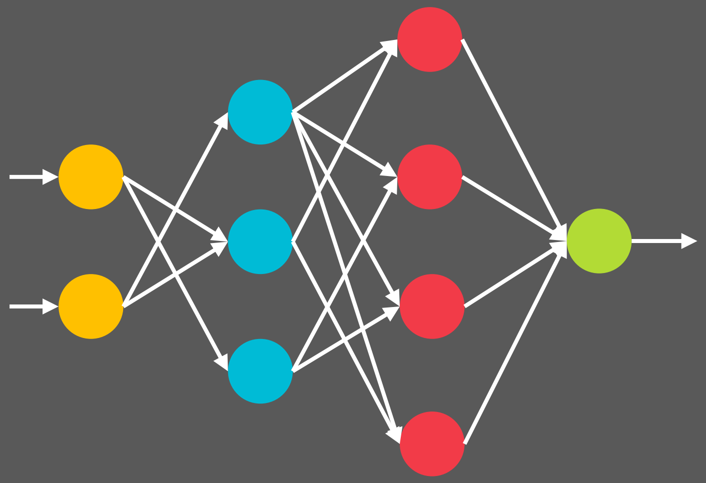
May I Have Your Attention Please ?

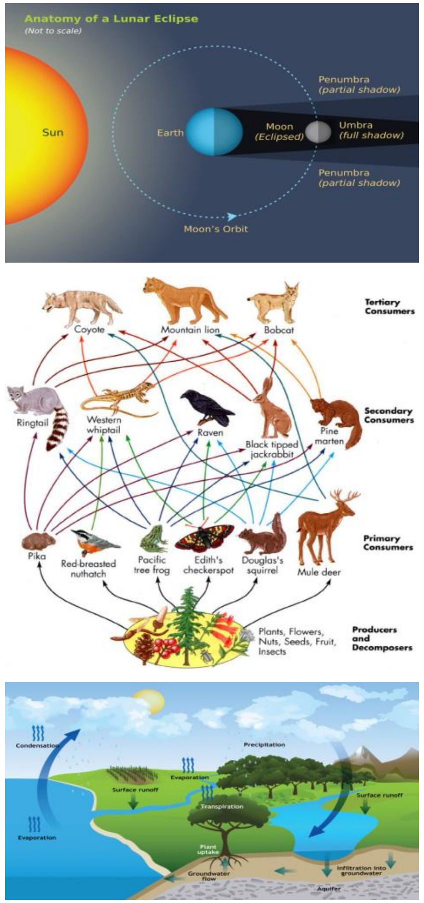
(said one neuron to another)



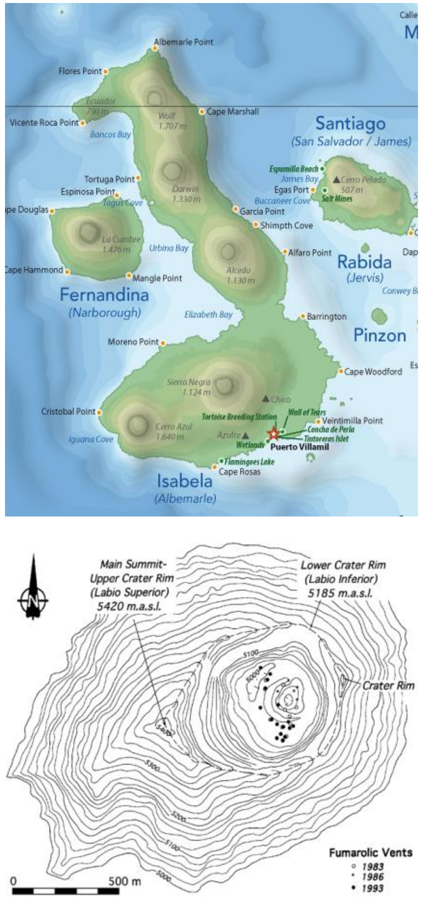
Ani Kembhavi

Allen Institute for Artificial Intelligence

The world of visual illustrations



Science Diagrams



Maps

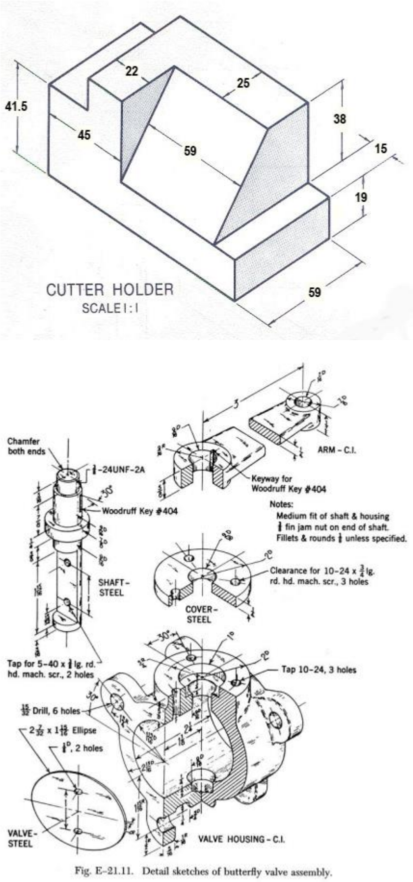
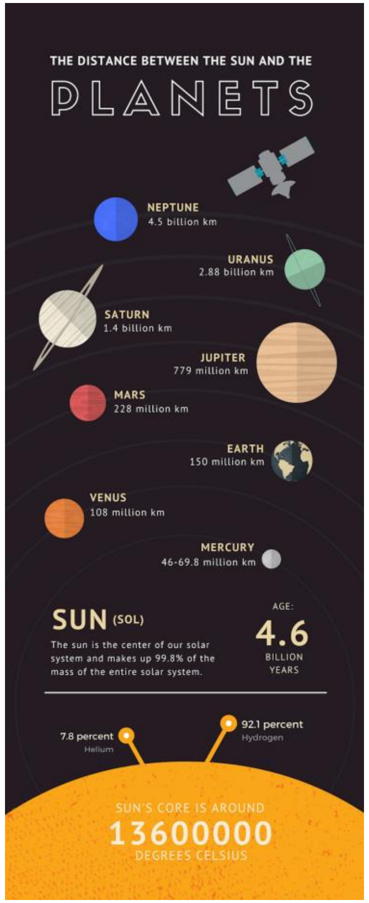

41.5
22
25
45
59
38
15
19
59
CUTTER HOLDER
SCALE 1:1
Chamfer both ends
24UNF-2A
Woodruff Key #404
SHAFT-STEEL
ARM - C.I.
Keyway for Woodruff Key #404
Notes:
Medium fit of shaft & housing
fin jam nut on end of shaft.
Fillets & rounds unless specified.
Clearance for 10-24 x lg. rd. hd. mach. scr., 3 holes
COVER-STEEL
Tap for 5-40 x lg. rd. hd. mach. scr., 2 holes
Drill, 6 holes
Ellipse
2 holes
Tap 10-24, 3 holes
VALVE-STEEL
VALVE HOUSING - C.I.

Fig. E-21.11. Detail sketches of butterfly valve assembly.

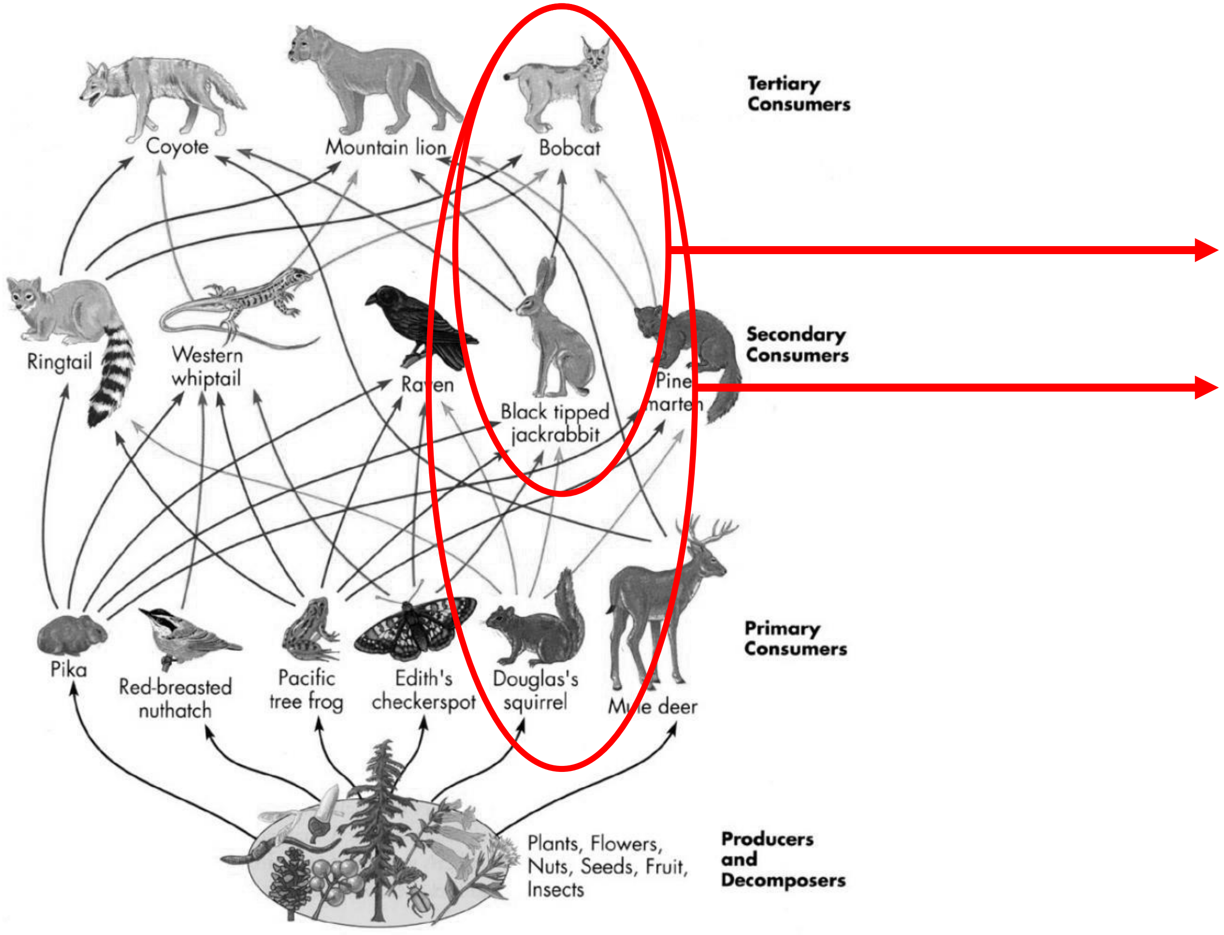
3d visualizations



Infographics

… and many more

Diagrams afford deep opportunities for reasoning

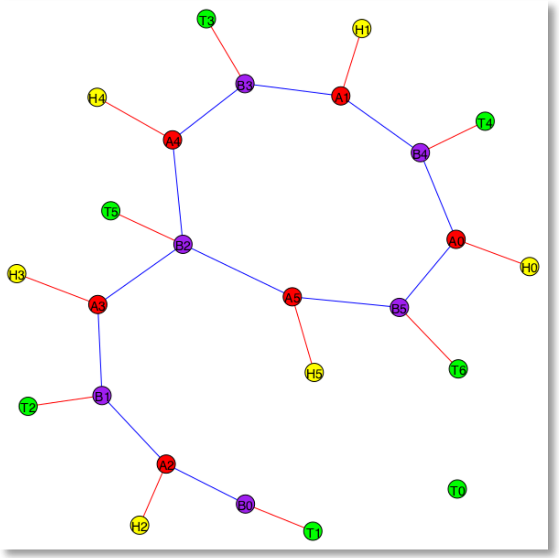


Which animal does the Bobcat eat ?

What is the effect on the population of Bobcats if the population of squirrel decreased ?

