Commonsense for Deep Learning

Niket Tandon

Allen Institute for AI, Seattle

Ontology summit 2019

Andrew Ng: What AI Can and Can't Do

If a task takes you less than one second of thought, a machine can probably do it.

hbr.org



Source: Tandon 2018, invited talk on "DL for AI and NLP".

Challenge # 1: Robustness



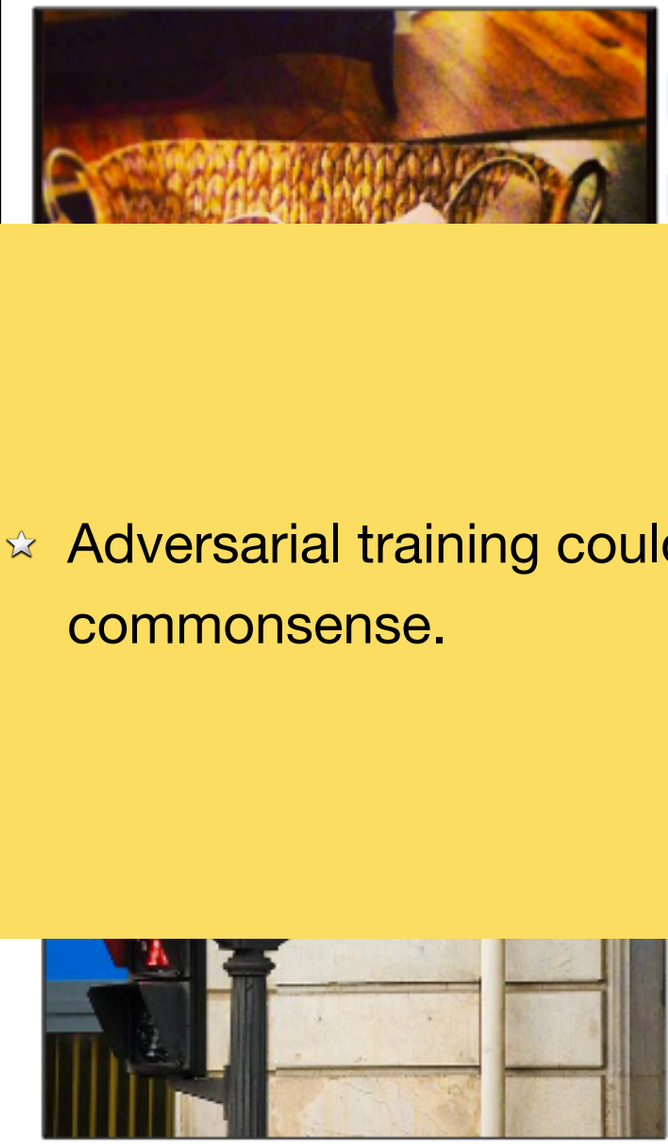
	Prediction
What is in the basket?	banana
What is contained in the basket?	pizza
What can be seen inside the basket?	remote
What does the basket mainly contain?	paper



Is it safe to turn left?	Yes
Can one safely turn left?	No
Would it be safe to turn left?	No
Would turning left considered safe in this picture?	Yes

[1]: Shah et. al 2019, Cycle-Consistency for Robust Visual Question Answering

Challenge # 1: Robustness



	Prediction
What is in the basket?	banana
What is contained in the basket?	pizza
Would it be safe to turn left?	No
Would turning left considered safe in this picture?	Yes

- Adversarial training could help. To generate adversarial data one route is commonsense.

[1]: Shah et. al 2019, Cycle-Consistency for Robust Visual Question Answering

Challenge # 2: A.Eye

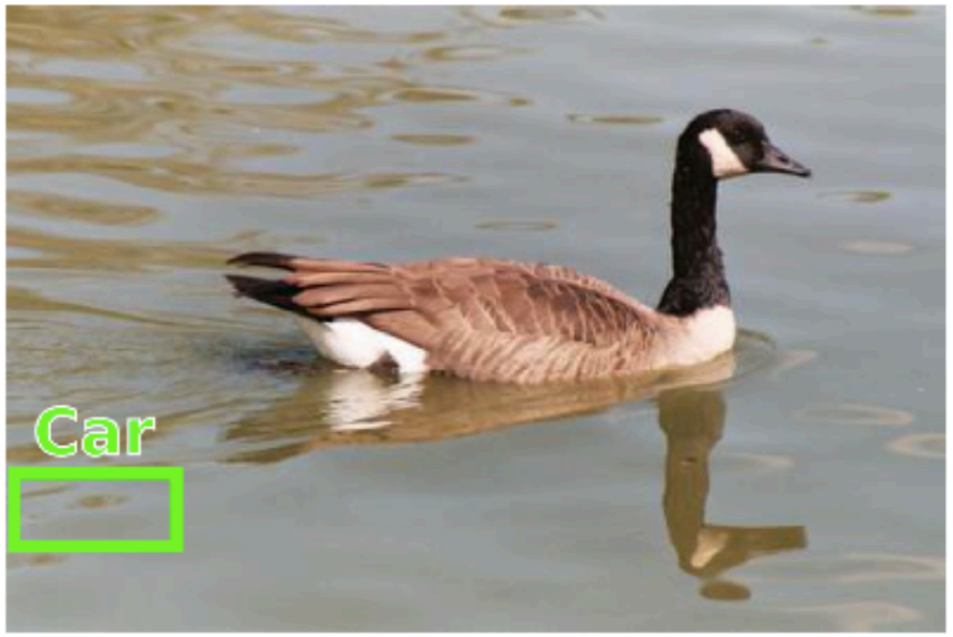
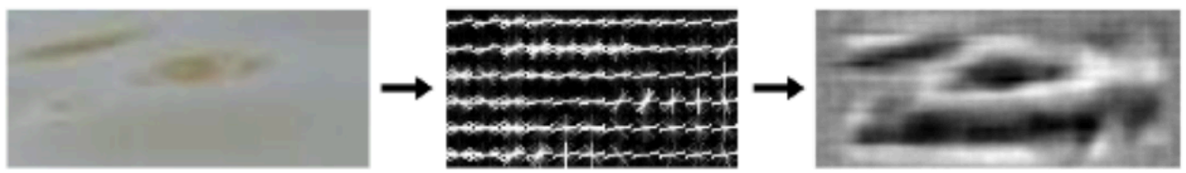
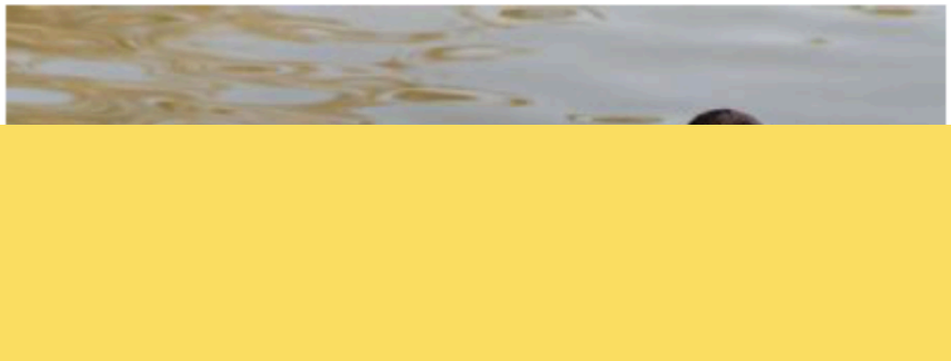


Figure 1: An image from PASCAL and a high scoring car detection from DPM [8]. Why did the detector fail?

