

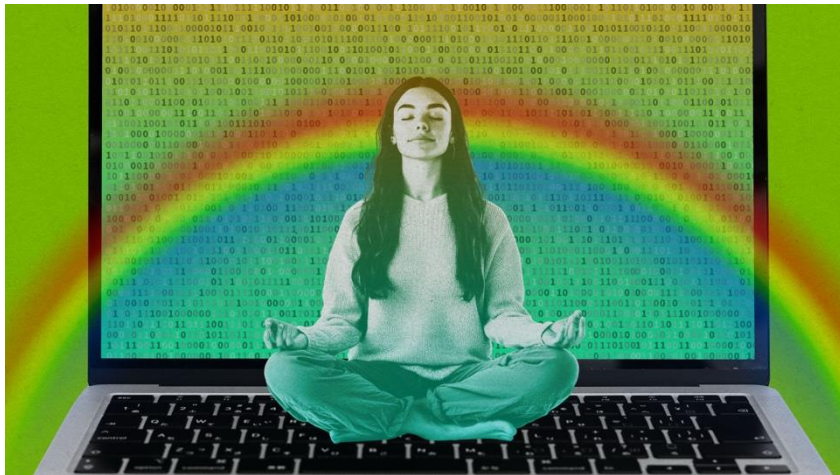
The background features a large, faint watermark of the Stanford University seal. The seal is circular and contains a redwood tree in the center, with the text 'STANFORD UNIVERSITY' at the top, '1891' at the bottom, and 'DIE LUFT DER FREIHEIT WEHT' around the inner border.

CS231n: Deep Learning for Computer Vision

Spring 2026

Lecture 1 – Part 2 – Overview

Vibe Coding??

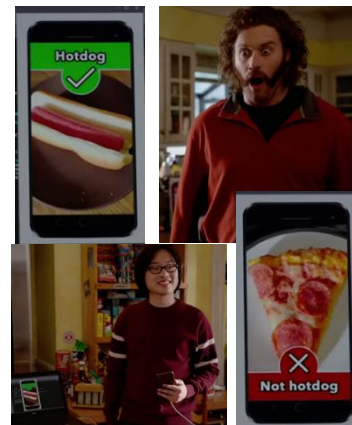


MIT Technology Reviews

CS231n Assignments??



"Hot Dog — Not Hot Dog" episode of Silicon Valley (Season 4, Episode 4, titled "Server Error") originally aired on May 14, 2017



Instructors



Fei-Fei Li



Ehsan Adeli



Zane Durante



Justin Johnson



Tiange Xiang

Teaching Assistants

Teaching Assistants



Aditesh Kumar
(Head CA)



Keshigeyan
Chandrasegaran



Mark Endo



Aniket Gupta



Wenlong Huang



Chaitanya Patel



Yash Shah



Karan Singh



Bailey Trang



Heng Yu



Koven Yu



Cristóbal Eyzaguirre



Fangrui Huang



Eris Zhang



Favour Nerrise



Yalcin Tur



June Zheng



Yang Zheng

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



Today's agenda

- A brief history of computer vision
- CS231n overview

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

This image by Nikita is
licensed under [CC-BY 2.0](https://creativecommons.org/licenses/by/2.0/)

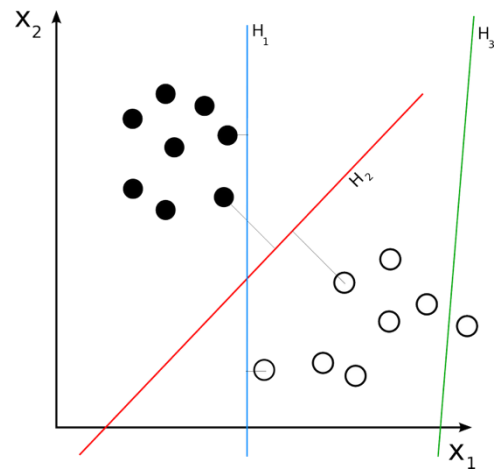
Deep Learning Basics

- Image Classification: A core task in Computer Vision



→ cat

This image by Nikita is licensed under [CC-BY 2.0](https://creativecommons.org/licenses/by/2.0/)



