CS 4803 / 7643: Deep Learning

Website: https://www.cc.gatech.edu/classes/AY2021/cs7643_spring/

Piazza: http://piazza.com/gatech/spring2021/cs48037643a
(code: DL2021)

Canvas: https://gatech.instructure.com/courses/172518 (4803)
https://gatech.instructure.com/courses/172536 (7643)

Gradescope: https://www.gradescope.com/courses/228228 (4803)
https://www.gradescope.com/courses/229744 (7643)

Zsolt Kira

School of Interactive Computing
Georgia Tech

Elephant in the room

- These times are filled with change and uncertainty
- Hope everyone is staying safe and healthy.
- Let’s make the best of it.

Are you in the right place?

- This is CS 4803(DL) / CS 7643
  - “On campus” class

- This is NOT CS 7643-O01/OAN/Q/R
  - Online class for OMSCS program

Spring 21 Delivery Format

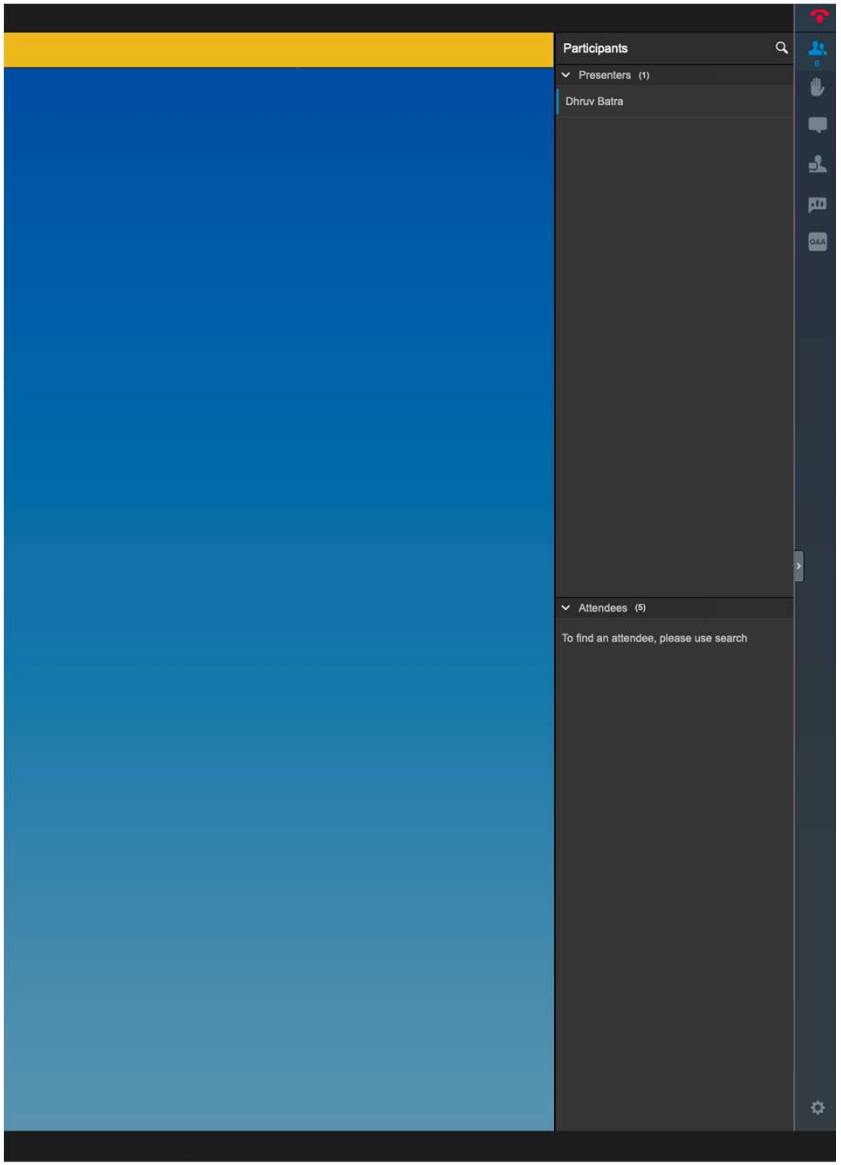
- Remote
  - No in-person interaction
  - Lectures, office hours, HW/project submissions online
    - No exam
- Sync
  - There is a scheduled “live” lecture time
- Recording
  - Lectures are recorded and available for viewing
  - We STRONGLY encourage you to attend the lectures
- **Remember: Content is free online.**
  - **You are here for the interaction and the insight.**

OMSCS Videos

- We will provide dropbox with OMSCS videos
- Nicely produced, clear lectures that allow you to go at your own pace
  - Note slides for this semester will be different from past as a result
- Currently these will be ***supplements*** for those that need it
  - Synchronized lectures will cover the same materials but with a little bit more depth (e.g. a few additional mathematical details/intuitions)
  - These will also be recorded
  - Depending on interest and time available, may experiment with adding research topics
- **Remember: Content is free online.**
  - **You are here for the interaction and the insight.**

How to interact

- Questions
  - Q&A
    - if one-off question
  - Chat
    - for back and forth
  - Raise hand
    - TA will elevate you to participant
  - We'll explicitly stop and take questions periodically
    - But feel free to ask in between
- BlueJeans Event
  - ~10 sec lag



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 are we here to discuss?

Some of the most exciting developments in

Machine Learning, Vision, NLP, Speech, Robotics & AI in general

in the last decade!

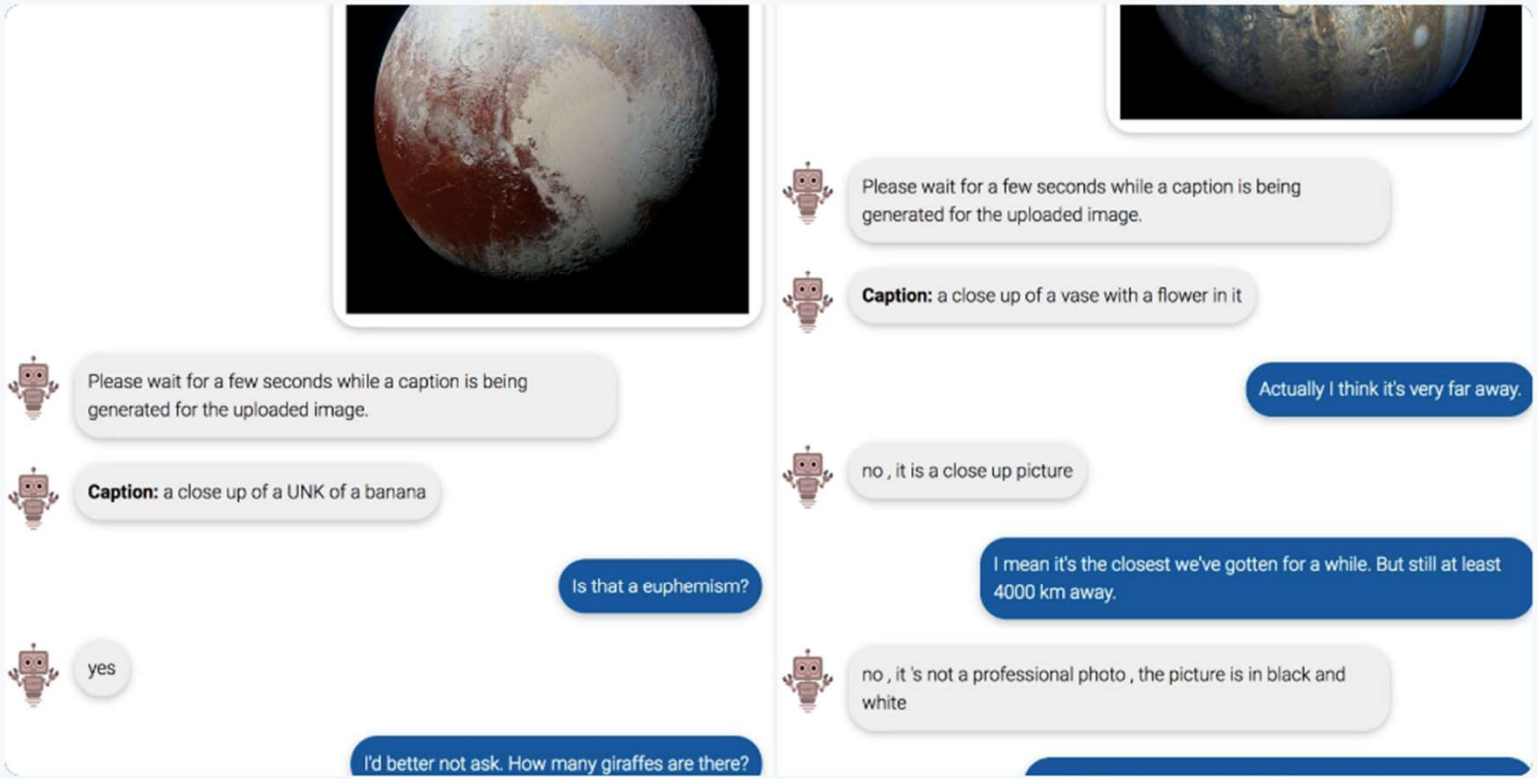
Demo time

vqa.cloudcv.org.

demo.visualdialog.org

Janelle Shane @JanelleCShane · Jun 24

One fun thing I discovered about Visual Chatbot.
It learned from answers that humans gave, and apparently nobody ever asked "how many giraffes are there?" when the answer was zero.
demo.visualdialog.org