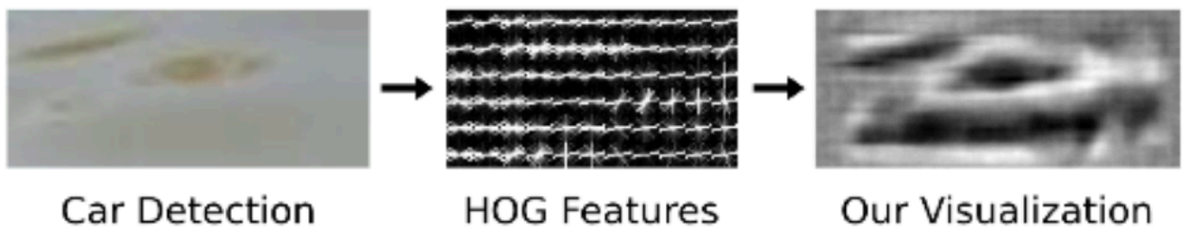


[2]: Deep neural networks are easily fooled: evolvingai.org/fooling

Challenge # 2: A.Eye

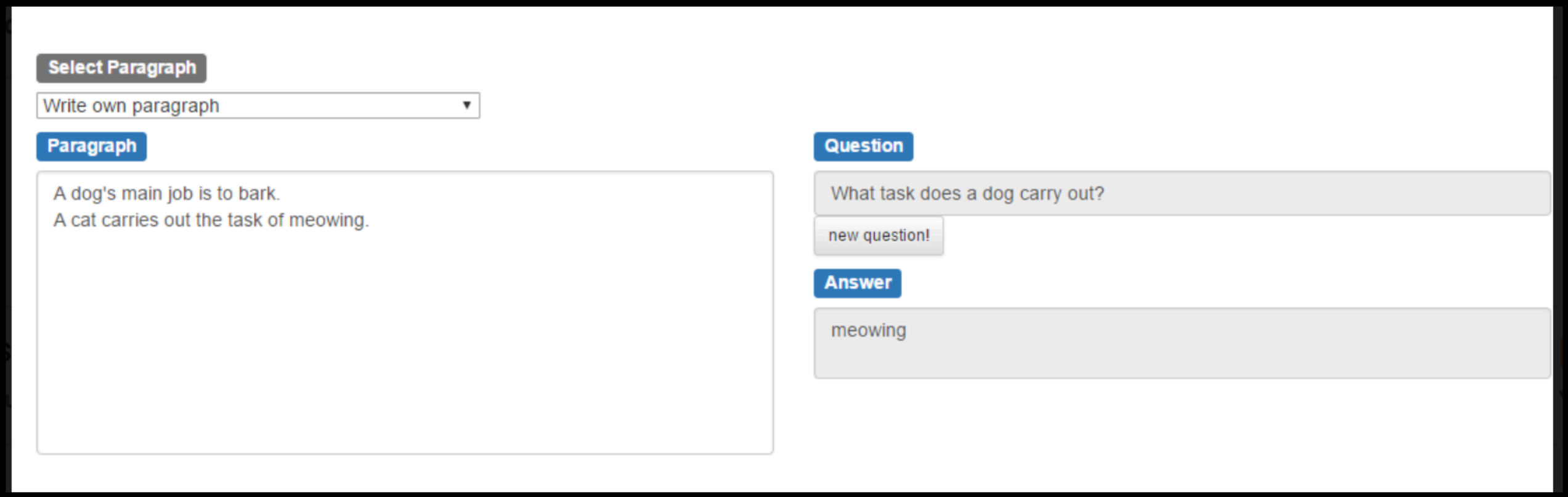


- Explanation would help during model debugging and decision making to correct the apparently unreasonable predictions through commonsense.



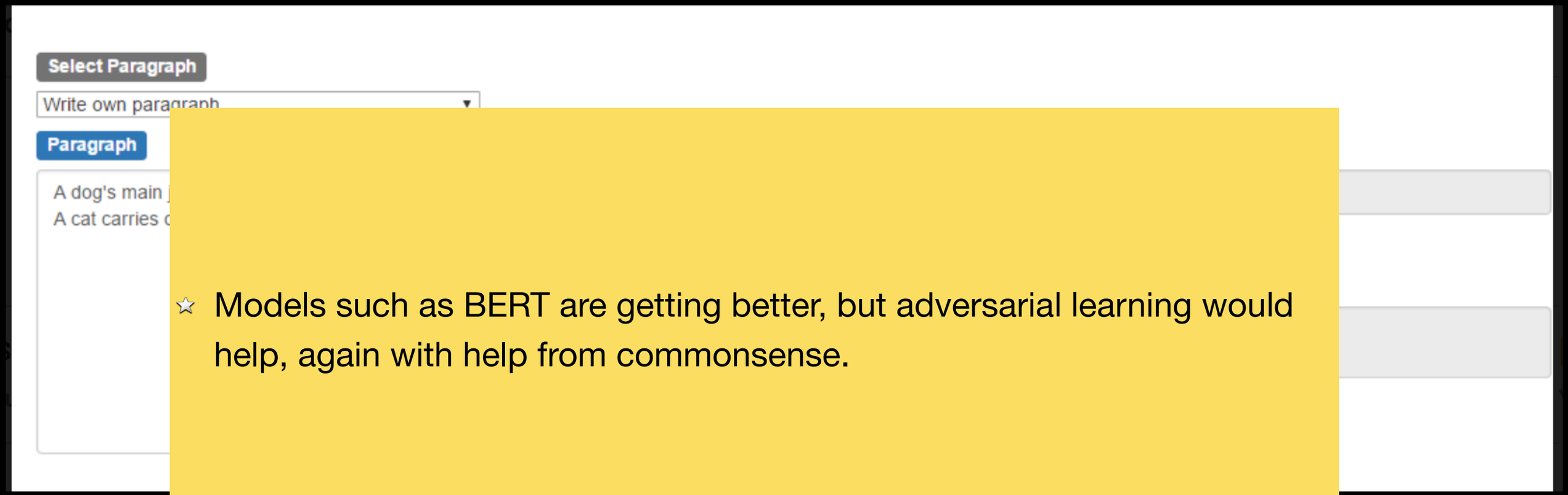
[2]: Deep neural networks are easily fooled: evolvingai.org/fooling

Challenge # 3: Flirt with words



Source: Peter Clark, AI2

Challenge # 3: Flirt with words



- Models such as BERT are getting better, but adversarial learning would help, again with help from commonsense.

Source: Peter Clark, AI2

Challenge # 4: Adversaries galore

Table 1: Failure examples from the SNLI dataset: negation (Top) and re-ordering (Bottom). **P** is premise, **H** is hypothesis, and **S** is prediction made by an entailment system (Parikh et al., 2016).

P: The dog did not eat all of the chickens.
H: The dog ate all of the chickens.
S: entails (score 56.5%)
P: The red box is in the blue box.
H: The blue box is in the red box.
S: entails (score 92.1%)

[3]: Kang et. al 2018, AdvEntuRe: Adversarial Training for Textual Entailment with Knowledge-Guided Examples

Challenge # 4: Adversaries galore

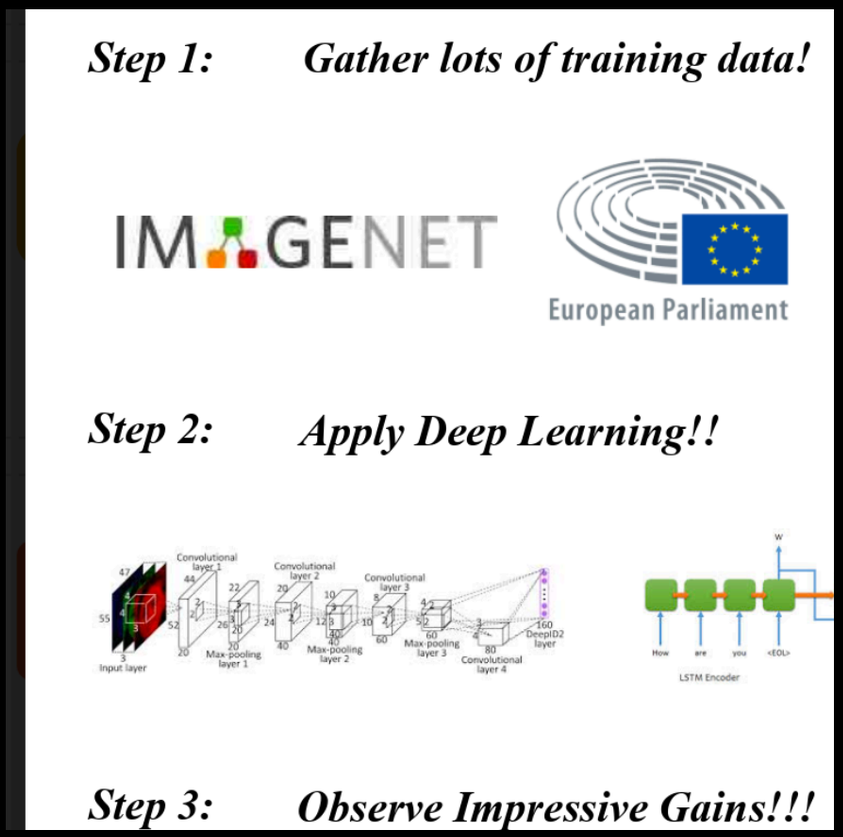
Table 1: Failure examples from the SNLI dataset: negation (Top) and re-ordering (Bottom). **P** is

S: entails (score 92.1%)

- Adversarial training and explanations would help. For both, commonsense would be useful.

[3]: Kang et. al 2018, AdvEntuRe: Adversarial Training for Textual Entailment with Knowledge-Guided Examples

Challenge # 5: Training sparsity

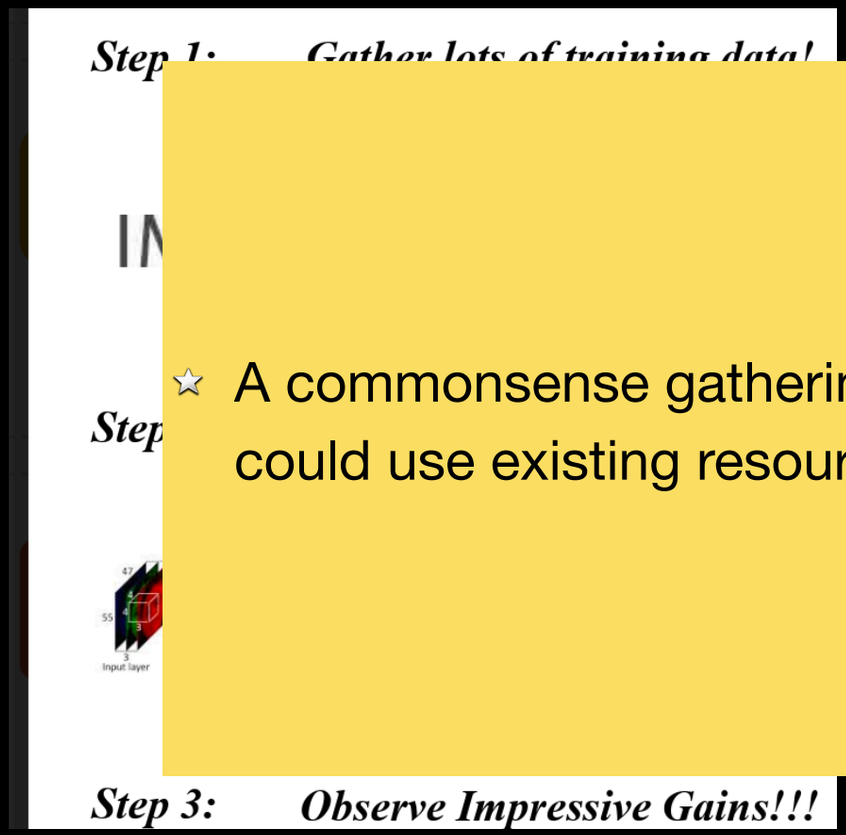


Difficult to find training data for all types of scenarios

- ★ Are shiny surfaces typically hard?
- ★ What's bigger the moon or a wolf?
- ★ If I put my socks in the drawer, will they still be there tomorrow?

