CS 7643: Deep Learning

www.cc.gatech.edu/classes/AY2018/cs7643_fall/

piazza.com/gatech/fall2017/cs7643

Canvas: gatech.instructure.com/courses/772

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

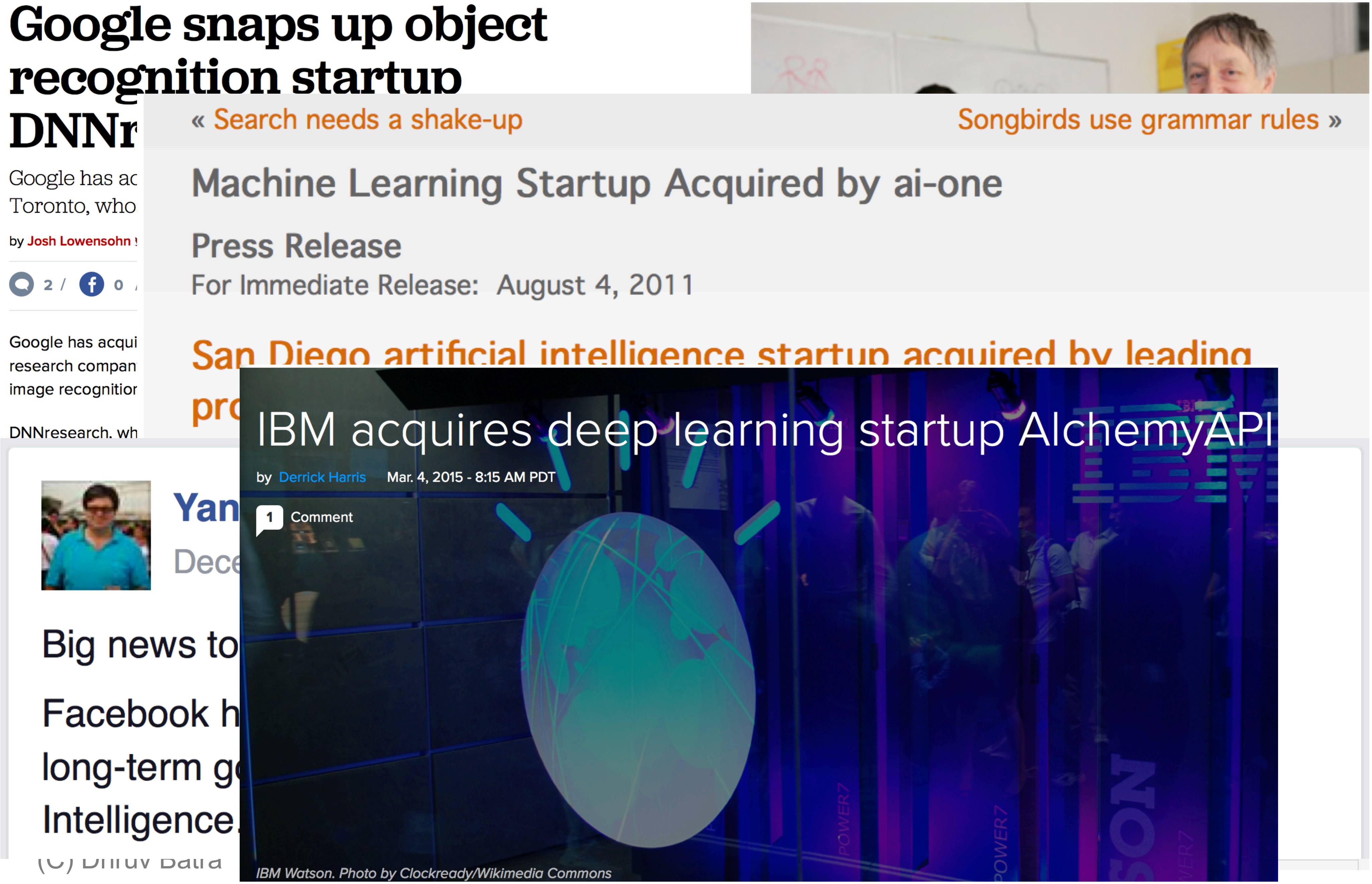
What is this class about?

Some of the most exciting developments in

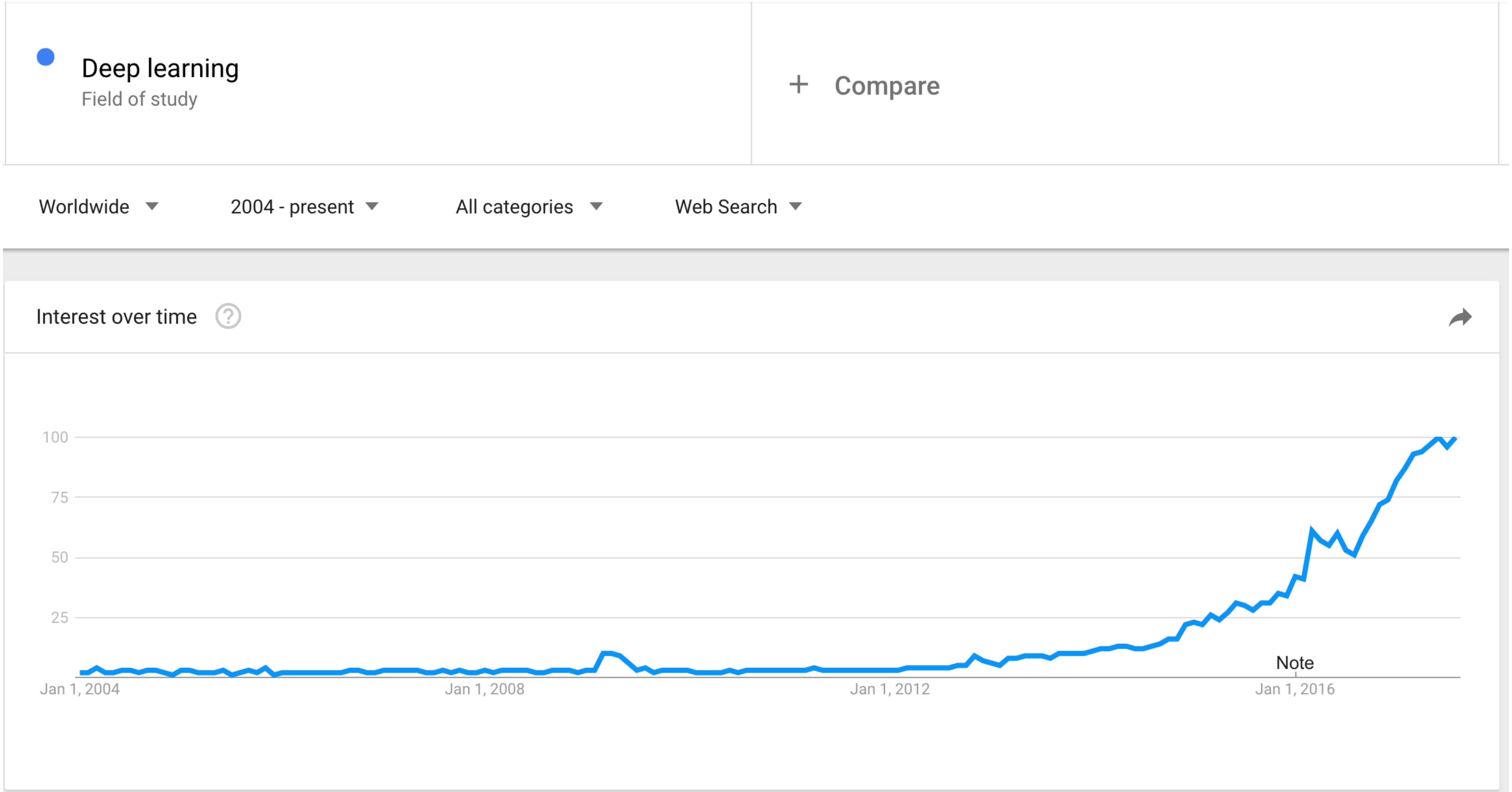
Machine Learning, Vision, NLP, Speech, Robotics & AI in general

in the last 5 years!

Acquisitions



Proxy for public interest



Microsoft researchers achieve new conversational speech recognition milestone

August 20, 2017 | Posted by Microsoft Research Blog

Share Tweet Share

By Xuedong Huang, Technical Fellow, Microsoft

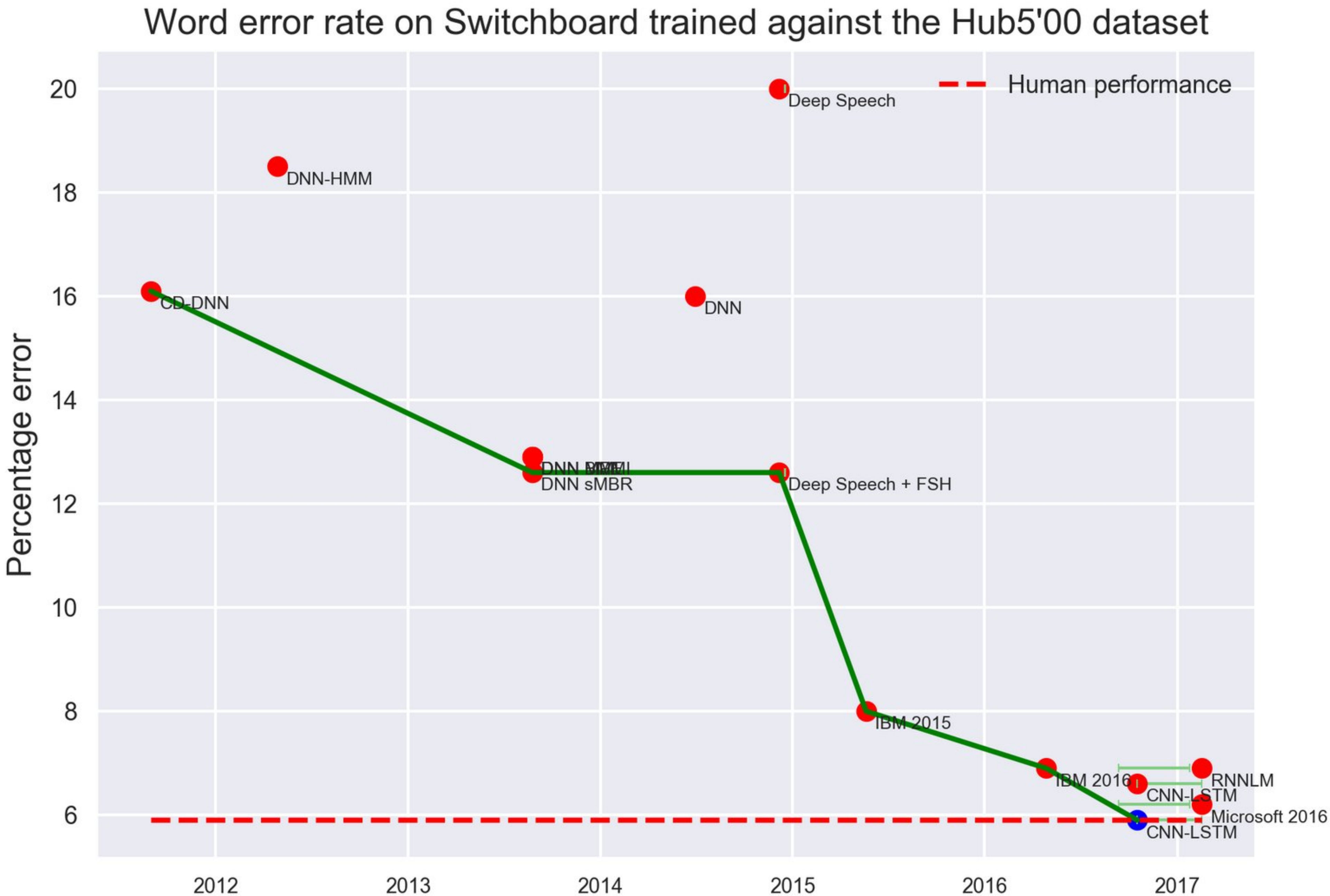
Last year, Microsoft's speech and dialog research group announced a milestone in reaching human parity on the Switchboard conversational speech recognition task, meaning we had created technology that recognized words in a conversation as well as professional human transcribers.

After our transcription system reached the 5.9 percent word error rate that we had measured for humans, other researchers conducted their own study, employing a more involved multi-transcriber process, which yielded a 5.1 human parity word error rate. This was consistent with prior research that showed that humans achieve higher levels of agreement on the precise words spoken as they expend more care and effort. Today, I'm excited to announce that our research team reached that 5.1 percent error rate with our speech recognition system, a new industry milestone, substantially surpassing the accuracy we achieved last year. A technical report published this weekend documents the details of our system.

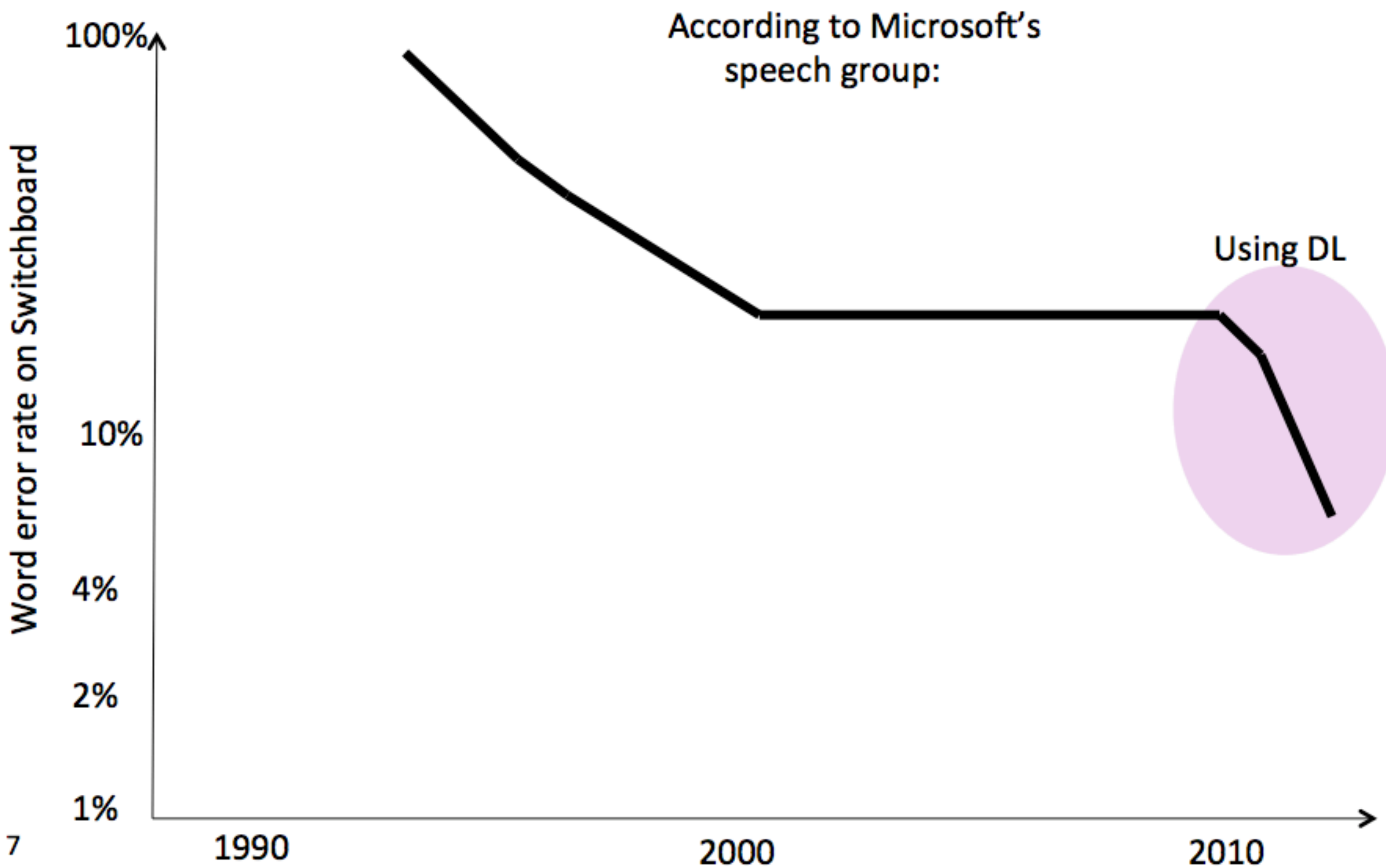
Chat

"Hey Cortana, tell me a joke."

Drop in WER


Word error rate on Switchboard trained against the Hub5'00 dataset
Human performance
Percentage error
20
18
16
14
12
10
8
6
2012
2013
2014
2015
2016
2017
Deep Speech
DNN-HMM
CD-DNN
DNN
DNN sMBR
Deep Speech + FSH
IBM 2015
IBM 2016
RNNLM
CNN-LSTM
Microsoft 2016
CNN-LSTM

Drop in WER



http://ptgmedia.pearsoncmg.com/images/art_sheil_namerecognition/elementLinks/art_sheil_namerecognition1_alt.png

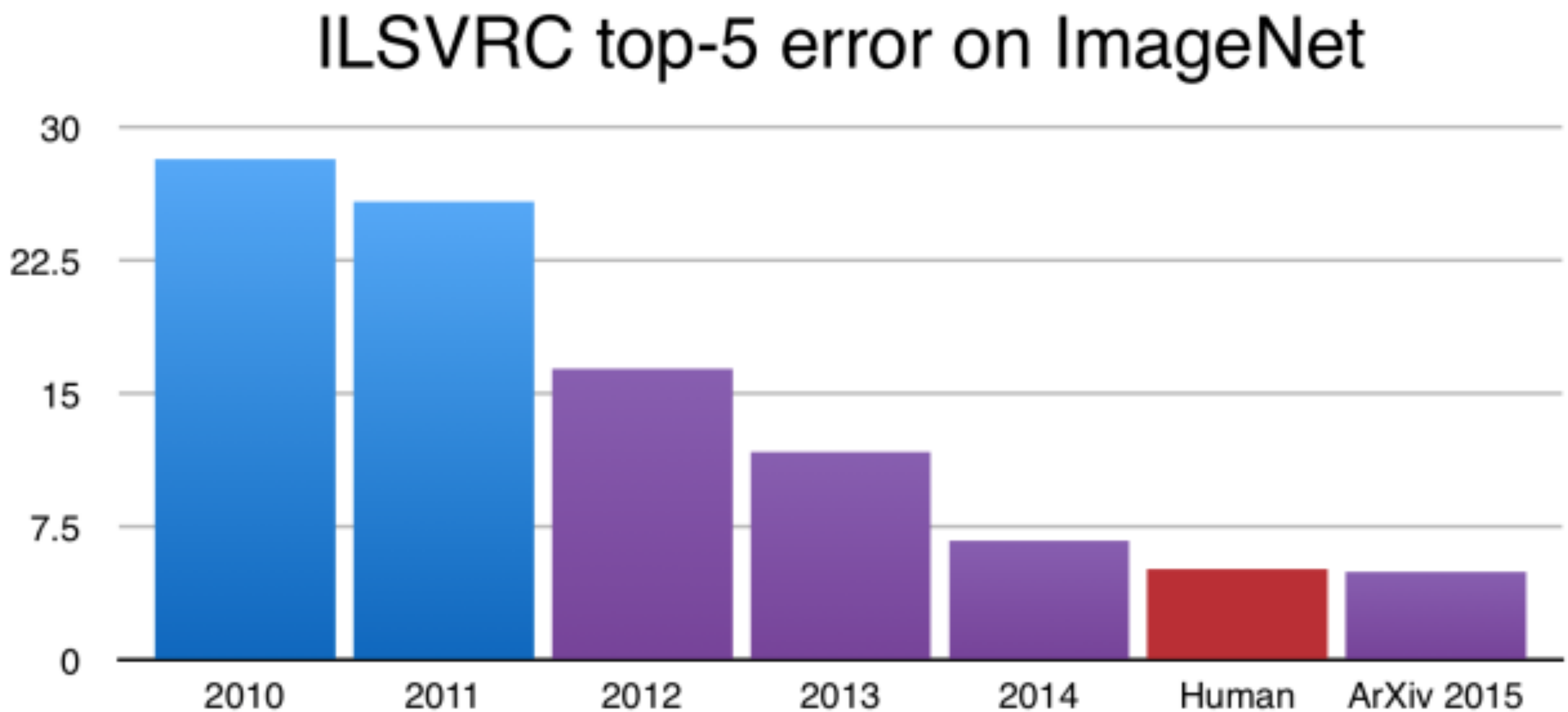
Image Classification

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



AlphaGo seals 4-1 victory over Go grandmaster Lee Sedol

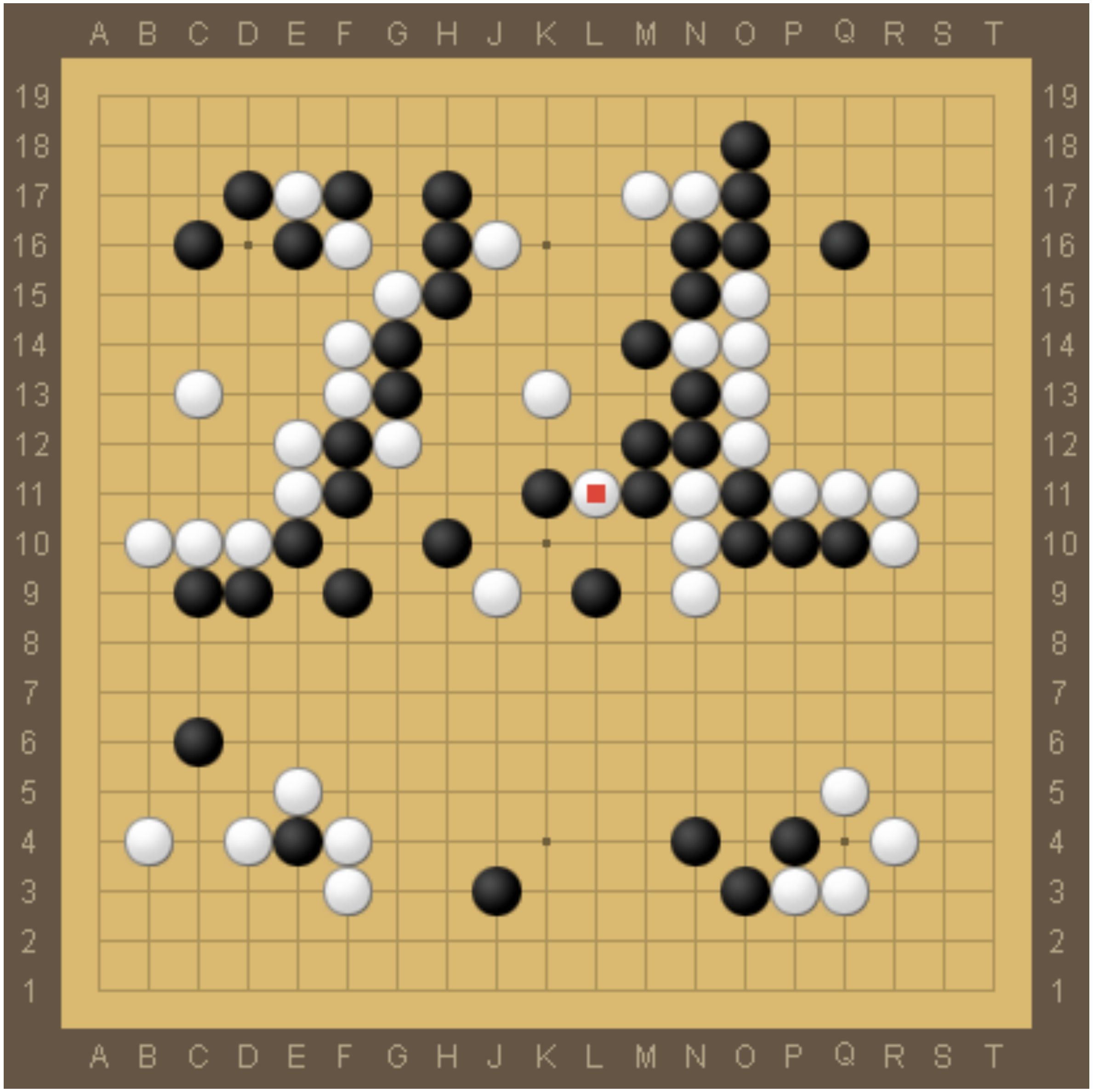
DeepMind's artificial intelligence astonishes fans to defeat human opponent and offers evidence computer software has mastered a major challenge

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.

AlphaGo vs Lee Sedol

- Match 4, Move 78



AlphaGo vs Ken Jie

FINAL DEFEAT

The awful frustration of a teenage Go champion playing Google's AlphaGo

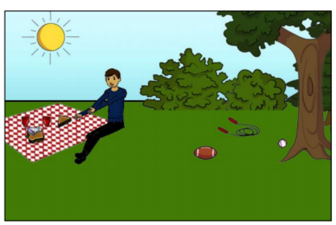
OUR PICKS | LATEST | POPULAR | QUARTZ | OBSESSIONS

Tasks are getting bolder

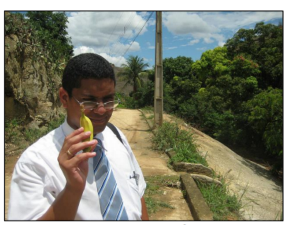
A group of young people playing a game of Frisbee
Vinyals et al., 2015

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

Image Captioning

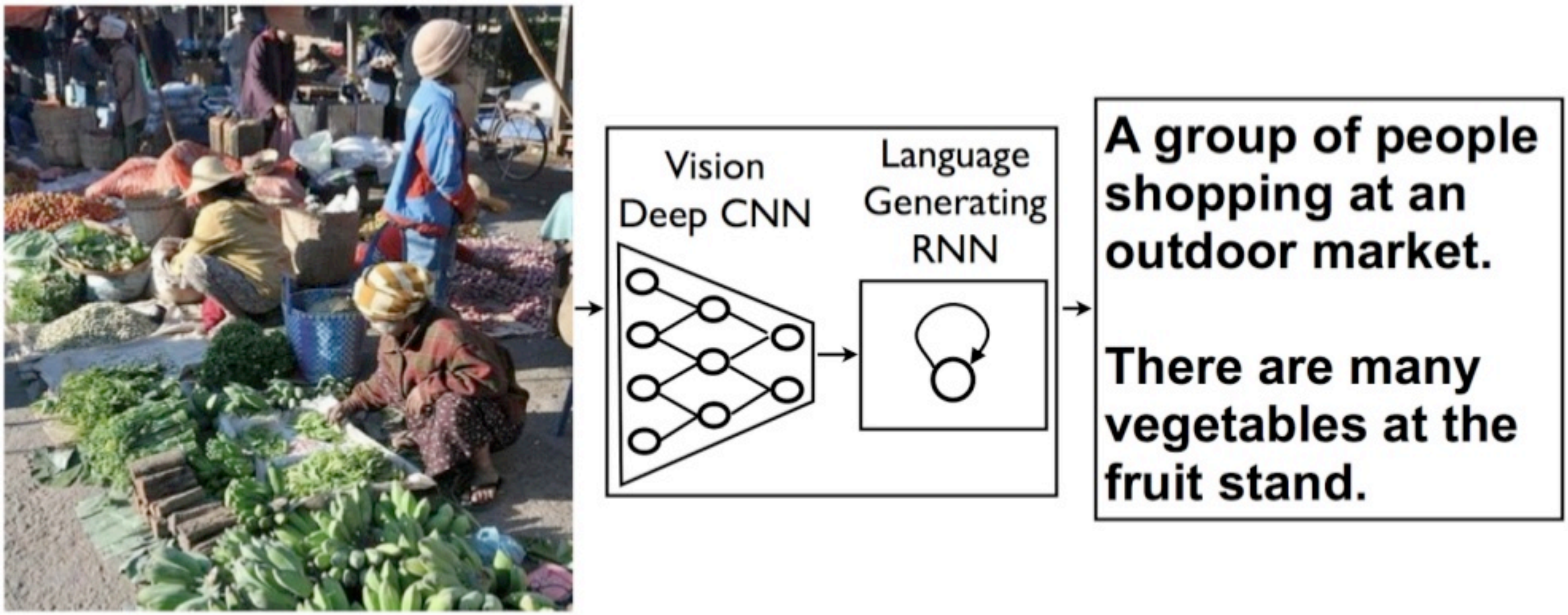
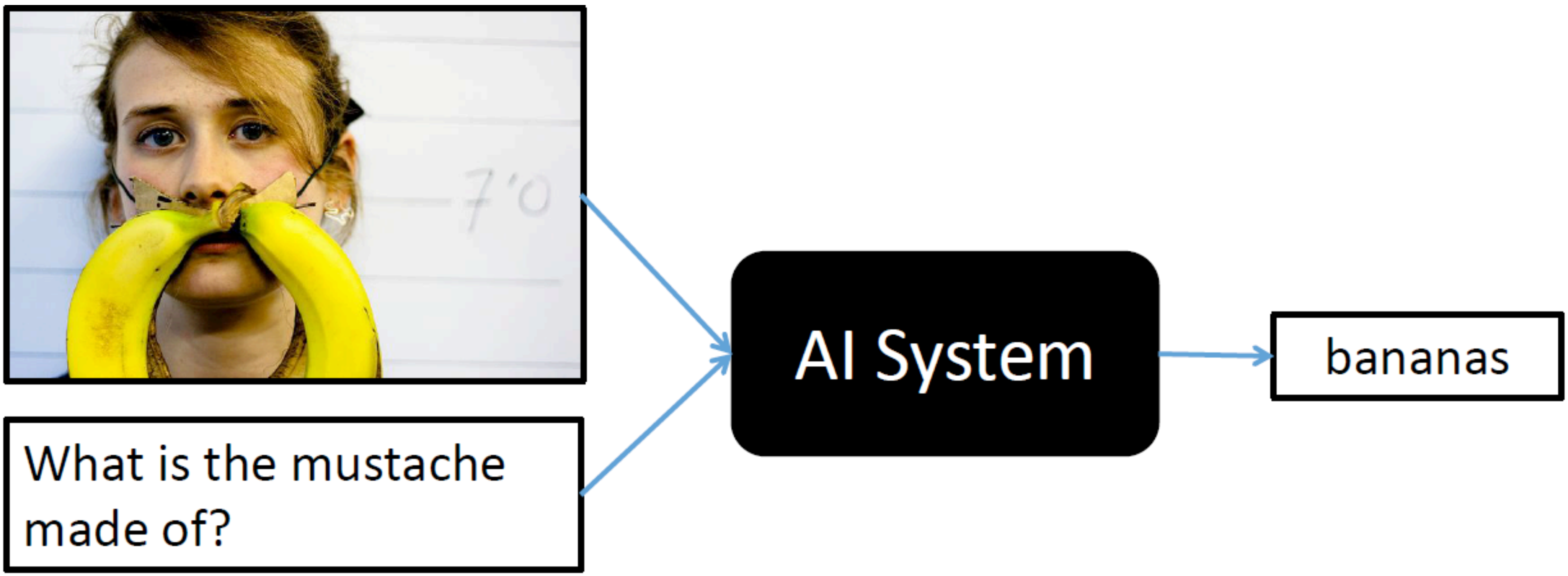
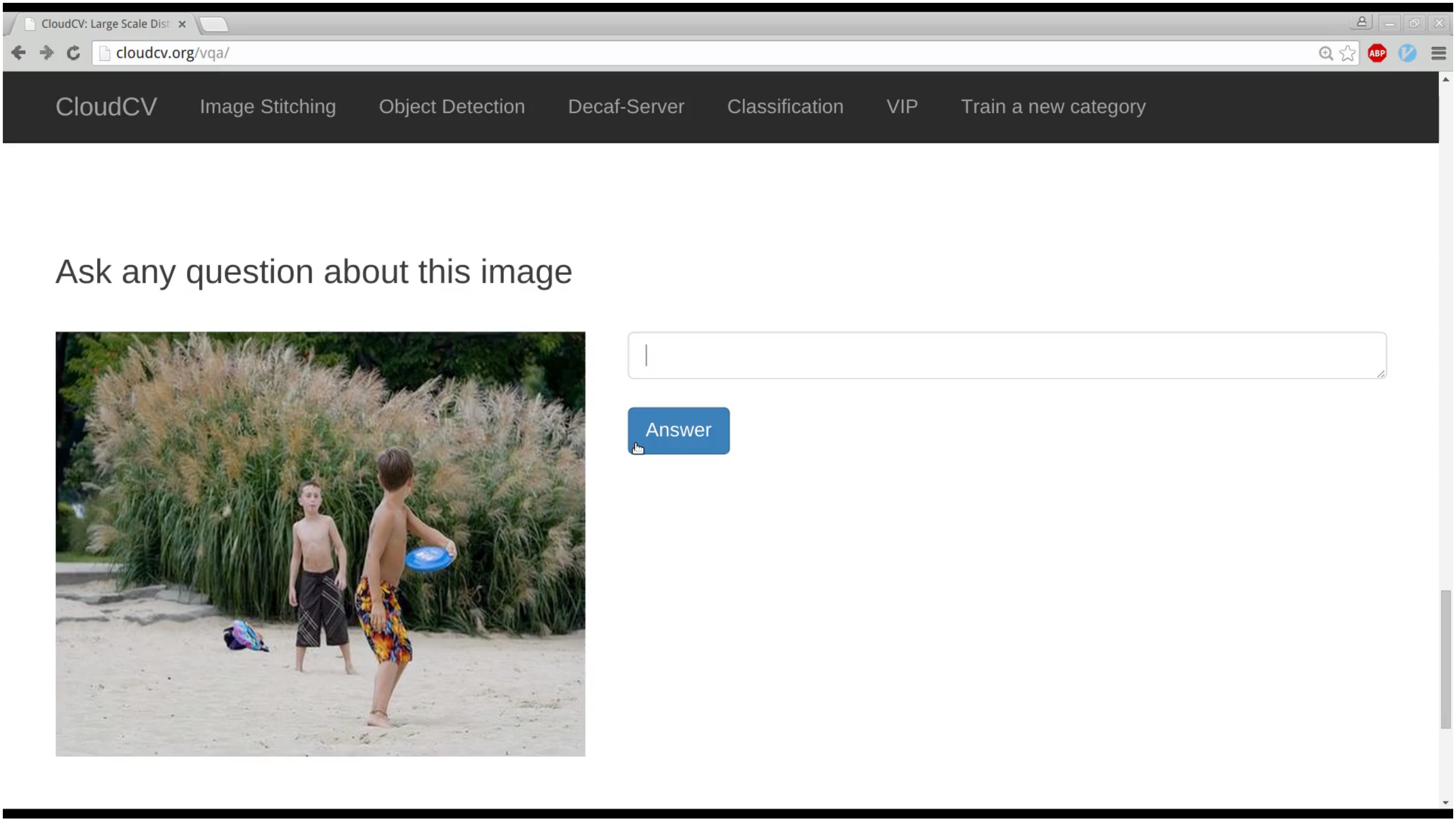


Image Credit: [Vinyals et al. CVPR15]

Visual Question Answering (VQA)



Visual Question Answering (VQA)



www.visualqa.org

Slide Credit: Stan Antol

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)

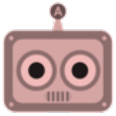
Devi Parikh
(Georgia Tech / FAIR)

Dhruv Batra
(Georgia Tech / FAIR)