Linear Classifier

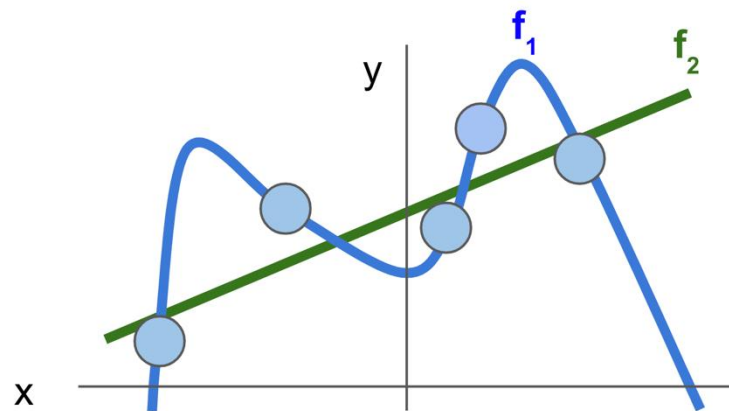
Deep Learning Basics

- Image Classification: A core task in Computer Vision



This image by Nikita is licensed under [CC-BY 2.0](https://creativecommons.org/licenses/by/2.0/)

→ cat



Regularization & Optimization

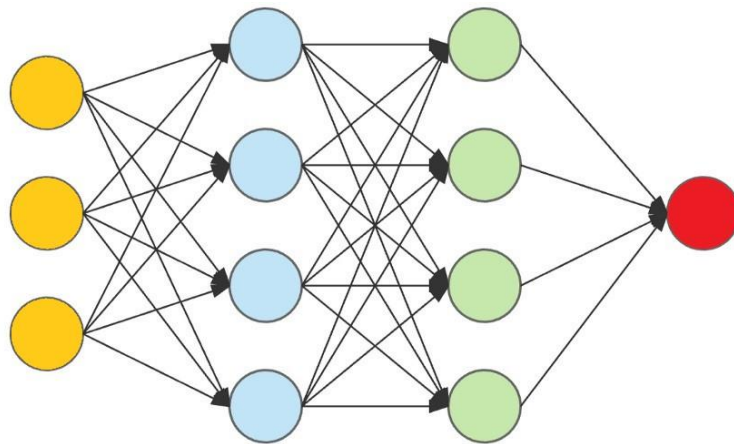
Deep Learning Basics

- Image Classification: A core task in Computer Vision



This image by Nikita is licensed under [CC-BY 2.0](https://creativecommons.org/licenses/by/2.0/)

→ cat



Neural Networks

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

A large orange circle containing the word "Tasks".

Tasks

A large blue circle containing the word "Models".

Models

Tasks Beyond Image Classification

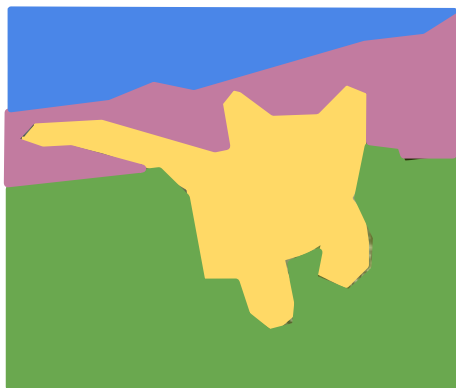
Classification



CAT

No spatial extent

Semantic Segmentation



GRASS, CAT, TREE, SKY

No objects, just pixels

Object Detection



DOG, DOG, CAT

Multiple Object

Instance Segmentation



DOG, DOG, CAT

[This image is CC0 public domain](#)

Tasks Beyond Image Classification

Video
Classification



Running? Jumping?

