Convolutional Neural Networks for Visual Recognition

Lecture 1 - Overview

Today's agenda

- A brief history of computer vision
- CS231n overview

Today's agenda

- A brief history of computer vision
- CS231n overview

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.

Image Classification: A core task in Computer Vision

→ cat


UNCLASSIFIED
Capture Left Iris


0.892 Burger
0.529 Cup
0.322 Plate


JORDI
ARENAS
I CLAVELL
DESCOBREIX MÉS
SOBRE L'AUTOR

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

This image is licensed under CC BY-NC-SA 2.0; changes made

Action recognition
bicycling

This image is licensed under CC BY-SA 3.0; changes made

Scene graph prediction
<person - holding - hammer>

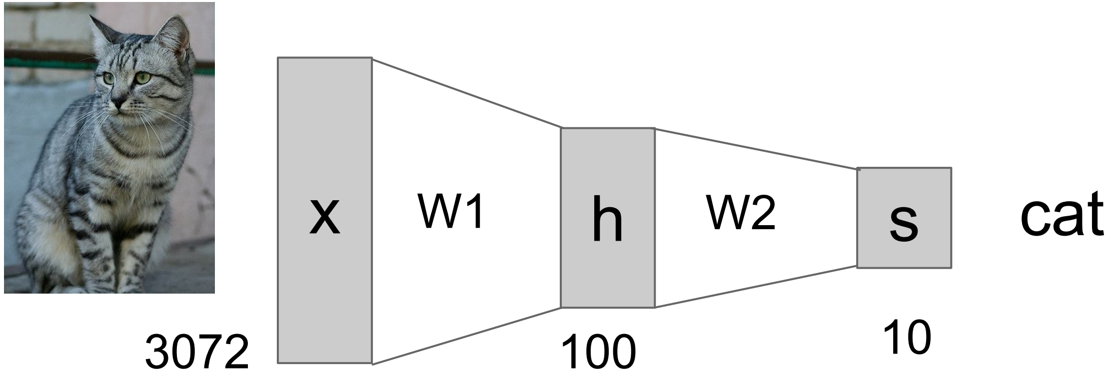
Captioning:
a person holding a hammer

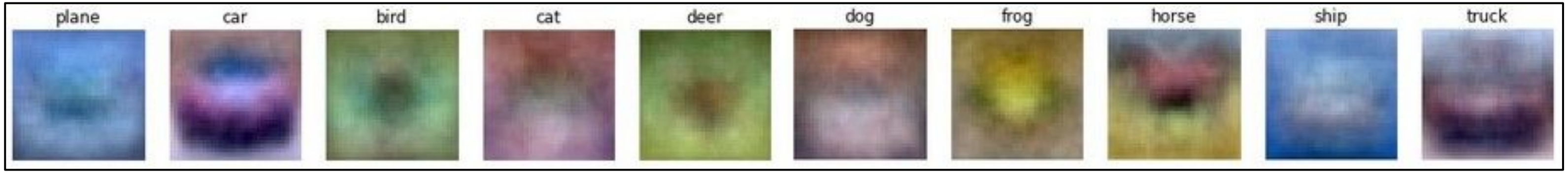
This image is licensed under CC BY-SA 3.0; changes made

Convolutional Neural Networks for Visual Recognition

Hierarchical computing systems with many "layers", that are very loosely inspired by Neuroscience

Last time: Neural Networks





Convolutional Neural Networks for Visual Recognition

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

Core ideas go back many decades!

The **Mark I Perceptron** machine was the first implementation of the perceptron algorithm.