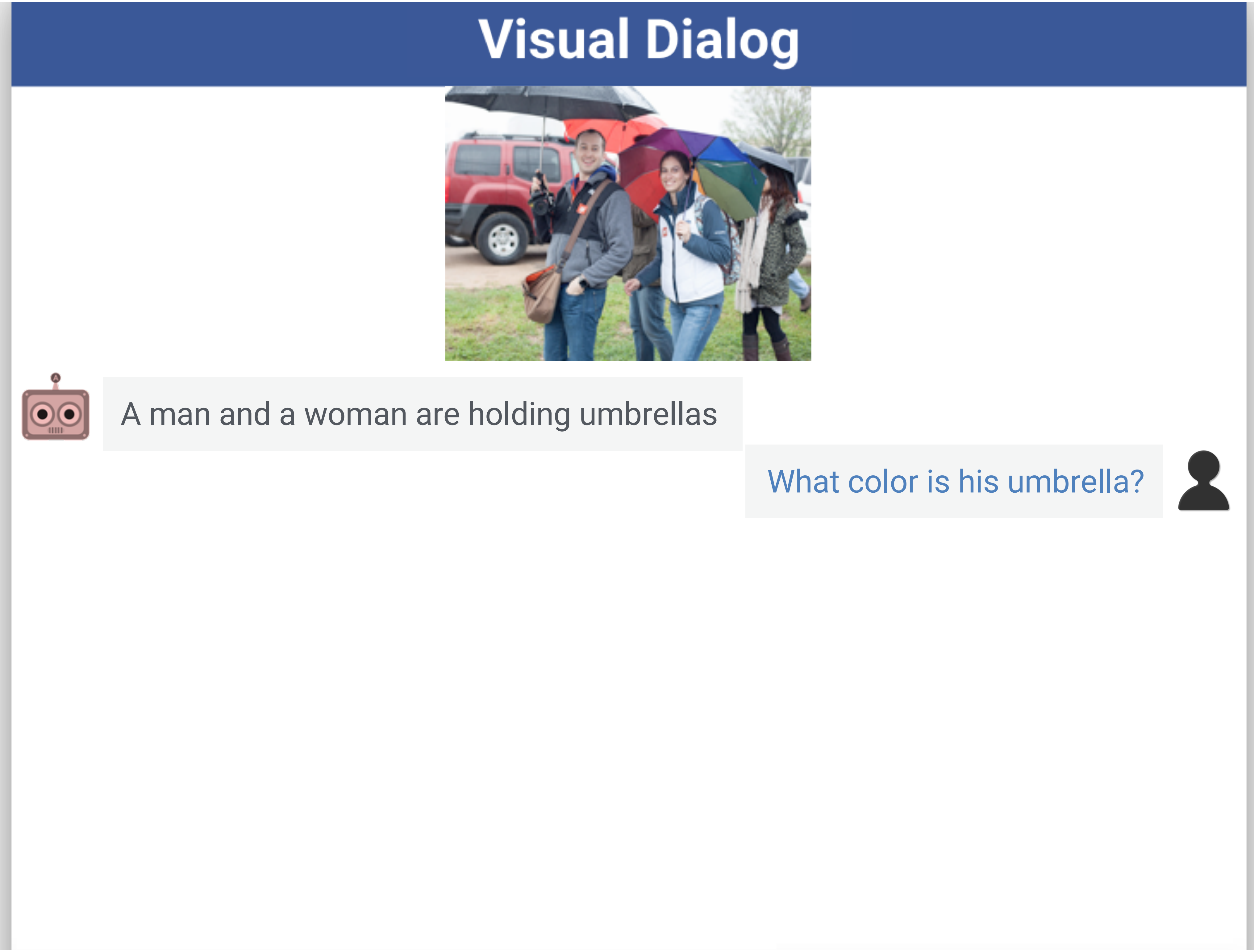
Visual Dialog

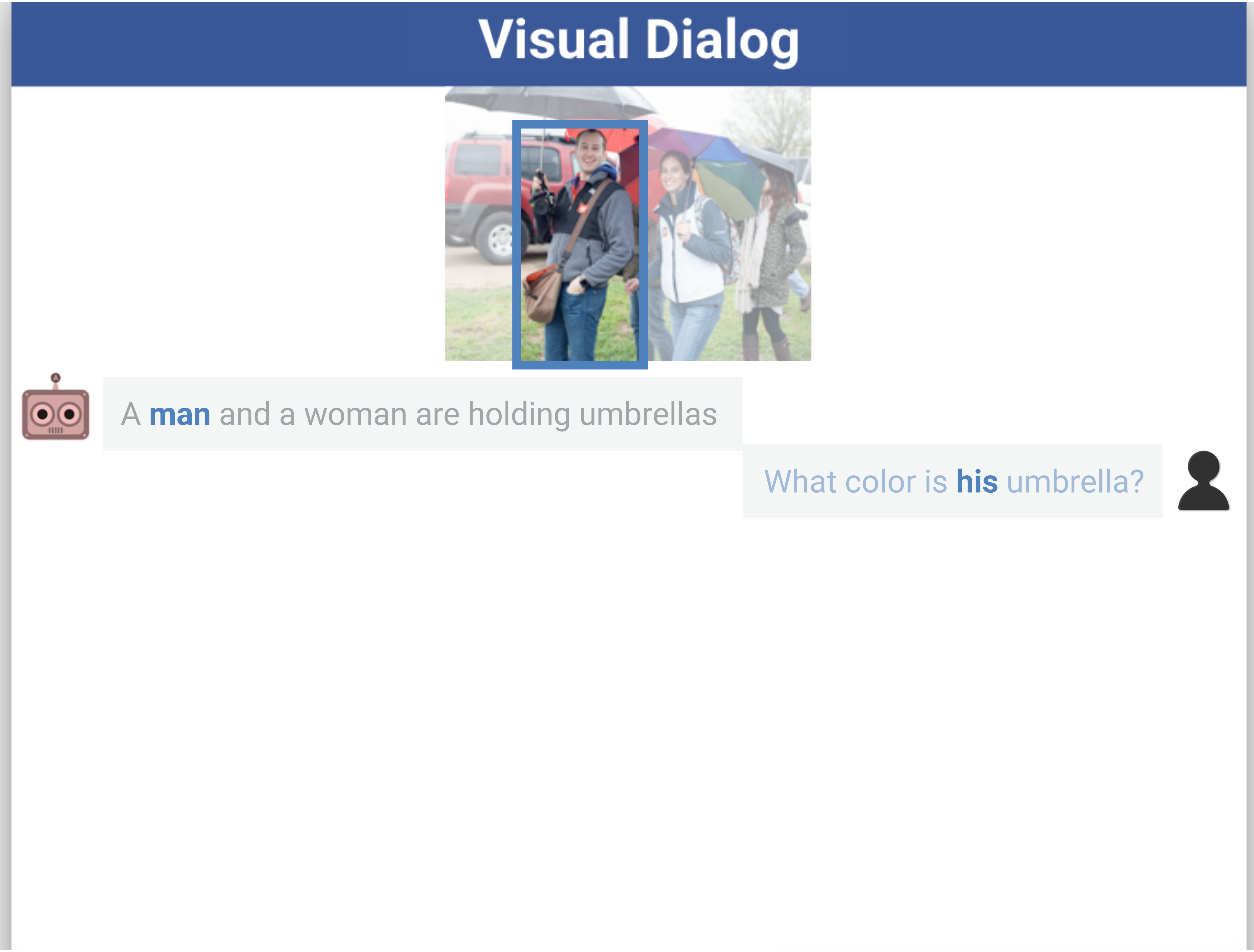
Visual Dialog

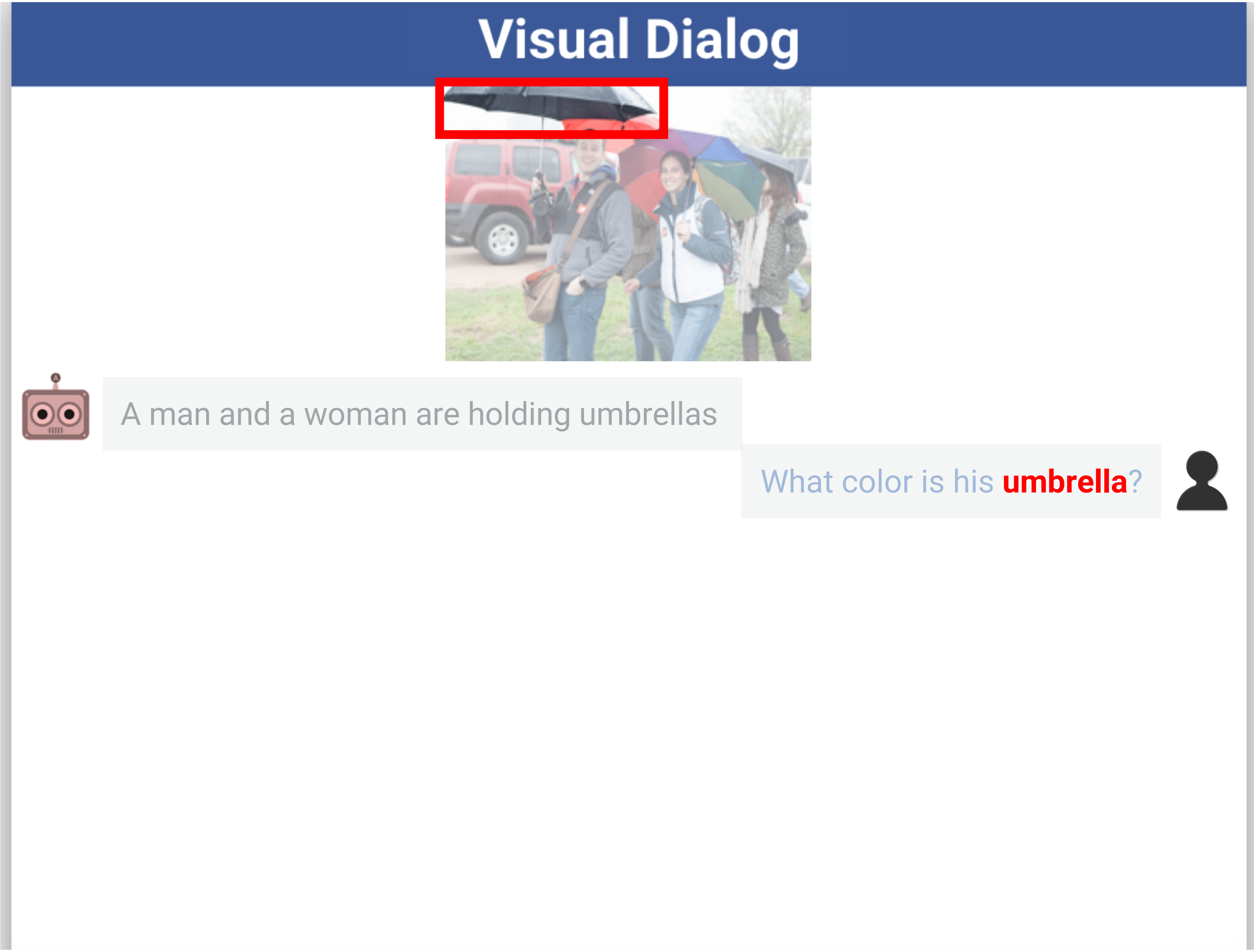
Visual Dialog

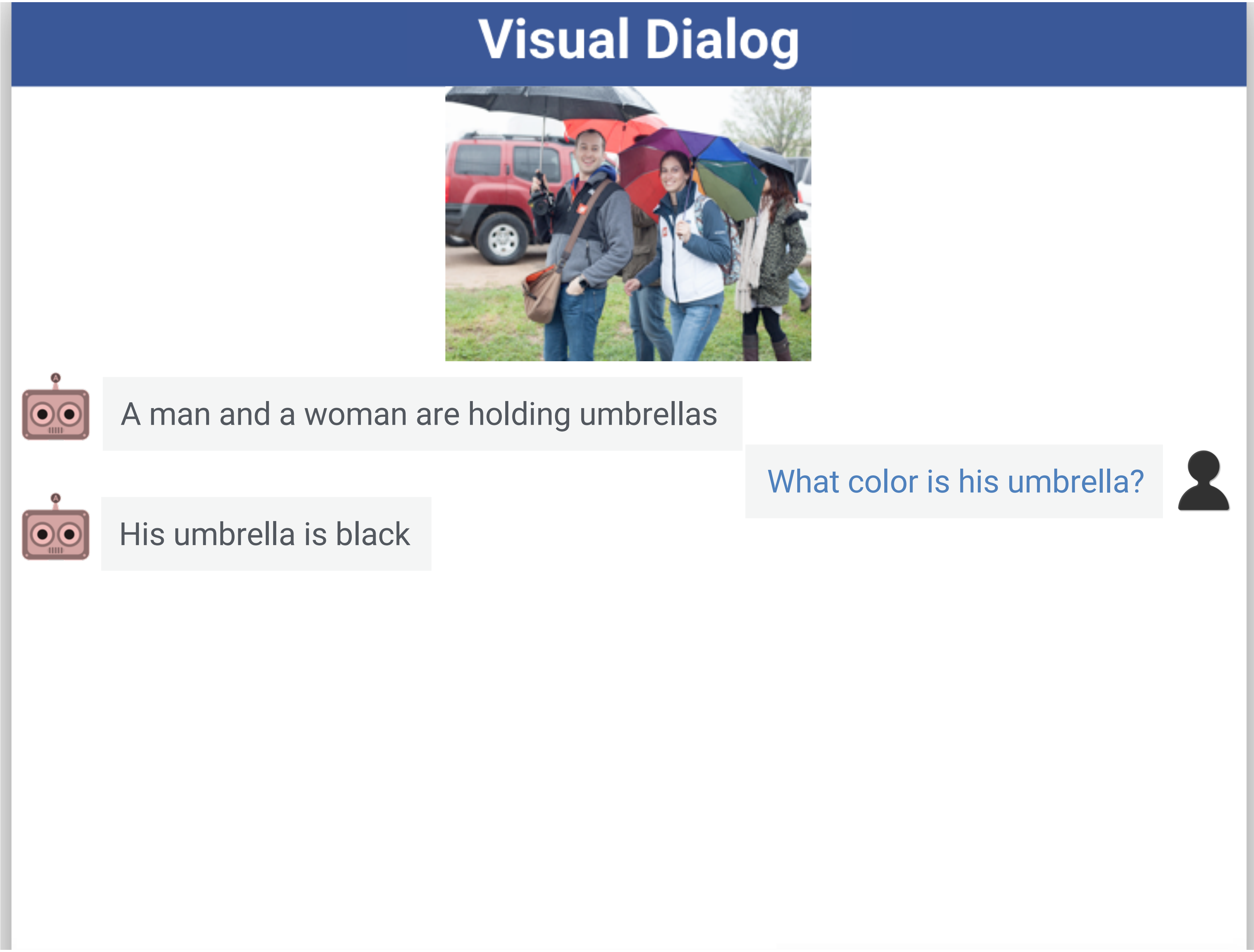


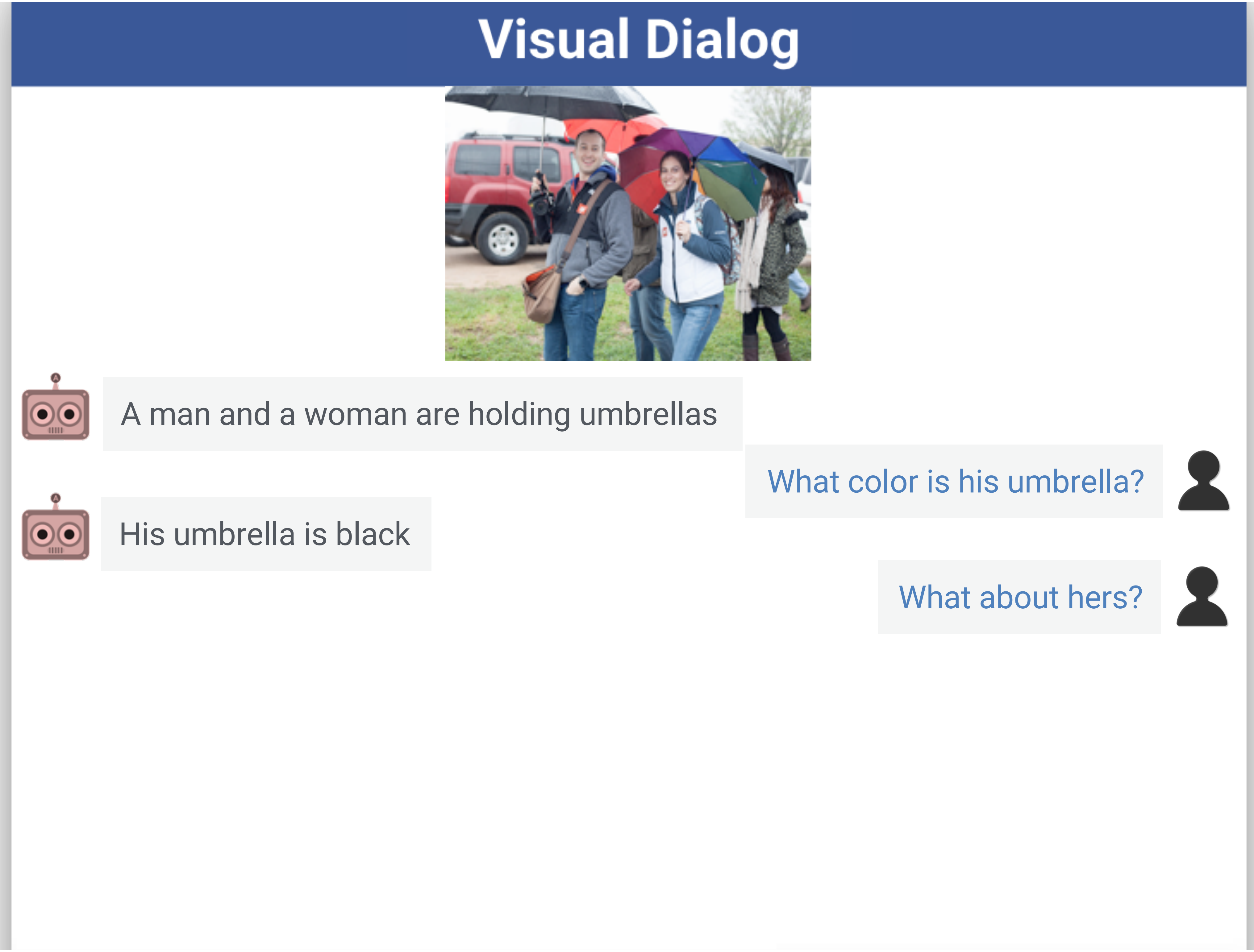
A man and a woman are holding umbrellas

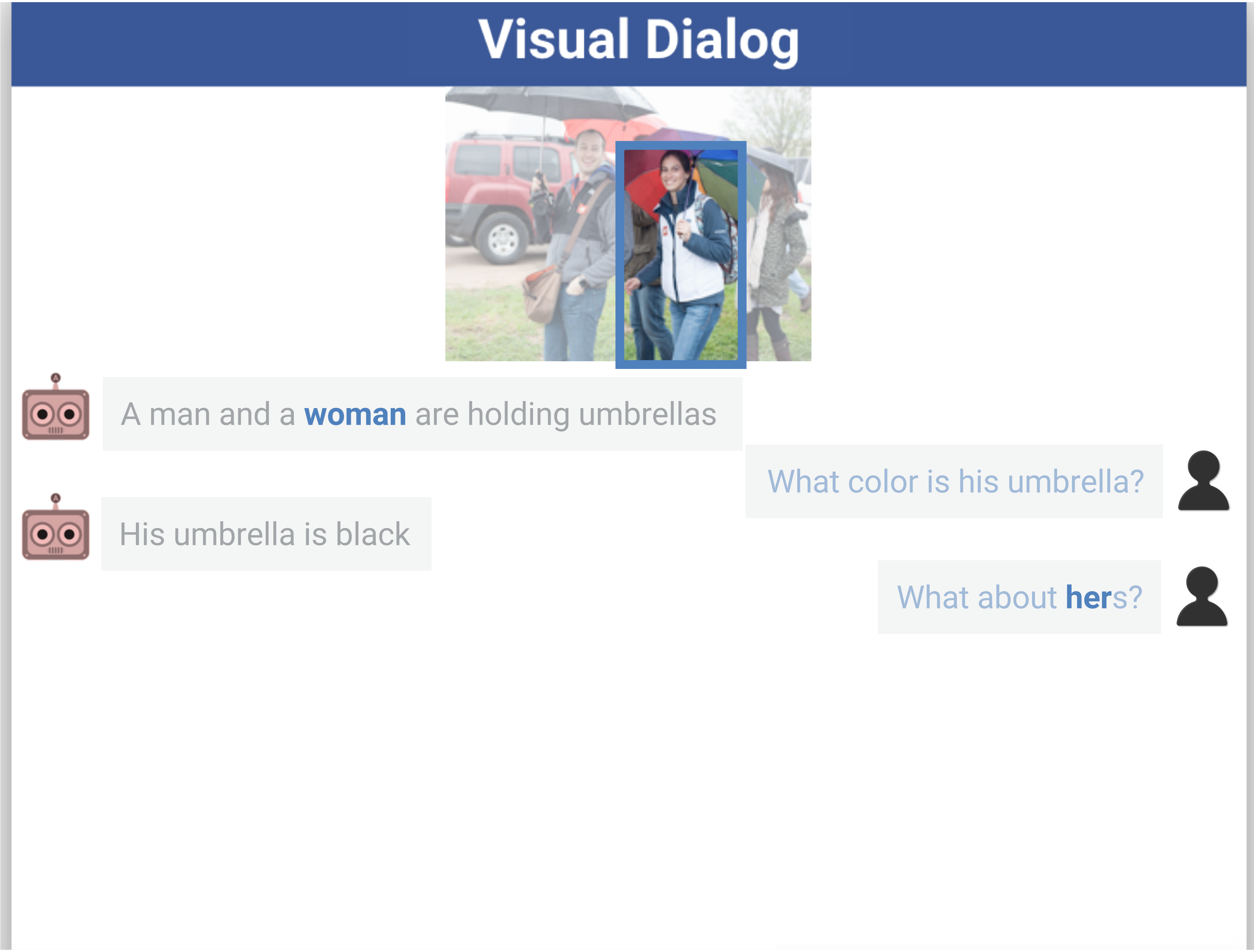

Visual Dialog
A man and a woman are holding umbrellas
What color is his umbrella?

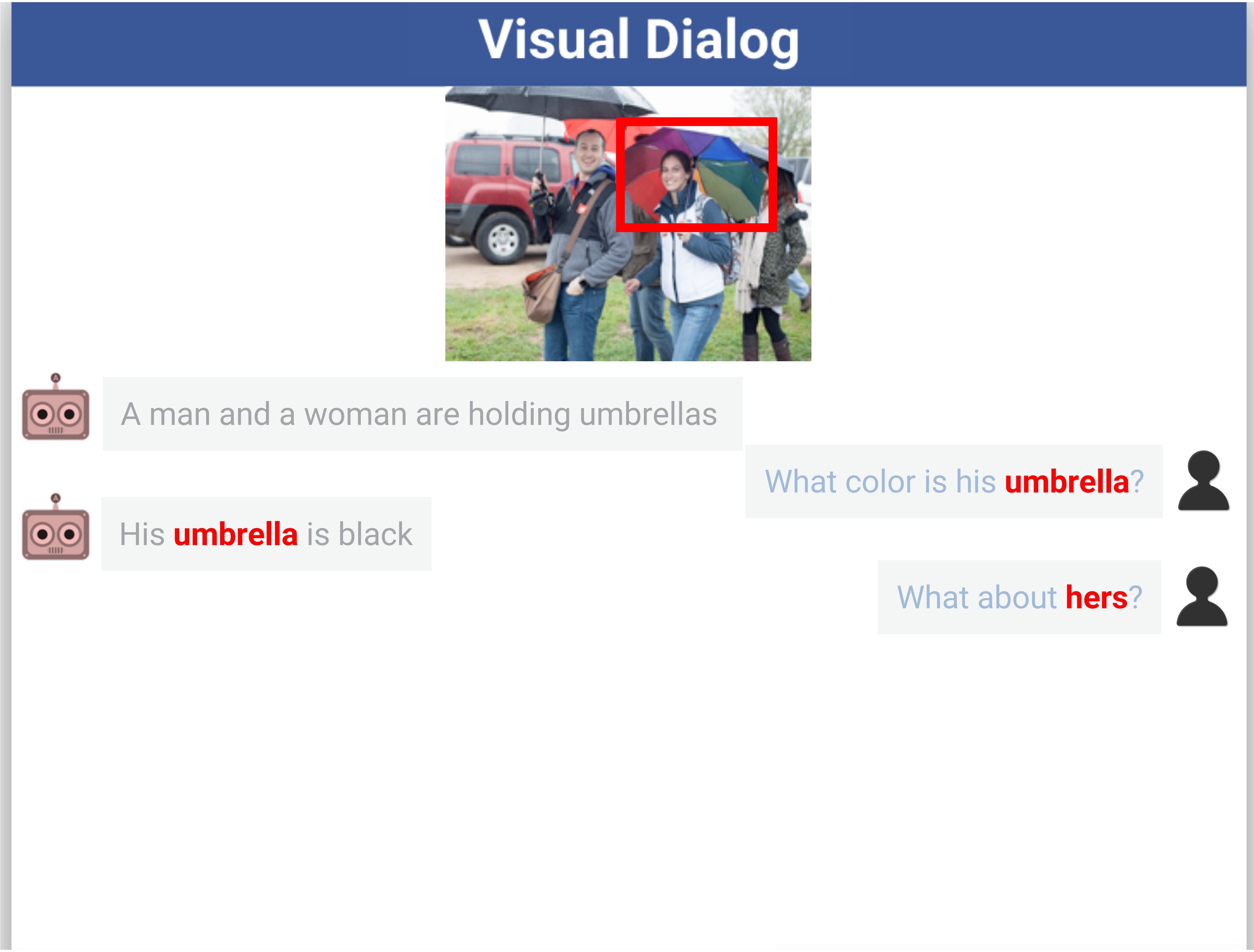

Visual Dialog
A man and a woman are holding umbrellas
What color is his umbrella?

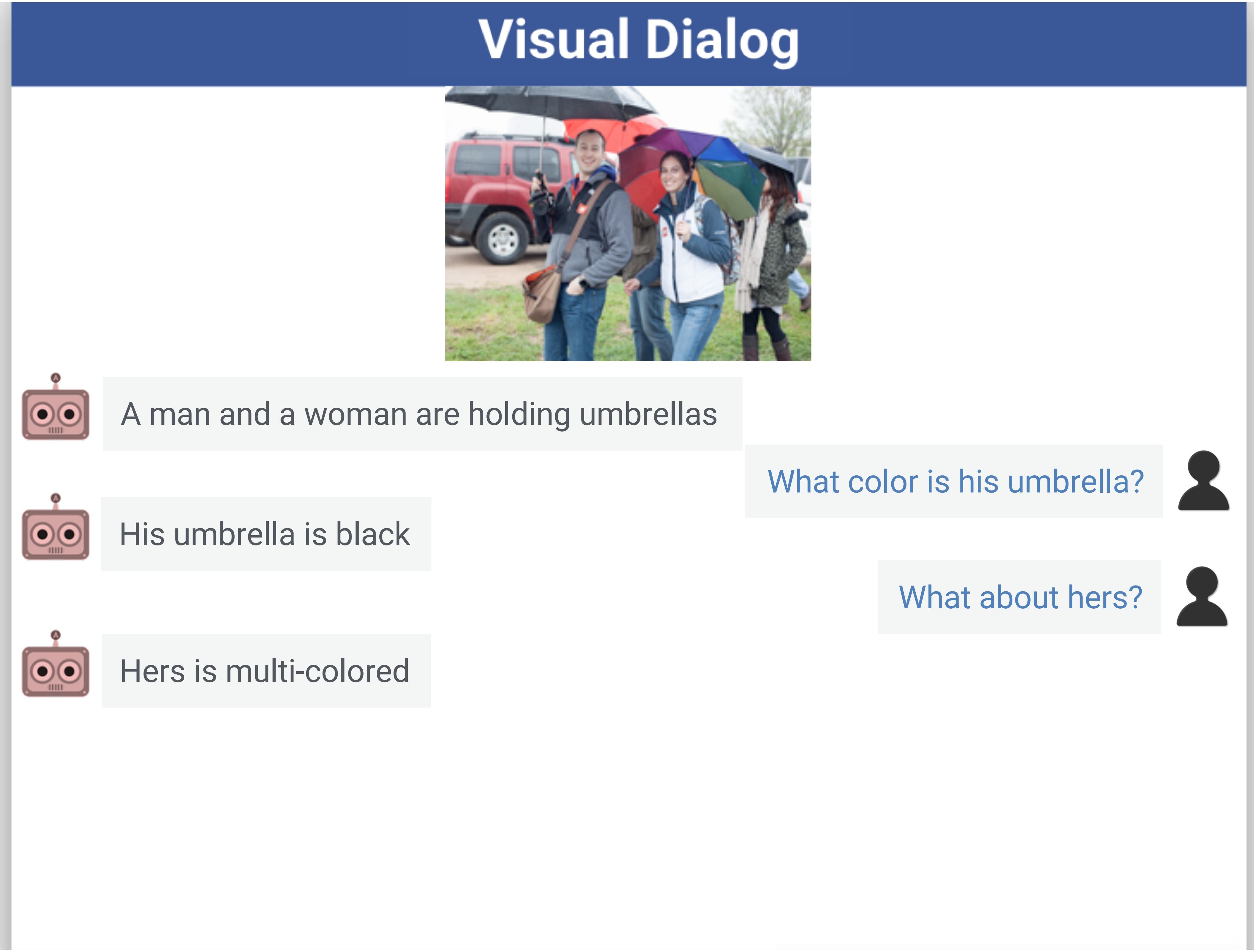

Visual Dialog
A man and a woman are holding umbrellas
What color is his **umbrella**?

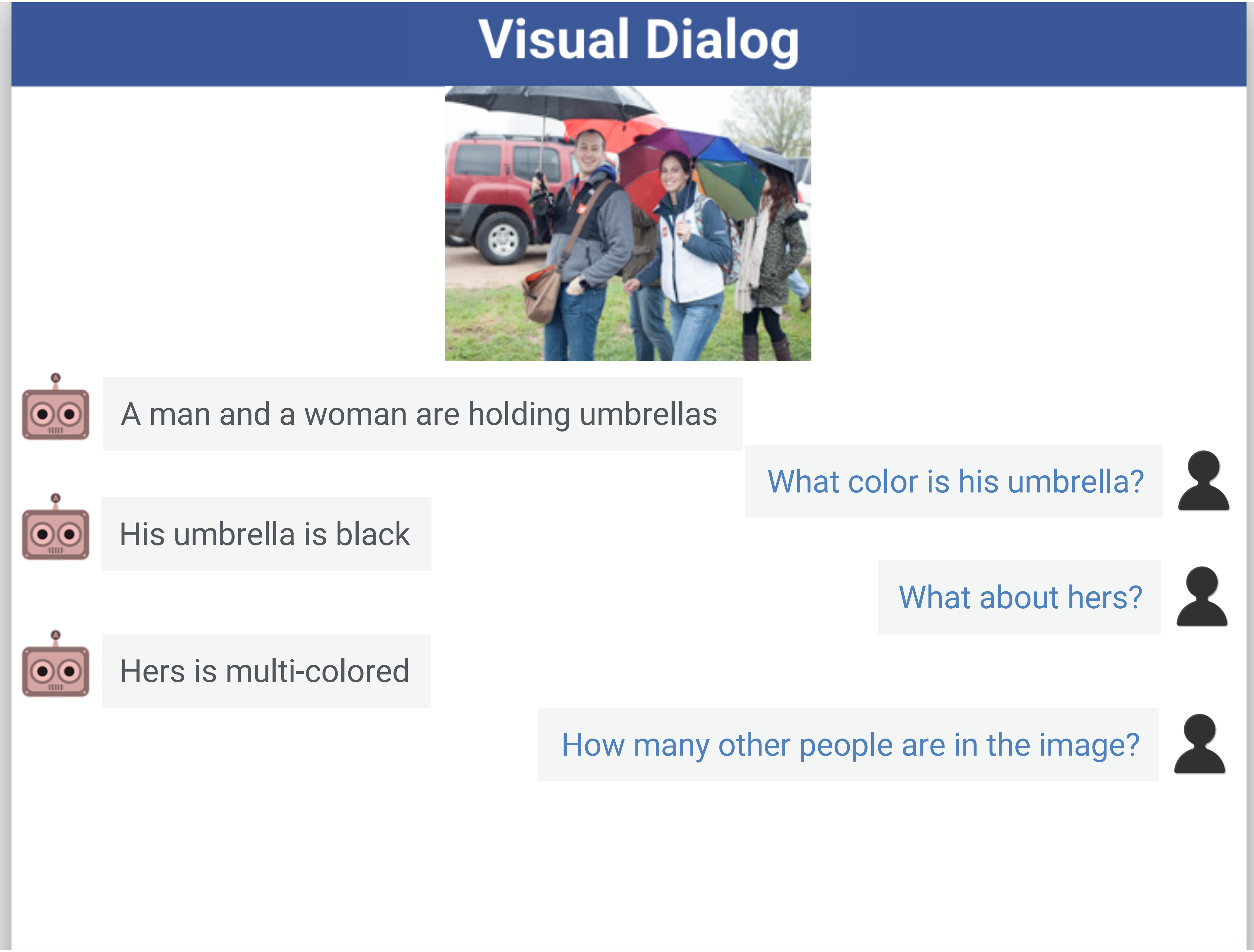

Visual Dialog
A man and a woman are holding umbrellas
What color is his umbrella?
His umbrella is black

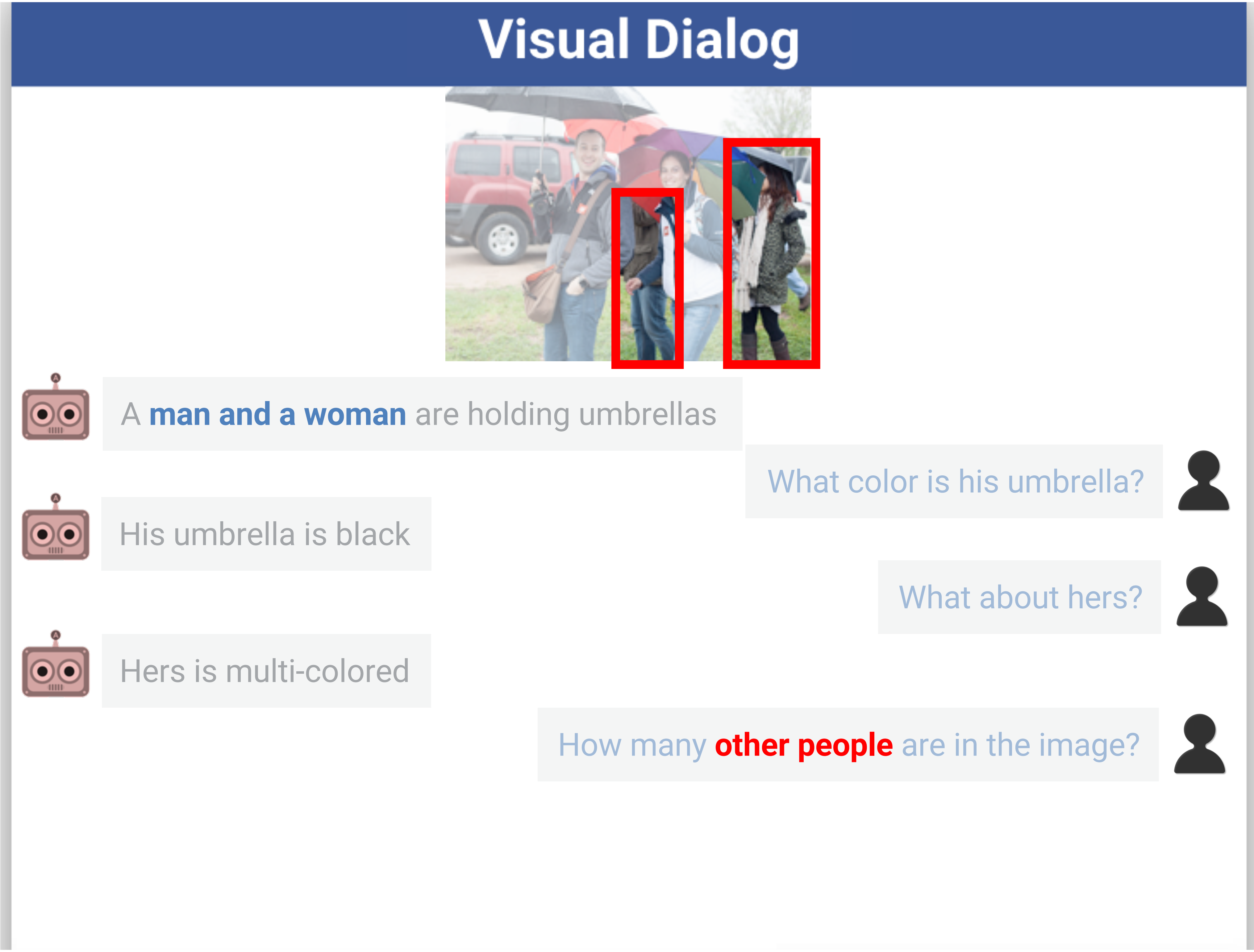

Visual Dialog
A man and a woman are holding umbrellas
What color is his umbrella?
His umbrella is black
What about hers?