Syntactic Parsing

Semantic Interpretation

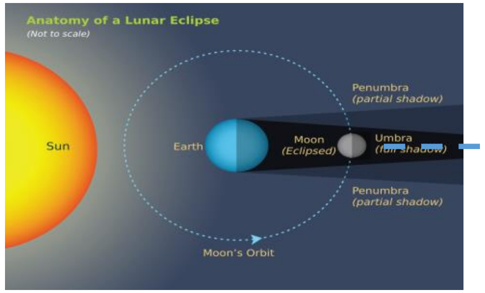


Detect Constituents

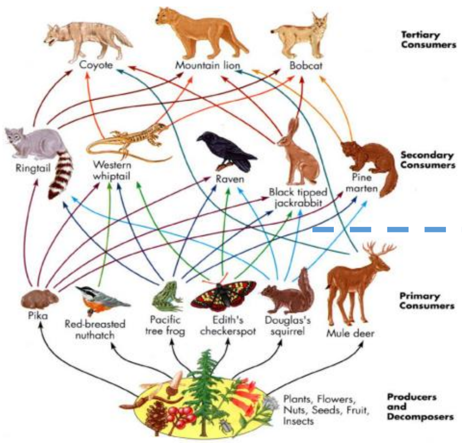
Objects, Text, Elements

Detect Relationships

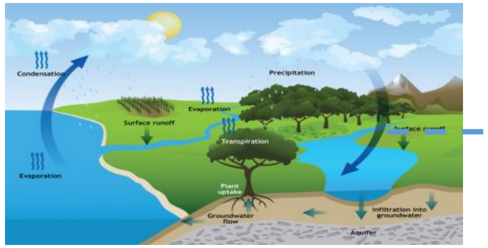
Label, Connections



Motion



Consumption



Phase change

Syntactic Parsing

Deep Sequential Diagram Parser

Structured Set Matching Networks

Diagram Question Answering

Bidirectional Attention Flow

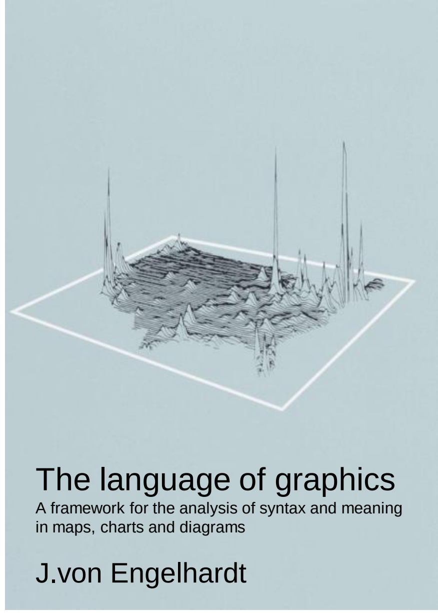
Textbook Question Answering

Semantic Interpretation

The language of diagrams

Prior work in the graphics community to represent visual illustrations

We build upon Engelhardt's representation of graphic

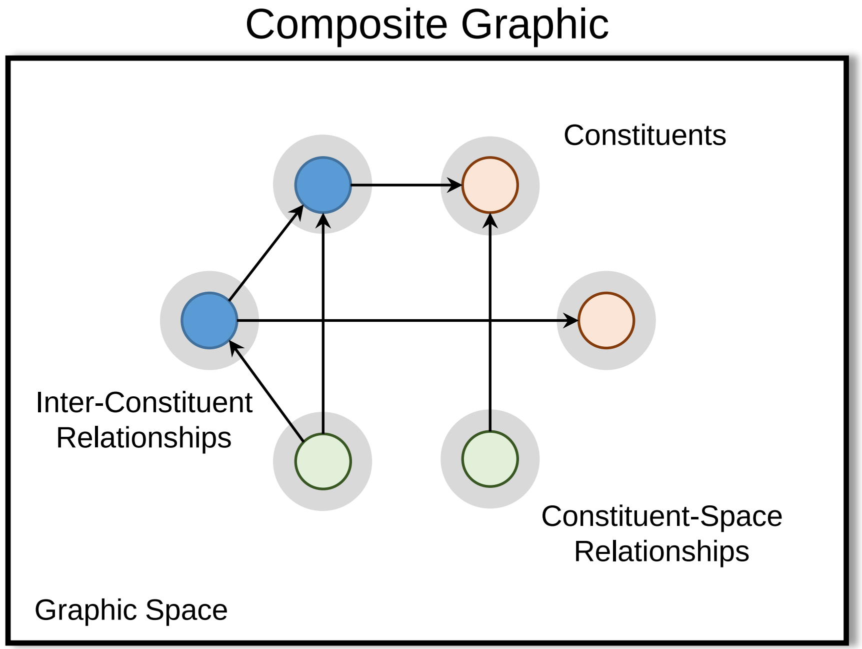


The language of graphics

A framework for the analysis of syntax and meaning in maps, charts and diagrams

J.von Engelhardt

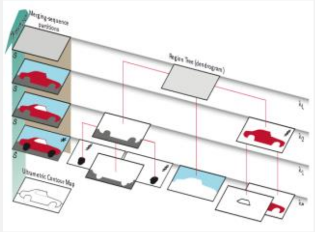
Syntactic decomposition of a diagram


Composite Graphic
Constituents
Inter-Constituent Relationships
Constituent-Space Relationships
Graphic Space

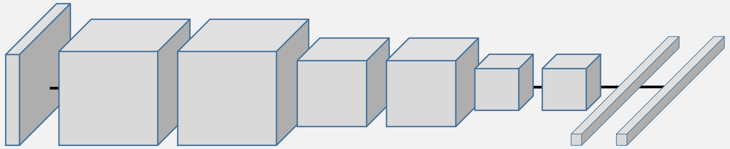
Generating candidates

Constituents

Segment Proposals

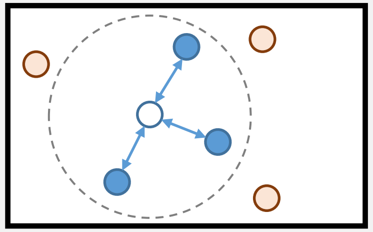


Convolutional Neural Networks

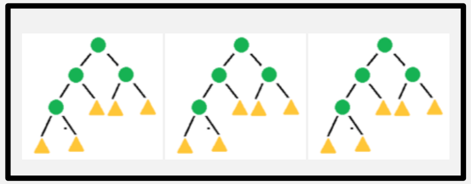


Inter-Constituent Relationships

Relationship Proposals

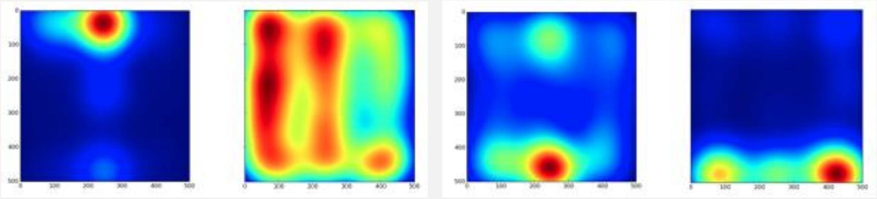


Random Forest Classifiers

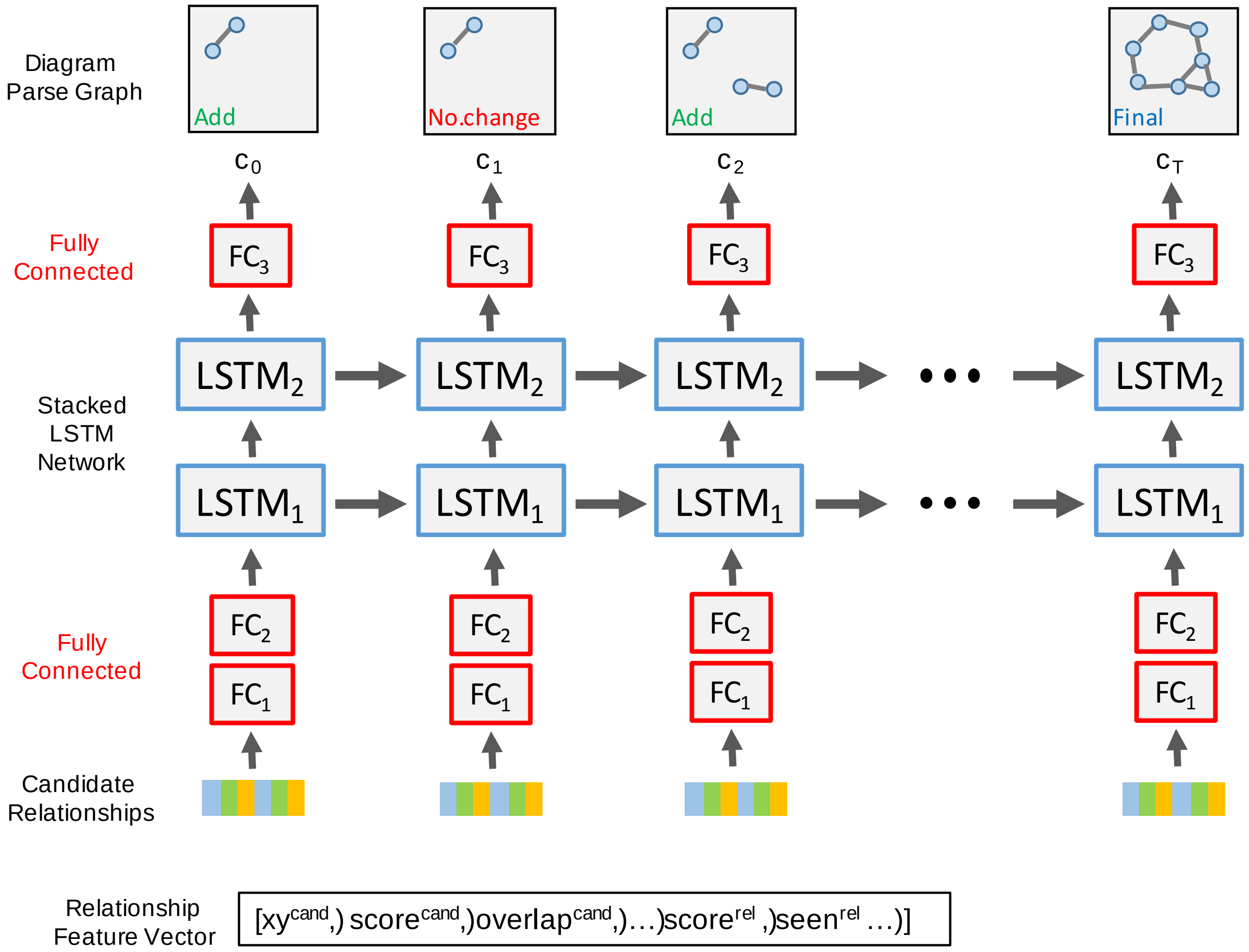


Constituent-Space Relationships

Kernel Density Estimates



Deep Sequential Diagram Parser

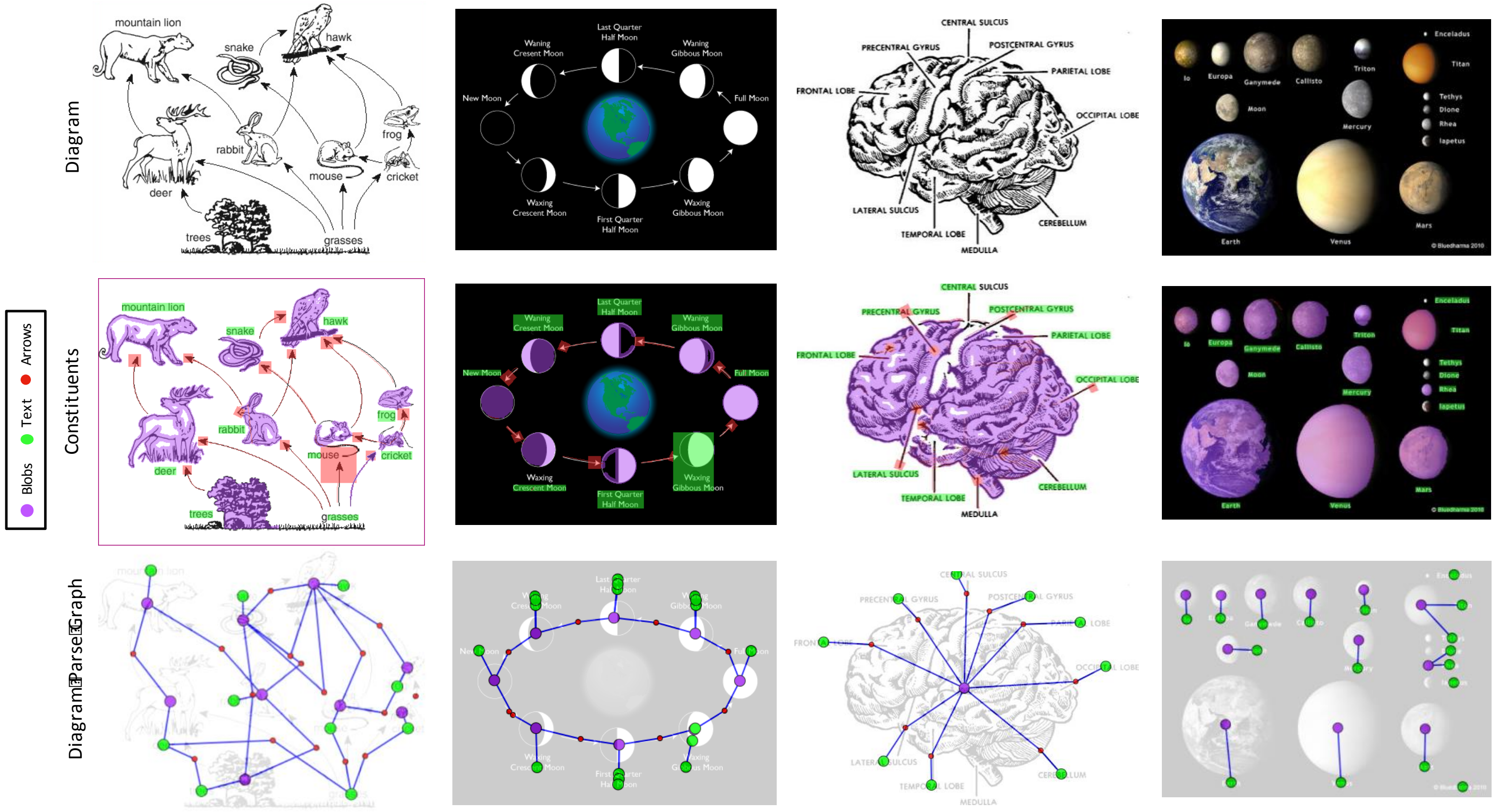


LSTMs require a lot of training data!

