CS231n: Deep Learning for Computer Vision

UPDATED SYLLABUS AND ASSIGNMENTS

Spring 2025

Lecture 1 – Part 2 – Overview

Instructors

Fei-Fei Li

Ehsan Adeli

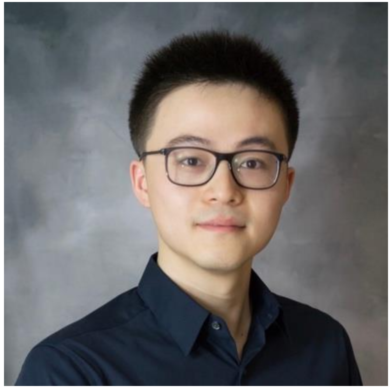
Zane Durante

Justin Johnson

Guest Lecturers

Jiajun Wu

Ranjay Krishna



Ruohan Gao

Yunzhu Li

CS231n overview

- Deep Learning Basics
- Perceiving and Understanding the Visual World
- Generative and Interactive Visual Intelligence
- Human-Centered Applications and Implications

Deep Learning Basics

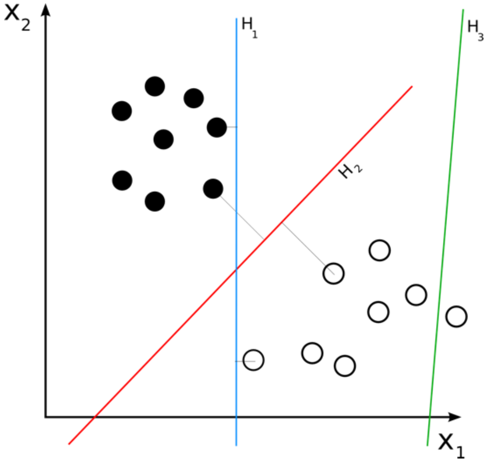
- Image Classification: A core task in Computer Vision

→ cat

Deep Learning Basics

- Image Classification: A core task in Computer Vision

→ cat

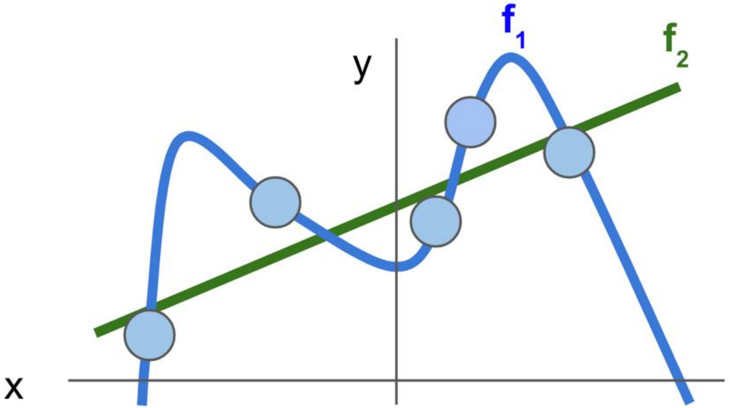


Linear Classifier

Deep Learning Basics

- Image Classification: A core task in Computer Vision

→ cat

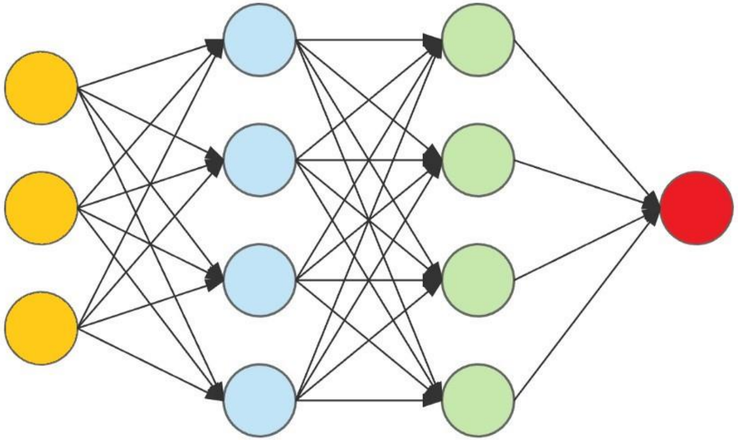


Regularization & Optimization

Deep Learning Basics

- Image Classification: A core task in Computer Vision

This image by Nikita is licensed under CC-BY 2.0



Neural Networks

CS231n overview

- Deep Learning Basics
- Perceiving and Understanding the Visual World
- Generative and Interactive Visual Intelligence
- Human-Centered Applications and Implications