Multimodal Video
Understanding



Visualization &
Understanding



Models Beyond Multi-Layer Perceptron

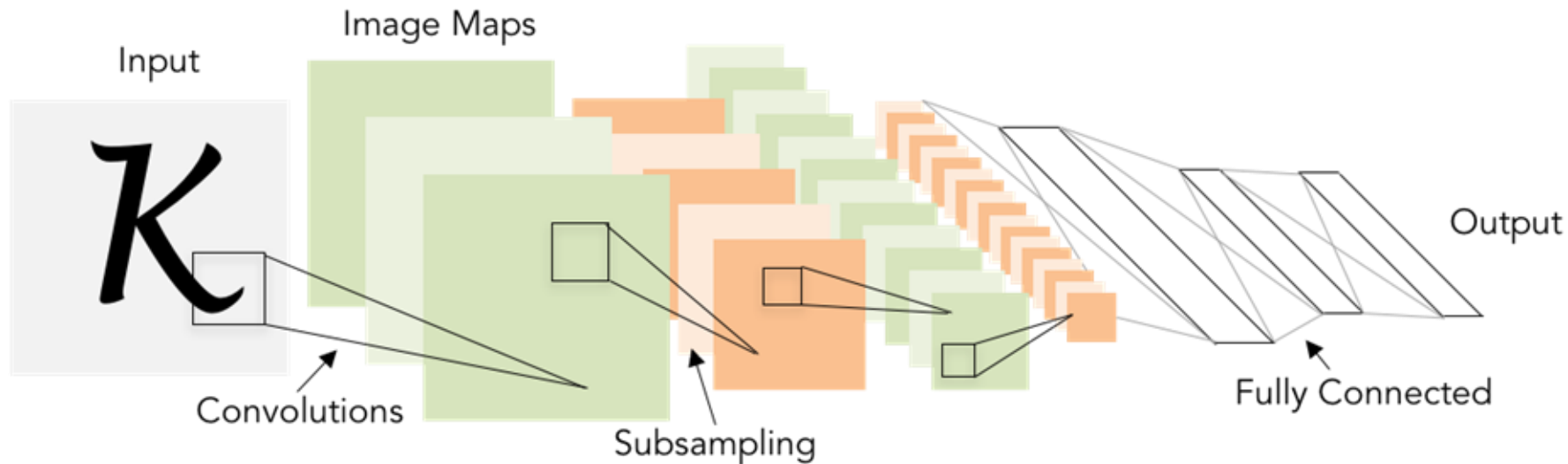
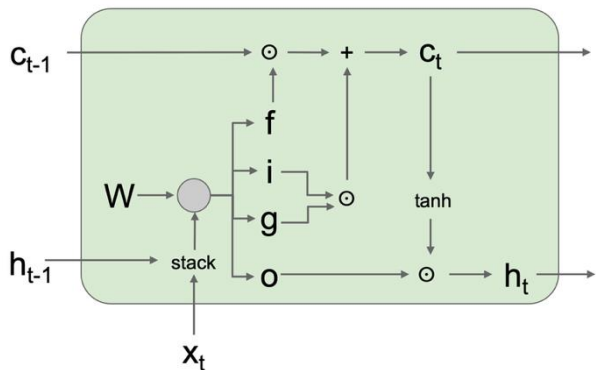
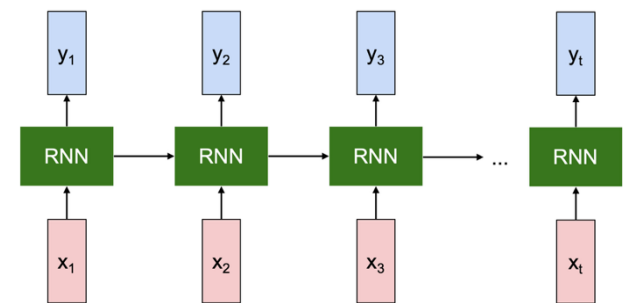