For each training image:
Sample 100s of relationship sequences
Sample without replacement
Relationship score as sampling weight

Test time:
Relationships sorted by proposal scores

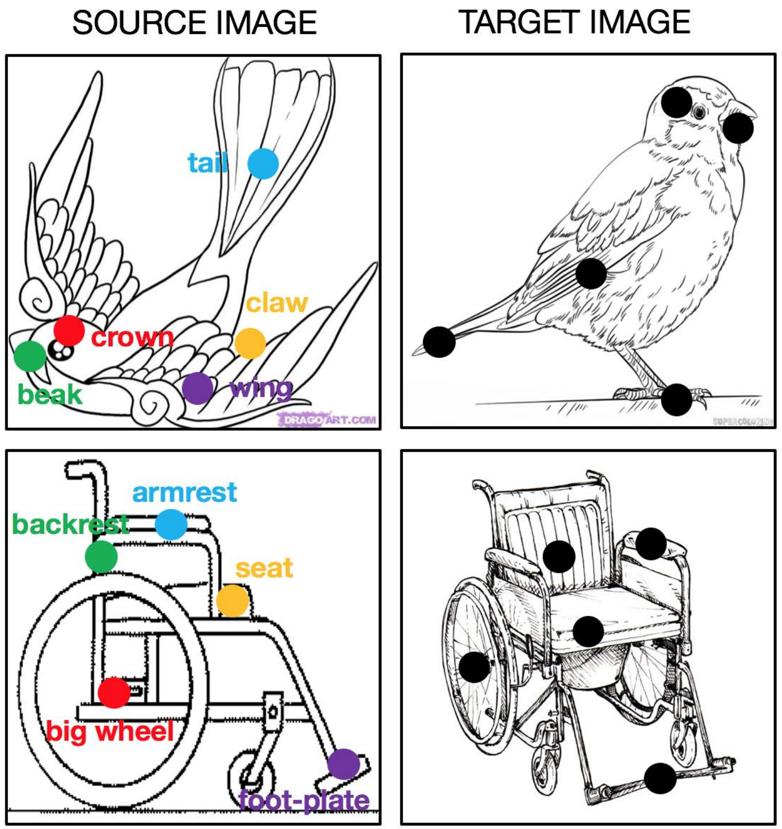
Parser Results



Method	JIG Score
Greedy Search	28.96
A* Search	41.02
DSDP-Net	**51.45**

Understanding diagrams can be partially addressed by matching

Scarce training data motivates a one-shot scenario



Must generalize to unseen categories

Cannot simply learn a classifier for each part

Absence of color and texture

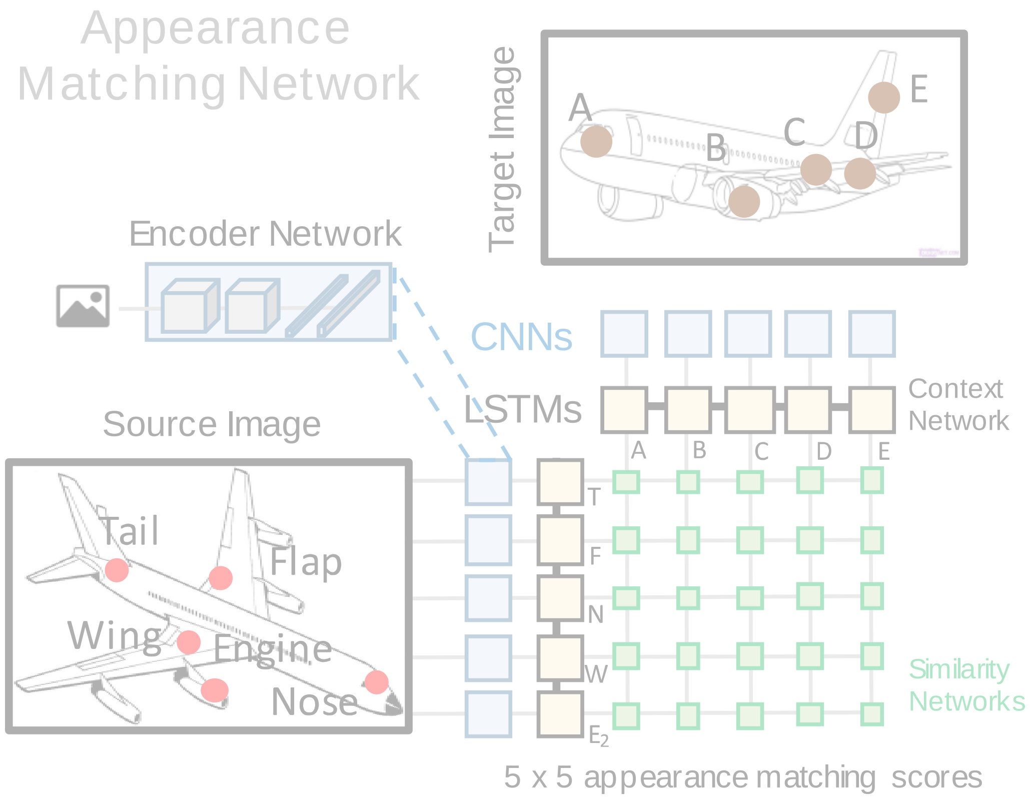
Local cues ambiguous

Pose variations between images

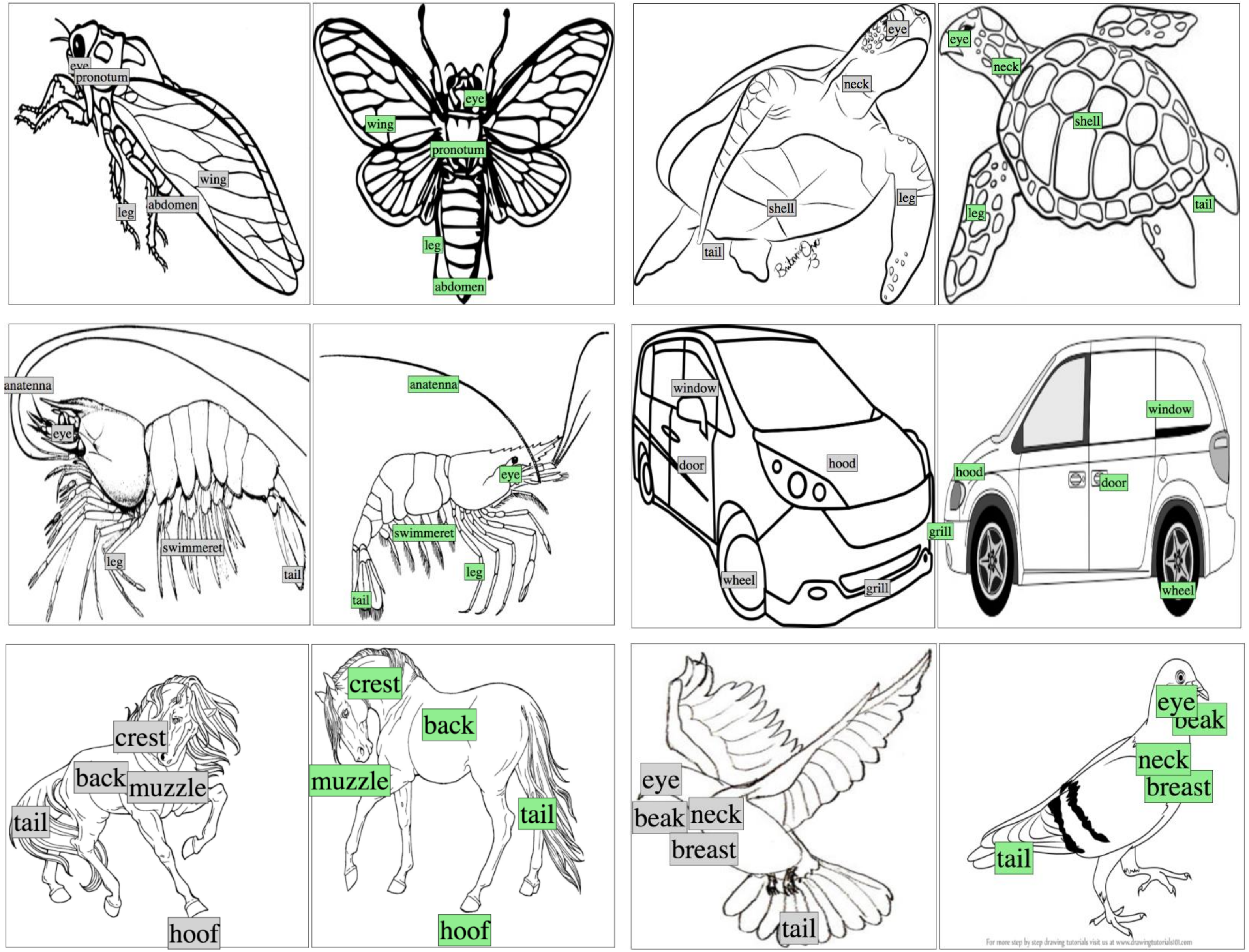
Absolute position ambiguous

Must enforce a 1:1 matching between parts

Structured Set Matching Network

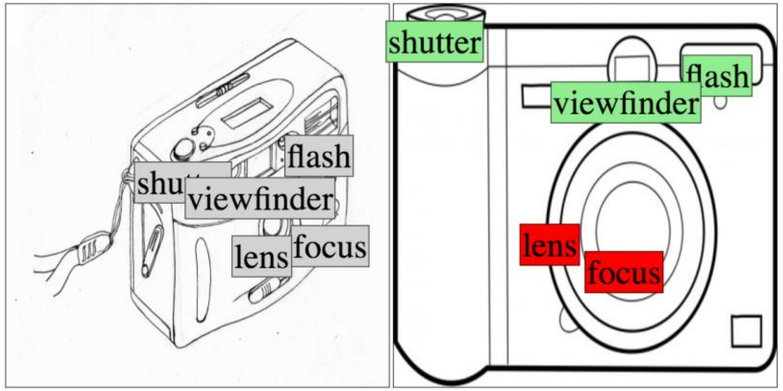


Results



Methods	Validation	Test
Random	20.0%	20.0%
Nearest Neighbor	41.4%	46.7%
MN-C	47.1%	51.0%
Affine Transform	54.0%	52.4%
Matching Network (MN) [44]	60.9%	67.6%
MN + Hungarian	69.2%	75.8%
SSMN (Ours)	**73.8%**	**79.3%**

