CS231n: Deep Learning for Computer Vision

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

Image Classification: A core task in Computer Vision

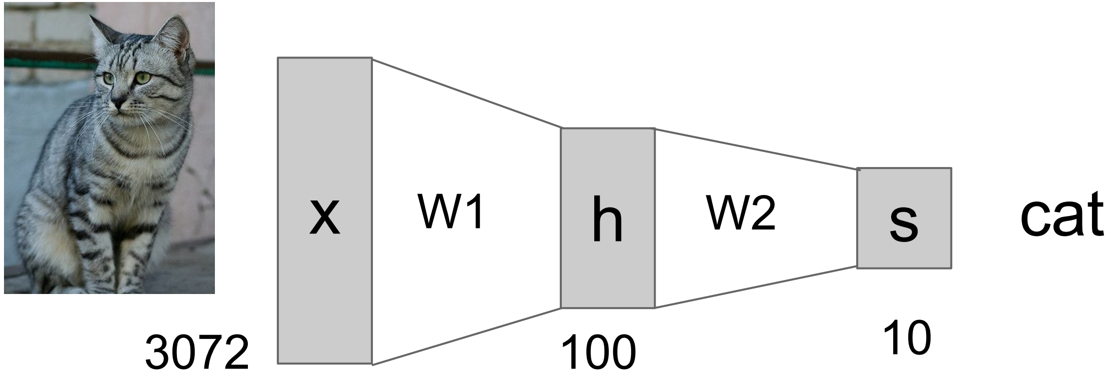
→ cat

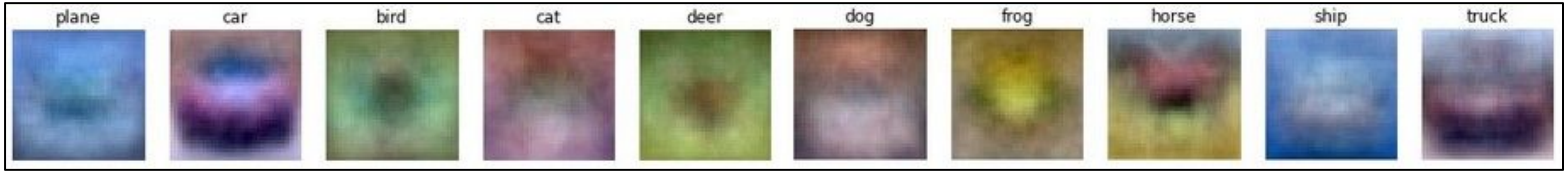
There are many visual recognition problems that are related to image classification, such as object detection, image captioning, image segmentation, visual question answering, visual instruction navigation, video understanding, etc.