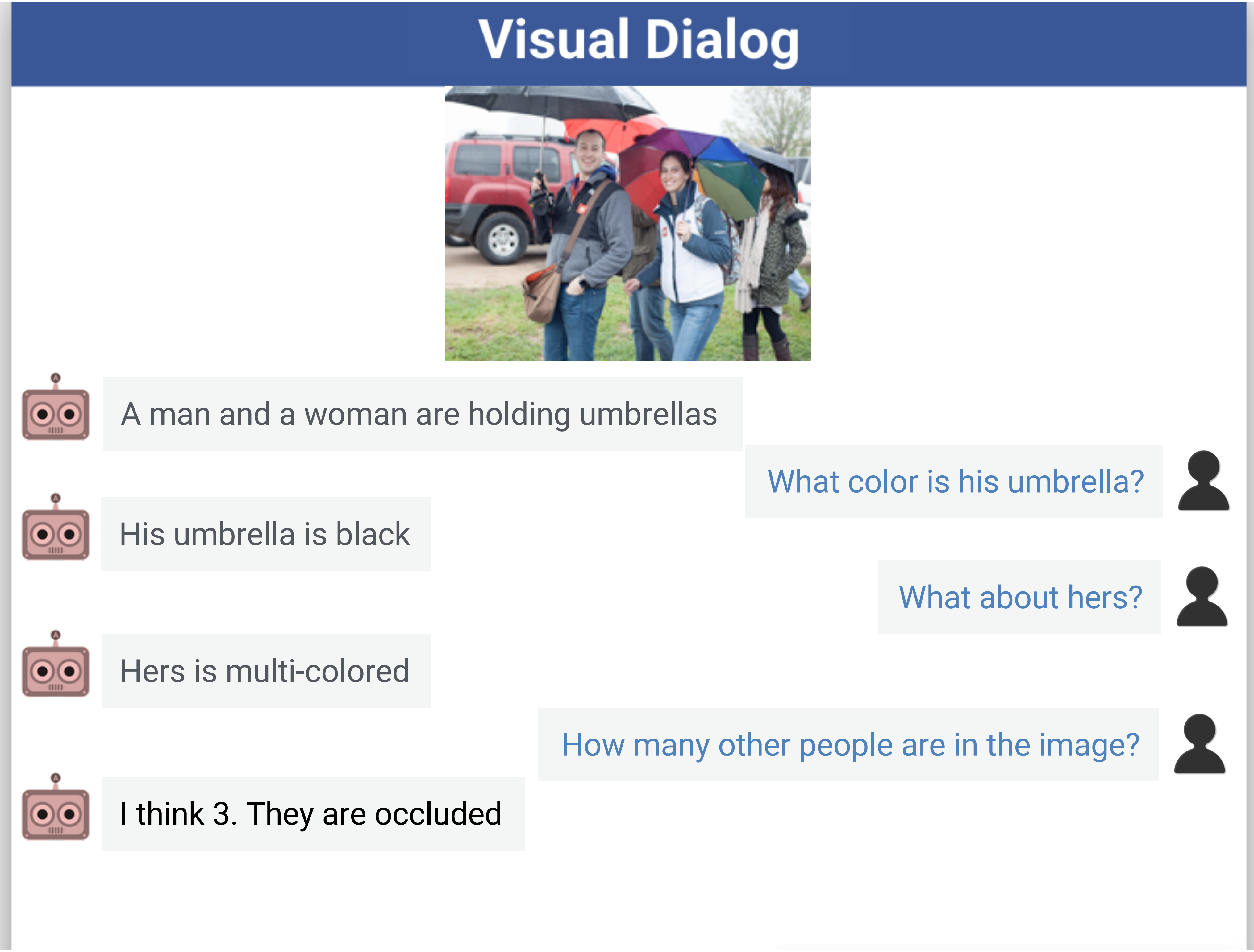

Visual Dialog
A man and a woman are holding umbrellas
What color is his umbrella?
His umbrella is black
What about hers?

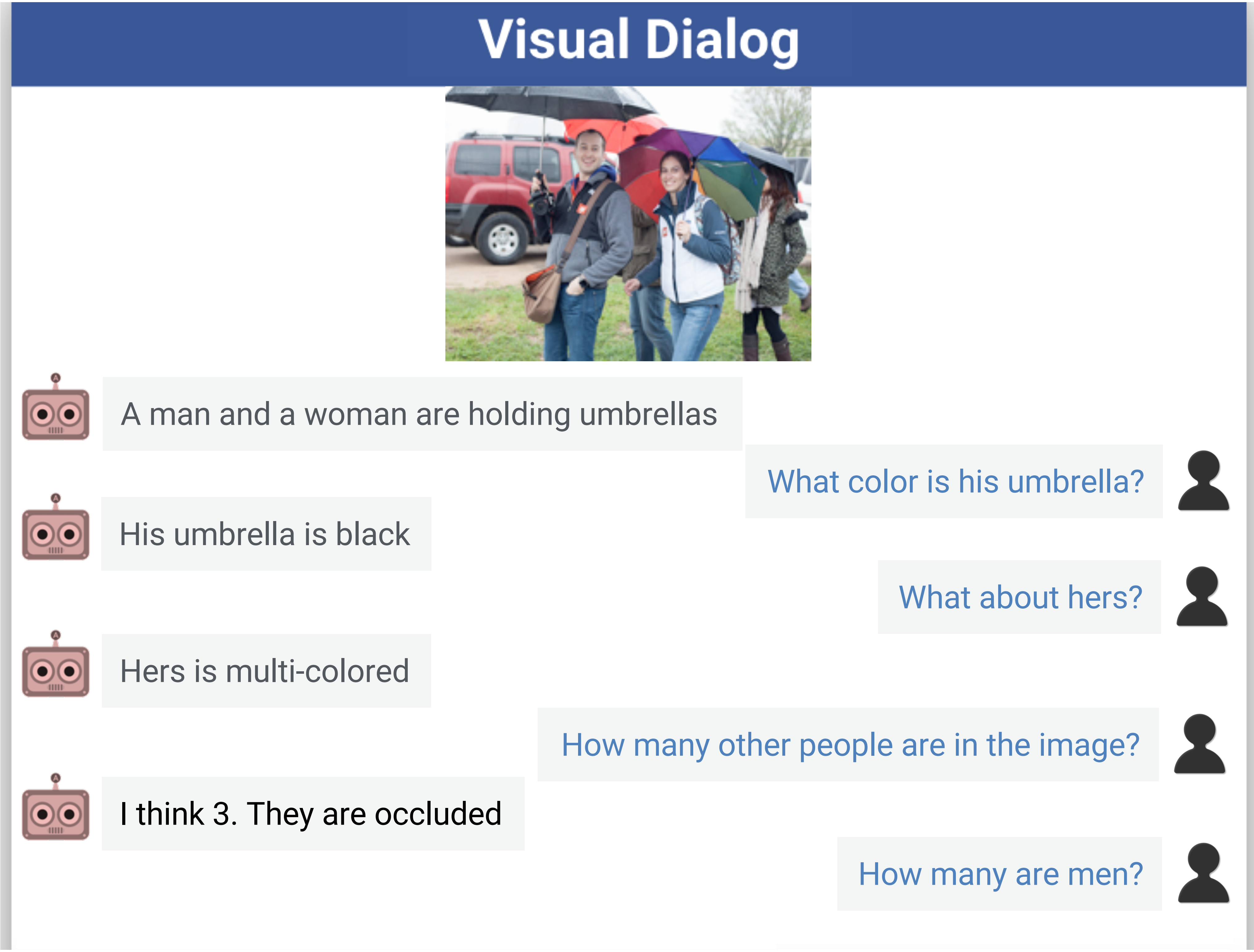

Visual Dialog
A man and a woman are holding umbrellas
What color is his umbrella?
His umbrella is black
What about hers?


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?


Visual Dialog
A cat drinking water out of a coffee mug.
What color is the mug?
White and red
Are there any pictures on it?
No, something is there can't tell what it is
Is the mug and cat on a table?
Yes, they are
Are there other items on the table?
Yes, magazines, books, toaster and basket, and a plate
Start typing question here ...


12:29
Visual Chatbot
Hi, I am a Visual Chatbot, capable of answering a sequence of questions about images. Please upload an image and fire away!
Start typing question here ...

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

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

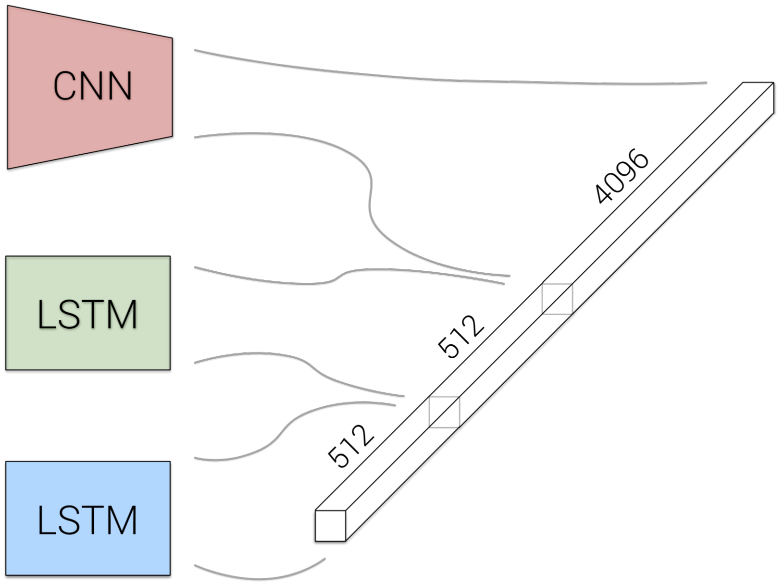
Image I

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.

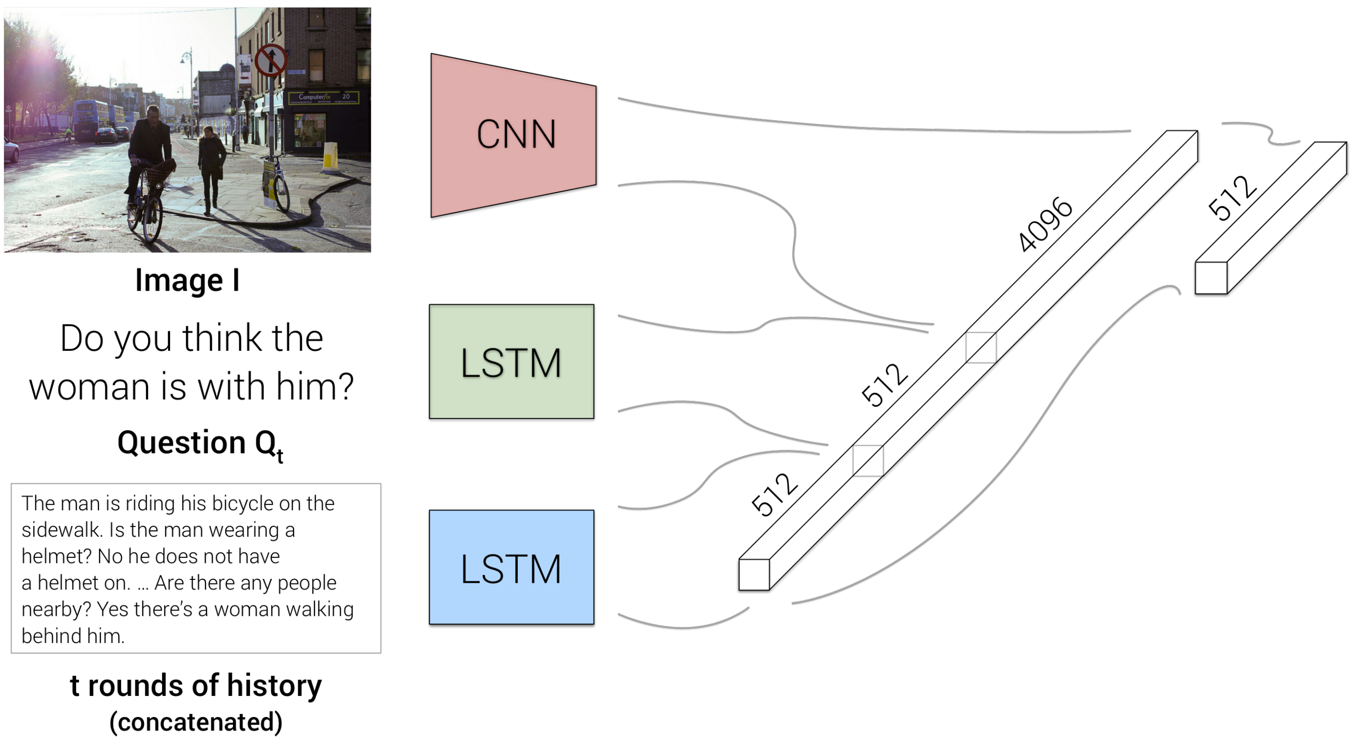
t rounds of history
(concatenated)



Late Fusion Encoder

Slide Credit: Abhishek Das

Visual Dialog Model #1



Late Fusion Encoder

Slide Credit: Abhishek Das

Visual Dialog Model #1

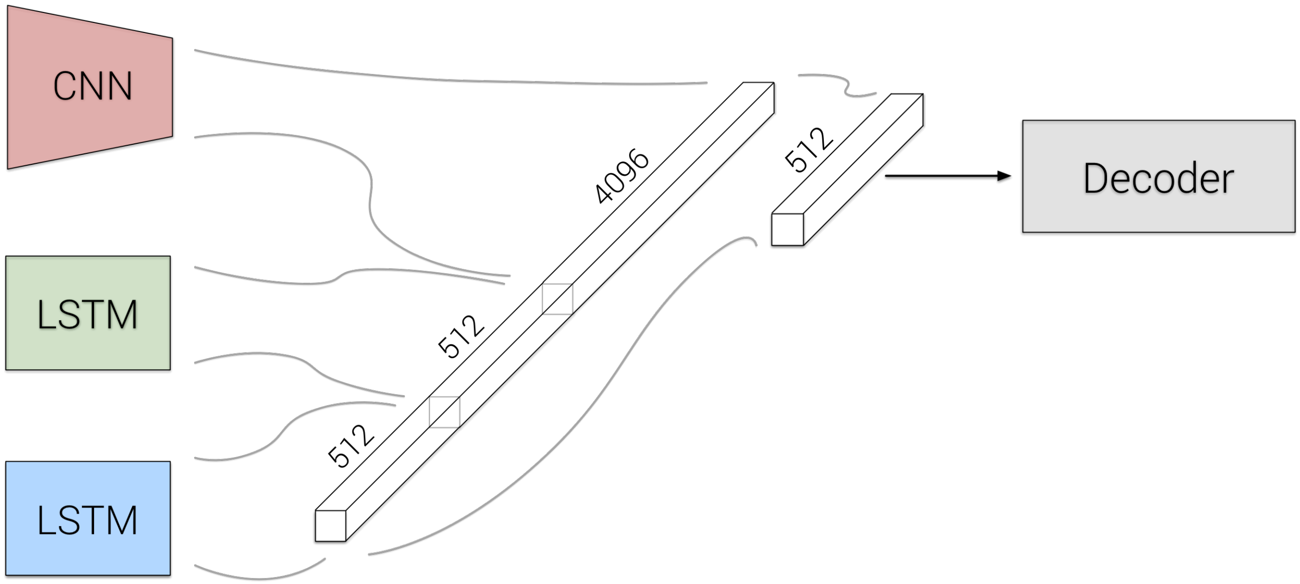
Image I

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.

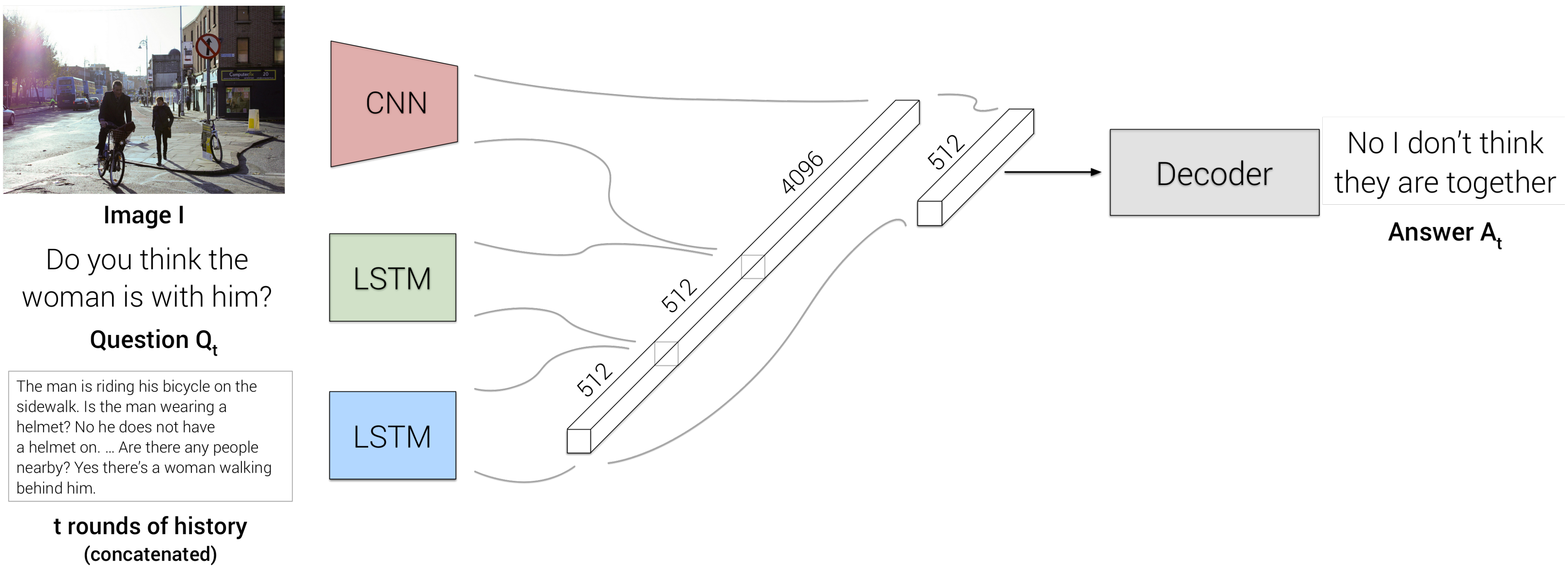
t rounds of history
(concatenated)



Late Fusion Encoder

Slide Credit: Abhishek Das

Visual Dialog Model #1



Late Fusion Encoder

Learning Cooperative Visual Dialog Agents with Deep Reinforcement Learning

[ICCV ‘17]


Abhishek Das*
(Georgia Tech)

Satwik Kottur*
(CMU)

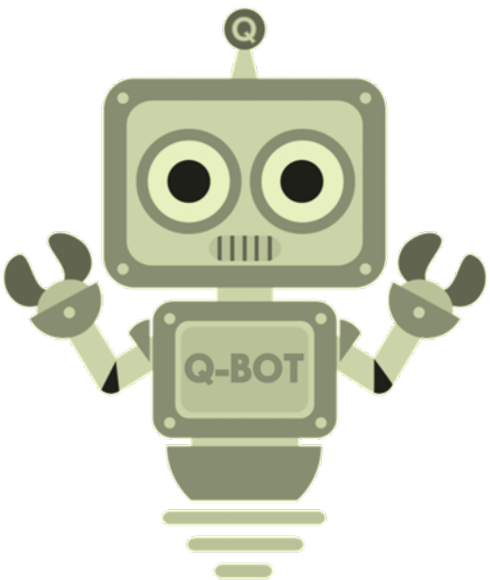


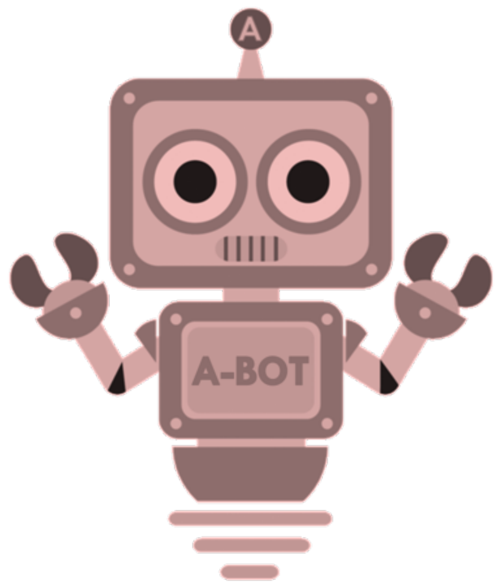
José Moura
(CMU)

Stefan Lee
(Virginia Tech)

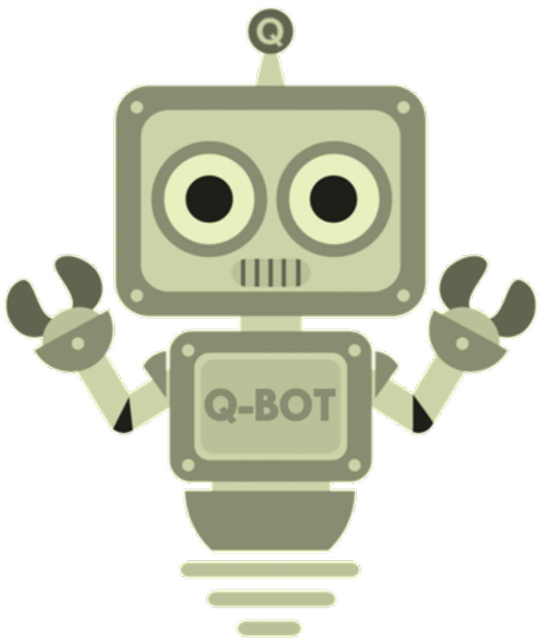
Dhruv Batra
(Georgia Tech / FAIR)