	Original Image	DT-image
Validation Accuracy	43.1%	47.1%

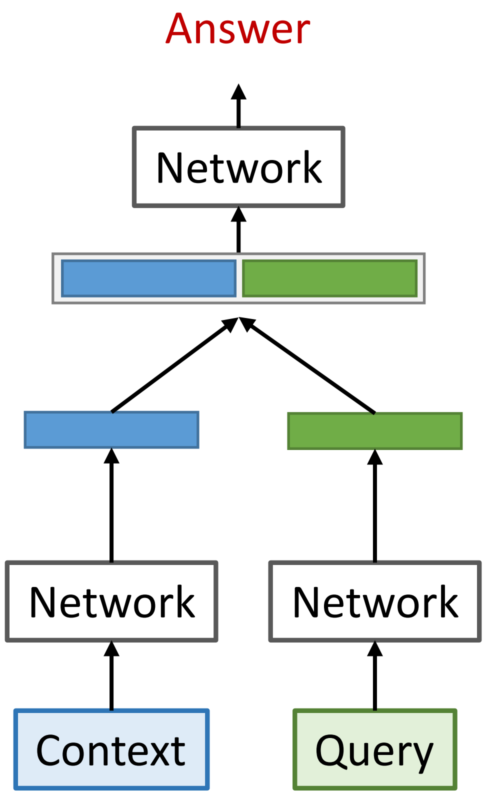


Semantic Interpretation

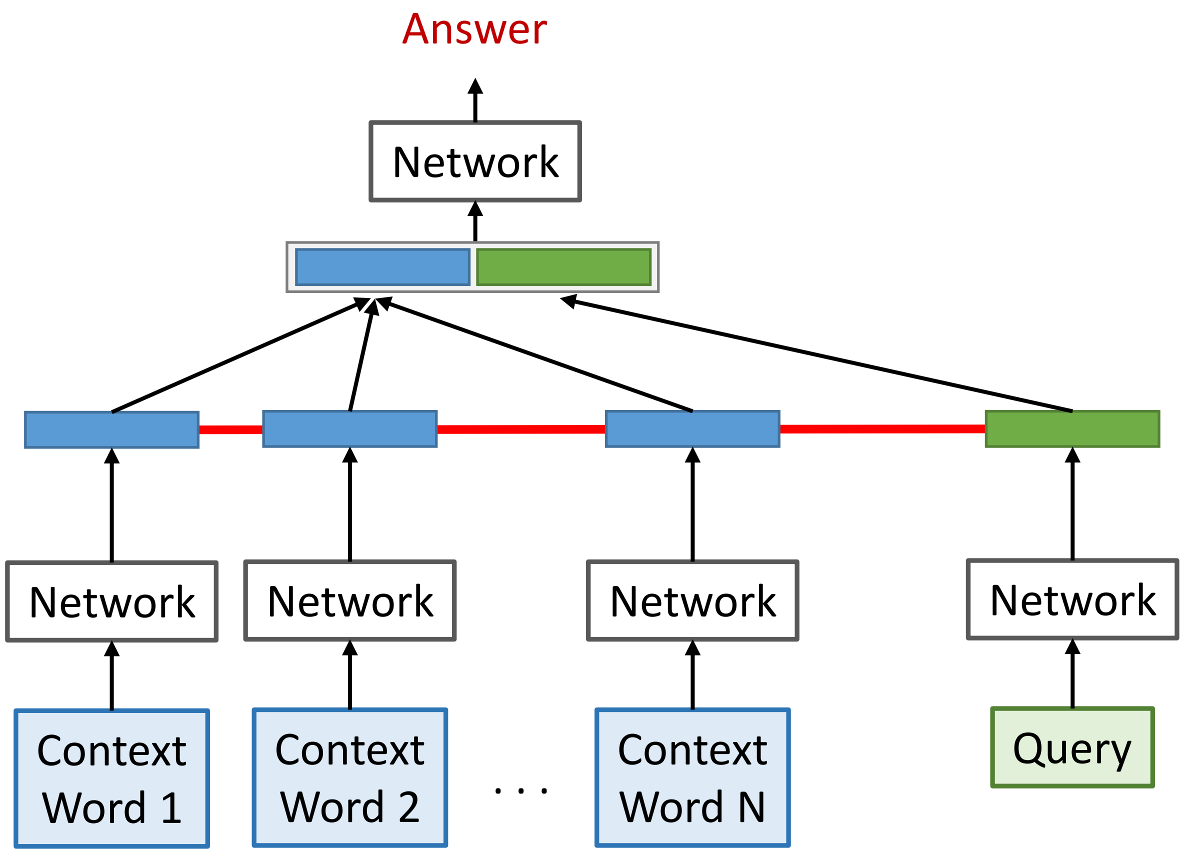
in the context of question answering

Neural Models for Machine Comprehension

Vanilla Architecture



Attention Architecture

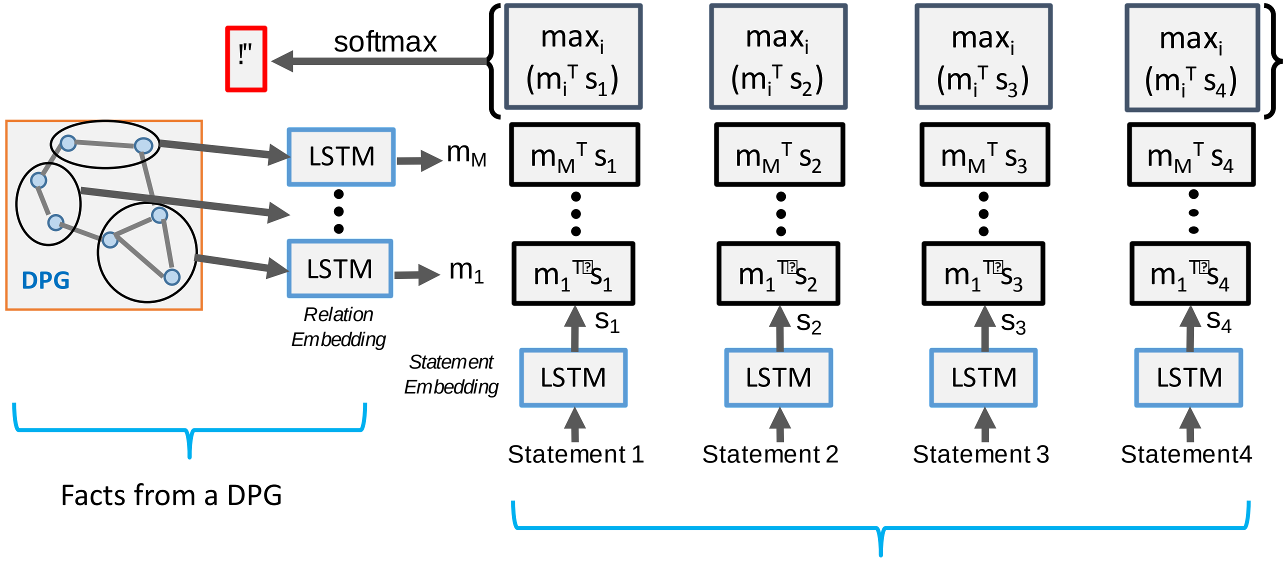


Attend over Diagram Parse Graph

Embed the question answer pair in a d-dim space

Embed each fact into the same space

Attention module learns to attend to the relevant fact, given a question

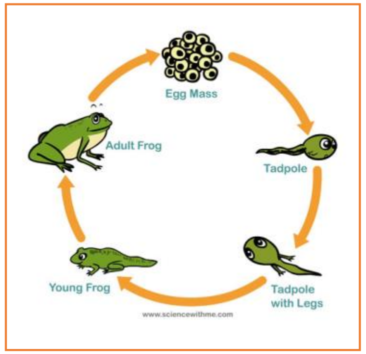


Facts from a DPG

Each question-answer pair into a statement

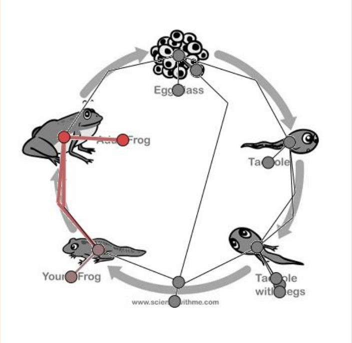
Results

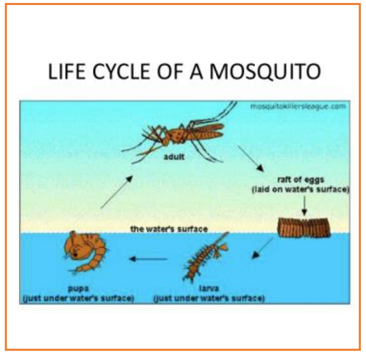
Method	Train Set	Accuracy
Q + I (VQA)	VQA	29.06
Q	AI2D	33.02
Q + I (VQA)	AI2D	32.90
Q + OCR	AI2D	34.21
Q + I + OCR	AI2D	34.02
DQA-Net	AI2D	38.47



The diagram depicts The life cycle of

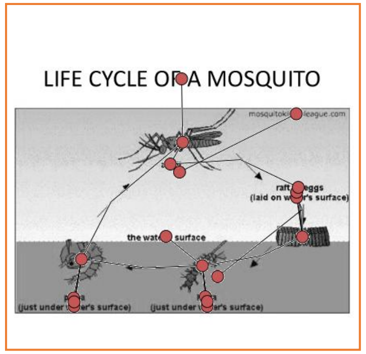
a) frog	**0.924**
b) bird	0.02
c) insecticide	0.054
d) insect	0.002

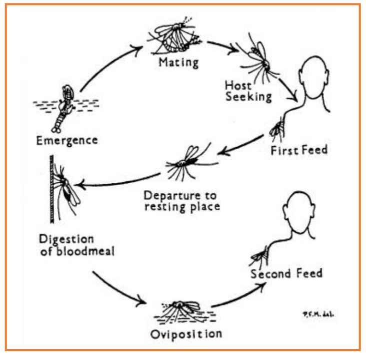



LIFE CYCLE OF A MOSQUITO
mosquitokillersleague.com
adult
raft of eggs (laid on water's surface)
the water's surface
pupa (just under water's surface)
larva (just under water's surface)

How many stages of Growth does the diagram Feature?

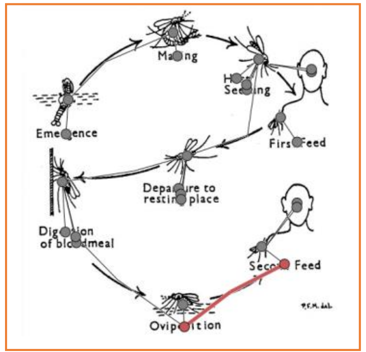
a) 4	**0.924**
b) 2	0.02
c) 3	0.054
d) 1	0.002





What comes before Second feed?

a) digestion	0.0
b) First feed	0.15
c) indigestion	0.0
d) oviposition	0.85



Neural Attention

Some characteristics of past attention models:

Attention weights used to summarize the modality into a single vector

Attended vectors allowed to *flow* through to the modelling layer

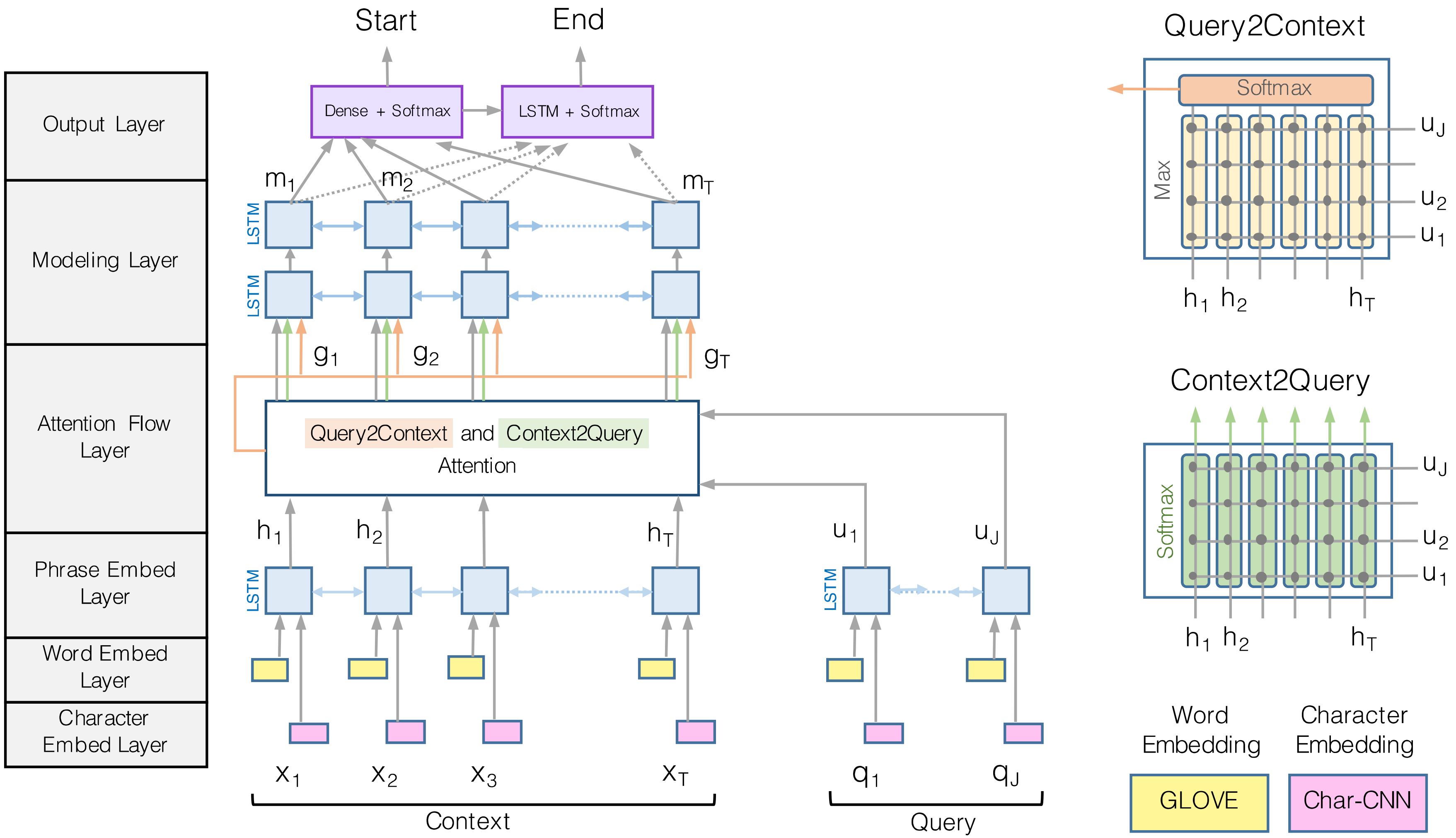
They are often temporally dynamic (attention at *t* affects attention at *t+1*)

