CS 4803 / 7643: Deep Learning

Website: www.cc.gatech.edu/classes/AY2019/cs7643_fall/

Piazza: piazza.com/gatech/fall2018/cs48037643

Canvas: gatech.instructure.com/courses/28059

Gradescope: gradescope.com/courses/22096

Dhruv Batra

School of Interactive Computing

Georgia Tech

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 Deep Learning?

Some of the most exciting developments in

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

in the last 5 years!

Proxy for public interest

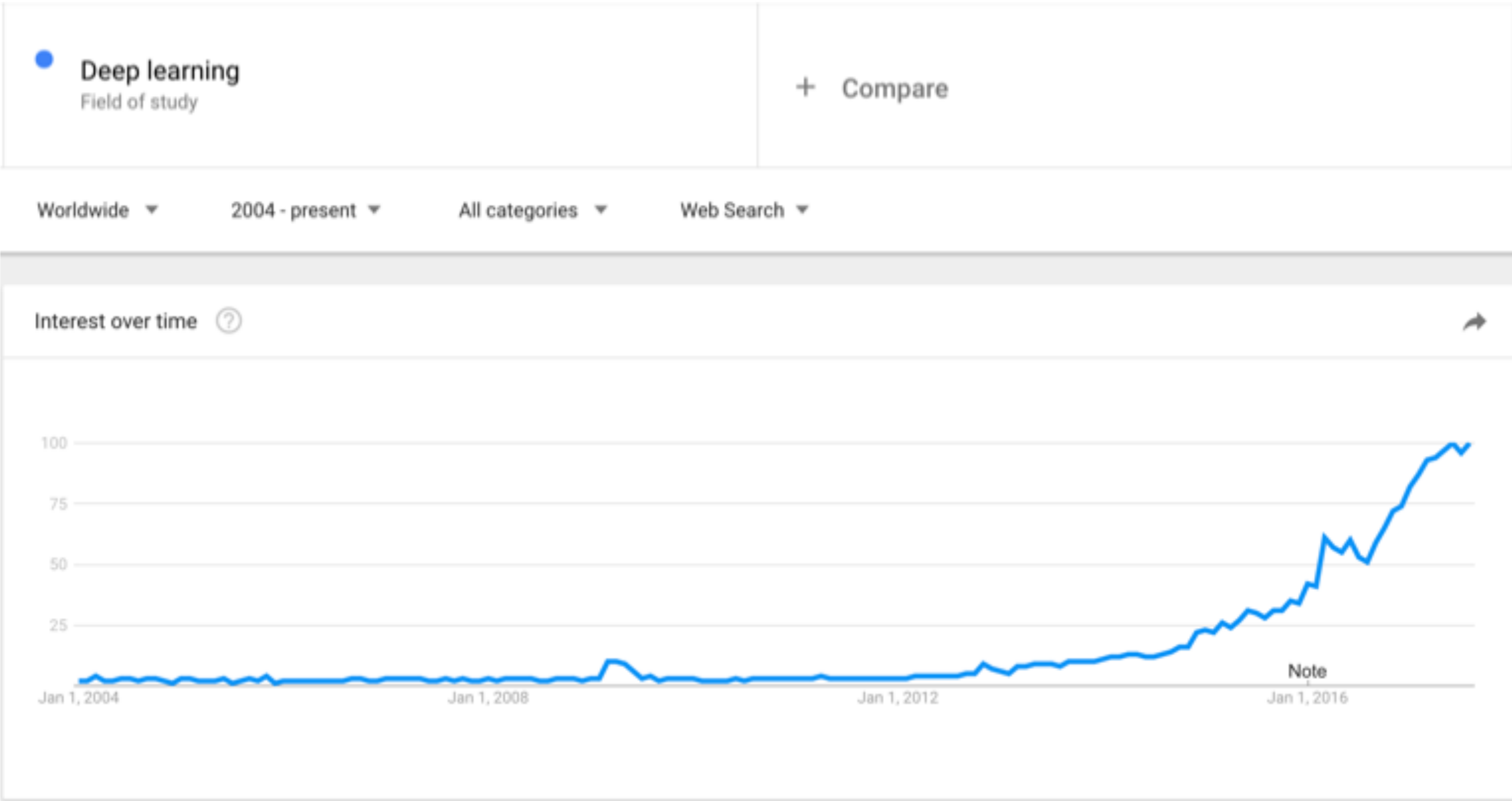


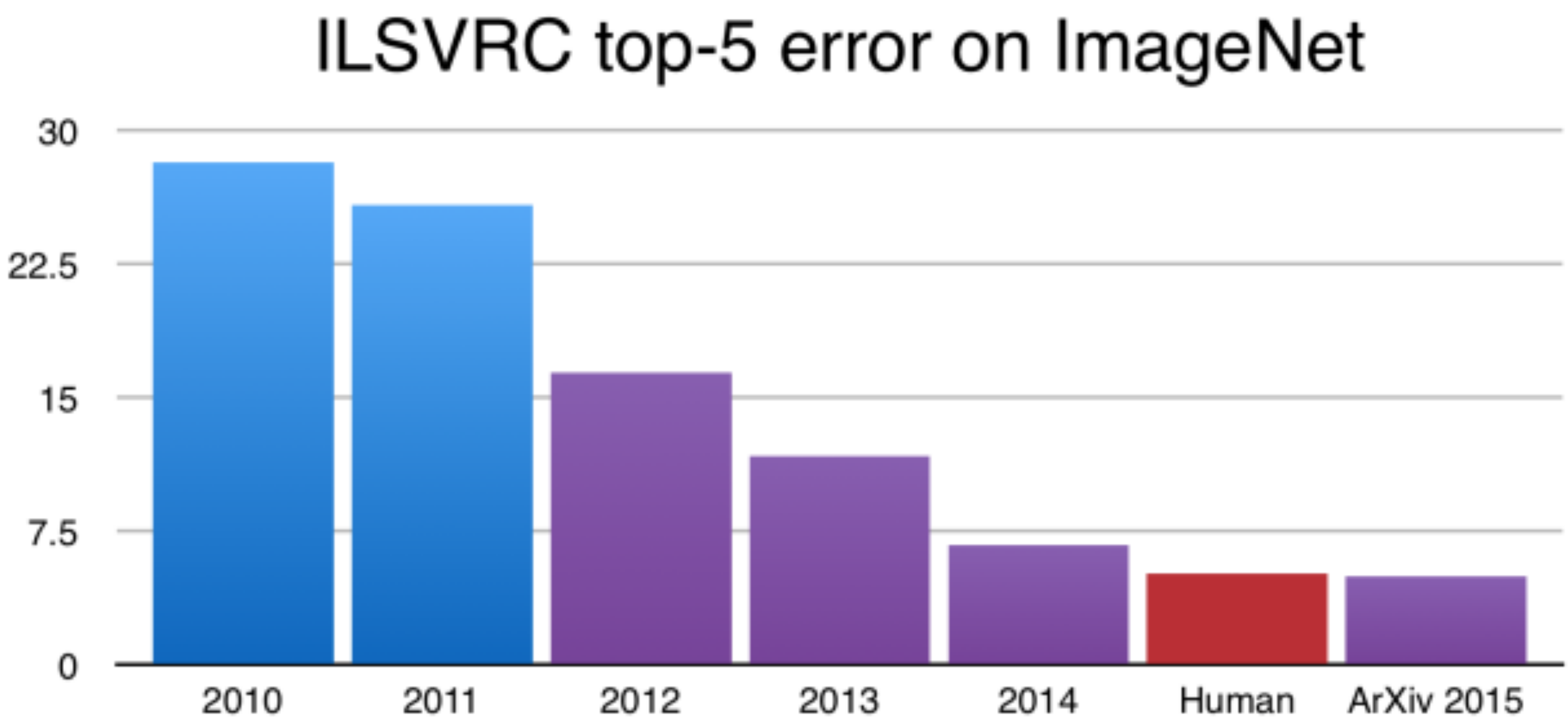
Image Classification

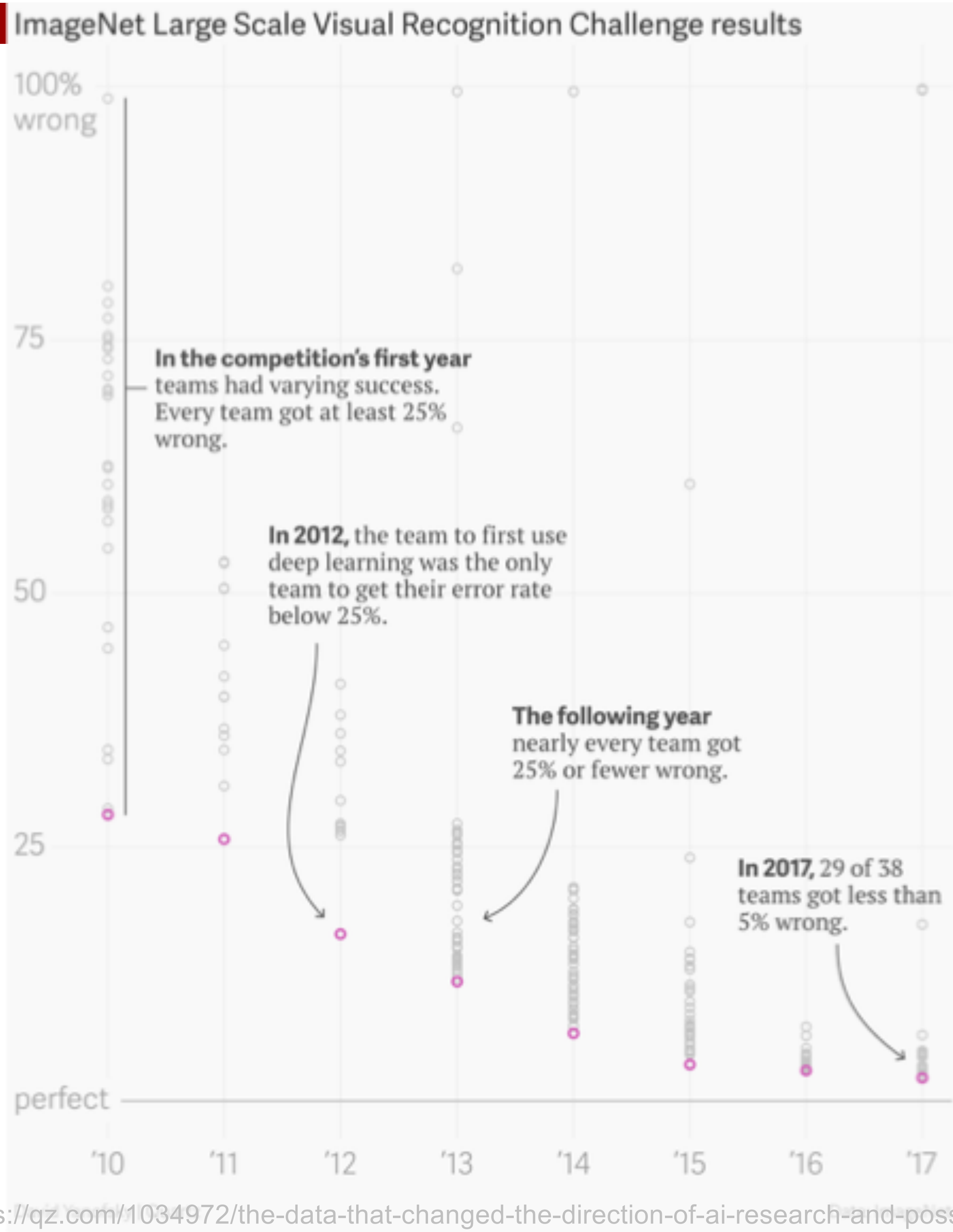
ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

1000 object classes **1.4M/50k/100k images**

http://image-net.org/challenges/LSVRC/{2010,…,2015}

Image Classification





https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world/

AlphaGo seals 4-1 victory over Go grandmaster Lee Sedol

DeepMind's artificial intelligence astonishes fans to defeat human opponent and offers evidence computer software has mastered a major challenge

The world's top Go player, Lee Sedol, lost the final game of the Google DeepMind challenge match. Photograph: Yonhap/Reuters

Google DeepMind's AlphaGo program triumphed in its final game against South Korean Go grandmaster Lee Sedol to win the series 4-1, providing further evidence of the landmark achievement for an artificial intelligence program.

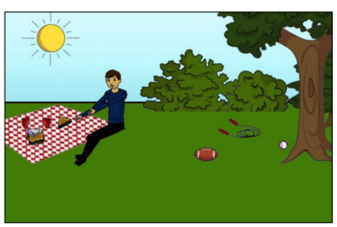
Tasks are getting bolder

A group of young people playing a game of Frisbee

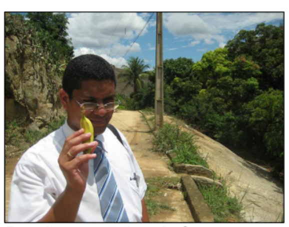
Vinyals et al., 2015

What color are her eyes?
What is the mustache made of?

How many slices of pizza are there?
Is this a vegetarian pizza?

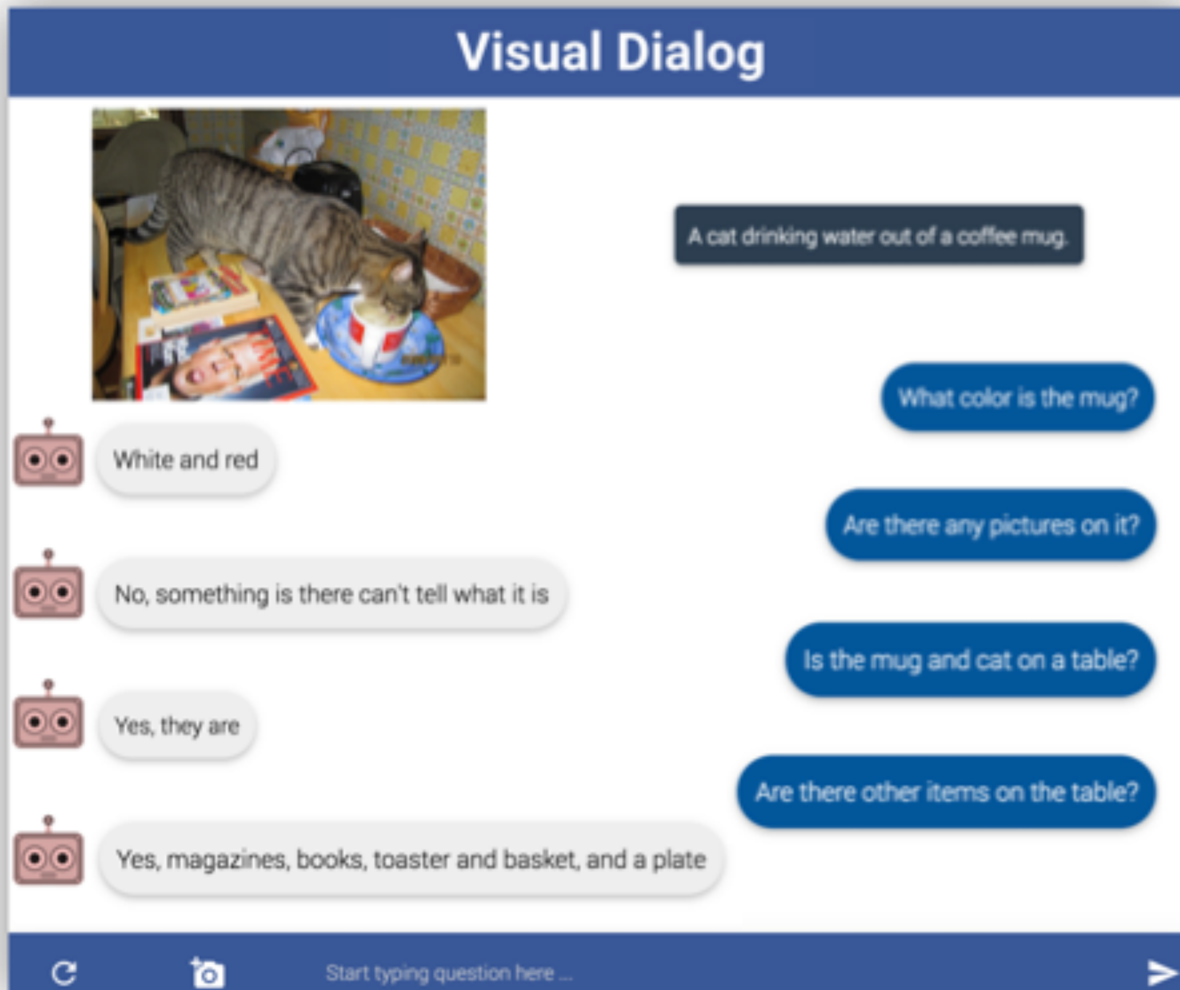


Is this person expecting company?
What is just under the tree?