CS231n overview

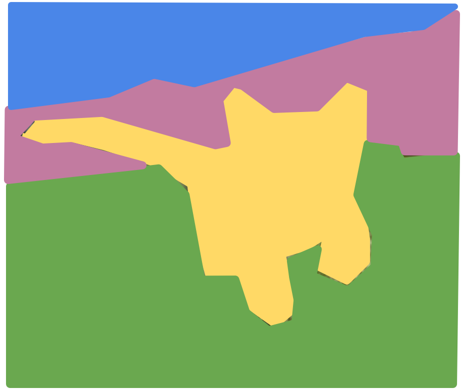
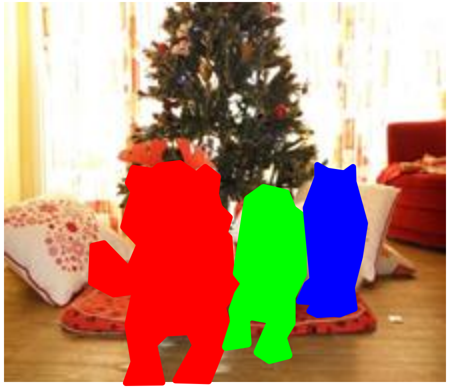
- Deep Learning Basics
- Perceiving and Understanding the Visual World
- Generative and Interactive Visual Intelligence
- Human-Centered Applications and Implications

Perceiving and Understanding the Visual World

Tasks

Models

Tasks 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	

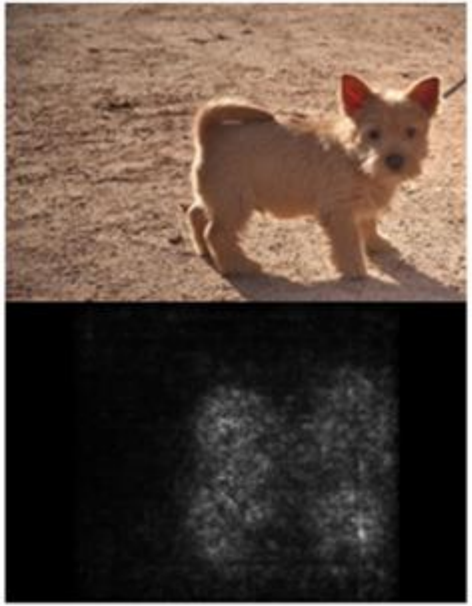
Tasks Beyond Image Classification

Video Classification

Running? Jumping?

Multimodal Video Understanding

Visualization & Understanding



Models Beyond Multi-Layer Perceptron

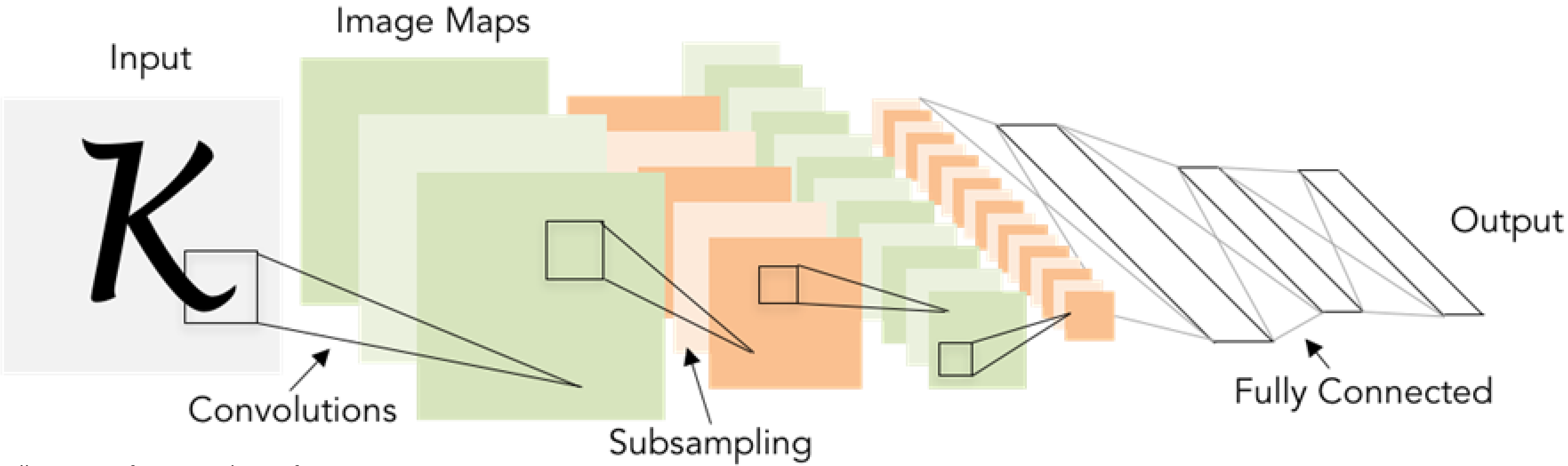
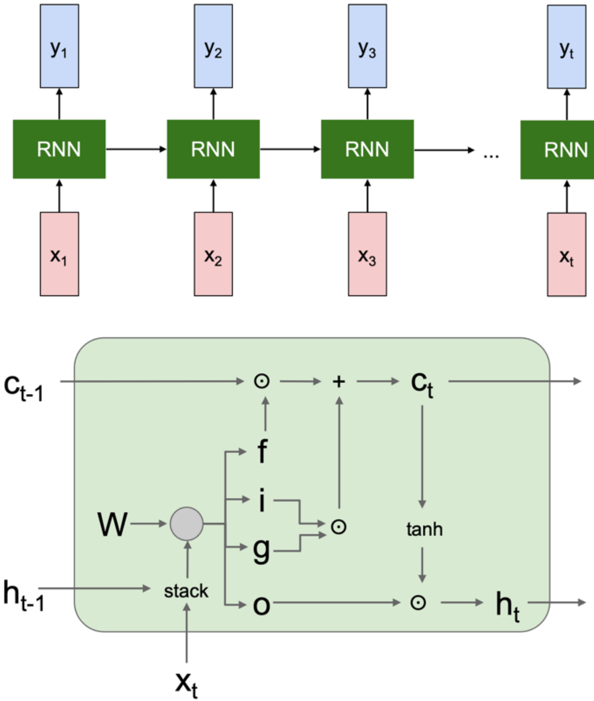


