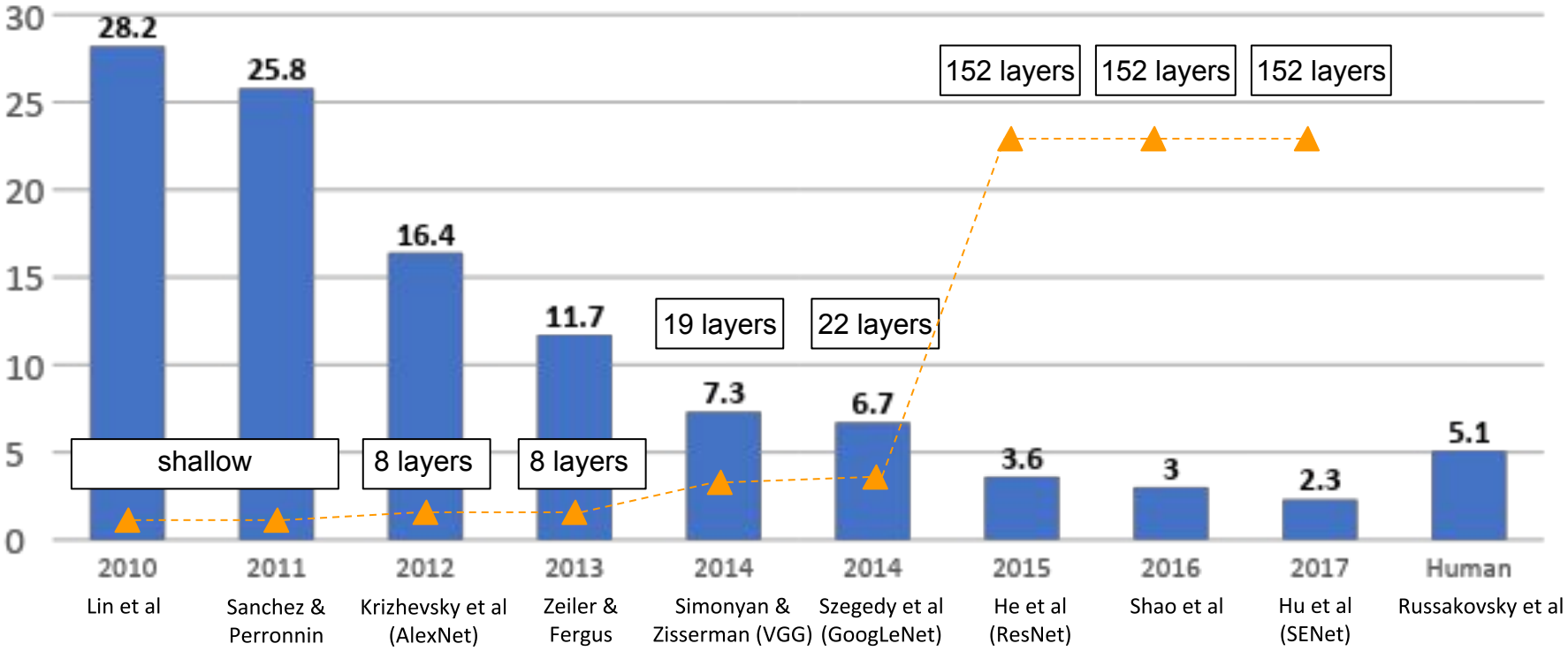
Action recognition
bicycling

Visual relationship detection
<person - holding - hammer>

Convolutional Neural Networks for Visual Recognition

Hierarchical computing systems with many “layers”, that are very loosely inspired by the brain