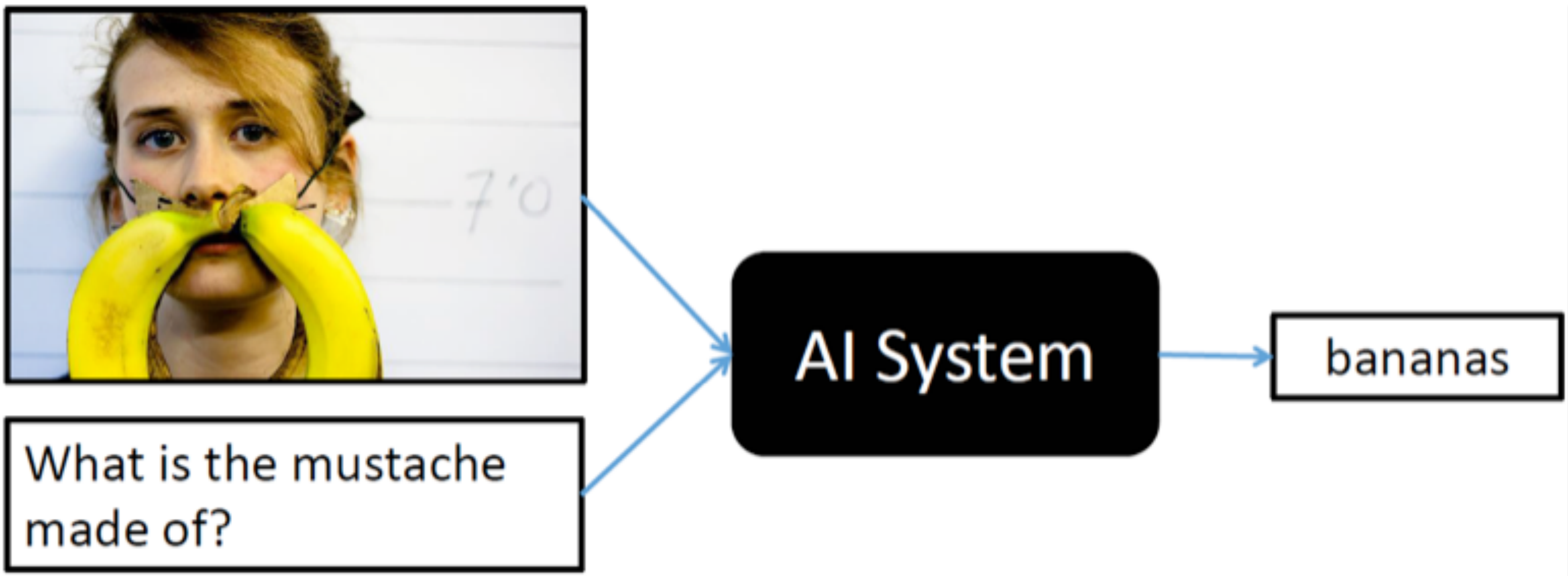
Does it appear to be rainy?
Does this person have 20/20 vision?

Antol et al., 2015



Das et al., 2017

Visual Question Answering (VQA)



Visual Dialog

[CVPR ‘17]

Abhishek Das
(Georgia Tech)

Satwik Kottur
(CMU)

Khushi Gupta
(CMU)

Avi Singh
(UC Berkeley)

Deshraj Yadav
(Virginia Tech)

José Moura
(CMU)

Devi Parikh
(Georgia Tech / FAIR)

Dhruv Batra
(Georgia Tech / FAIR)

Visual Dialog

Visual Dialog

Visual Dialog

A man and a woman are holding umbrellas

Visual Dialog

A man and a woman are holding umbrellas

What color is his umbrella?

Visual Dialog
A man and a woman are holding umbrellas
What color is his umbrella?


Visual Dialog
A man and a woman are holding umbrellas
What color is his umbrella?

Visual Dialog

A man and a woman are holding umbrellas

What color is his umbrella?

His umbrella is black

Visual Dialog

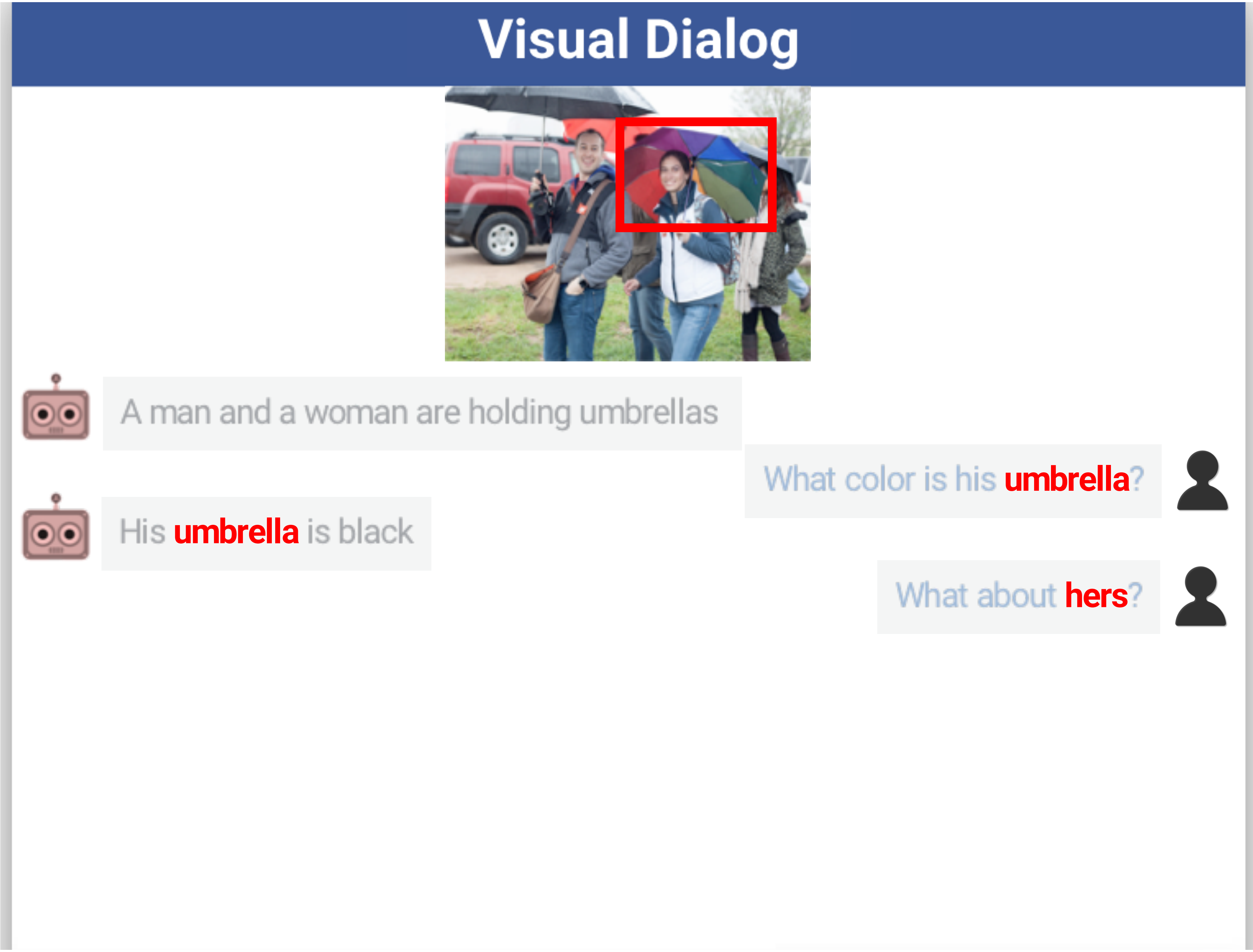
A man and a woman are holding umbrellas

What color is his umbrella?

His umbrella is black

What about hers?


Visual Dialog
A man and a woman are holding umbrellas
What color is his umbrella?
His umbrella is black
What about hers?


Visual Dialog
A man and a woman are holding umbrellas
What color is his umbrella?
His umbrella is black
What about hers?

Visual Dialog

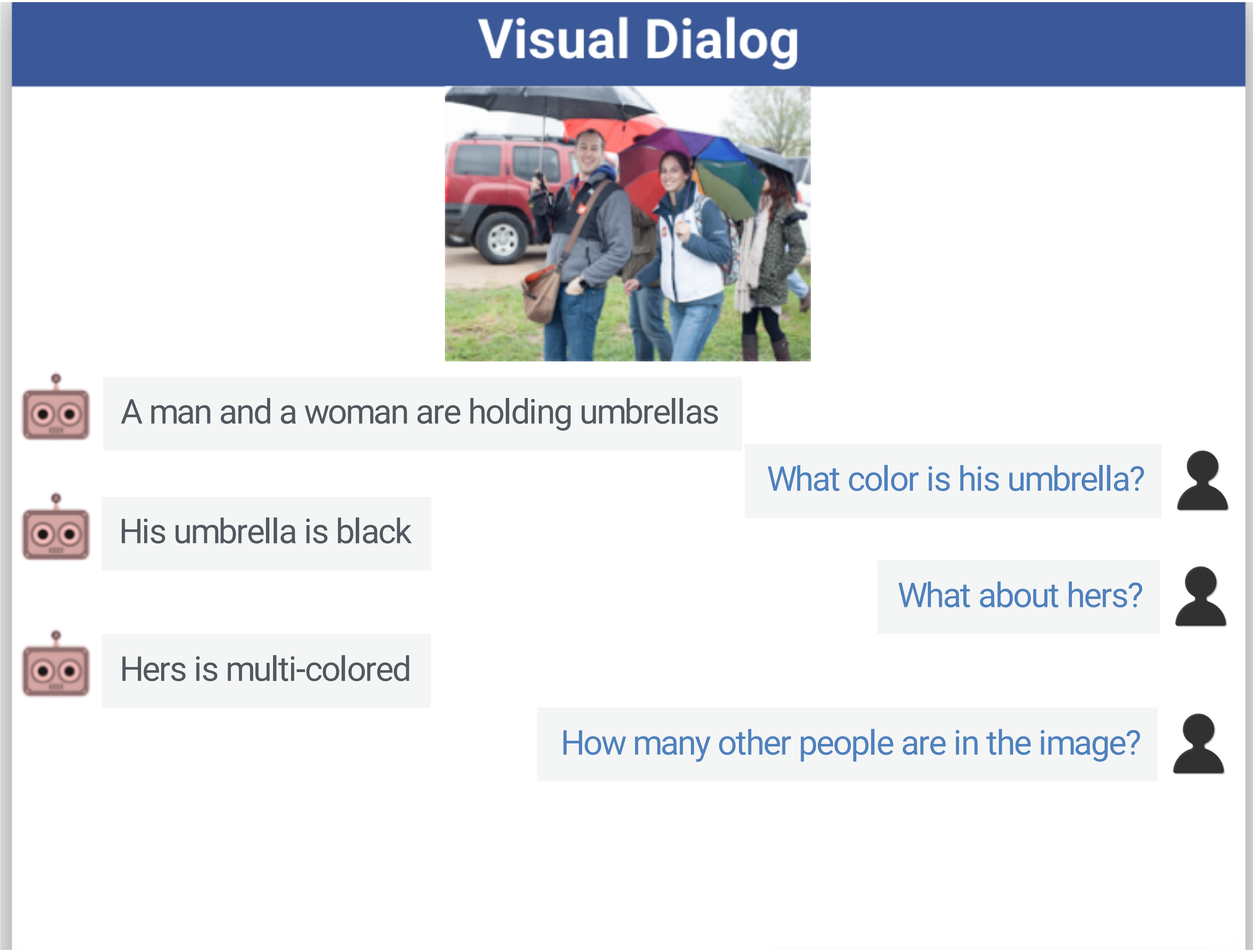
A man and a woman are holding umbrellas

What color is his umbrella?

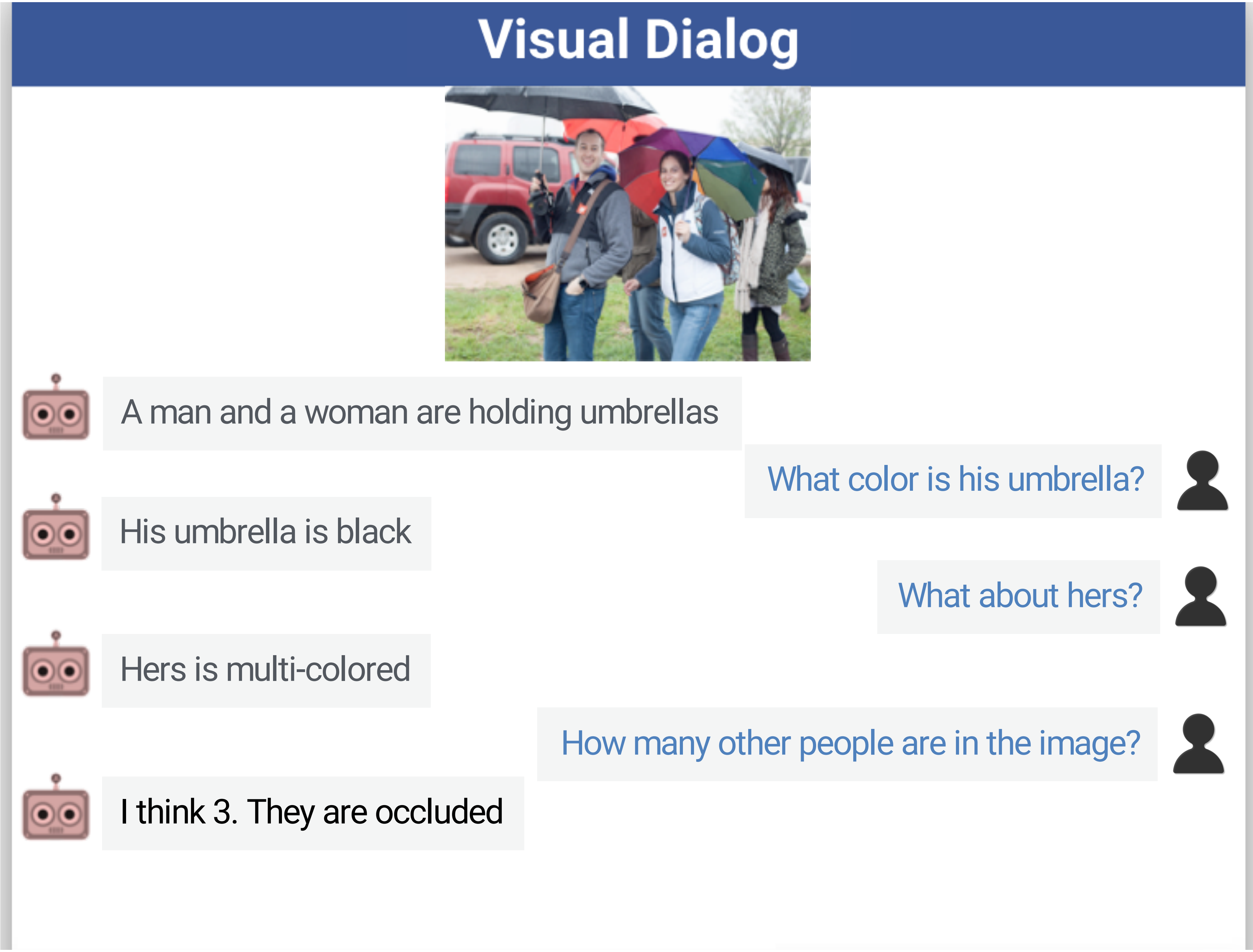
His umbrella is black

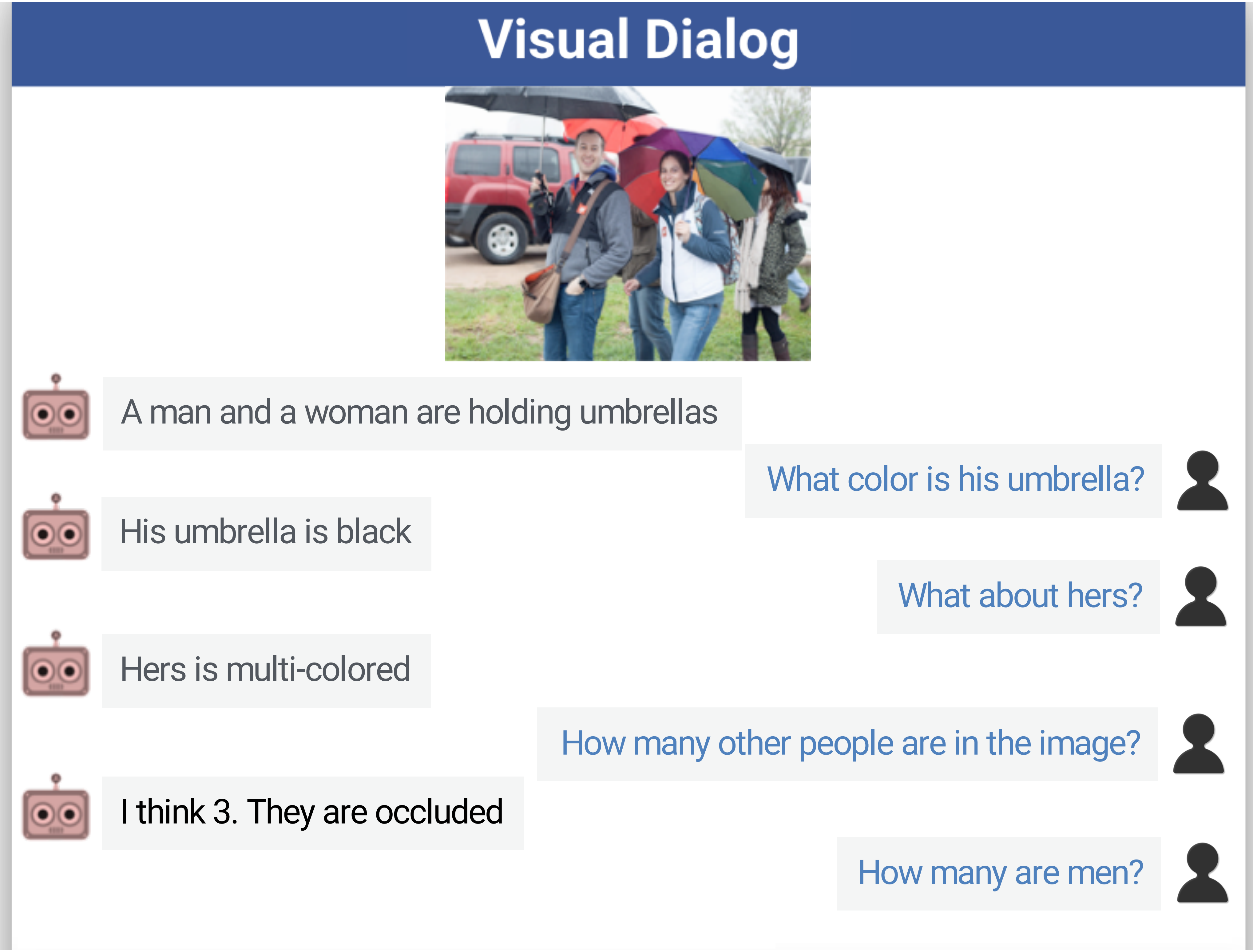
What about hers?

