CS 4644 / 7643: Deep Learning

Website: https://www.cc.gatech.edu/classes/AY2023/cs7643_fall/
Piazza: https://piazza.com/gatech/fall2022/cs46447643
Code: GTDL@2022
Canvas: https://gatech.instructure.com/courses/286512 (4644)
https://gatech.instructure.com/courses/275392 (7643)
Gradescope: https://www.gradescope.com/courses/415232 (4644)
https://www.gradescope.com/courses/415719 (7643)

Danfei Xu

School of Interactive Computing
Georgia Tech

Are you in the right place?

- This is CS 4644(DL) / CS 7643
  - “On campus” class

- This is NOT CS 7643-O01/OAN/Q/R
  - Online class for OMSCS program (Prof. Zsolt Kira)

Fall 22 Delivery Format

- In-person
  - Paper Tricentennial

- Streaming & Recording
  - We STRONGLY encourage you to attend the lectures in person.
  - Lectures will be streamed over zoom (link on Canvas).
  - Lectures are recorded and available for viewing

- **Remember: Content is free online.**
  - **You are here for the interactive experience.**

Outline for Today

- What is Deep Learning, the field, about?

- What is this class about?
  - What to expect?
  - Logistics

- FAQ

Survey

Undergrad?

M.S.?

Ph.D.?

CS (CoC) / ECE?

Other Engineering?

Math / Natural Science?

Others?

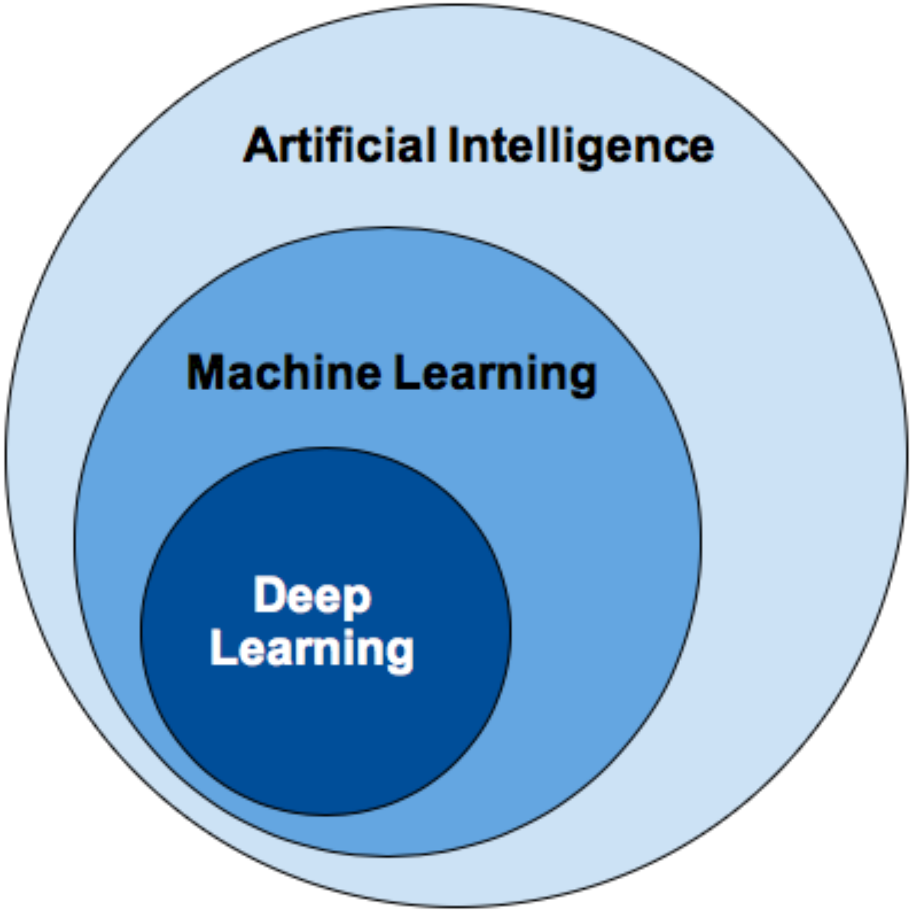
Outline

- What is Deep Learning, the field, about?

- What is this class about?
  - What to expect?
  - Logistics

- FAQ

Concepts



"Deep Learning is part of a broader family of **machine learning methods** based on **artificial neural networks**"

--- https://en.wikipedia.org/wiki/Deep_learning

Image Credit: https://www.sumologic.com/blog/machine-learning-deep-learning/

What is Artificial Intelligence?

- Boring textbook answer

Intelligence demonstrated by machines

  – Wikipedia

- What others say:

The science and engineering of making computers behave in ways that, until recently, we thought required human intelligence.

  – Andrew Moore, CMU

So what *is* Deep Learning?

- **Objective**: Representation Learning
  - Automatically discover useful features/representations for a **task** from raw data

- **Model**: (Deep) Artificial Neural Networks

- **Learning Method**: Supervised/Unsupervised/Reinforcement/Generative
  …
  Learning

- **Simply**: Deep Learning

So what *is* Deep Learning?

Ways to think about Deep Learning:

- (Hierarchical) Compositionality
  - Cascade of non-linear transformations
  - Multiple layers of representations

- End-to-End Learning
  - Learning (goal-driven) representations
  - Learning to feature extraction

Hierarchical Compositionality

VISION

pixels → edge → texton → motif → part → object

NLP

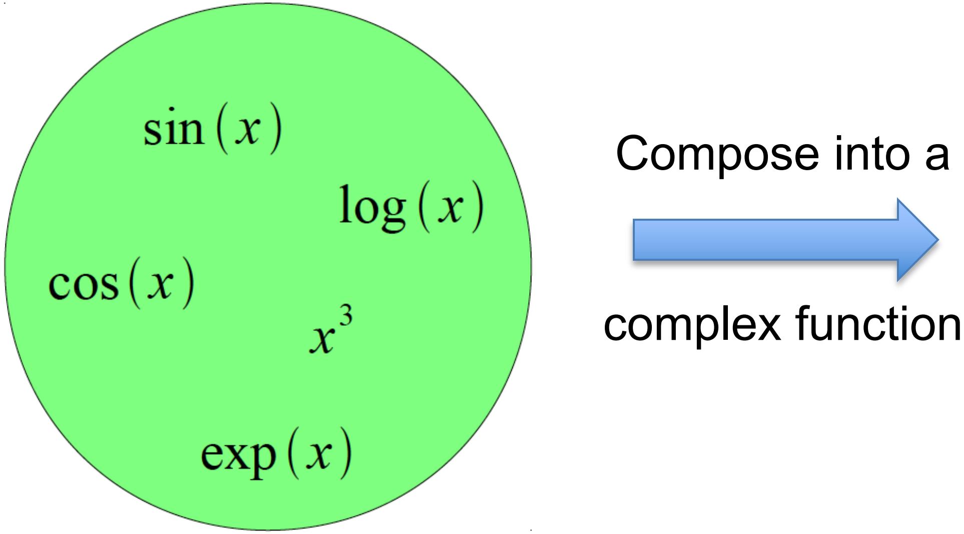
character → word → NP/VP/.. → clause → sentence → story