Inspired by Oren Etzioni's talk on commonsense, Slides from Luke Zettlemoyer, UW

Challenge # 5: Training sparsity



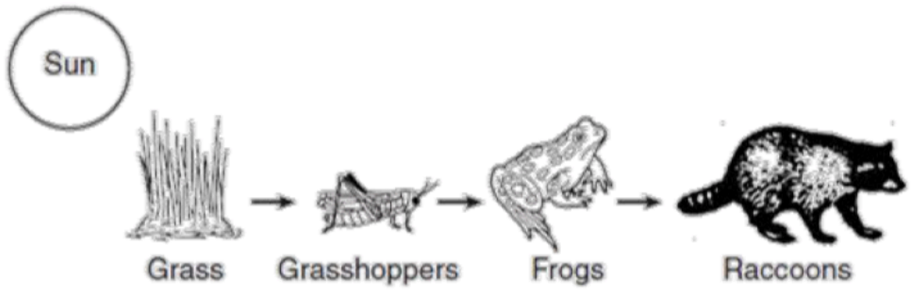
Difficult to find training data for all

ard?
wolf?
er, will
y?

- ☆ A commonsense gathering agent can learn from surrounding, or, we could use existing resources of commonsense.

Inspired by Oren Etzioni's talk on commonsense, Slides from Luke Zettlemoyer, UW

Challenge # 6: unseen situations



If all the frogs died, the raccoon population would most likely
(A) **decrease** (B) increase (C) remain the same



For roller-skate race, what is the best surface?
(A) sand (B) grass (C) **blacktop**

Picture depicts food web. Arrow indicates consumes.
If frogs die, raccoons won't get food and die-- so their population will decrease.

Humans

Human- Machine Knowledge Gap

Machines

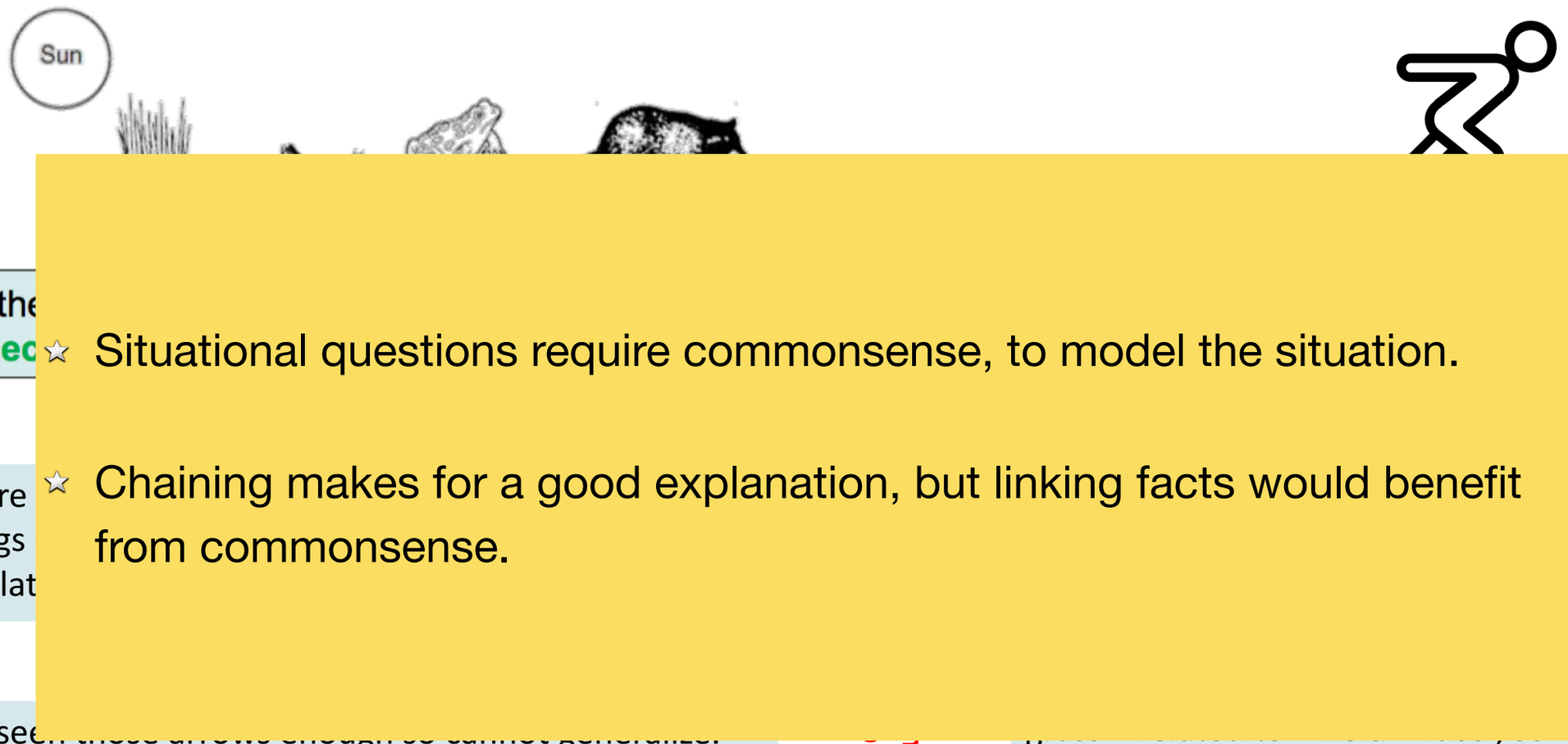
Roller skate is best on a smooth surface. Blacktop surfaces are shiny and shiny surfaces are smooth.

Not seen those arrows enough so cannot generalize.
Haven't seen raccoons, frogs in a sentence frequently

grass – related to – field – race , so "grass"
blacktop is not related to race.

[4] Peter Clark, Elementary School Science and Math Tests as a Driver for AI: Take the Aristo Challenge!

Challenge # 6: unseen situations



- ☆ Situational questions require commonsense, to model the situation.
- ☆ Chaining makes for a good explanation, but linking facts would benefit from commonsense.

Commonsense for Deep Learning

- For these reasons, commonsense aware models may:
  - help to create adversarial learning training data
  - generalize to other novel situations and domains
  - compensate for limited training data
  - be amicable to explanation e.g. with intermediate structures
- Underlying assumption: commonsense representation is DL friendly

Commonsense for Deep Learning

- For these reasons, commonsense aware models may:
  - help to create adversarial learning training data
  - generalize to other novel situations and domains
  - compensate for limited training data
  - be amicable to explanation
- Underlying assumption: commonsense representation is DL friendly

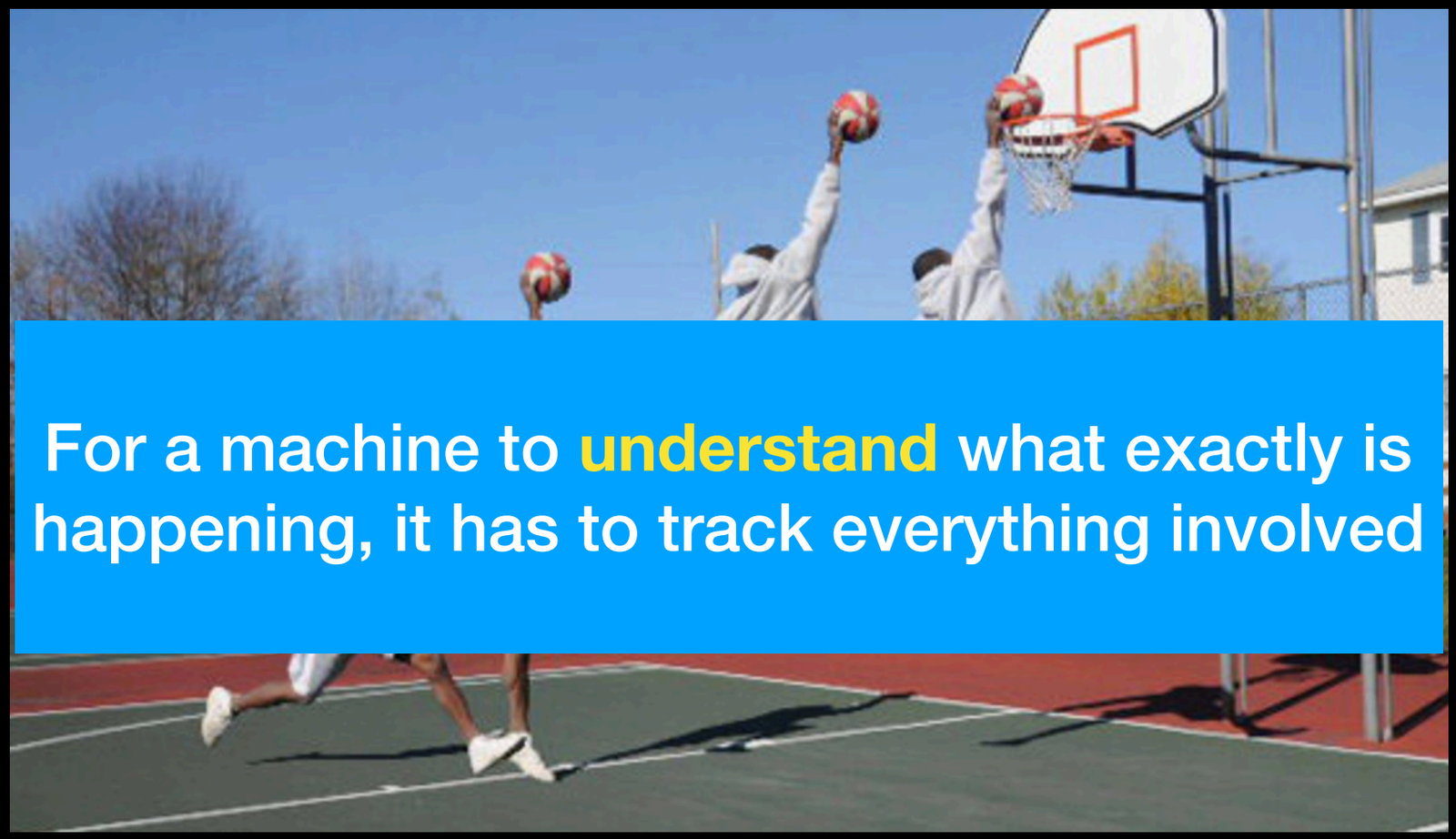
Explored in this talk.