Hers is multi-colored


Visual Dialog
A man and a woman are holding umbrellas
What color is his umbrella?
His umbrella is black
What about hers?
Hers is multi-colored
How many other people are in the image?

Visual Dialog
A man and a woman are holding umbrellas
What color is his umbrella?
His umbrella is black
What about hers?
Hers is multi-colored
How many other people are in the image?


Visual Dialog
A man and a woman are holding umbrellas
What color is his umbrella?
His umbrella is black
What about hers?
Hers is multi-colored
How many other people are in the image?
I think 3. They are occluded


Visual Dialog
A man and a woman are holding umbrellas
What color is his umbrella?
His umbrella is black
What about hers?
Hers is multi-colored
How many other people are in the image?
I think 3. They are occluded
How many are men?


Visual Dialog
A man and a woman are holding umbrellas
What color is his umbrella?
His umbrella is black
What about hers?
Hers is multi-colored
How many other people are in the image?
I think 3. They are occluded
How many are men?

Live demo at

vqa.cloudcv.org.

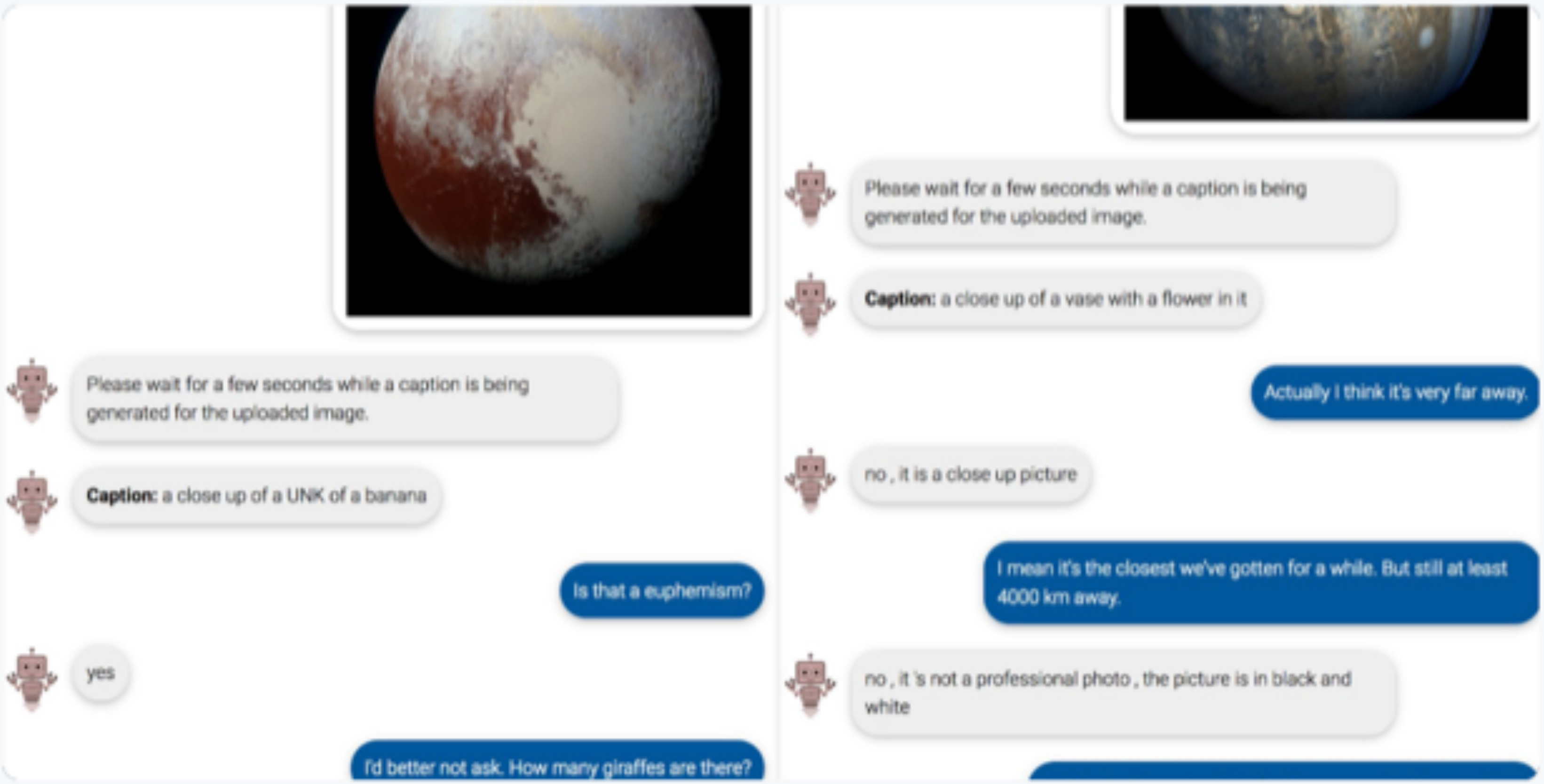
demo.visualdialog.org

Janelle Shane @JanelleCShane · Jun 24

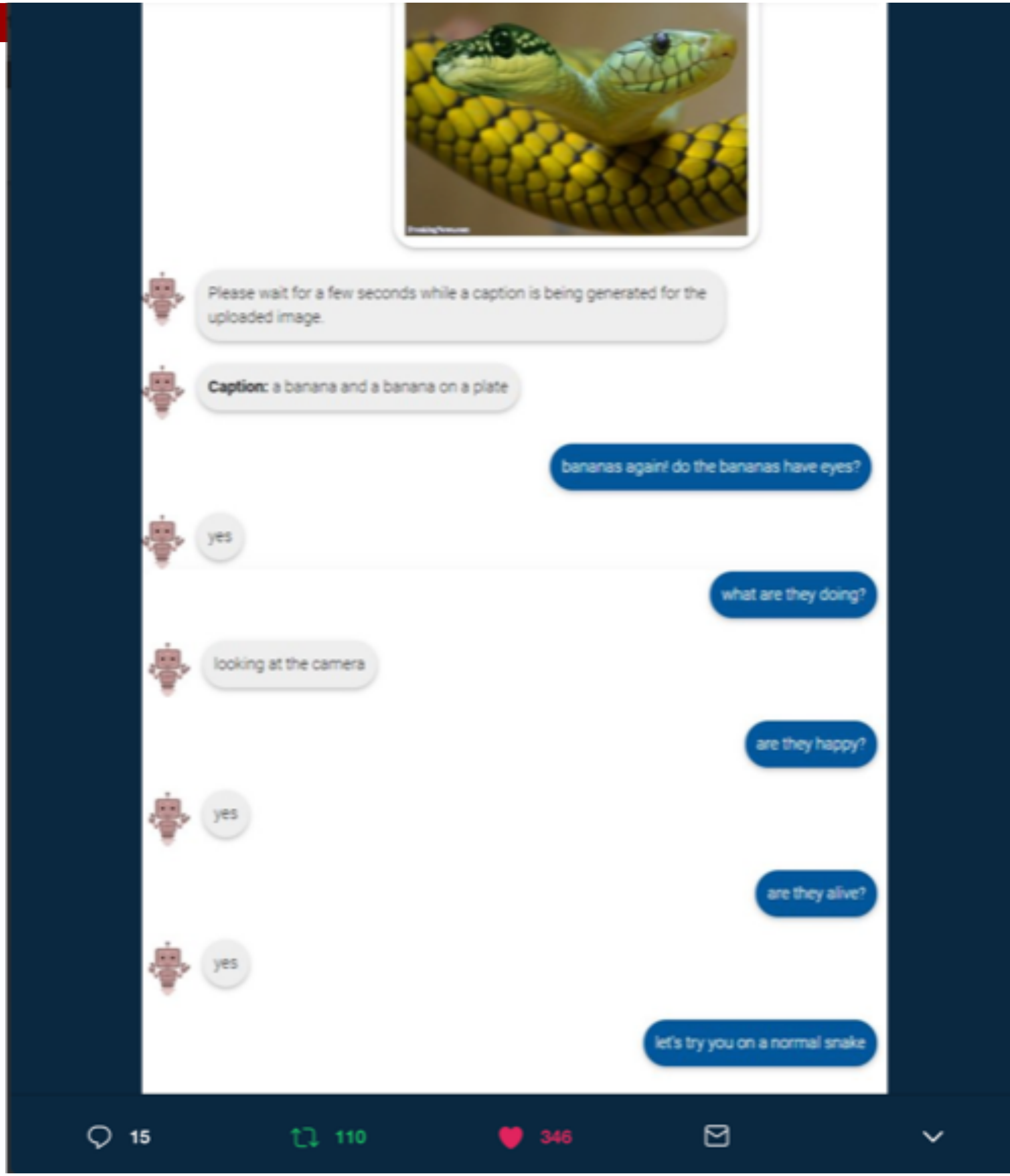
One fun thing I discovered about Visual Chatbot.

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


Please wait for a few seconds while a caption is being generated for the uploaded image.
Caption: a banana and a banana on a plate
bananas again! do the bananas have eyes?
yes
what are they doing?
looking at the camera
are they happy?
yes
are they alive?
yes
let's try you on a normal snake
15
110
346

Embodied Question Answering
[CVPR ’18 Oral]


Abhishek Das
(Georgia Tech)


Samyak Datta
(Georgia Tech)


Georgia Gkioxari
(FAIR)


Stefan Lee
(Georgia Tech)


Devi Parikh
(Georgia Tech / FAIR)


Dhruv Batra
(Georgia Tech / FAIR)


What is to the left of the shower?
Cabinet

What color is the car? – AI Challenges

- Language Understanding
  – What is the question asking?
- Vision
  – What does a ‘car’ look like?
- Active Perception
  – Agent must navigate by perception
- Common sense
  – Where are ‘cars’ generally located in the house?
- Credit Assignment
  – (forward, forward, turn-right, forward, . . . , turn-left, ‘red’)

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

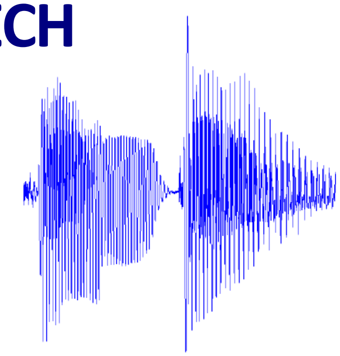
Traditional Machine Learning

VISION

hand-crafted features SIFT/HOG → your favorite classifier → "car"

fixed | learned

SPEECH



hand-crafted features MFCC → your favorite classifier → \ˈd ē p\

fixed | learned

NLP

This burrito place is yummy and fun!

hand-crafted features Bag-of-words → your favorite classifier → "+"

fixed | learned

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

Hierarchical Compositionality

VISION

pixels ▶ edge ▶ texton ▶ motif ▶ part ▶ object

SPEECH

sample ▶ spectral band ▶ formant ▶ motif ▶ phone ▶ word

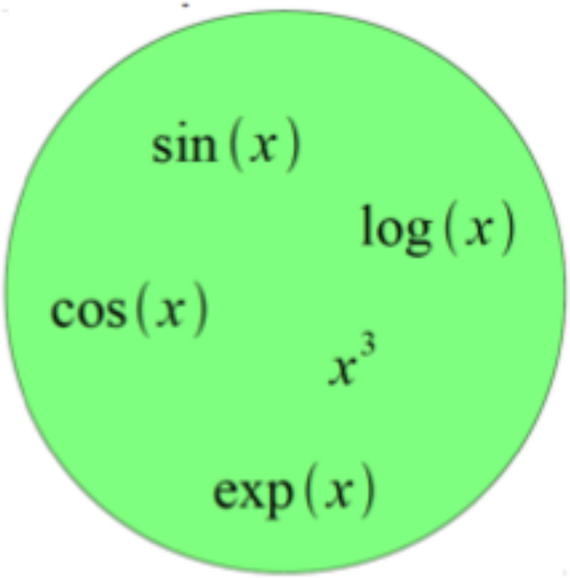
NLP

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

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

Building A Complicated Function

Given a library of simple functions

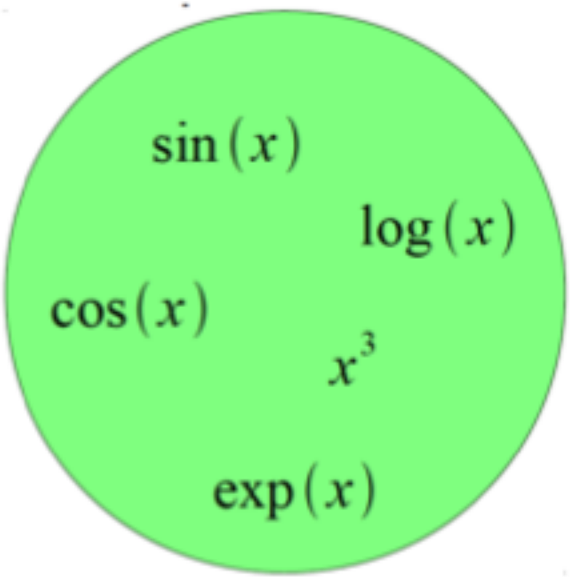


Compose into a

complicate function

Building A Complicated Function

Given a library of simple functions



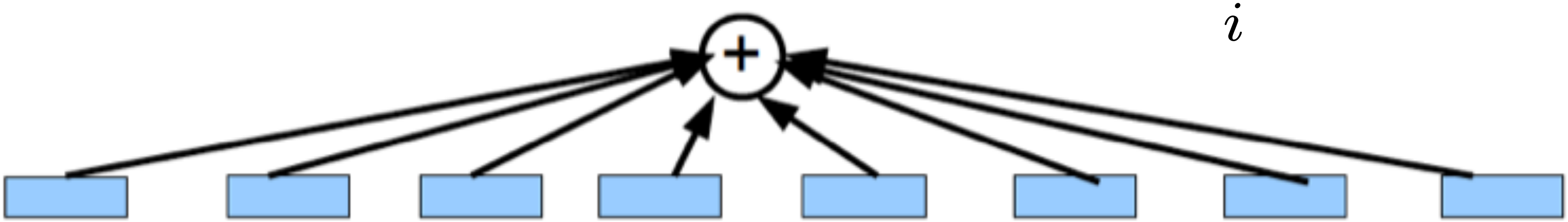
Compose into a

complicate function

Idea 1: Linear Combinations

- Boosting
- Kernels
- …

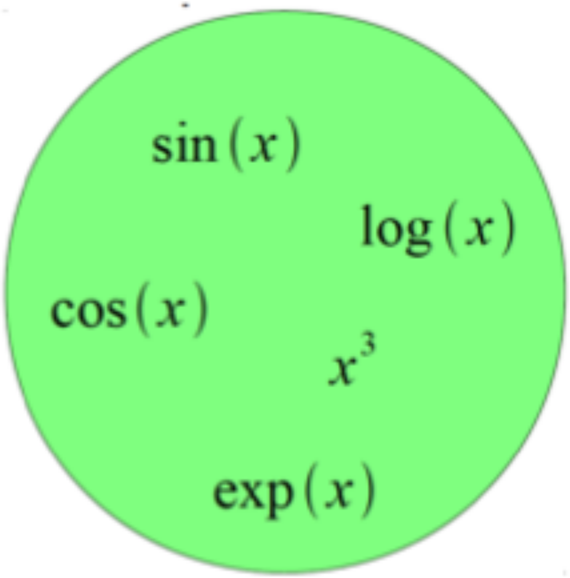
$$f(x) = \sum_i \alpha_i g_i(x)$$



Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

Building A Complicated Function

Given a library of simple functions



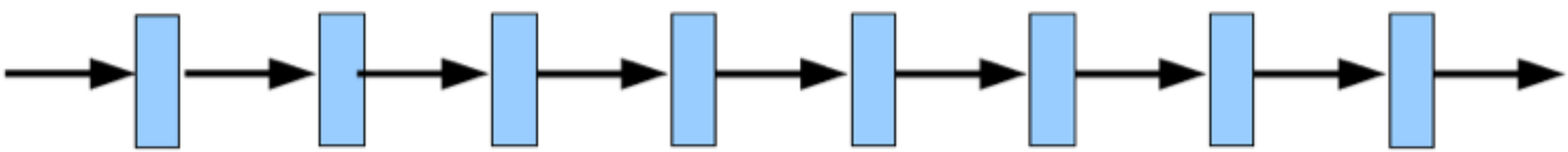
Compose into a

complicate function

Idea 2: Compositions

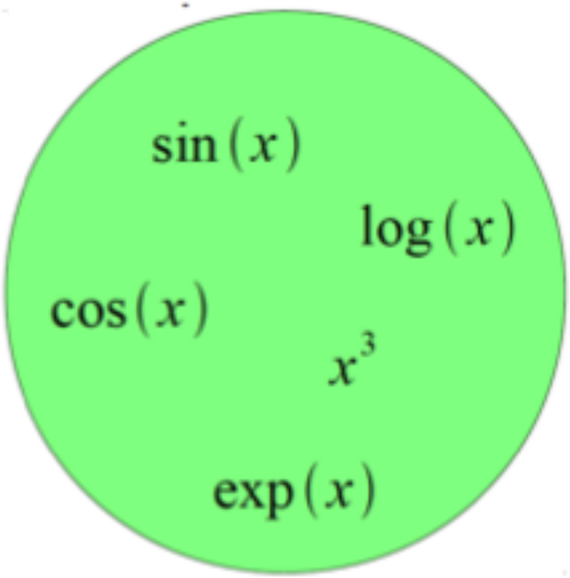
- Deep Learning
- Grammar models
- Scattering transforms…