Simple Functions -> Complex Functions

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

Building A Complicated Function

Given a library of simple functions



$\sin(x)$

$\log(x)$

$\cos(x)$

x^3

$\exp(x)$

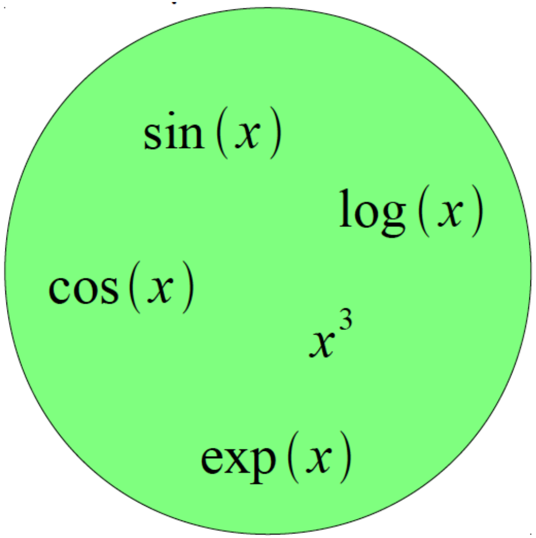
Compose into a

complex function

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

Building A Complicated Function

Given a library of simple functions



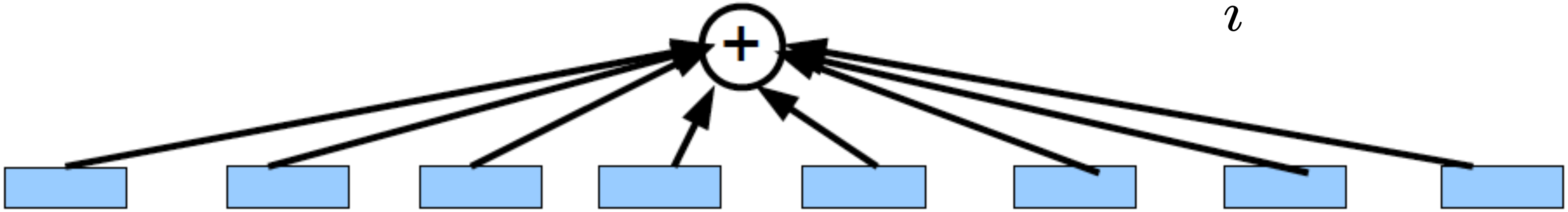
Compose into a

complex function

Idea 1: Linear Combinations

- Boosting
- Kernels
- …

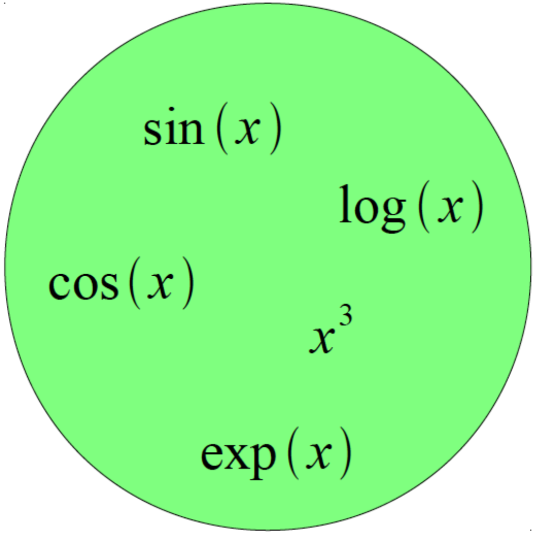
$$f(x) = \sum_i \alpha_i g_i(x)$$



Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

Building A Complicated Function

Given a library of simple functions



Compose into a complex function

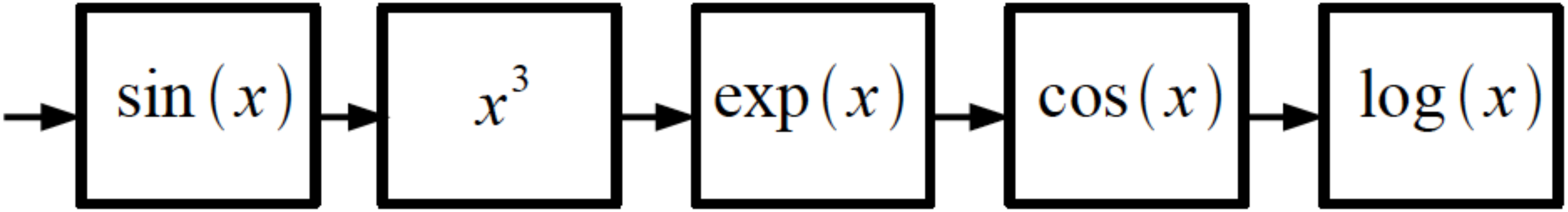
Idea 2: Compositions

Compose a set of functions (layers) through which the input data get transformed.

More layers = “Deeper”

Can we make it more expressive?

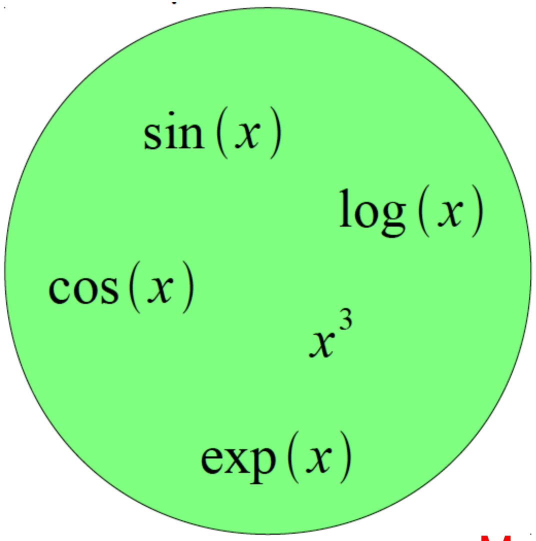
$$f(x) = \log(\cos(\exp(\sin^3(x))))$$



Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

Building A Complicated Function

Given a library of simple functions



Compose into a

complex function

Idea 2: Compositions

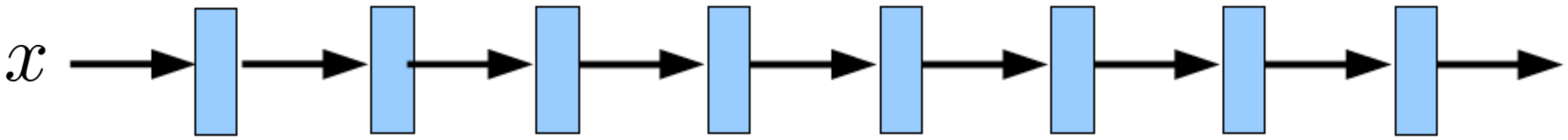
Compose a set of functions (layers) through which the input data get transformed.