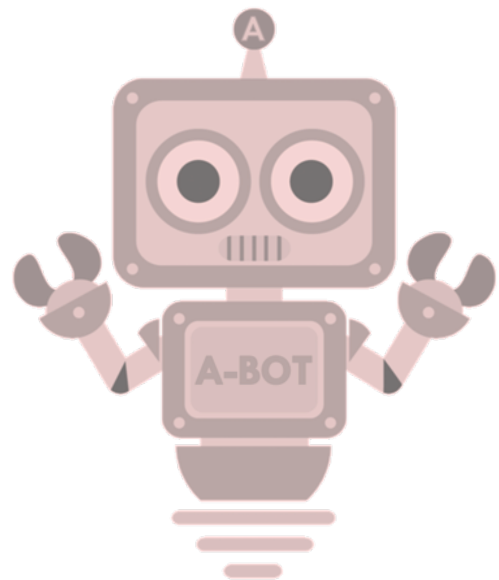
GuessWhich: Image Guessing Game





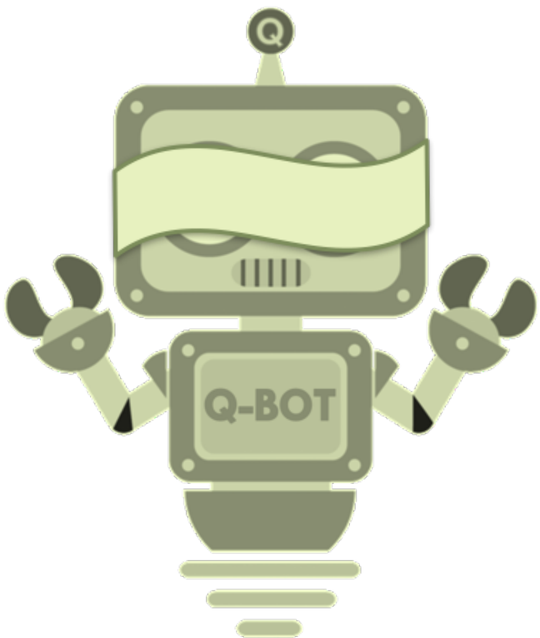
GuessWhich: Image Guessing Game

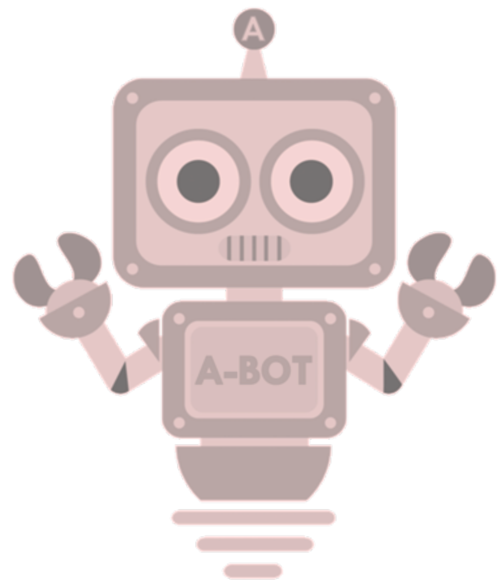




Q-Bot asks questions

GuessWhich: Image Guessing Game

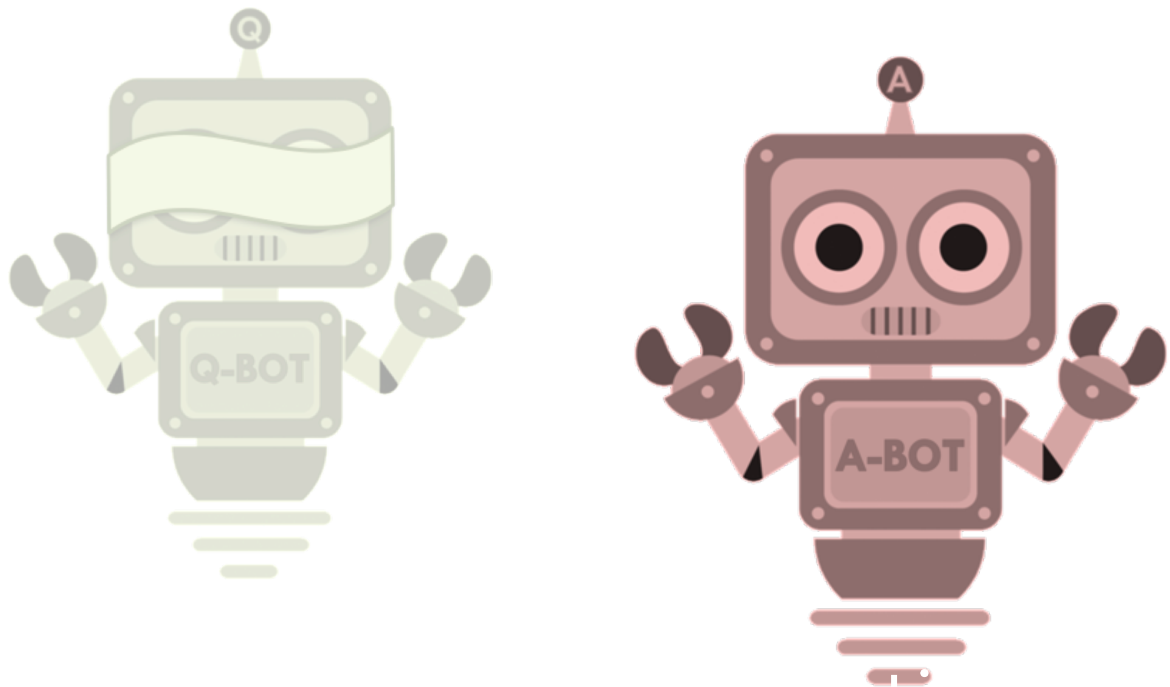




Q-Bot is blindfolded

Slide Credit: Abhishek Das

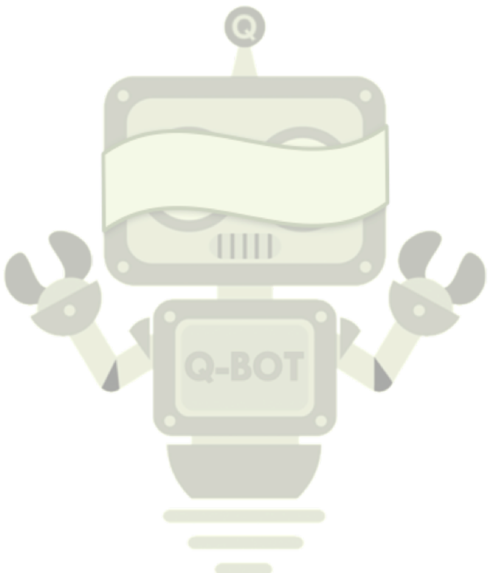
GuessWhich: Image Guessing Game

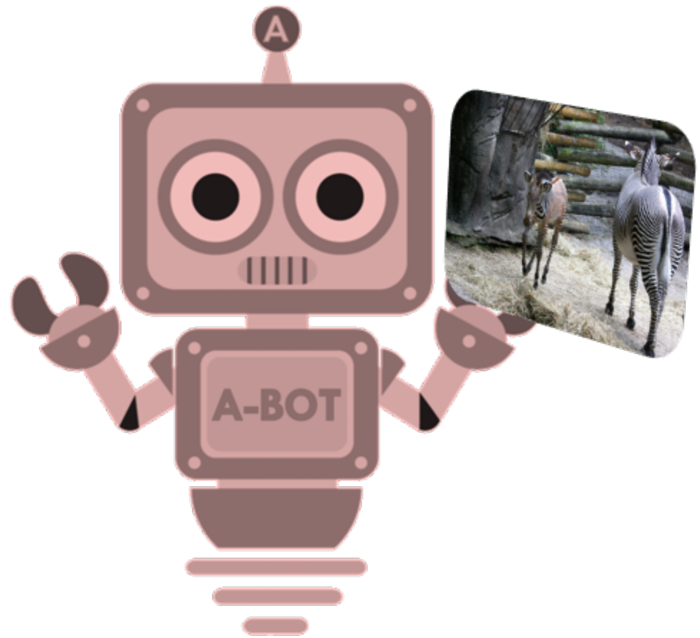


A-Bot answers questions

Slide Credit: Abhishek Das

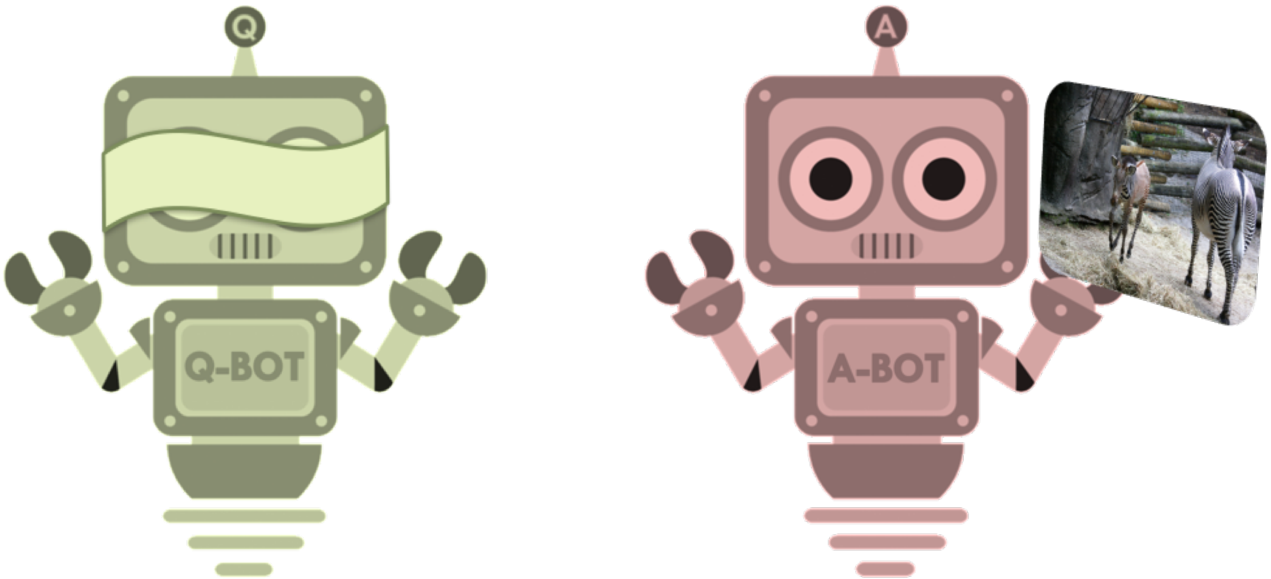
GuessWhich: Image Guessing Game





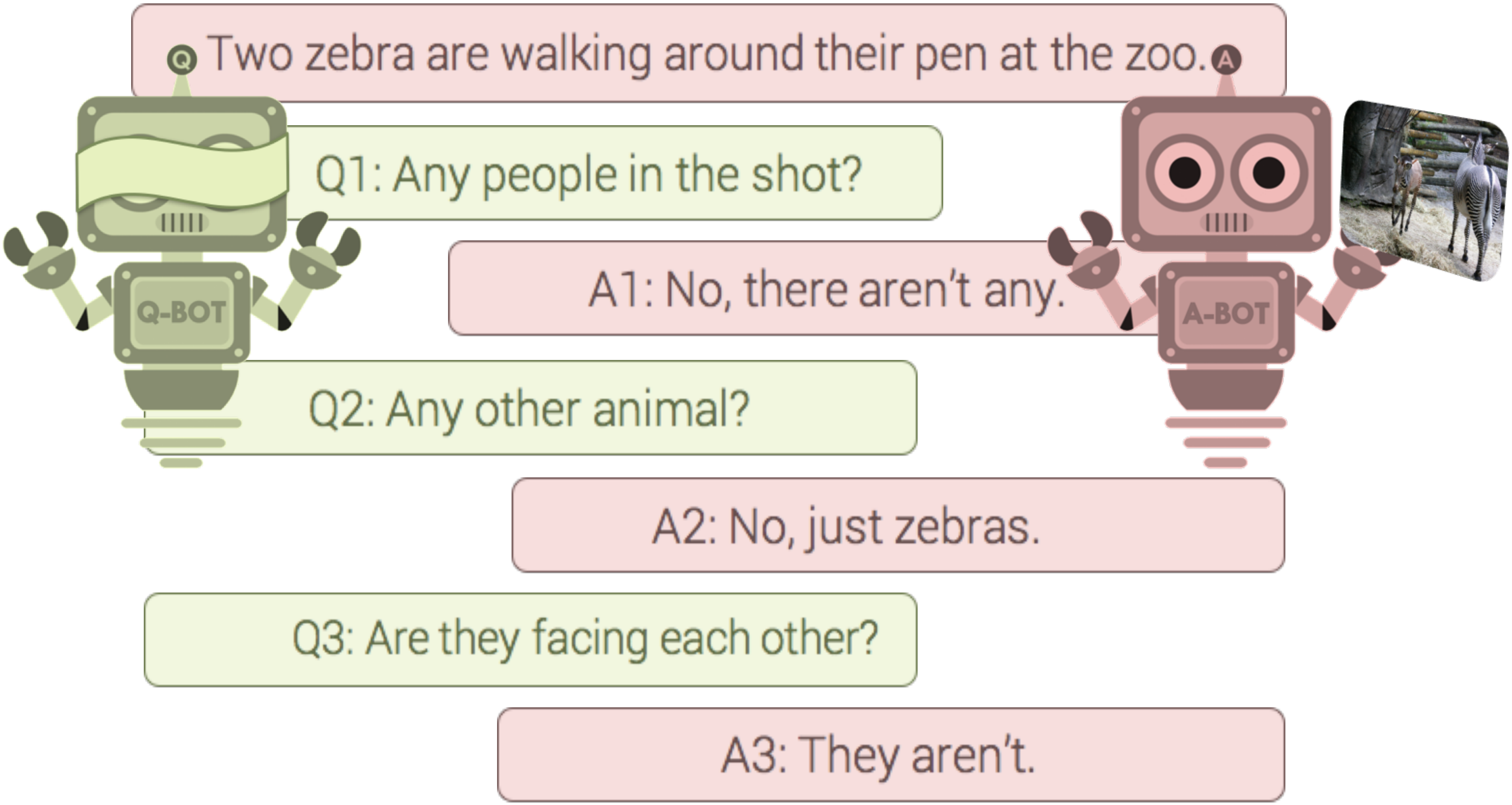
A-Bot sees an image

GuessWhich: Image Guessing Game



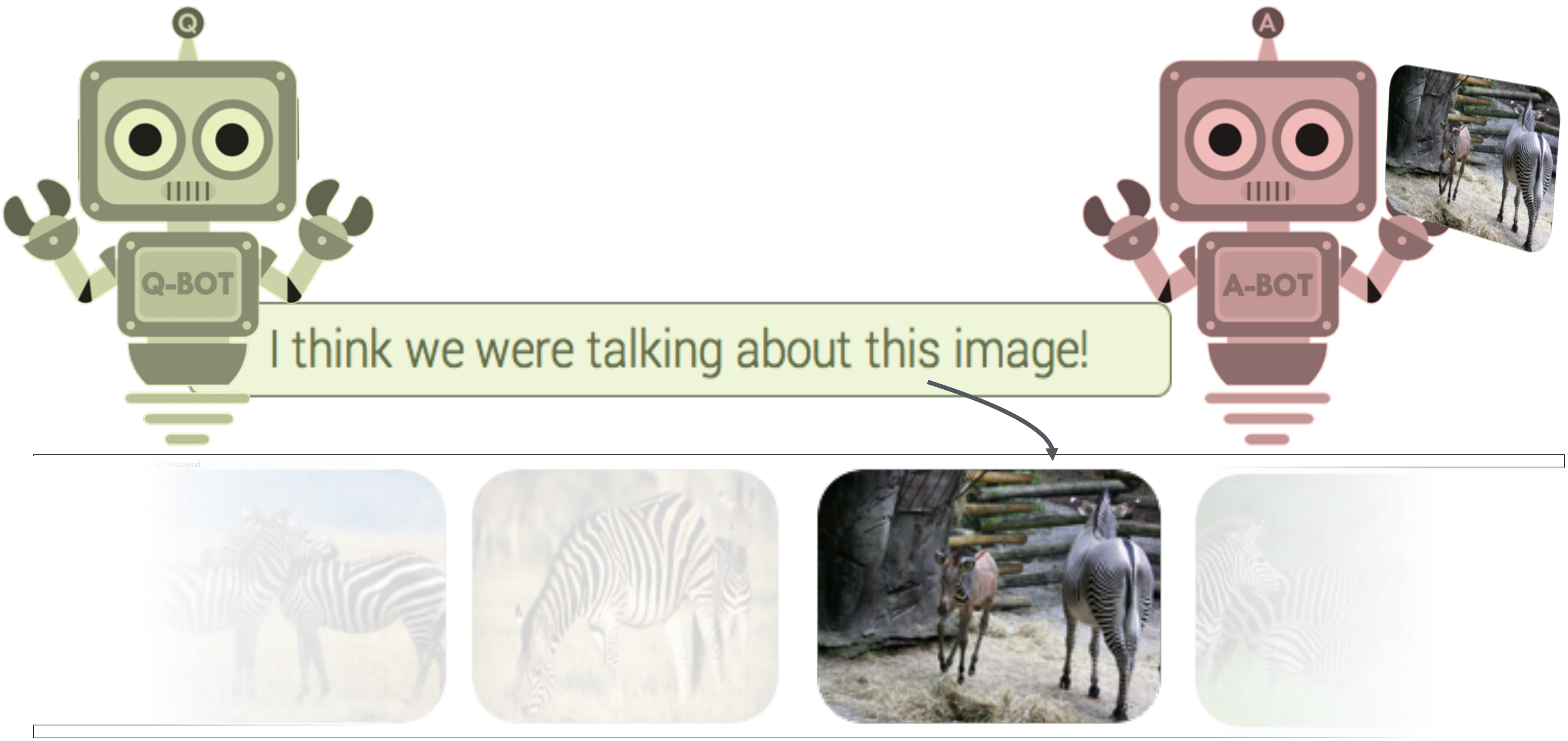
Slide Credit: Abhishek Das

GuessWhich: Image Guessing Game



GuessWhich: Image Guessing Game

A3: They aren't.



Turing Test

How old is this man?	He looks like late teens
Does he have short hair or long hair?	Medium
What color is his hair?	Brown
Is he wearing glasses?	No
Does he have facial hair?	No
How is he dressed?	He's white dress shirt with red tie ... school uniform
Is he larger man or is he skinny?	He's skinny
Is this picture taken outside?	Inside
Can you tell if it's daytime?	I'm not sure
Can you tell what room he is standing in?	He's in hallway

A man making the live long and prosper sign from star trek.

Where is man located?	Looks like classroom of some sort
What is man wearing?	Black t shirt and jeans
Are there any other people in picture?	Yes, quite few people in background
What color is table?	Picture is black and white, but it's wood table
Is there anything else on table?	Not that I can see
What are other people doing?	They are standing in front of him, but I don't ...
Are there any windows?	Not that I can see
What color are walls?	I can't see walls
What time of day do you think it is?	I can't tell
What is man wearing?	Black t shirt and jeans

Natural Language Does Not Emerge 'Naturally' in Multi-Agent Dialog

[EMNLP '17]

Best Paper Award

Satwik Kottur*
(CMU)

José Moura
(CMU)

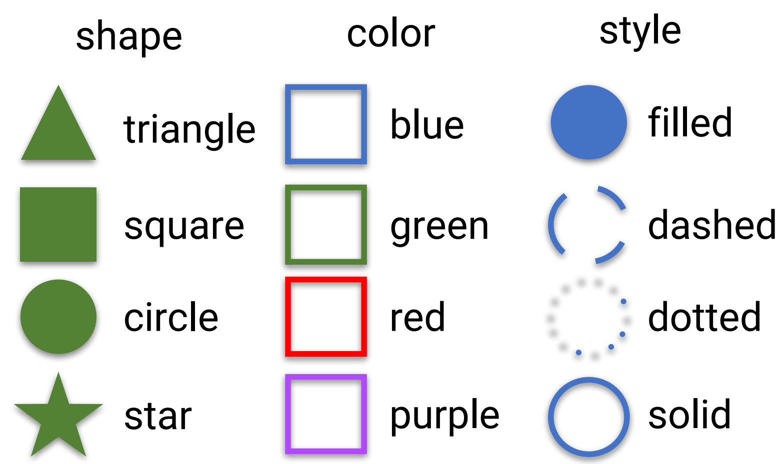
Stefan Lee
(Virginia Tech)

Dhruv Batra
(Georgia Tech / FAIR)

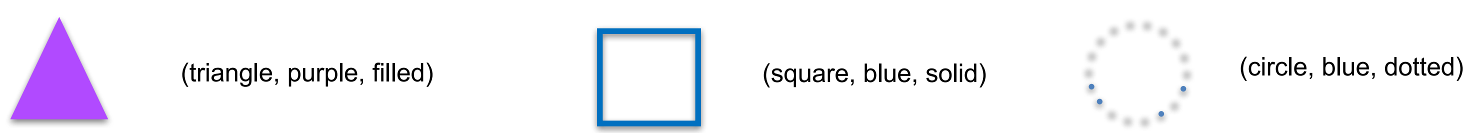
Toy World

- Sanity check

- Simple, synthetic world
  - Instances - (shape, color, style)
  - Total of $4^3 (64)$ instances

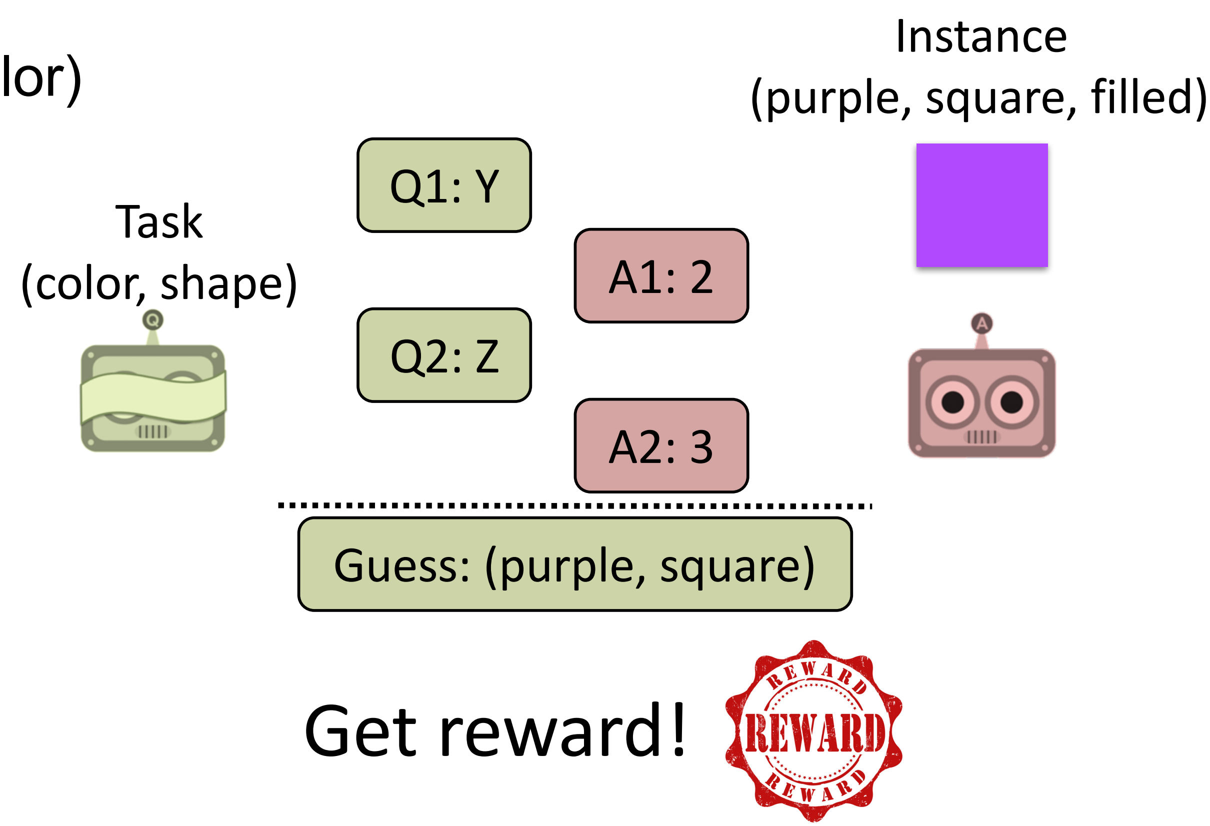


  - Example instances:

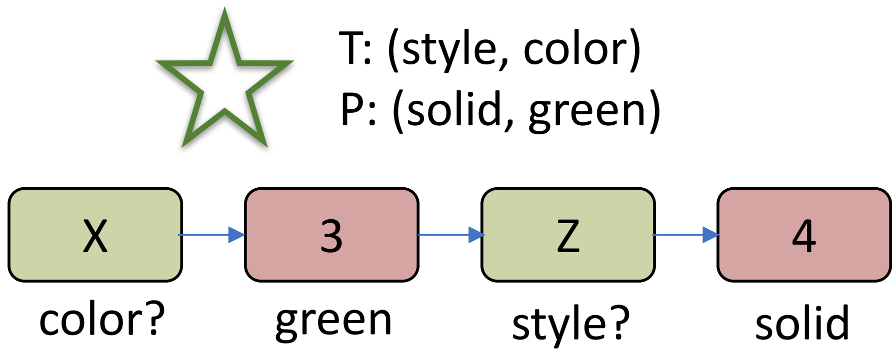


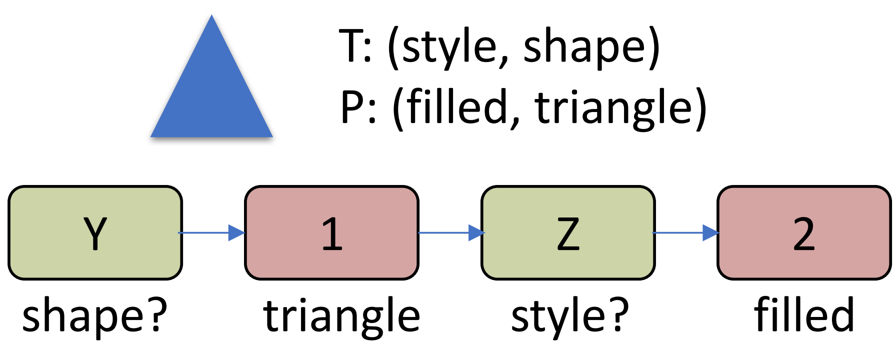
Task & Talk

- Task (G)
  - Inquire pair of attributes
  - (color, shape), (shape, color)
- Talk
  - Single token per round
  - Two rounds
- Q-bot guesses a pair
  - Reward : +1 / -1
  - Prediction order matters!



Emergence of Grounded Dialog





Emergence of Grounded Dialog

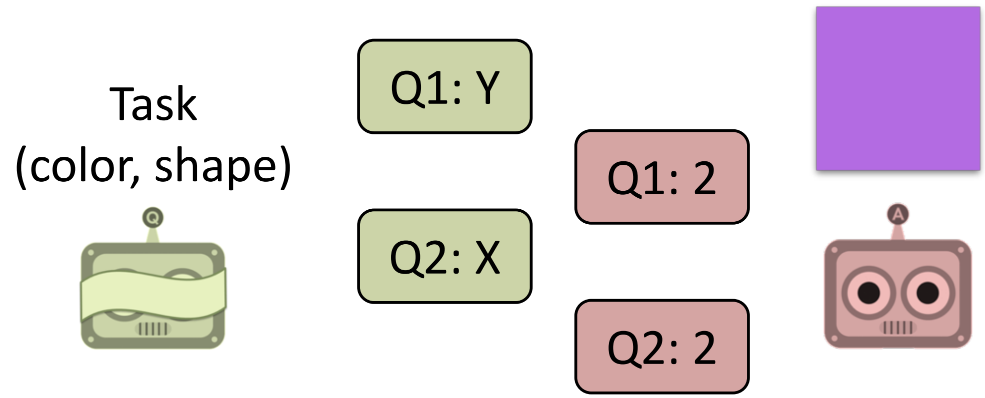
- Compositional grounding
- Predict dialog for unseen instances

	Attributes		
	color	*shape*	*style*
V_A	X	Y	Z
1	*blue*	*triangle*	*dotted*
2	*purple*	*square*	*filled*
3	*green*	*circle*	*dashed*
4	*red*	*start*	*solid*

(a) A-BOT

Task	q_1, q_2
(color, shape)	*Y, X*
(shape, color)	
(shape, style)	*Y, Z*
(style, shape)	
(color, style)	*Z, X*
(style, color)	*X, Z*

(b) Q-BOT



Summary of findings

Setting	Vocabulary		Memory		Generalization	Characteristics
	$\lvert V_Q \rvert$	$\lvert V_A \rvert$	Q-bot	A-bot		
A. Over-complete	64	64	Yes	Yes	25.6 %	• Non-compositional language • Q-bot insignificant • Inconsistent A-bot grounding • Poor generalization
B. Attribute	3	12	Yes	Yes	38.5 %	• Non-compositional language • Q-bot uses one round to convey task • Inconsistent A-bot grounding • Poor generalization
C. Minimal	3	4	Yes	No	74.4 %	• Compositional language • Q-bot uses both rounds • Consistent A-bot grounding • Good generalization

Deep Multi-Agent Communication

- NIPS '16
  - [DeepMind] Learning to Communicate with Deep Multi-Agent Reinforcement Learning. Jakob N. Foerster, Yannis M. Assael, Nando de Freitas, Shimon Whiteson. NIPS '16.
  - [NYU / FAIR] Learning Multiagent Communication with Backpropagation. Sainbayar Sukhbaatar, Arthur Szlam, Rob Fergus. NIPS '16.
- Arxiv '17
  - [OpenAI] Emergence of Grounded Compositional Language in Multi-Agent Populations. Igor Mordatch, Pieter Abbeel.
  - [FAIR] Multi-Agent Cooperation and the Emergence of (Natural) Language. Angeliki Lazaridou, Alexander Peysakhovich, Marco Baroni.
  - Learning to play guess who? and inventing a grounded language as a consequence. Emilio Jorge, Mikael Kågebäck, and Emil Gustavsson.
  - Emergence of language with multi-agent games: Learning to communicate with sequences of symbols. Serhii Havrylov and Ivan Titov.
  - [Berkeley] Translating neuralese. Jacob Andreas, Anca Dragan and Dan Klein. ACL 2017.

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

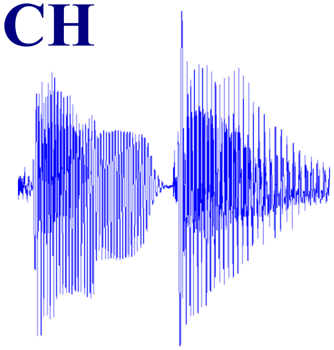
Traditional Machine Learning

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

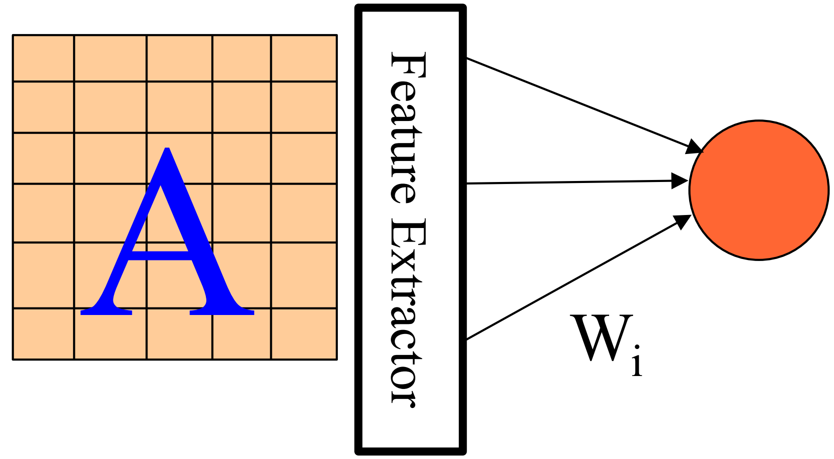
This burrito place is yummy and fun!

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

fixed | learned

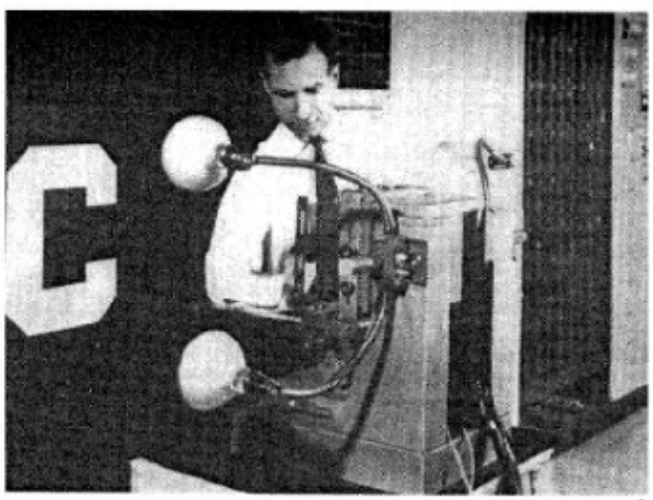
It’s an old paradigm

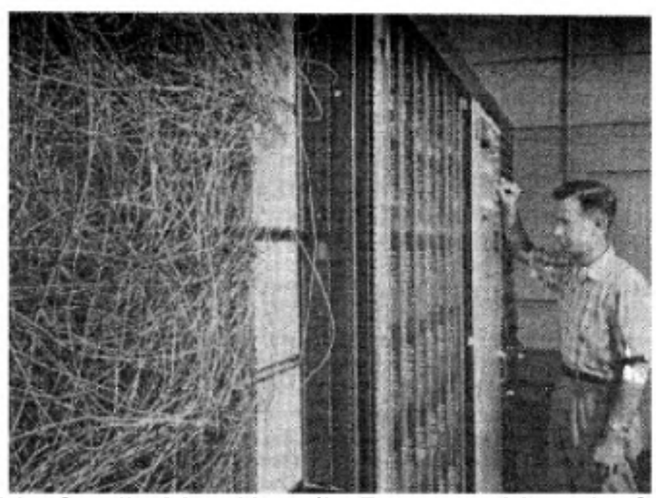
- The first learning machine: the Perceptron
  - Built at Cornell in 1960
- The Perceptron was a linear classifier on top of a simple feature extractor

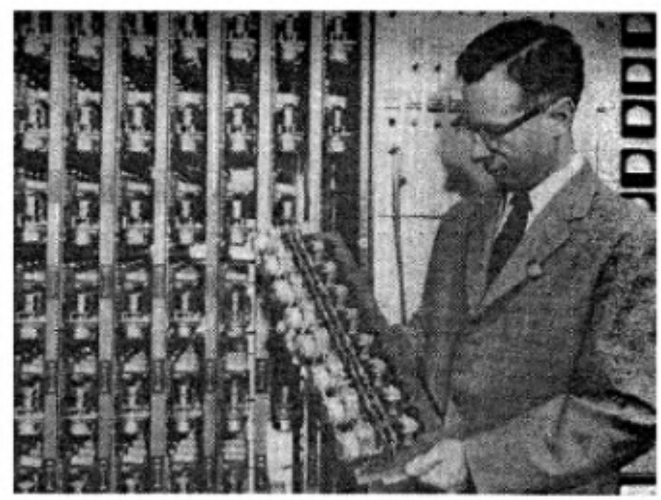


- The vast majority of practical applications of ML today use glorified linear classifiers or glorified template matching.
- Designing a feature extractor requires considerable efforts by experts.

$$y = sign\left(\sum_{i=1}^{N} W_i F_i(X) + b\right)$$







Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

Hierarchical Compositionality

VISION

pixels → edge → texton → motif → part → object

SPEECH

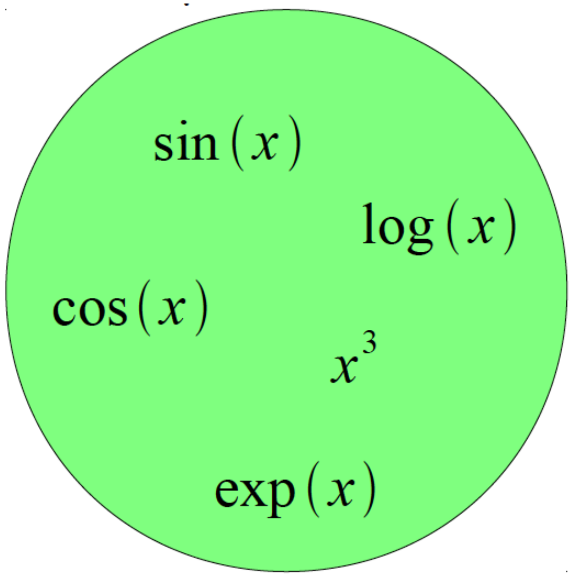
sample → spectral band → formant → motif → phone → word

NLP

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

Building A Complicated Function

Given a library of simple functions

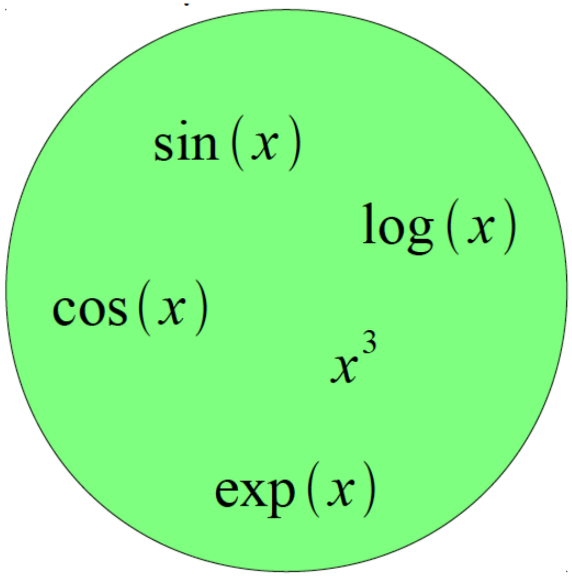


Compose into a

complicate function

Building A Complicated Function

Given a library of simple functions



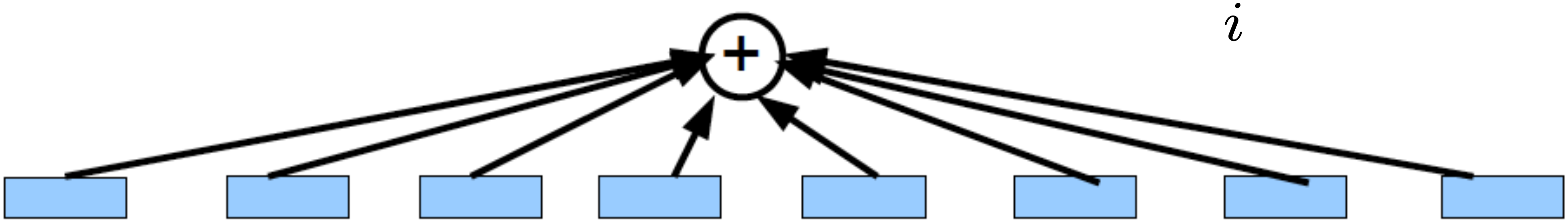
Compose into a

complicate function

Idea 1: Linear Combinations

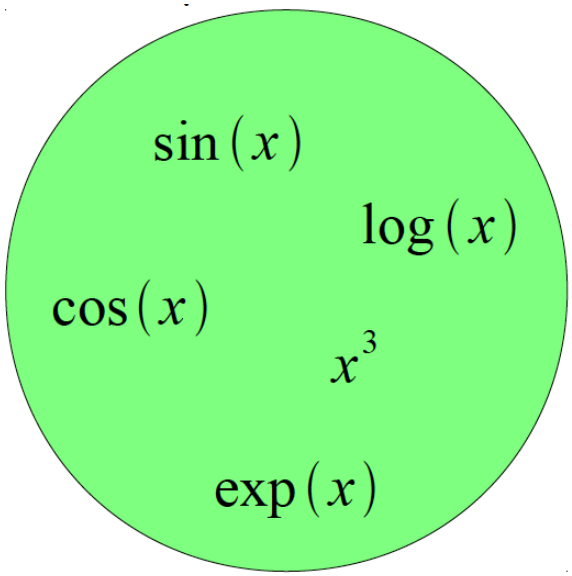
- Boosting
- Kernels
- …

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



Building A Complicated Function

Given a library of simple functions

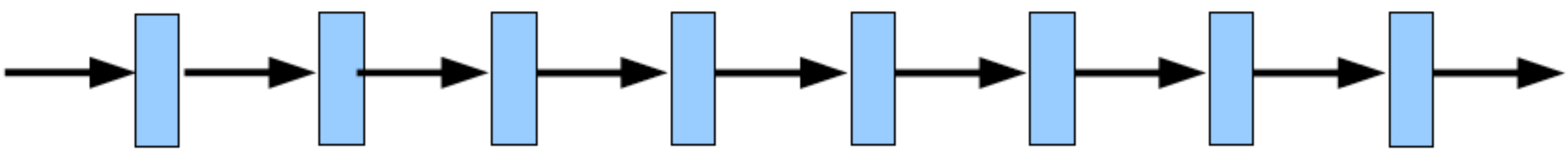


Compose into a complicate function

Idea 2: Compositions

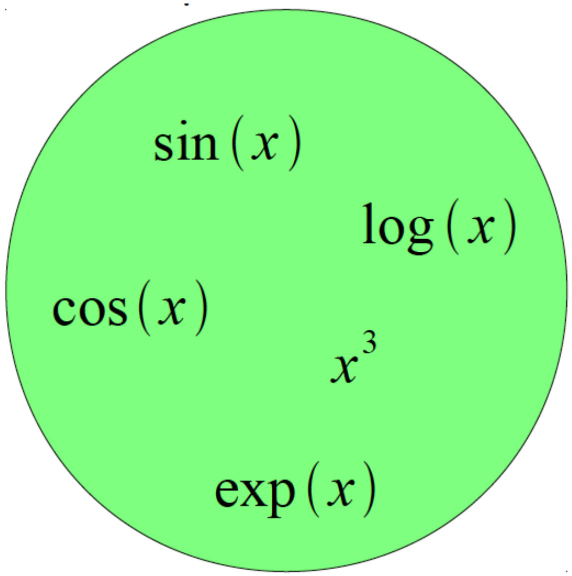
- Deep Learning
- Grammar models
- Scattering transforms…

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



Building A Complicated Function

Given a library of simple functions



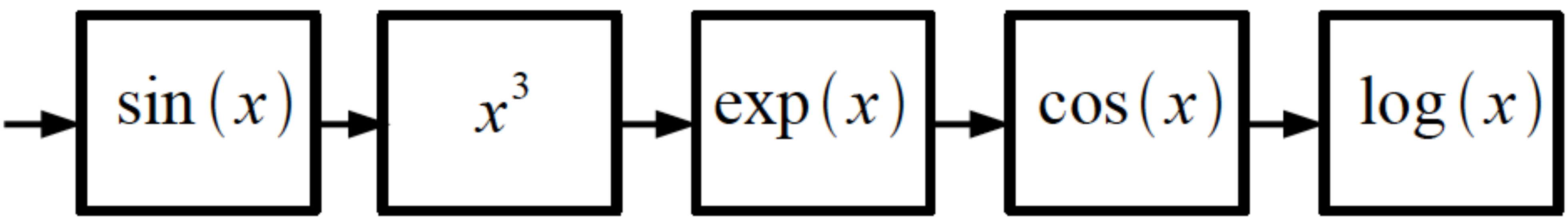
Compose into a

complicate function

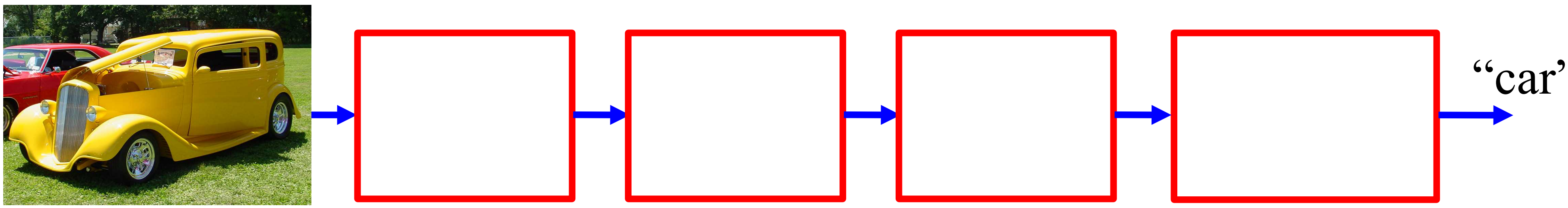
Idea 2: Compositions

- Deep Learning
- Grammar models
- Scattering transforms…

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

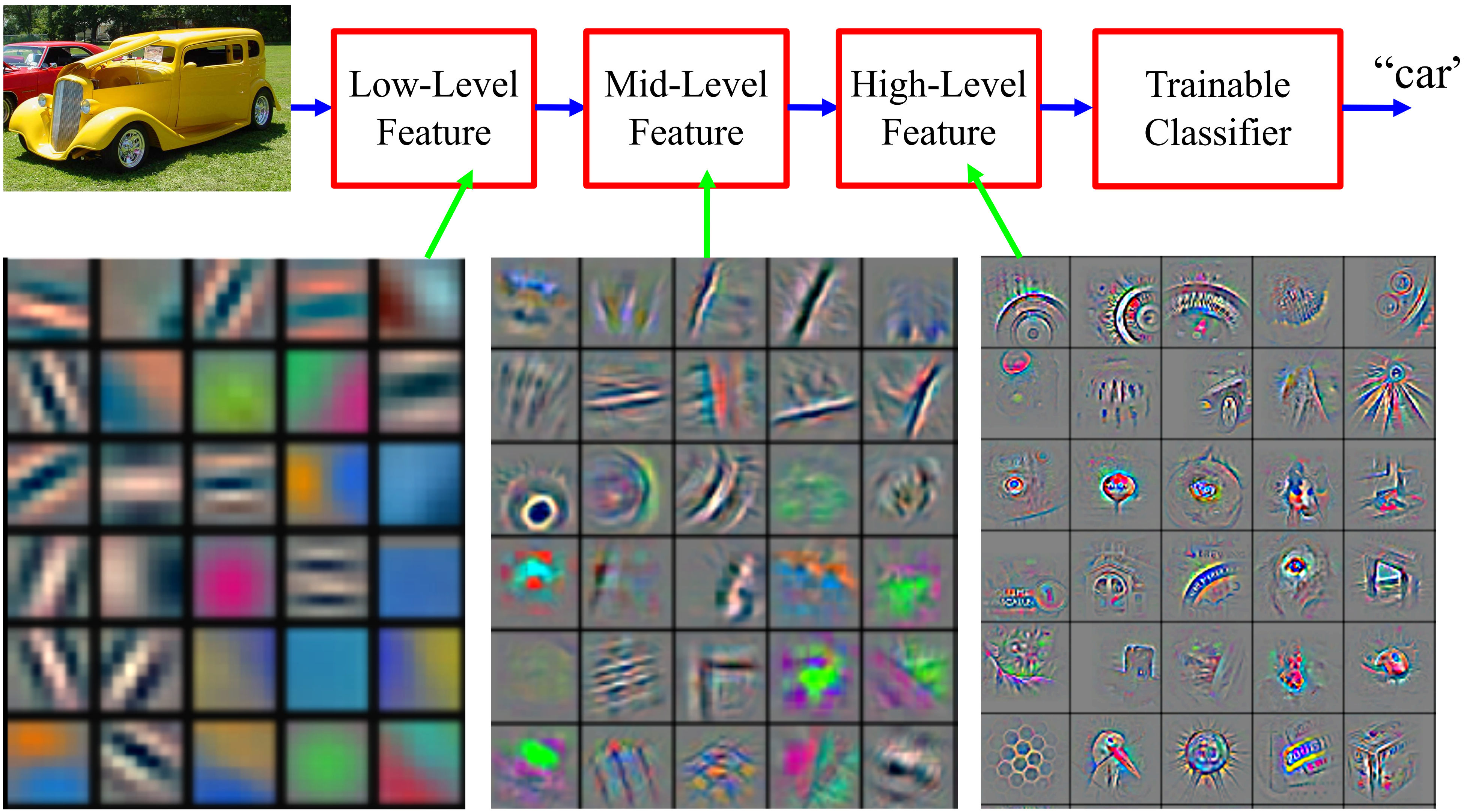


Deep Learning = Hierarchical Compositionality



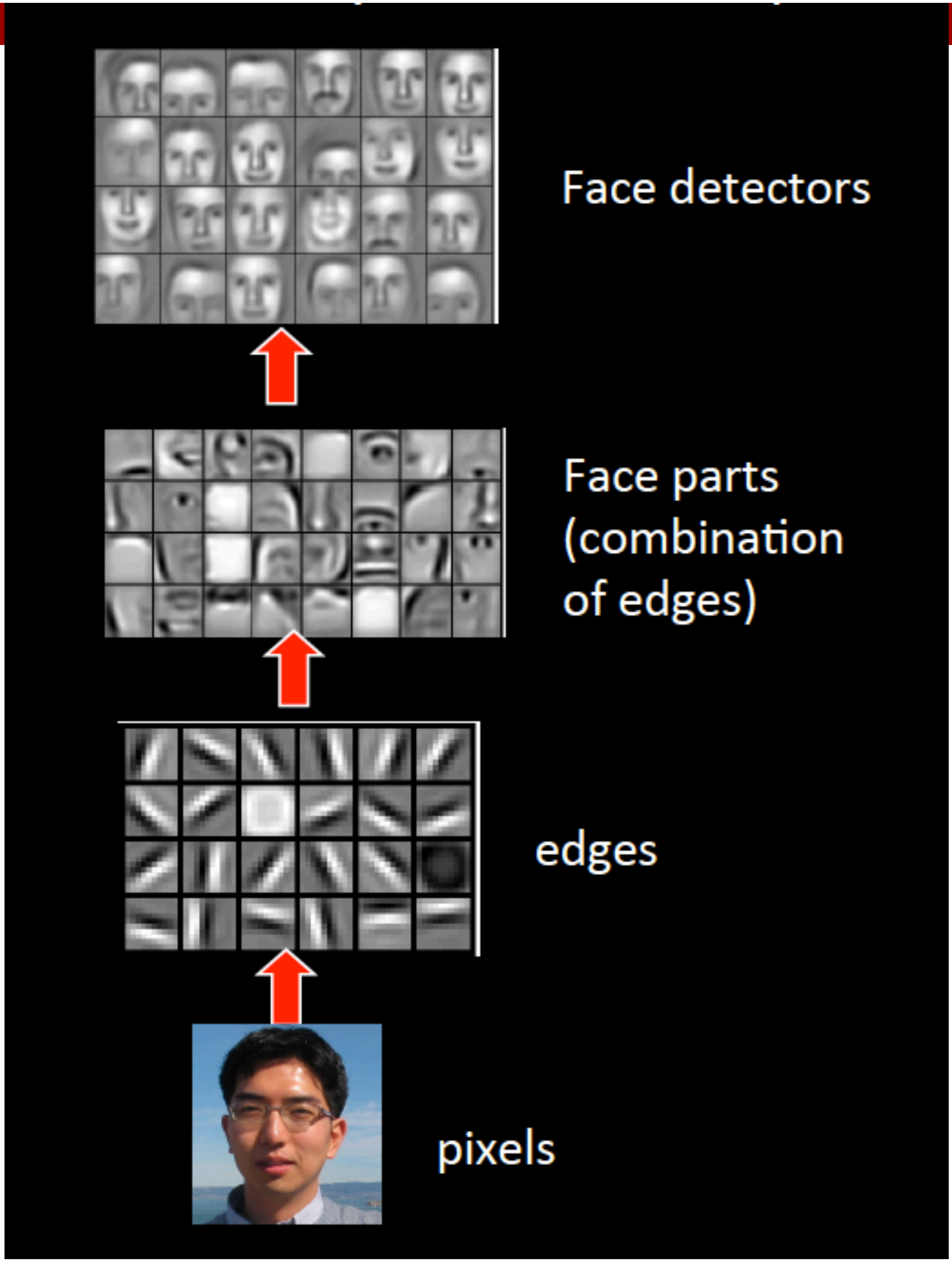
Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