Deep Learning for Computer Vision

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

Neural Networks

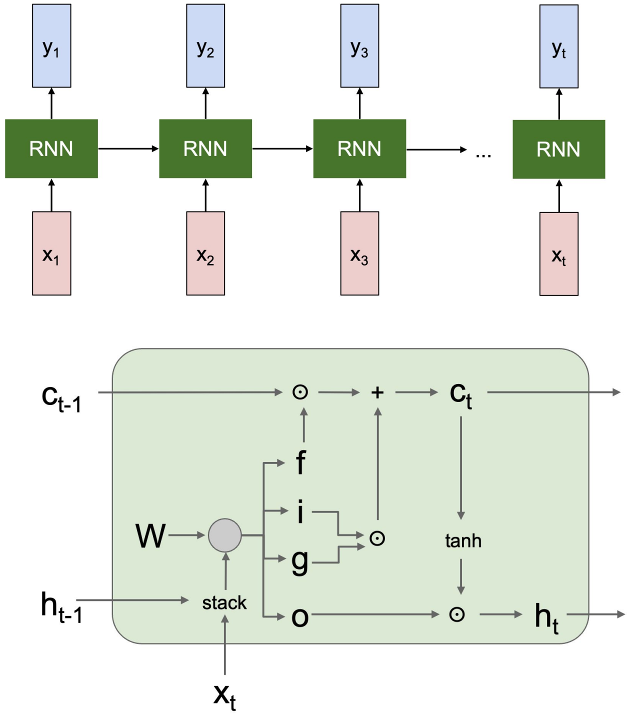




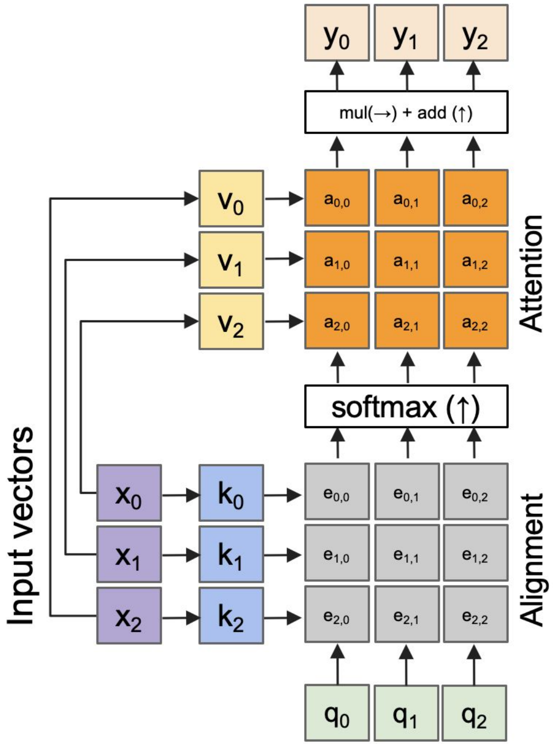
Convolutional Neural Networks for Visual Recognition

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

Beyond Convolutional Neural Networks

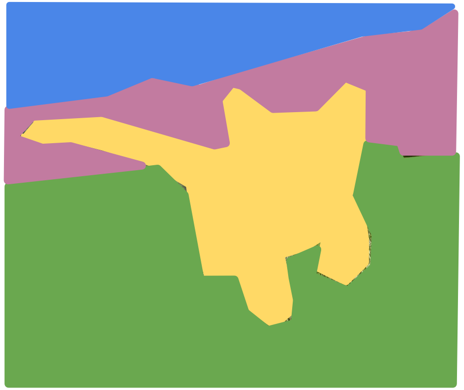
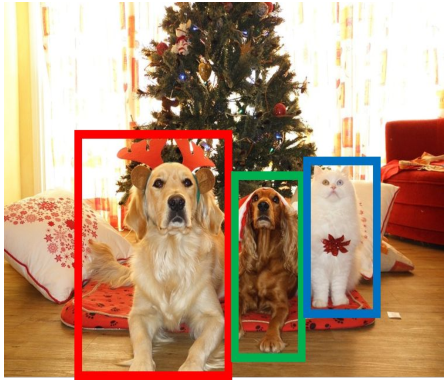
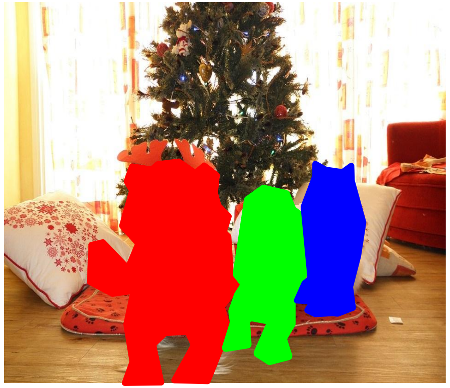


Recurrent neural network



Attention mechanism / Transformers

Beyond Image Classification

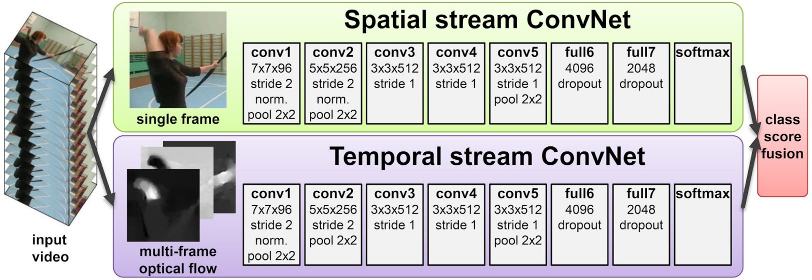
Classification	Semantic Segmentation	Object Detection	Instance Segmentation
			