24 159 515

Concepts

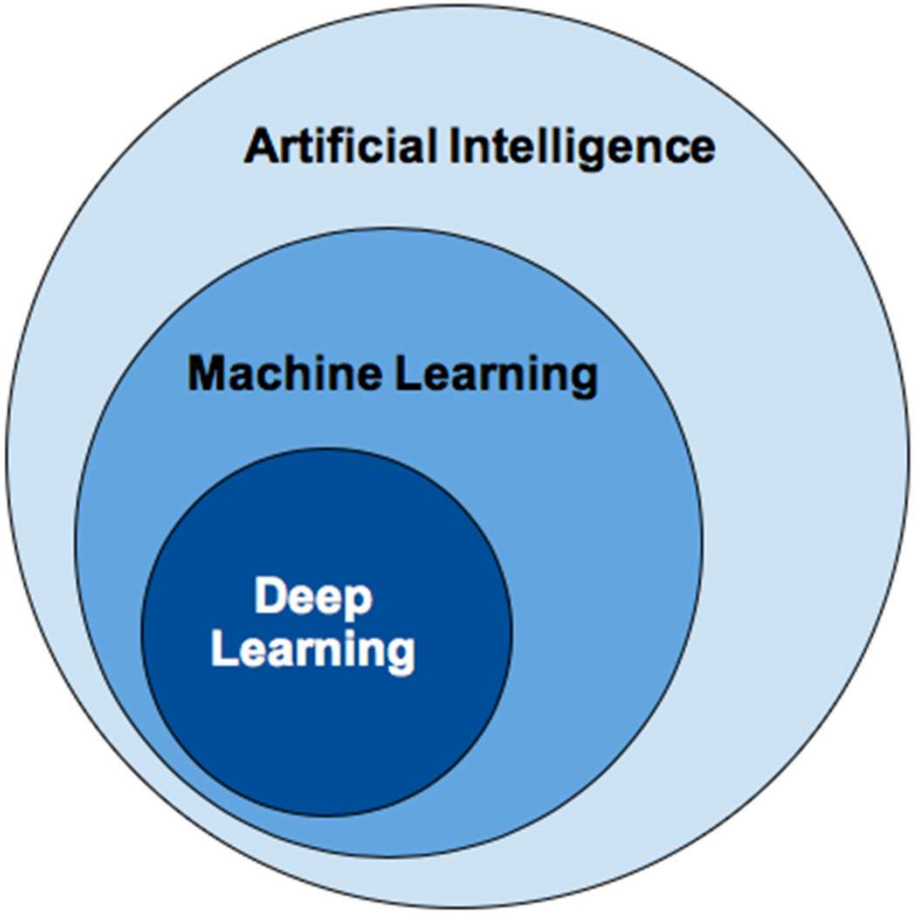


Image Credit: https://www.sumologic.com/blog/machine-learning-deep-learning/

What is (general) intelligence?

- Boring textbook answer

The ability to acquire and apply knowledge and skills

  – Dictionary

- A favorite

The ability to navigate in problem space

  – Siddhartha Mukherjee, Columbia

What is artificial intelligence?

- Boring textbook answer

Intelligence demonstrated by machines

  – Wikipedia

- A favorite:

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

  – Andrew Moore, CMU

What is machine learning?

- A favorite

Study of algorithms that
improve their performance (P)
at some task (T)
with experience (E)

  – Tom Mitchell, CMU

So what *is* Deep (Machine) Learning?

- Representation Learning
- Neural Networks
- Deep Unsupervised/Reinforcement/Structured/ <insert-qualifier-here> Learning
- Simply: Deep Learning

So what *is* Deep (Machine) Learning?

- A few different ideas:

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

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

- Distributed Representations
  - No single neuron “encodes” everything
  - Groups of neurons work together

Hierarchical Compositionality

VISION

pixels → edge → texton → motif → part → object

SPEECH

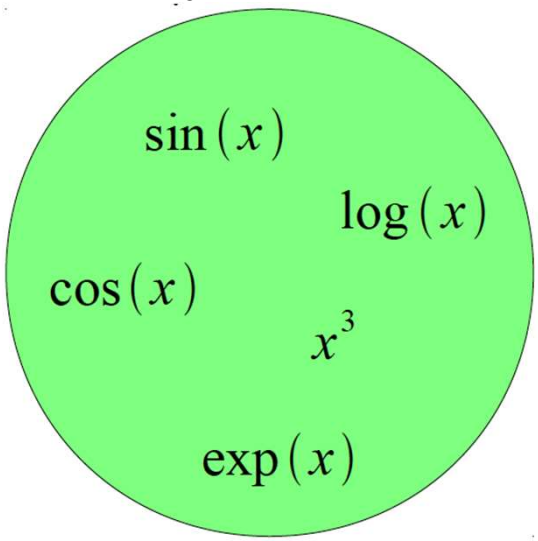
sample → spectral band → formant → motif → phone → word

NLP

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

Building A Complicated Function

Given a library of simple functions



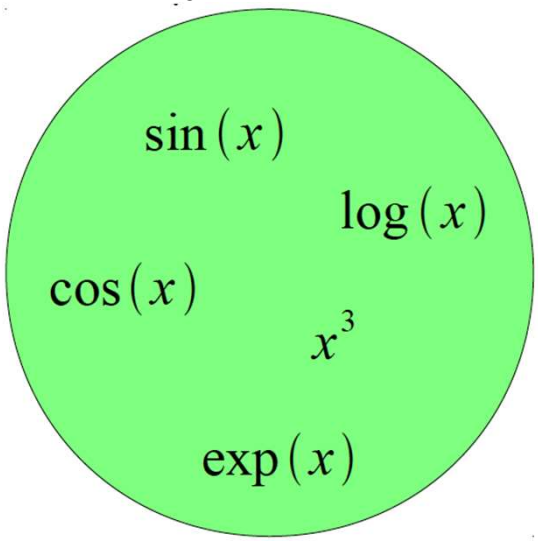
Compose into a

complicate function

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

Building A Complicated Function

Given a library of simple functions



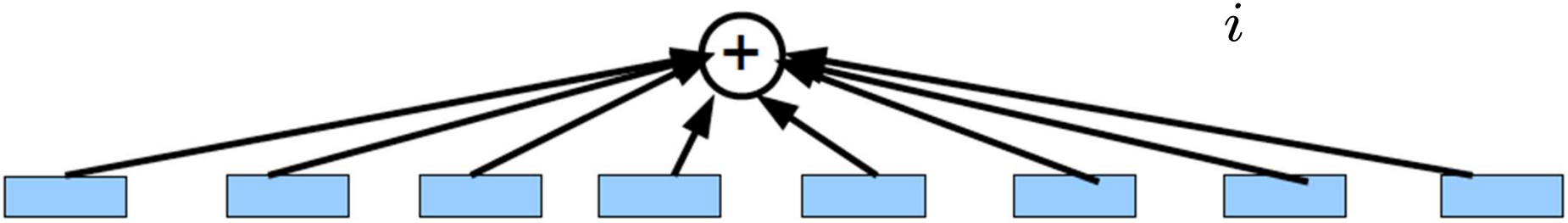
Compose into a

complicate function

Idea 1: Linear Combinations

- Boosting
- Kernels
- …

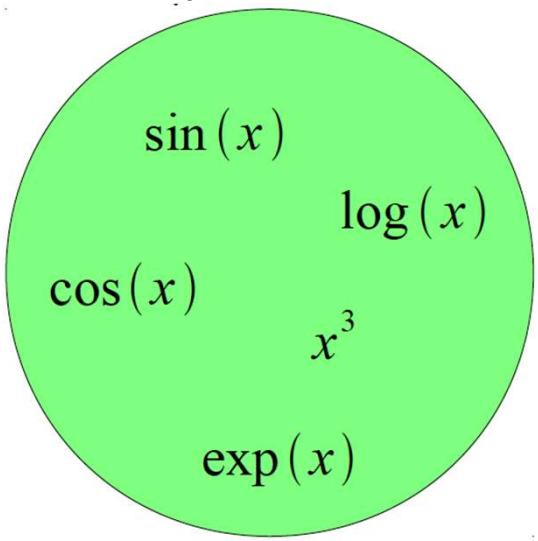
$$f(x) = \sum_i \alpha_i g_i(x)$$



Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

Building A Complicated Function

Given a library of simple functions



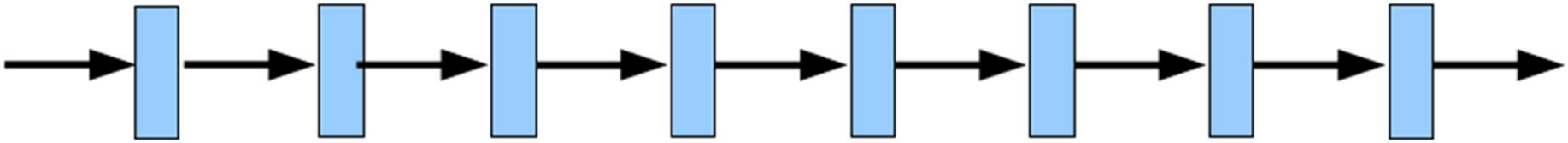
Compose into a

complicate function

Idea 2: Compositions

- Deep Learning
- Grammar models
- Scattering transforms…

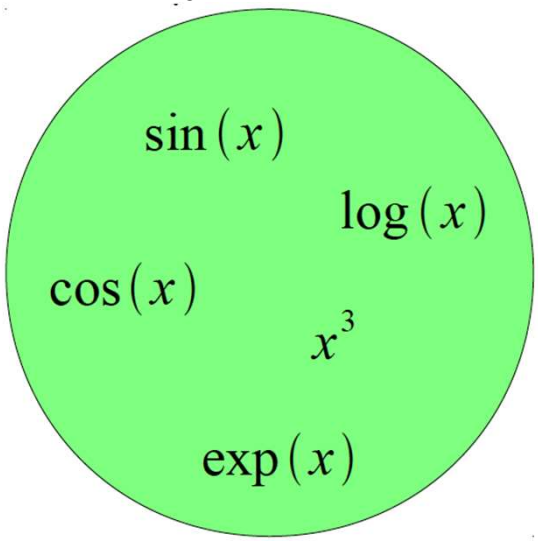
$$f(x) = g_1(g_2(\ldots(g_n(x)\ldots))$$



Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

Building A Complicated Function

Given a library of simple functions



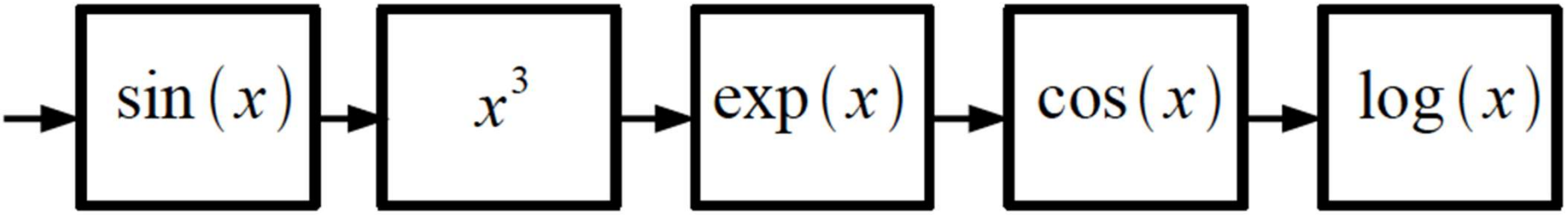
Compose into a

complicate function

Idea 2: Compositions

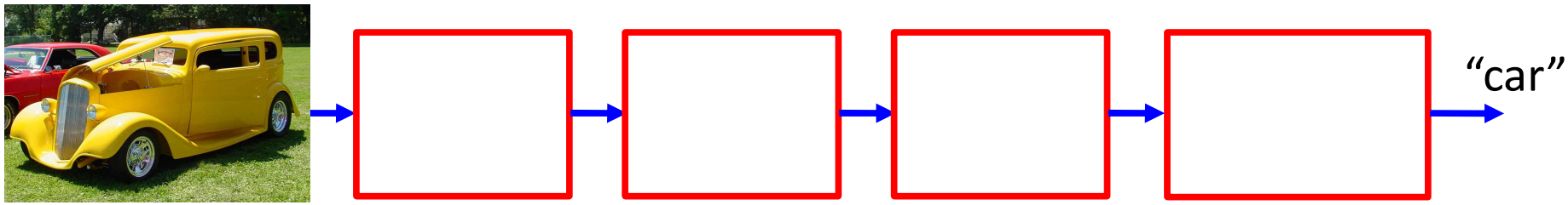
- Deep Learning
- Grammar models
- Scattering transforms…

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



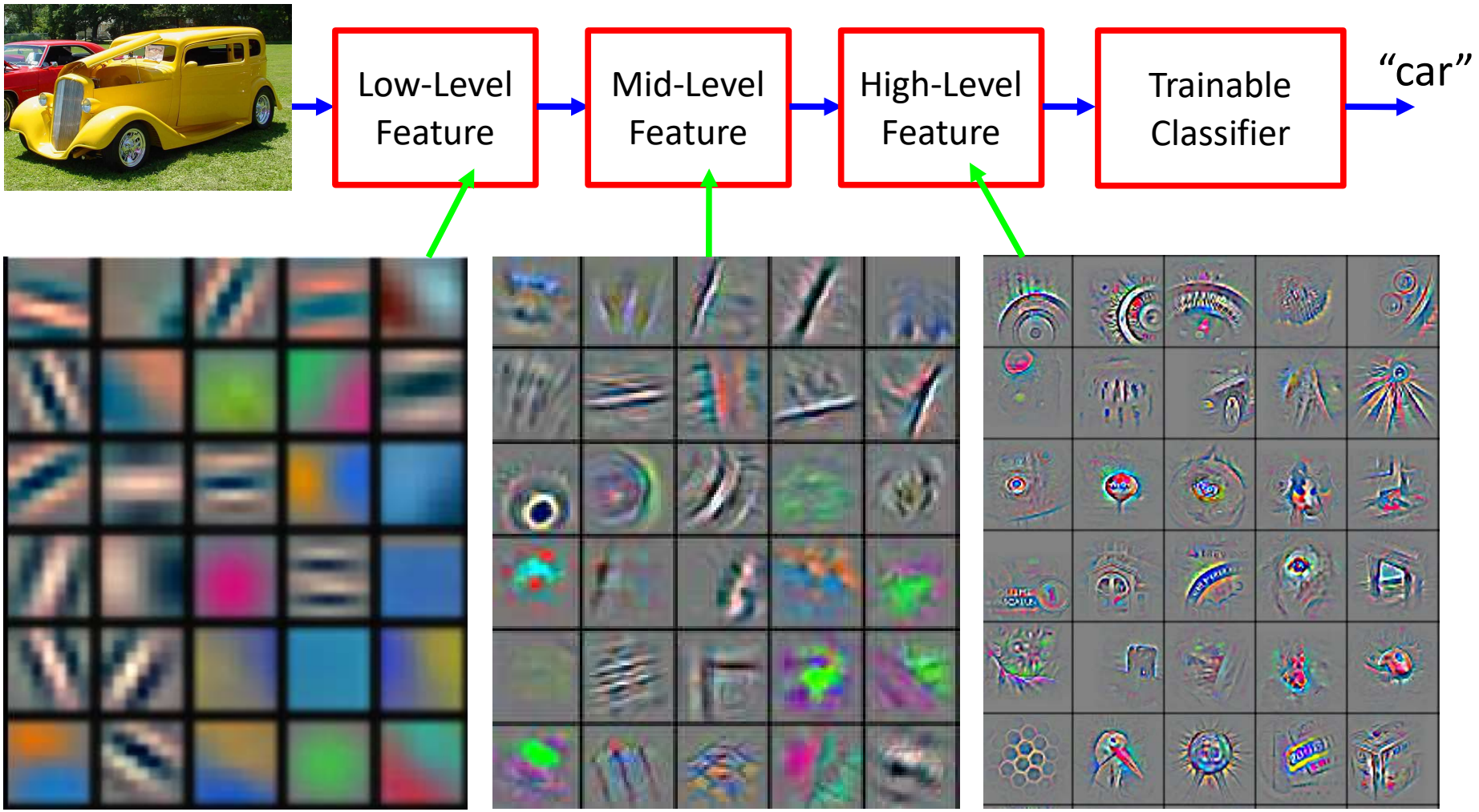
Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

Deep Learning = Hierarchical Compositionality



Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

Deep Learning = Hierarchical Compositionality



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

So what *is* Deep (Machine) Learning?

- A few different ideas:

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

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

- Distributed Representations
  - No single neuron “encodes” everything
  - Groups of neurons work together

Traditional Machine Learning

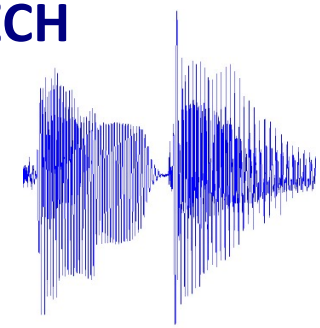
VISION

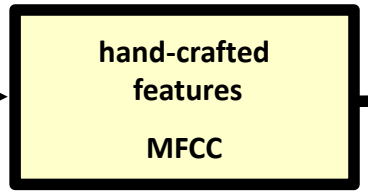
fixed

learned

"car"

SPEECH





fixed

learned

\ˈd ē p\

NLP

This burrito place

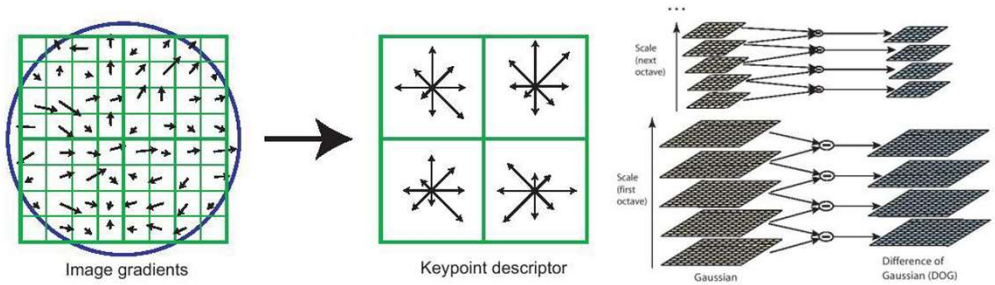
is yummy and fun!

fixed

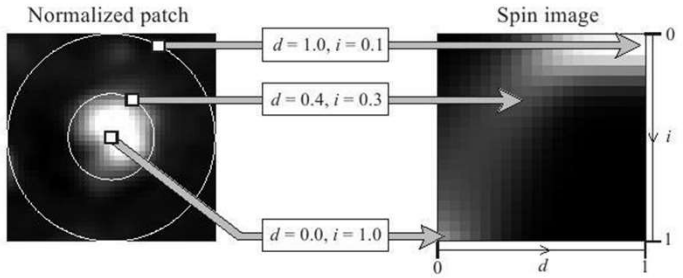
learned

"+"