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



Building A Complicated Function

Given a library of simple functions

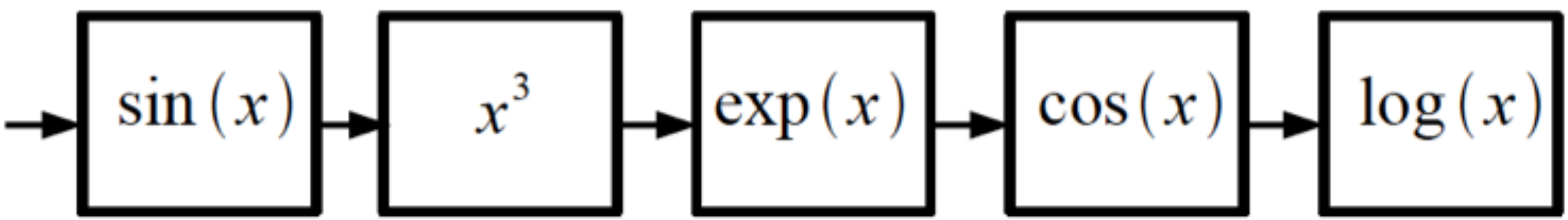


Compose into a complicate function

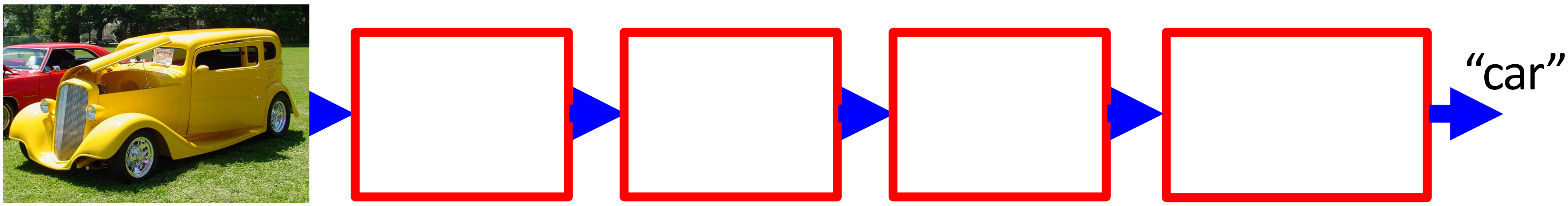
Idea 2: Compositions

- Deep Learning
- Grammar models
- Scattering transforms…

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

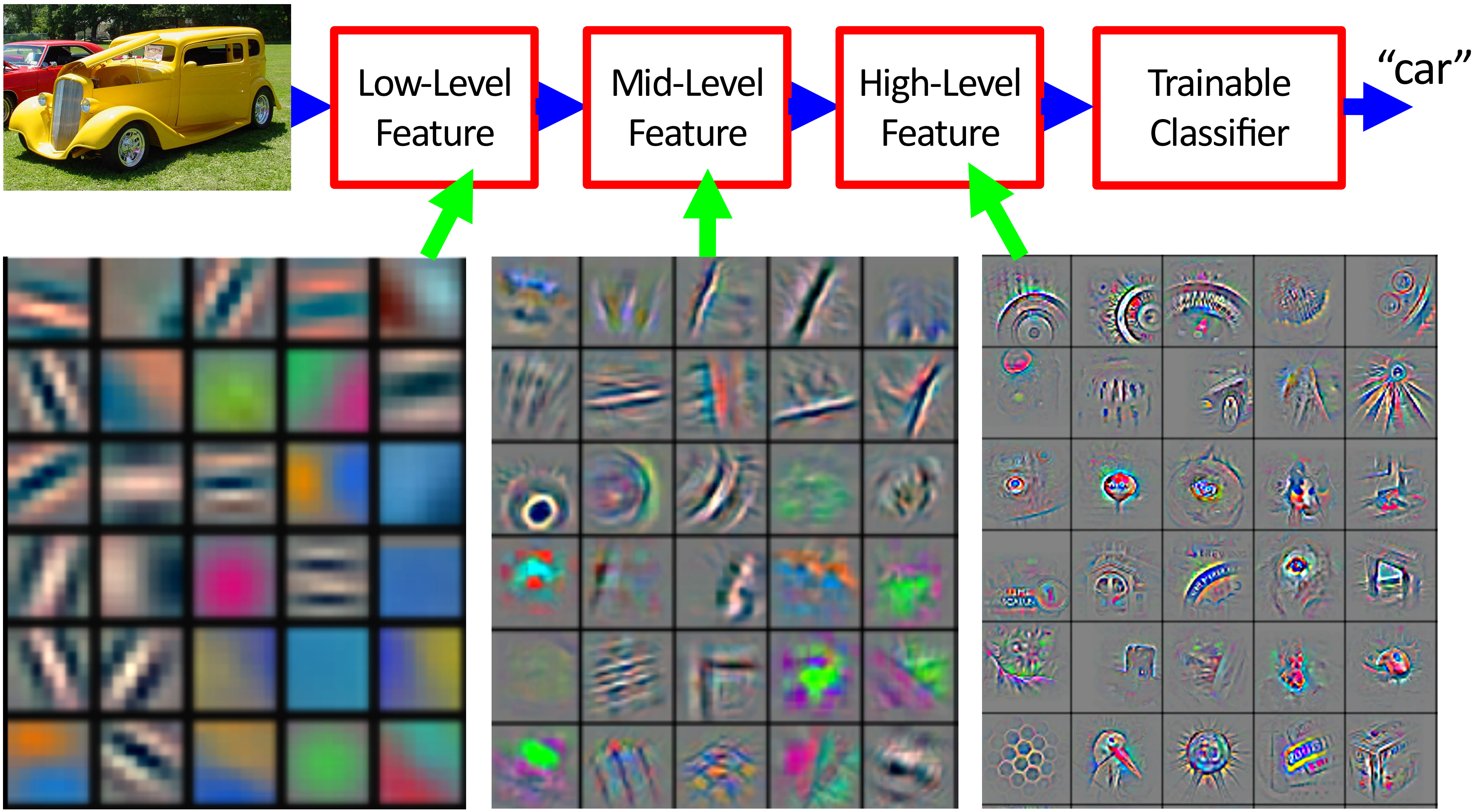


Deep Learning = Hierarchical Compositionality



Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

Deep Learning = Hierarchical Compositionality



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

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

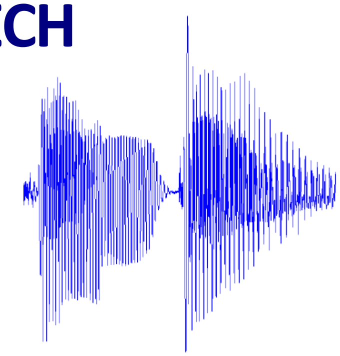
Traditional Machine Learning

VISION

hand-crafted features SIFT/HOG → your favorite classifier → "car"

fixed | learned

SPEECH



hand-crafted features MFCC → your favorite classifier → \ˈd ē p\

fixed | learned

NLP

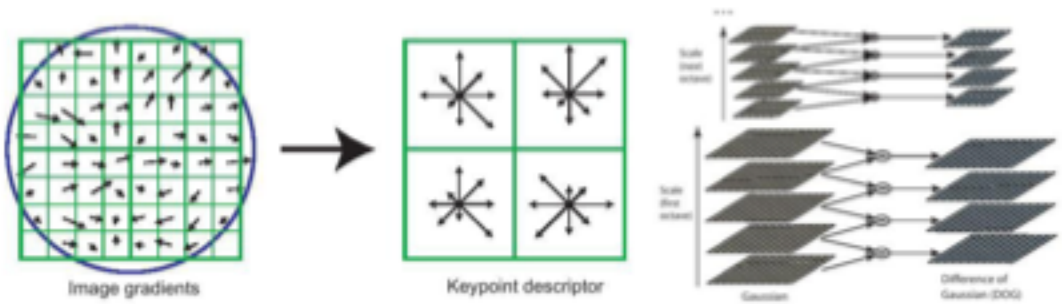
This burrito place is yummy and fun!

→ hand-crafted features Bag-of-words → your favorite classifier → "+"

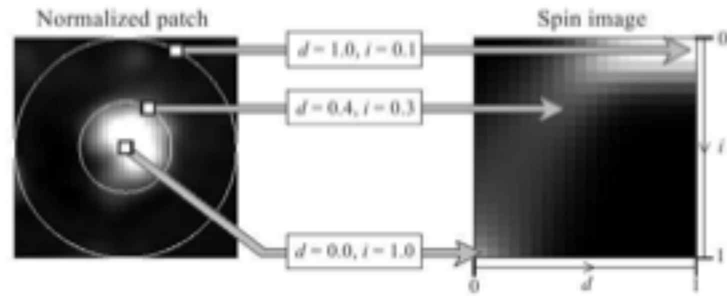
fixed | learned

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

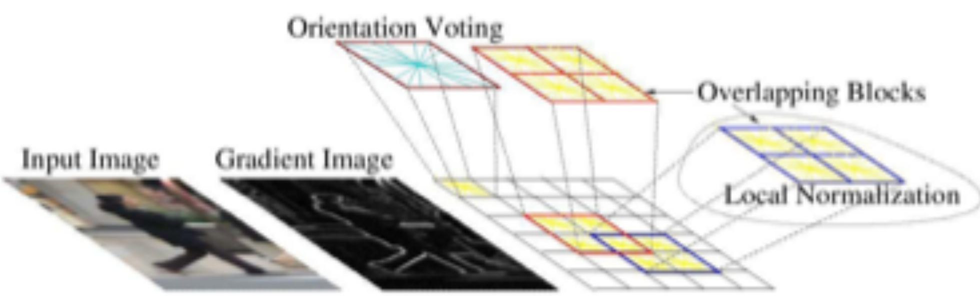
Feature Engineering



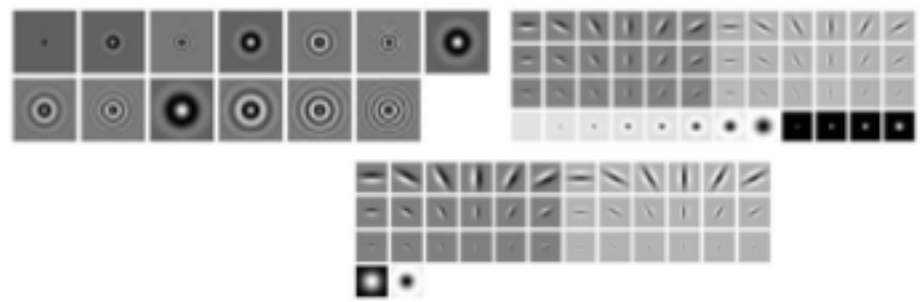
SIFT



Spin Images



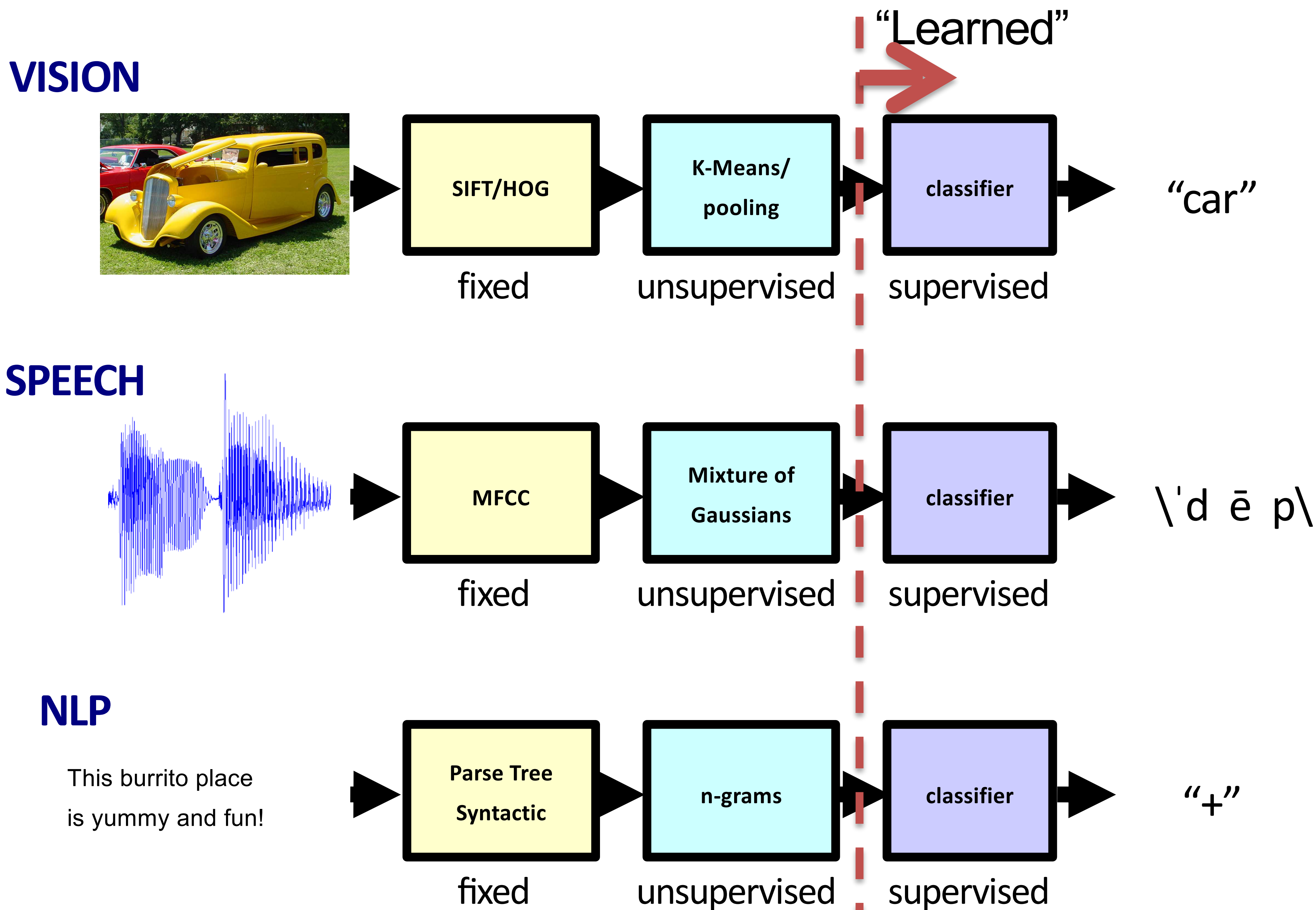
HoG



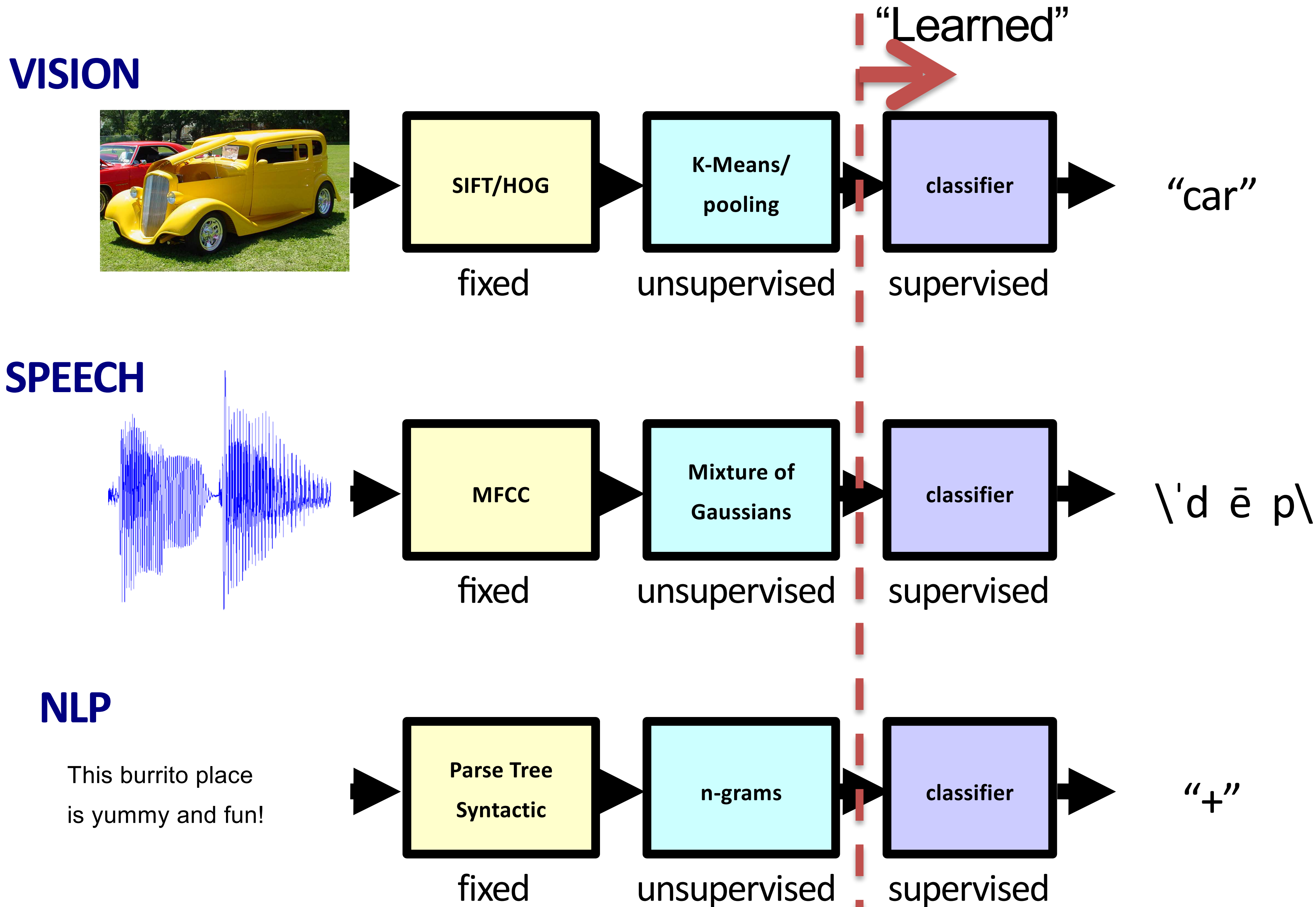
Textons

and many many more….