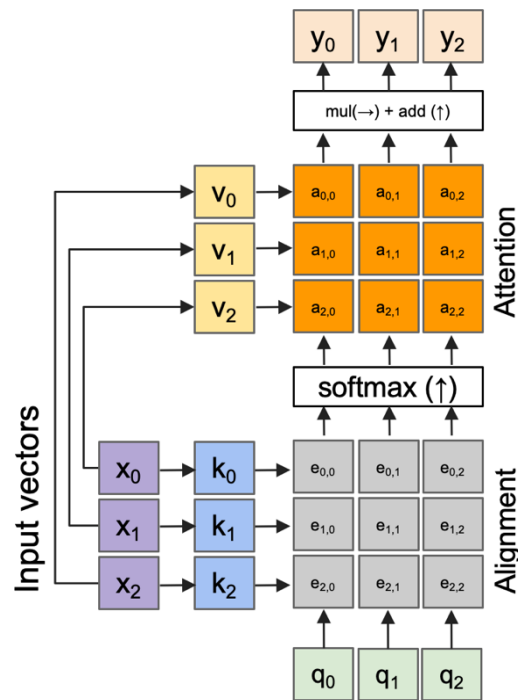
Illustration of LeCun et al. 1998 from CS231n 2017 Lecture 1

Convolutional neural network

Models Beyond Multi-Layer Perceptron



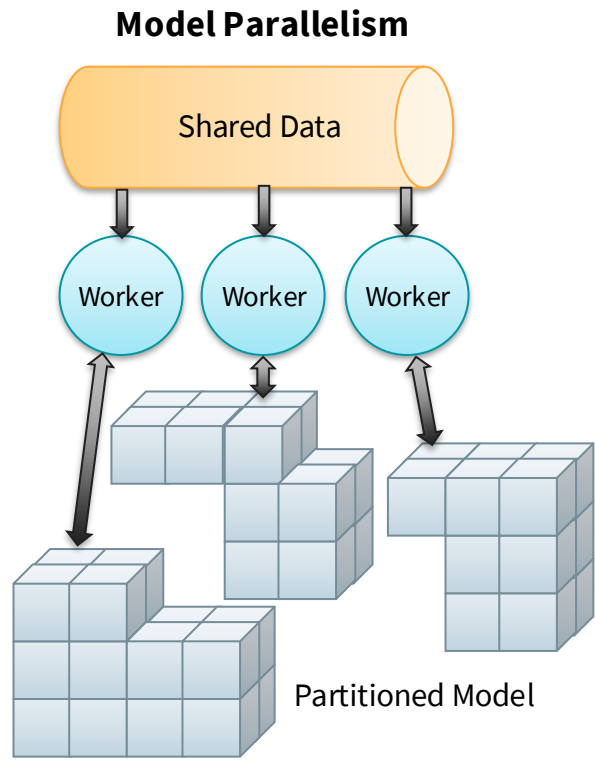
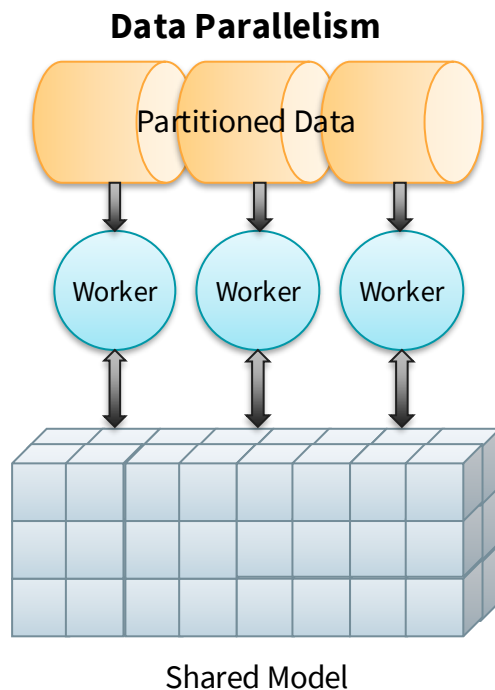
Recurrent neural network