More layers = “Deeper”

Yes! Parametric functions

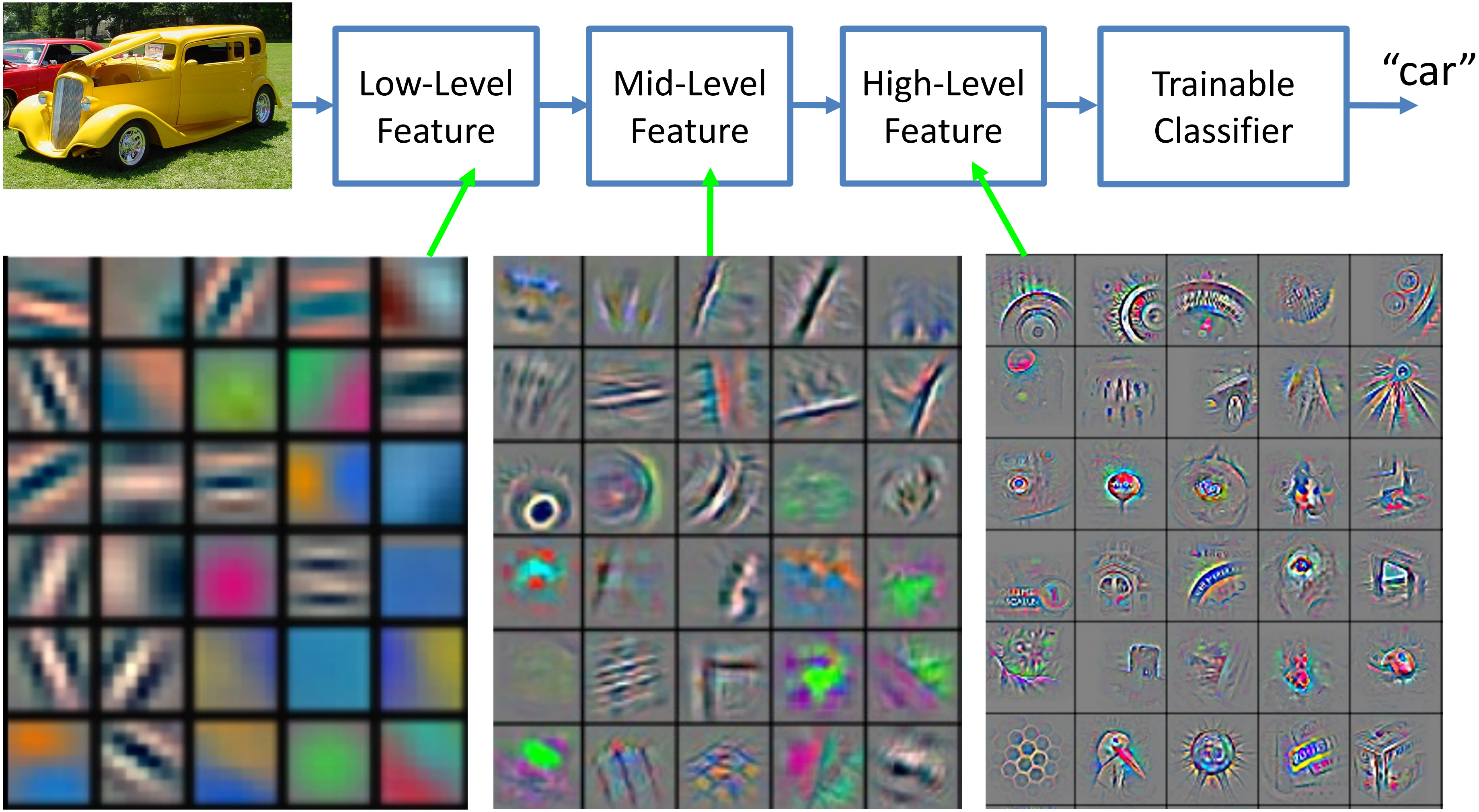
Modern DNNs have huge # of parameters, on the orders of bn’s

$$f_\theta(x) = \overbrace{g_{\theta_n}}^{\text{Parametric functions}}(\ldots g_{\theta_2}(g_{\theta_1}(x)\ldots))$$



Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

Deep Learning = Hierarchical Compositionality



Low-Level Feature

Mid-Level Feature

High-Level Feature

Trainable Classifier

“car”

Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

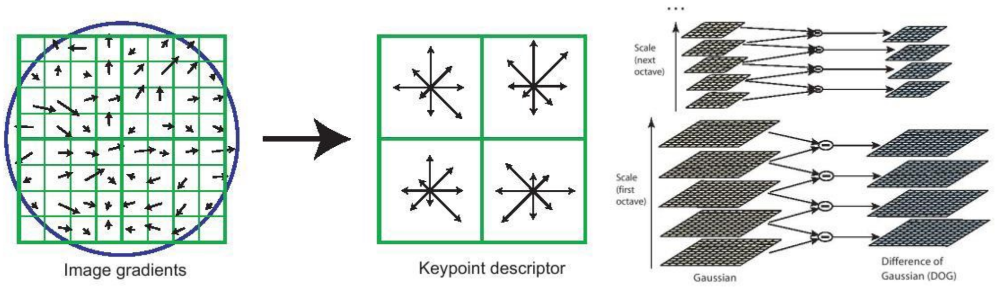
So what *is* Deep Learning?

Ways to think about Deep Learning:

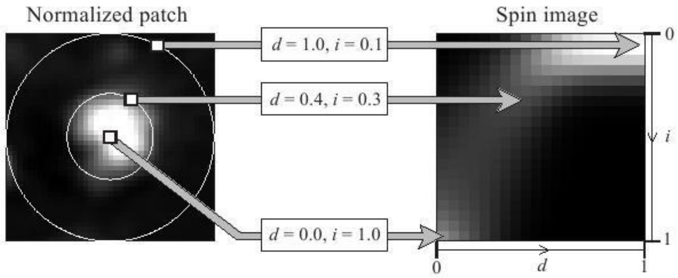
- (Hierarchical) Compositionality
  - Cascade of non-linear transformations
  - Multiple layers of representations

- End-to-End Learning
  - Learning (goal-driven) representations
  - Learning to feature extraction