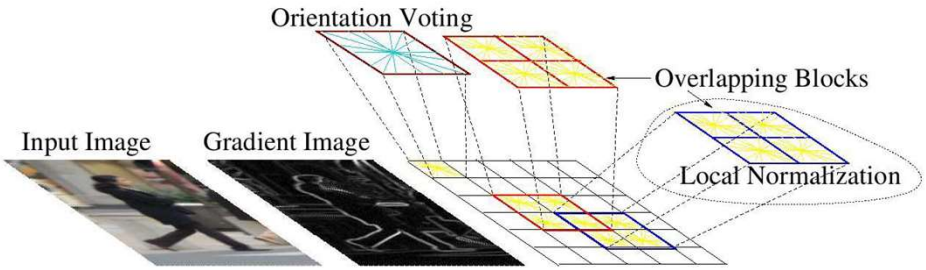
Feature Engineering



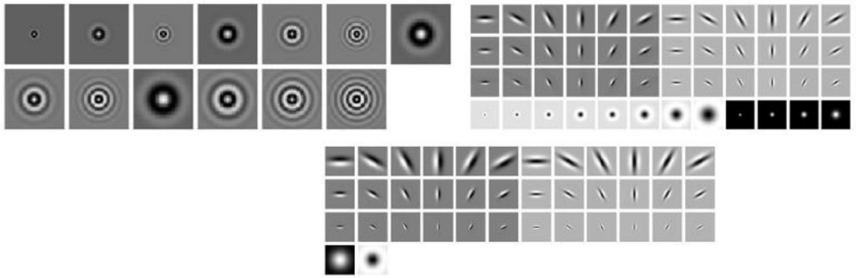
SIFT



Spin Images



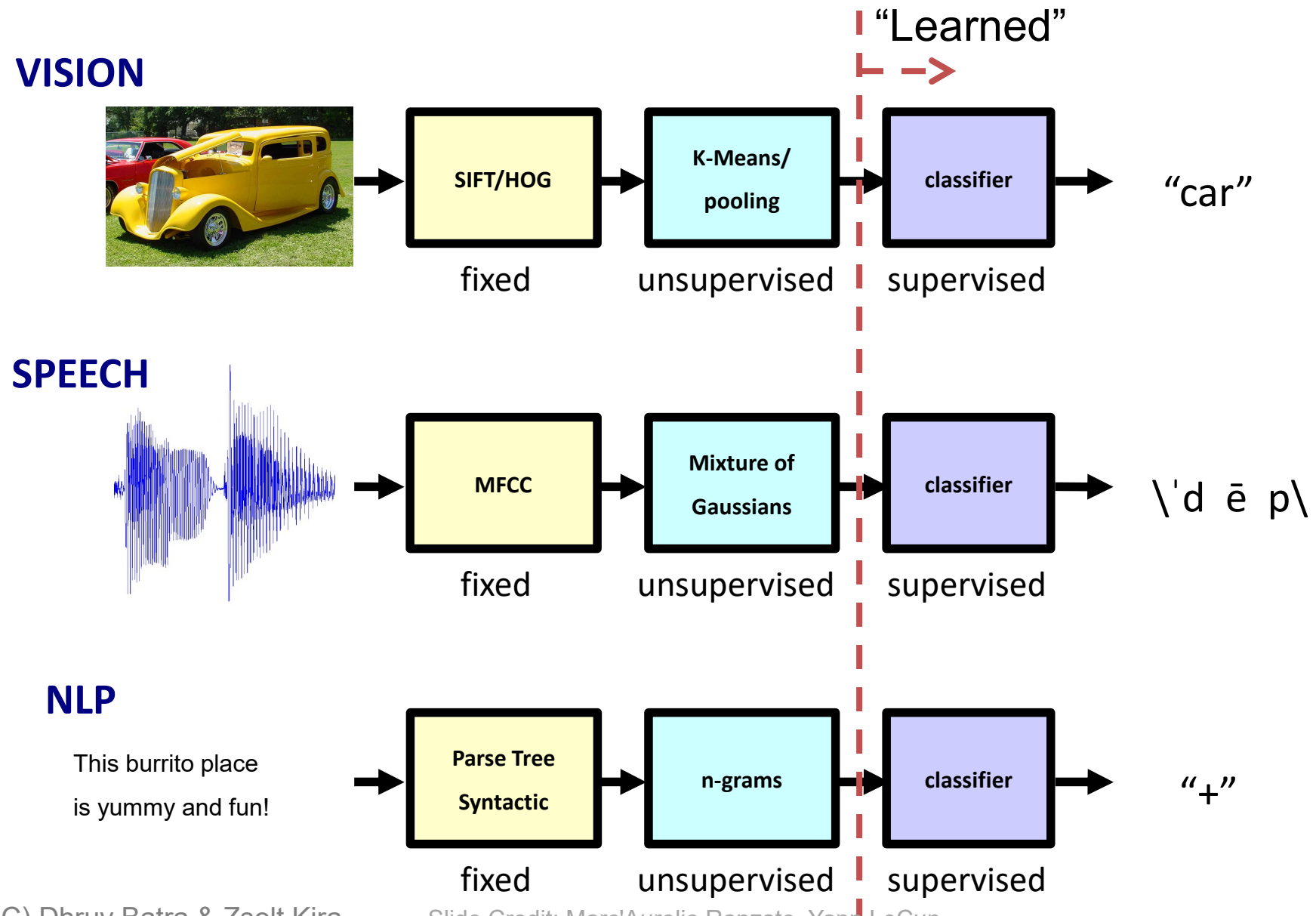
HoG



Textons

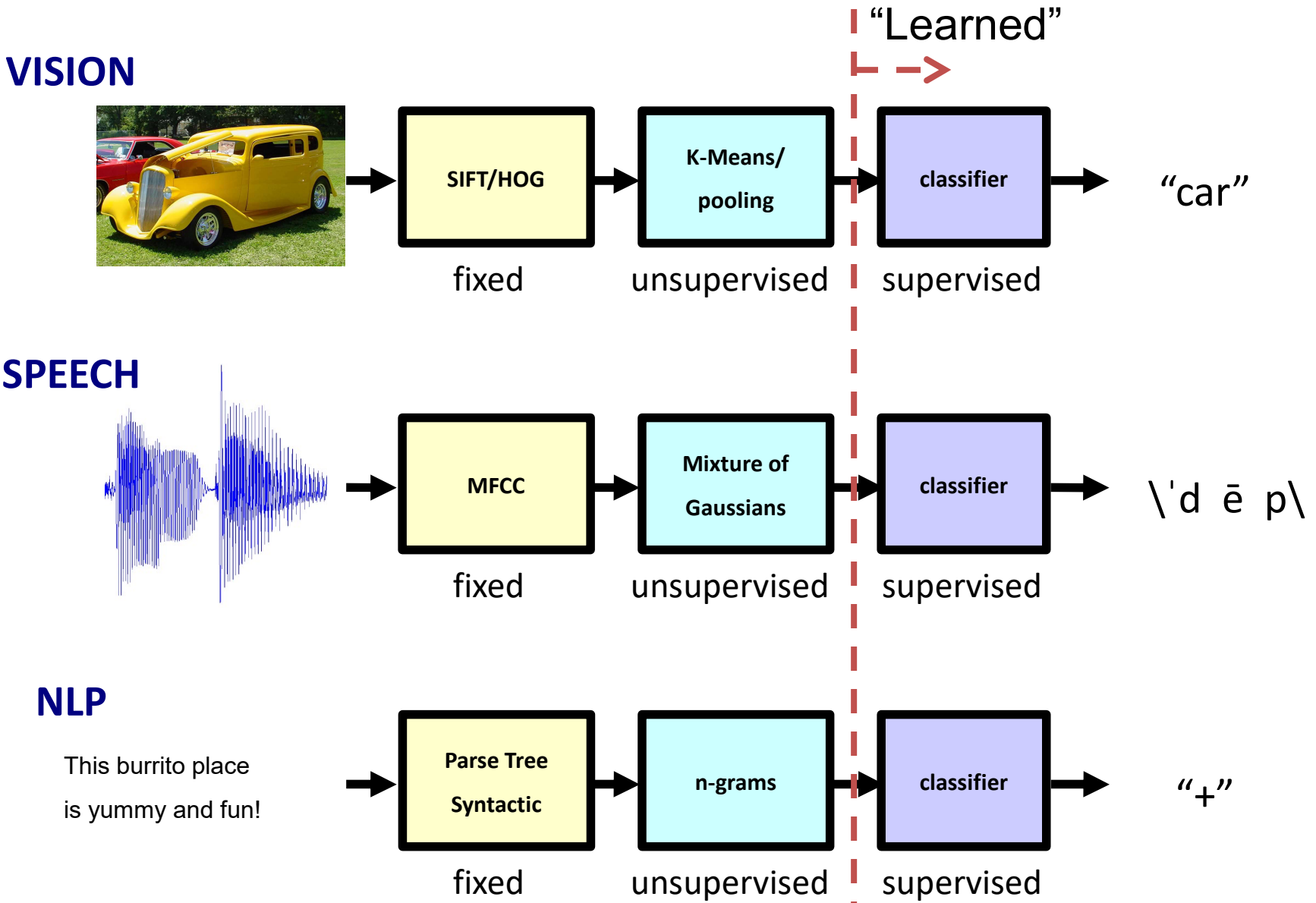
and many many more….