Part I: CSK compensates for limited training data

Part II: Explanations and CSK

Part III: DL friendly commonsense representations

Hey computer, what is going on?



For a machine to **understand** what exactly is happening, it has to track everything involved

He charges forward

The ball is in his hands.

And a big leap

The player is in the air.

He made a basket

The ball is in the hoop.

Task: State change prediction

Actions

He charges forward → **And a big leap** → **He made a basket**

States

The ball is in his hands.

Location (ball) = player's hand

The player is in the air.

Location (player) = air

The ball is in the hoop.

Location (ball) = hoop

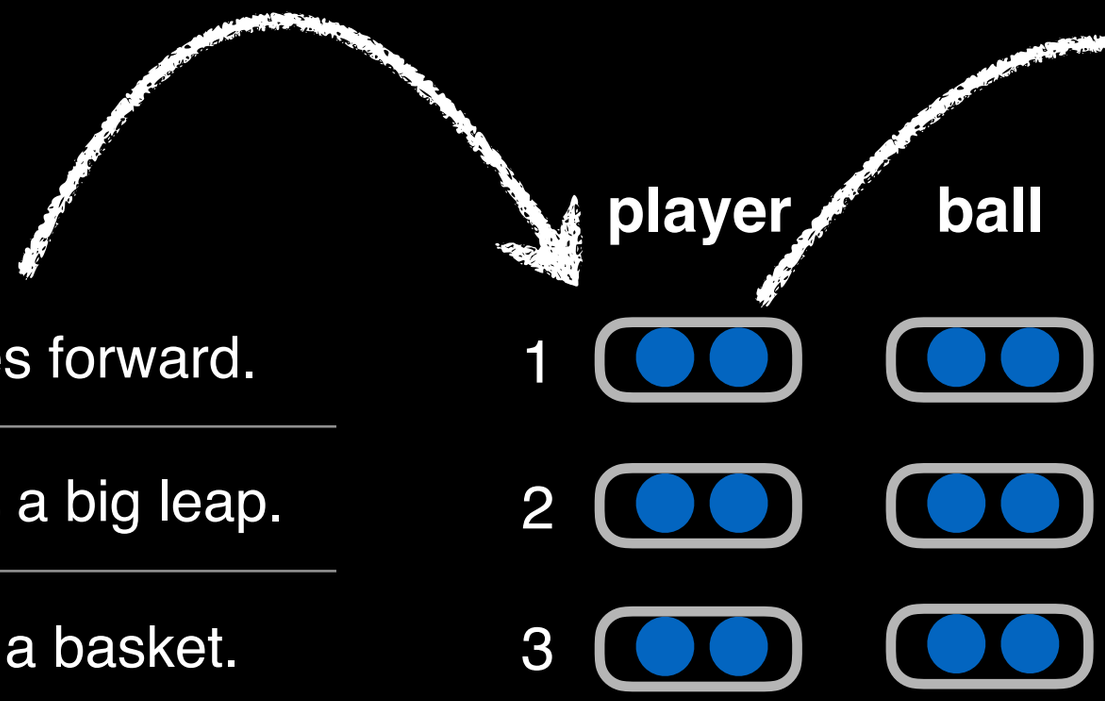
Training data is expensive
> 1$ to annotate one paragraph imperfectly

		player	**ball**
Step 1	He charges forward.	UNK	hand
Step 2	And makes a big leap.	air	hand
Step 3	He made a basket.	air	hoop

[1.1] Dalvi, Tandon, Clark: ProPara task: data.allenai.org/propara

Basic model

1	He charges forward.
2	And makes a big leap.
3	He made a basket.



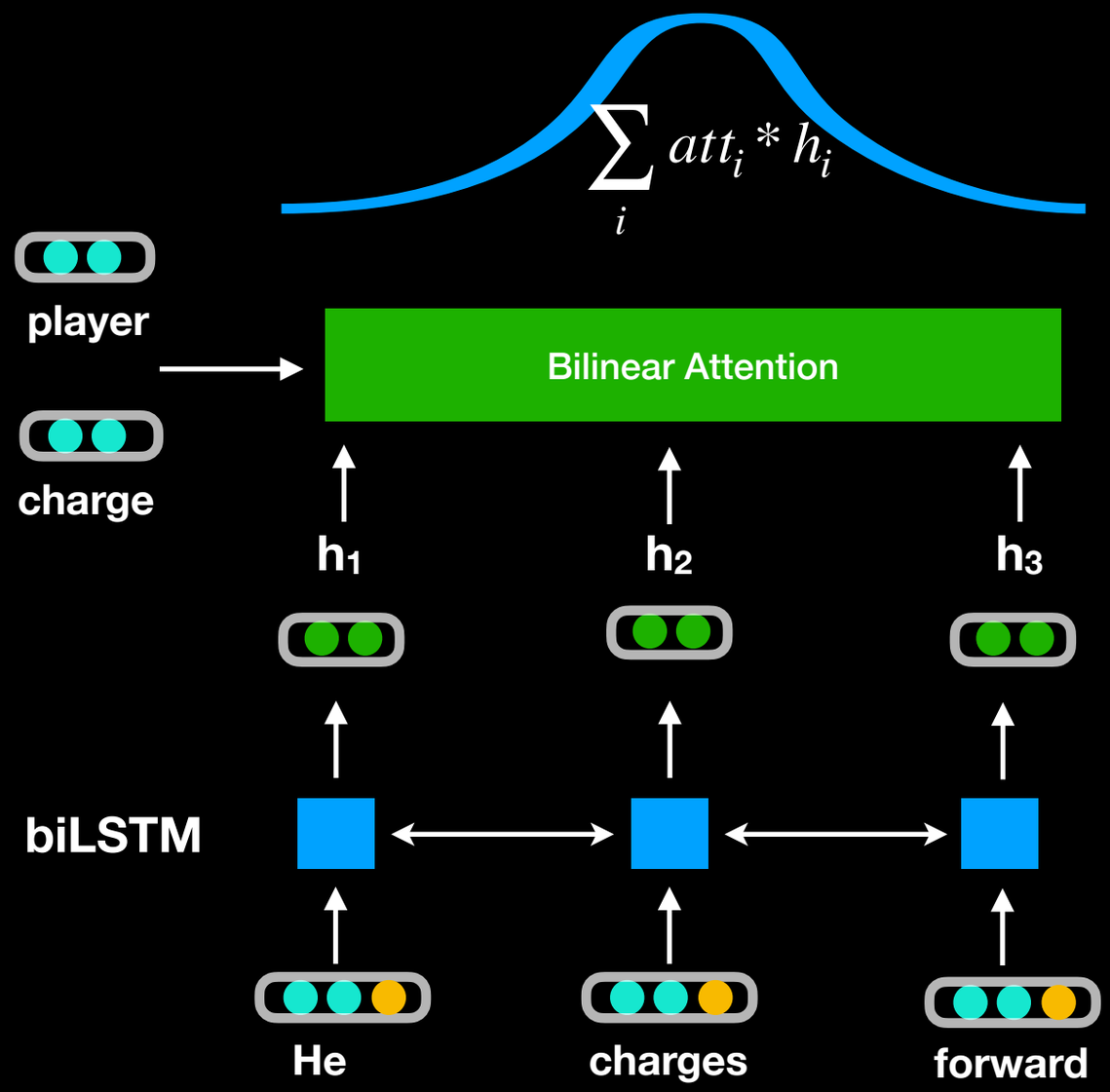
player	ball
MOVE unk air	
	MOVE hand hoop

action encodings

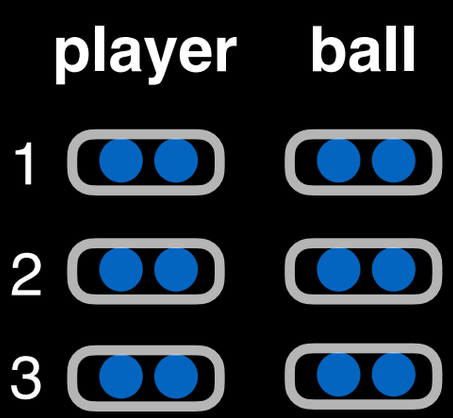
state change predictions

Encoder


player
ball
1
2
3

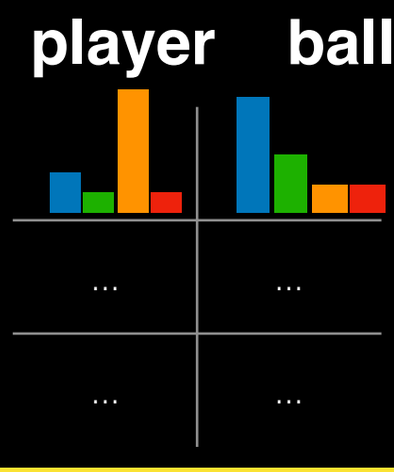

$\sum_i att_i * h_i$
player
charge
Bilinear Attention
h_1
h_2
h_3
biLSTM
He
charges
forward

Basic model



action encodings

MLP



label distribution

Inference

player	ball
	MOVE
MOVE	DESTROY
DESTROY	CREATE
...	

choose from many possible candidate structures

#P complete problem

player	ball
	CREATE
MOVE	
MOVE	MOVE

Basic model can predict non-sense

Actions

He charges forward

And a big leap

He made a basket

Model has seen insufficient data to learn these correlations, so we use commonsense to steer away from unrealistic states.