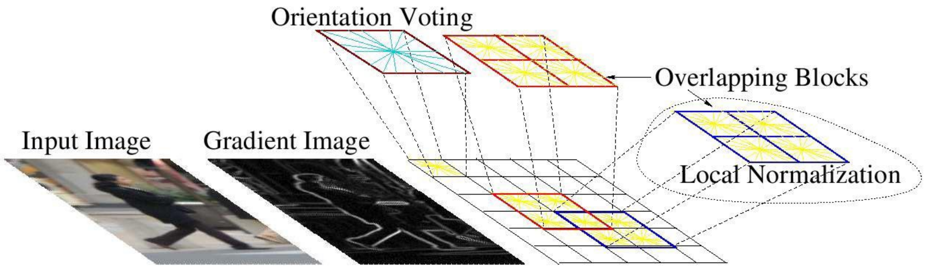
Pre-Deep Learning: Feature Engineering



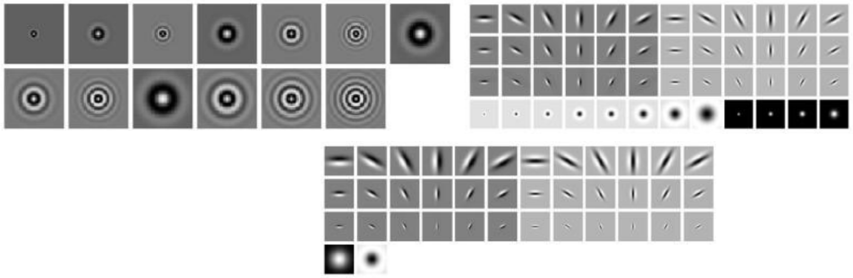
SIFT



Spin Images



HoG

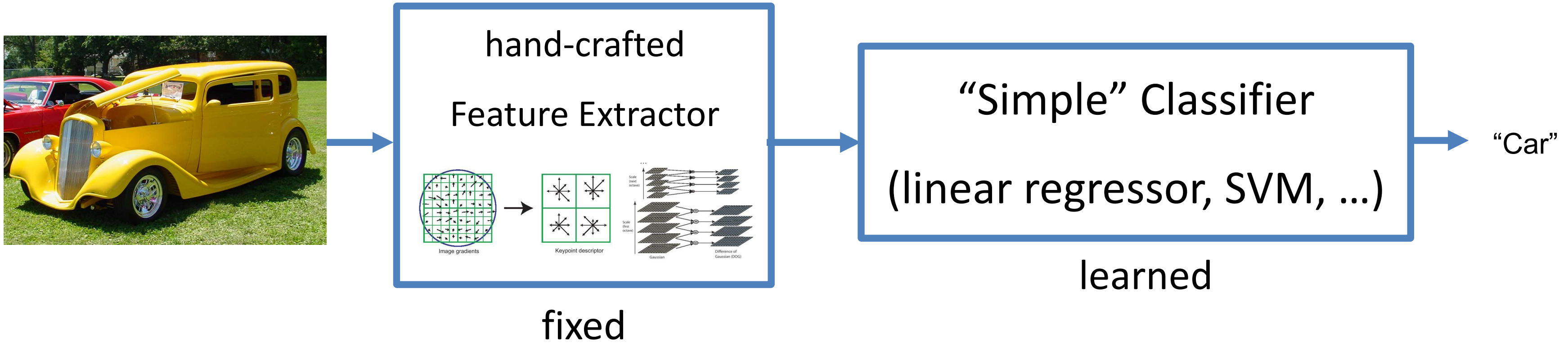


Textons

and many many more….

“Shallow” vs Deep Learning

- “Shallow” models



hand-crafted

Feature Extractor

fixed

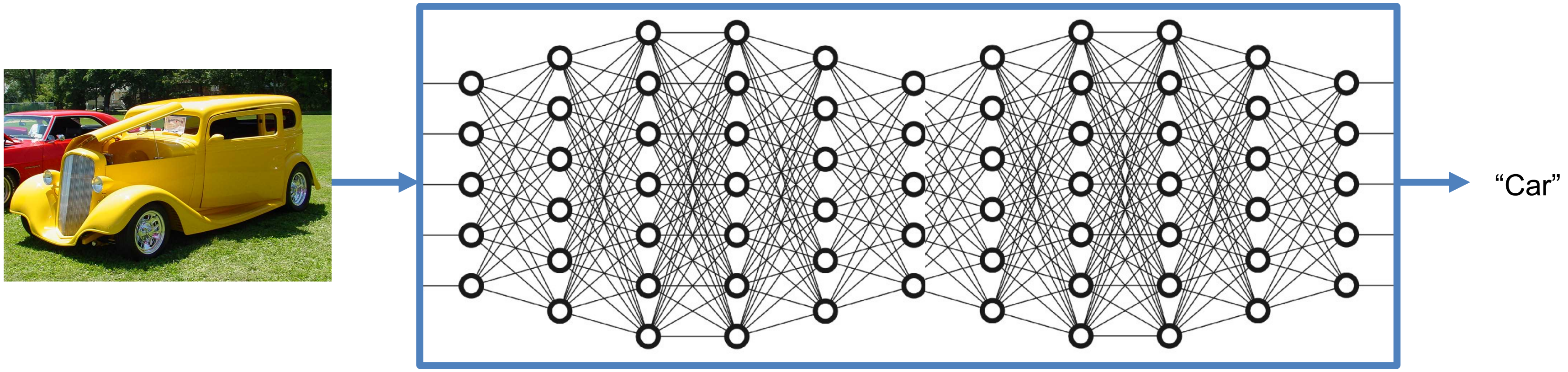
“Simple” Classifier

(linear regressor, SVM, …)

learned

“Car”

- Deep models



“Car”

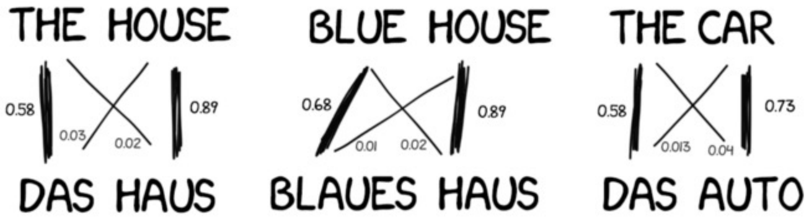
Learned

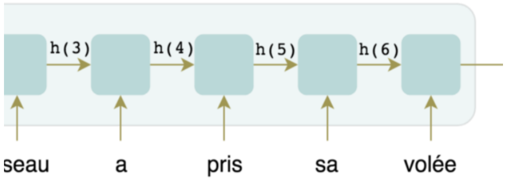
(C) D. Batra, Z. Kira, D. Xu

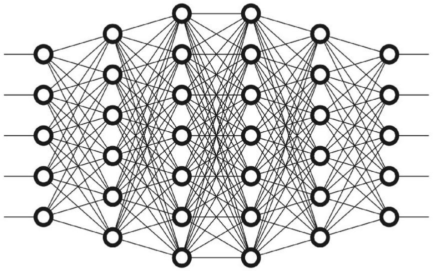
Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