CAT	**GRASS, CAT, TREE, SKY**	**DOG, DOG, CAT**	**DOG, DOG, CAT**

No spatial extent

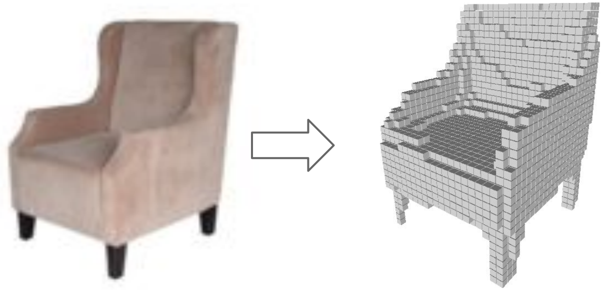
No objects, just pixels

Multiple Object

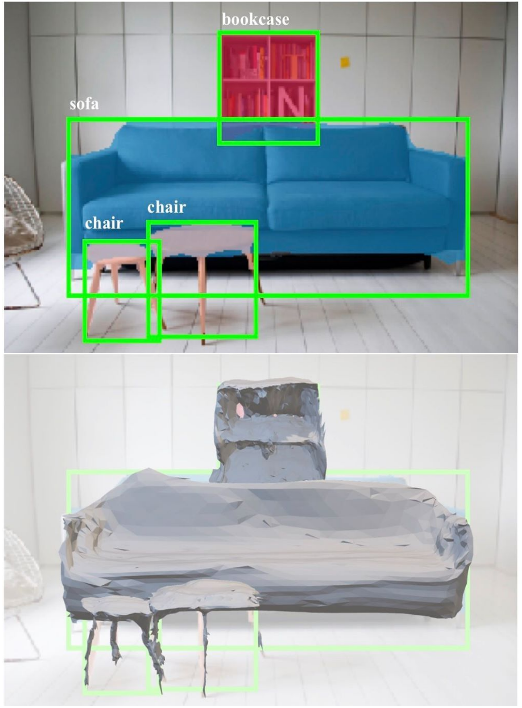
Beyond 2D Images



Simonyan and Zisserman, “Two-stream convolutional networks for action recognition in videos”, NeurIPS 2014

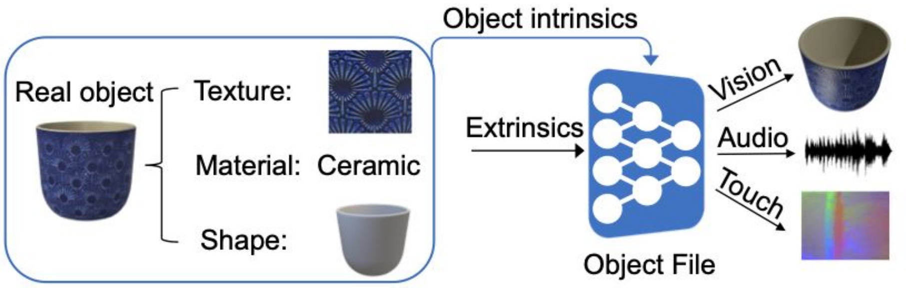


Choy et al., 3D-R2N2: Recurrent Reconstruction Neural Network (2016)



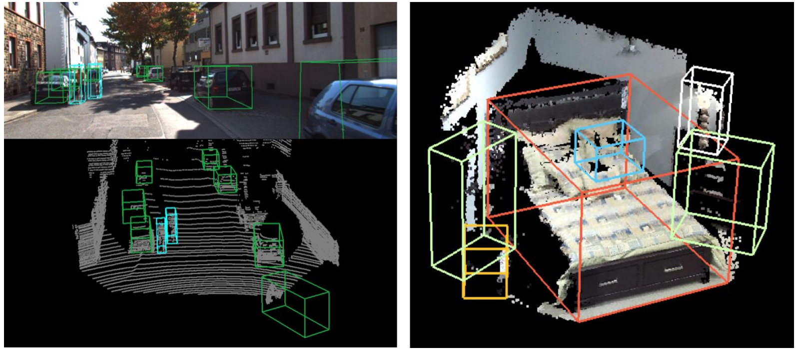
Gkioxari et al., “Mesh R-CNN”, ICCV 2019