States

Commonsense

Runs with ball in hand.

Speed (player) = ↑

Location (ball) = player's hand

The player jumps in air.

Location (player) = air

Existence (ball) = destroyed

The ball is in the hoop.

Location (ball) = hoop

Location (player) = hoop

Commonsense aware model

player ball

	MOVE
MOVE	DESTROY
DESTROY	CREATE

player ball

1
2
3

action encodings

MLP

player ball

label distribution

Inference

...

The decoder performs beam search to find globally consistent results.

During search, probability mass moves away from implausible states.

player ball

	CREATE
MOVE	
MOVE	MOVE

Acquiring Commonsense Knowledge

Runs with ball in hand.
Speed (player) = ↑
Location (ball) = player's hand

The player jumps in air.
Location (player) = air
Existence (ball) = destroyed

The ball is in the hoop.
Location (ball) = hoop
Location (player) = hoop



**If the ball is destroyed,
how can it magically reappear in the hoop.**

We design a handful of universally applicable rules:

- An entity must exist before it can be moved or destroyed.
- An entity cannot be created if it already exists.
- ...

Acquiring Commonsense Knowledge

Runs with ball in hand.
Speed (player) = ↑
Location (ball) = player's hand

The player jumps in air.
Location (player) = air
Existence (ball) = destroyed

The ball is in the hoop.
Location (ball) = hoop
Location (player) = hoop



**If the ball is destroyed,
how can it magically reappear in the hoop.**

We design a handful of universally applicable rules:

- Our hard commonsense constraints (CS1 - CS3) are directly derivable from SUMO rules: MakingFn, DestructionFn, MotionFn, indicating that ontologies such as SUMO could be good guides for producing a handful of generic hard constraints in new domains.

Acquiring Commonsense Knowledge

Runs with ball in hand.	The player jumps in air.	**The ball is in the hoop.**
Speed (player) = ↑	**Location (player) = air**	**Location (ball) = hoop**
Location (ball) = player's hand	Existence (ball) = destroyed	**Location (player) = hoop**



This is not common, isn't it? :-)

It is very unlikely in a basketball like scenario for a player to be inside a hoop






Though it may be totally possible in another scenario!

Commonsense aware model

player ball

	MOVE
MOVE	DESTROY
DESTROY	CREATE

player ball

1
2
3

action encodings

MLP

player ball

label distribution

Inference

...

The decoder performs beam search to find globally consistent results.

During search, probability mass moves away from implausible states.

player ball

	CREATE
MOVE	
MOVE	MOVE

step 1

step 2

step 3

Inference

player	ball
	MOVE

player	ball
	MOVE
MOVE	DESTROY

player	ball
	MOVE
MOVE	DESTROY
DESTROY	CREATE

...

player	ball
DESTROY	MOVE

...

player	ball
DESTROY	MOVE
MOVE	

Hard constraints violation

...

player	ball
	CREATE
MOVE	
MOVE	MOVE

step 1

step 2

step 3

Inference

player **ball**

	MOVE

player **ball**

	MOVE
MOVE	DESTROY

player **ball**

	MOVE
MOVE	DESTROY
DESTROY	CREATE

Less likely for a ball to be destroyed in basketball

...

...

player **ball**

DESTROY	MOVE

player **ball**

DESTROY	MOVE
MOVE	

player **ball**

	CREATE
MOVE	
MOVE	MOVE

Hard constraints violation

Model confidence at step 1 step 2 step 3

$$\phi'(\pi_t) = \sum_{j=1}^{|E|} \left(\lambda \; logit(\pi_t^j) \quad +(1-\lambda) \; \log P(\pi_t^j \mid e_j, topic) \right)$$

Commonsense aware scoring

player	ball
	MOVE
MOVE	DESTROY
DESTROY	CREATE

player	ball
	MOVE
MOVE	DESTROY

player	ball
MOVE	

Less likely for a ball to be destroyed in basketball

Inference

...

player	ball
DESTROY	MOVE

player	ball
DESTROY	MOVE
MOVE	

Hard constraints violation

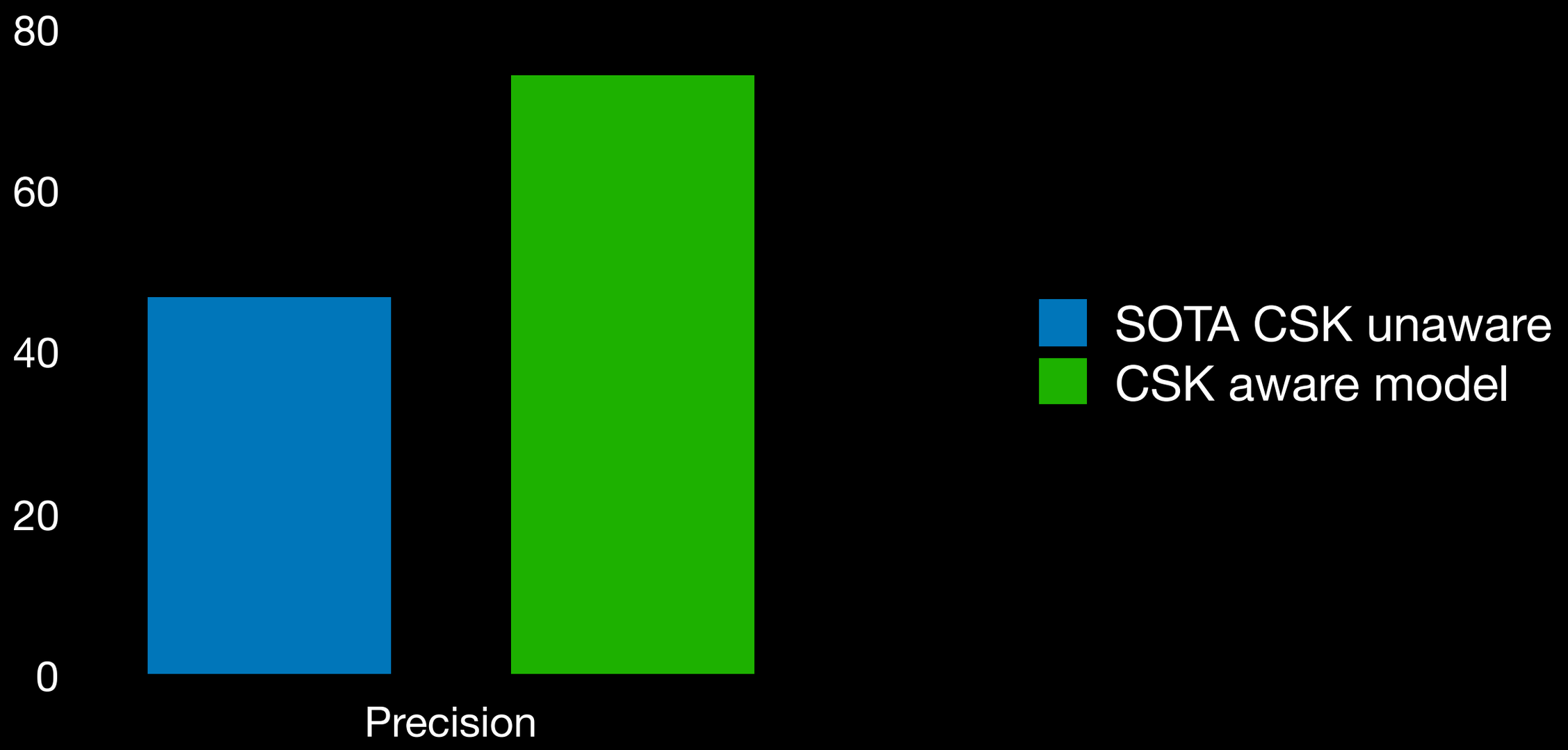
player	ball
	CREATE
MOVE	
MOVE	MOVE

...