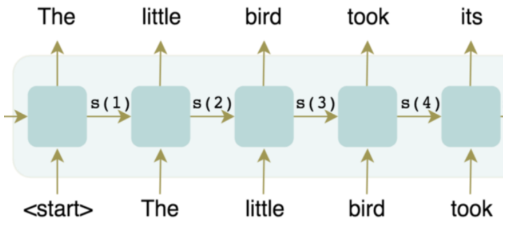
“Shallow” vs Deep Learning

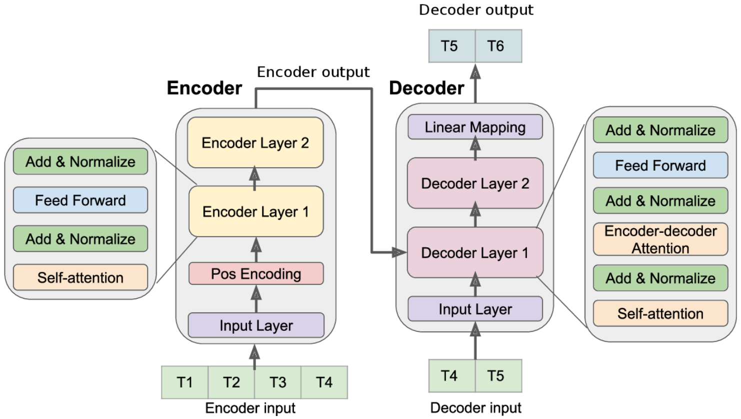
“Shallow” vs. deep language models









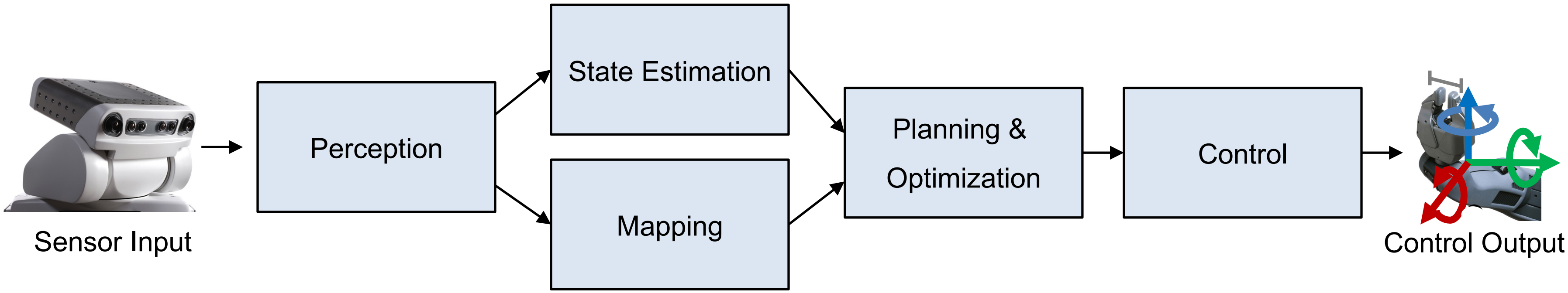


Transformer Models
(Vaswani *et al.*, 2017)

GPT3 large language model
(Brown *et al.*, 2020)

https://www.freecodecamp.org/news/a-history-of-machine-translation-from-the-cold-war-to-deep-learning-f1d335ce8b5/

“Pipelining” vs. “End-to-End Learning”



Sensor Input

Perception

State Estimation

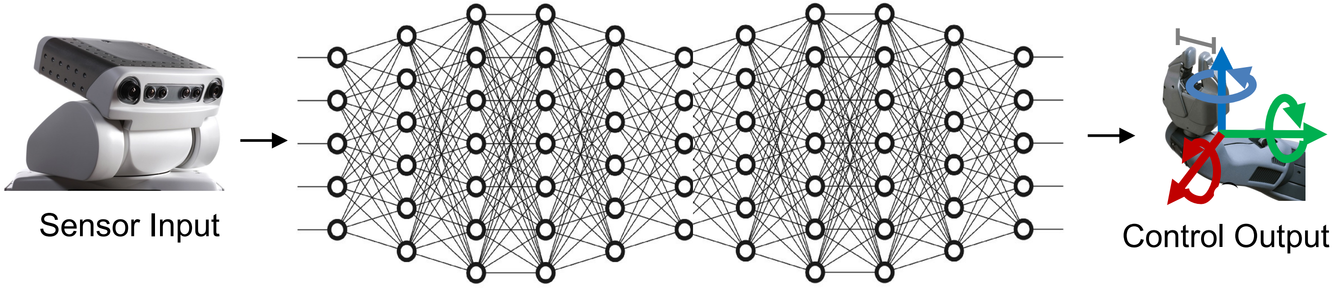
Mapping

Planning & Optimization

Control

Control Output

Hand-engineered pipelines



Sensor Input

Control Output

End-to-end learning (“pixel-to-torque”)

So what *is* Deep Learning?

Ways to think about Deep Learning:

- (Hierarchical) Compositionality
  - Cascade of non-linear transformations
  - Multiple layers of representations
- End-to-End Learning
  - Learning (goal-driven) representations
  - Learning to feature extraction

Benefits of Deep Learning

- (Usually) Better Performance
  - Caveats: given enough data, similar train-test distributions, non-adversarial evaluation, etc, etc.

- New domains without experts
  - RGBD/Lidar
  - Language data
  - Gene-expression data
  - Complex controlling problem
  - Unclear how to hand-engineer

- New abilities emerge with more data and compute
- "Homogenization" of model design

“Expert” intuitions can be misleading

- *“Every time I fire a linguist, the performance of our speech recognition system goes up”*
  - Fred Jelinik, IBM ’98

- *“Because gradient descent is better than you”*
  - *Yann LeCun,* CVPR ‘13

“The Bitter Lesson”

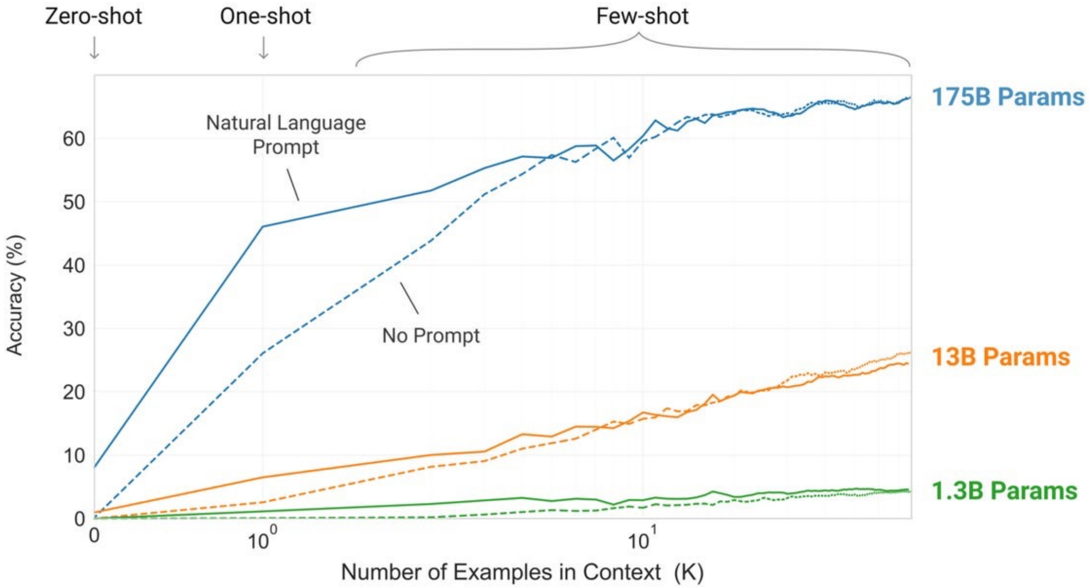
“The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin. The ultimate reason for this is Moore's law, or rather its generalization of continued exponentially falling cost per unit of computation.” (Sutton, 2019)

http://www.incompleteideas.net/IncIdeas/BitterLesson.html

Emergence of new behaviors

Emergence means that the behavior of a system is implicitly induced rather than explicitly constructed. For Deep Learning, emergence is often induced by larger model & more data.