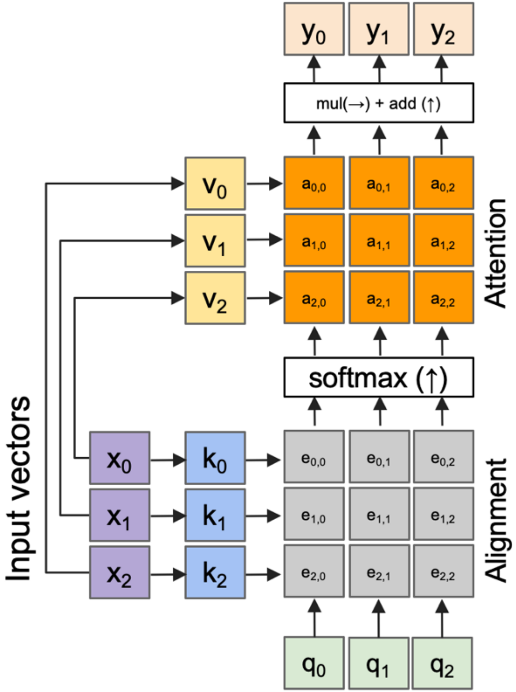
Illustration of LeCun et al. 1998 from CS231n 2017 Lecture 1

Convolutional neural network

Models Beyond Multi-Layer Perceptron



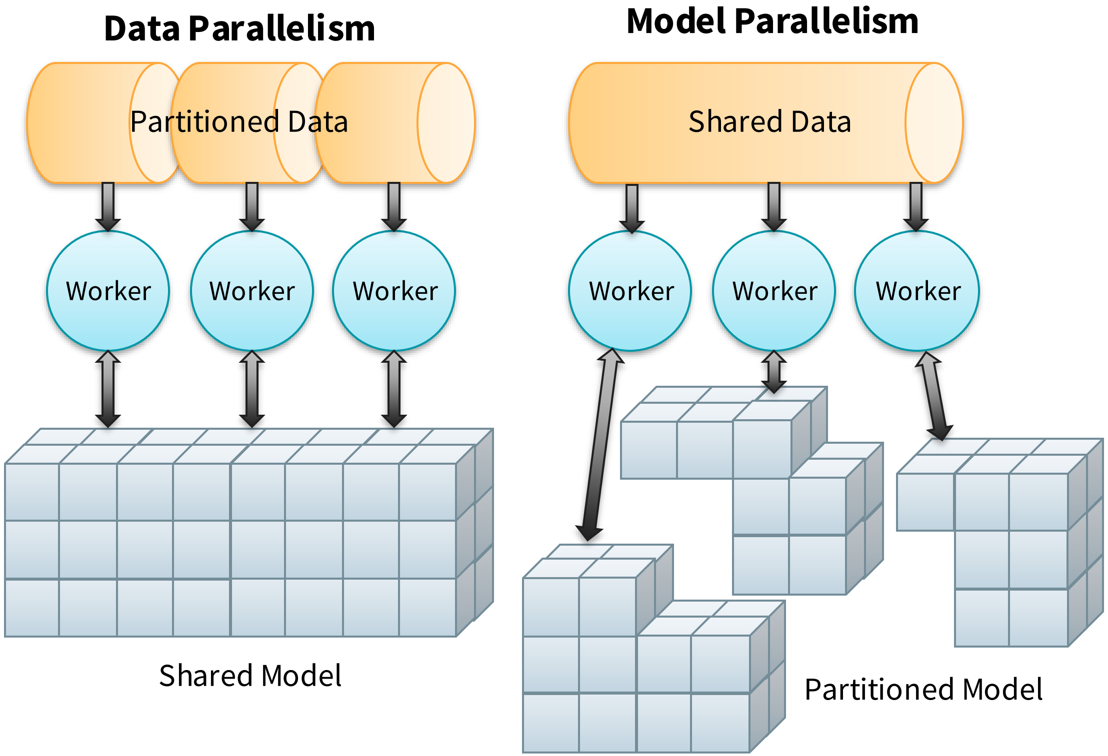
Recurrent neural network



Attention mechanism / Transformers

Large Scale Distributed Training

- Train Large Models on big datasets faster
- Scale beyond single GPU/machine limitations
- How?
  - Data Parallelism: Copy the model to all workers, split the data
  - Model Parallelism: Split model across devices
  - Synchronous vs. Asynchronous gradient updates