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

Beyond 2D Recognition

Beyond 2D Recognition: Self-supervised Learning

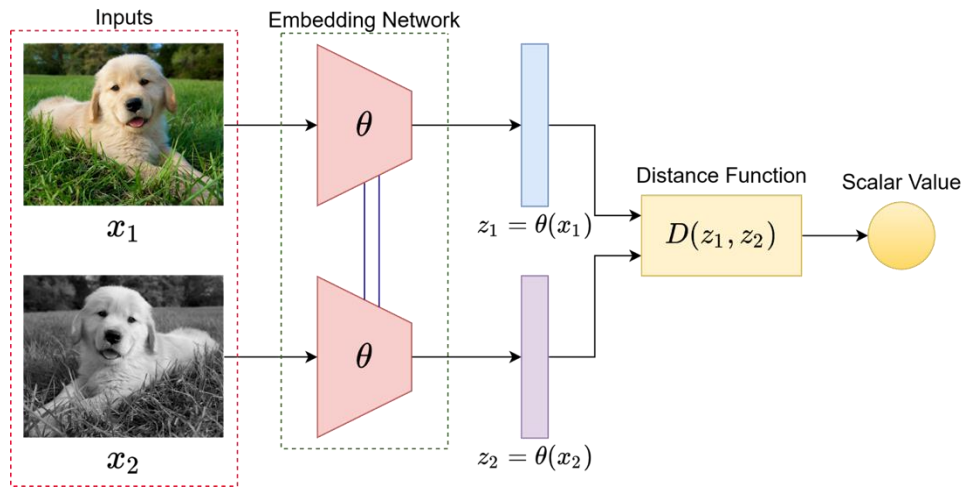


Image Courtesy of Rohit Kundu

No manually annotated labels required!



Meta

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

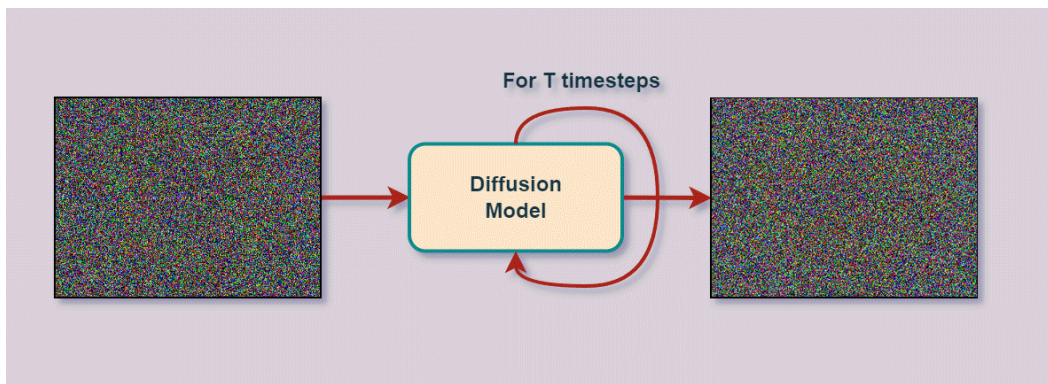


“Rendered in an intricate origami art style using complex, angular folds and crisp creases. A multi-layered diorama depicts a cute neighborhood street entirely from folded paper—houses with sharp rooflines, precise white picket fences, and layered, geometric flowers and rose bushes in vibrant paper hues. Focused lighting enhances the dimensionality. A vibrant origami cat, its body segmented by distinct, sharp folds, moves with articulated, deliberate steps along the paper sidewalk. Its limbs shift segment by segment, maintaining crisp creases as it progresses. The viewpoint tracks smoothly alongside the cat, revealing successive layers of the detailed papercraft neighborhood scrolling past, enhancing the scene’s geometric depth and dimensionality.”