Beyond Vision

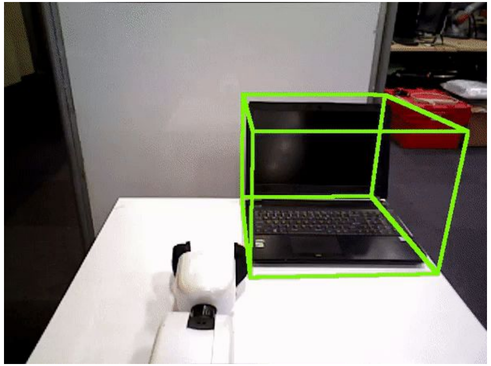


Gao et al., ObjectFolder 2.0: A Multisensory Object Dataset for Sim2Real Transfer (2022)

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



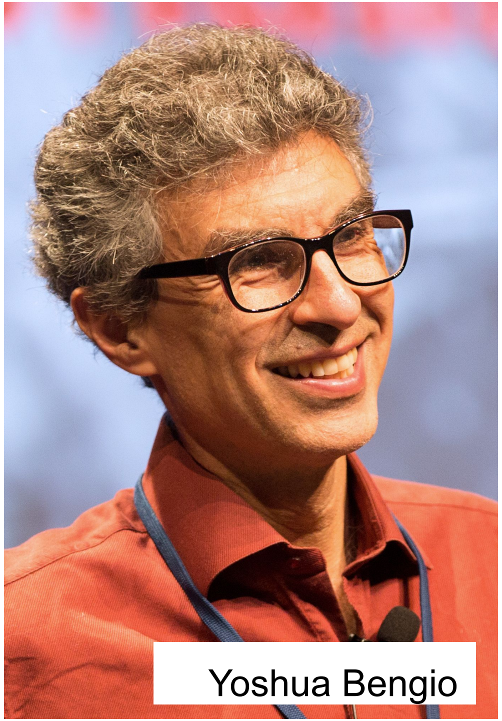
Xu et al., PointFusion: Deep Sensor Fusion for 3D Bounding Box Estimation (2018)



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

2018 Turing Award for deep learning

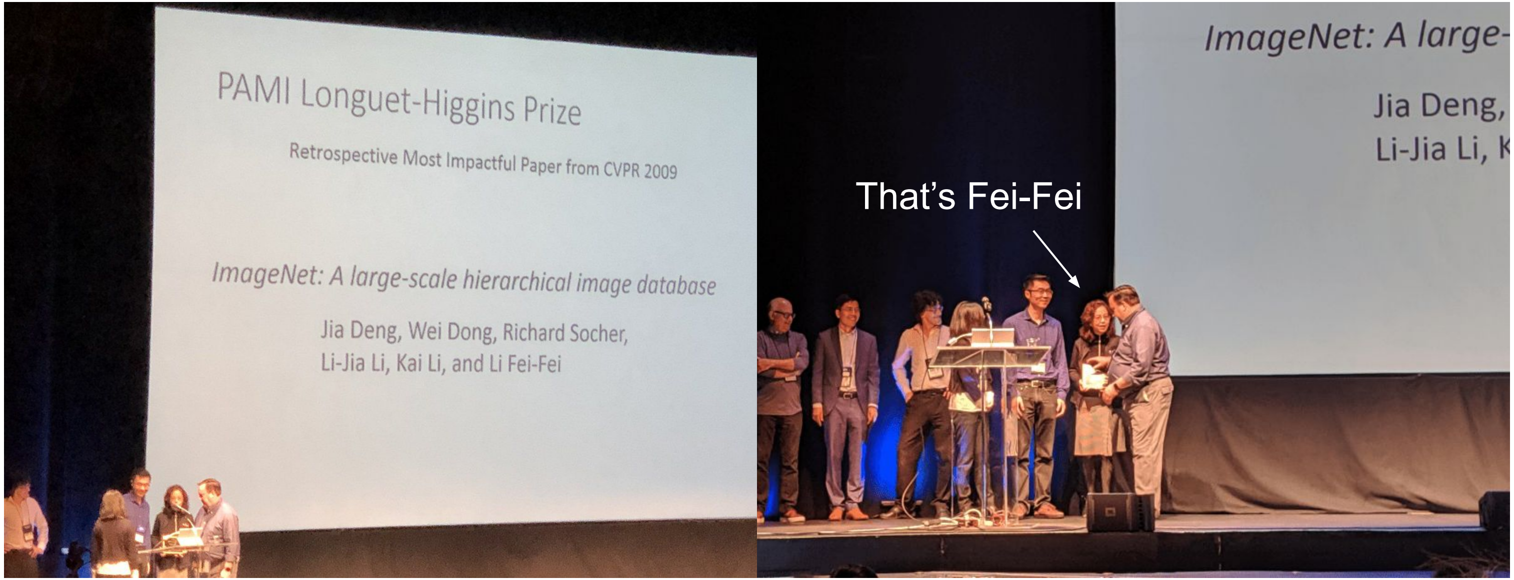
most prestigious technical award, is given for major contributions of lasting importance to computing.