Traditional Machine Learning (more accurately)



Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

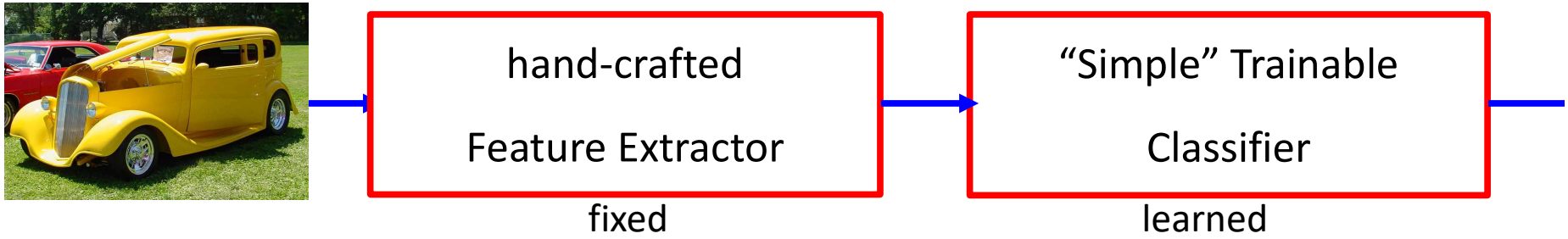
Deep Learning = End-to-End Learning



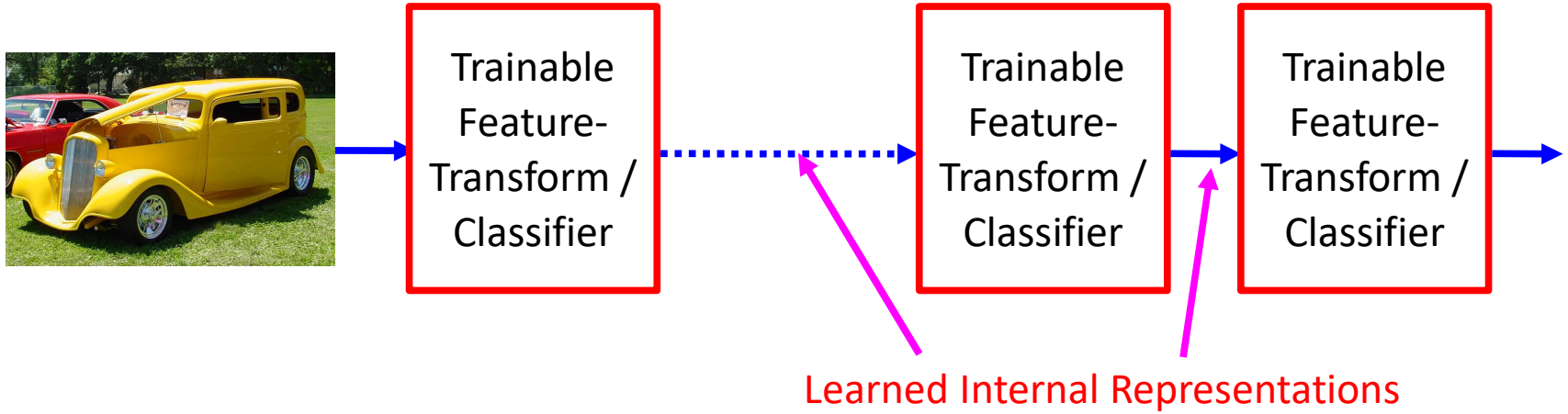
Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

“Shallow” vs Deep Learning

- “Shallow” models



- Deep models



Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

So what *is* Deep (Machine) Learning?

- A few different ideas:

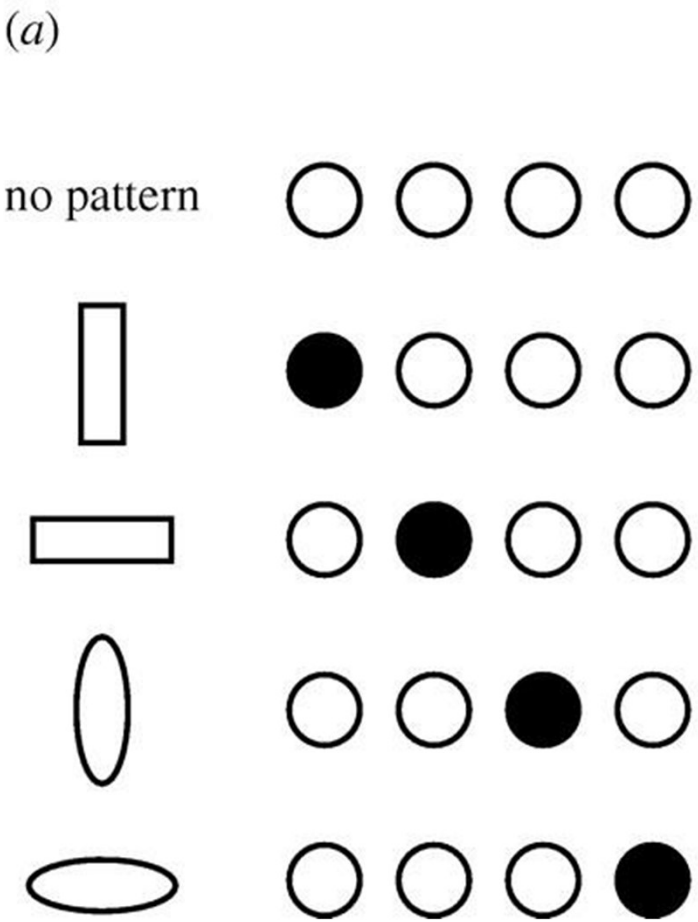
- (Hierarchical) Compositionality
  - Cascade of non-linear transformations
  - Multiple layers of representations

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

- Distributed Representations
  - No single neuron “encodes” everything
  - Groups of neurons work together

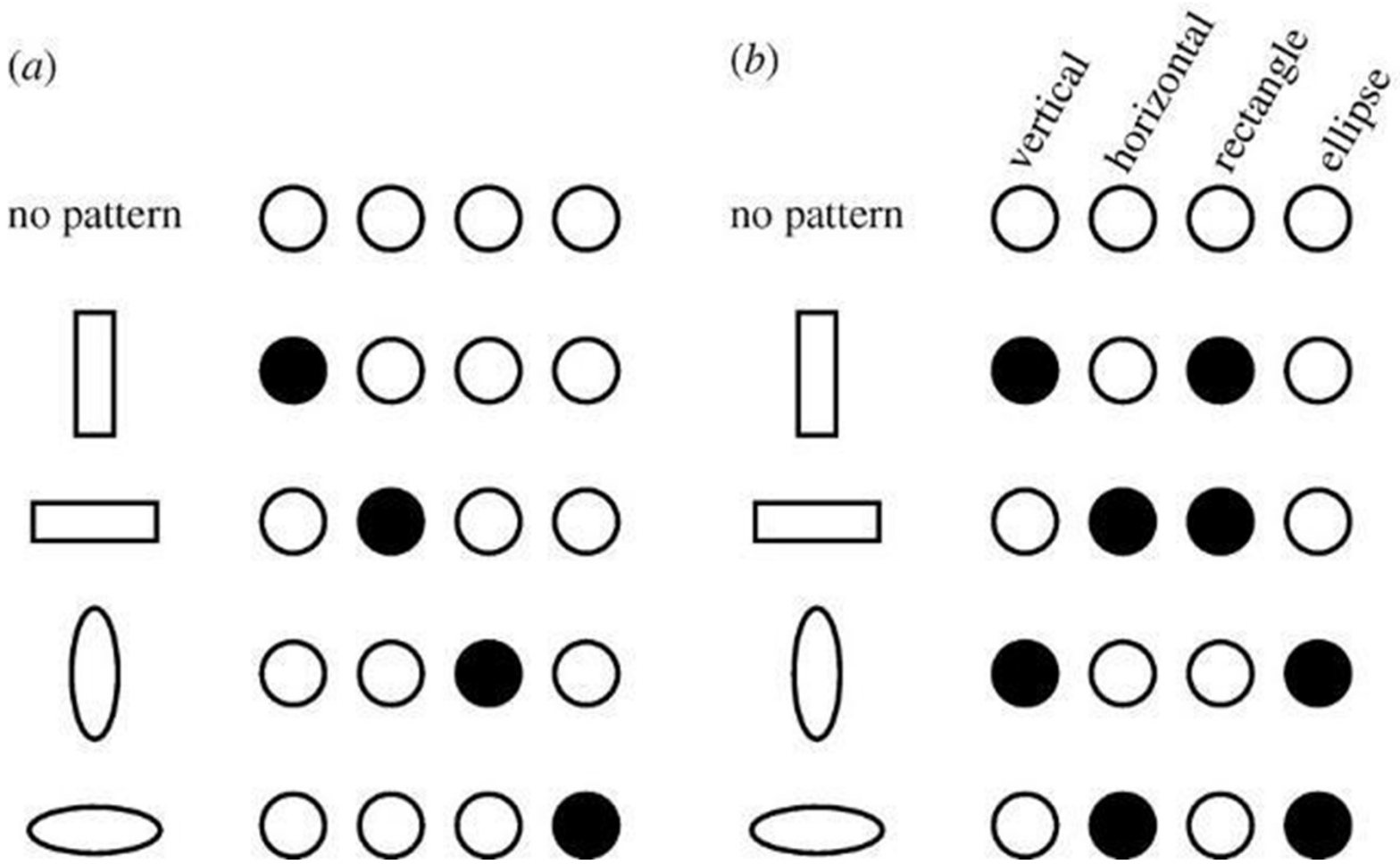
Distributed Representations Toy Example

- Local vs Distributed

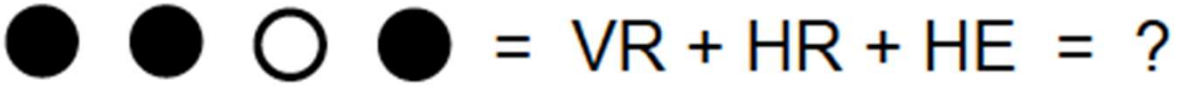


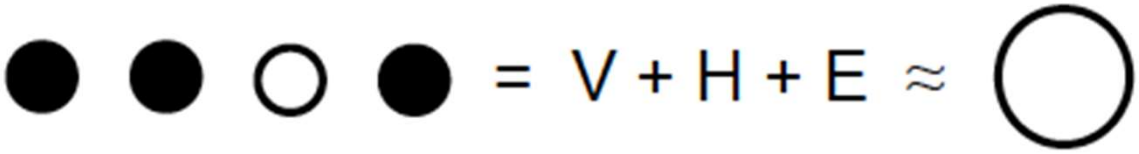
Distributed Representations Toy Example

- Can we interpret each dimension?