Veo

Beyond 2D Recognition: Generative Modeling

Image Generation using Diffusion Models

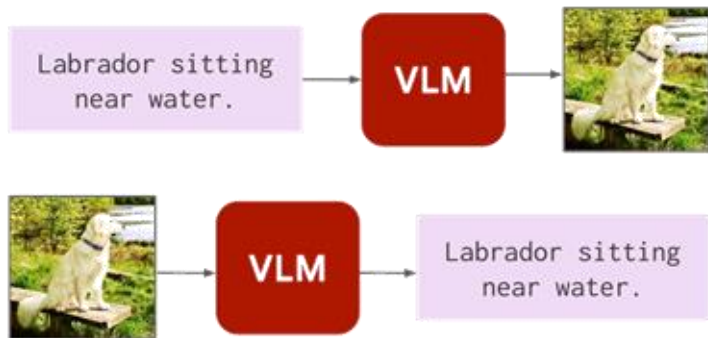


<https://leamopencv.com/image-generation-using-diffusion-models/>

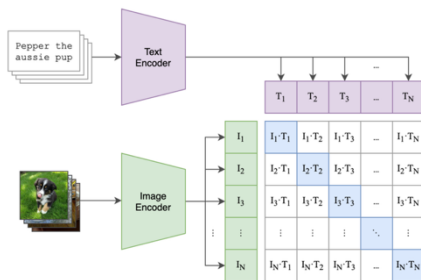
You will learn and implement a generative model in Assignment 3 that generates emojis from text inputs



Beyond 2D Recognition: Vision Language Models



Yasunaga, Michihiro, et al. "Retrieval-augmented multimodal language modeling." arXiv preprint arXiv:22.11.12561 (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>

Applications of VLMs

Query

Point to the Santa dog.



Response

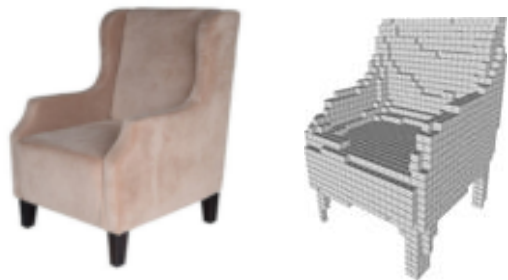
`<point x="74.8" y="31.5" alt="Santa dog">Santa dog</point>`

Segmentation

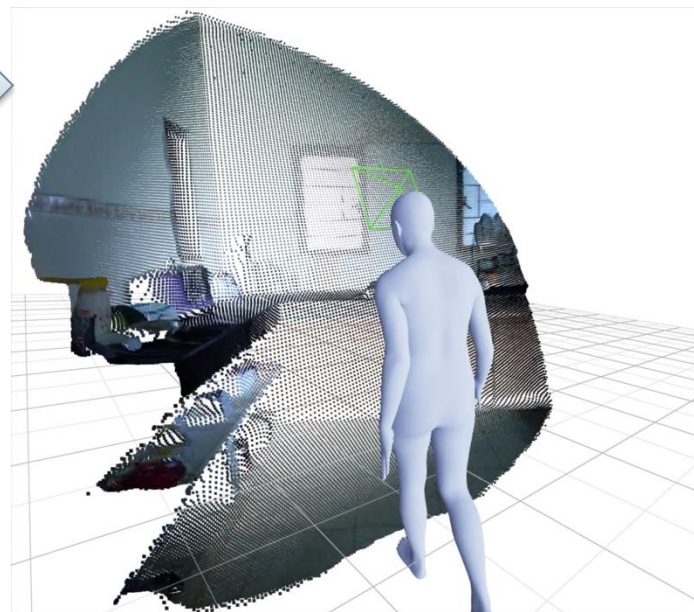
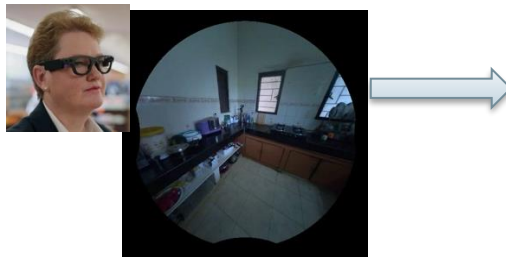