Deep Learning = Hierarchical Compositionality



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

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun



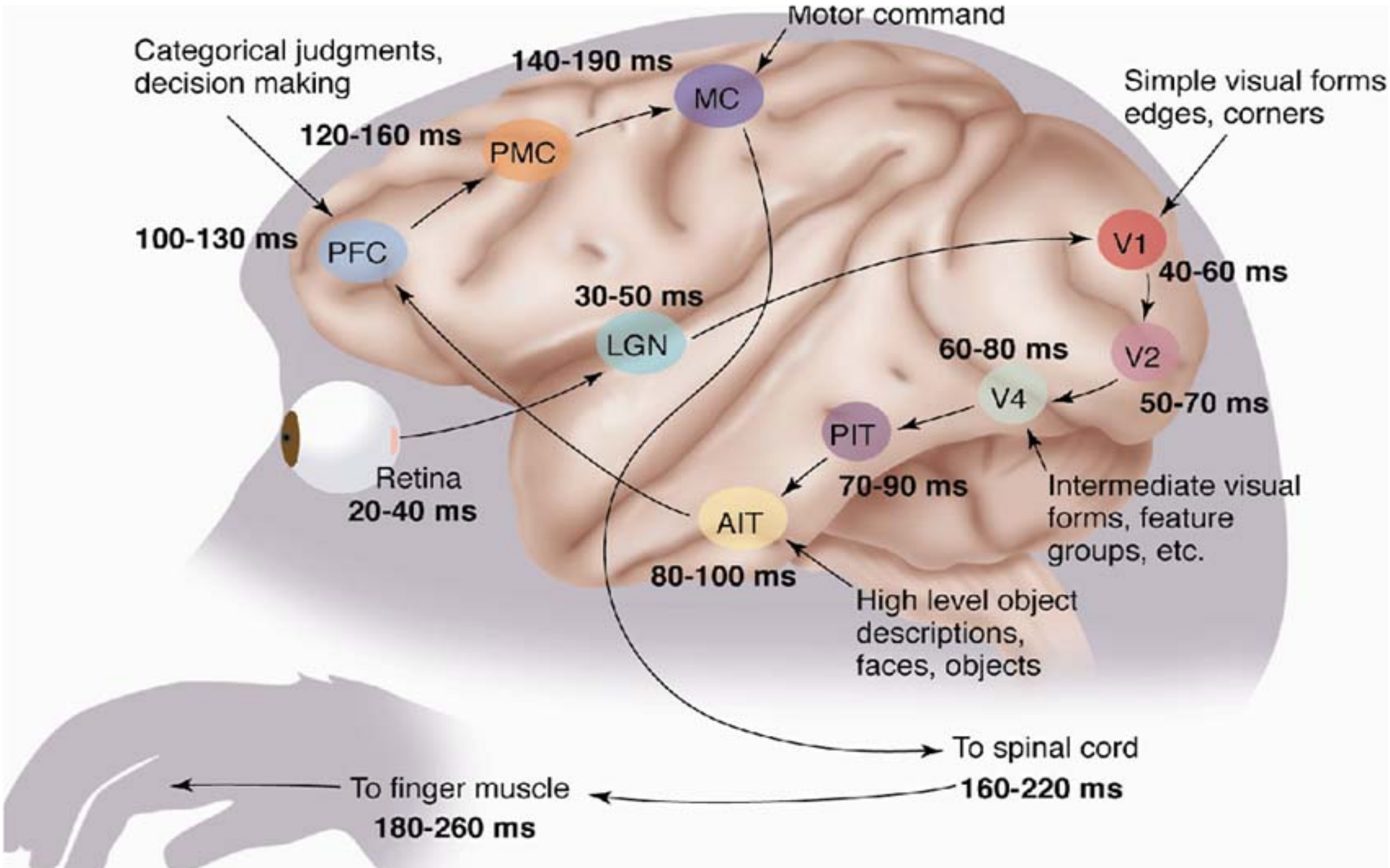
Sparse DBNs

[Lee et al. ICML '09]

Figure courtesy: Quoc Le

The Mammalian Visual Cortex is Hierarchical

- The ventral (recognition) pathway in the visual cortex



[picture from Simon Thorpe]

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

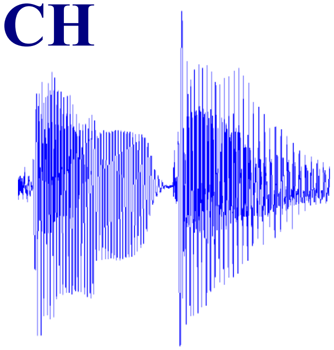
Traditional Machine Learning

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

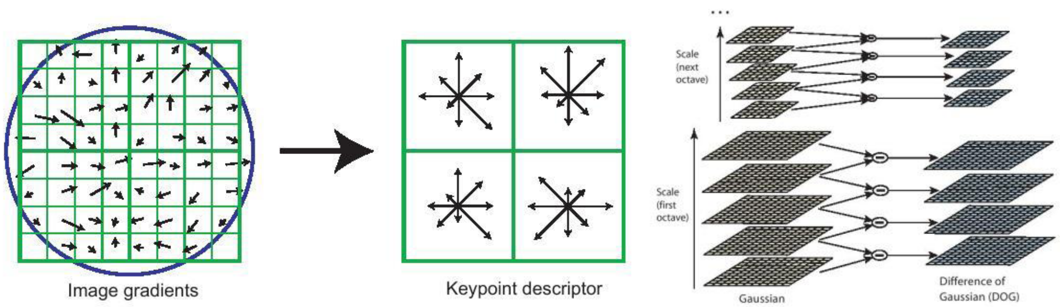
This burrito place is yummy and fun!

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

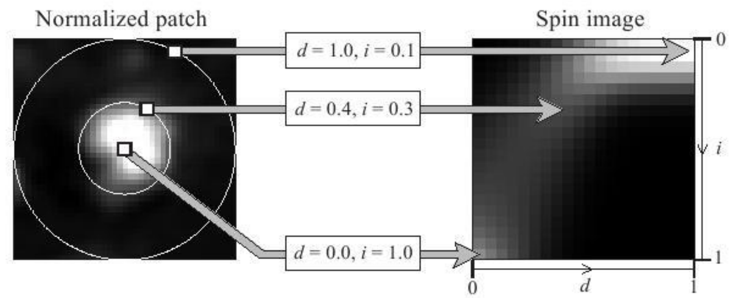
fixed | learned

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

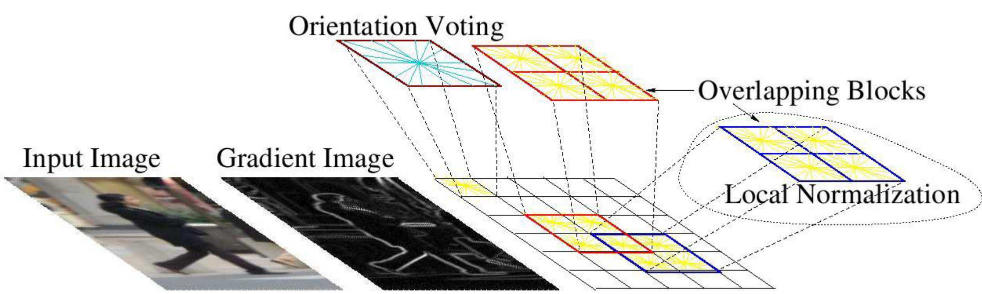
Feature Engineering



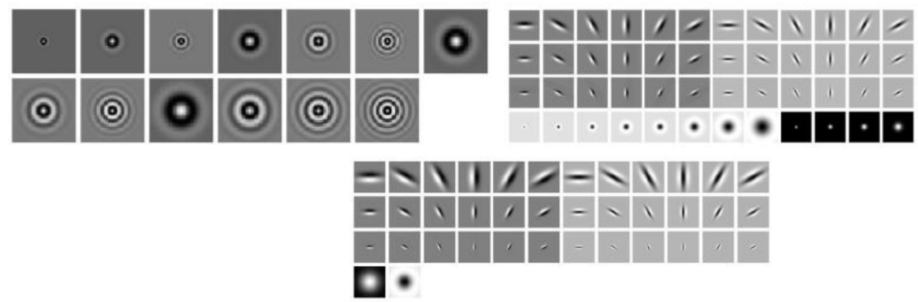
SIFT



Spin Images



HoG



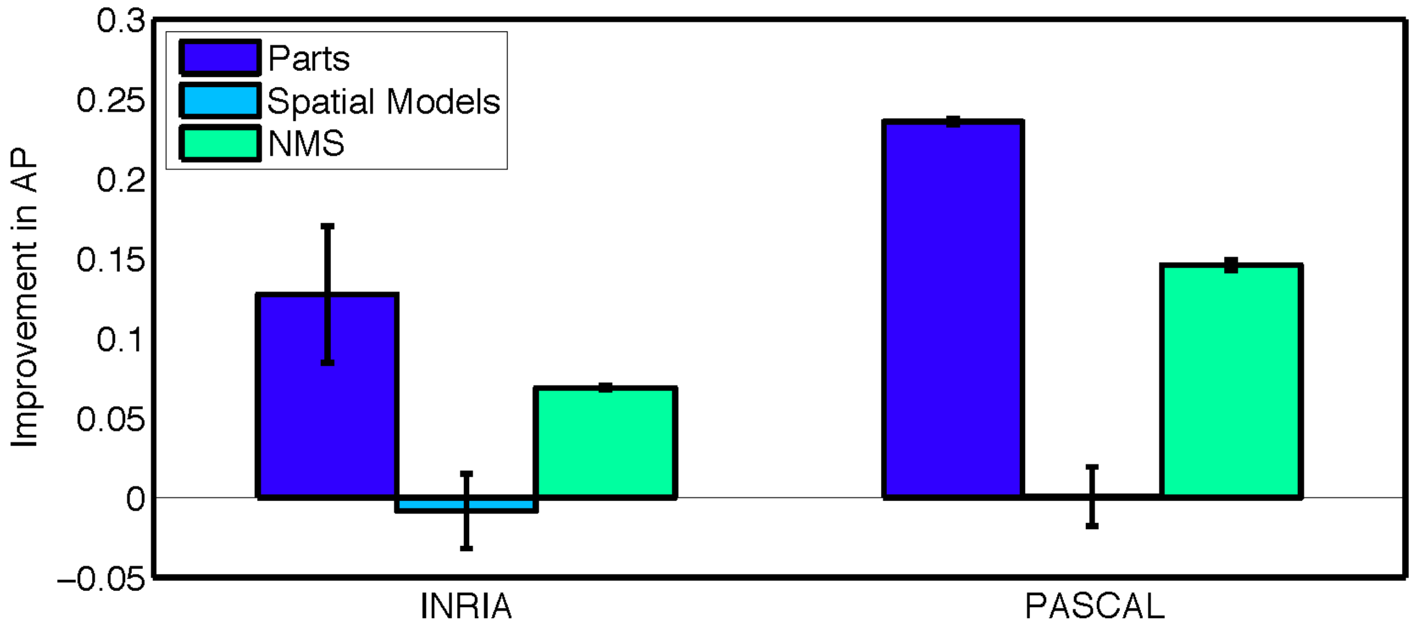
Textons

and many many more….

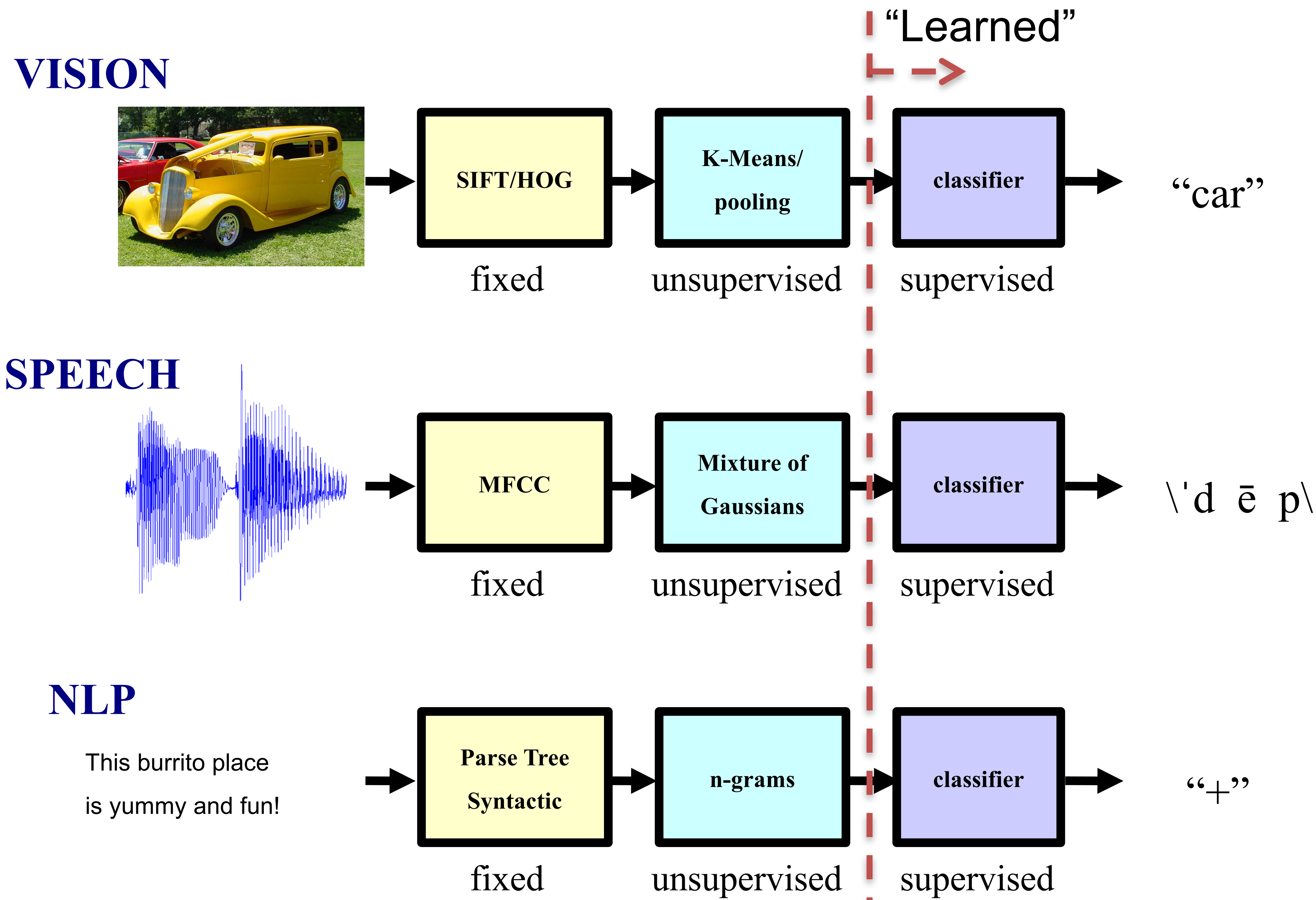
What are the current bottlenecks?

- Ablation studies on DPM [Parikh & Zitnick, CVPR10]
  - Replace every "component" in the model with a human

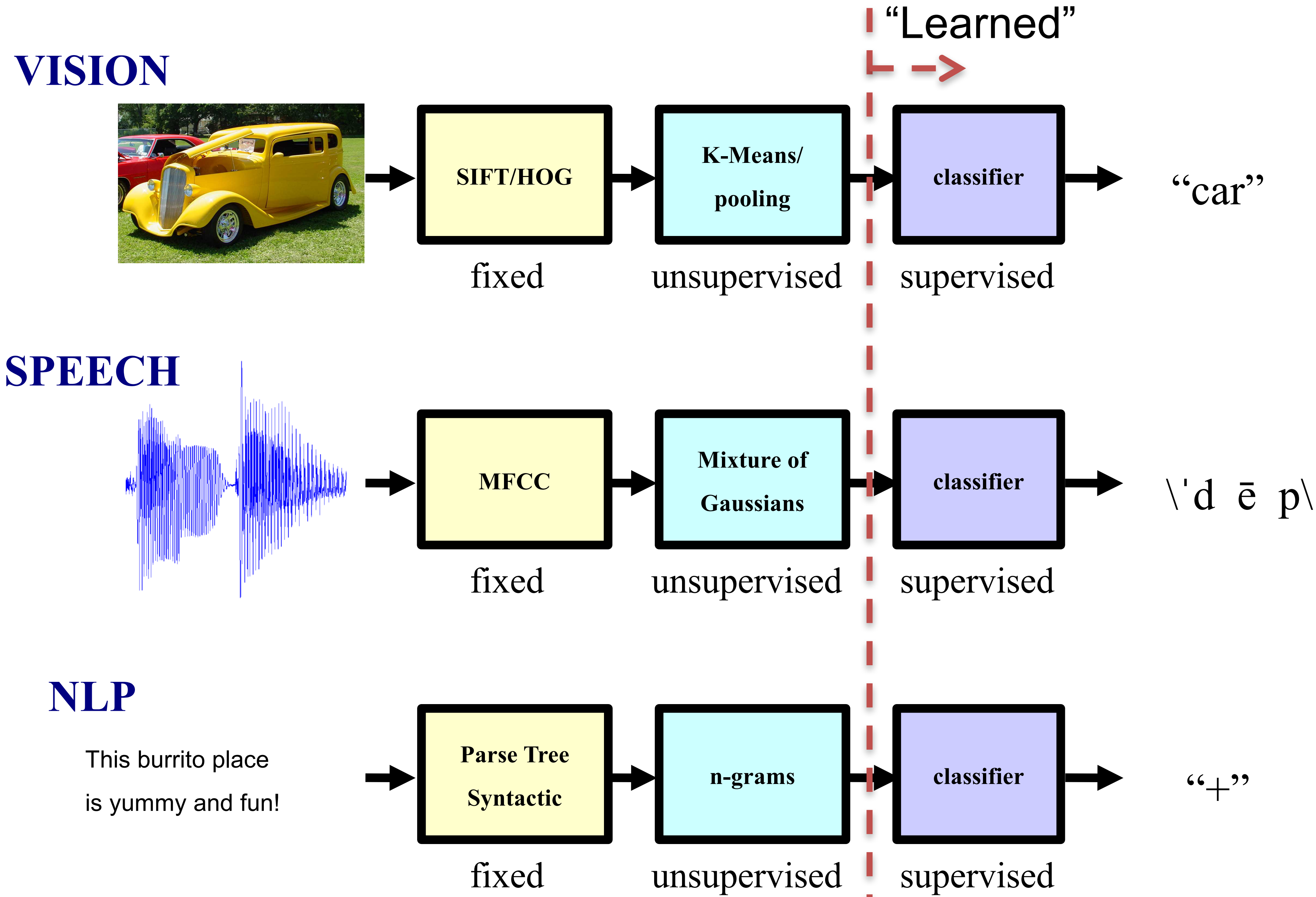
- Key takeaway: "parts" or features are the most important!



Traditional Machine Learning (more accurately)

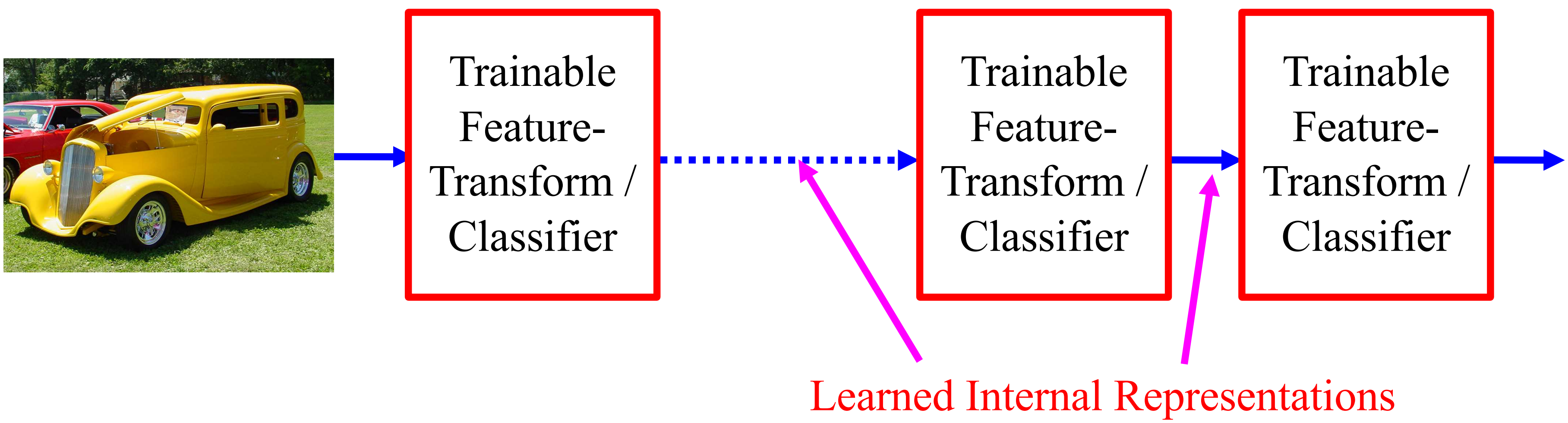


Deep Learning = End-to-End Learning



Deep Learning = End-to-End Learning

- A hierarchy of trainable feature transforms
  - Each module transforms its input representation into a higher-level one.
  - High-level features are more global and more invariant
  - Low-level features are shared among categories



Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

"Shallow" vs Deep Learning

- "Shallow" models

hand-crafted Feature Extractor → "Simple" Trainable Classifier

fixed | 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:

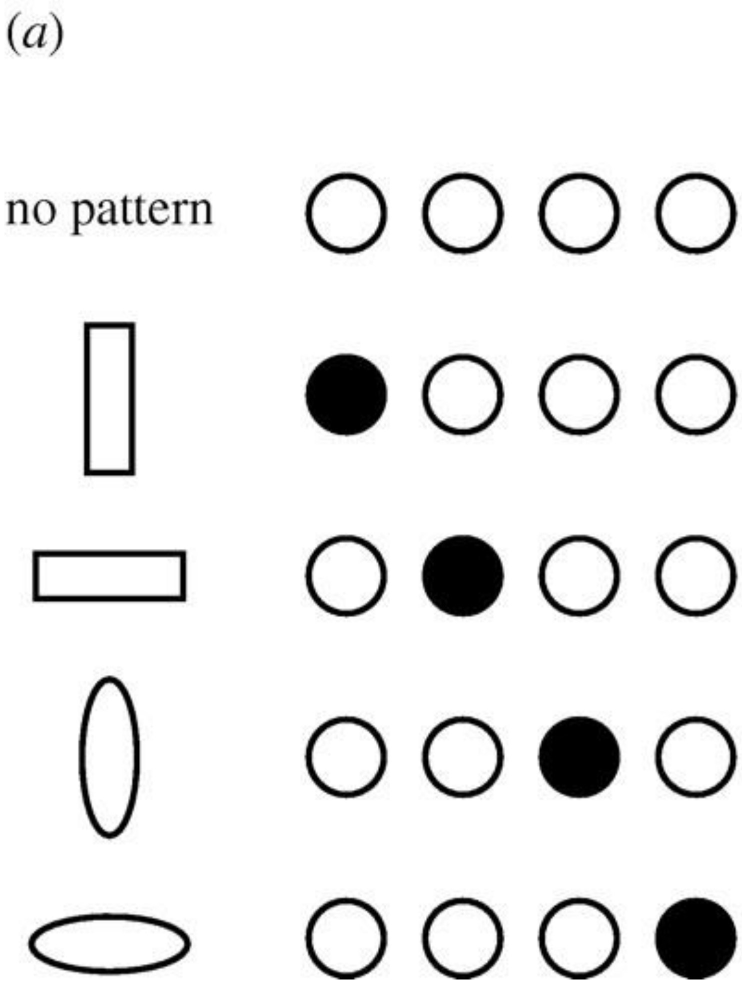
- (Hierarchical) Compositionality
  - Cascade of non-linear transformations
  - Multiple layers of representations

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

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

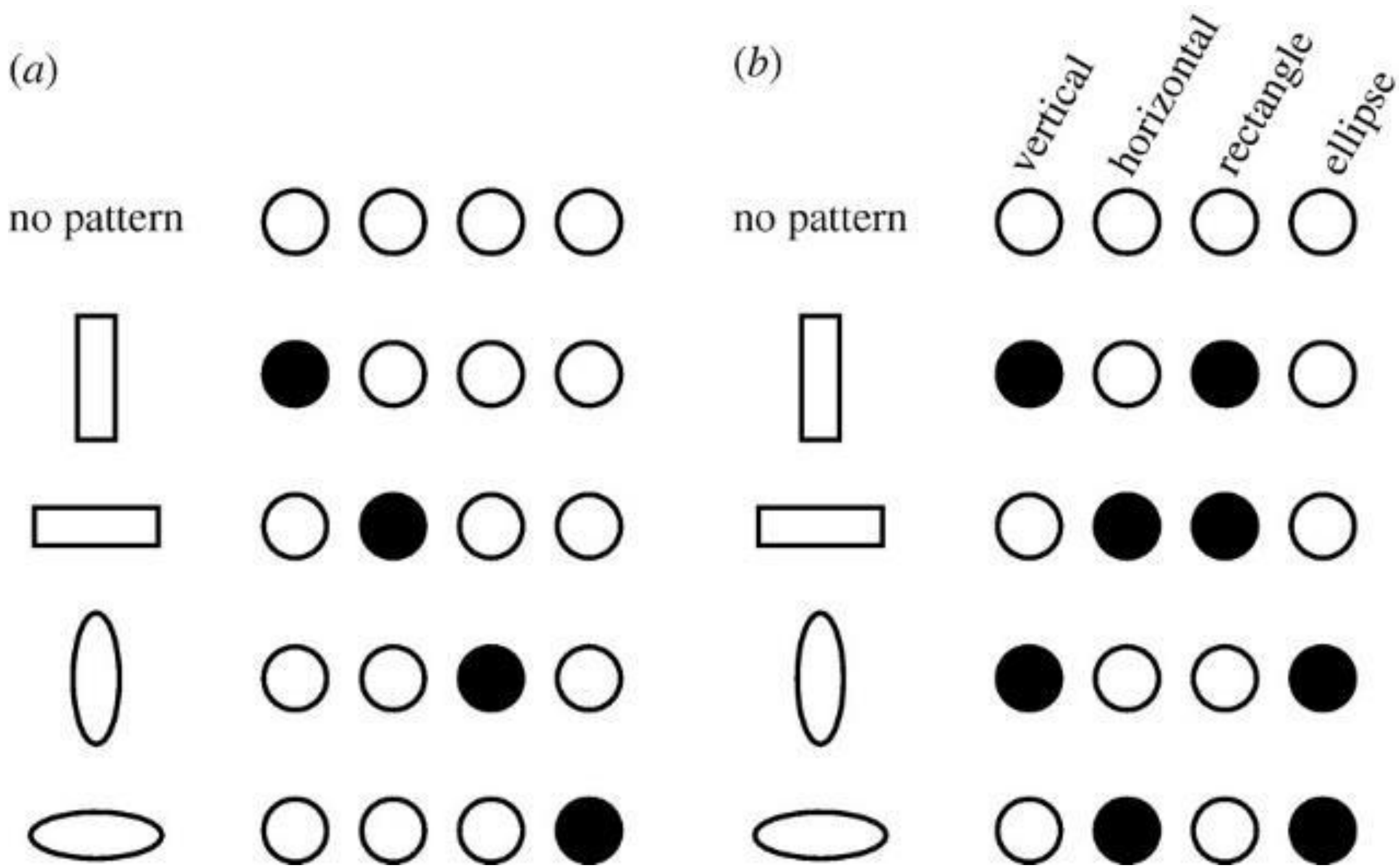
Distributed Representations Toy Example

- Local vs Distributed



Distributed Representations Toy Example

- Can we interpret each dimension?



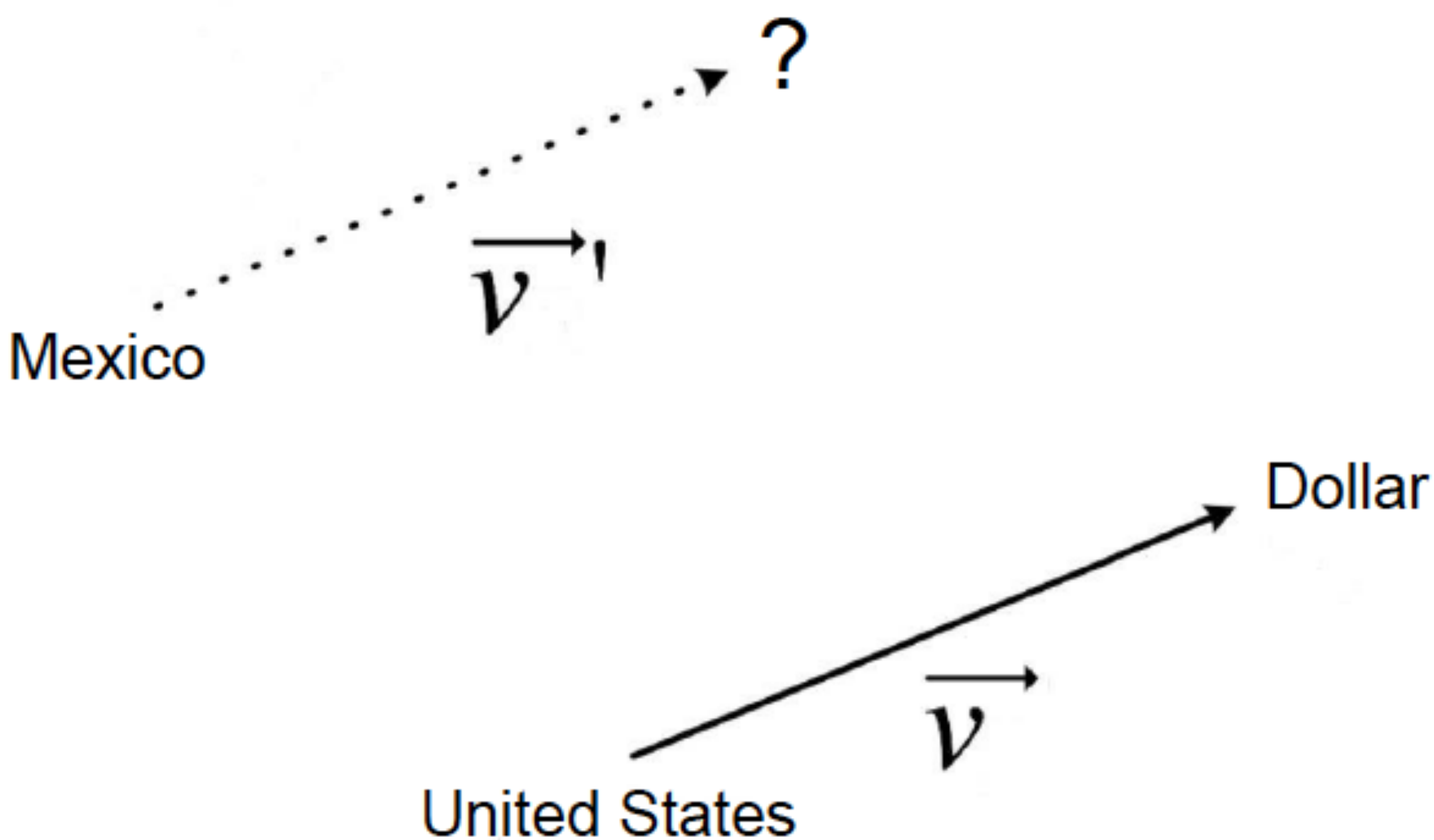
Power of distributed representations!