Our attention mechanism is memory-less

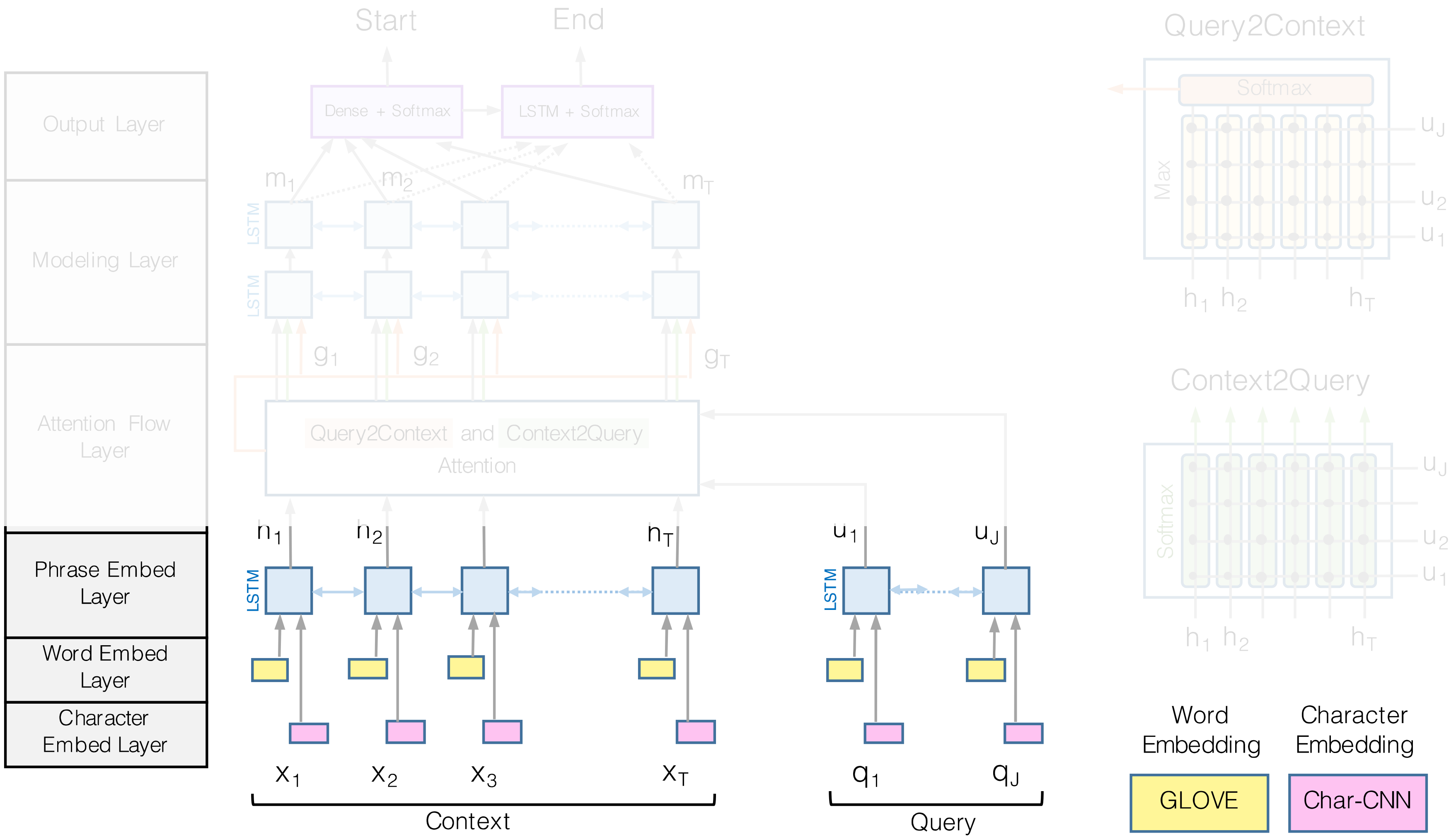
They are usually uni-directional

We use bi-directional attention: Query-to-context & Context-to-query

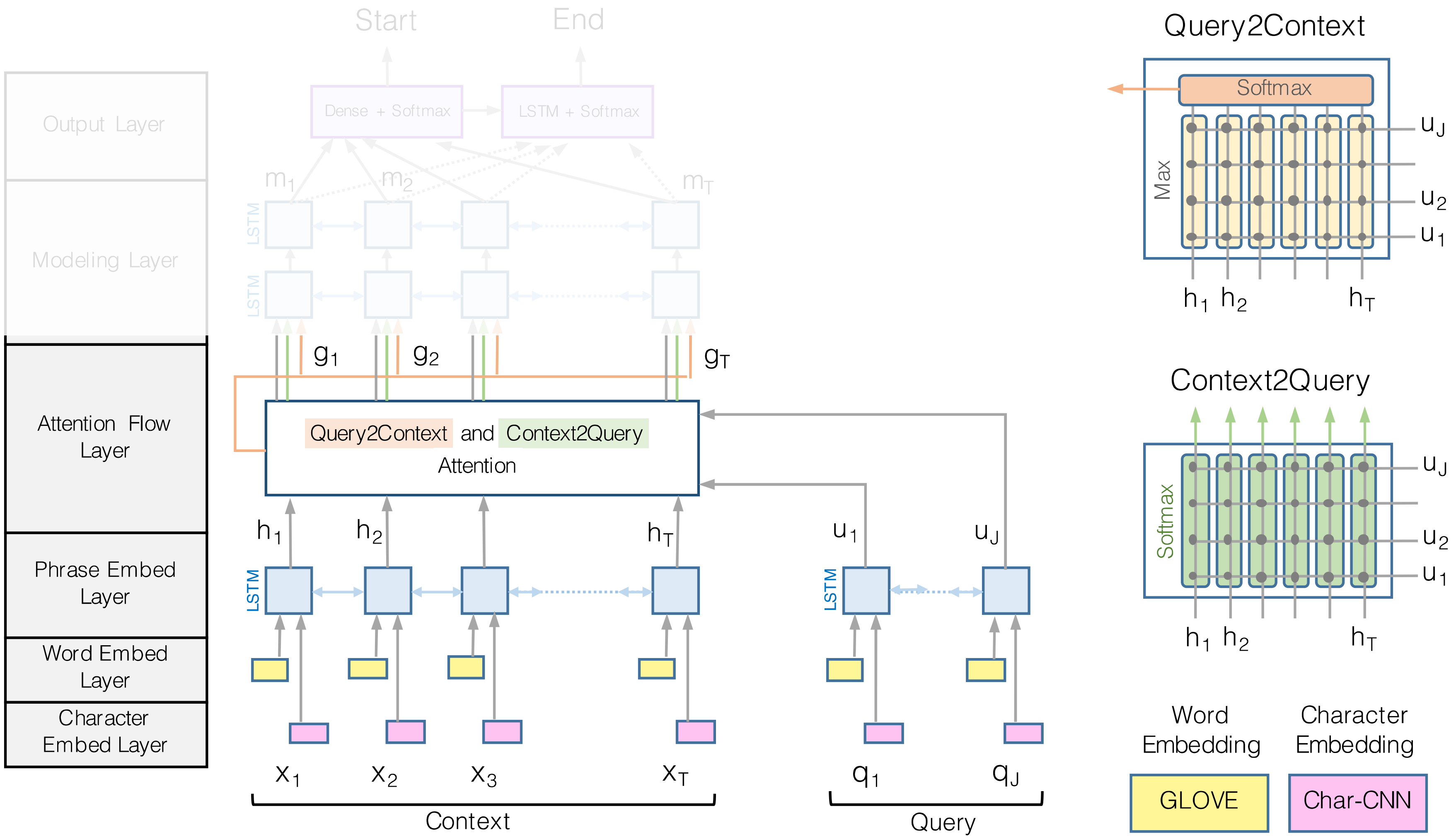
Bidirectional Attention Flow (BiDAF) Model



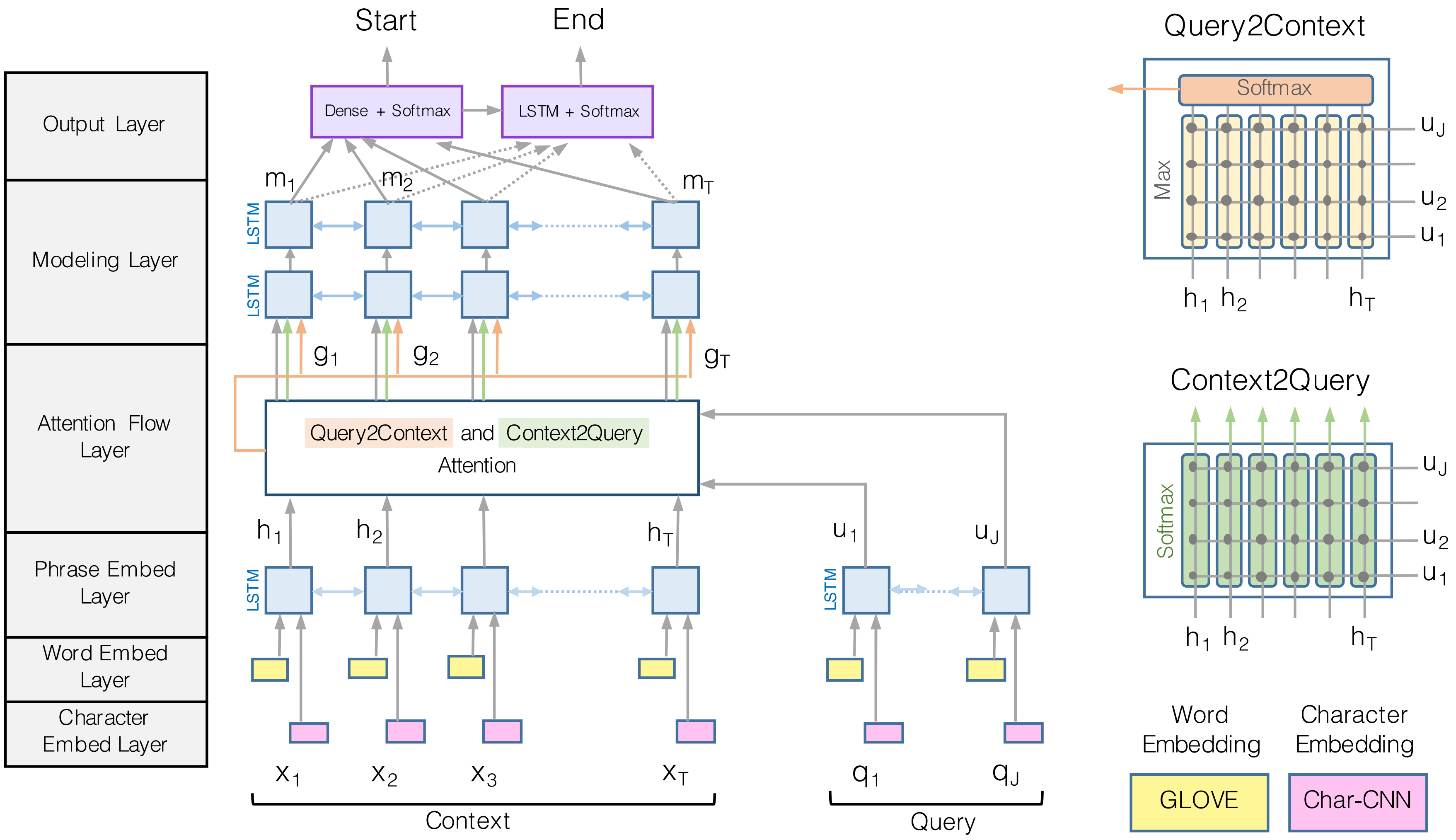
Bidirectional Attention Flow (BiDAF) Model



Bidirectional Attention Flow (BiDAF) Model

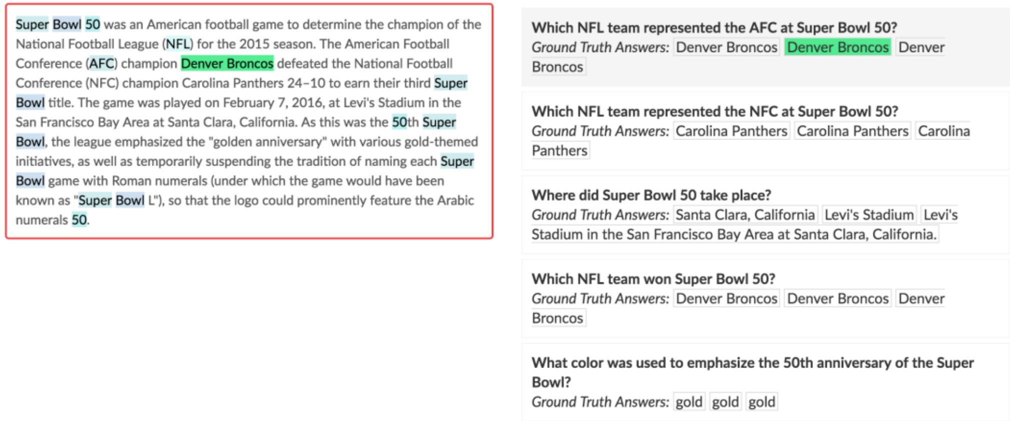


Bidirectional Attention Flow (BiDAF) Model



Machine Comprehension Task

SQuAD
The Stanford Question Answering Dataset



Super Bowl 50 was an American football game to determine the champion of the National Football League (NFL) for the 2015 season. The American Football Conference (AFC) champion Denver Broncos defeated the National Football Conference (NFC) champion Carolina Panthers 24–10 to earn their third Super Bowl title. The game was played on February 7, 2016, at Levi's Stadium in the San Francisco Bay Area at Santa Clara, California. As this was the 50th Super Bowl, the league emphasized the "golden anniversary" with various gold-themed initiatives, as well as temporarily suspending the tradition of naming each Super Bowl game with Roman numerals (under which the game would have been known as "Super Bowl L"), so that the logo could prominently feature the Arabic numerals 50.

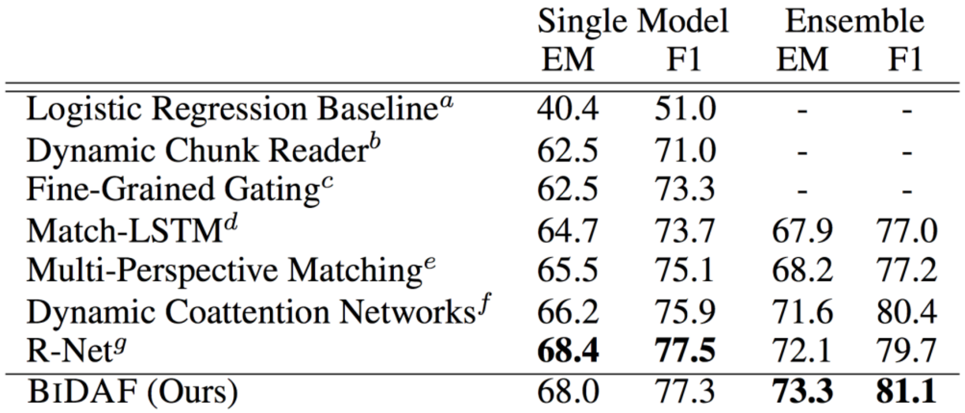
Which NFL team represented the AFC at Super Bowl 50?
Ground Truth Answers: Denver Broncos Denver Broncos Denver Broncos

Which NFL team represented the NFC at Super Bowl 50?
Ground Truth Answers: Carolina Panthers Carolina Panthers Carolina Panthers

Where did Super Bowl 50 take place?
Ground Truth Answers: Santa Clara, California Levi's Stadium Levi's Stadium in the San Francisco Bay Area at Santa Clara, California.