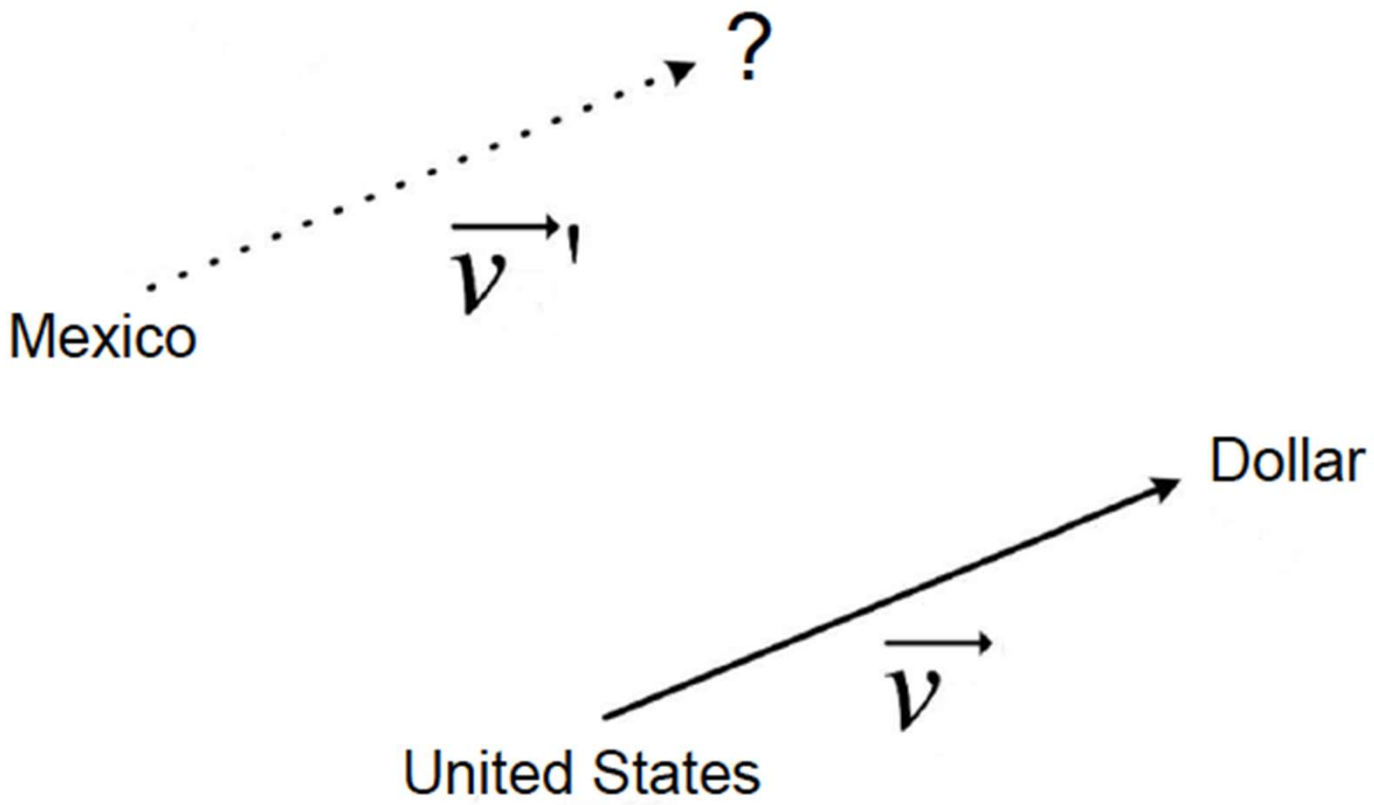
Power of distributed representations!

Local 

Distributed 

Power of distributed representations!

- United States:Dollar :: Mexico:?



ThisPlusThat.me

Image Credit: http://blog.kiradatascience.com/blog/thisplusthat_a_search_engine_that_lets_you_add_words_as_vectors.html

So what *is* Deep (Machine) Learning?

- A few different ideas:

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

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

- Distributed Representations
  - No single neuron “encodes” everything
  - Groups of neurons work together

Benefits of Deep/Representation Learning

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

- New domains without “experts”
  - RGBD/Lidar
  - Multi-spectral data
  - Gene-expression data
  - Unclear how to hand-engineer

“Expert” intuitions can be misleading

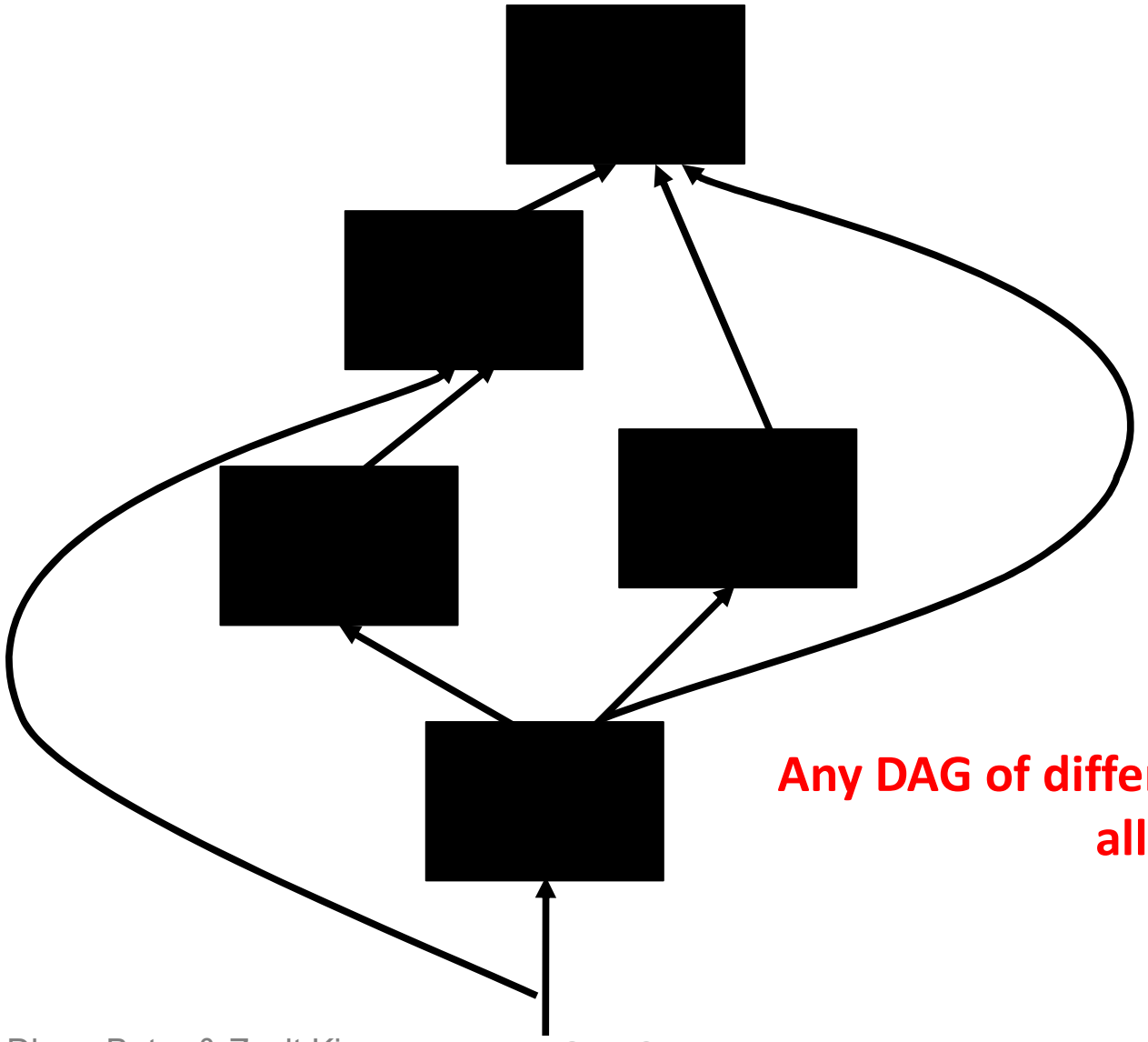
- *“Every time I fire a linguist, the performance of our speech recognition system goes up”*
  - Fred Jelinik, IBM ’98

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

Benefits of Deep/Representation Learning

- Modularity!
- Plug and play architectures!

Differentiable Computation Graph



Any DAG of differentialble modules is allowed!