Local ● ● ○ ● = VR + HR + HE = ?

Distributed ● ● ○ ● = V + H + E $\approx$ ◯

Power of distributed representations!

- United States:Dollar :: Mexico:?



ThisPlusThat.me

the matrix - thoughtful + dumb

Search

How it Works

mbiguated into *+1 the_matrix -1 thoughtful +1 dumb* in 0.0 seconds from ip-10-32-114-31

FILM, W FILM, NETFLIX TITLE,

Blade II

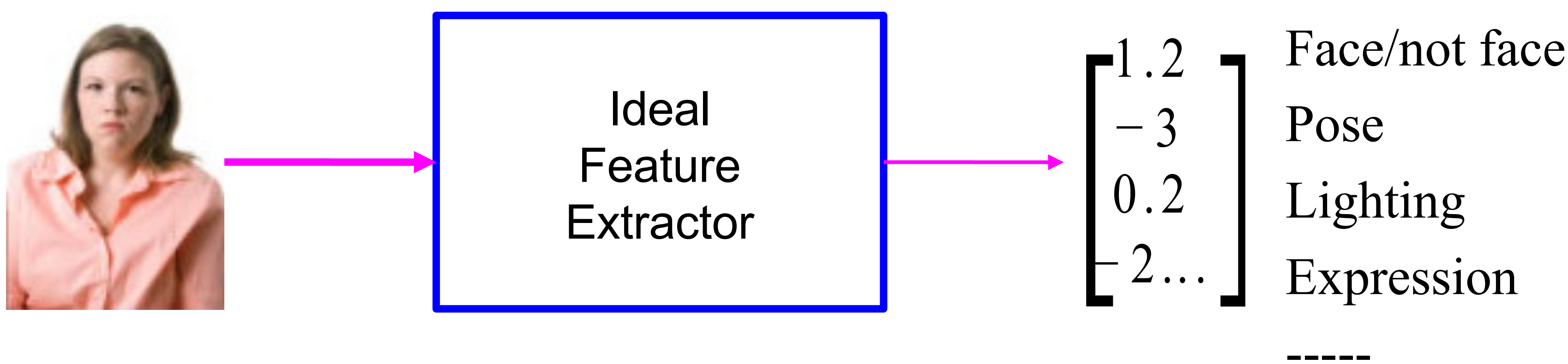
Blade II is a 2002 American vampire superhero action film base
Marvel Comics character Blade. It is the sequel of the first film a
part of the Blade film series. It was written by David S. Goyer, w
previous film. Guillermo del Toro was signed in to d...

Horror Film

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

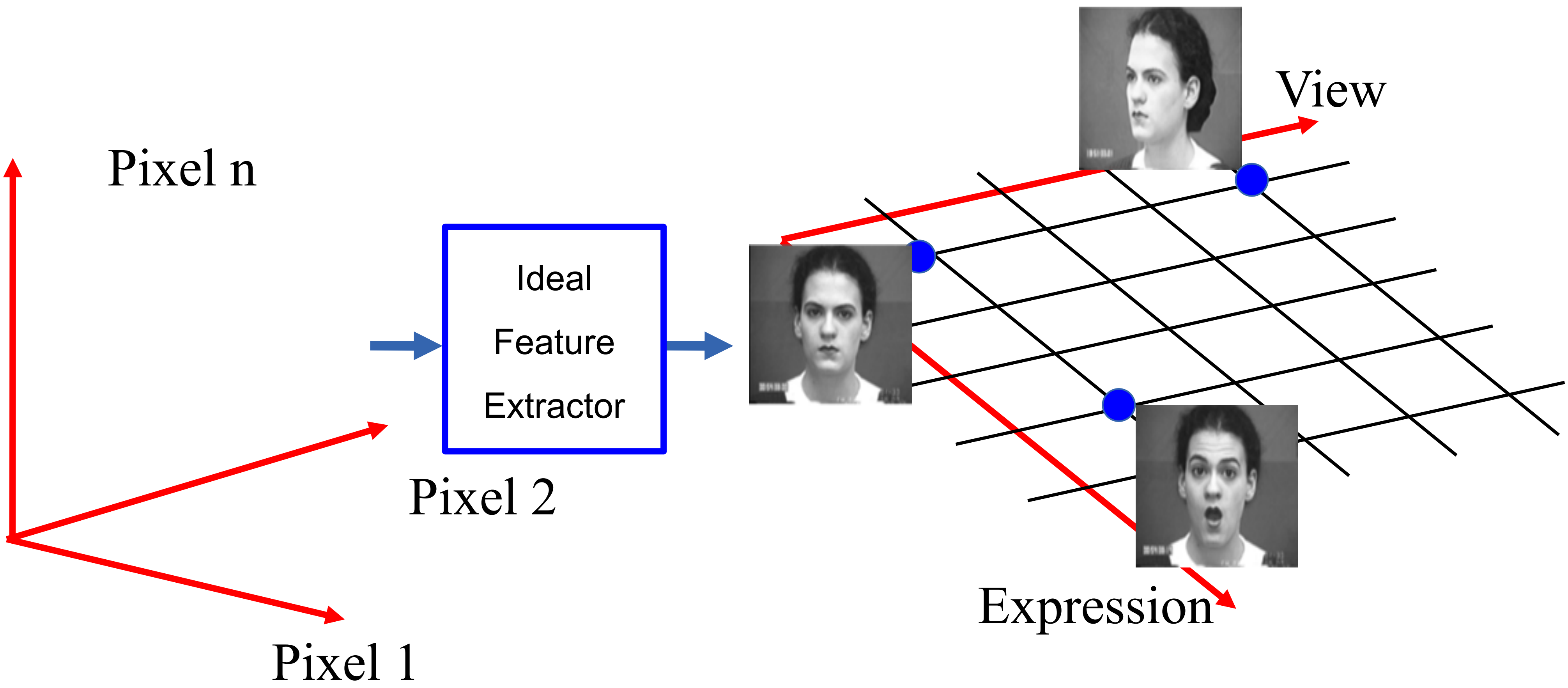
Power of distributed representations!

- Example: all face images of a person
  - 1000x1000 pixels = 1,000,000 dimensions
  - But the face has 3 cartesian coordinates and 3 Euler angles
  - And humans have less than about 50 muscles in the face
  - Hence the manifold of face images for a person has <56 dimensions
- The perfect representations of a face image:
  - Its coordinates on the face manifold
  - Its coordinates away from the manifold



Power of distributed representations!

The Ideal Disentangling Feature Extractor



Distributed Representations

- Q: What objects are in the image? Where?

Ideal Feature Extractor

- window, top-left
- clock, top-middle
- shelf, left
- drawing,middle
- statue, bottom left
- …
- hat, bottom right

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

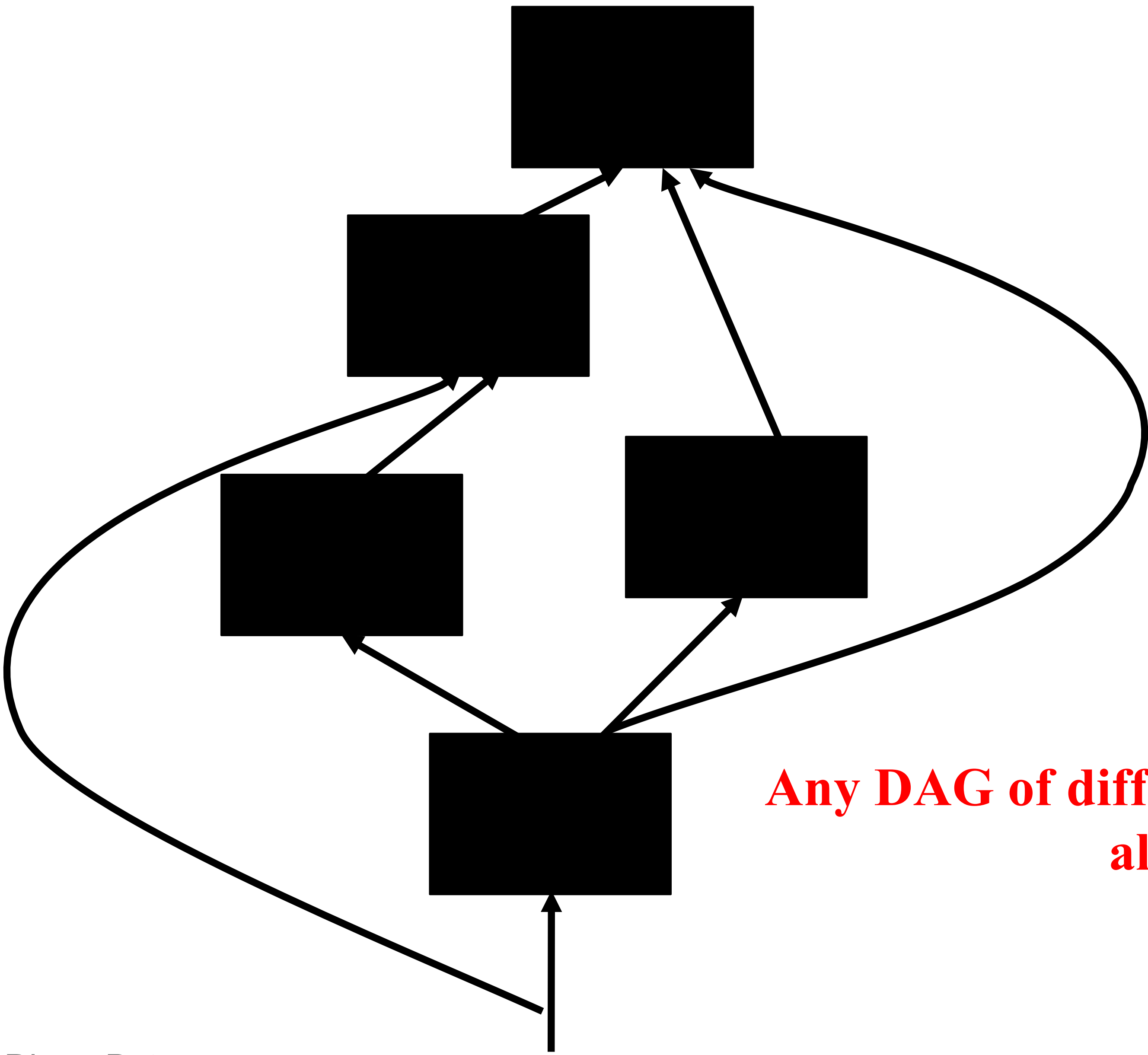
- *"Maybe the molecule didn't go to graduate school"*
  - Will Welch defending the success of his approximate molecular screening algorithm, given that he's a computer scientist, not a chemist

Database Screening for HIV Protease Ligands: The Influence of Binding-Site Conformation and Representation on Ligand Selectivity", Volker Schnecke, Leslie A. Kuhn, Proceedings of the Seventh International Conference on Intelligent Systems for Molecular Biology, Pages 242-251, AAAI Press, 1999.

Benefits of Deep/Representation Learning

- Modularity!
- Plug and play architectures!

Differentiable Computation Graph



Any DAG of differentialble modules is allowed!

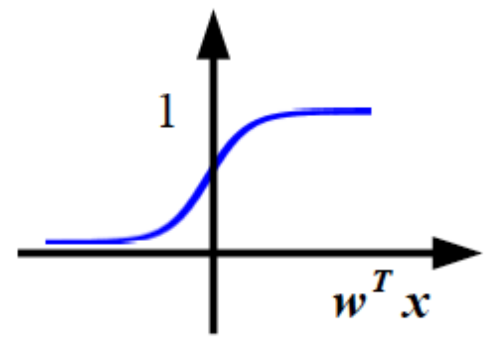
Linear Classifier: Logistic Regression

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

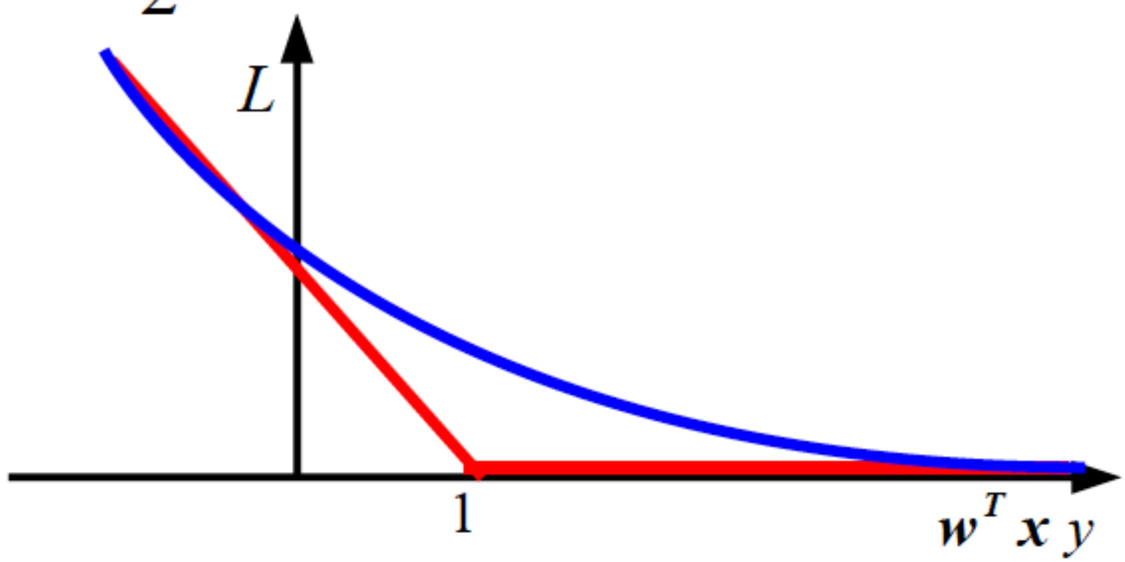
Binary label: $y \in \{-1, +1\}$

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

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



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

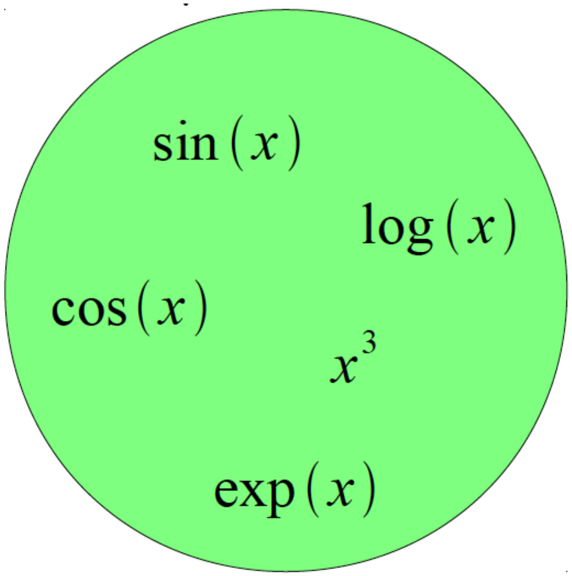


Log Loss

Ranzato

Logistic Regression as a Cascade

Given a library of simple functions



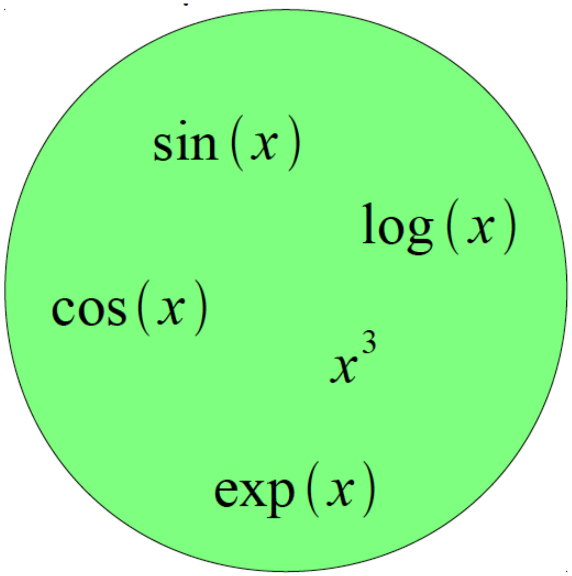
Compose into a complicate function

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

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

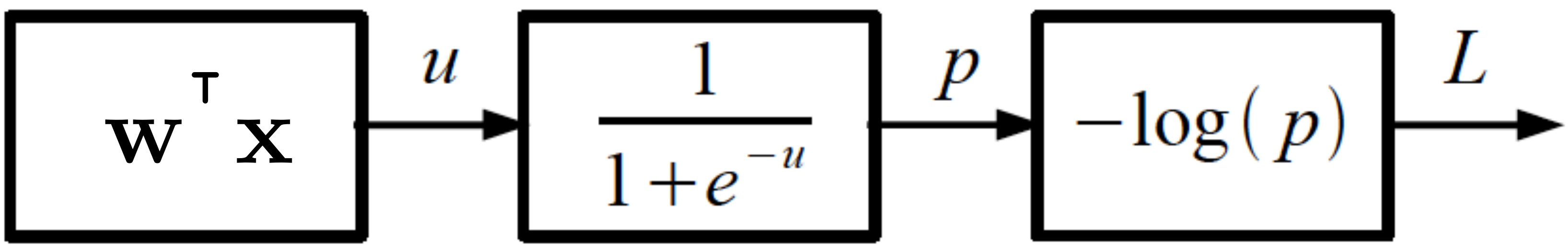
Logistic Regression as a Cascade

Given a library of simple functions

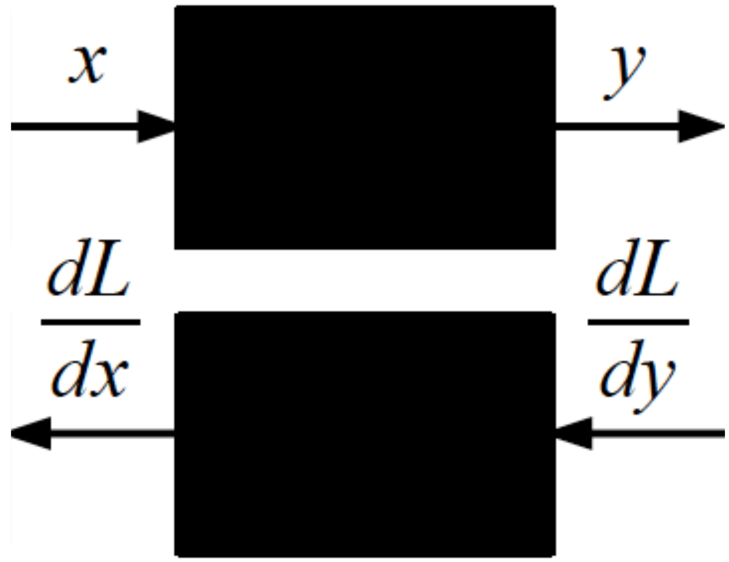


Compose into a complicate function

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



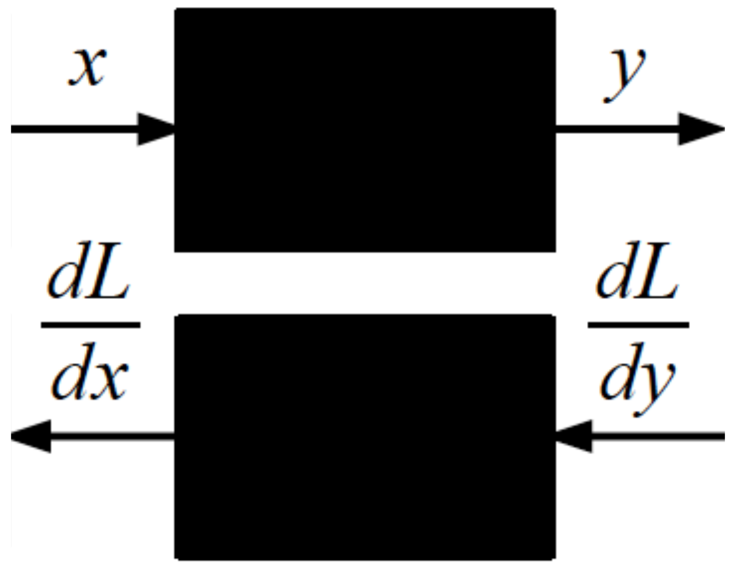
Chain Rule



Given $y(x)$ and dL/dy,

What is dL/dx ?

Chain Rule

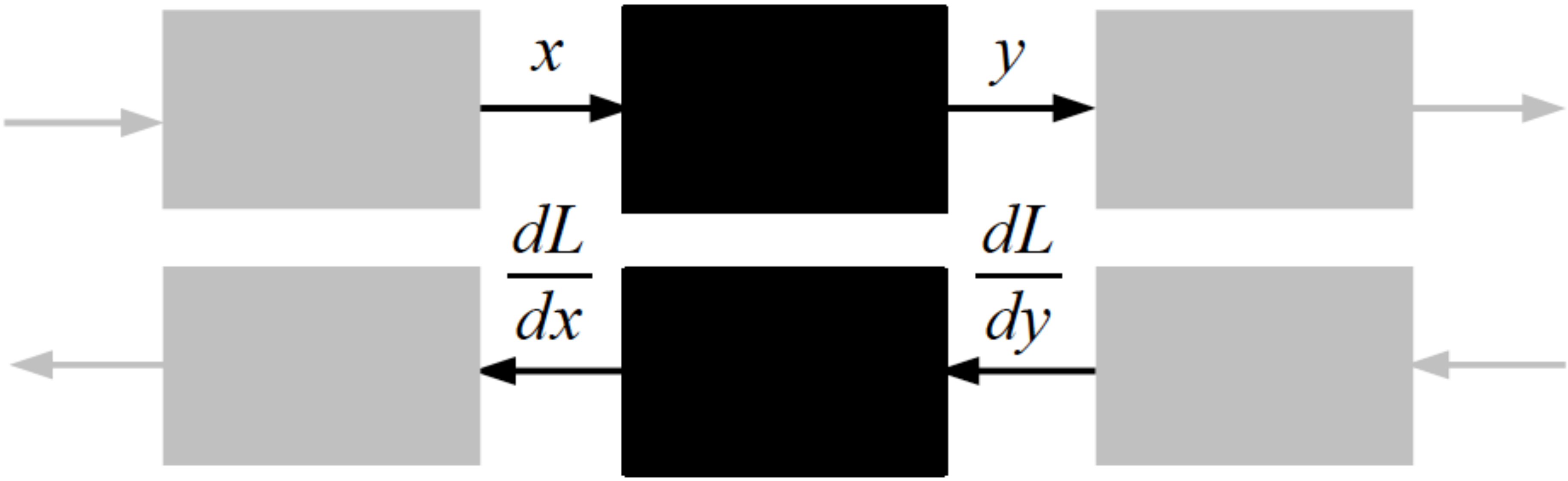


Given $y(x)$ and dL/dy,

What is dL/dx ?

$$\frac{dL}{dx} = \frac{dL}{dy} \cdot \frac{dy}{dx}$$

Chain Rule: All local

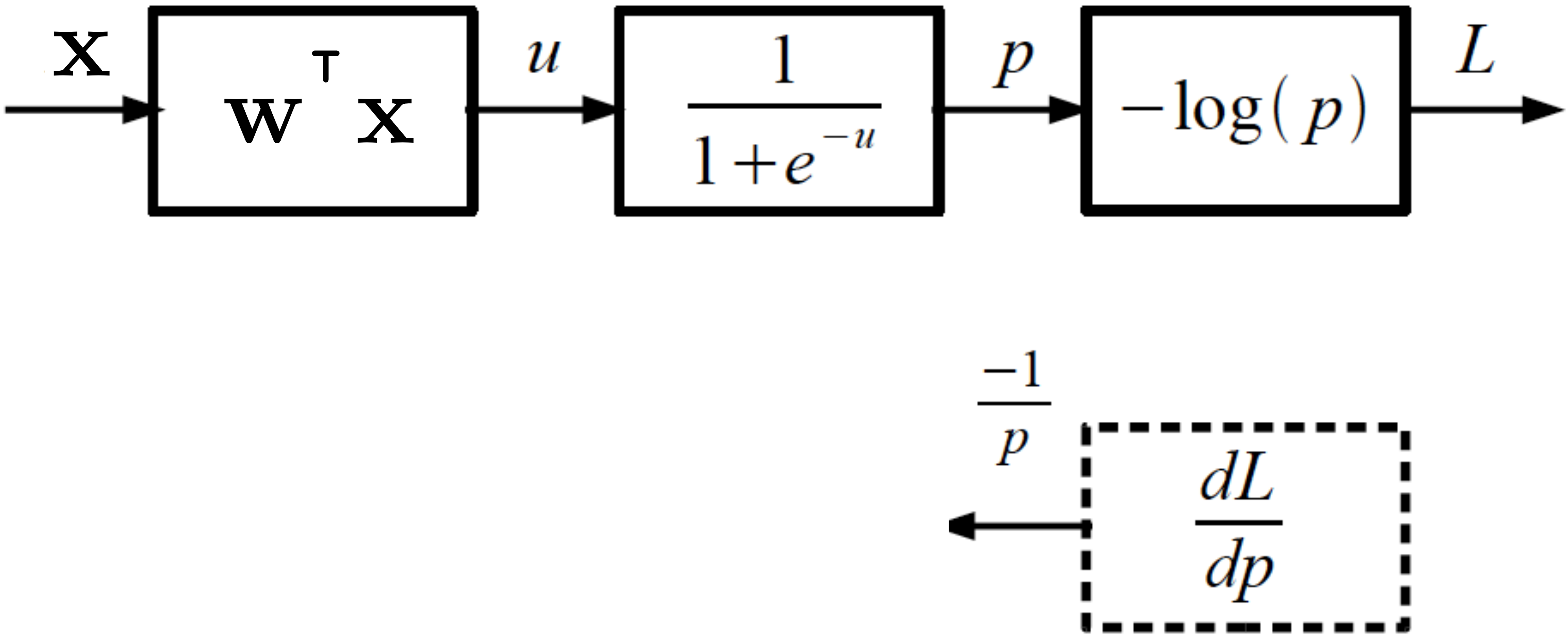


Given $y(x)$ and dL/dy,

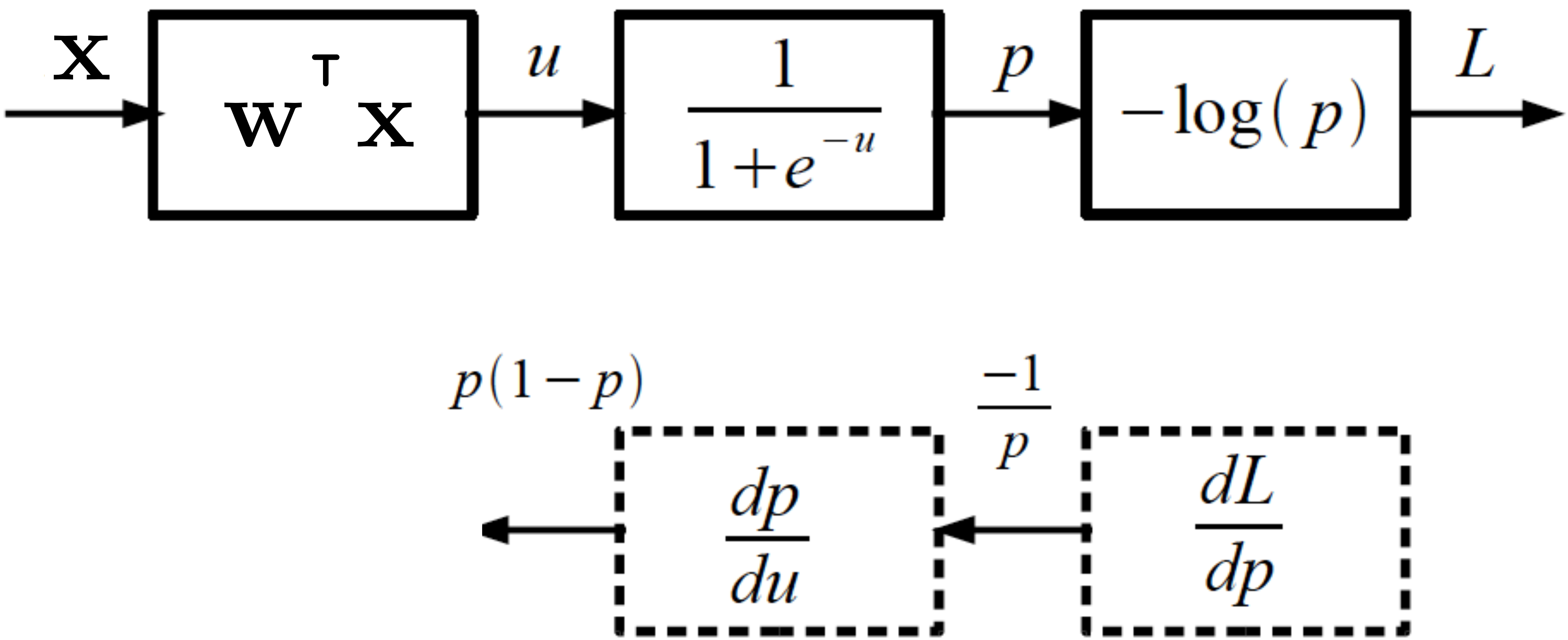
What is dL/dx ?

$$\frac{dL}{dx} = \frac{dL}{dy} \cdot \frac{dy}{dx}$$

Logistic Regression as a Cascade

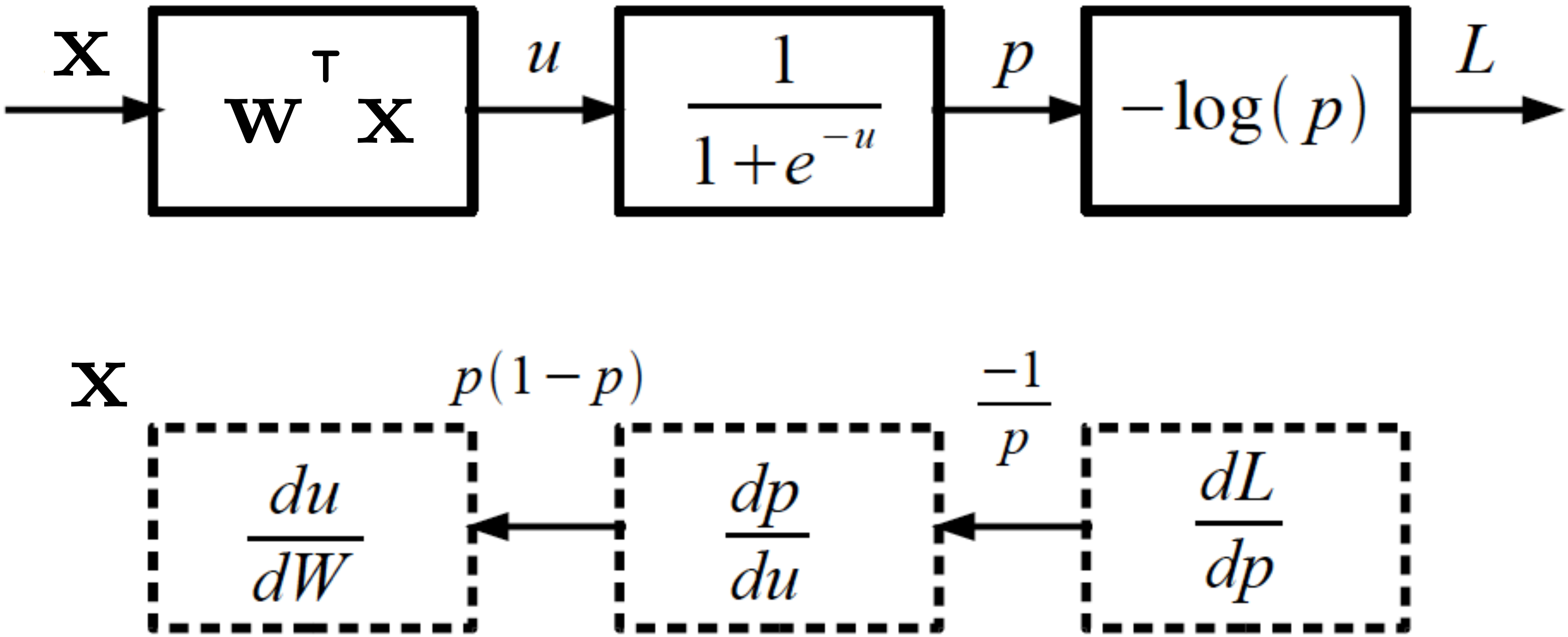


Logistic Regression as a Cascade



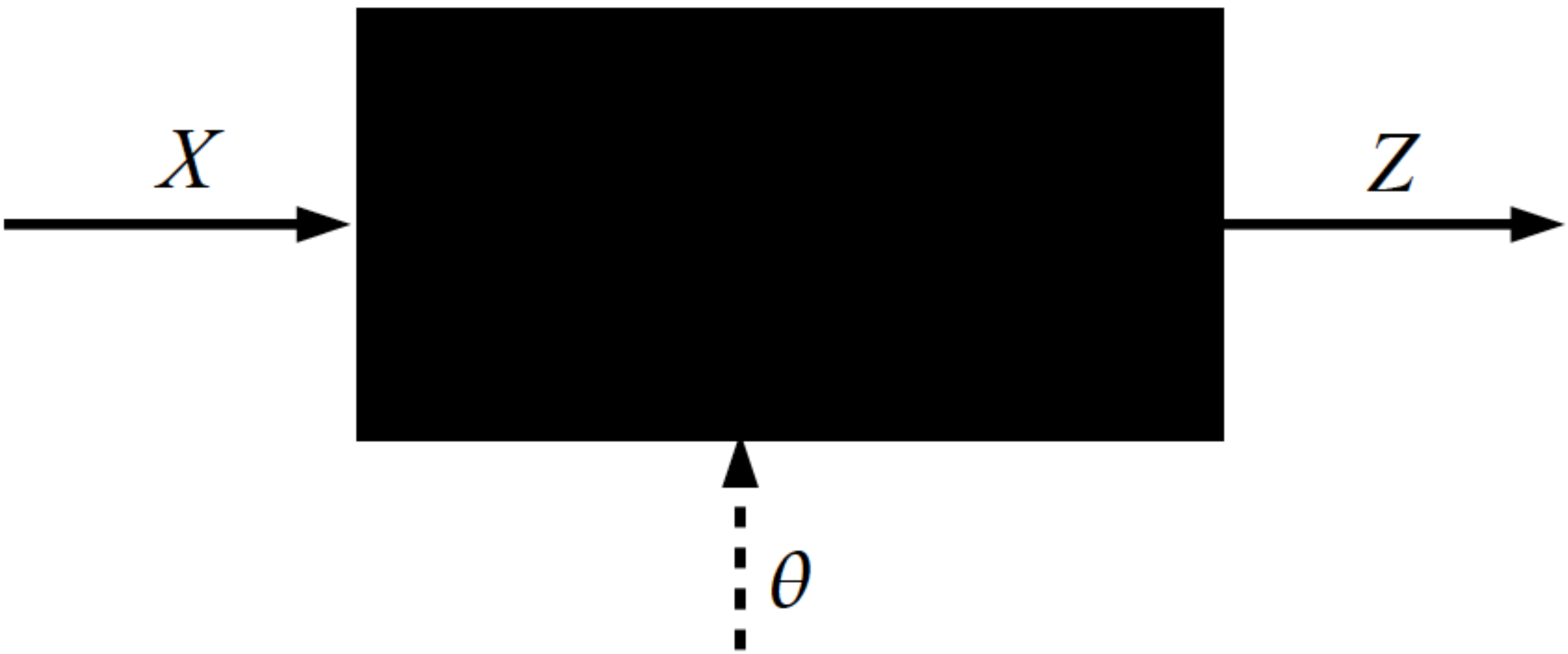
 Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