Loss = - (score of gold path)

$$\mathscr{L} = -\sum_{t=1}^{T} \log \phi'(\pi_t)$$

Commonsense biased search improves precision by nearly 30% over SOTA



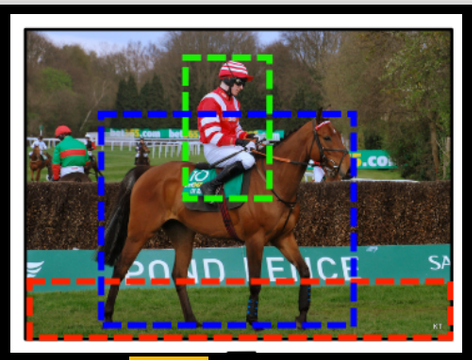
[1.2]: Tandon et. al 2018, Reasoning about Actions and State Changes by Injecting Commonsense Knowledge

Part I: CSK compensates for limited training data

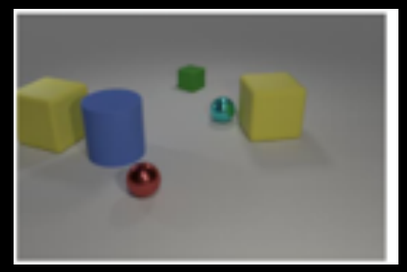
Part II: Explanations and CSK

Part III: DL friendly commonsense representations

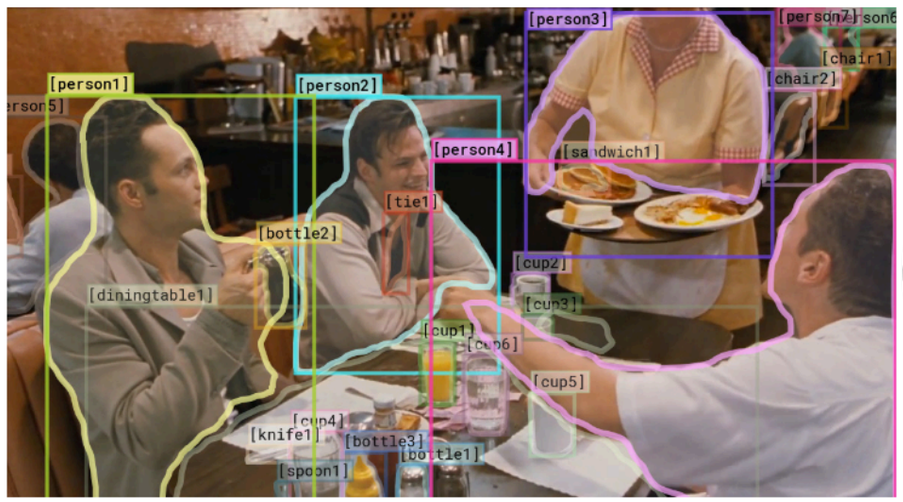
Vision datasets



[D1] **VQA:**
What is the man riding on?



[D2] **CLEVR:**
What is the shape of the object to the left of the blue cylinder?

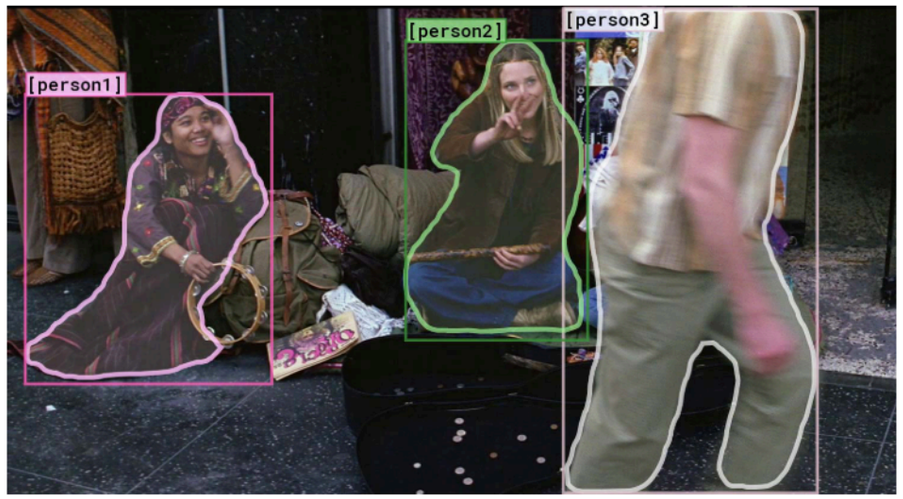


Why is [person4] pointing at [person1]?

a) He is telling [person3] that [person1] ordered the pancakes.
b) He just told a joke.
c) He is feeling accusatory towards [person1].
d) He is giving [person1] directions.

*I chose **a)** because...*

a) [person1] has the pancakes in front of him.
b) [person4] is taking everyone's order and asked for clarification.
c) [person3] is looking at the pancakes and both she and [person2] are smiling slightly.
d) [person3] is delivering food to the table, and she might not know whose order is whose.



How did [person2] get the money that's in front of her?

a) [person2] is selling things on the street.
b) [person2] earned this money playing music.
c) She may work jobs for the mafia.
d) She won money playing poker.

*I chose **b)** because...*

a) She is playing guitar for money.
b) [person2] is a professional musician in an orchestra.
c) [person2] and [person1] are both holding instruments, and were probably busking for that money.
d) [person1] is putting money in [person2]'s tip jar, while she plays music.

[D3]
VCR

e-SNLI [D4]

Premise: An adult dressed in black holds a stick.
Hypothesis: An adult is walking away, empty-handed.
Label: contradiction
Explanation: Holds a stick implies using hands so it is not empty-handed.

Premise: A child in a yellow plastic safety swing is laughing as a dark-haired woman in pink and coral pants stands behind her.
Hypothesis: A young mother is playing with her daughter in a swing.
Label: neutral
Explanation: Child does not imply daughter and woman does not imply mother.

Premise: A man in an orange vest leans over a pickup truck.
Hypothesis: A man is touching a truck.
Label: entailment
Explanation: Man leans over a pickup truck implies that he is touching it.

Figure 1: Examples from e-SNLI. Annotators were given the premise, hypothesis, and label. They highlighted the words that they considered essential for the label and provided the explanations.

HOTPOT-QA [D5]

Paragraph A, Return to Olympus:
[1] *Return to Olympus is the only album by the alternative rock band Malfunkshun.* [2] *It was released after the band had broken up and after lead singer Andrew Wood (later of Mother Love Bone) had died of a drug overdose in 1990.* [3] Stone Gossard, of Pearl Jam, had compiled the songs and released the album on his label, Loosegroove Records.

Paragraph B, Mother Love Bone:
[4] *Mother Love Bone was an American rock band that formed in Seattle, Washington in 1987.* [5] The band was active from 1987 to 1990. [6] *Frontman Andrew Wood's personality and compositions helped to catapult the group to the top of the burgeoning late 1980s/early 1990s Seattle music scene.* [7] *Wood died only days before the scheduled release of the band's debut album, "Apple", thus ending the group's hopes of success.* [8] The album was finally released a few months later.

Q: What was the former band of the member of Mother Love Bone who died just before the release of "Apple"?
A: Malfunkshun
Supporting facts: 1, 2, 4, 6, 7

Aristo exam

When S. marcescens, a bacterium, is grown in a refrigerator, it produces red-colored colonies. However, if the bacterium is grown at room temperature, the colonies are white. The best explanation for this situation is that
(1) refrigeration changes the structure of genes
(2) room temperature stimulates the synthesis of a red pigment
(3) temperature has an effect on the expression of genes
(4) only temperature is responsible for the expression of a trait

Question: Which event involves a consumer and a producer in a food chain ?

Process Roles Table

	PROCESS NAME		ACTOR		ROLE		ACTION	PATIENT		PURPOSE
In the	food chain	process, a	green plant	has the role of	producer	which	creates	food	for	consumers
In the	food chain	process, an	animal	has the role of	consumer	which	eats	producers/other animals	for	food
In the	food chain	process, a	bacteria	has the role of	decomposer	which	recycles	nutrients		
In the	tree reproduction	process, a	squirrel	has the role of	seed disperser	which	relocates	seeds		

Taxonomy Table

	HYPONYM		SCOPE	HYPERNYM
A	deer	is a kind of		animal
	green	is a kind of		color
	shelter	is a kind of	protective	covering
An	electromagnet	is a kind of	electric	magnet