Slide Credit: Marc'Aurelio Ranzato

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

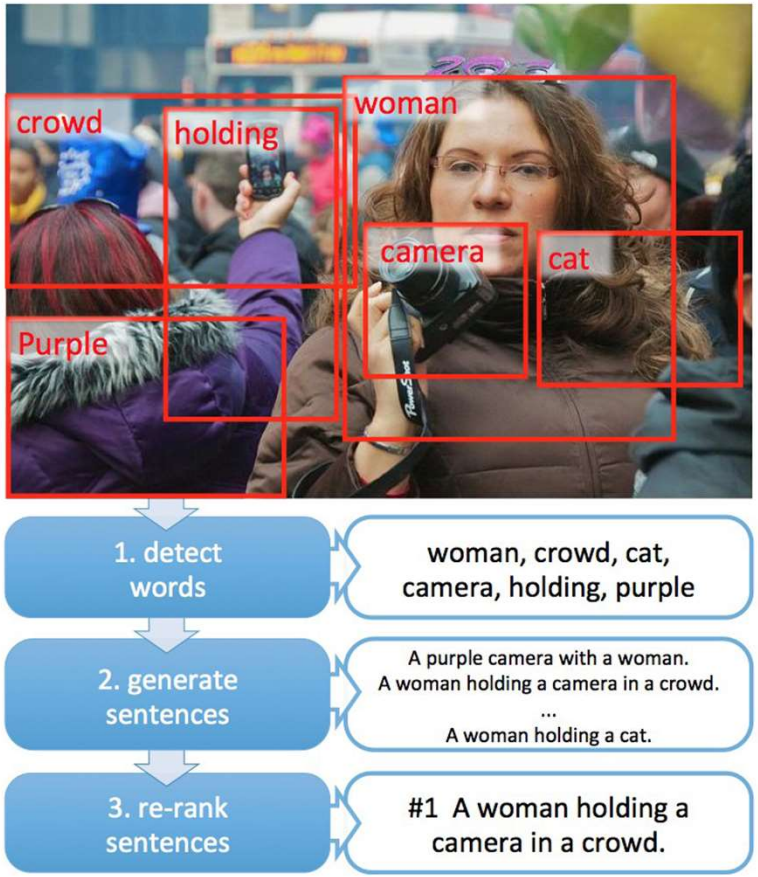
- Standard response #2
  - "Yes, but it often works!"

Problems with Deep Learning

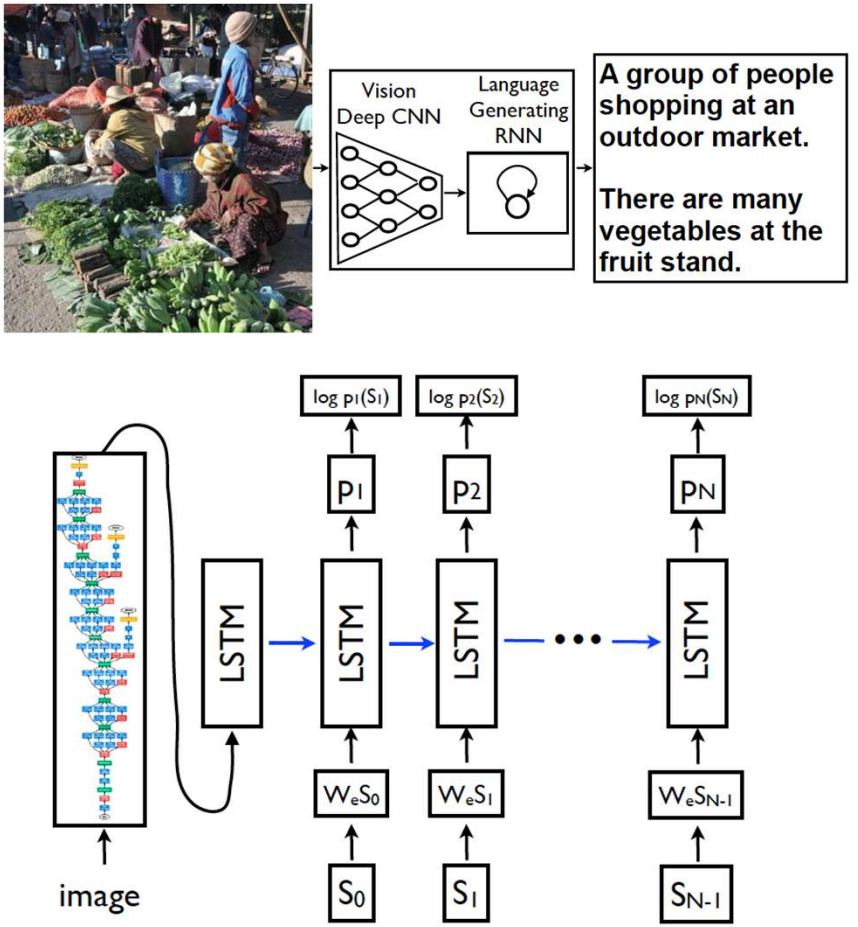
- Problem#2: Lack of interpretability
  - Hard to track down what's failing
  - Pipeline systems have expected performances at each step
  - In end-to-end systems, it's hard to know why things are not working

Problems with Deep Learning

- Problem#2: Lack of interpretability



[Fang et al. CVPR15]



[Vinyals et al. CVPR15]

Pipeline

End-to-End

Problems with Deep Learning

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

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

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

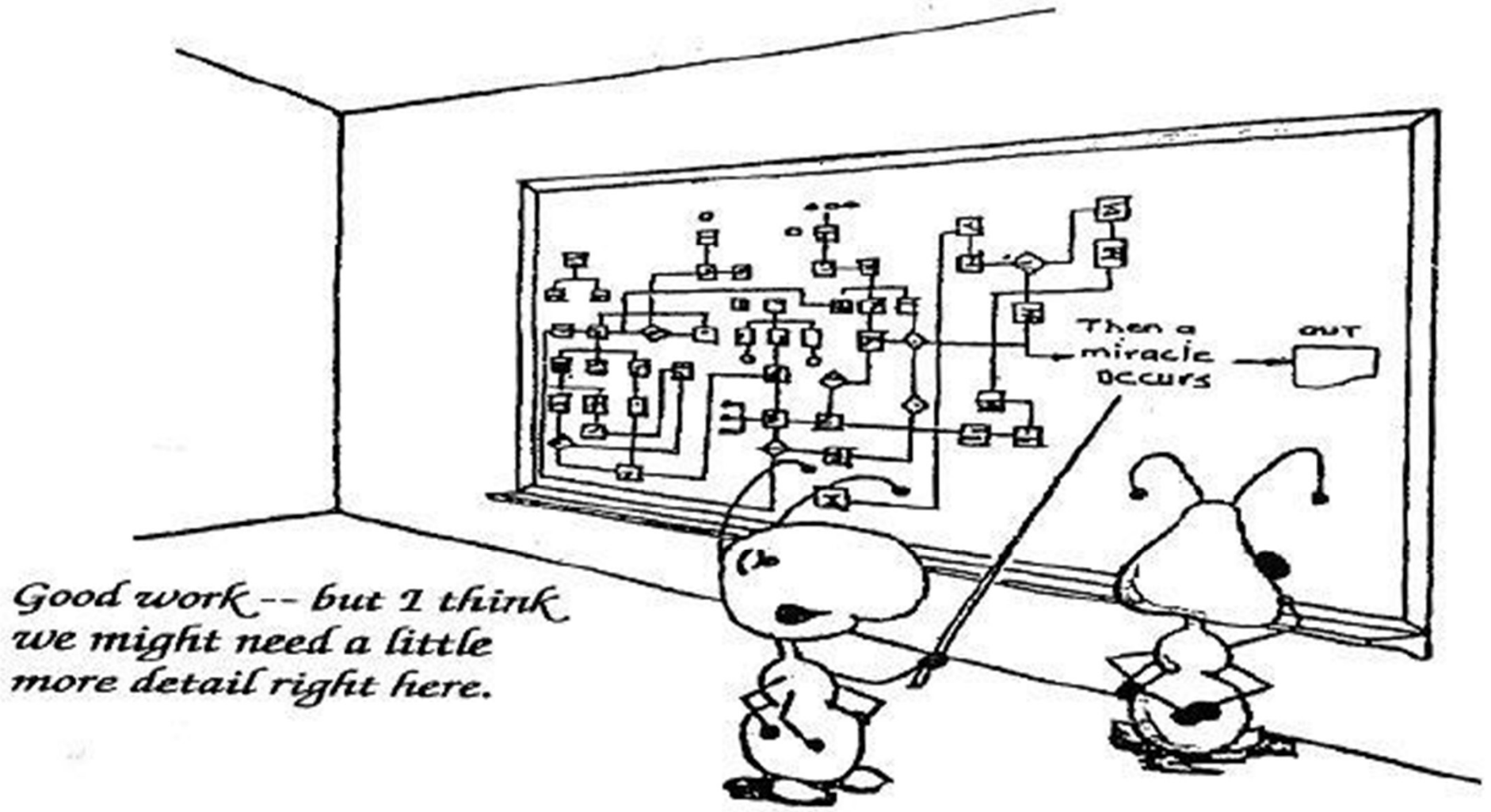
Problems with Deep Learning

- Problem#3: Lack of easy reproducibility
  - Direct consequence of stochasticity & non-convexity
    - different initializations → different local minima

- 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?