Which NFL team won Super Bowl 50?
Ground Truth Answers: Denver Broncos Denver Broncos Denver Broncos

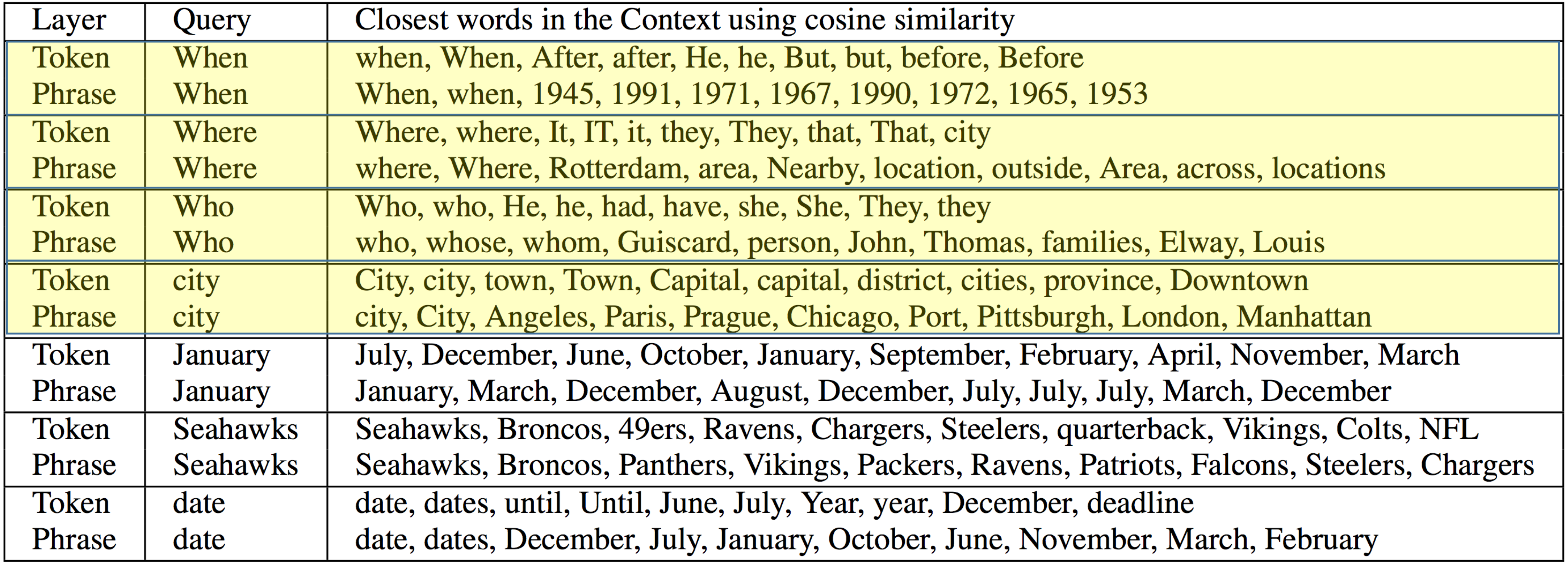
What color was used to emphasize the 50th anniversary of the Super Bowl?
Ground Truth Answers: gold gold gold



	Single Model		Ensemble	
	EM	F1	EM	F1
Logistic Regression Baseline[a]	40.4	51.0	-	-
Dynamic Chunk Reader[b]	62.5	71.0	-	-
Fine-Grained Gating[c]	62.5	73.3	-	-
Match-LSTM[d]	64.7	73.7	67.9	77.0
Multi-Perspective Matching[e]	65.5	75.1	68.2	77.2
Dynamic Coattention Networks[f]	66.2	75.9	71.6	80.4
R-Net[g]	**68.4**	**77.5**	72.1	79.7
BIDAF (Ours)	68.0	77.3	**73.3**	**81.1**

Over 100,000 question-answer tuples

Visualizations: Word vs Phrase Spaces



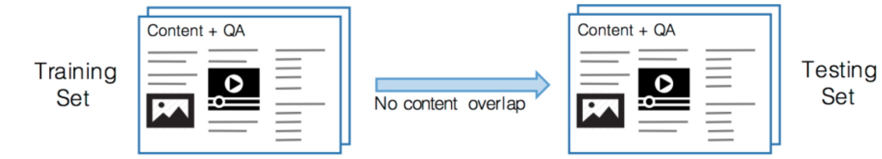
Layer	Query	Closest words in the Context using cosine similarity
Token	When	when, When, After, after, He, he, But, but, before, Before
Phrase	When	When, when, 1945, 1991, 1971, 1967, 1990, 1972, 1965, 1953
Token	Where	Where, where, It, IT, it, they, They, that, That, city
Phrase	Where	where, Where, Rotterdam, area, Nearby, location, outside, Area, across, locations
Token	Who	Who, who, He, he, had, have, she, She, They, they
Phrase	Who	who, whose, whom, Guiscard, person, John, Thomas, families, Elway, Louis
Token	city	City, city, town, Town, Capital, capital, district, cities, province, Downtown
Phrase	city	city, City, Angeles, Paris, Prague, Chicago, Port, Pittsburgh, London, Manhattan
Token	January	July, December, June, October, January, September, February, April, November, March
Phrase	January	January, March, December, August, December, July, July, July, March, December
Token	Seahawks	Seahawks, Broncos, 49ers, Ravens, Chargers, Steelers, quarterback, Vikings, Colts, NFL
Phrase	Seahawks	Seahawks, Broncos, Panthers, Vikings, Packers, Ravens, Patriots, Falcons, Steelers, Chargers
Token	date	date, dates, until, Until, June, July, Year, year, December, deadline
Phrase	date	date, dates, December, July, January, October, June, November, March, February

BiDAF Demo

https://allenai.github.io/bi-att-flow/

Textbook QA Challenge

Multi-modal Machine Comprehension (M³C)



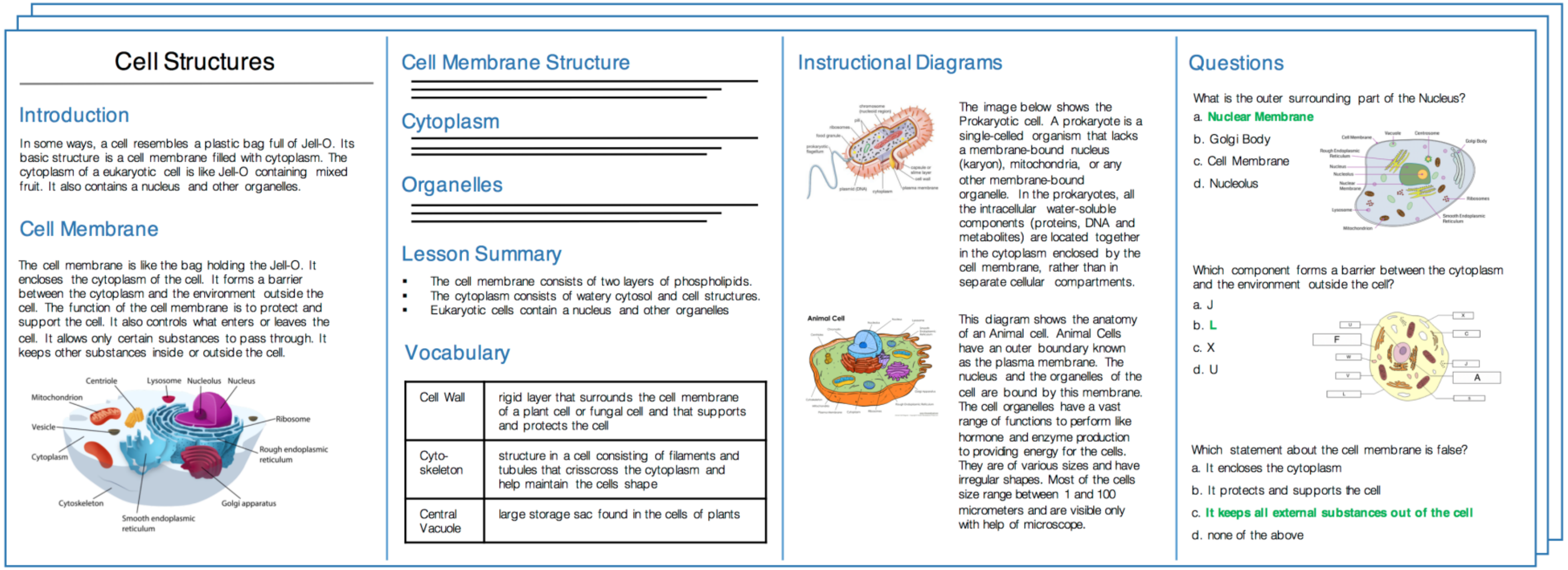
Textbook Question Answering (TQA)

1076 lessons from middle school curricula

Life Science	Earth Science	Physical Science

78,338 sentences
3,455 images
26,260 questions

Lessons in TQA



Cell Structures

Introduction

In some ways, a cell resembles a plastic bag full of Jell-O. Its basic structure is a cell membrane filled with cytoplasm. The cytoplasm of a eukaryotic cell is like Jell-O containing mixed fruit. It also contains a nucleus and other organelles.

Cell Membrane

The cell membrane is like the bag holding the Jell-O. It encloses the cytoplasm of the cell. It forms a barrier between the cytoplasm and the environment outside the cell. The function of the cell membrane is to protect and support the cell. It also controls what enters or leaves the cell. It allows only certain substances to pass through. It keeps other substances inside or outside the cell.

Cell Membrane Structure

Cytoplasm

Organelles

Lesson Summary

- The cell membrane consists of two layers of phospholipids.
- The cytoplasm consists of watery cytosol and cell structures.
- Eukaryotic cells contain a nucleus and other organelles

Vocabulary

Cell Wall	rigid layer that surrounds the cell membrane of a plant cell or fungal cell and that supports and protects the cell
Cyto-skeleton	structure in a cell consisting of filaments and tubules that crisscross the cytoplasm and help maintain the cells shape
Central Vacuole	large storage sac found in the cells of plants

Instructional Diagrams

The image below shows the Prokaryotic cell. A prokaryote is a single-celled organism that lacks a membrane-bound nucleus (karyon), mitochondria, or any other membrane-bound organelle. In the prokaryotes, all the intracellular water-soluble components (proteins, DNA and metabolites) are located together in the cytoplasm enclosed by the cell membrane, rather than in separate cellular compartments.

This diagram shows the anatomy of an Animal cell. Animal Cells have an outer boundary known as the plasma membrane. The nucleus and the organelles of the cell are bound by this membrane. The cell organelles have a vast range of functions to perform like hormone and enzyme production to providing energy for the cells. They are of various sizes and have irregular shapes. Most of the cells size range between 1 and 100 micrometers and are visible only with help of microscope.

Questions

What is the outer surrounding part of the Nucleus?

a. **Nuclear Membrane**

b. Golgi Body

c. Cell Membrane

d. Nucleolus

Which component forms a barrier between the cytoplasm and the environment outside the cell?

a. J

b. **L**

c. X

d. U

Which statement about the cell membrane is false?

a. It encloses the cytoplasm

b. It protects and supports the cell

c. **It keeps all external substances out of the cell**

d. none of the above

Complex parsing and reasoning

(a) Rich Diagram Parsing

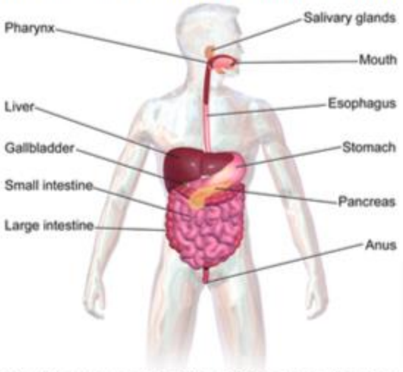
Q: This is the long narrow tube that carries food from the pharynx to the stomach.

a. mouth

b. salivary glands

c. liver

d. esophagus



The Components of the Digestive System

(b) Multiple Sentences

Q: when are most of nadh and fadh2 generated

a) during glycolysis

b) during the krebs cycle

c) during the electron transport chain

d) during cellular respiration

The Krebs Cycle

In the presence of oxygen, under aerobic conditions, pyruvate enters the mitochondria to proceed into the Krebs cycle. The second stage of cellular respiration is the transfer of the energy in pyruvate, which is the energy initially in glucose, into two energy carriers, NADH and FADH2 . A small amount of ATP is also made during this process. This process occurs in a continuous cycle, named after its discover, Hans Krebs. The Krebs cycle uses a 2-carbon molecule (acetyl-CoA) derived from pyruvate and produces carbon dioxide.