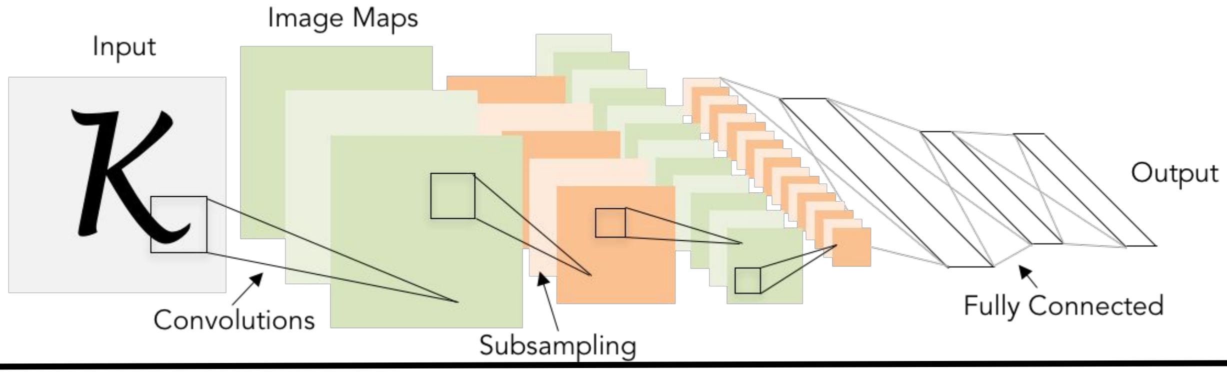
The machine was connected to a camera that used 20×20 cadmium sulfide photocells to produce a 400-pixel image.

recognized
letters of the alphabet

Frank Rosenblatt, ~1957: Perceptron

1998 LeCun et al.

提出CNN的概念，并应用于手写数字识别



# of transistors

10^6

# of pixels used to train:

10^7 NIST

2012 Krizhevsky et al.

AlexNet，大大加深CNN的深度

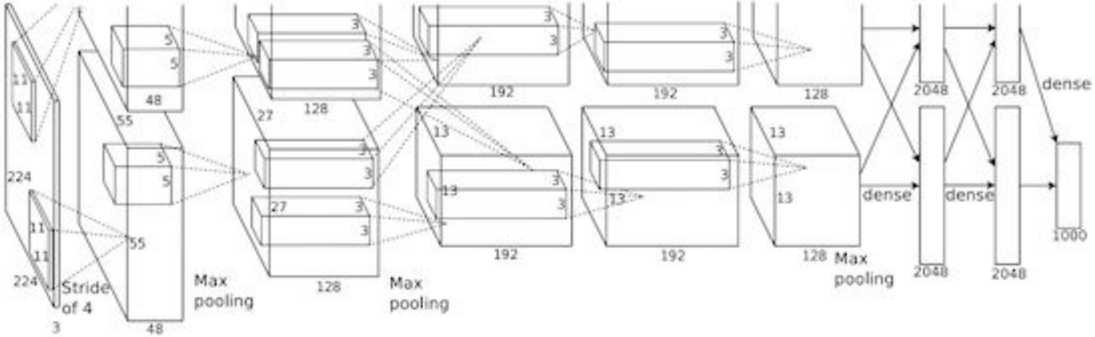


Figure copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.