Traditional Machine Learning (more accurately)



Deep Learning = End-to-End Learning



“Shallow” vs Deep Learning

- “Shallow” models

hand-crafted Feature Extractor

fixed

“Simple” Trainable Classifier

learned

- Deep models

Trainable Feature-Transform / Classifier

Trainable Feature-Transform / Classifier

Trainable Feature-Transform / Classifier

Learned Internal Representations

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

So what *is* Deep (Machine) Learning?

- A few different ideas:

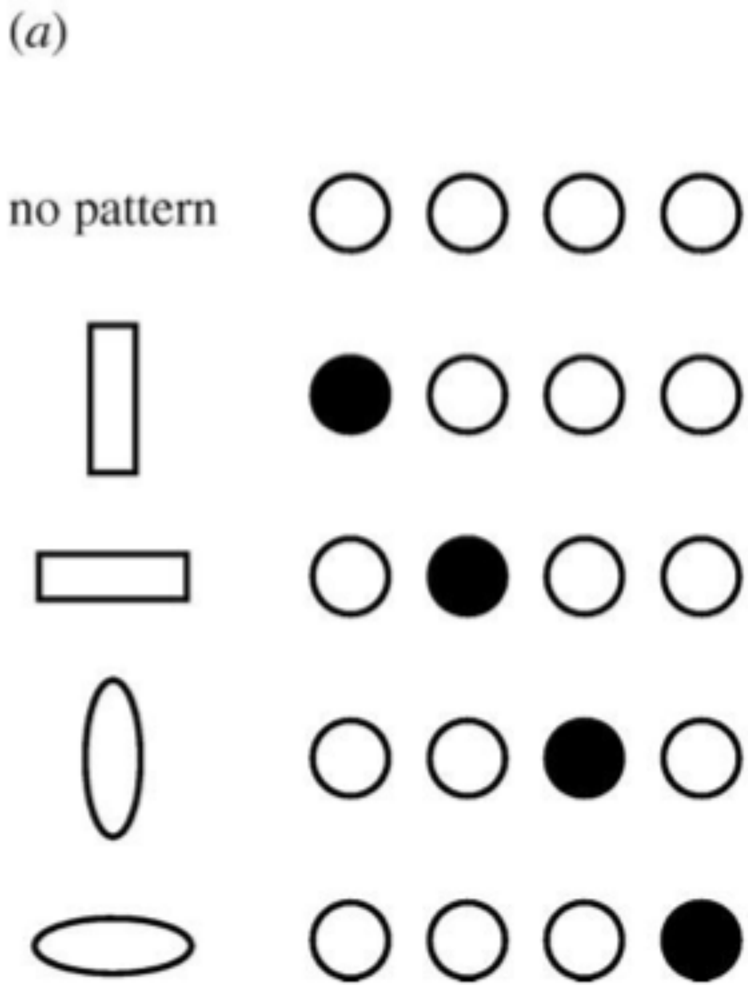
- (Hierarchical) Compositionality
  - Cascade of non-linear transformations
  - Multiple layers of representations

- 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

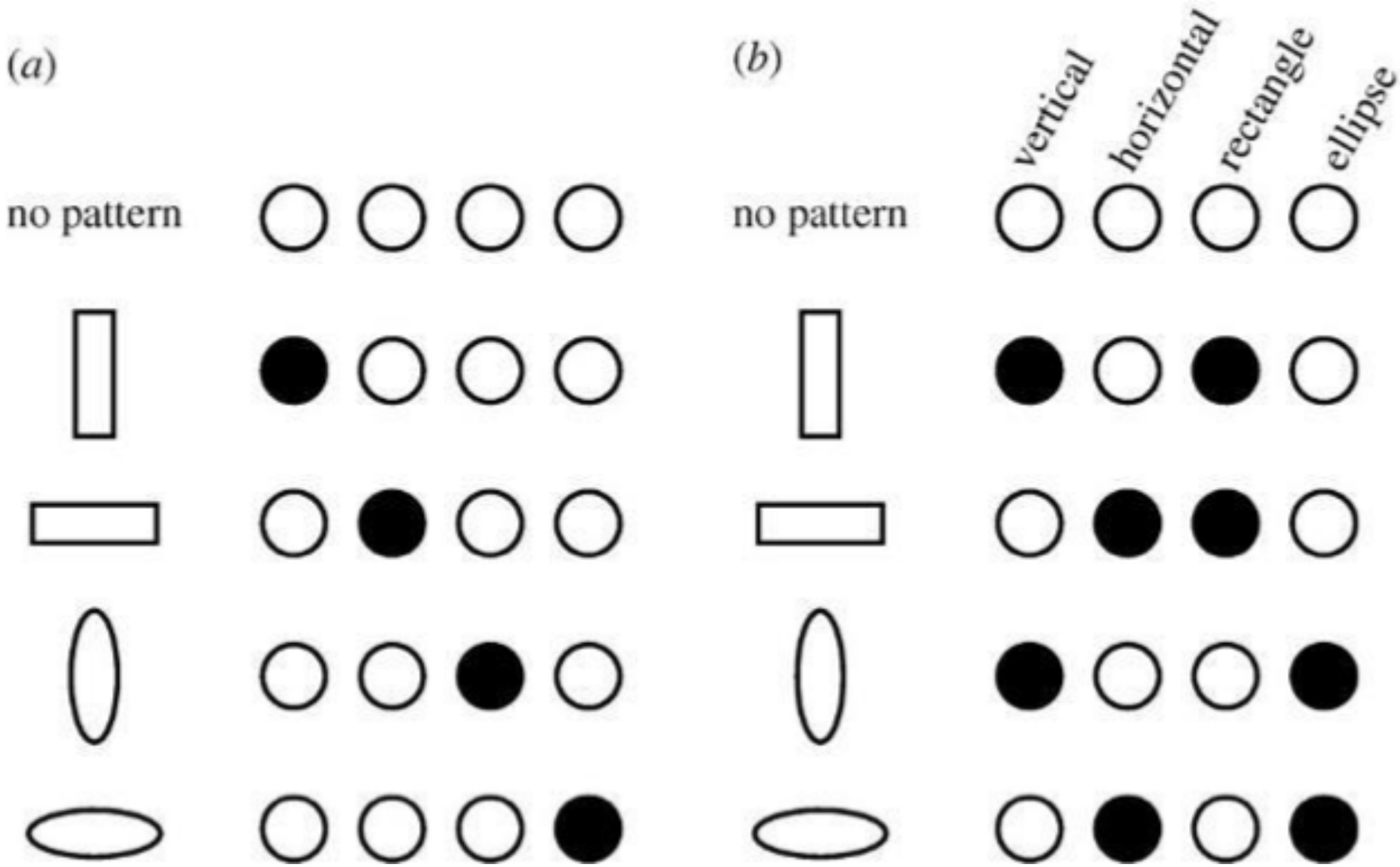
- Local vs Distributed



Slide Credit: Moontae Lee

Distributed Representations Toy Example

- Can we interpret each dimension?



Slide Credit: Moontae Lee

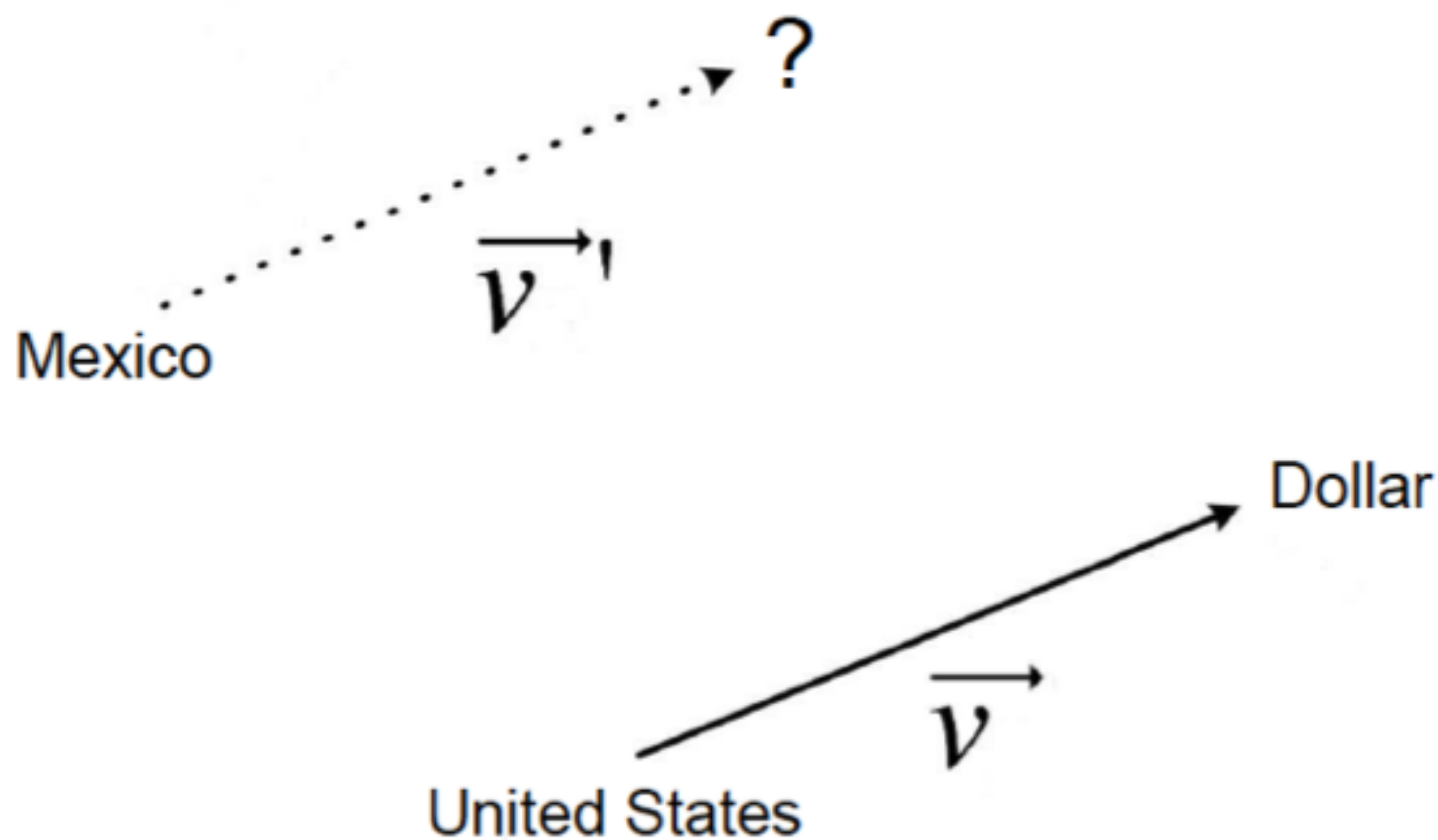
Power of distributed representations!

Local ● ● ○ ● = VR + HR + HE = ?

Distributed ● ● ○ ● = V + H + E ≈ ○

Power of distributed representations!

- United States:Dollar :: Mexico:?



ThisPlusThat.me

Image Credit: http://insightdatascience.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
  - *“Because gradient descent is better than you”*
    Yann LeCun

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

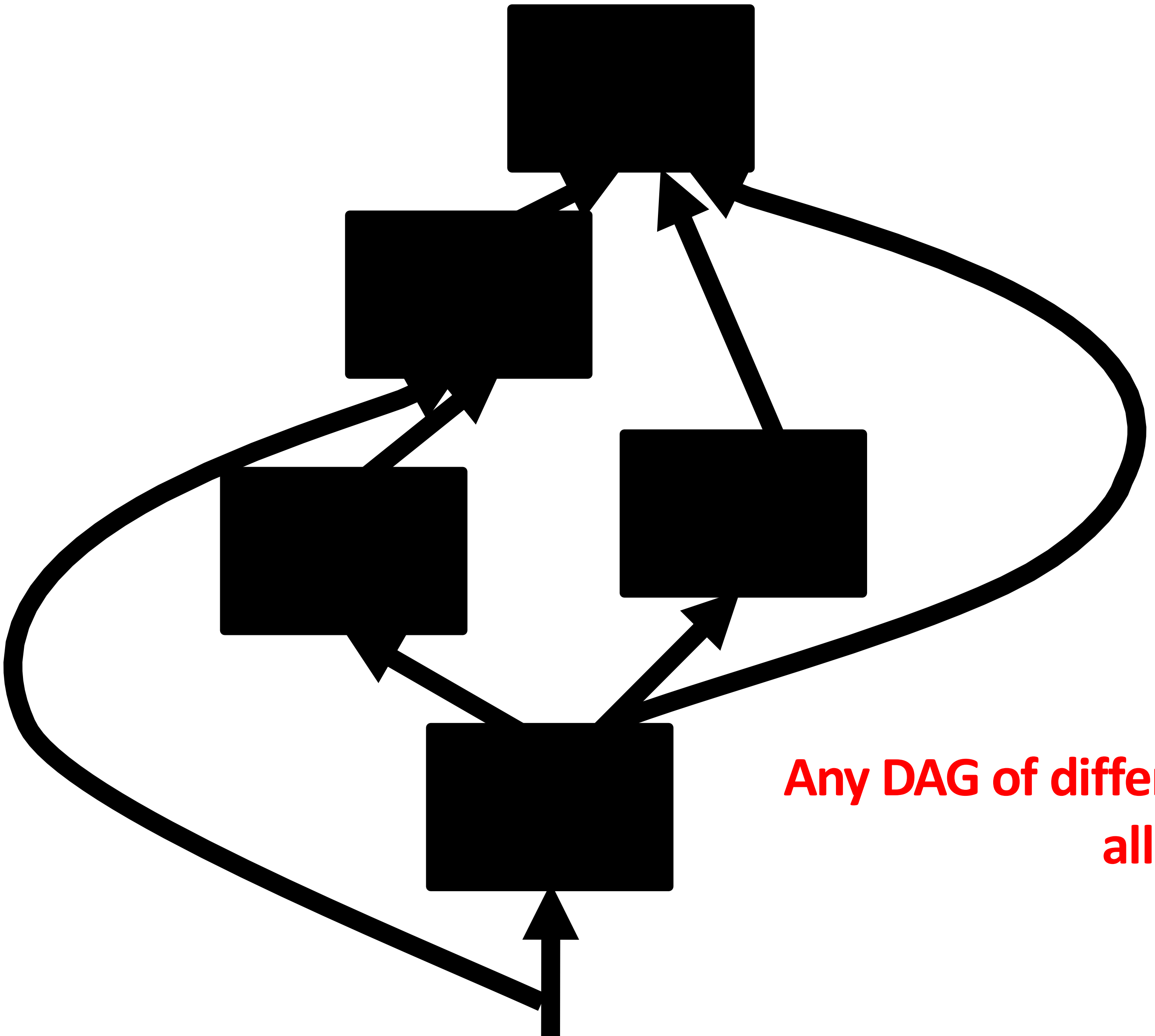
“Expert” intuitions can be misleading

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

Benefits of Deep/Representation Learning

- Modularity!
- Plug and play architectures!