(c) Text and Diagram

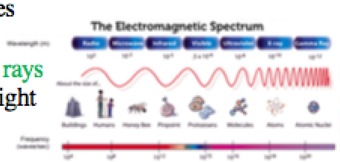
Q: Which of the following choices lists electromagnetic waves from lower to higher frequencies?

a. Radio waves, infrared light, microwaves

b. Ultraviolet light, infrared light, X rays

c. Infrared light, ultraviolet light, gamma rays

d. Visible light, microwaves, ultraviolet light



Light

Radio waves have the longest wavelengths and lowest frequencies of all electromagnetic waves. … On the right side of the diagram are X rays and gamma rays. They have the shortest wavelengths and highest frequencies of all electromagnetic waves.

(d) Order of Events

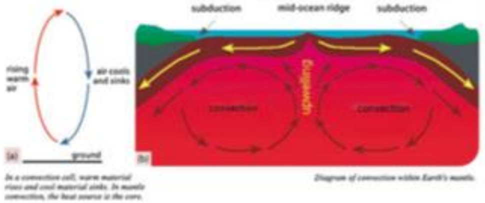
Q: put in order of how convection currents in the mantle move. i. the material that moved up cools and sinks back down into the mantle. ii. the bottom layer of the mantle material rises and spreads horizontally. iii. the mantle material near the core is heated. iv. the bottom layer of the mantle becomes less dense.

a) iv, iii, ii, i

b) iii, iv, ii, i

c) i, ii, iii, iv

d) iii, i, iv, ii



Heat Flow

Scientists know … 2. Convection: … Convection in the mantle is the same as convection in a pot of water on a stove. …

(e) 'N of Above' Answer

Q: What organ(s) do amphibians use to obtain oxygen?

a. gills

b. lungs

c. skin

d. all of the above

Amphibian Skin

… America to poison the tips of their hunting arrows. Amphibian skin contains keratin, a protein that is also found in the outer covering of most other four-legged vertebrates. The keratin in amphibians is not too tough to allow gases and water to pass through their skin. Most amphibians breathe with gills as larvae and with lungs as adults. However, extra oxygen is absorbed through the skin.

(f) Hypothetical Question

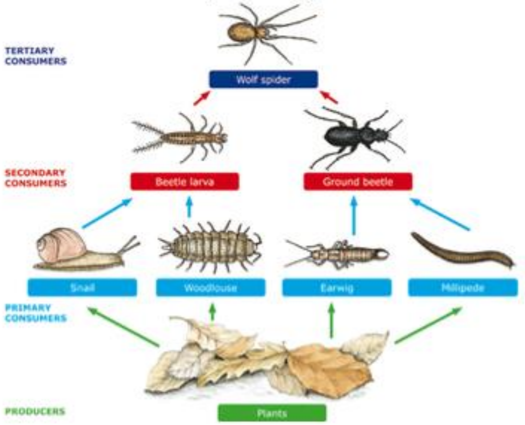
Q: If the population of beetle larva decreases, what happens with the snail population?

a. Decreases

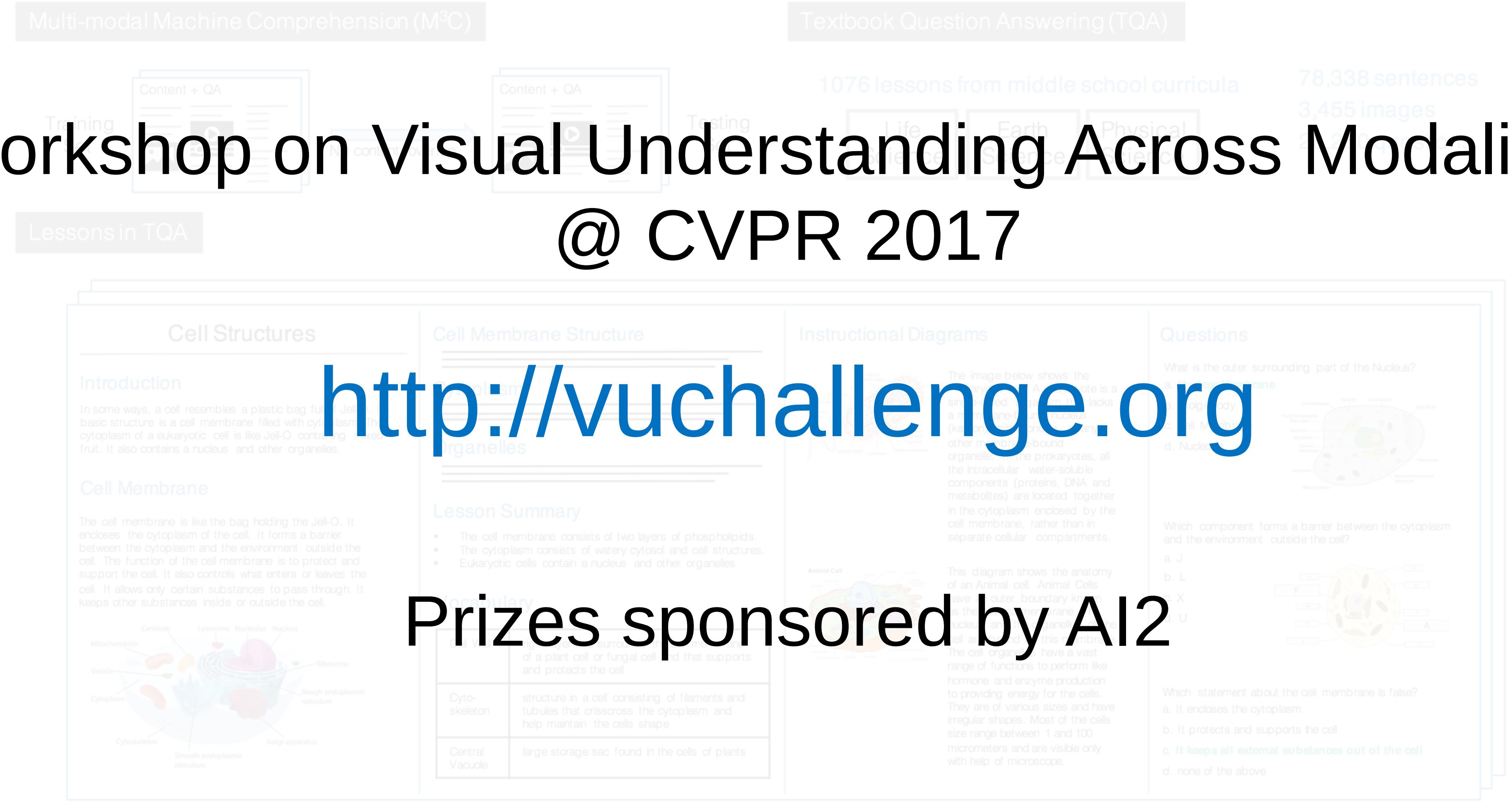
b. Increases

c. Decreases slightly

d. Stays the same



Textbook QA Challenge a part of



Workshop on Visual Understanding Across Modalities
@ CVPR 2017

http://vuchallenge.org

Prizes sponsored by AI2

Newtonian Image Understanding

Unfolding the dynamics of objects in static images

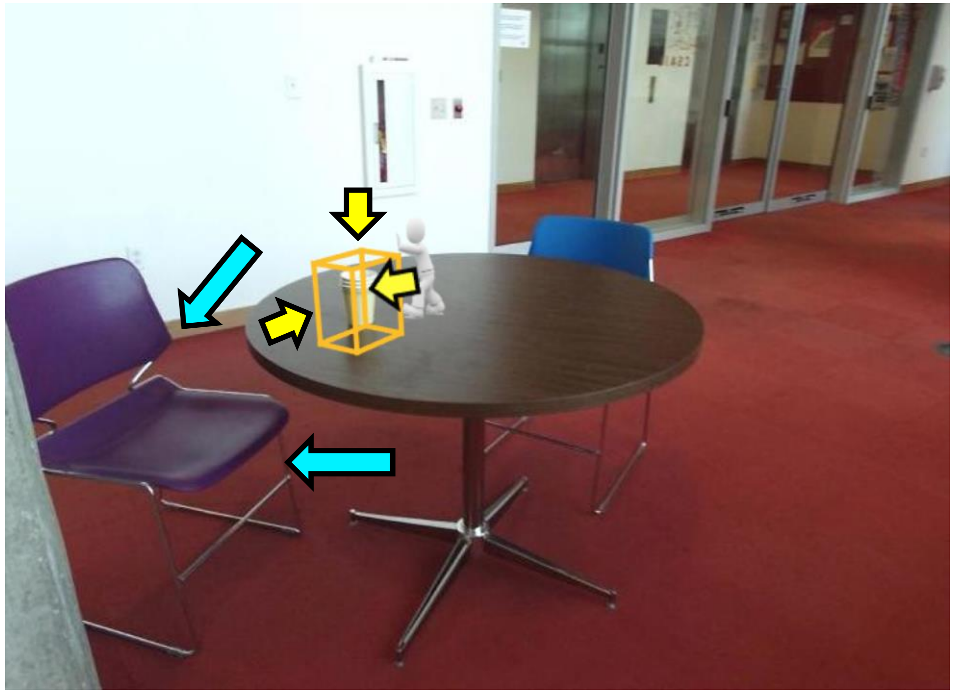
What happens if ...?

Predicting the effect of forces in images

Unfolding Object Dynamics

Predicting Effects of Forces

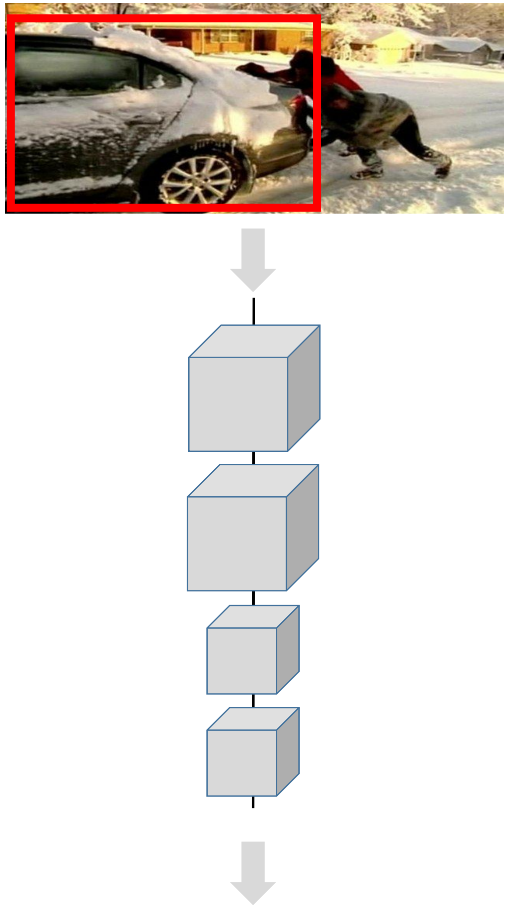
What happens if I push this cup ?



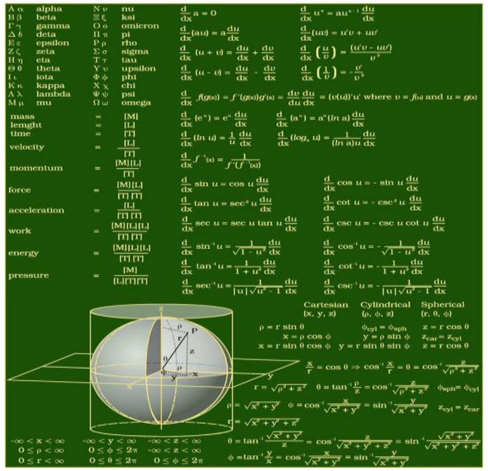
Spectrum of approaches

Let neural networks figure it out!

Estimate friction, mass, etc.
Then solve some equations.



Predicted trajectory

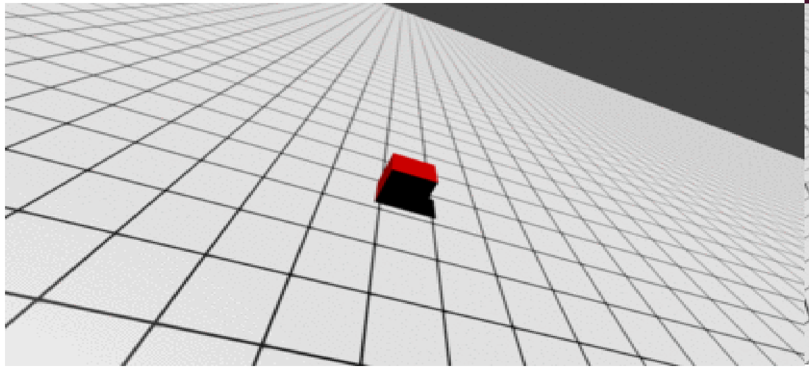


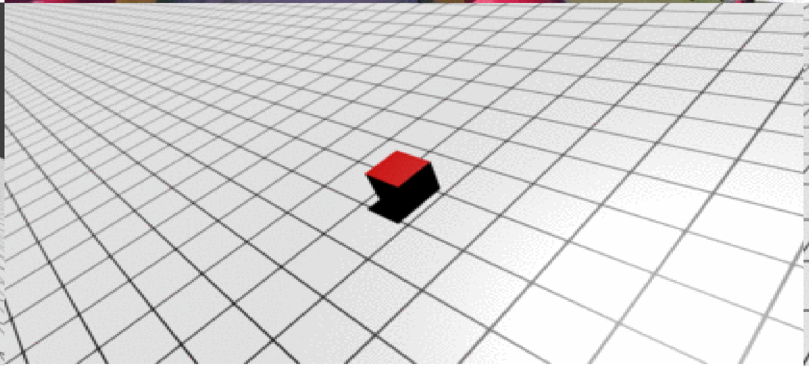
Spectrum of approaches

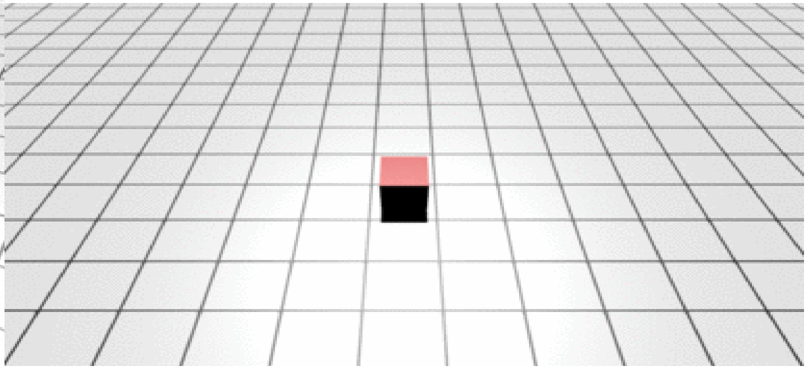
Let neural networks figure it out!