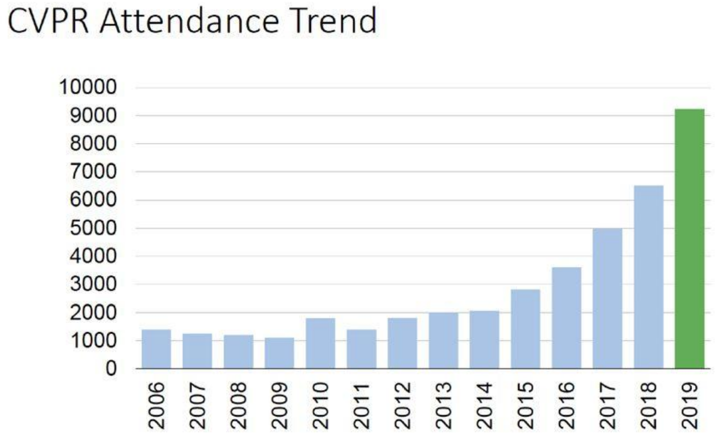


IEEE PAMI Longuet-Higgins Prize

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

At CVPR 2019, it was awarded to the 2009 original ImageNet paper





>8k submissions, 2,067 accepted papers

Logistics

Instructors


Fei-Fei Li


Jiajun Wu


Ruohan Gao

Teaching Assistants


Moo Jin Kim
(Head TA)


Dhruva Bansal


Shyamal Buch


Gokul Dharan


Agrim Gupta


Zhuoyi Huang


Drew Kaul


Sumith Kulal


Mihir Vipul Patel


Hongyu Ren


Manasi Sharma


William Shen


Haochen Shi


Stephen Su


Bohan Wu


Yinan Zhang

Lectures

- **Tuesdays** and **Thursdays** between **1:30pm to 3:00pm at NVIDIA Auditorium**

- Slides will be posted on the course website shortly before each lecture

- All lectures will be recorded and uploaded to Canvas after the lecture under the "Panopto Course Videos" Tab.

Course website

CS231n: Deep Learning for Computer Vision

Stanford - Spring 2022

Schedule

- **Lectures** will occur Tuesday/Thursday from 1:30-3:00pm Pacific Time at NVIDIA Auditorium.
- **Discussion** sections will (generally) occur Friday from 11:30am-12:30pm Pacific Time on Zoom. Check Ed for any exceptions.

Updated lecture slides will be posted here shortly before each lecture. For ease of reading, we have color-coded the lecture category titles in blue, discussion sections (and final project poster session) in yellow, and the midterm exam in red. Note that the schedule is subject to change as the quarter progresses.

Date	Description	Course Materials	Events	Deadlines
03/29	**Lecture 1: Introduction** Computer vision overview Historical context Course logistics			
---	**Deep Learning Basics**			
03/31	**Lecture 2: Image Classification with Linear Classifiers** The data-driven approach K-nearest neighbor Linear Classifiers Algebraic / Visual / Geometric viewpoints SVM and Softmax loss	Image Classification Problem Linear Classification		
04/01	Python / Numpy Review Session [Colab] [Tutorial]	11:30am-12:30pm	Assignment 1 out	

Friday Discussion Sections

6 Discussion sections **Fridays 1:30pm - 2:30pm over Zoom**

04/01	Python / Numpy Review Session
04/08	Backprop Review Session
04/15	Final Project Overview and Guidelines
04/22	PyTorch / TensorFlow Review Session
04/29	Detection software & RNNs
05/06	Midterm Review Session

Hands-on tutorials, with more practical details than the main lecture

Check canvas for the Zoom link of the discussion sessions!