Example: Compared to GPT-2's 1.5B parameter parameter model, GPT-3's 175-billion model permits "prompting", i.e., adapting to a new task simply by describing task. (Try prompting yourself)

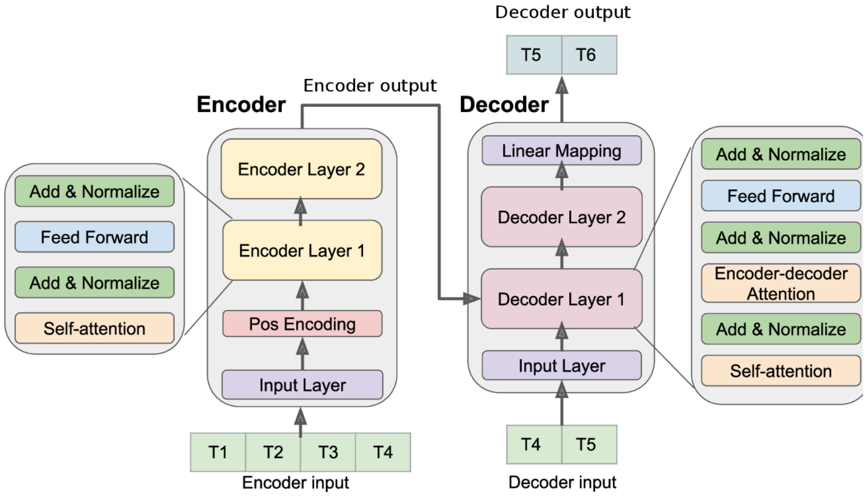


On the Opportunities and Risks of Foundation Models (Bommasani et al., 2021)

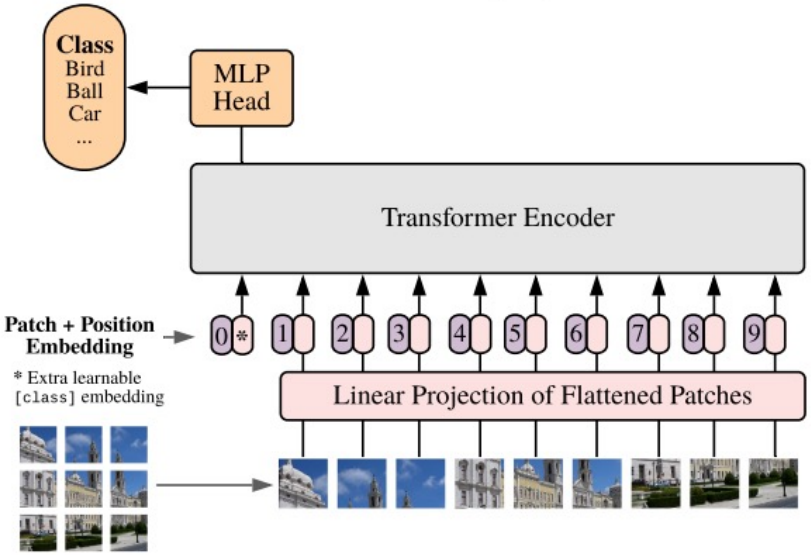
Homogenization of Deep Learning

Homogenization is the **consolidation** of methodologies for building machine learning systems across a wide range of applications.

Example: The Transformer Models (Vaswani *et al.*, 2017)



Transformer Models originally designed for NLP



Class
Bird
Ball
Car
...
MLP Head
Transformer Encoder
Patch + Position Embedding
* Extra learnable [class] embedding
0 *
1
2
3
4
5
6
7
8
9
Linear Projection of Flattened Patches

Almost identical model (Visual Transformers) can be applied to Computer Vision tasks

On the Opportunities and Risks of Foundation Models (Bommasani et al., 2021)

Problems with Deep Learning

- Problem#1: Lack of a formal understanding
  - Non-Convex! Non-Convex! Non-Convex!
    - Depth>=3: most losses non-convex in parameters
  - Worse still, existing intuitions from classical statistical learning theory don’t seem to carry over.
  - Theoretically, we are stumbling in the dark here

- Standard response #1
  - “Yes, but this just means there’s new theory to be constructed”
  - “All interesting learning problems are non-convex”
    - For example, human learning
      - Order matters → wave hands → non-convexity

- Standard response #2
  - “Yes, but it often works!”

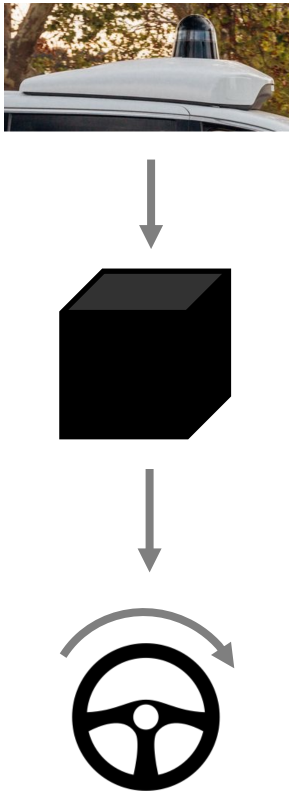
Problems with Deep Learning

- Problem#2: Lack of interpretability
  - Hard to track down what’s failing
  - Pipeline systems have expected performances at each step
  - In end-to-end systems, it’s hard to know why things are not working

Problems with Deep Learning

- Problem#2: Lack of interpretability

Why did the robot do that?



Problems with Deep Learning

- Problem#2: Lack of interpretability
  - Hard to understand why a decision is made
  - In pipeline systems, one can debug by isolating components
  - In end-to-end systems, it's hard to triage an error

- Standard response #1
  - Tricks of the trade: visualize features, add losses at different layers, pre-train to avoid degenerate initializations…
  - "MOOOORE DATA!"
  - "We're working on it"

- Standard response #2
  - "Yes, but it often works!"