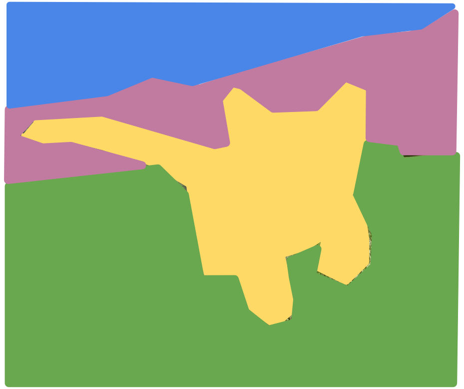
# of transistors

10^9

# of pixels used to train:

10^{14} IMAGENET

Beyond recognition: Segmentation, 2D/3D Generation

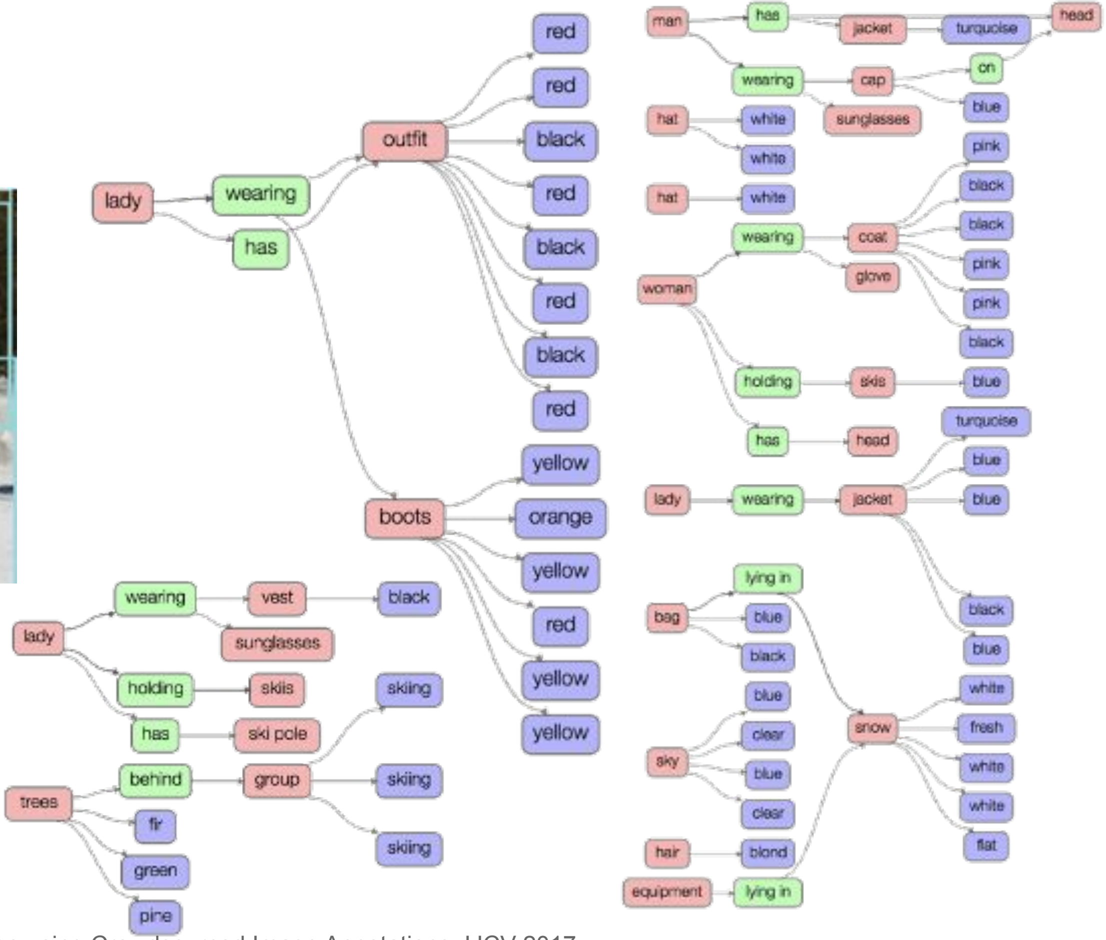


Progressive GAN, Karras 2018.

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

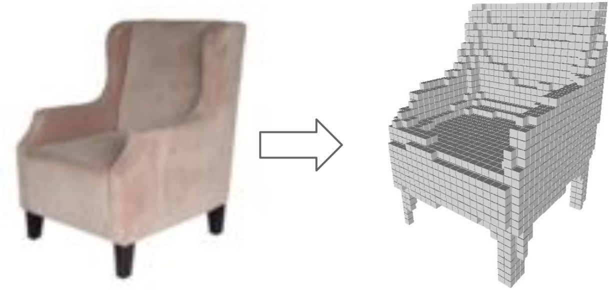
Krishna et al., Visual Genome: Connecting Vision and Language using Crowdsourced Image Annotations, IJCV 2017

Spatio-temporal scene graphs

将动作信息转化为时空场景图，通过对动作进行分解识别并学习到其时序上的改变来完成对象以及相应关系的识别从而构造场景图。

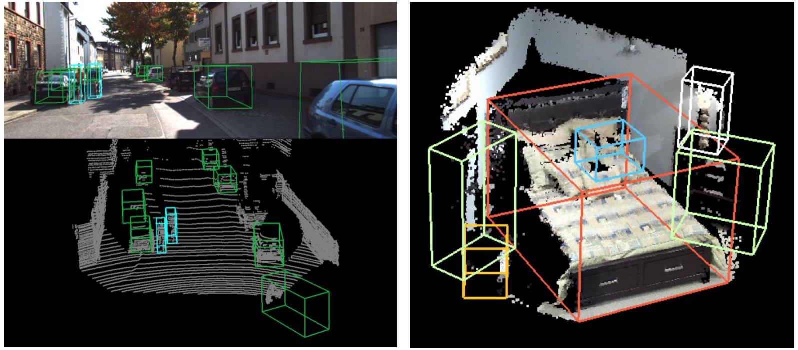
Ji, Krishna et al., Action Genome: Actions as Composition of Spatio-temporal Scene Graphs, CVPR 2020