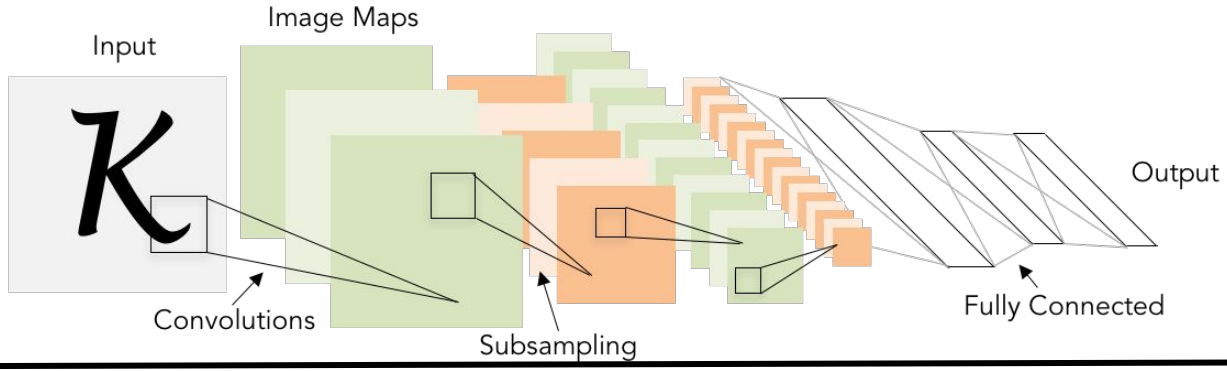
ImageNet Large Scale Visual Recognition Challenge (ILSVRC) winners



Convolutional Neural Networks for Visual Recognition

A class of Neural Networks that have become an important tool for visual recognition

1998 LeCun et al.



# of transistors

10^6

# of pixels used to train:

10^7 NIST

2012 Krizhevsky et al.

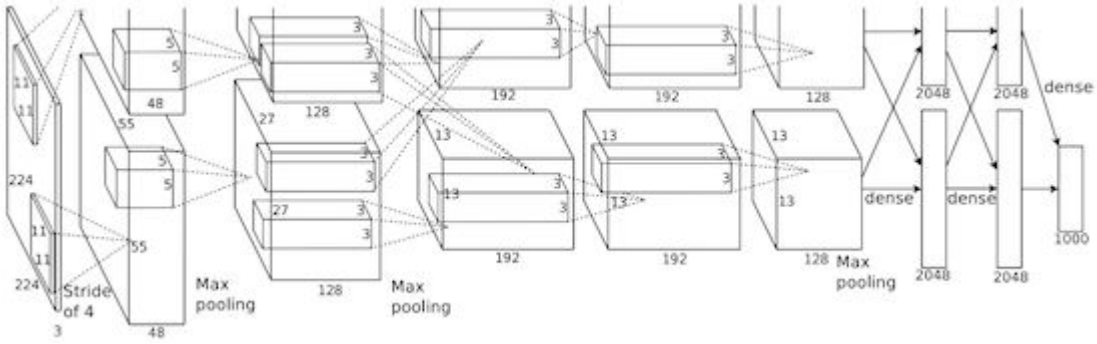


Figure copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.

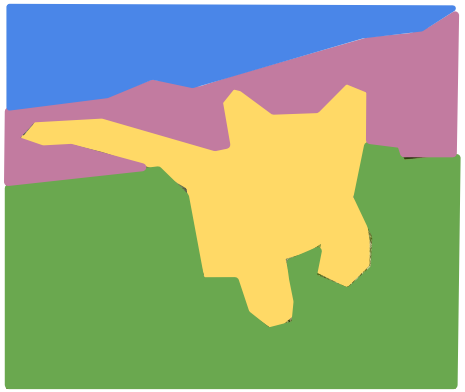
# of transistors

10^9

# of pixels used to train:

10^{14} IMAGENET

Beyond recognition: Segmentation, 2D/3D Generation



Progressive GAN, Karras 2018.