Logistic Regression as a Cascade

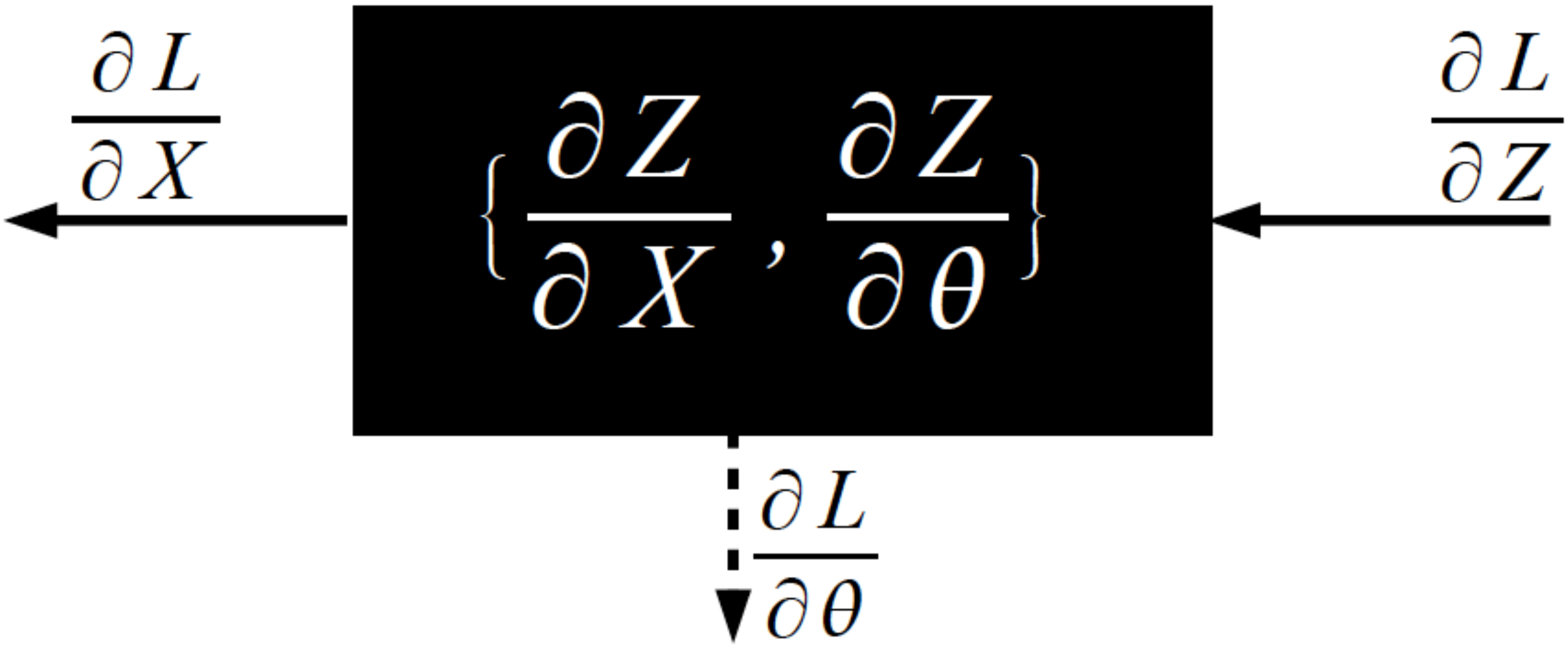


$$\frac{dL}{dW} = \frac{dL}{dp} \cdot \frac{dp}{du} \cdot \frac{du}{dW} = (p-1)\,\mathbf{x}$$

Key Computation: Forward-Prop

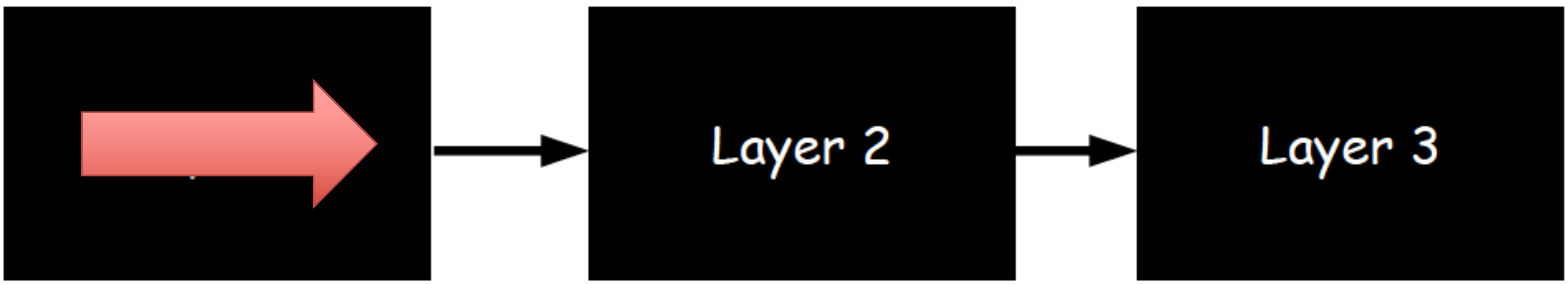


Key Computation: Back-Prop



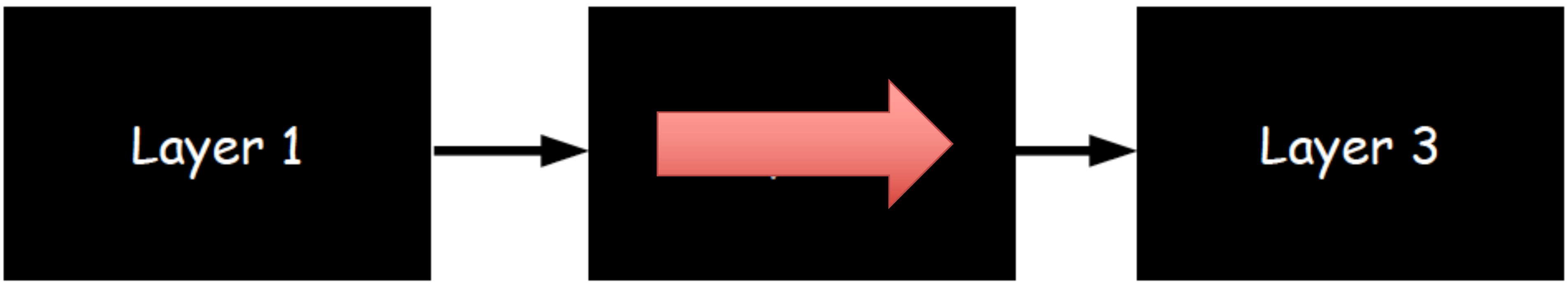
Neural Network Training

- Step 1: Compute Loss on mini-batch [F-Pass]



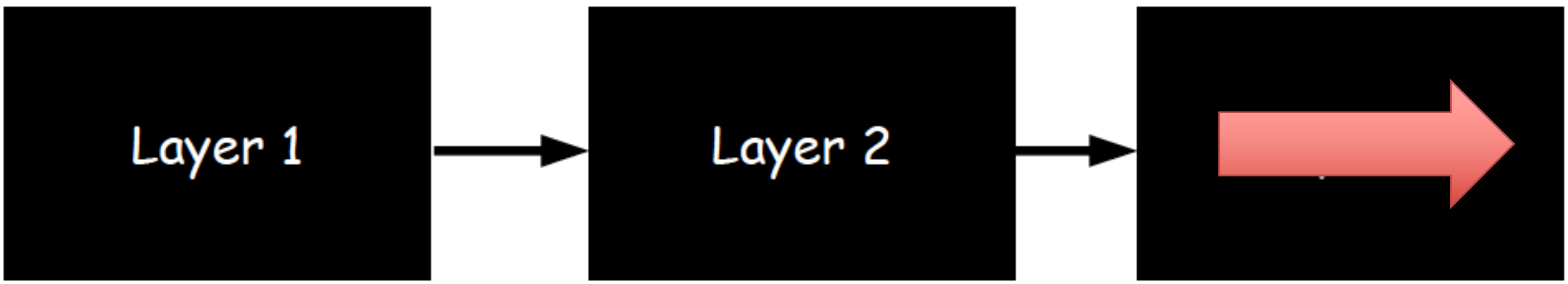
Neural Network Training

- Step 1: Compute Loss on mini-batch [F-Pass]



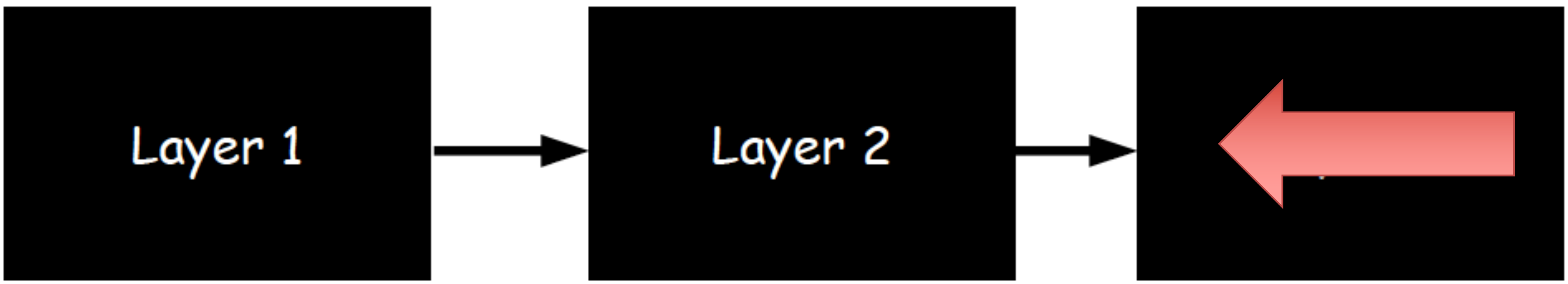
Neural Network Training

- Step 1: Compute Loss on mini-batch [F-Pass]



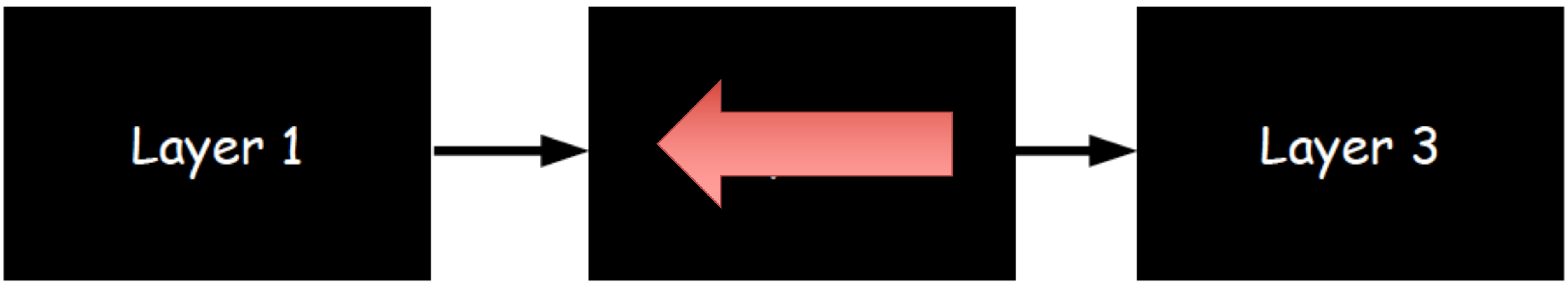
Neural Network Training

- Step 1: Compute Loss on mini-batch [F-Pass]
- Step 2: Compute gradients wrt parameters [B-Pass]



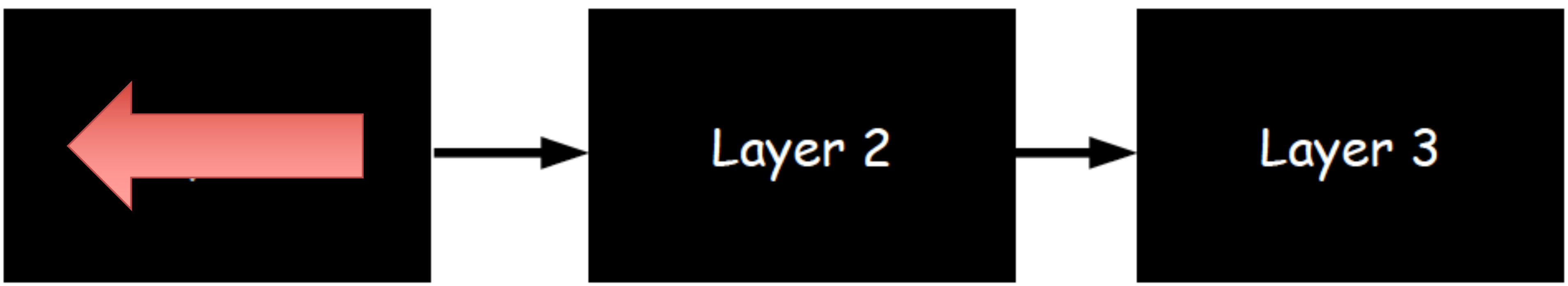
Neural Network Training

- Step 1: Compute Loss on mini-batch [F-Pass]
- Step 2: Compute gradients wrt parameters [B-Pass]



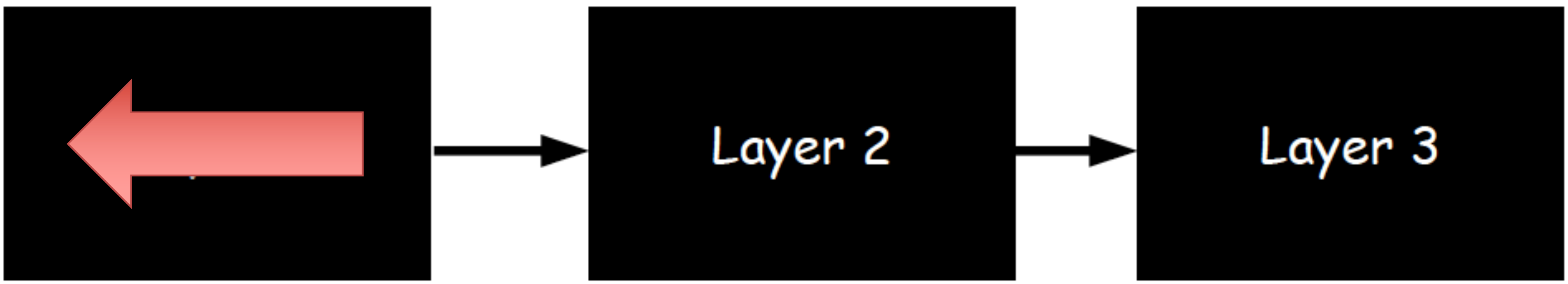
Neural Network Training

- Step 1: Compute Loss on mini-batch [F-Pass]
- Step 2: Compute gradients wrt parameters [B-Pass]



Neural Network Training

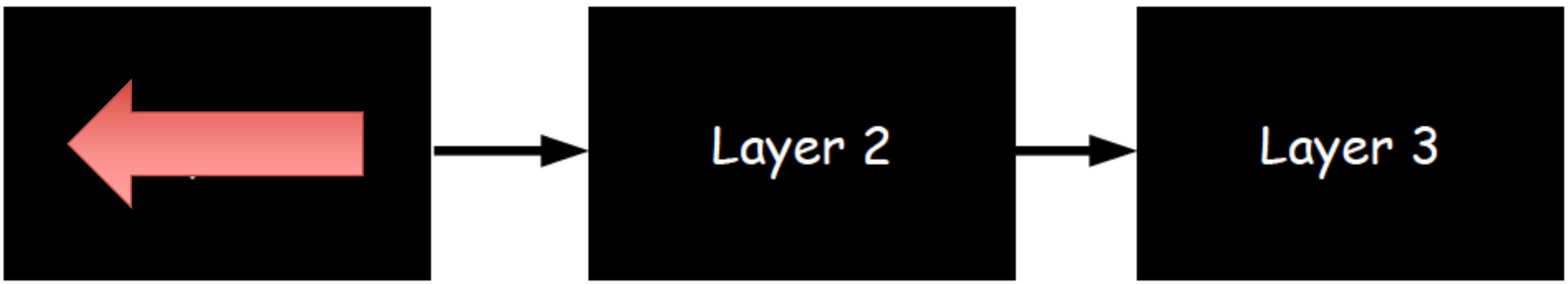
- Step 1: Compute Loss on mini-batch [F-Pass]
- Step 2: Compute gradients wrt parameters [B-Pass]
- Step 3: Use gradient to update parameters



$$\theta \leftarrow \theta - \eta \frac{dL}{d\theta}$$

Neural Network Training

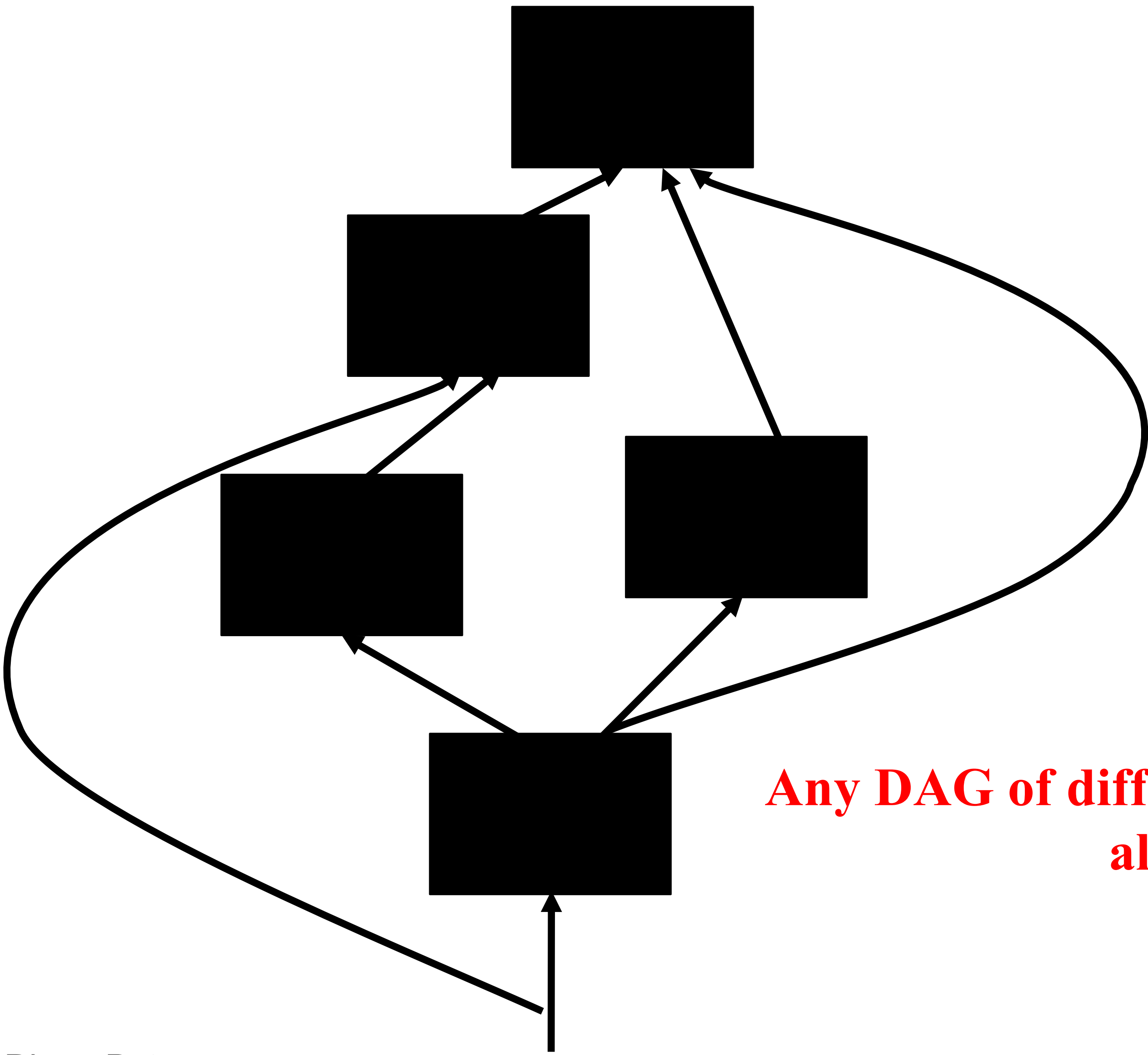
- Step 1: Compute Loss on mini-batch [F-Pass]
- Step 2: Compute gradients wrt parameters [B-Pass]
- Step 3: Use gradient to update parameters
  - With momemtum



$$\theta \leftarrow \theta - \eta \Delta$$

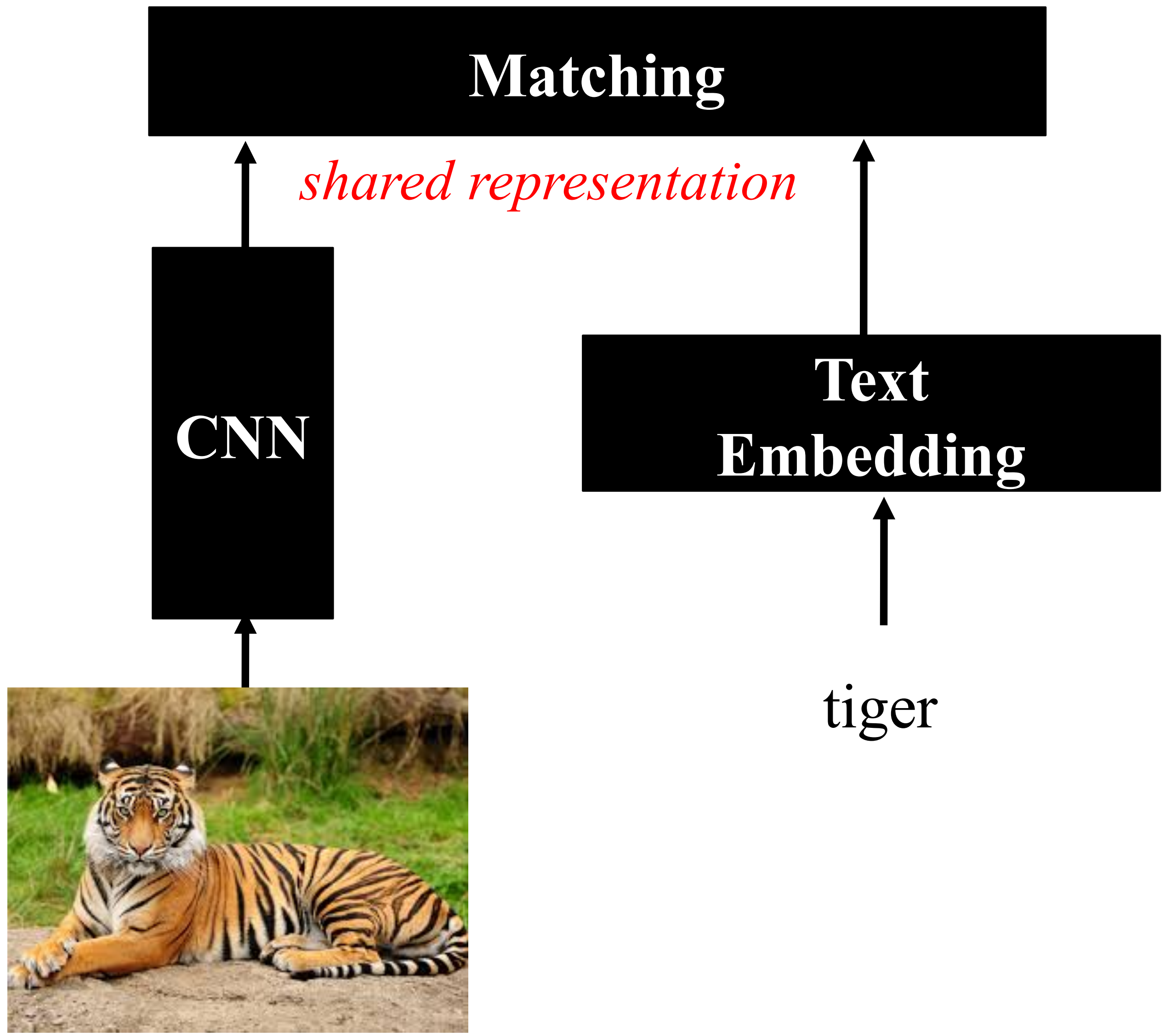
$$\Delta \leftarrow 0.9 \Delta + \frac{\partial L}{\partial \theta}$$

Differentiable Computation Graph



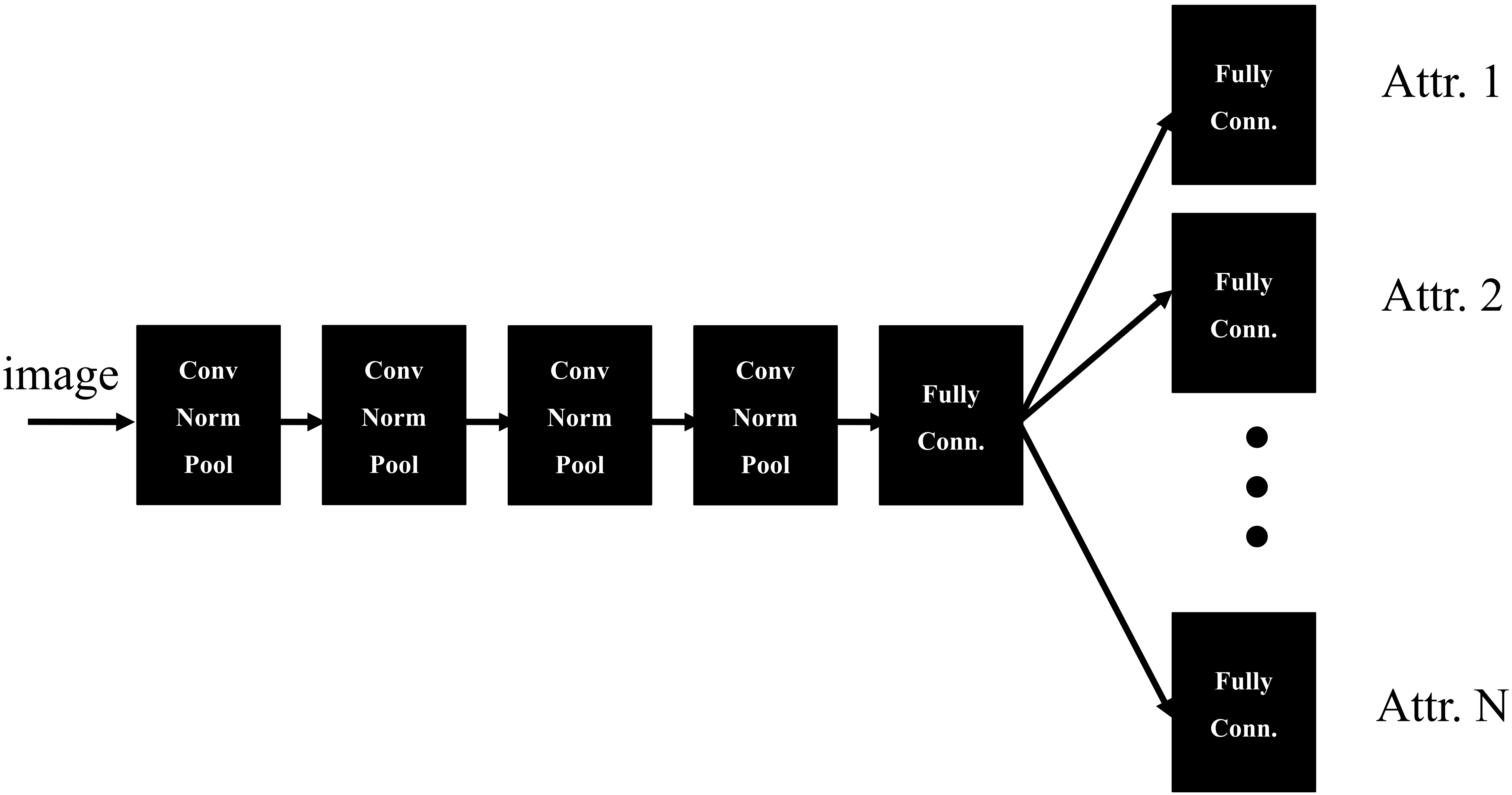
Any DAG of differentialble modules is allowed!

Fancier Architectures: Multi-Modal



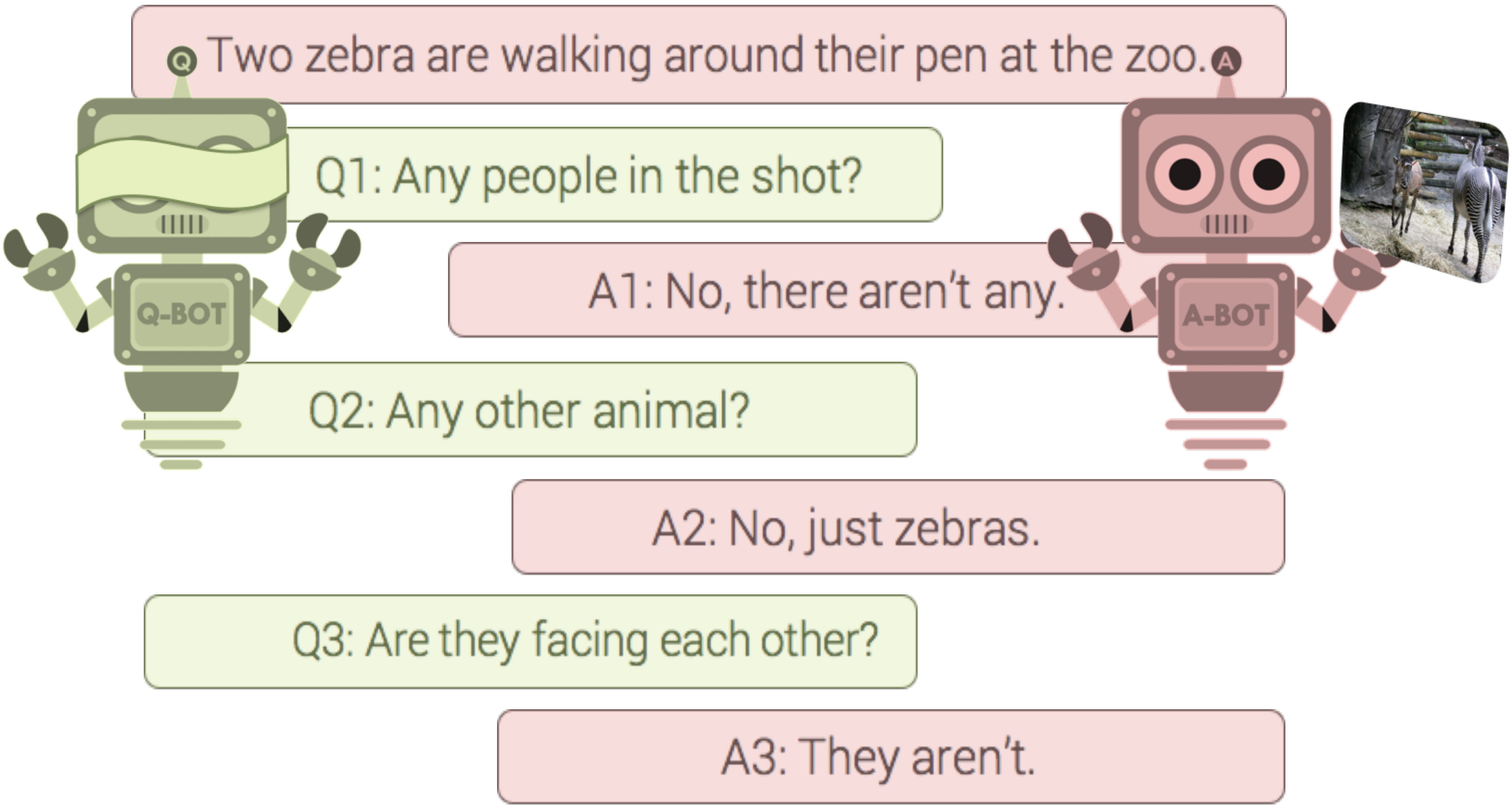
Frome et al. “Devise: a deep visual semantic embedding model” NIPS 2013