3D Vision & Robotic Vision



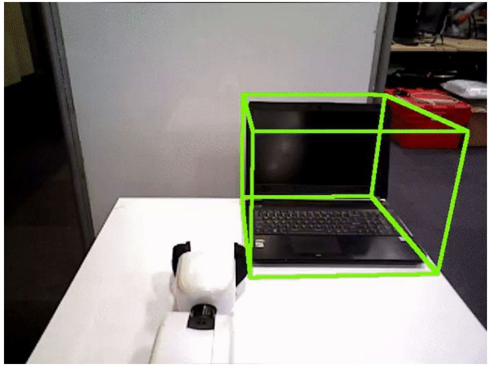
Choy et al., 3D-R2N2: Recurrent Reconstruction Neural Network (2016)

三维场景重构



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

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



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

Human vision

PT = 500ms

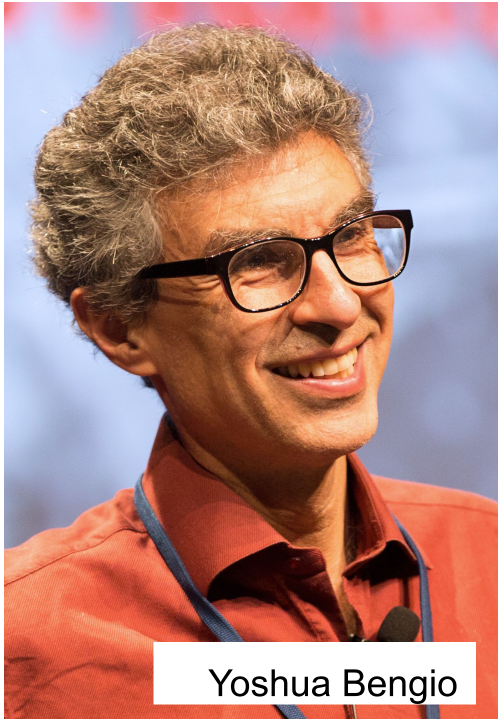
Some kind of game or fight. Two groups of two men? The man on the left is throwing something. Outdoors seemed like because i have an impression of grass and maybe lines on the grass? That would be why I think perhaps a game, rough game though, more like rugby than football because they pairs weren't 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

2018 Turing Award for deep learning

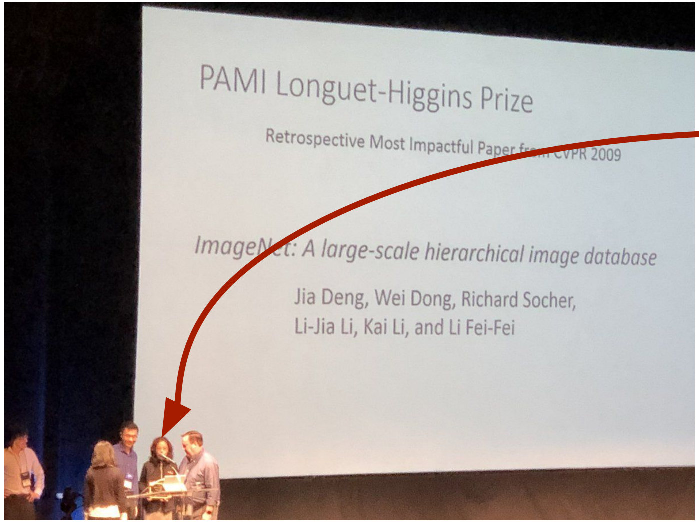
most prestigious technical award, is given for major contributions of lasting importance to computing.