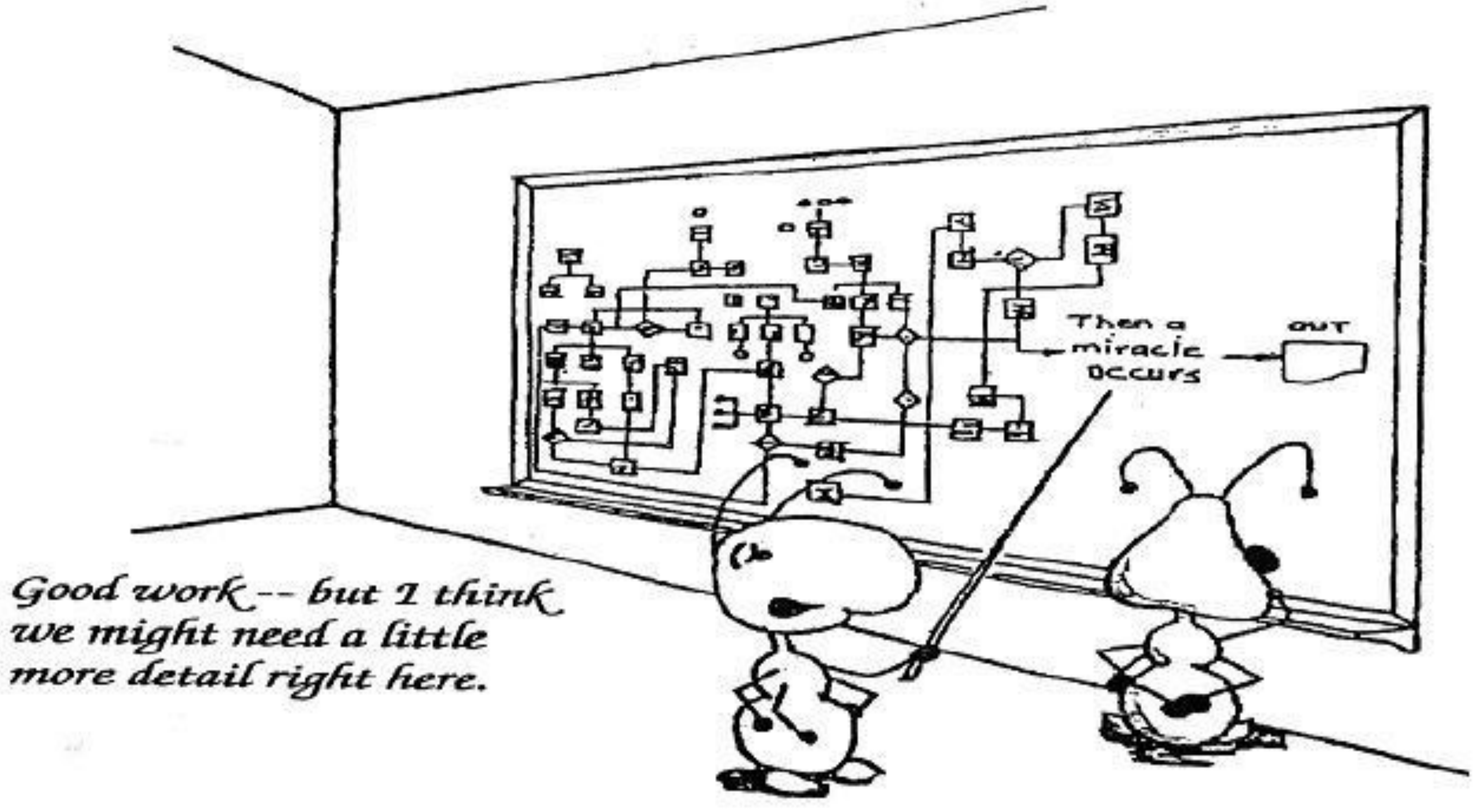
Problems with Deep Learning

- Problem#3: Lack of easy reproducibility
  - Direct consequence of stochasticity & non-convexity
    - different initializations → different local minima
  - Other stochasticity in the training pipeline: parallel data loading, distributed training, numerical precision on GPU…

- Standard response #1
  - It's getting much better
  - Standard toolkits/libraries/frameworks now available
  - PyTorch, TensorFlow, MxNet…

- Standard response #2
  - "Yes, but it often works!"

Yes it works, but how?


Then a miracle occurs
OUT
Good work -- but I think we might need a little more detail right here.

Outline

- What is Deep Learning, the field, about?
  - Highlight of some recent projects from my lab

- What is this class about?
  - What to expect?
  - Logistics

- FAQ

Outline

- What is Deep Learning, the field, about?
  - Highlight of some recent projects from my lab

- What is this class about?
  - What to expect?
  - Logistics

- FAQ

What is this class about?

- Introduction to Deep Learning

- Goal:
  - After finishing this class, you should be ready to get started on your first DL research / engineering project.
    - CNNs (image data)
    - RNNs / Transformers (sequence data)
    - Generative Models (unsupervised learning)
    - Deep Reinforcement Learning (decision making)
    - (Glimpses of) cutting-edge research
  - Work on fun projects with your peers!

- Target Audience:
  - Senior undergrads, MS-(CS, ML, …), and new PhD students

What this class is NOT

- NOT the target audience:
  - Students without sufficient background knowledge (Python, linear algebra, calculus, basic probability & statistics)
  - Advanced grad-students already working in ML/DL areas
  - People looking for an in-depth understanding of a research area that uses deep learning (3D Vision, Large Language Models, Deep RL, etc.).
- NOT the goal:
  - Intro to Machine Learning / Optimization

Caveat

- This is an ADVANCED Machine Learning class
  - This should NOT be your first introduction to ML
  - You will need a formal class; not just self-reading/coursera
  - If you took CS 7641/ISYE 6740/CSE 6740 @GT, you're in the right place
  - If you took an equivalent class elsewhere, see list of topics taught in CS 7641 to be sure.

Prerequisites

- Python Programming
  - Basic knowledge of numerical computations & tools (e.g., numpy)
  - You will write a lot of code!

- Intro Machine Learning
  - Classifiers, regressors, loss functions, MLE, MAP

- Linear Algebra
  - Matrix multiplication, eigenvalues, positive semi-definiteness…

- Calculus
  - Multi-variate gradients, hessians, jacobians…

- Must read (on W3 reading list): Matrix calculus for deep learning
  - https://explained.ai/matrix-calculus/index.html

Your Teaching Team

- Instructor: Prof. Danfei Xu
- Starting Fall 2022 --- I'm new!
- Ph.D. in CS from Stanford (2015-2021)
  - Gap year as a Research Scientist at NVIDIA AI
- Research in Robotics & Machine Learning
  - Some 2D / 3D Vision
- Thesis on Robot Learning
  - On my website, if anyone is interested.

Your Teaching Team

Head TA: Adi Singh

Anshul Ahluwalia

Amogh Dabholkar

Charlie Gunn

Anshul Gupta

Yash Jakhotia

Hoon Lee

Zach Minot

Aaditya Singh

Ningyuan Yang

Office Hour

TA Office Hours:

- Virtual over zoom
- Check course website for OH slots and zoom links
- Start next week

Danfei’s Office Hours:

- In-person (Klaus 1314) or zoom
- **No assignment (PS/HW) questions**
- Lecture content / project ideas / administrative / career advice, …