Differentiable Computation Graph



Any DAG of differentialble modules is allowed!

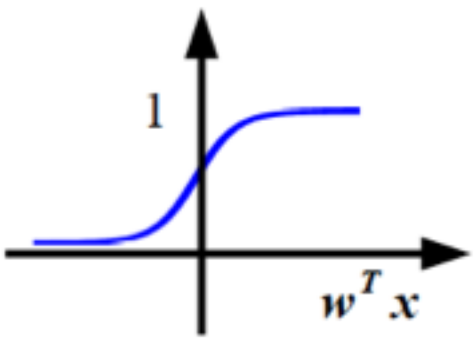
Linear Classifier: Logistic Regression

Input: $\boldsymbol{x} \in R^D$

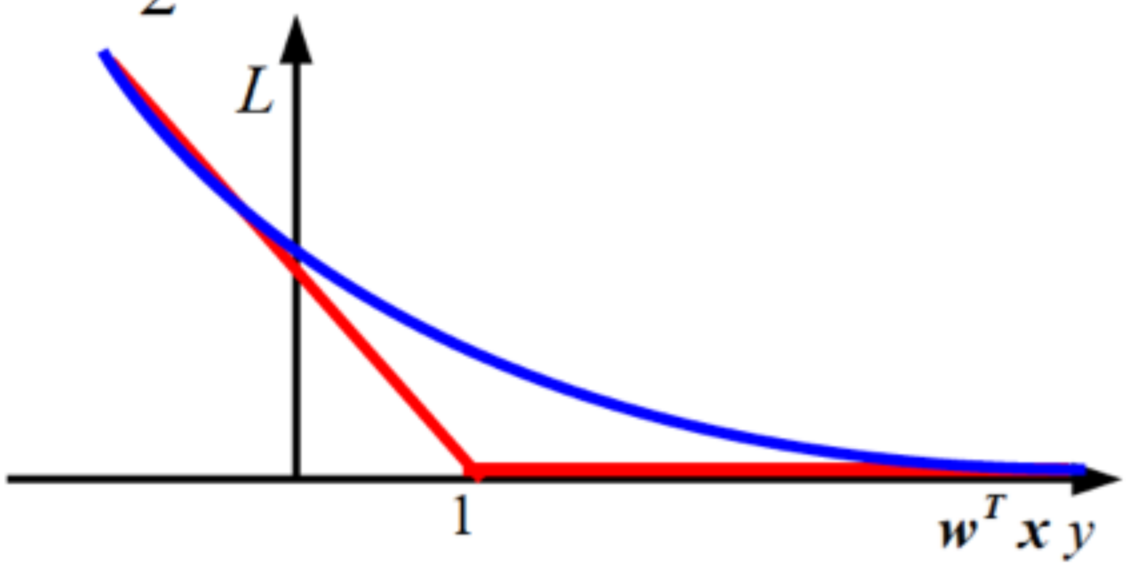
Binary label: $y \in \{-1, +1\}$

Parameters: $\boldsymbol{w} \in R^D$

Output prediction: $p(y=1|\boldsymbol{x}) = \frac{1}{1+e^{-\boldsymbol{w}^T \boldsymbol{x}}}$



Loss: $L = \frac{1}{2}\|\boldsymbol{w}\|^2 - \lambda \log(p(y|\boldsymbol{x}))$

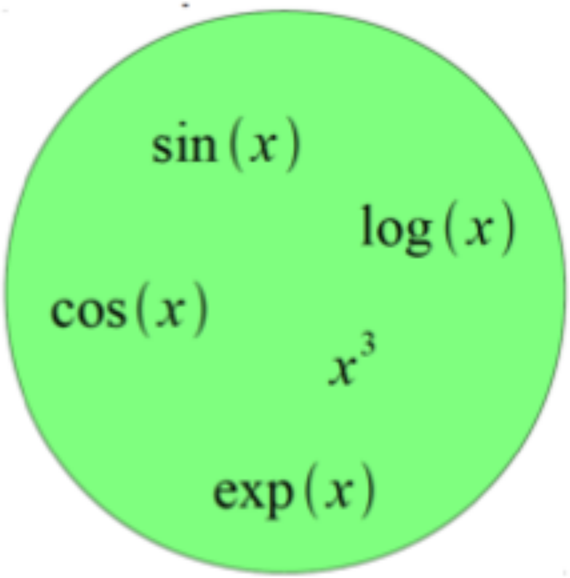


Log Loss

Ranzato

Logistic Regression as a Cascade

Given a library of simple functions

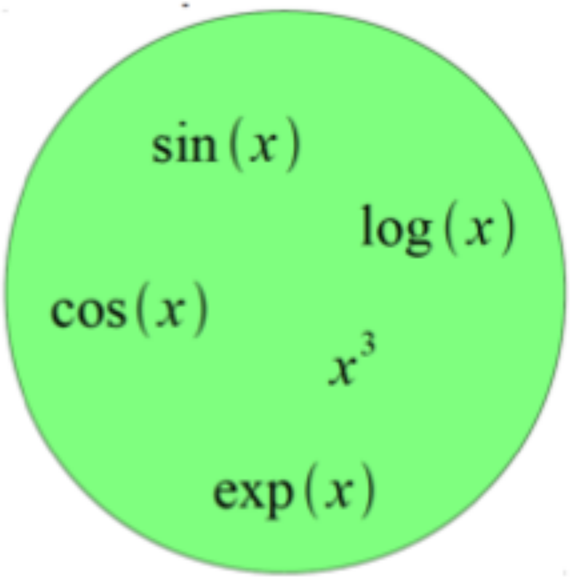


Compose into a complicate function

$$-\log\left(\frac{1}{1+e^{-\mathbf{w}^\intercal \mathbf{x}}}\right)$$

Logistic Regression as a Cascade

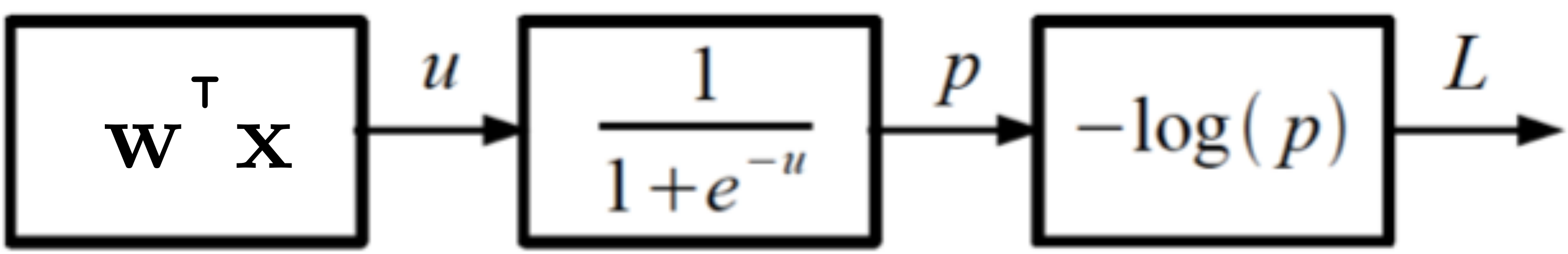
Given a library of simple functions



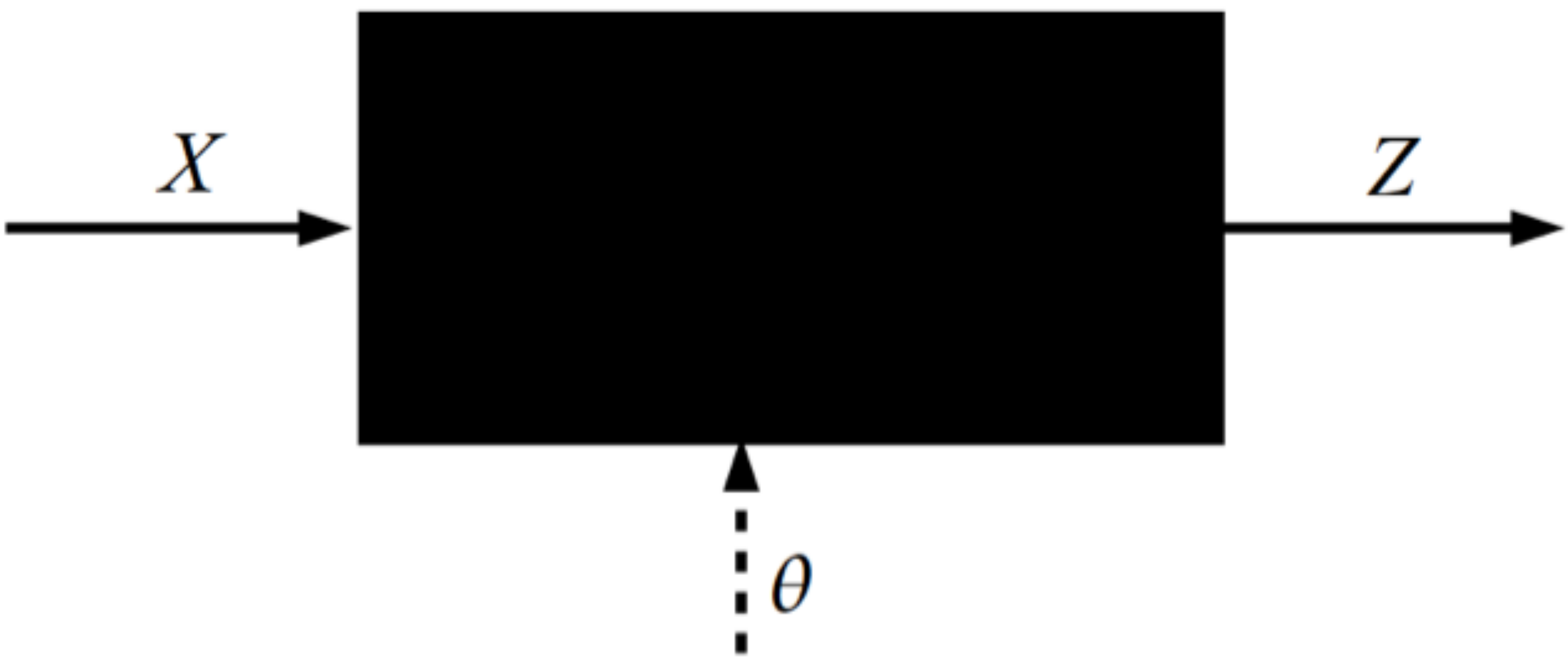
Compose into a

complicate function

$$-\log\left(\frac{1}{1+e^{-\mathbf{w}^\intercal\mathbf{x}}}\right)$$

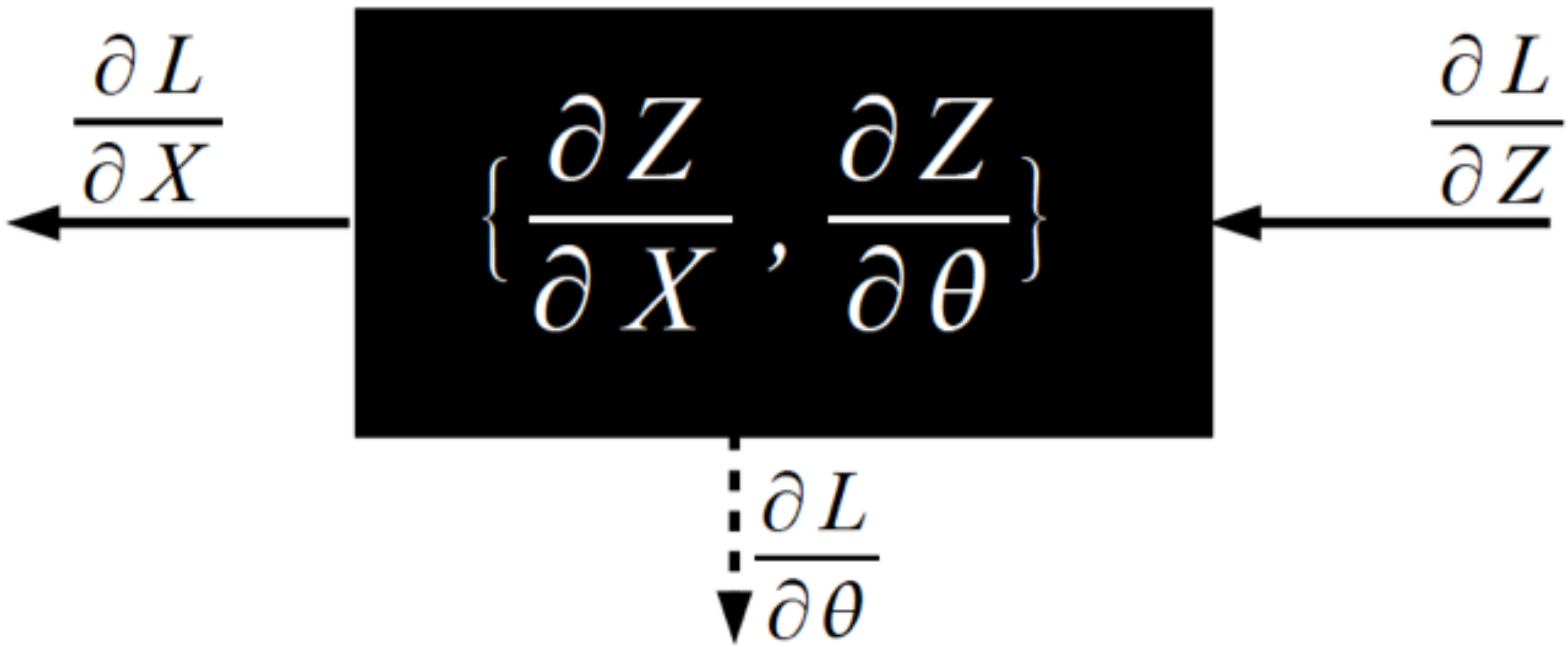


Key Computation: Forward-Prop

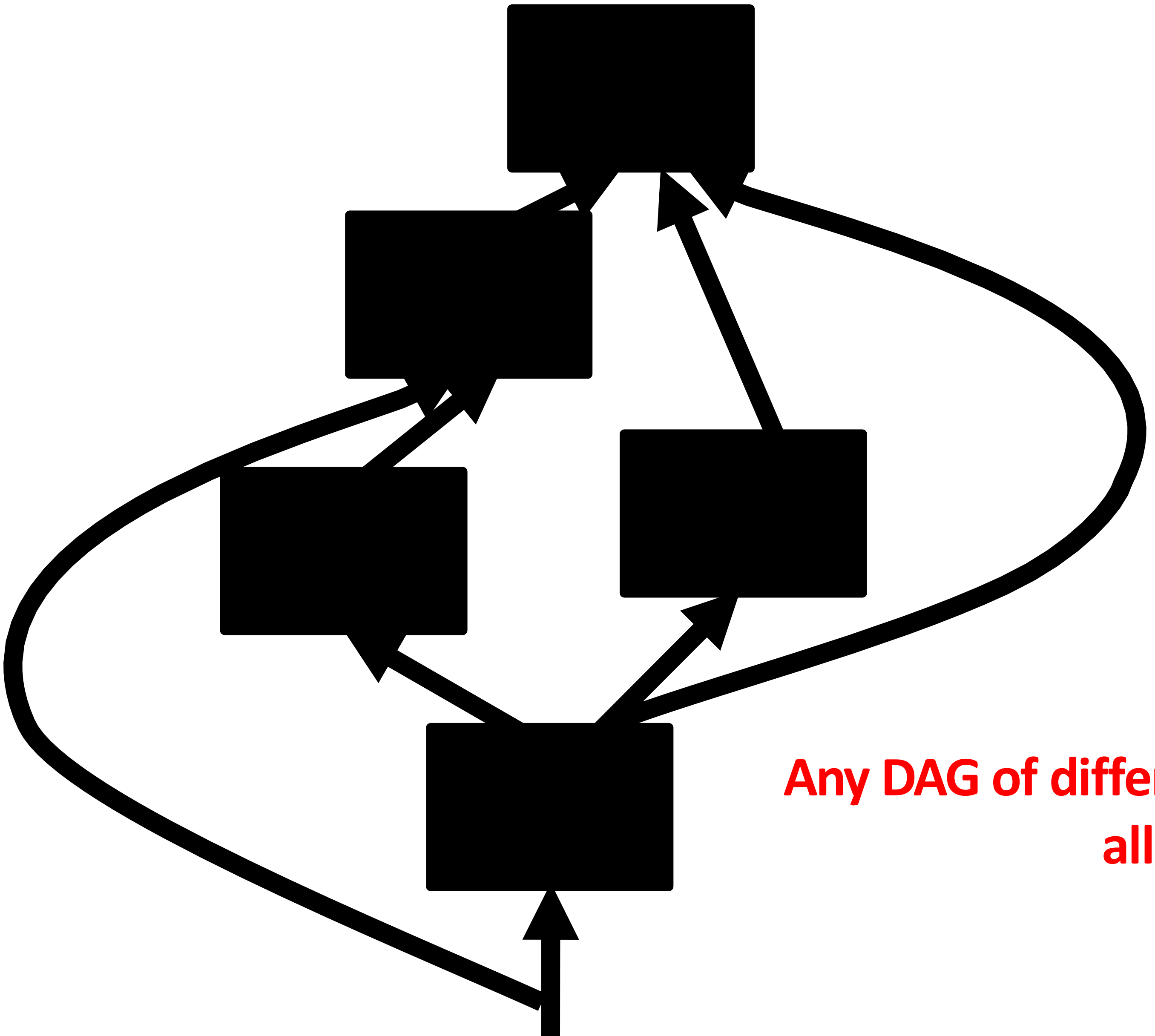


Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

Key Computation: Back-Prop



Differentiable Computation Graph



Any DAG of differentialble modules is allowed!