Fancier Architectures: Multi-Task

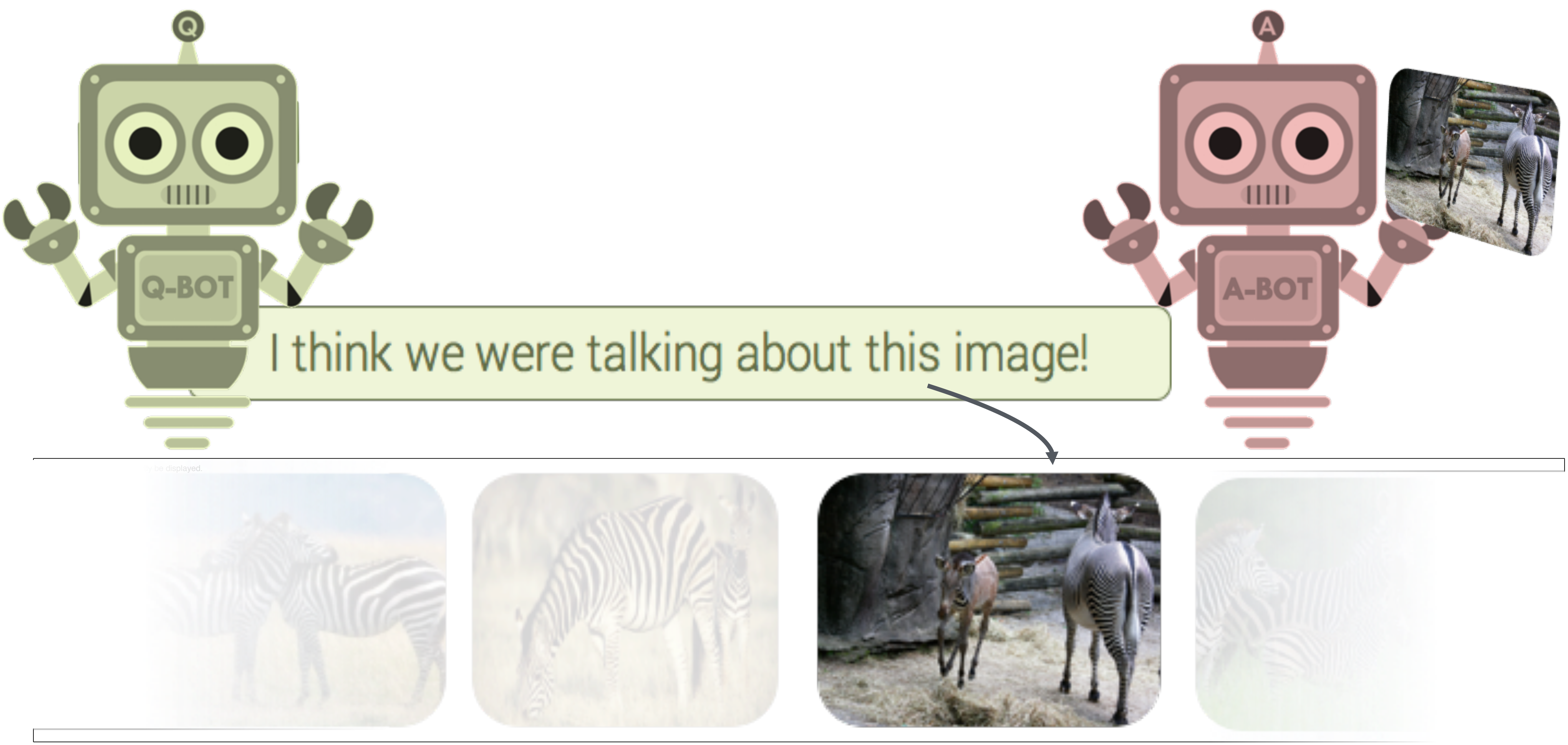


Zhang et al. "PANDA.." CVPR 2014

GuessWhich: Image Guessing Game



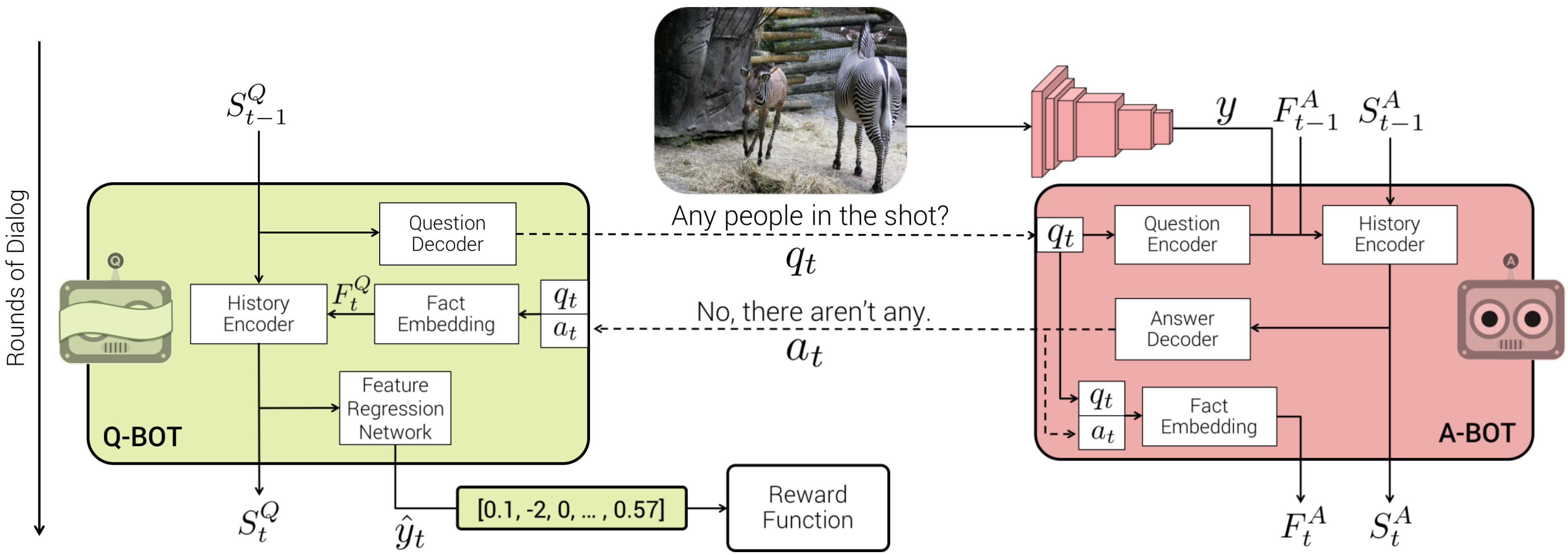
GuessWhich: Image Guessing Game



Policy Networks

Q-Bot
$\pi_Q(q_t|S^Q_{t-1})$

A-Bot
$\pi_A(a_t|S^A_{t-1})$



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

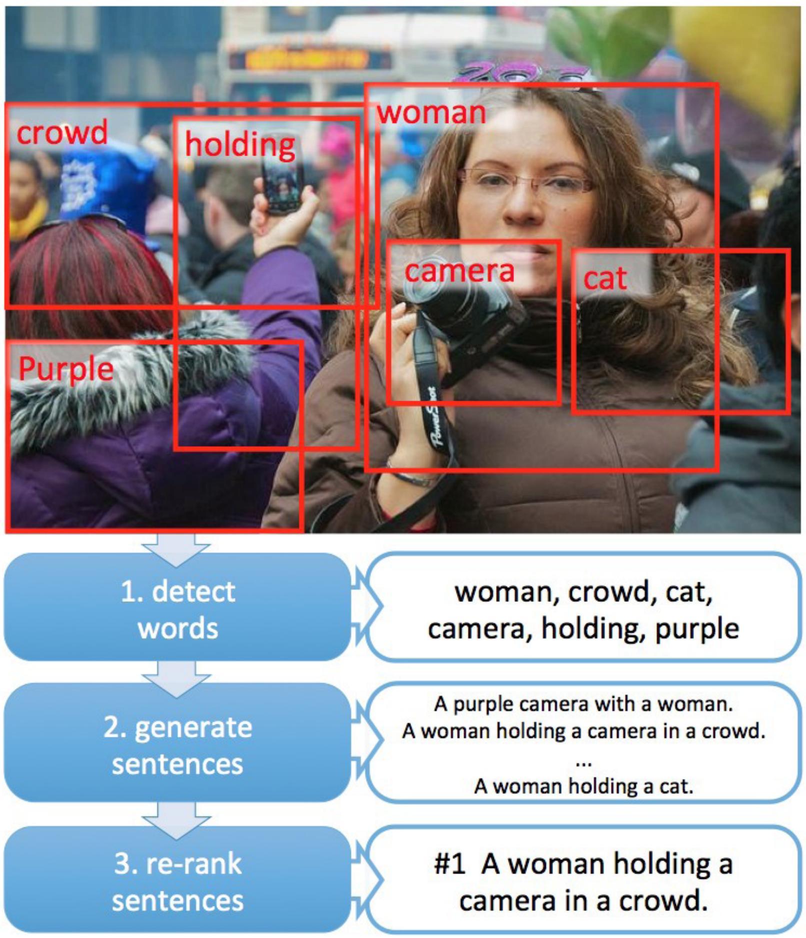
- Standard response #2
  - "Yes, but it often works!"

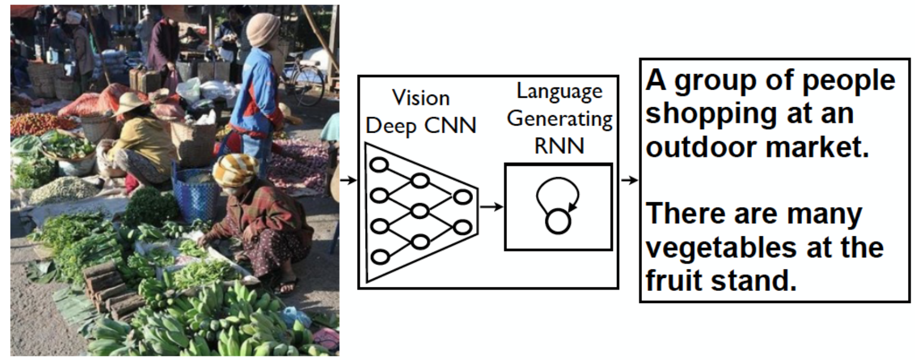
Problems with Deep Learning

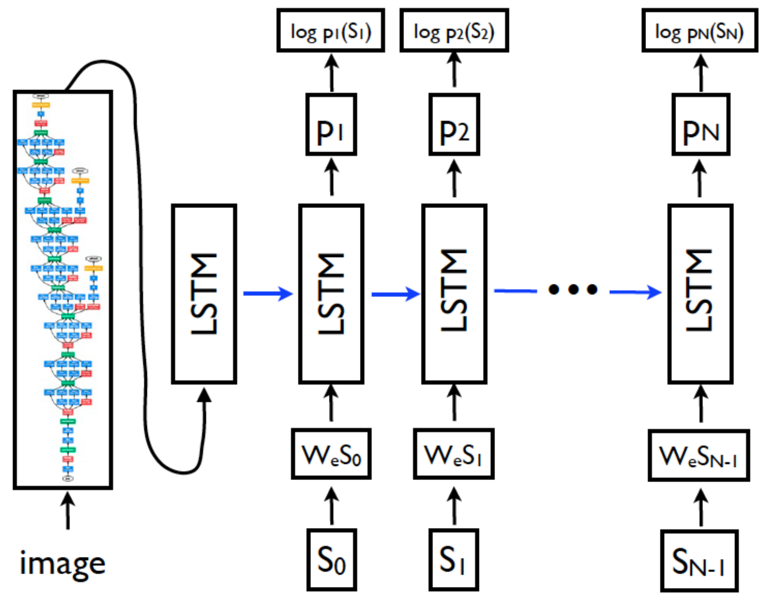
- Problem#2: Lack of interpretability
  - Hard to track down what’s failing
  - Pipeline systems have “oracle” performances at each step
  - In end-to-end systems, it’s hard to know why things are not working

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!"

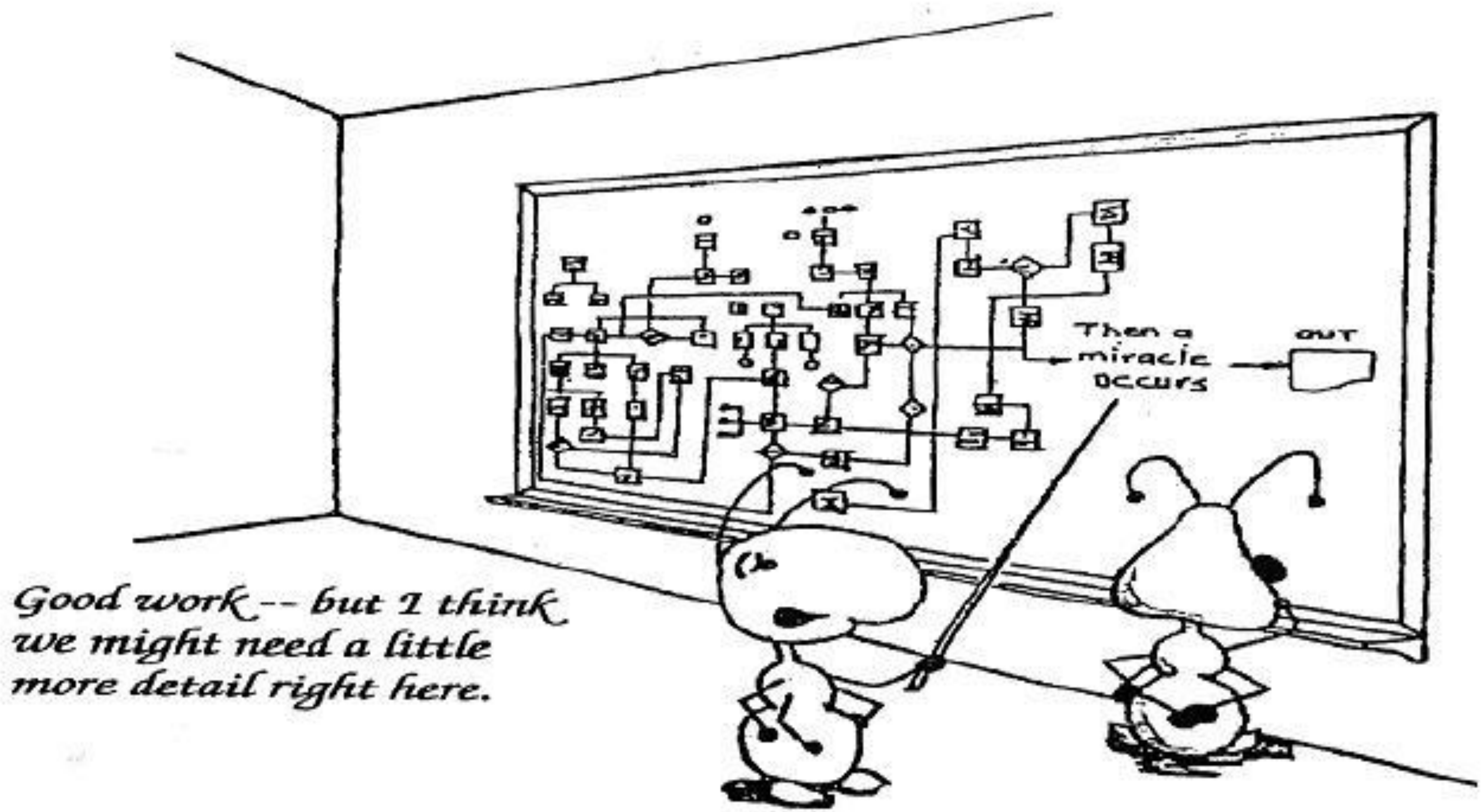
Problems with Deep Learning

- 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

What is this class about?

- Firehose of arxiv

Arxiv Fire Hose

PhD Student

Deep Learning papers

Cornell University Library

arXiv.org

So, what *is* this class?

- Goal:
  - After taking this class, you should be able to pick up the latest Arxiv paper and easily understand it.
    - CNNs
    - RNNs
    - Deep Reinforcement Learning
    - Variational Auto Encoders
    - GANs
    - Vision, Language, Agents in Environment

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

    (think ICLR/CVPR papers as outcomes)

What this class is NOT

- NOT the goal:
  - Teaching a toolkit. “Intro to TensorFlow/PyTorch”
  - Intro to Machine Learning
  - “How to apply Deep Learning to your domain”

- NOT the target audience:
  - Undergraduate/Masters students looking to graduate with a DL class on their resume.

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.

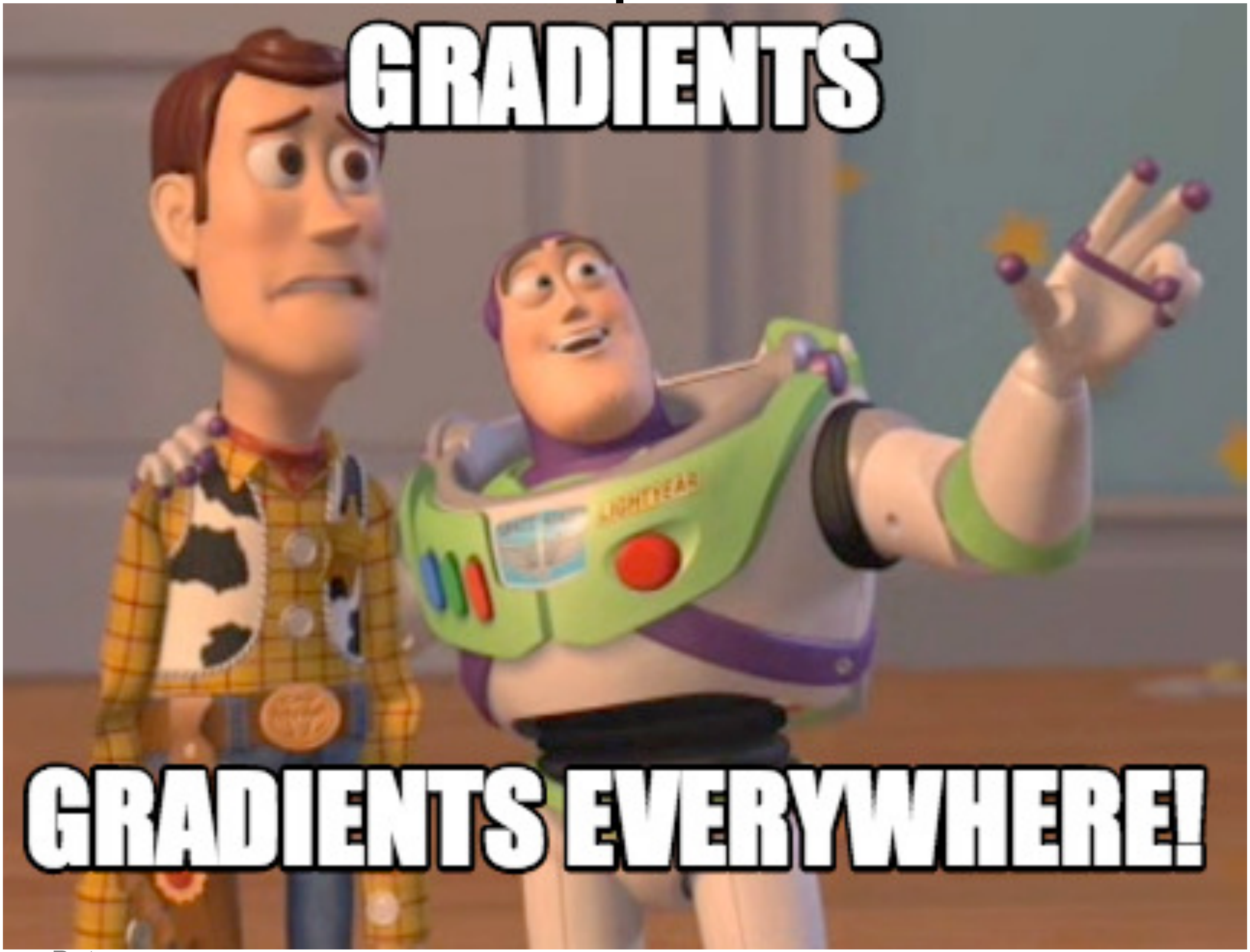
Topics Covered in Intro to ML

- Basics of Statistical Learning
  - Loss function, MLE, MAP, Bayesian estimation, bias-variance tradeoff, overfitting, regularization, cross-validation
- Supervised Learning
  - Nearest Neighbour, Naïve Bayes, Logistic Regression, Support Vector Machines, Kernels, Neural Networks, Decision Trees
  - Ensemble Methods: Bagging, Boosting
- Unsupervised Learning
  - Clustering: k-means, Gaussian mixture models, EM
  - Dimensionality reduction: PCA, SVD, LDA
- Applications
  - Vision, Natural Language Processing

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

- I
- L
- C
- F

Syllabus

- Background & Basics
  - Neural Networks, Backprop, Optimization (SGD)
- Module 1: Convolutional Neural Networks (CNNs)
  - Architectures, Pre-training, Fine-tuning
  - Visualizations, Fooling CNSS, Adversarial examples
  - Different tasks: detection CNNs, segmentation CNNs
- Module 2: Recurrent Neural Networks (RNNs)
  - Difficulty of learning; “Vanilla” RNNs, LSTMs, GRU
  - RNNs for Sequence-to-Sequence (machine translation & image captioning, VQA, Visual Dialog)
- Module 3: Deep Reinforcement Learning
  - Overview, policy gradients, deep Q learning
  - Optimizing Neural Sequence Models for goal-driven rewards
- Module 4: Deep Structured Prediction
  - Crash course on Bayes Nets, Variational Inference
  - Variational Auto Encoders (VAEs)
- Module 5: Advanced Topics
  - GANs, Adversarial Learning
  - Gumbel-Softmax

Syllabus

- You will learn about the methods you heard about
- But we are not teaching “how to use a toolbox”
- You will understand algorithms, theory, applications, and implementations
- **It’s going to be FUN and HARD WORK ☺**

Course Information

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

Machine Learning & Perception Group

Dhruv Batra
Assistant Professor

PhD

Qing Sun

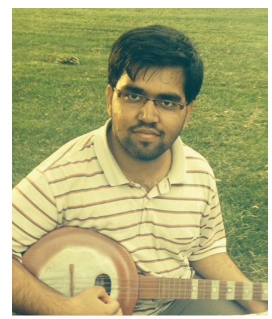
Aishwarya Agrawal

Yash Goyal

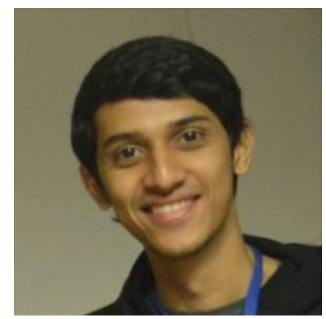
Michael Cogswell

Abhishek Das

Ashwin Kalyan



Nirbhay Modhe

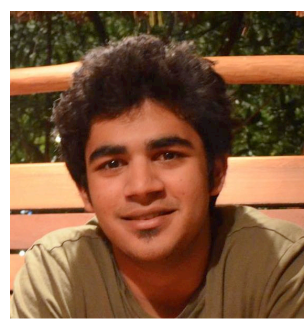


Research Scientist

Stefan Lee

Masters

Akrit Mohapatra



Deshraj Yadav

TAs

Michael Cogswell

3rd year CS PhD student

http://mcogswell.io/

Abhishek Das

2nd year CS PhD student

http://abhishekdas.com/

Zhaoyang Lv

3rd year CS PhD student

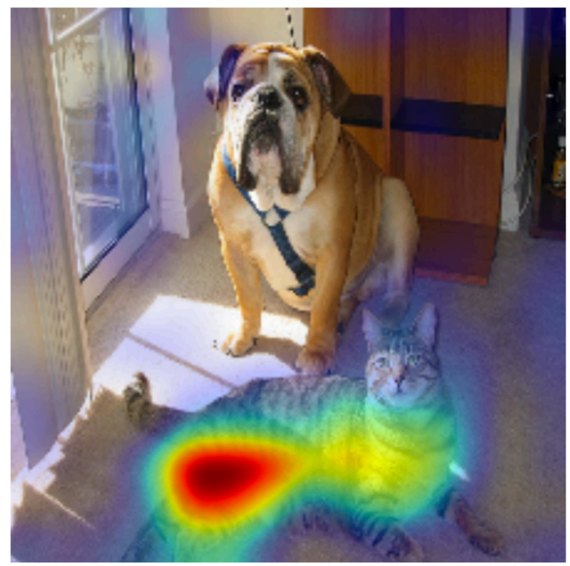
https://www.cc.gatech.edu/~zlv30

TA: Michael Cogswell

- PhD student working with Dhruv

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

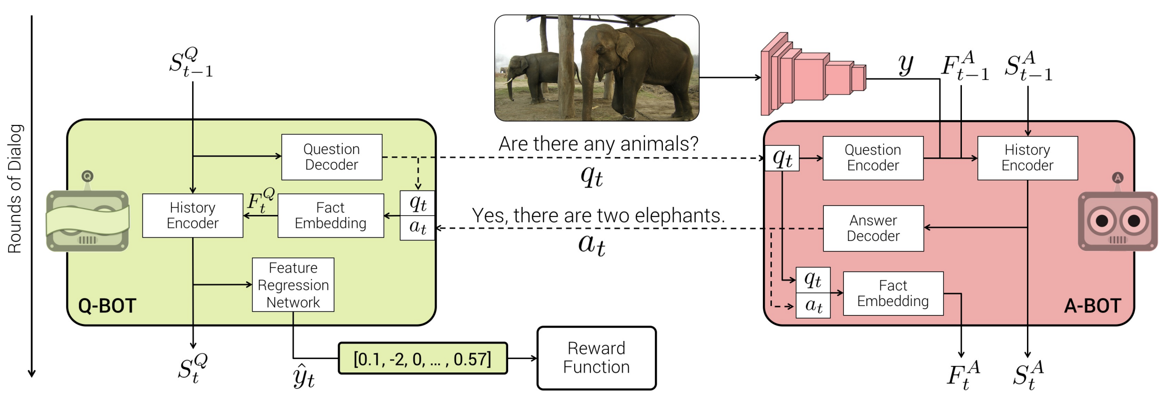
$$\mathcal{L}_{\texttt{DeCov}}$$

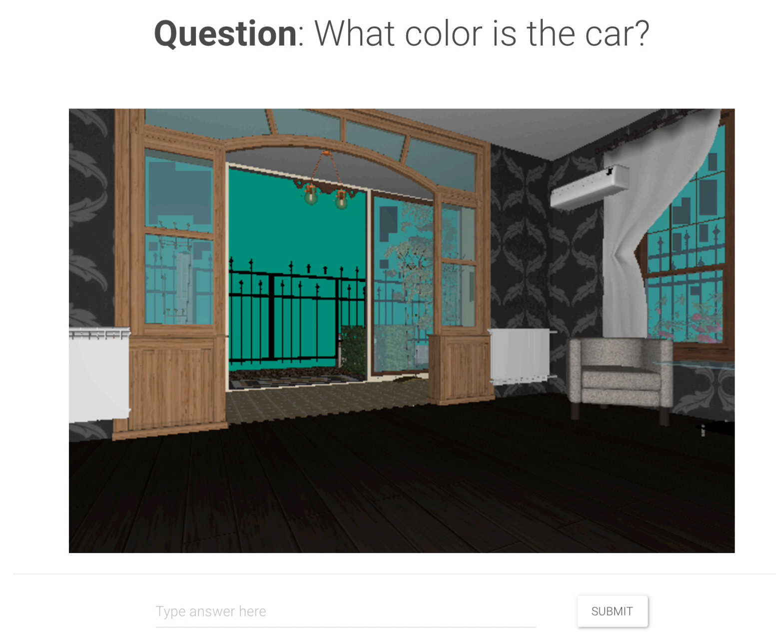


- I also Fence (mainly foil)

TA: Abhishek Das

- 2nd year CS PhD student
- Research interests:
  - Agents that can see, talk and act





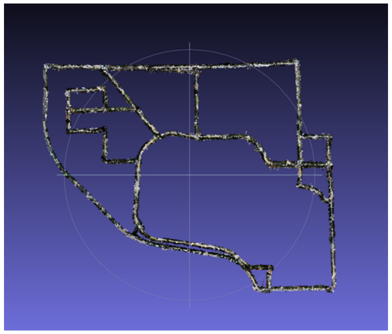
TA: Zhaoyang Lv

Ph.D. student in robotics, school of IC

Advisors: James Rehg, Frank Dellaert (co-advised)

Research Interests:

3D Vision: SLAM, reconstruction



Motion Understanding

Video

Optical flow

Scene flow

Organization & Deliverables

- 3 homeworks + 4-6 Problem Sets (50%)
  - 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
- Paper Reviews (10%)
  - Read 1 paper per class
  - Submit summary before class
- Paper Presentations (15%)
  - [Tentative] 1 presentation in the semester
  - Practice run with a TA 1 week before scheduled date
- Final project (20%)
  - Projects done in groups of say 2-3 (exceptions okay)
- Class Participation (5%)
  - Contribute to class discussions on Piazza
  - Ask questions, answer questions

Invited Talks

- Nathan Silberman on TensorFlow-Slim
  - Butterfly Networks, Previously Google Research
  - (Tentative) Sept 5, in class

Google Research Blog

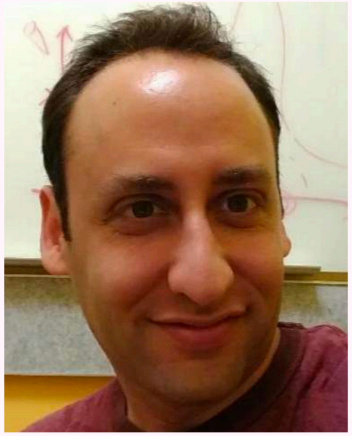
The latest news from Research at Google

TF-Slim: A high level library to define complex models in TensorFlow

Tuesday, August 30, 2016

Posted by Nathan Silberman and Sergio Guadarrama, Google Research

Earlier this year, we released a TensorFlow implementation of a state-of-the-art image classification model known as Inception-V3. This code allowed users to train the model on the ImageNet classification dataset via synchronized gradient descent, using either a single local machine or a cluster of machines. The Inception-V3 model was built on an experimental TensorFlow library called TF-Slim, a lightweight package for defining, training and evaluating models in TensorFlow. The TF-Slim library provides common abstractions which enable users to define models quickly and concisely, while keeping the model architecture transparent and its hyperparameters explicit.



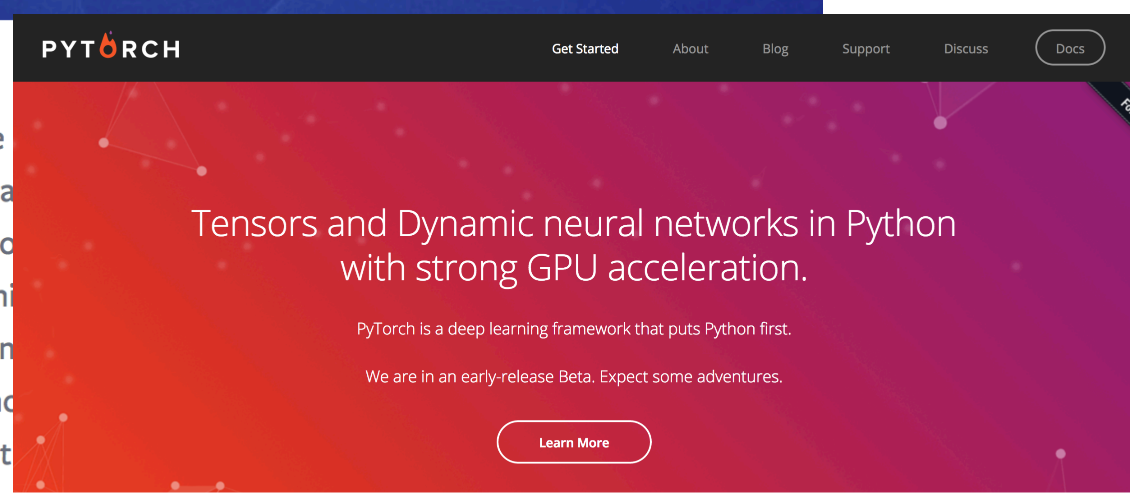
st at 4Catalyzer where I work on a variety of health care related projects. My machine
tation, detection and reinforcement learning and how to best apply these areas to

g various projects, I co-wrote TensorFlow-Slim, now a major component of the

Invited Talks

- Soumith Chintala on PyTorch
  - Facebook AI Research
  - Co-located as ML@GT Seminar, Sep 6 12-1pm

Soumith Chintala is a Researcher at Facebook AI Research, where
deep learning, reinforcement learning, generative image models, a
games and large-scale high-performance deep learning. Prior to jo
in August 2014, he worked at MuseAmi, where he built deep learni
music and vision targeted at mobile devices. He holds a Masters in
and spent time in Yann LeCun's NYU lab building deep learning mo
pedestrian detection, natural image OCR, depth-images among ot



Invited Talks

- Ross Girshick on Object Detection and VQA
  - Facebook AI Research
  - TBD

Ross Girshick (rbg)
Research Scientist
Facebook AI Research (FAIR)
r...@gmail.com
arXiv / Google scholar / cv

Problem Sets vs Homeworks

- PS: All theory questions
  - Due in 1 week

- HW: All implementation questions
  - Due in 2 weeks

- PS and HW are hard, start early!
  - Due via Canvas (Assignments tool)

- "Free" Late Days
  - 7 late days for the semester
    - Use for HW, PS
    - Cannot use for HW0, reviews, or presentations

  - After free late days are used up:
    - 25% penalty for each late day

HW0

- Out today; due Thursday (08/24)
  - Available on class webpage + Canvas
- Grading
  - Does not count towards grade.
  - BUT Pass/Fail.
  - <=90% 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

Paper Reviews

- Length
  - 200-400 words.
- Due: Midnight before class on Piazza

- Organization
  - Summary:
    - What is this paper about? What is the main contribution? Describe the main approach & results. Just facts, no opinions yet.
  - List of positive points / Strengths:
    - Is there a new theoretical insight? Or a significant empirical advance? Did they solve a standing open problem? Or is a good formulation for a new problem? Or a faster/better solution for an existing problem? Any good practical outcome (code, algorithm, etc)? Are the experiments well executed? Useful for the community in general?
  - List of negative points / Weaknesses:
    - What would you do differently? Any missing baselines? missing datasets? any odd design choices in the algorithm not explained well? quality of writing? Is there sufficient novelty in what they propose? Has it already been done? Minor variation of previous work? Why should anyone care? Is the problem interesting and significant?
  - Reflections
    - How does this relate to other papers we have read? What are the next research directions in this line of work?

Presentations

- Frequency
  - [Tentative] Once in the semester
- Expectations
  - Present details of 1 paper in detail
    - Describe formulation, experiment, approaches, datasets
    - Encouraged to present a broad picture
    - Show results; demo code if possible
  - Please clearly cite the source of each slide that is not your own.
  - Meet with TA 1 week before class to dry run presentation
    - Worth 40% of presentation grade

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

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.

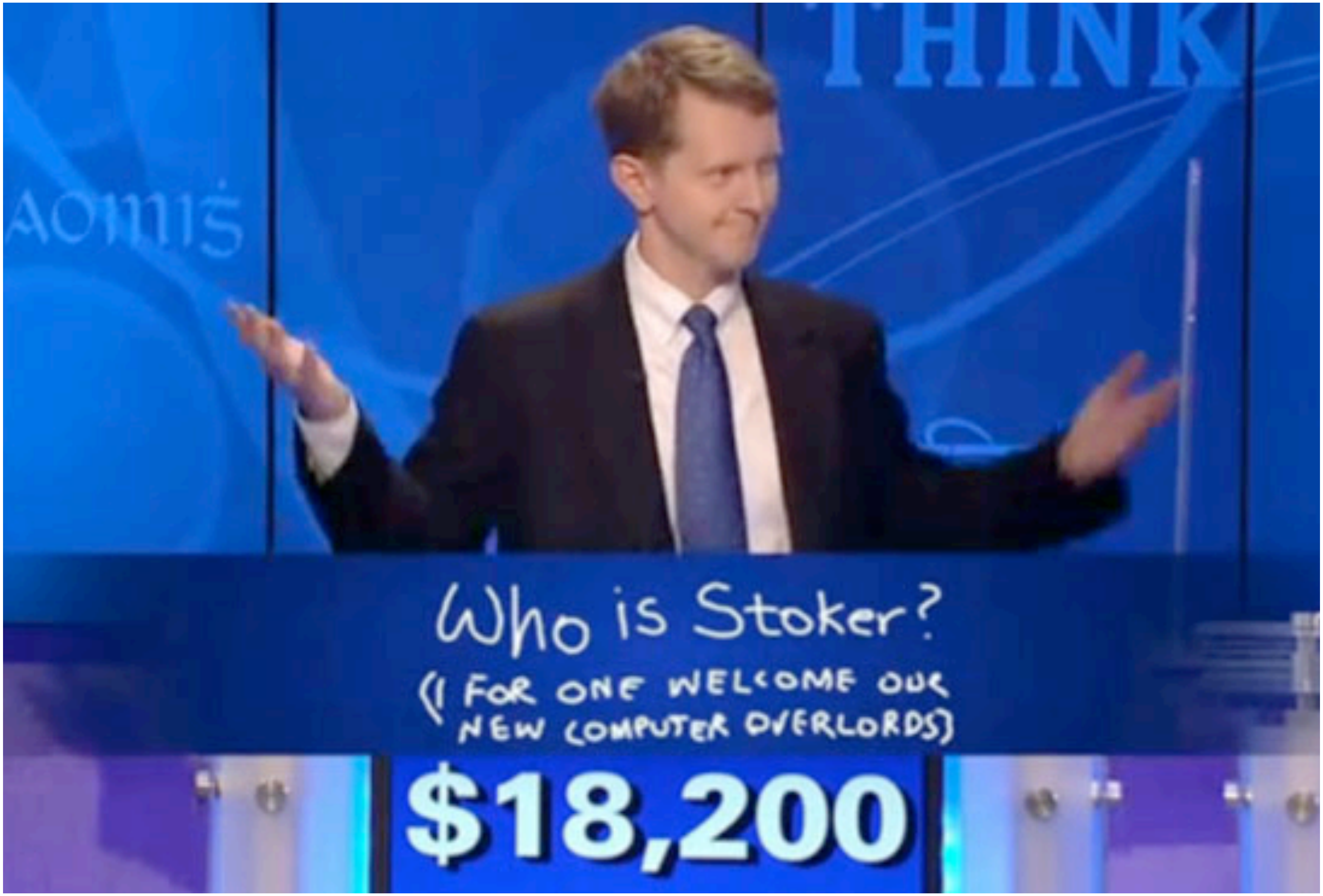
- Staff Mailing List
  - cs-7643-f17-staff@googlegroups.com

- Class websites:
  - https://www.cc.gatech.edu/classes/AY2018/cs7643_fall/
  - gatech.instructure.com/courses/772
  - piazza.com/gatech/fall2017/cs7643

Todo

- HW0
  - Due Thursday Aug 24 11:55pm

Welcome


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