PartOf Table

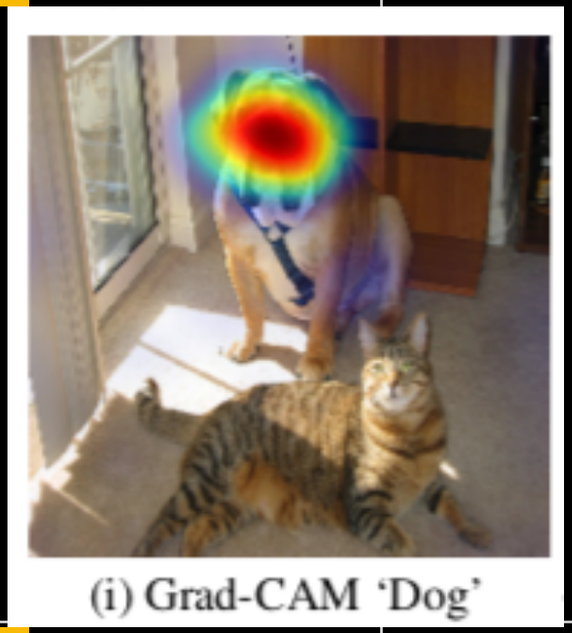
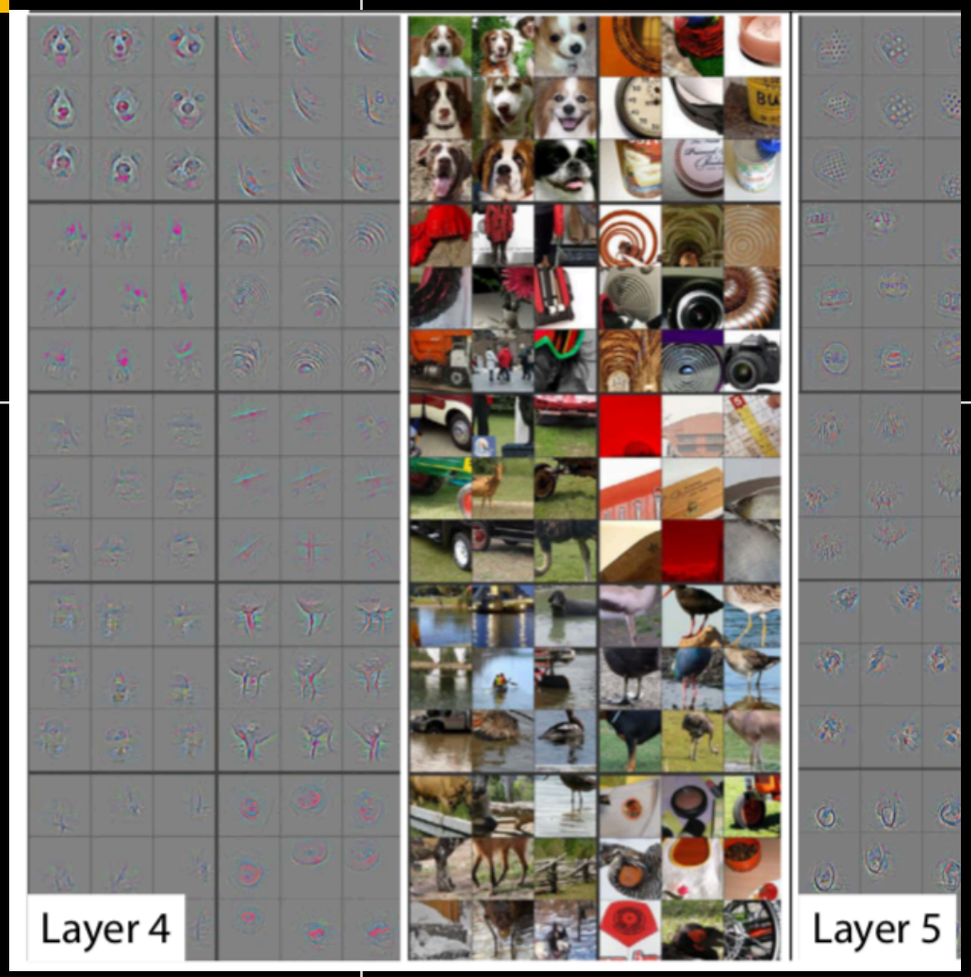
	PART		WHOLE
A	leaf	is a part of a	green plant
	roots	are a part of a	plant
	pedals	are a part of a	bicycle
A	cell wall	is a part of a	plant cell

Answer Candidates: [A] a cat eats a mouse [*B] a deer eats a leaf [C] a hawk eats a mouse [D] a snake eats a rat

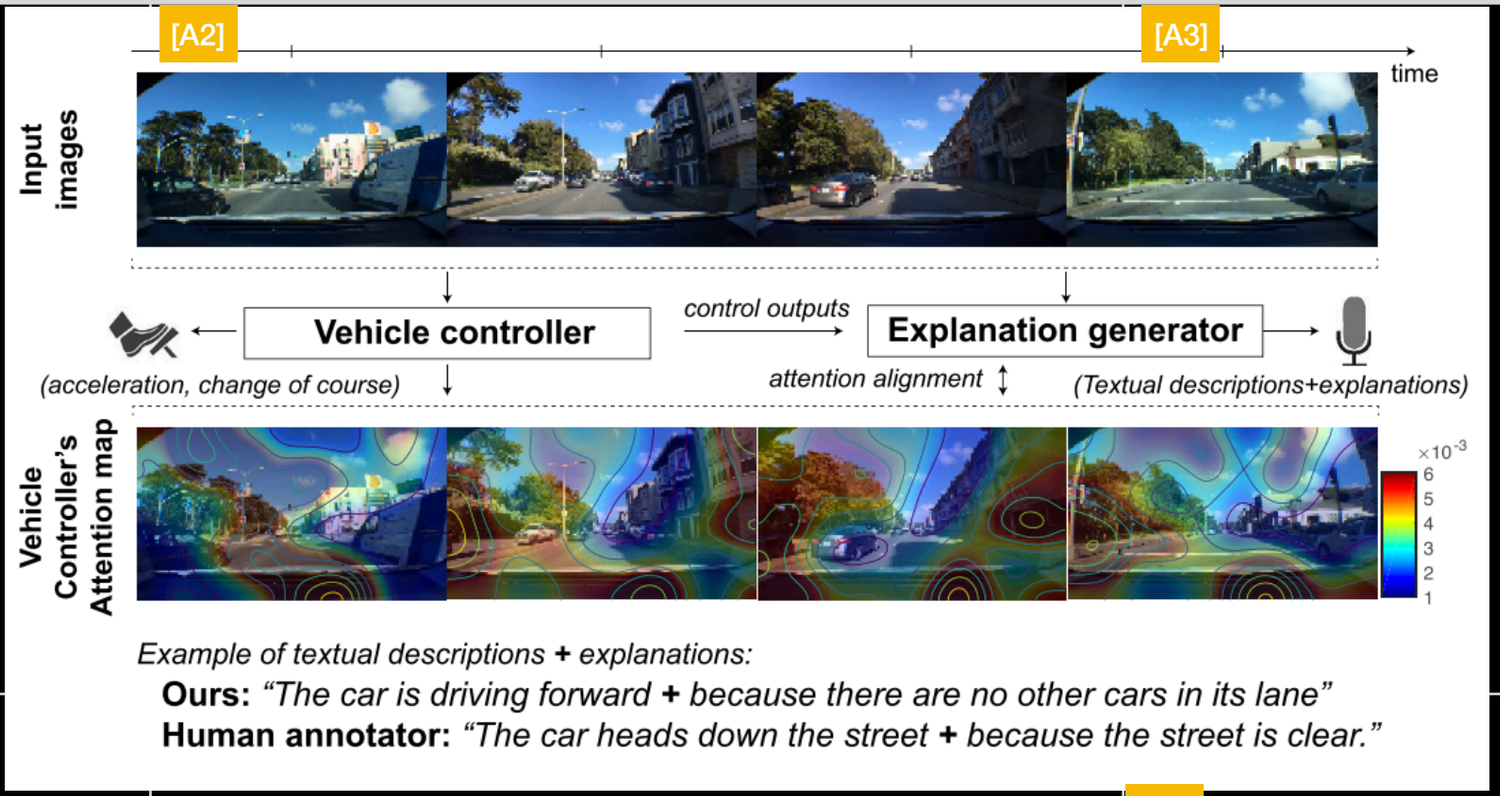
Figure 2: Examples of tables and table rows from the tablestore, grounded in an example question and explanation. Table columns define the primary roles or arguments for a given relation (e.g. *process name, actor, role, etc*). Unlabeled "filler" columns allow each row to be used as a stand-alone natural language sentence. Note that for clarity only 4 example rows per table are shown.[3]

[D6]

WorldTree

	Attention	Natural Language	Structure
Justification	[A1]  (i) Grad-CAM 'Dog'	[A2]	[A3]
Introspective	[A4] 		[A6]
J & I			[A9]

	Attention	Natural Language	Structure
Justification	[A1]	[A2] [A3]	
Introspective	[A4]	[A6] [A8]	
J & I			