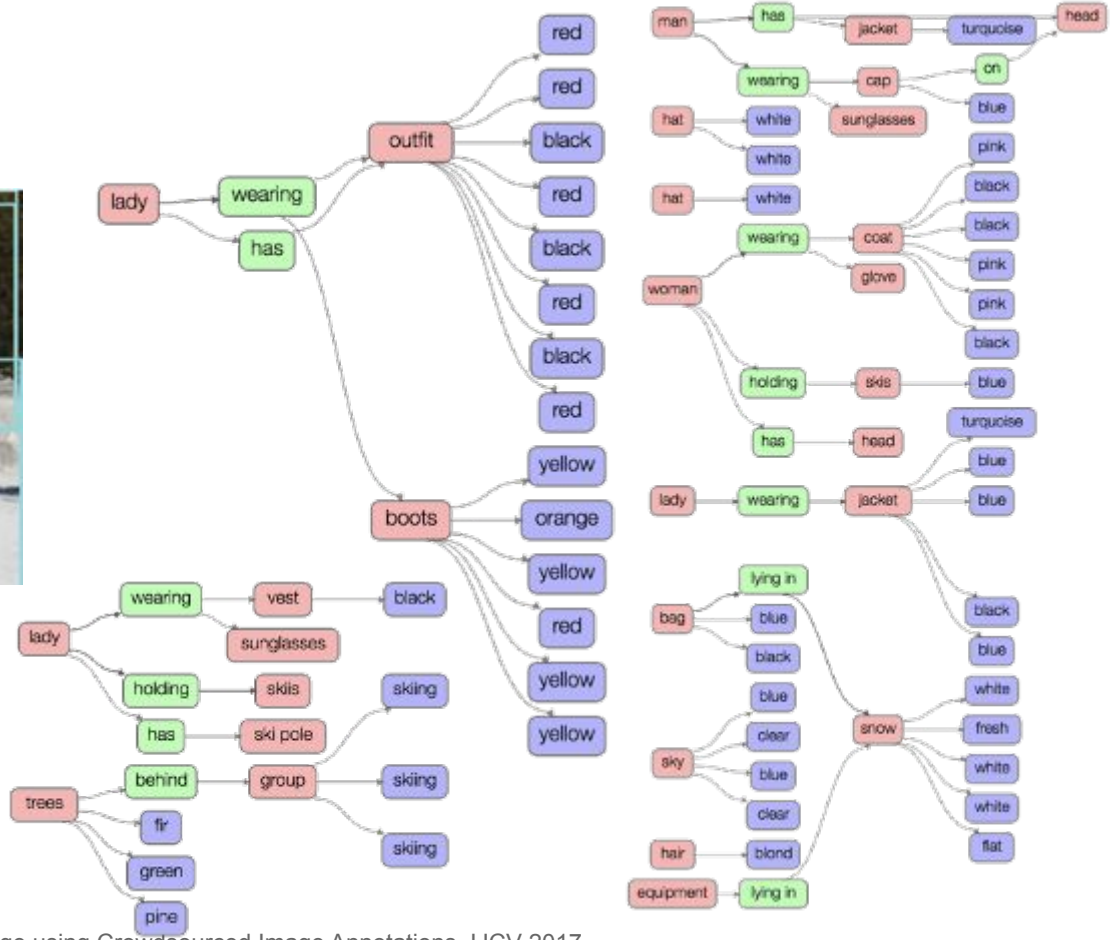
Wang et al, “Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images”, ECCV 2018

Scene Graphs

This image is CC0 public domain

[Three Ways Computer Vision Is Transforming Marketing]

- Forbes Technology Council

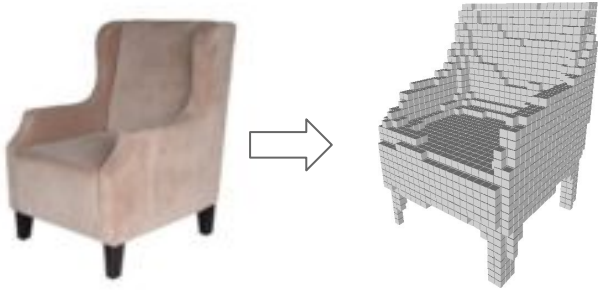


Krishna et al., Visual Genome: Connecting Vision and Language using Crowdsourced Image Annotations, IJCV 2017

Spatio-temporal scene graphs

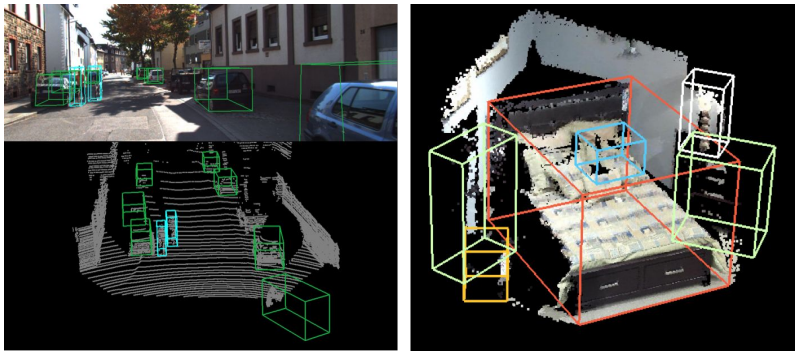
Ji, Krishna et al., Action Genome: Actions as Composition of Spatio-temporal Scene Graphs, CVPR 2020