This Friday: Python / numpy / Colab (Presenter: Manasi Sharma)

Ed

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

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

Office Hours

We'll be using Zoom to hold office hours and QueueStatus to setup queues

- please see Canvas or Ed for the QueueStatus link
- TAs will admit students to their Zoom meeting rooms for 1-1 conversations when it's your turn using QueueStatus.
- Office hours is listed on the course webpage!

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

All assignments will be completed using Google Colab

Assignment 1: Will be out Friday, due 4/15 by 11:59pm

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

Grading

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

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 Assignments: 10% + 20% + 15% = 45%

In-Class Midterm Exam: 20%

Course Project: 35%

- Project Proposal: 1%
- Milestone: 2%
- Final Project Report: 29%
- Poster & Poster Session: 3%

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

AWS

We will have AWS Cloud credits available for projects

- Not for HWs (only for final projects)

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

We will have a tutorial for walking through the AWS setup

Overview on communication

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

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

Ed:

- Use this for most communication with course staff
- Ask questions about homework, grading, logistics, etc
- Use private questions only if your post will violate honor code if you release publicly.

Gradescope:

- For turning in homework and receiving grades

Canvas:

- For watching recorded lectures
- For watching recorded discussion sessions

Prerequisites

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)

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 PyTorch and TensorFlow

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?

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 2022 conference
- ICCV 2021 conference

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

- Perception team at Google AI, Vision at Google Cloud
- Vision at Meta AI
- Vision at Amazon AWS
- Nvidia, Tesla, Apple, Salesforce, ……

General interest

CS231n: Deep Learning for Computer Vision

- Deep Learning Basics (Lecture 2 – 4)
- Perceiving and Understanding the Visual World (Lecture 5 – 12)
- Reconstructing and Interacting with the Visual World (Lecture 13 – 16)
- Human-Centered Artificial Intelligence (Lecture 17 – 18)

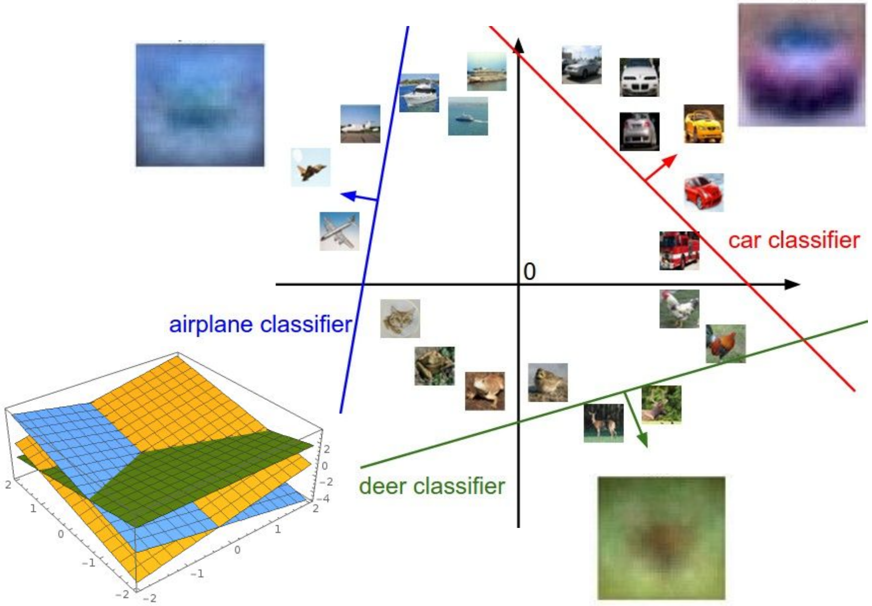
Syllabus

Deep Learning Basics	Convolutional Neural Networks	Computer Vision Applications
Data-driven learning	Convolutions	RNNs / Attention / Transformers
Linear classification & kNN	PyTorch / TensorFlow	Image captioning
Loss functions	Activation functions	Object detection and segmentation
Optimization	Batch normalization	Style transfer
Backpropagation	Transfer learning	Video understanding
Multi-layer perceptrons	Data augmentation	Generative models
Neural Networks	Momentum / RMSProp / Adam	Self-supervised learning
	Architecture design	3D vision
		Human-centered AI
		Fairness & ethics

Next time: Image classification with Linear Classifiers

k- nearest neighbor

Linear classification



Plot created using Wolfram Cloud