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 project.
    - CNNs
    - RNNs / Transformers
    - Deep Reinforcement Learning
    - Generative Models (VAEs, GANs)

- Target Audience:
  - Senior undergrads, MS-ML, and new PhD students

What this class is NOT

- NOT the target audience:
  - Advanced grad-students already working in ML/DL areas
  - People looking to understand latest and greatest cutting-edge research (e.g. GANs, AlphaGo, etc)
  - Undergraduate/Masters students looking to graduate with a DL class on their resume.
- NOT the goal:
  - Teaching a toolkit. “Intro to TensorFlow/PyTorch”
  - Intro to Machine Learning

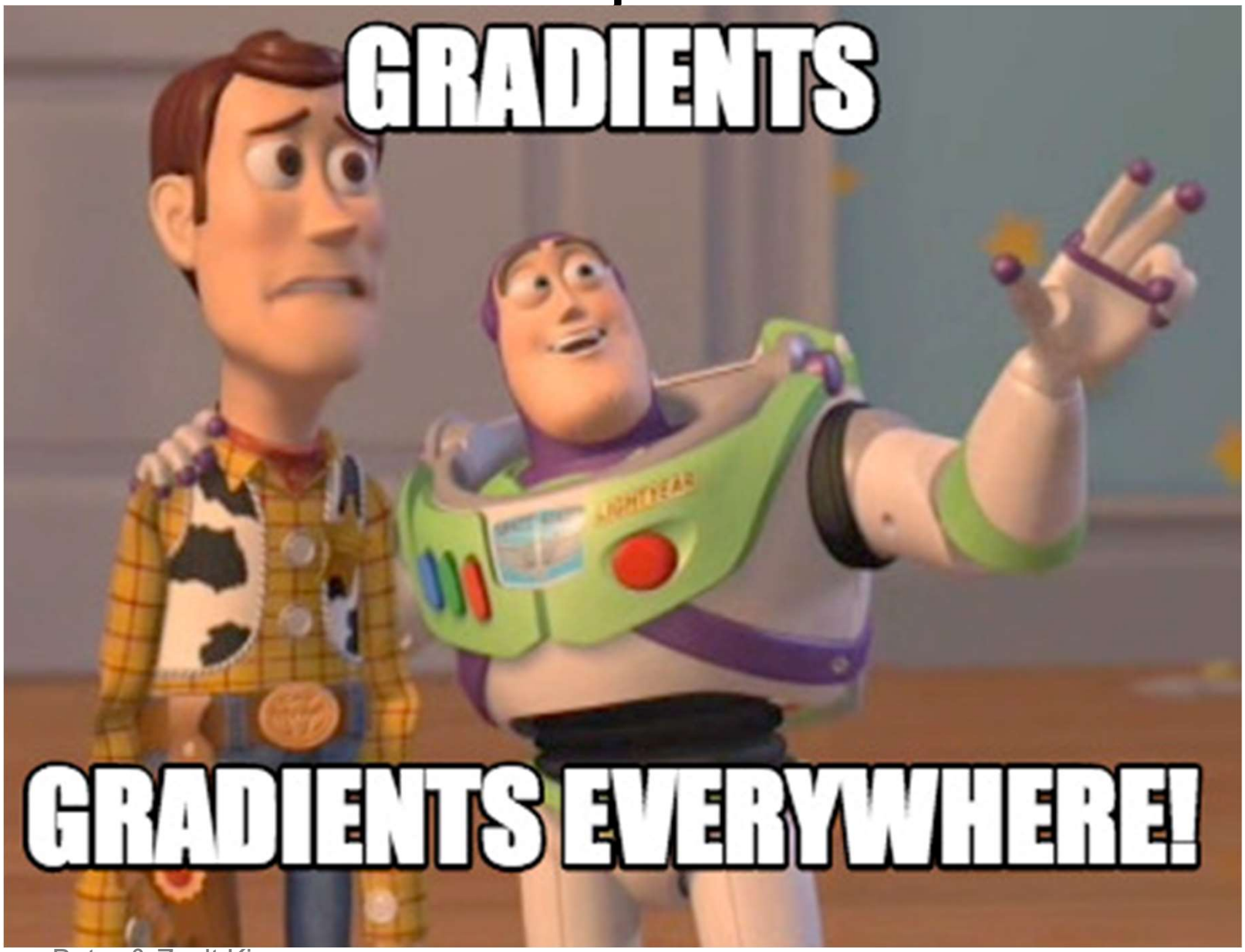
Caveat

- 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

- 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

Prerequisites



Prerequisites

- Intro Machine Learning
  - Classifiers, regressors, loss functions, MLE, MAP
- Linear Algebra
  - Matrix multiplication, eigenvalues, positive semi-definiteness…
- Calculus
  - Multi-variate gradients, hessians, jacobians…
- Programming!
  - Homeworks will require Python!
  - Libraries/Frameworks: PyTorch
  - HW0+5 (pure python), HW1 (python + PyTorch), HW2+3+4 (PyTorch)
  - Your language of choice for project

Course Information

- Instructor: Zsolt Kira
  - zkira@gatech dot edu **(use piazza for most things!)**

TAs

Manas Sahni (Head TA)

Yihao Chen

Rahul Duggal

Program: PhD CS (3rd Year)
Research Interests: Model compression, Neural architecture search.

Hrishikesh Kale

Man Xie

PhD in Interactive Computing, working with Dr. Frank Dellaert

Michael Piseno

- 5th year studying CS and Math
- Third semester TAing for 4803/7643
- Research interests are multi-agent RL and robotics

Interests: Snowboarding, Learning Languages

Zhuoran Yu

Organization & Deliverables

- 4 problem-sets+homeworks (80%)
  - 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, Start early, Start early, Start early, Start early
- Final project (20%)
  - Projects done in groups of 3-4
- (Bonus) Class Participation (3%)
  - Contribute to class discussions on Piazza
  - Ask questions, answer questions

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

PS0

- Out today; due Wed. Jan 20th
  - Will be available on class webpage + Canvas
- Grading
  - Not counted towards your final grade, but required
  - <=75% means that you might not be prepared for the class
  - If you submit after Wed. we will not grade before registration ends
- Topics
  - PS: probability, calculus, convexity, proving things

Computing

- Major bottleneck
  - GPUs
- Options
  - Your own / group / advisor's resources
  - Google Colab
    - jupyter-notebook + free GPU instance
  - Google Cloud credits (details TBD)

4803 vs 7643

- Level differentiation
- HWs
  - Extra credit questions for 4803 students, necessary for 7643
- Project
  - Higher expectations from 7643

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. Size will not increase further.
  - Do PS0. 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. We will give preference to people taking class for credit.
- Sitting in
  - Welcome to. Talk to instructor.

What is the re-grading policy?

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

- This is an advanced grad class.
  - The goal is understanding the material and making progress towards our research.

What is the collaboration policy?

- Collaboration
  - Only on HWs and project (not allowed in PS0).
  - 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

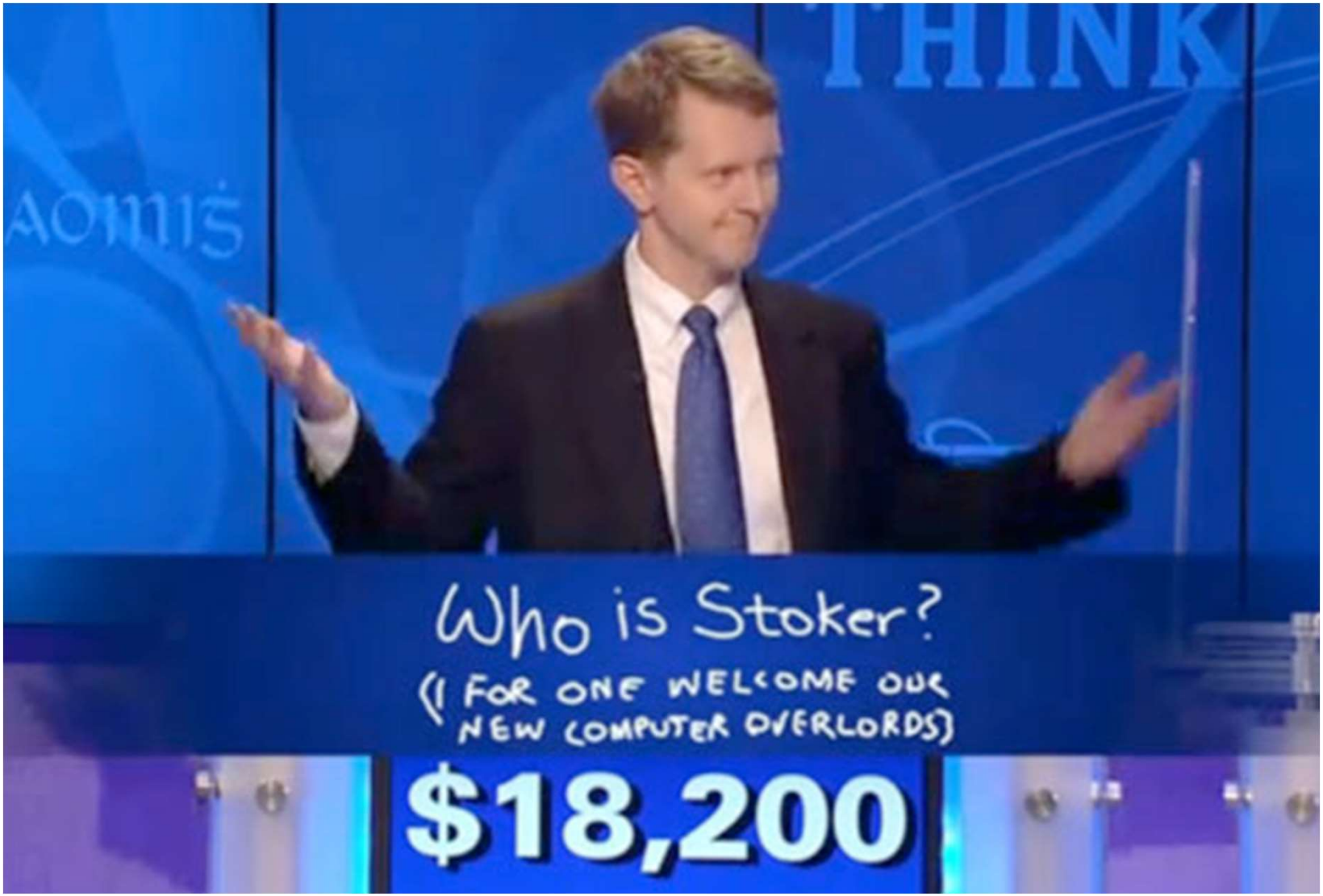
Research

- "Can I work with your group for funding/credits/neither?"
  - I am not taking new advising duties.
  - If you can find one of my students to supervise you, I am happy to sign off on the paperwork.
  - Your responsibility to approach them and ask. It will help if you know what they are working on.

Todo

- PS0
  - Due: Jan 20 11:59pm

Welcome


THINK
Who is Stoker?
(I FOR ONE WELCOME OUR NEW COMPUTER OVERLORDS)
$18,200