IEEE PAMI Longuet-Higgins Prize

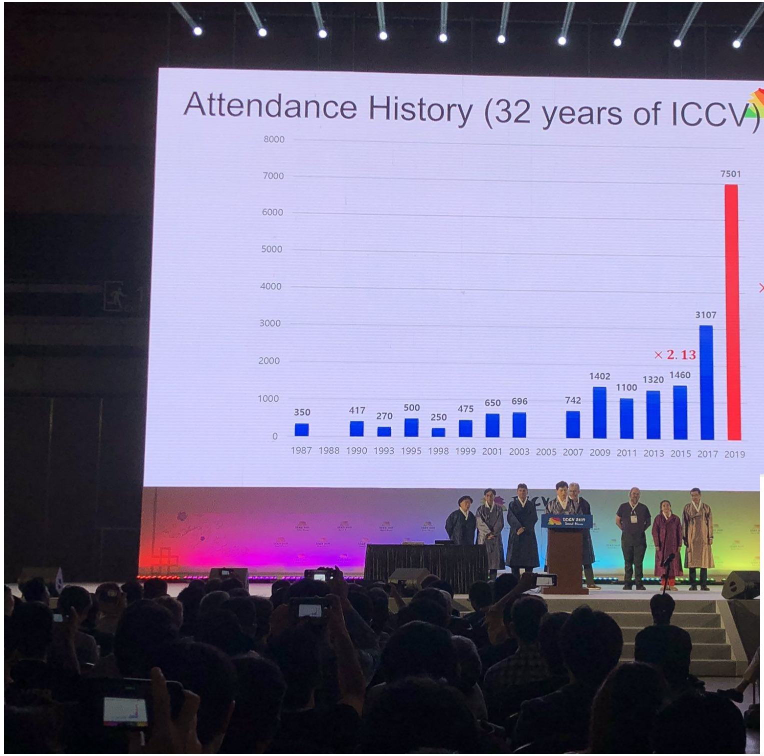
Award recognizes ONE Computer Vision paper from **ten years ago** with **significant impact on computer vision** research.

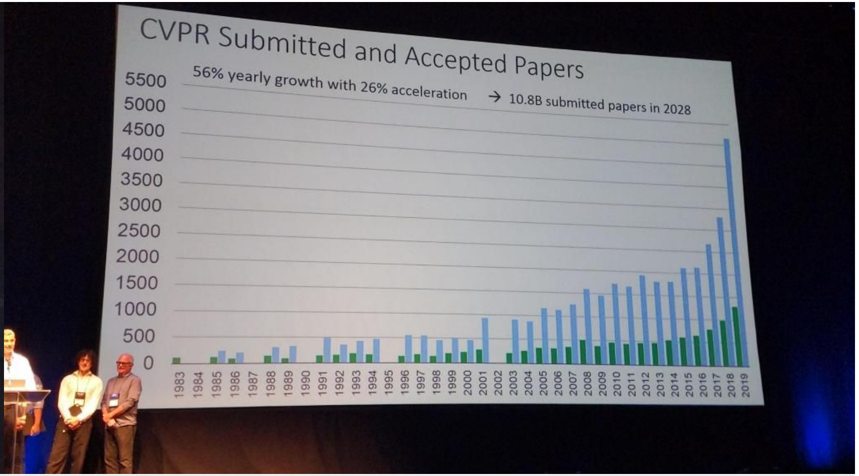
In 2019, it was awarded to the 2009 original ImageNet paper



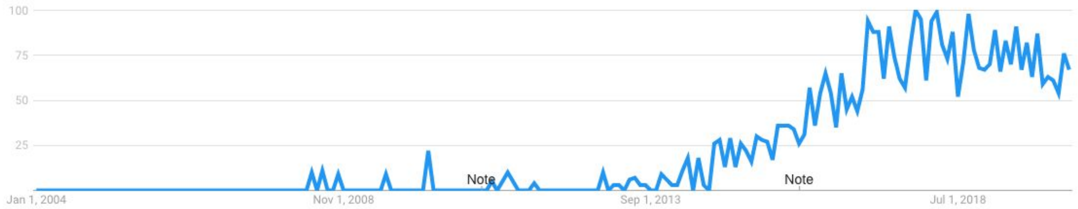
That's Fei-Fei

Why is this such a large class?