CS231n overview

- Deep Learning Basics
- Perceiving and Understanding the Visual World
- Generative and Interactive Visual Intelligence
- Human-Centered Applications and Implications

CS231n overview

- Deep Learning Basics
- Perceiving and Understanding the Visual World
- Generative and Interactive Visual Intelligence
- Human-Centered Applications and Implications

Beyond 2D Recognition

Beyond 2D Recognition: Self-supervised Learning

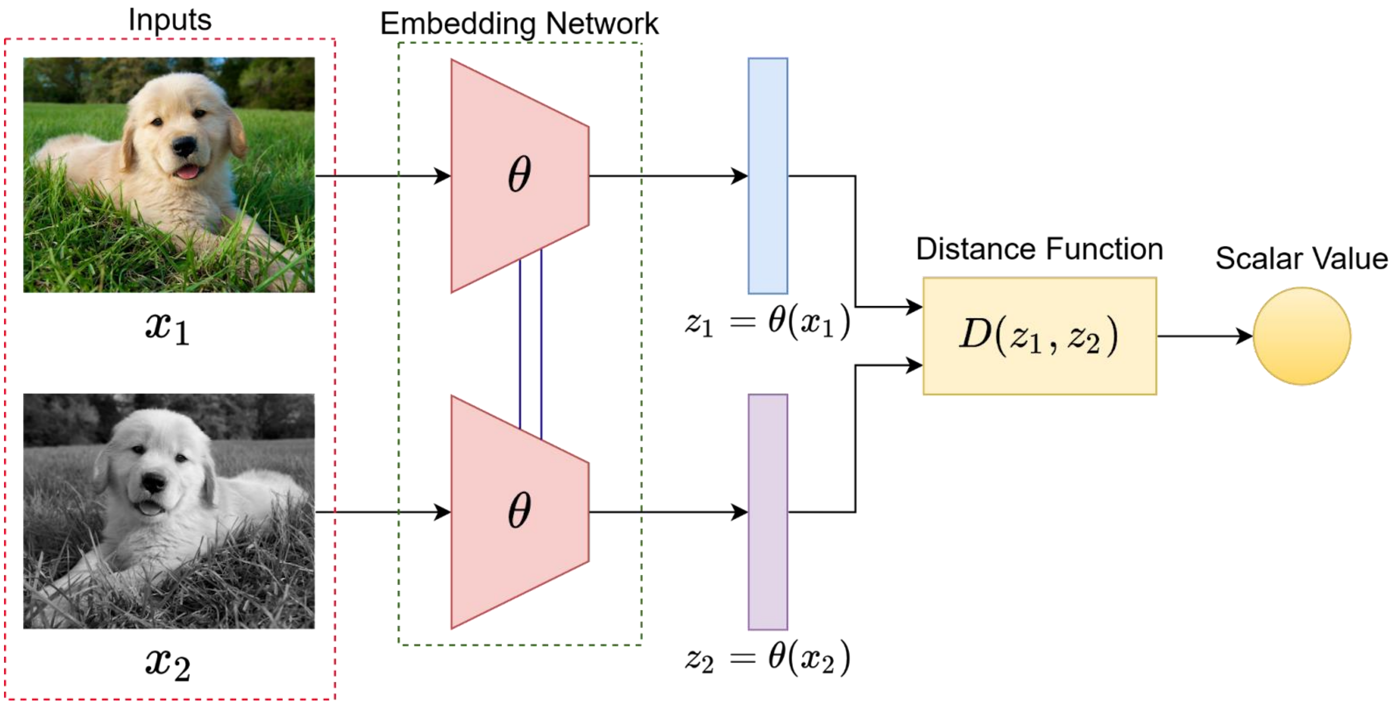


Image Courtesy of Rohit Kundu

Beyond 2D Recognition: Generative Modeling

Style Transfer

Beyond 2D Recognition: Generative Modeling

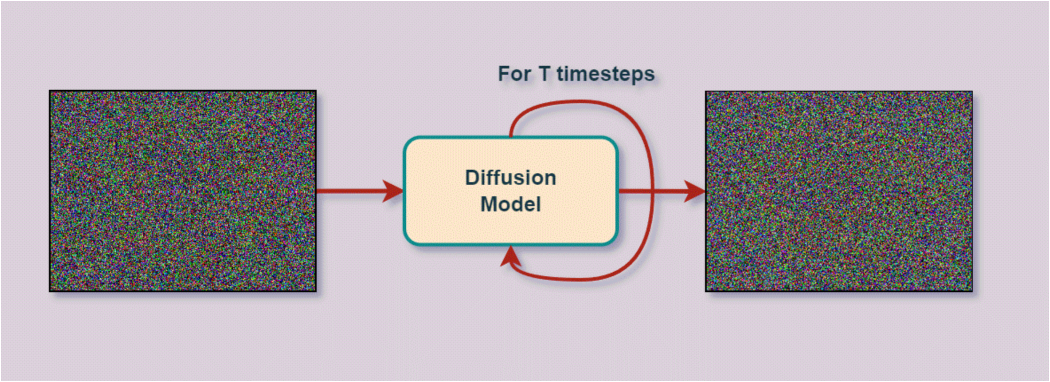
This image is public domain

"Teddy bears working on new AI research underwater with 1990s technology"