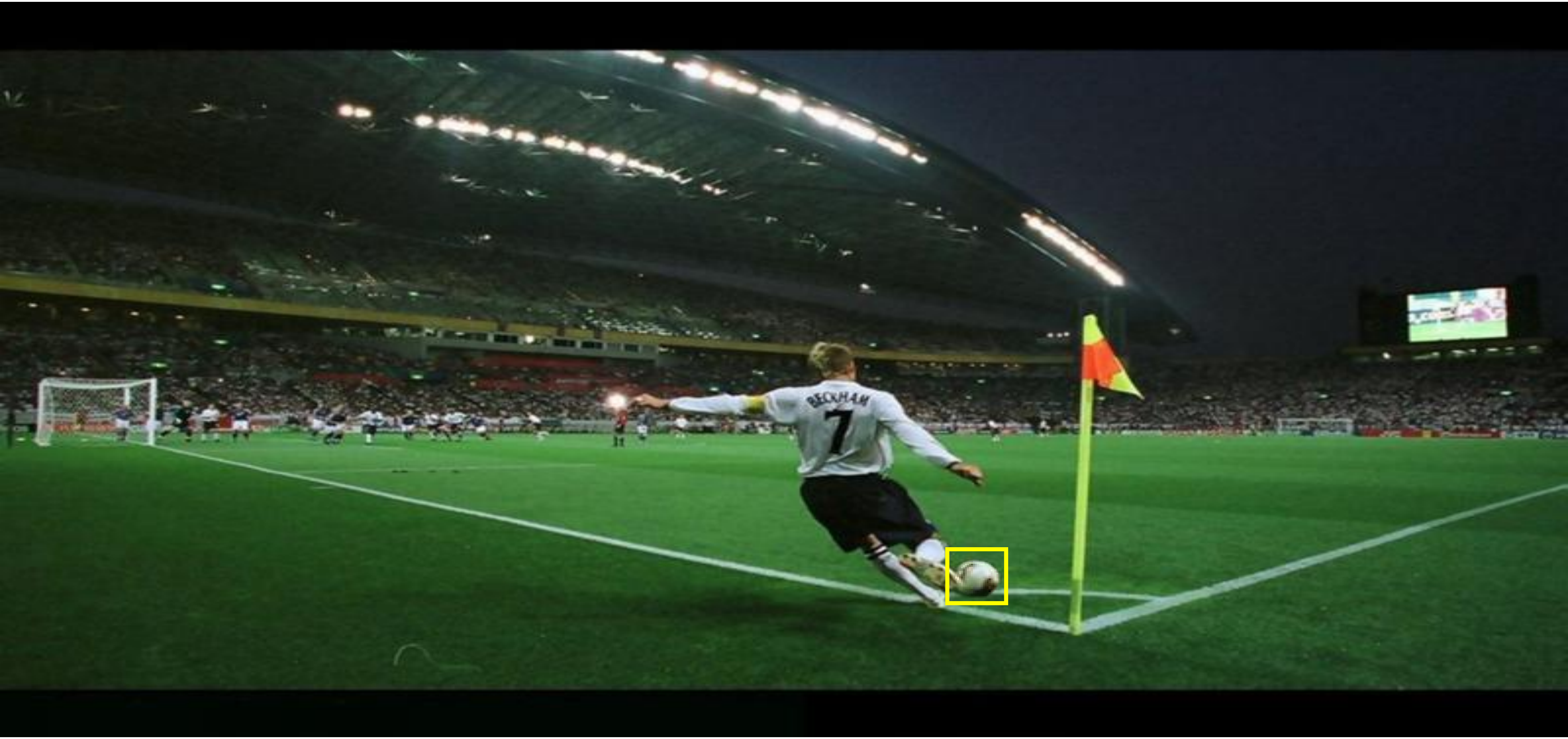
Intermediate Representation
Game Engine

Estimate friction, mass, etc.
Then solve some equations.

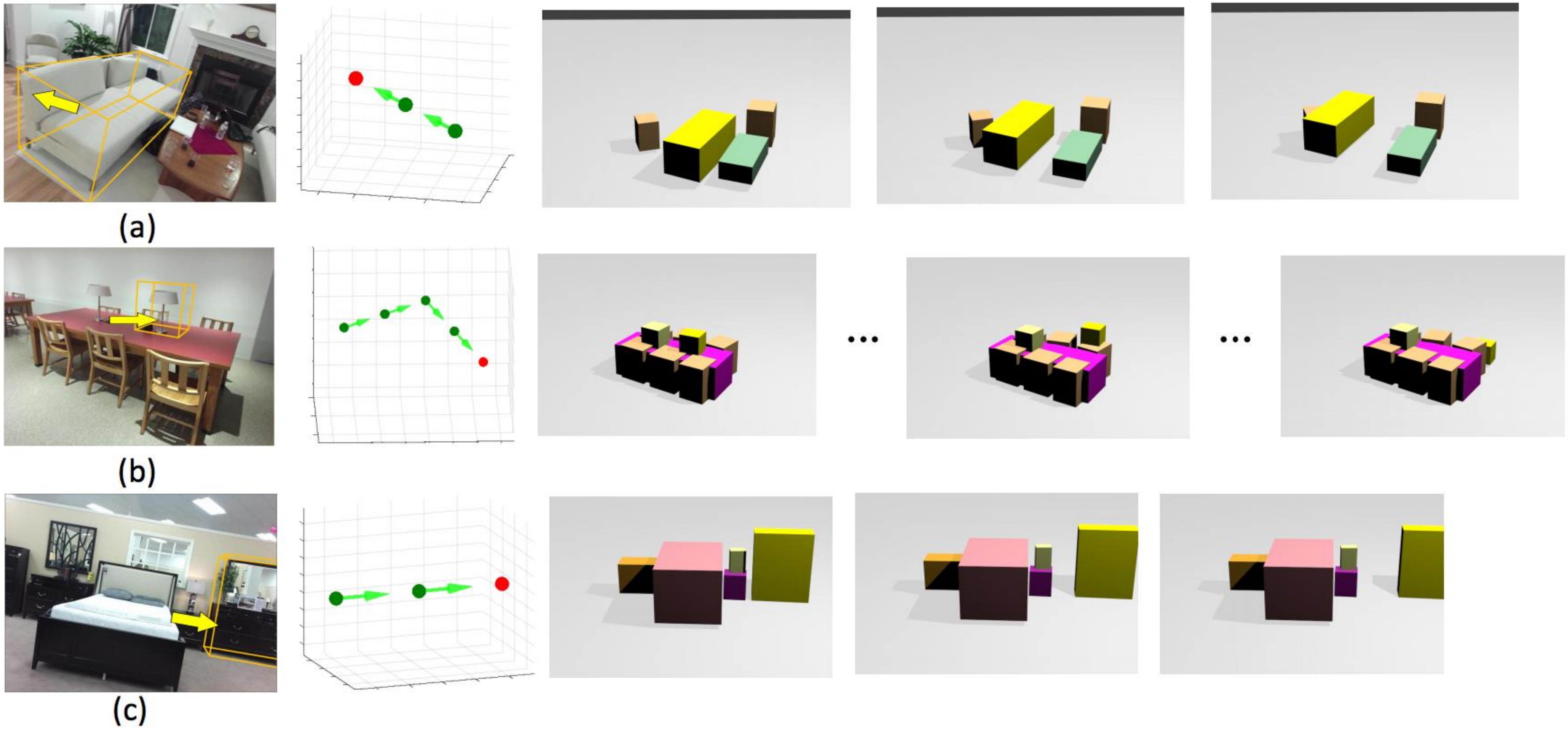







BECKHAM
7

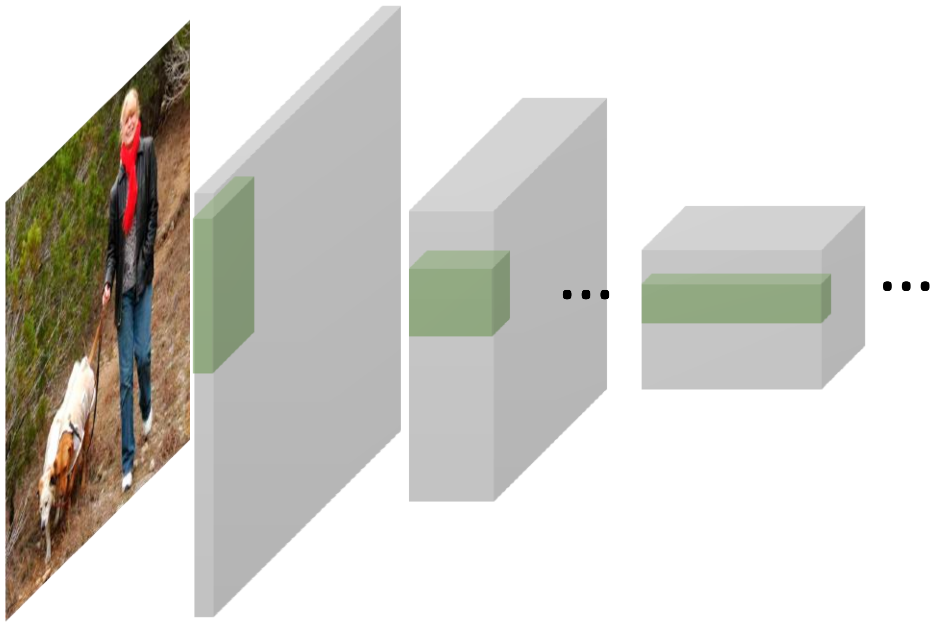
More results

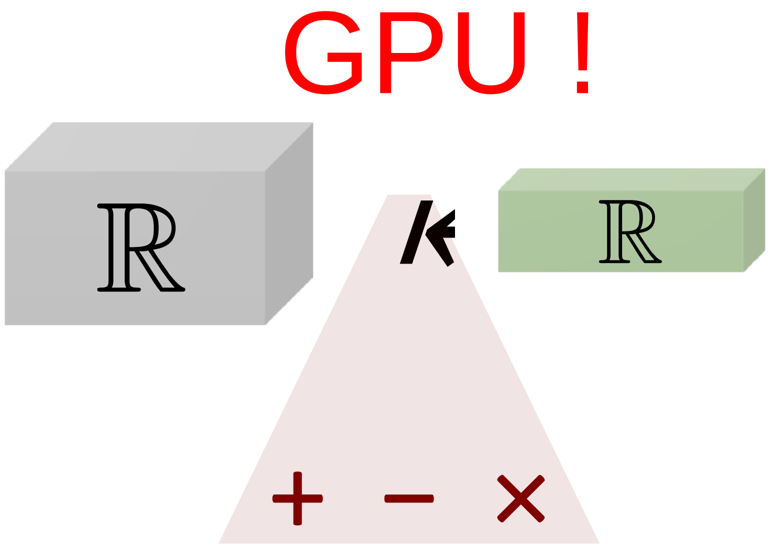


XNOR-Net

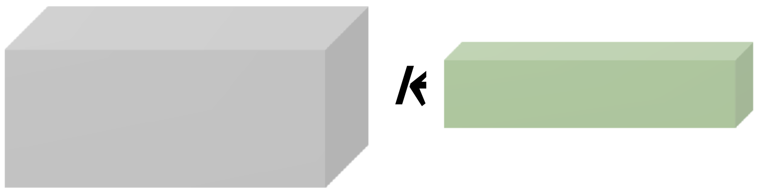
Image Classification using Binary CNNs

Convolutional Neural Networks




GPU !

Network	# operations	Inference (CPU)
AlexNet	1.5B FLOPs	~3 fps
VGG	19.6B FLOPs	~0.25 fps

	Operations	Memory	Computation
$\mathbb{R} * \mathbb{R}$	+ − ×	1x	1x
$\mathbb{R} * \mathbb{B}$	+ −	~32x	~2x
$\mathbb{B} * \mathbb{B}$	XNOR Bit-count	~32x	~58x

XNOR-NET Demo

On the iPhone!

Thank you!

Collaborators

Minjoon Seo, Eric Kolve, Mike Salvato

Jonghyun Choi, Jayant Krishnamurthy, Dustin Schwenk

Hannaneh Hajishirzi, Ali Farhadi

Projects by AI2 colleagues

Roozbeh Mottaghi, Mohammad Rastegari, Ali Farhadi