Google search trends for convolutional neural networks



Logistics

Instructors


Fei-Fei Li


Ranjay Krishna


Danfei Xu

Teaching Assistants


Kevin Zakka (Head TA)


Haofeng Chen


Rachel Gardner


Samuel Kwong


Yichen Li


Sean Liu


Mandy Lu


Nishant Rai


Geet Sethi


Lin Shao


Guanzhi Wang


Chris Waites


Jiequan Zhang


Russel Xie

Course Coordinator


Yosefa Gilon

Lectures

Live Zoom Webinar

- Links will be shared via email and canvas: cs231n.stanford.edu
    - Due to security reasons, please do not share zoom links publicly
- **Tuesdays** and **Thursdays** between **1pm to 2:20pm**
    - To watch the lectures, you must login to Zoom using your SUNETID@stanford.edu accounts.
- Q/A functionality - a dedicated TA will answer questions live

- All lectures will be recorded and uploaded to Canvas
- 2 new lectures were added last year.
- 2 more new lectures will be added this year.

Friday Discussion Sections

(Most) Fridays 11:30am - 12:30pm

Hands-on tutorials, with more practical detail than main lecture

We may not have discussion sections every Friday, check our syllabus!

Zoom meetings (not webinars) - there will be more student-student interactions

This Friday: Python / numpy / Google Cloud (Presenter: Rachel Gardner)

Piazza

For questions about midterm, projects, logistics, etc, use Piazza!

SCPD students: Use your @stanford.edu address to register for Piazza; contact scpd-customerservice@stanford.edu for help.

Office Hours

Will occur through Nooks

- Join Nooks and add your name to a queue for a particular office hours
- TAs will take you into a private room for 1-1 conversations when it’s your turn
- Office hours will be listed here by Friday!

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

Grading

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

New since last year: an **auto-grading system**

- A consistent grading scheme,
- Public tests:
  - Students see results of public tests immediately
- Private tests
  - Generalizations of the public tests to thoroughly test your implementation

Grading

3 Problem Sets: 10% + 20% + 20% = 50%

Take home 24hr Midterm Exam: 15%

Course Project: 35%

- Project Proposal: 1%
- Milestone: 2%
- Video presentation: 10%
- Project Report: 22%

Participation Extra Credit: up to 3%

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

Overview on communication

Course Website: http://cs231n.stanford.edu/

- Syllabus, lecture slides, links to assignment downloads, etc

Piazza:

- Use this for most communication with course staff
- Ask questions about homework, grading, logistics, etc
- Use private questions if you want to post code

Gradescope:

- For turning in homework and receiving grades

Canvas:

- For watching lecture videos

Zoom:

- For watching live lectures and discussion sections and for participating!

Assignments

All assignments will be completed using Google Colab

Assignment 1: Will be out Friday, due 4/16 by 11:59pm

- K-Nearest Neighbor
- Linear classifiers: SVM, Softmax
- Two-layer neural network
- Image features

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

No longer need CS229 (Machine Learning)

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

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

Learning objectives

Formalize computer vision applications into tasks

- Formalize inputs and outputs for vision-related problems
- Understand what data and computational requirements you need to train a model

Develop and train vision models

- Learn to code, debug, and train convolutional neural networks.
- Learn how to use software frameworks like TensorFlow and PyTorch

Gain an understanding of where the field is and where it is headed

- What new research has come out in the last 0-5 years
- What are open research challenges?
- What ethical and societal considerations should we consider before deployment?

What you should expect from us

Fun.

- We will discuss fun applications like image captioning, visual question answering, style transfer

What we expect from you

Patience.

- This is new for us as much as it is new for you
- Things will break; we will experience technical difficulties
- Bear with us and trust us to listen to you

Contribute

- Build a community on slack
- Help one another - discuss topics you enjoy
- Give us (annonymous) feedback

Why should you take this class?

Become a vision researcher (an incomplete list of conferences)

- Get involved with vision research at Stanford: apply using this form.
- CVPR 2020 conference
- ICCV 2020 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