OpenCV

Beyond 2D Recognition: 3D Vision



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



Patel et al. (2026)



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

Beyond 2D Recognition: Embodied Intelligence



<https://www.pi.website/blog/pi05>



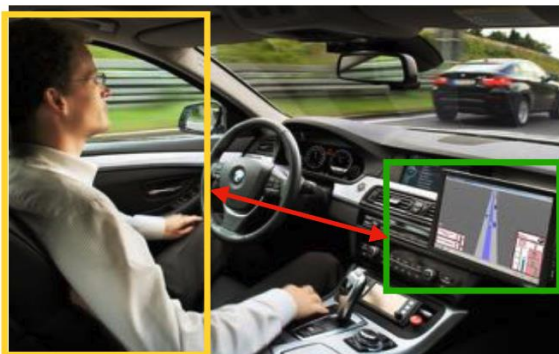
Mandlekar and Xu et al., Learning to Generalize Across Long-Horizon Tasks from Human Demonstrations (2020)

CS231n overview

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

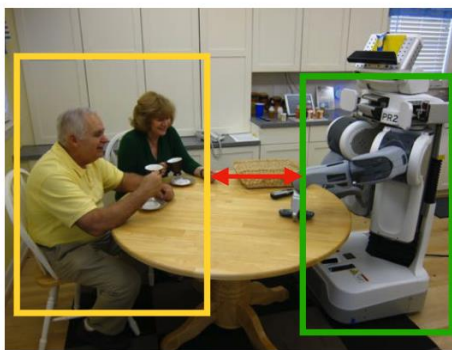
Human-Centered Applications and Implications

Self-Driving



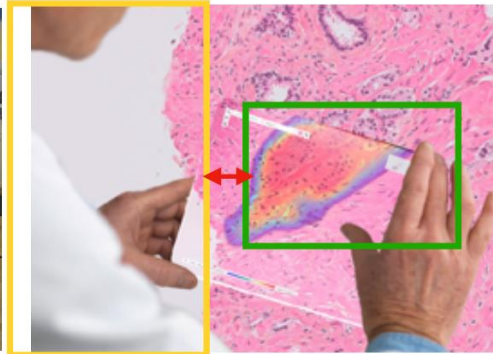
Driver and Autopilot

Home Robot



Elders and Robot

Healthcare



Physician and Medical AI

build trustworthy interactions between humans and AI

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 Tuesdays and Thursdays from 12:00-1:20pm Pacific Time at NVIDIA Auditorium.
- **Discussion** sections will (generally) occur on Fridays. Details TBD.

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
Mar 31	Lecture 1: Introduction Computer vision overview Course overview Course logistics			
Deep Learning Basics				
Apr 02	Lecture 2: Image Classification with Linear Classifiers The data-driven approach K-nearest neighbor Linear Classifiers Algebraic / Visual / Geometric viewpoints Softmax loss	Image Classification Problem Linear Classification	Assignment 1 out	
Apr 03	Python / Numpy Review Session [Colab] [Tutorial]	🕒 TBD		
Apr 07	Lecture 3: Regularization and Optimization Regularization Stochastic Gradient Descent Momentum, AdaGrad, Adam Learning rate schedules	Optimization		
Apr 09	Lecture 4: Neural Networks and Backpropagation Multi-layer Perceptron Backpropagation	Backprop Linear backprop example Suggested Readings: 1. Why Momentum Really Works 2. Derivatives notes 3. Efficient backprop 4. More backprop references: [1], [2], [3]		
Apr 10	Backprop Review Session	🕒 TBD		
Perceiving and Understanding the Visual World				
Apr 14	Lecture 5: Image Classification with CNNs History Higher-level representations, image features Convolution and pooling	Convolutional Networks		