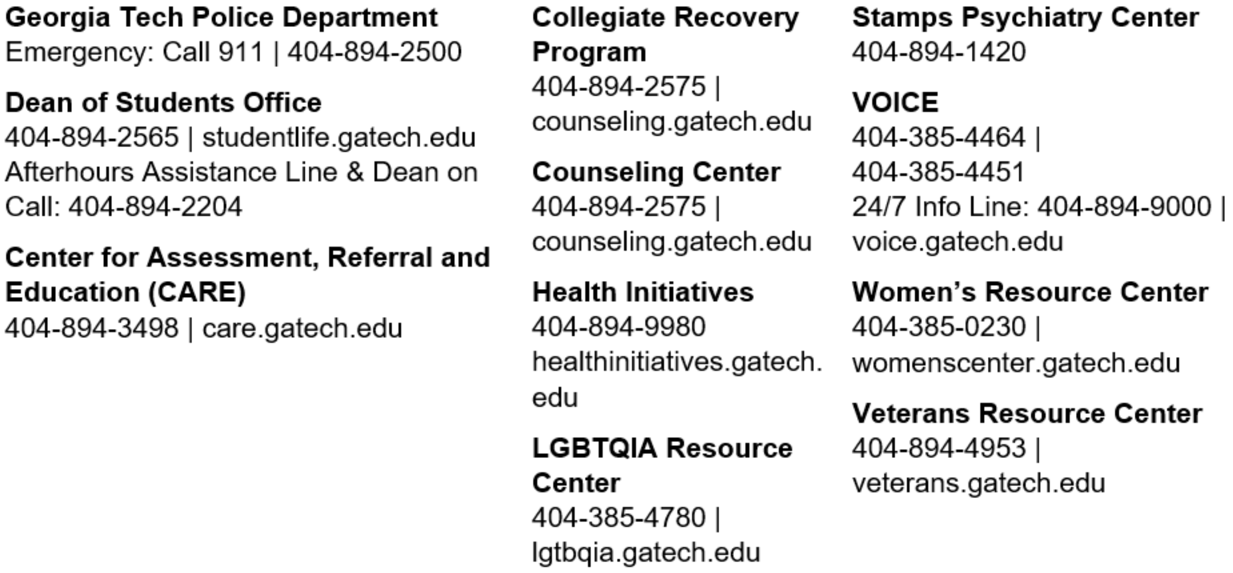
Organization & Deliverables

- 4 problem-sets+homeworks (64%)
  - Mix of theory (PS) and implementation (HW)
  - First one goes out next week
    - Start early, Start early, Start early, Start early, Start early, Start early
- Course project (36%)
  - Projects done in groups of 2-4
  - You need a good reason to do a solo project.
  - Proposal (1%), Milestone Report (5%), Final Report (25%), Poster Session (5%)
  - **Find a team ASAP! Talk to people, use Piazza “find a teammate” post.**
  - Ideas & scope: http://cs231n.stanford.edu/reports.html
- (Bonus) Class Participation (1%)
  - Top (endorsed) contributors on Piazza

Plenty of “buffer” built in

- Grace period
  - 2 days grace period
    - Intended for *checking* submission NOT to replace due date
    - No need to ask for grace, no penalty for turning it in within grace period
    - Can NOT use for PS0
  - After grace period, you get a 0 (no excuses except medical)
    - Send all medical requests to dean of students (https://studentlife.gatech.edu/)
    - Form: https://gatech-advocate.symplicity.com/care_report/index.php/pid224342?
  - **DO NOT SEND US ANY MEDICAL INFORMATION!** We do not need any details, just a confirmation from dean of students

GT Resources for Mental Health



Georgia Tech Police Department
Emergency: Call 911 | 404-894-2500

Dean of Students Office
404-894-2565 | studentlife.gatech.edu
Afterhours Assistance Line & Dean on Call: 404-894-2204

Center for Assessment, Referral and Education (CARE)
404-894-3498 | care.gatech.edu

Collegiate Recovery Program
404-894-2575 | counseling.gatech.edu

Counseling Center
404-894-2575 | counseling.gatech.edu

Health Initiatives
404-894-9980
healthinitiatives.gatech.edu

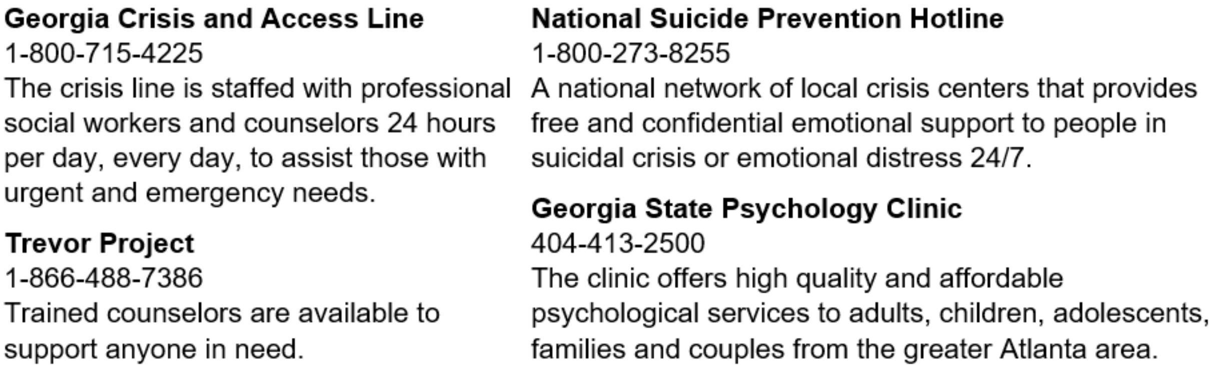
LGBTQIA Resource Center
404-385-4780 | lgtbqia.gatech.edu

Stamps Psychiatry Center
404-894-1420

VOICE
404-385-4464 | 404-385-4451
24/7 Info Line: 404-894-9000 | voice.gatech.edu

Women's Resource Center
404-385-0230 | womenscenter.gatech.edu

Veterans Resource Center
404-894-4953 | veterans.gatech.edu



Georgia Crisis and Access Line
1-800-715-4225
The crisis line is staffed with professional social workers and counselors 24 hours per day, every day, to assist those with urgent and emergency needs.

Trevor Project
1-866-488-7386
Trained counselors are available to support anyone in need.

National Suicide Prevention Hotline
1-800-273-8255
A national network of local crisis centers that provides free and confidential emotional support to people in suicidal crisis or emotional distress 24/7.

Georgia State Psychology Clinic
404-413-2500
The clinic offers high quality and affordable psychological services to adults, children, adolescents, families and couples from the greater Atlanta area.

PS0

- Will be out today. Due Tuesday Aug 30th
  - Will be available on class webpage
  - If not registered yet (on waitlist), see webpage FAQ for form to request gradescope access

- Grading
  - Not counted towards your final grade, but required
  - <=75% means that you might not be prepared for the class
  - If you submit after Thursday, we will not grade before registration ends

- Topics
  - PS: probability, calculus, convexity
  - HW: Python + Numpy

Project

- Goal
  - Chance to try Deep Learning in practice
  - Encouraged to apply to your research (computer vision, NLP, robotics,…)
  - Must be done this semester.
  - Can combine with other classes, but **separate thrust**
    - get permission from both instructors; delineate different parts
  - 2-4 members (outside of this requires approval)

Computing

- Major bottleneck
  - GPUs

- Options
  - Your own / group / advisor’s resources

  - Google Colab
    - jupyter-notebook + free GPU instance

  - Google Cloud credits (details TBA)
    - Tutorial on setting up gloud: https://github.com/cs231n/gcloud

4644 vs 7643

- Level differentiation
- Separate grade curves calculation
  - As a result, 4644 and 7643 may have different letter grade cut-offs.

Outline

- What is Deep Learning, the field, about?
  - Highlight of some recent projects from my lab

- What is this class about?
  - What to expect?
  - Logistics

- FAQ

Waitlist / Audit / Sit in

- Waitlist
  - Class is full. Capacity change unlikely
  - Do PS0 **NOW**. Come to first few classes.
  - Hope people drop.
- “I need this class to graduate”
  - Talk to your degree program advisor. They control the process of making sure you have options to graduate on time.
- Audit or Pass/Fail
  - No.

What is the re-grading policy?

- Homework assignments
  - **Within 1 week** of receiving grades: see the TAs

What is the collaboration policy?

- Collaboration
  - Only on HW (coding) and project.
  - You may discuss the questions
  - Each student writes their own answers
  - Write on your homework anyone with whom you collaborate
  - Each student must write their own code for the programming part

- Zero tolerance on plagiarism
  - Neither ethical nor in your best interest
  - Always credit your sources
  - Don’t cheat. We will find out.

How do I get in touch?

- Primary means of communication -- Piazza
  - No direct emails to Instructor **unless private information**
  - Instructor/TAs can provide answers to everyone on forum
  - Class participation credit for answering questions!
  - No posting answers. We will monitor.
  - Stay respectful and professional

Questions?