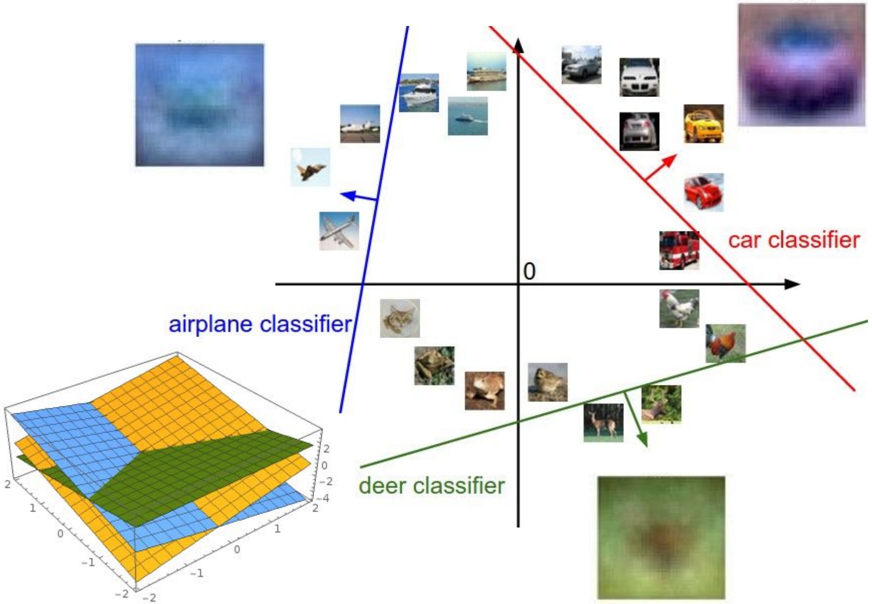
Syllabus

Neural Network Fundamentals	Convolutional Neural Networks	Computer Vision Applications
Data-driven learning	Convolutions	RNNs / LSTMs / Transformers
Linear classification & kNN	Pytorch 1.4 / Tensorflow 2.0	Image captioning
Loss functions	Activation functions	Interpreting neural networks
Optimization	Batch normalization	Style transfer
Backpropagation	Transfer learning	Adversarial examples
Multi-layer perceptrons	Data augmentation	Fairness & ethics
Neural Networks	Momentum / RMSProp / Adam	Human-centered AI
	Architecture design	3D vision
		Deep reinforcement learning
		Scene graphs
		Self-supervised learning

Next time: Image classification

k- nearest neighbor

Linear classification



Plot created using [Wolfram Cloud]

References

•Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 1. IEEE, 2005. [PDF]
•Felzenszwalb, Pedro, David McAllester, and Deva Ramanan. "A discriminatively trained, multiscale, deformable part model." Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008 [PDF]
•Everingham, Mark, et al. "The pascal visual object classes (VOC) challenge." International Journal of Computer Vision 88.2 (2010): 303-338. [PDF]
•Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009. [PDF]
•Russakovsky, Olga, et al. "Imagenet Large Scale Visual Recognition Challenge." arXiv:1409.0575. [PDF]
•Lin, Yuanqing, et al. "Large-scale image classification: fast feature extraction and SVM training." Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011. [PDF]
•Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012. [PDF]
•Szegedy, Christian, et al. "Going deeper with convolutions." arXiv preprint arXiv:1409.4842 (2014). [PDF]
•Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014). [PDF]
•He, Kaiming, et al. "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition." arXiv preprint arXiv:1406.4729 (2014). [PDF]
•LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324. [PDF]
•Fei-Fei, Li, et al. "What do we perceive in a glance of a real-world scene?." Journal of vision 7.1 (2007): 10. [PDF]

References

•Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 1. IEEE, 2005. [PDF]

•Felzenszwalb, Pedro, David McAllester, and Deva Ramanan. "A discriminatively trained, multiscale, deformable part model." Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008 [PDF]

•Everingham, Mark, et al. "The pascal visual object classes (VOC) challenge." International Journal of Computer Vision 88.2 (2010): 303-338. [PDF]

•Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009. [PDF]

•Russakovsky, Olga, et al. "Imagenet Large Scale Visual Recognition Challenge." arXiv:1409.0575. [PDF]

•Lin, Yuanqing, et al. "Large-scale image classification: fast feature extraction and SVM training." Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011. [PDF]

•Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012. [PDF]

•Szegedy, Christian, et al. "Going deeper with convolutions." arXiv preprint arXiv:1409.4842 (2014). [PDF]

•Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014). [PDF]

•He, Kaiming, et al. "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition." arXiv preprint arXiv:1406.4729 (2014). [PDF]

•LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324. [PDF]

•Fei-Fei, Li, et al. "What do we perceive in a glance of a real-world scene?." Journal of vision 7.1 (2007): 10. [PDF]