Visual Dialog Model #1

Image I

Late Fusion Encoder

Slide Credit: Abhishek Das

Visual Dialog Model #1

Image I

Do you think the woman is with him?

Question Q_t

Late Fusion Encoder

Slide Credit: Abhishek Das

Visual Dialog Model #1

Image I

Do you think the woman is with him?

Question Q_t

The man is riding his bicycle on the sidewalk. Is the man wearing a helmet? No he does not have a helmet on. ... Are there any people nearby? Yes there's a woman walking behind him.

t rounds of history
(concatenated)

Late Fusion Encoder

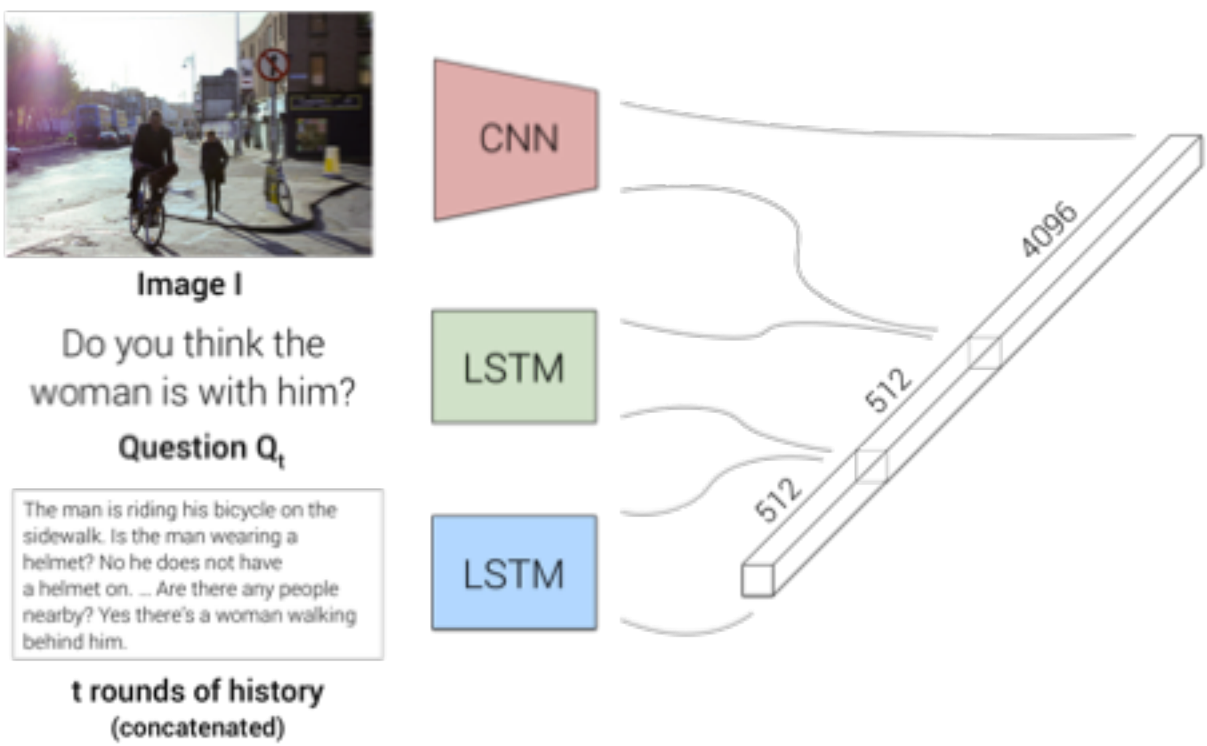
Slide Credit: Abhishek Das

Visual Dialog Model #1

Late Fusion Encoder

Slide Credit: Abhishek Das

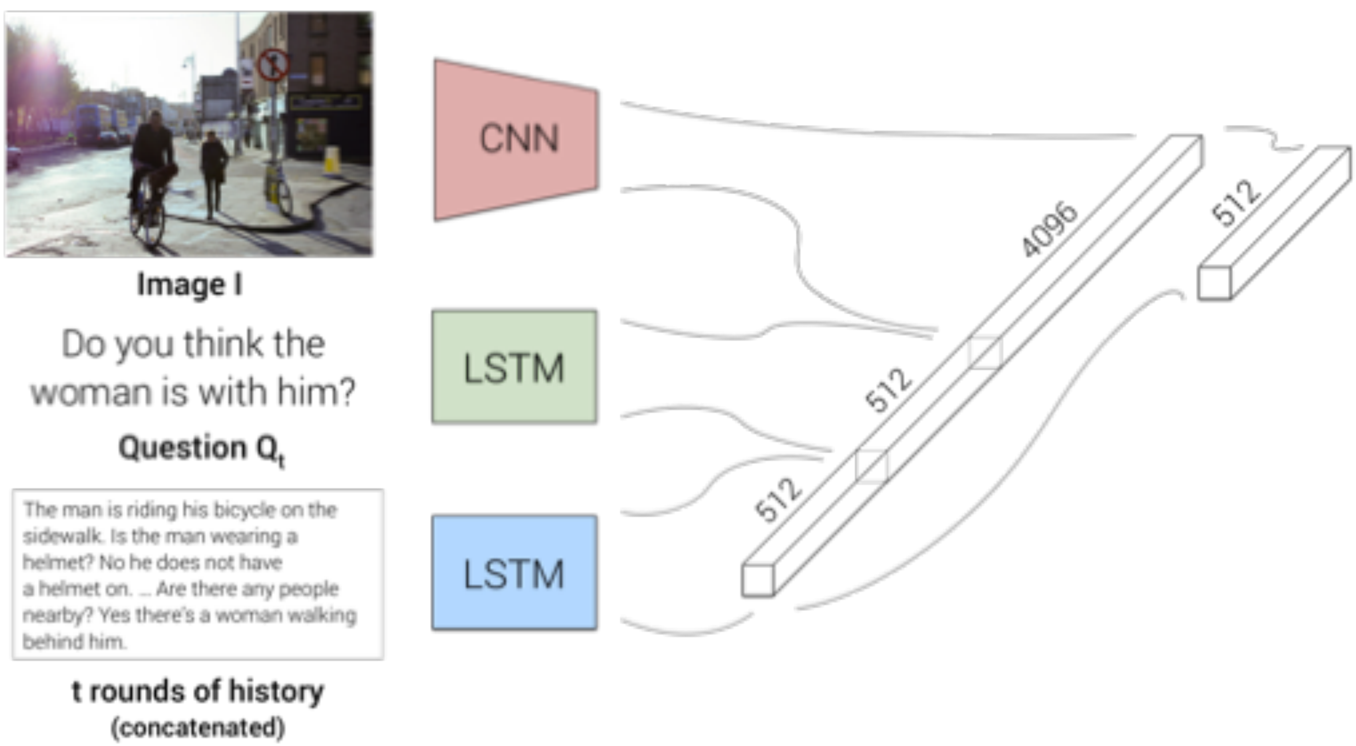
Visual Dialog Model #1



Late Fusion Encoder

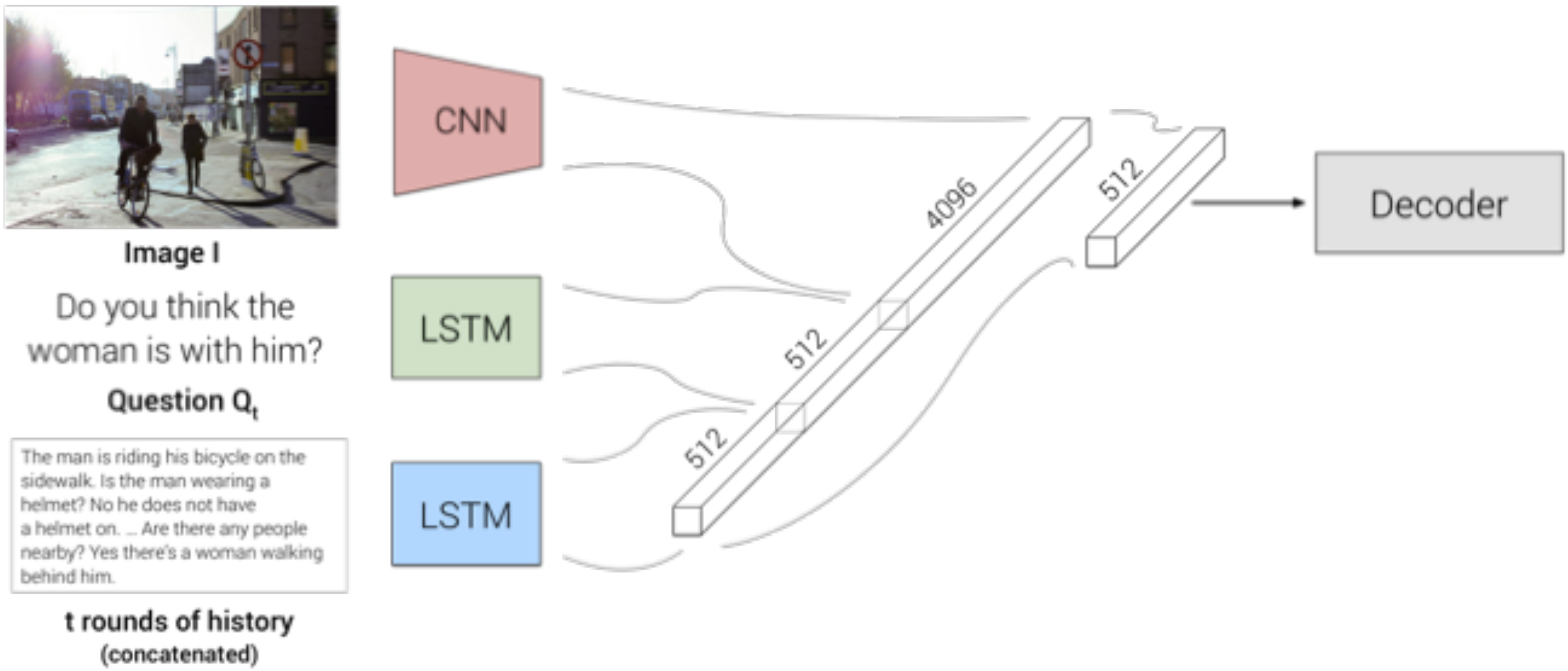
Slide Credit: Abhishek Das

Visual Dialog Model #1



Late Fusion Encoder

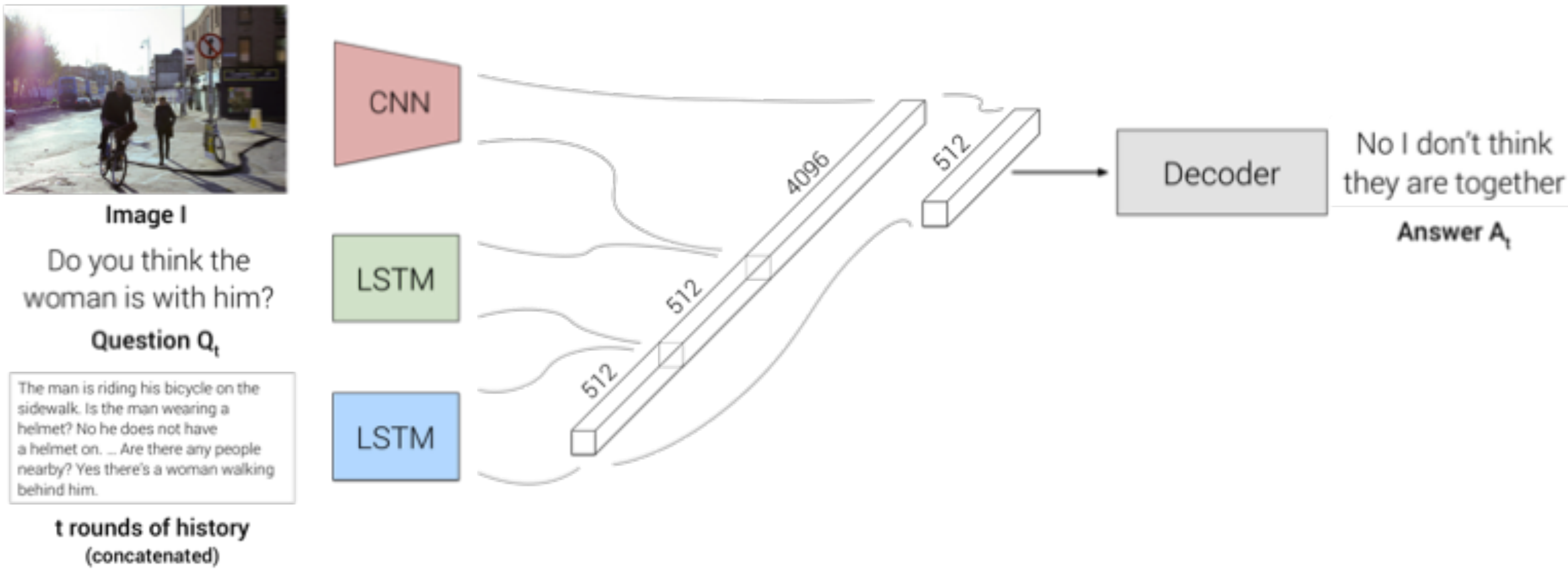
Visual Dialog Model #1



Late Fusion Encoder

Slide Credit: Abhishek Das

Visual Dialog Model #1



Late Fusion Encoder

Slide Credit: Abhishek Das

Problems with Deep Learning

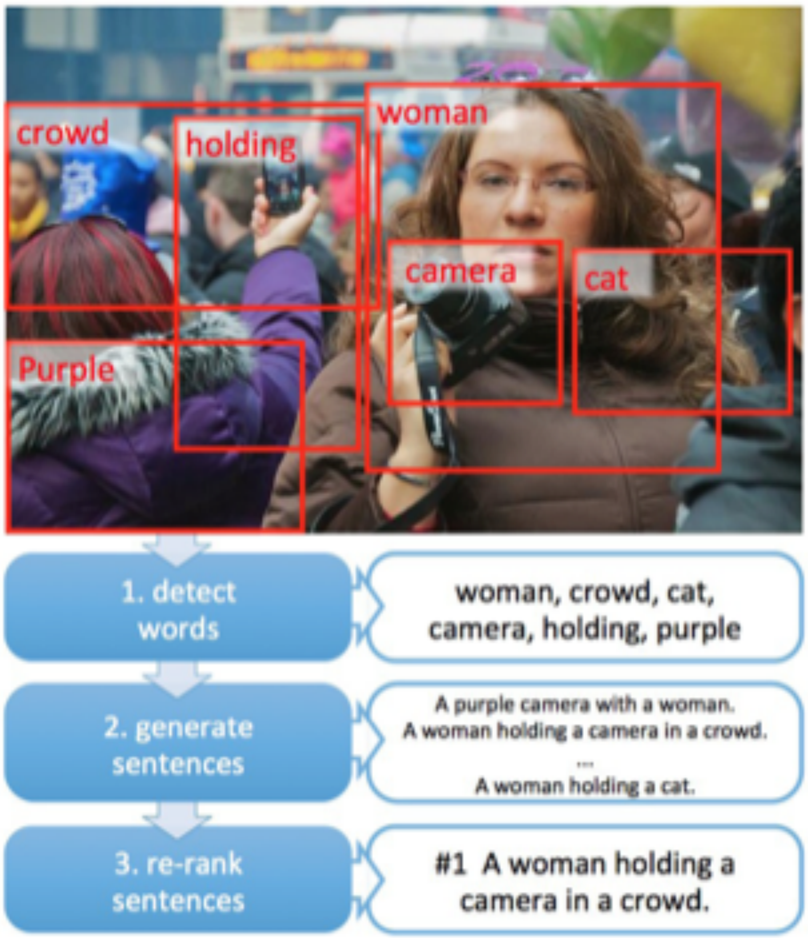
- Problem#1: Non-Convex! Non-Convex! Non-Convex!
  - Depth>=3: most losses non-convex in parameters
  - Theoretically, all bets are off
  - Leads to stochasticity
    - different initializations → different local minima
- Standard response #1
  - "Yes, but all interesting learning problems are non-convex"
  - For example, human learning
    - Order matters → wave hands → non-convexity
- Standard response #2
  - "Yes, but it often works!"