DALL-E 2

Beyond 2D Recognition: Generative Modeling

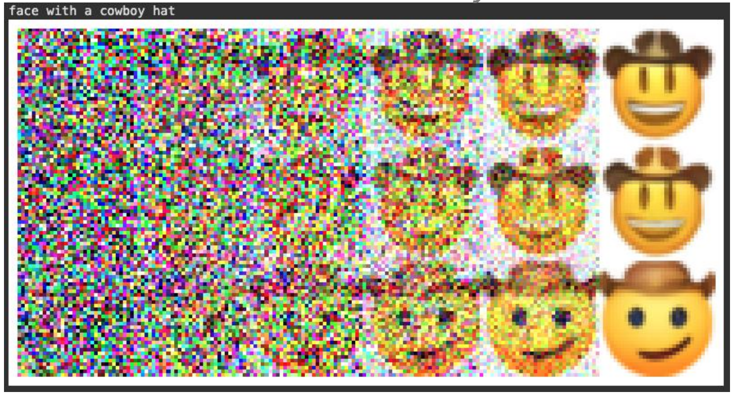
Image Generation using Diffusion Models



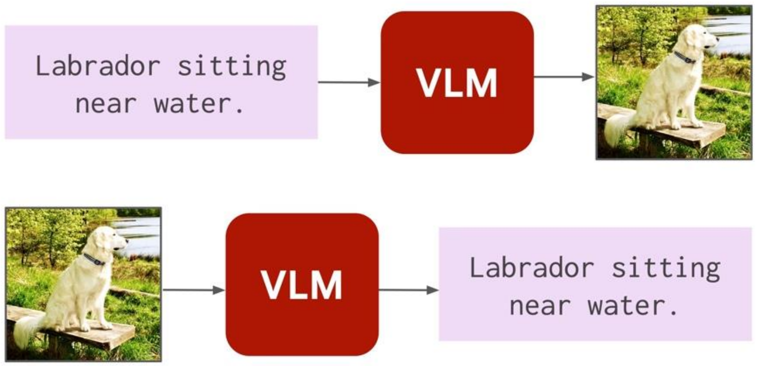
https://learnopencv.com/image-generation-using-diffusion-models/

You will learn and implement a generative model in Assignment 3 that generates emojis from text inputs

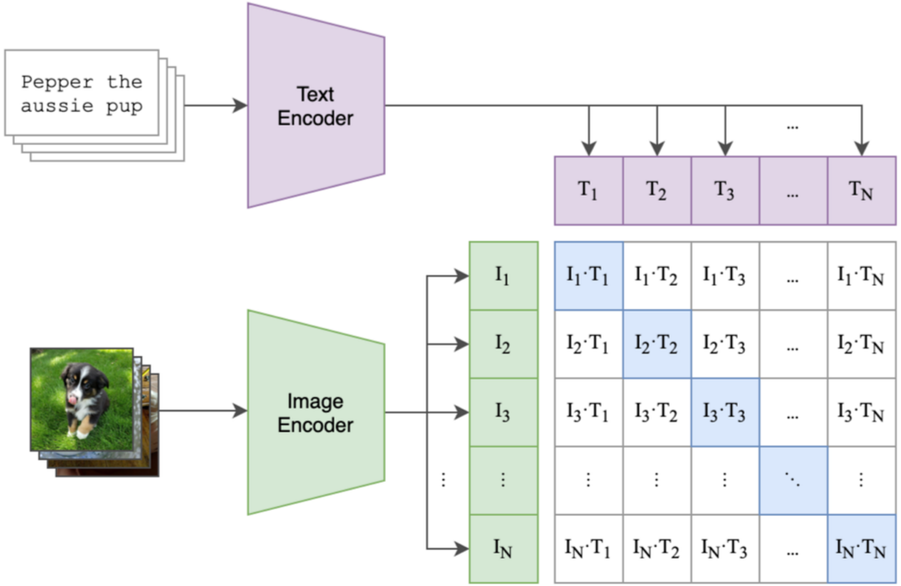
Face with a cowboy hat



Beyond 2D Recognition: Vision Language Models

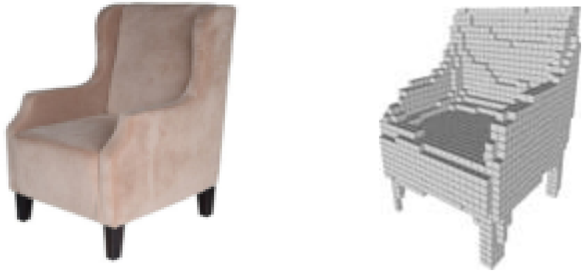


Yasunaga, Michihiro, et al. "Retrieval-augmented multimodal language modeling." arXiv preprint arXiv:2211.12561 (2022).