Friday Discussion Sections

6 Discussion sections Fridays (Check the course website for time & location)

04/03	Python / Numpy Review Session
04/10	Backprop Review Session
04/17	Final Project Overview and Guidelines
04/24	PyTorch / TensorFlow Review Session
05/01	RNNs & Transformers
05/08	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 April 2nd)

- Location
 - In-person: Huang Basement, check for CS231n signs, Check the course website and Canvas
 - Remote: Zoom and QueueStatus to set up 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-spr26@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 Thursday 4/2, due 4/16 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](#).

- **All assignments have a written and a coding section**
- The code section is intended to get students' hands-on experience with the gritty implementation details of widely used algorithms in AI/CV. The submission will be **autograded**.
- The written submission is intended to build intuition and conceptual understanding about the algorithms. The submission will be **graded by the TAs**.

Grading

3 Assignments: $12\% + 18\% + 15\% = 45\%$

In-Class Midterm Exam: 20%

Course Project: 35%

- Project Proposal: 1%
- Milestone Check-Ins: 3% each = 9%
- Final Project Report: 20%
- Poster & Poster Session: 5%

New



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 (including AI tools); 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 2026 conference](#)
- [ECCV 2026 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

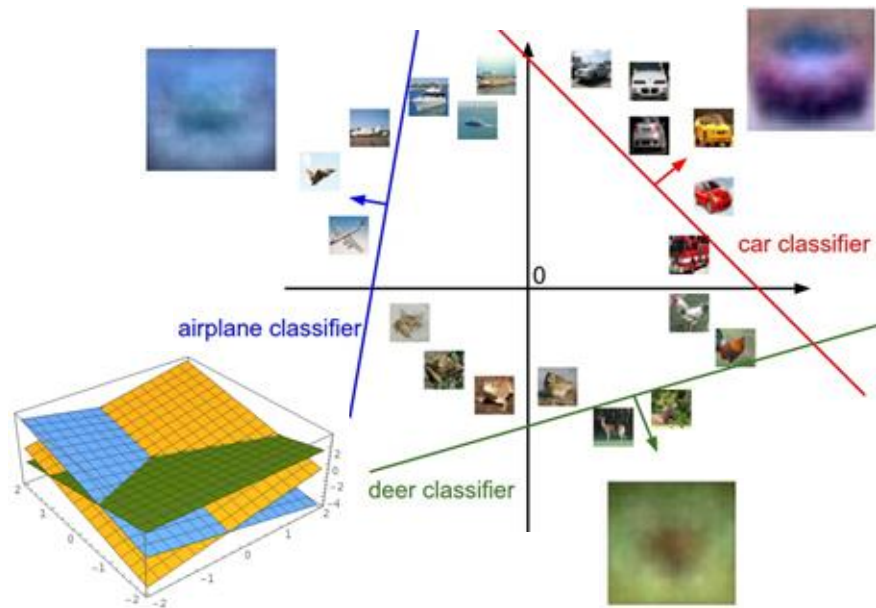
- Deep Learning Basics (Lecture 2 – 4)
- Perceiving and Understanding the Visual World (Lecture 5 – 12)
- Reconstructing and Interacting with the Visual World (Lecture 13 – 17)
- Human-Centered Artificial Intelligence (Lecture 18)

Next time: Image classification with Linear Classifiers

k- nearest neighbor



Linear classification



Plot created using [Wolfram Cloud](#)

cs231n-staff-spr26@stanford.edu