Problems with Deep Learning

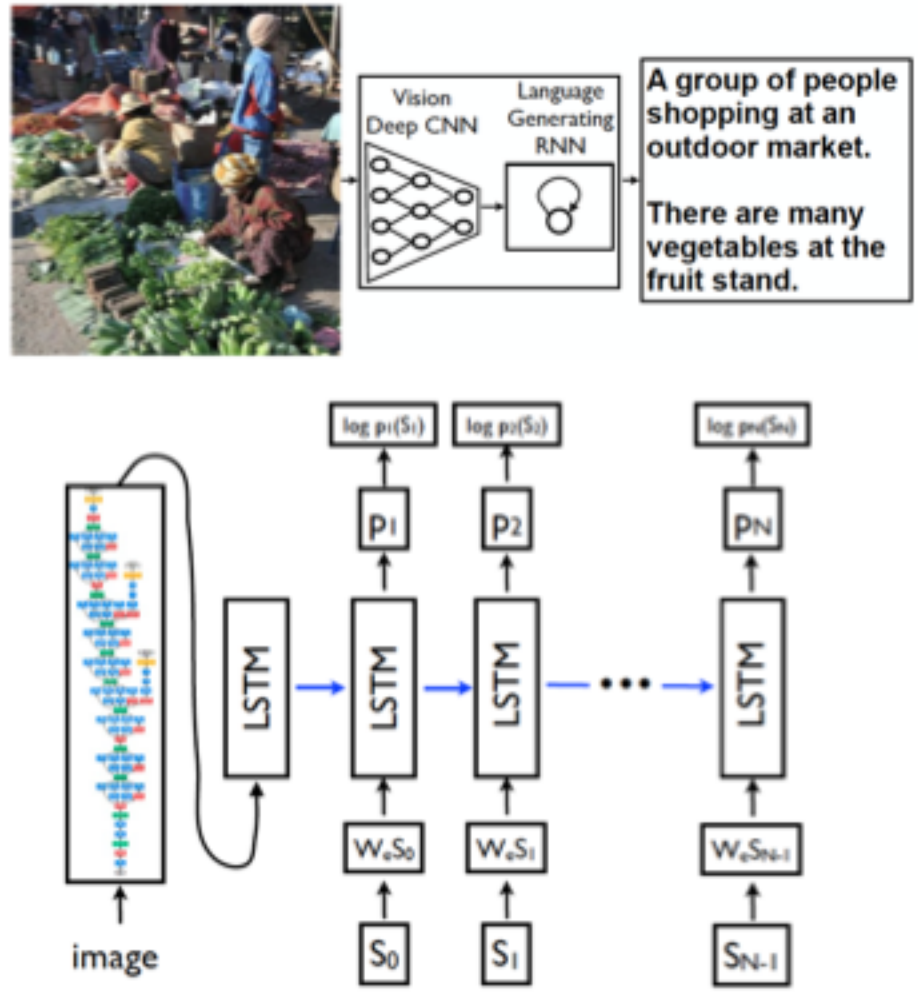
- Problem#2: Lack of interpretability
  - Hard to track down what's failing
  - Pipeline systems have "oracle" performances at each step
  - In end-to-end systems, it's hard to know why things are not working

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!"

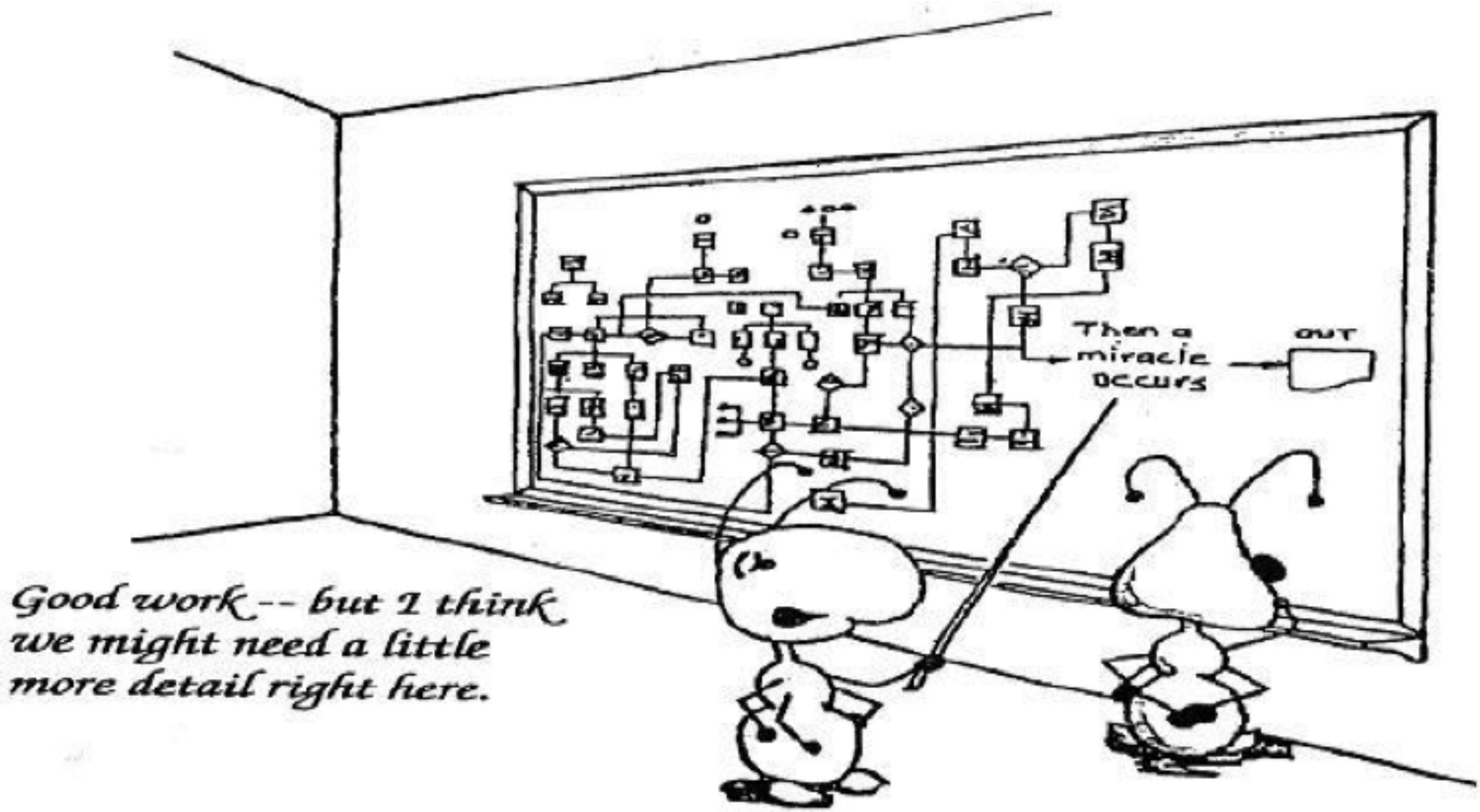
Problems with Deep Learning

- Problem#3: Lack of easy reproducibility
  - Direct consequence of stochasticity & non-convexity

- Standard response #1
  - It's getting much better
  - Standard toolkits/libraries/frameworks now available
  - Caffe, Theano, (Py)Torch

- 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?

What was F17 DL class about?

- Firehose of arxiv

Arxiv Fire Hose

PhD Student

Deep Learning papers

Cornell University Library

arXiv.org

What was F17 DL class about?

- Goal:
  – After taking this class, you should be able to pick up the latest Arxiv paper, easily understand it, & implement it.

- Target Audience:
  – Junior/Senior PhD students who want to *conduct research and publish in Deep Learning.*

    (think ICLR/CVPR papers as outcomes)

What is the F18 DL 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
    - 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…

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, C++!
  - Libraries/Frameworks: PyTorch
  - HW0 (pure python), HW1 (python + PyTorch), HW2+3 (PyTorch)
  - Your language of choice for project

Course Information

- Instructor: Dhruv Batra
  - dbatra@gatech
  - Location: 219 CCB

Machine Learning & Perception Group

Dhruv Batra
Assistant Professor

PhD

Qing Sun
(2012 – Present)

Aishwarya Agrawal
(2014 – Present)

Yash Goyal
(2014 – Present)

Michael Cogswell
(2015 – Present)

Abhishek Das
(2016 – Present)

Ashwin Kalyan
(2016 – Present)

Nirbhay Modhe
(2017 – Present)

Research Scientist

Stefan Lee

MS

Akrit Mohapatra

Deshraj Yadav

TAs

Michael Cogswell

3^{rd} year CS PhD student

http://mcogswell.io/

Erik Wijmans

2^{nd} year CS PhD student

http://wijmans.xyz/

Nirbhay Modhe

2^{nd} year CS PhD student

https://nirbhayjm.github.io/

Harsh Agrawal

1^{st} year CS PhD student

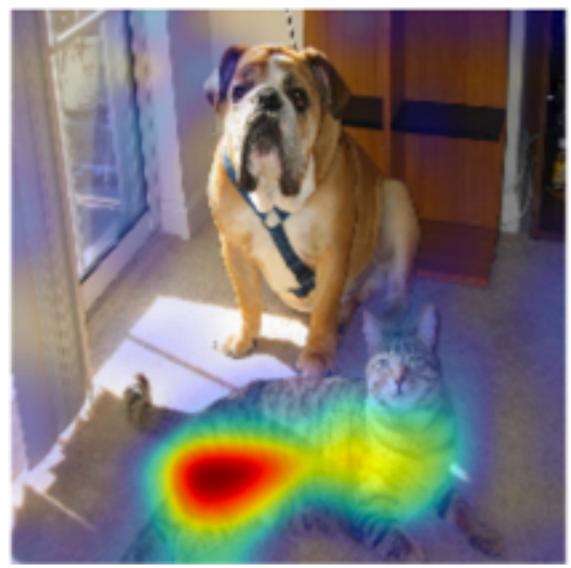
https://dexter1691.github.io/

TA: Michael Cogswell

- PhD student working with Dhruv

- Research work/interest:
  - Deep Learning
  - applications to Computer Vision and AI

$$\mathcal{L}_{\text{DeCov}}$$



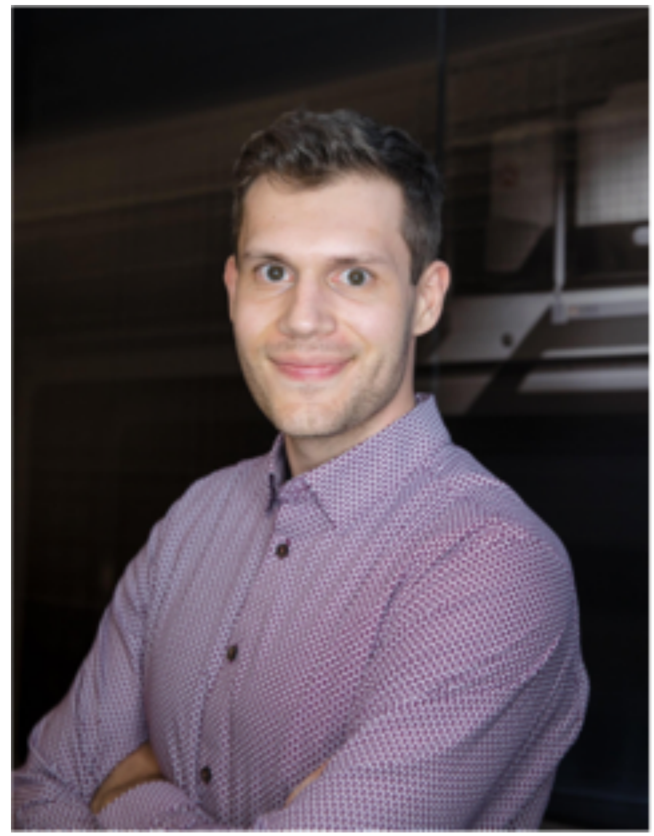
- I also Fence (mainly foil)

TA: Erik Wijmans

PhD student in CS

Research Interests

- Scene Understanding
- Embodied Agents
- 3D Computer Vision



TA: Nirbhay Modhe

2nd Year PhD Student

Research Interests:

- Visual Dialog
- Bayesian Machine Learning
- Generative Modeling

TA: Harsh Agrawal

- 1st year CS PhD student
- Previously at Snapchat Research
- Research at the intersection of vision and language

Sorting jumbled story elements into coherent story

Adding new logos without re-training the model

Organization & Deliverables

- 4 homeworks (80%)
  - Mix of theory and implementation
  - 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 (5%)
  - Contribute to class discussions on Piazza
  - Ask questions, answer questions

Late Days

- “Free” Late Days
  - 7 late days for the semester
    - Use for HWs
    - Cannot use for project related deadlines
  - After free late days are used up:
    - 25% penalty for each late day

HW0

- Out today; due Sept 5 (09/05)
  - Available on class webpage + Canvas

- Grading
  - <=80% means that you might not be prepared for the class

- Topics
  - PS: probability, calculus, convexity, proving things
  - HW: Implement training of a soft-max classifier via SGD

Project

- Goal
  - Chance to try Deep Learning
  - Encouraged to apply to your research (computer vision, NLP, robotics,…)
  - Must be done this semester.
  - Can combine with other classes
    - get permission from both instructors; delineate different parts
  - Extra credit for shooting for a publication

- Main categories
  - Application/Survey
    - Compare a bunch of existing algorithms on a new application domain of your interest
  - Formulation/Development
    - Formulate a new model or algorithm for a new or old problem
  - Theory
    - Theoretically analyze an existing algorithm

Computing

- Major bottleneck
  – GPUs

- Options
  – Your own / group / advisor's resources

  – Google Cloud Credits
    - $50 credits to every registered student courtesy Google

  – Minsky cluster in IC

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 HW0. Come to first few classes.
  - Hope people drop.
- Audit or Pass/Fail
  - We will give preference to people taking class for credit.
- Sitting in
  - Talk to instructor.

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.

Collaboration Policy

- Collaboration
  - Only on HWs and project (not allowed in HW0).
  - 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.

Communication Channels

- 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.

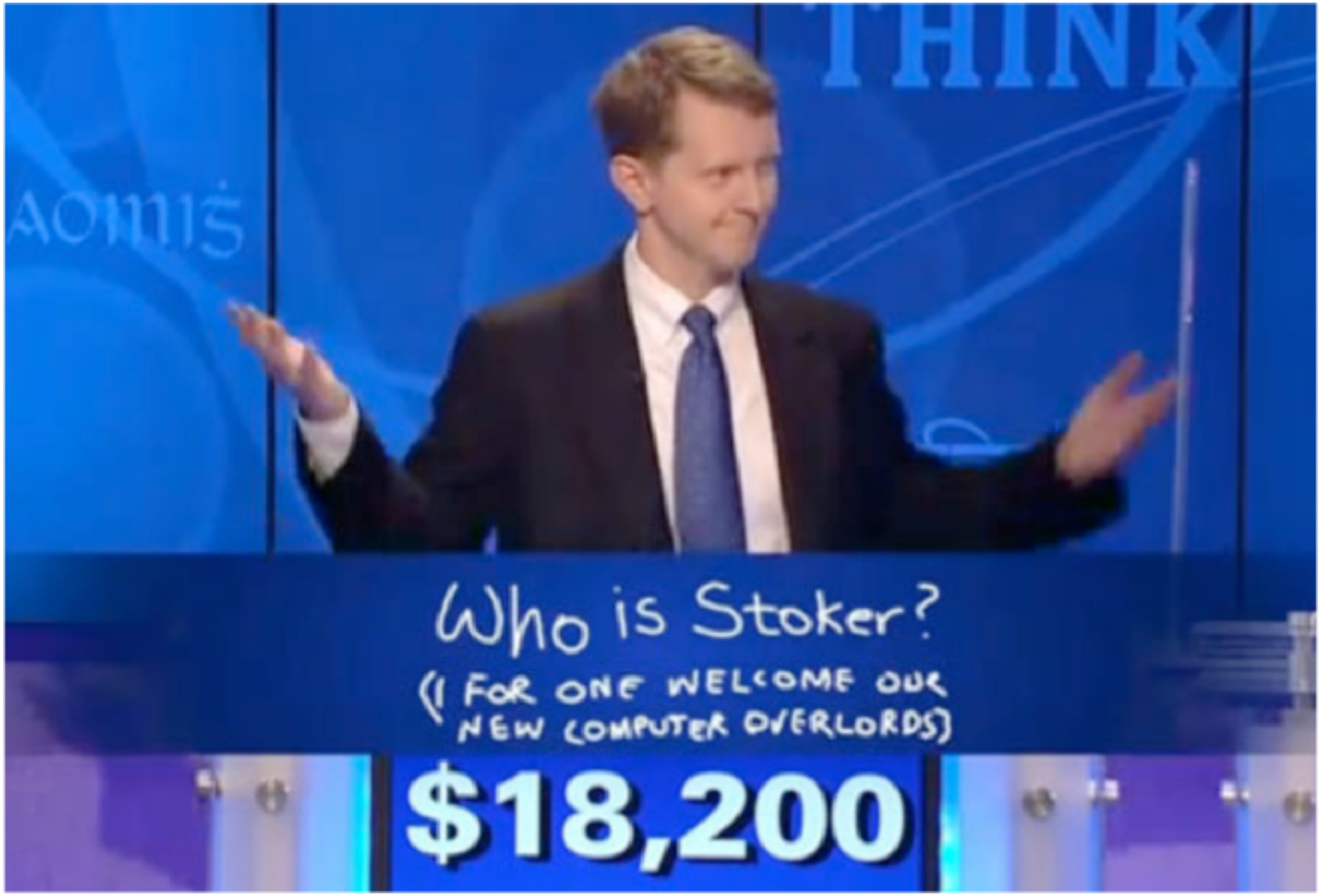
- Staff Mailing List
  - cs4803-7643-f18-staff@googlegroups.com

- Links:
  - Website: www.cc.gatech.edu/classes/AY2019/cs7643_fall/
  - Piazza: piazza.com/gatech/fall2018/cs48037643
  - Canvas: gatech.instructure.com/courses/28059
  - Gradescope: gradescope.com/courses/22096

Todo

- HW0
  - Due Wed Sept 5 11:55pm

Welcome


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