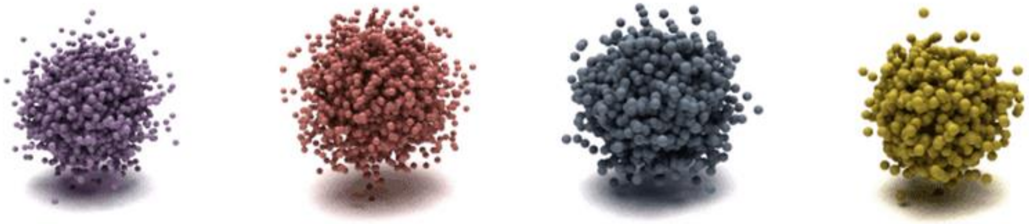


Contrastive pre-training in CLIP. The blue squares are the pairs for which we want to optimize the similarity. Image derived from https://github.com/openai/CLIP

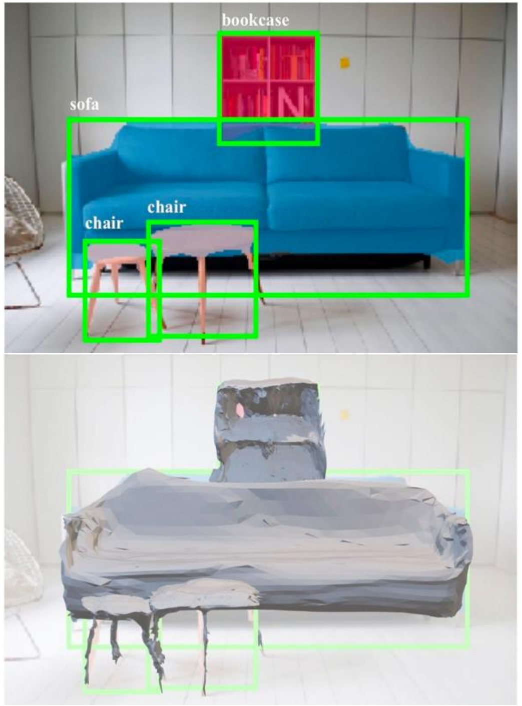
Beyond 2D Recognition: 3D Vision



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



Zhou et al., 3D Shape Generation and Completion through Point-Voxel Diffusion (2021)



Gkioxari et al., "Mesh R-CNN", ICCV 2019

Beyond 2D Recognition: Embodied Intelligence

Li et al., BEHAVIOR-1K: A Benchmark for Embodied AI with 1,000 Everyday Activities and Realistic Simulation (2022)

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

CS231n overview

- Deep Learning Basics
- Perceiving and Understanding the Visual World
- Generative and Interactive Visual Intelligence
- Human-Centered Applications and Implications

CS231n overview

- Deep Learning Basics
- Perceiving and Understanding the Visual World
- Generative and Interactive Visual Intelligence
- Human-Centered Applications and Implications