3D Vision & Robotic Vision

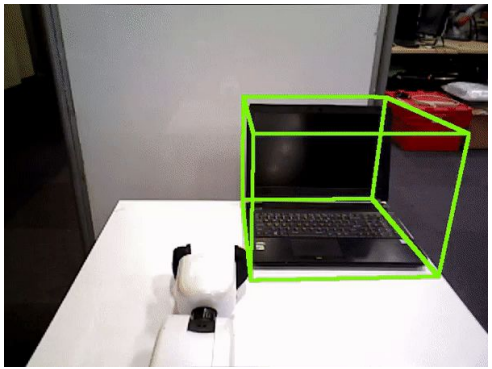


Choy et al., 3D-R2N2: Recurrent Reconstruction Neural Network (2016)

Mandlekar and Xu et al., Learning to Generalize Across Long-Horizon Tasks from Human Demonstrations (2020)



Xu et al., PointFusion: Deep Sensor Fusion for 3D Bounding Box Estimation (2018)



Wang et al., 6-PACK: Category-level 6D Pose Tracker with Anchor-Based Keypoints (2020)

Human vision

PT = 500ms

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.

Fei-Fei, Iyer, Koch, Perona, *JoV,* 2007

This image is copyright-free United States government work
Example credit: Andrej Karpathy

Optional textbook resources

- *Deep Learning*
  - by Goodfellow, Bengio, and Courville
  - Here is a free version
- Mathematics of deep learning
  - Chapters 5, 6 7 are useful to understand vector calculus and continuous optimization
  - Free online version
- Dive into deep learning
  - An interactive deep learning book with code, math, and discussions, based on the NumPy interface.
  - Free online version

Assignments

Two alternative ways of completing assignments

- On local machines
- On Google Colab

See (https://cs231n.github.io/assignments2020/assignment1/) for more details.

Grading

All assignments, coding and written portions, will be submitted via Gradescope.

New this year: an **auto-grading system**

- a consistent grading scheme,
- Public tests:
  - Students see results of public tests immediately
- Private tests
  - More thorough and used to thoroughly test your implementation

Grading

3 Problem Sets: 15% x 3 = 45%

Take home 24hr Midterm Exam: 20%

Course Project: 35%

- Project Proposal: 1%
- Milestone: 2%
- Video presentation: 7%
  - Uploaded to YouTube
- Project Report: 25%

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

Pre-requisite

Proficiency in Python

- All class assignments will be in Python (and use numpy)
- Later in the class, you will be using Pytorch and TensorFlow
- 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.

Google Cloud

We have Google Cloud credits available for projects
- Not for HWs (only for final projects)

We will be distributing coupons to all enrolled students who need it

See our tutorial here for walking through Google Cloud setup:
https://github.com/cs231n/gcloud

Why should you take this class?

Become a vision researcher (an incomplete list of conferences)

- CVPR 2019 conference
- ICCV 2019 conference

Become a vision engineer in industry (an incomplete list of industry teams)

- Perception team at Google AI
- Vision at Google Cloud
- Vision at Facebook AI

General interest

Syllabus

Neural Network Fundamentals → Convolutional Neural Networks → Computer Vision Applications

Data-driven learning
Linear classification & kNN
Loss functions
Optimization
Backpropagation
Multi-layer perceptrons
Neural Networks

Convolutions
Pytorch 1.4 / Tensorflow 2.0
Activation functions
Batch normalization
Transfer learning
Data augmentation
Momentum / RMSProp / Adam
Architecture design

RNNs / LSTMs
Image captioning
Interpreting neural networks
Style transfer
Adversarial examples
Fairness & ethics
Human-centered AI
3D vision
Deep reinforcement learning
Scene graphs
Self-supervised learning

GigaFLOPs per Dollar