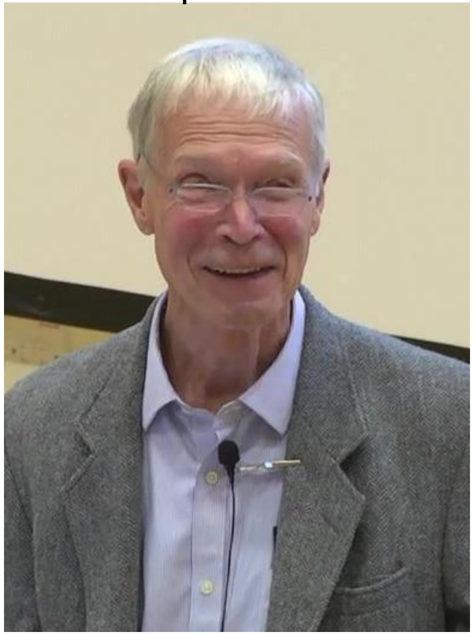
2018 Turing Award for deep learning

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

2024 Nobel Prize in Physics

Hinton speaking at the Nobel Prize Lectures in Stockholm in 2024

John Hopefield in 2016

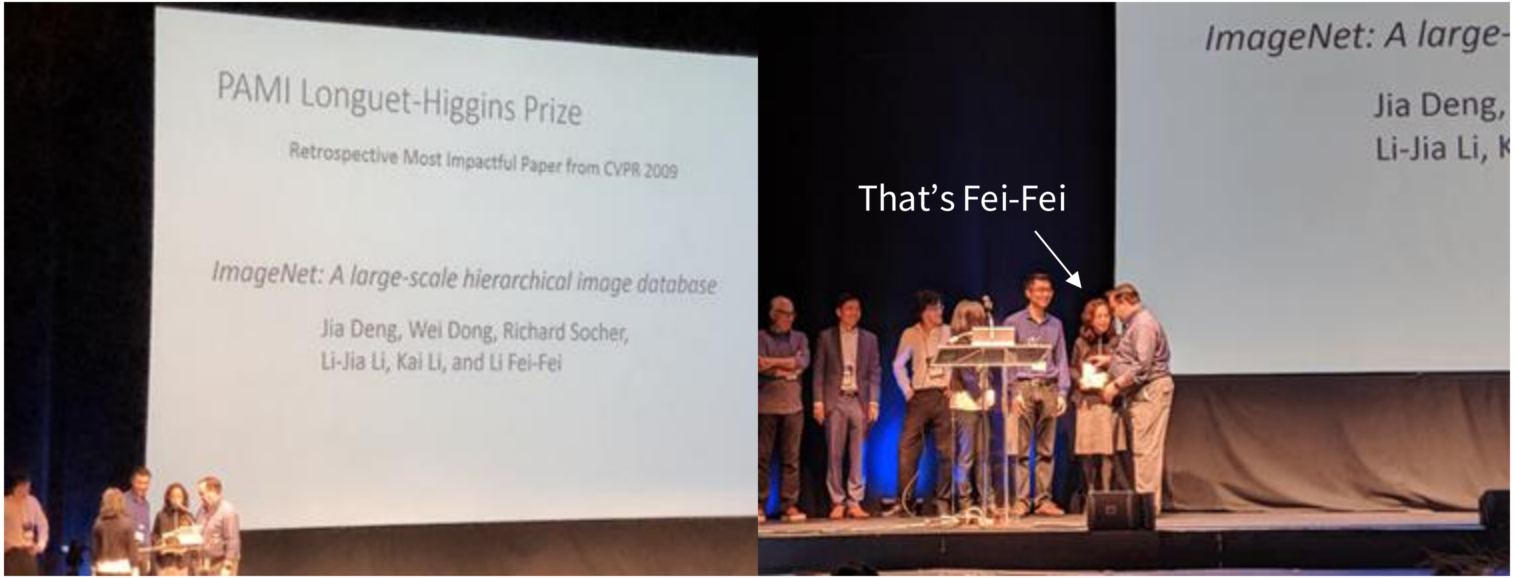


In 2024, he was jointly awarded the Nobel Prize in Physics with John Hopfield "for foundational discoveries and inventions that enable machine learning with artificial neural networks."

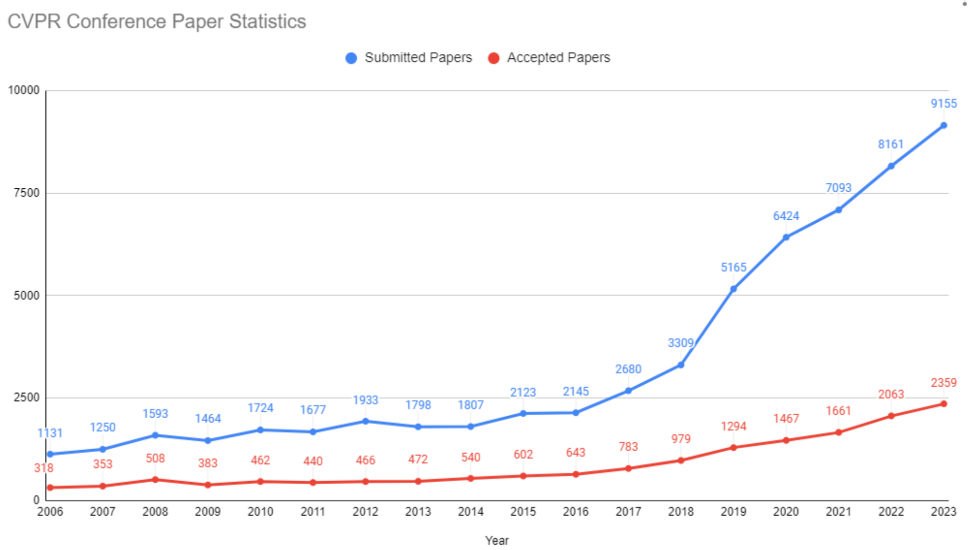
IEEE PAMI Longuet-Higgins Prize

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



CVPR Nashville JUNE 11-15, 2025



Logistics

Lectures

- Tuesdays and Thursdays between 12:00 PM to 1:20 PM at NVIDIA Auditorium
- Lectures will not be streamed on Zoom but will be broadcast live via Panopto
- 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 [http://cs231n.stanford.edu/] - Refresh!

Schedule

- **Lectures** will occur Tuesday/Thursday from 12:00-1:20pm Pacific Time at NVIDIA Auditorium.
- **Discussion** sections will (generally) occur on Fridays from 12:30-1:20pm Pacific Time at NVIDIA Auditorium. 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
04/01	**Lecture 1: Introduction** Computer vision overview Course overview Course logistics			
---	**Deep Learning Basics**			
04/03	**Lecture 2: Image Classification with Linear Classifiers** The data-driven approach K-nearest neighbor Linear Classifiers Algebraic / Visual / Geometric viewpoints Softmax loss			
04/04	Python / Numpy Review Session	12:30-1:20pm PT		
04/08	**Lecture 3: Regularization and Optimization** Regularization Stochastic Gradient Descent Momentum, AdaGrad, Adam Learning rate schedules			
04/09			Assignment 1 **out**	
04/10	**Lecture 4: Neural Networks and Backpropagation** Multi-layer Perceptron Backpropagation			
04/11	Backprop Review Session	12:30-1:20pm PT		

Friday Discussion Sections

6 Discussion sections Fridays 12:30-1:20 pm, NVIDIA Auditorium

04/04	Python / Numpy Review Session
04/11	Backprop Review Session
04/18	Final Project Overview and Guidelines
04/25	PyTorch / TensorFlow Review Session
05/02	RNNs & Transformers
05/09	Midterm Review Session

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

Check Canvas for the Zoom link for the discussion sessions! Recordings will be available on Canvas.

This Friday: Python / numpy / Colab

Ed

For questions about assignments, final project, midterm, logistics, etc, use Ed!

Access: Canvas -> Deep Learning for Computer Vision -> Ed Discussion

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

Office Hours

We'll be hosting both in-person and remote office hours. (starting week 2)

- Location
    - In-person: Huang Basement, check for CS231n signs, check the course website and Canvas
    - Remote: Zoom 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.
- The office hour schedule is on the course website

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 the honor code if you release it publicly

Course email address:

- cs231n-staff-spr25@stanford.edu

Gradescope:

- For turning in homework and receiving grades

Canvas:

- For watching recorded lectures
- For watching recorded discussion sessions

Assignments

All assignments will be completed using Google Colab

Assignment 1: Will be out Wednesday 4/9, due 4/23 by 11:59 PM

- K-Nearest Neighbor
- Linear classifiers: Softmax
- Two-layer neural network
- Image features
- Deep neural network and optimizers

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: 12% + 18% + 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
- Afterward, 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
- Rule 4: Do not submit AI-generated responses

Turning in something late/incomplete is better than violating the honor code

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

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

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 2025 conference
- ICCV 2025 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, Apple, Microsoft, OpenAI, Salesforce, ……

Apply computer vision to solve problems in other fields of science & engineering

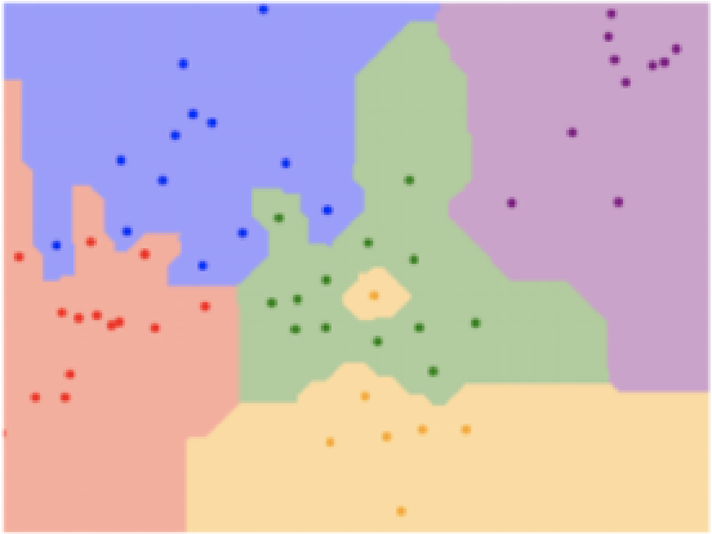
General interest

CS231n: Deep Learning for Computer Vision

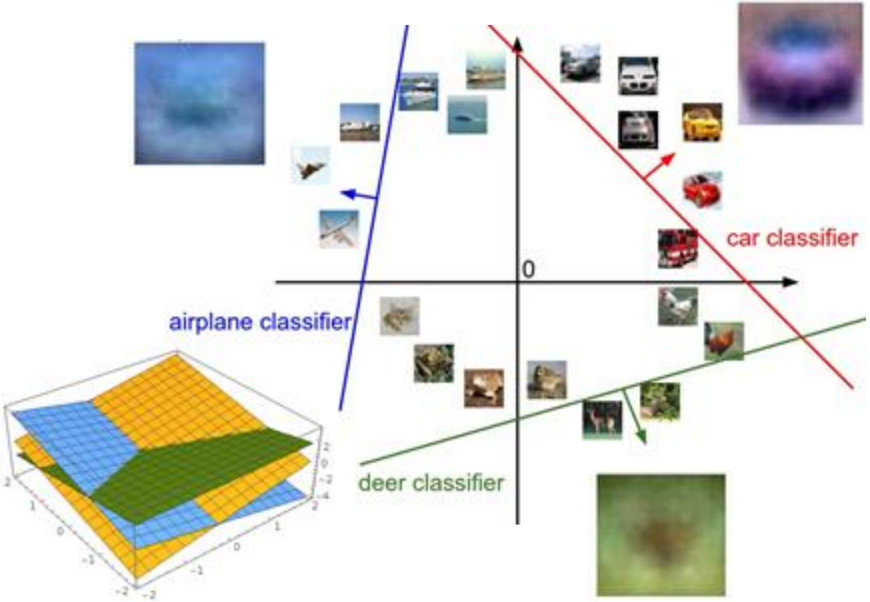
- Deep Learning Basics (Lecture 2 – 4)
- Perceiving and Understanding the Visual World (Lecture 5 – 12)
- Reconstructing and Interacting with the Visual World (Lecture 13 – 17)
- Human-Centered Artificial Intelligence (Lecture 18)

Next time: Image classification with Linear Classifiers

k- nearest neighbor



Linear classification



Plot created using Wolfram Cloud cs231n-staff-spr25@stanford.edu