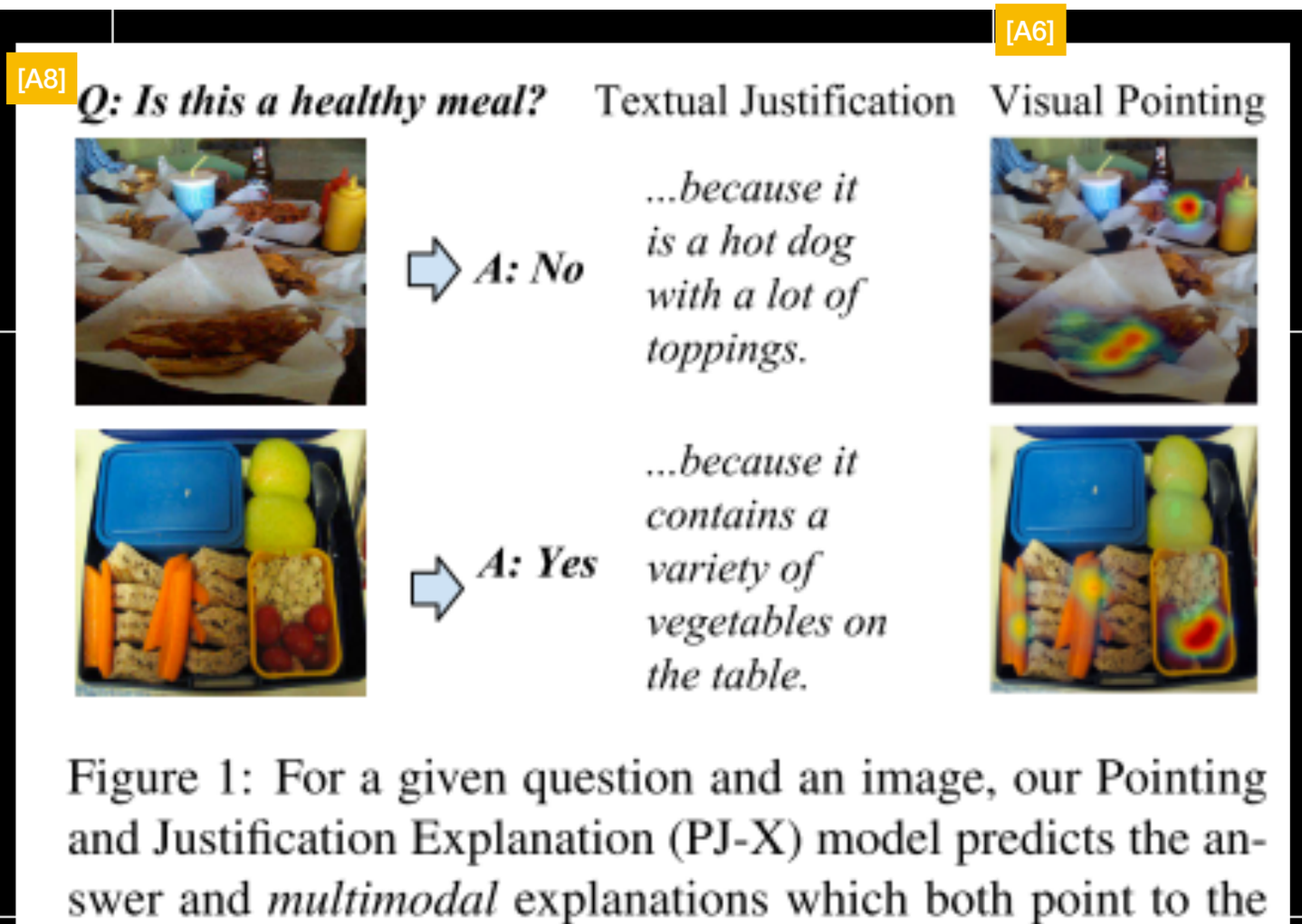
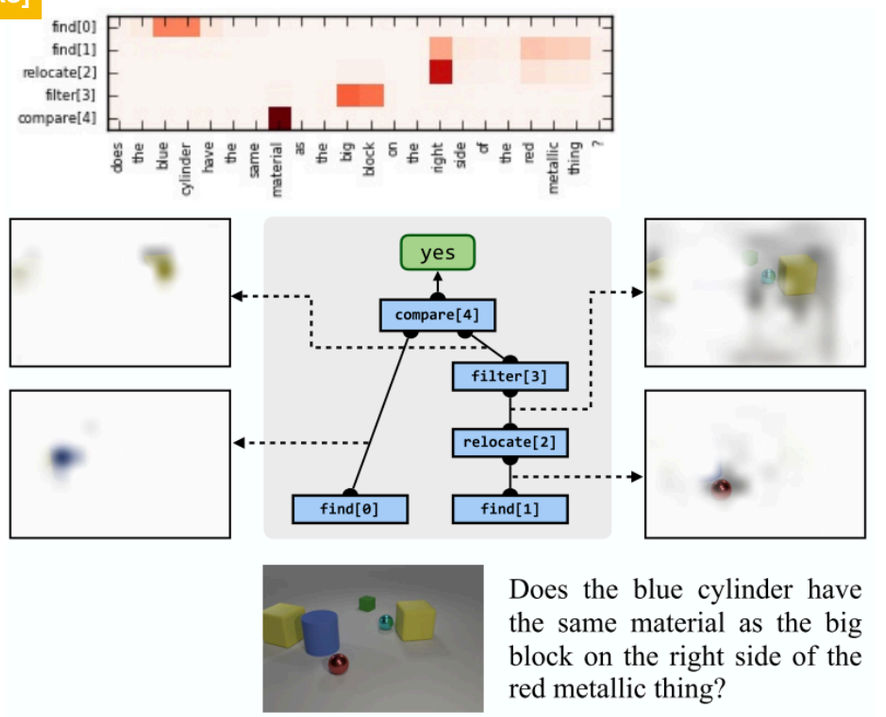
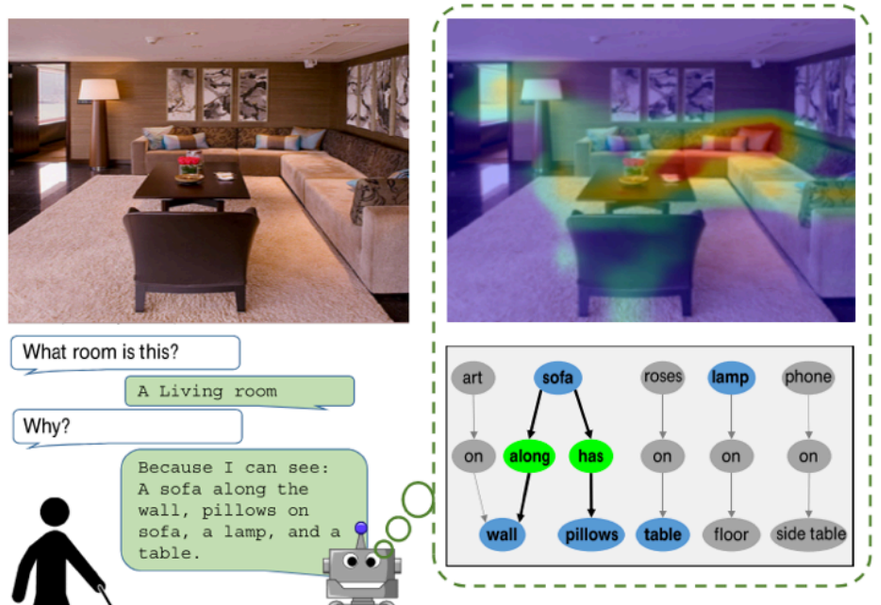


Figure 1: For a given question and an image, our Pointing and Justification Explanation (PJ-X) model predicts the answer and *multimodal* explanations which both point to the

	Attention	Natural Language	Structure
Justification	[A1]	[A2]	[A3]
Introspective	[A4]		[A6]
J & I	[A8]		[A9]



Paragraph (seq. of steps):		Participants: water	light	CO2	mixture	sugar
	state0	soil	sun	?	-	-
Roots absorb water from soil						
	state1	roots	sun	?	-	-
The water flows to the leaf.						
	state2	leaf	sun	?	-	-
Light from the sun and CO2 enter the leaf.						
	state3	leaf	leaf	leaf	-	-
The light, water, and CO2 combine into a mixture.						
	state4	-	-	-	leaf	-
Mixture forms sugar.						
	state5	-	-	-	-	leaf



Giving a voice to structures

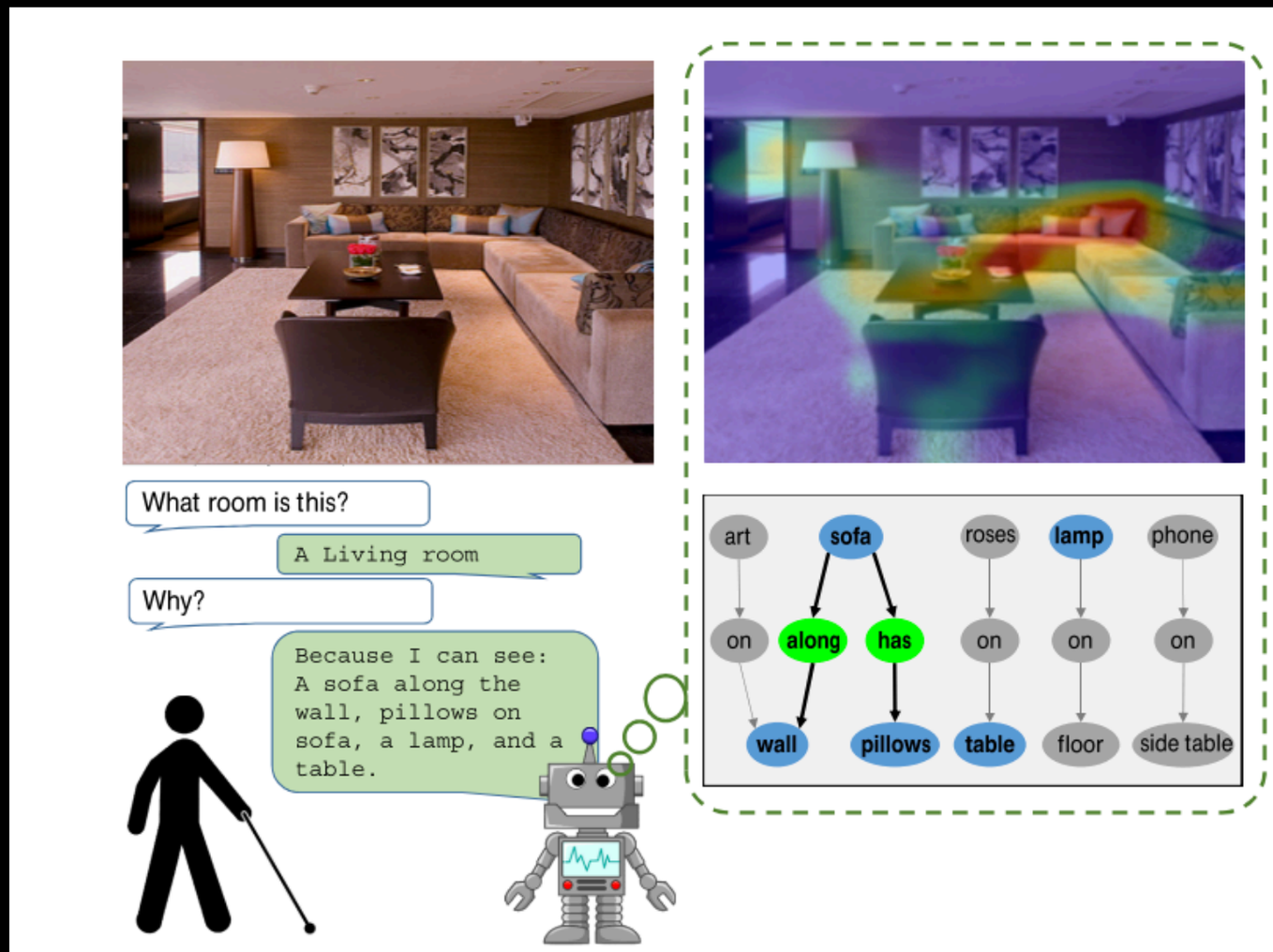
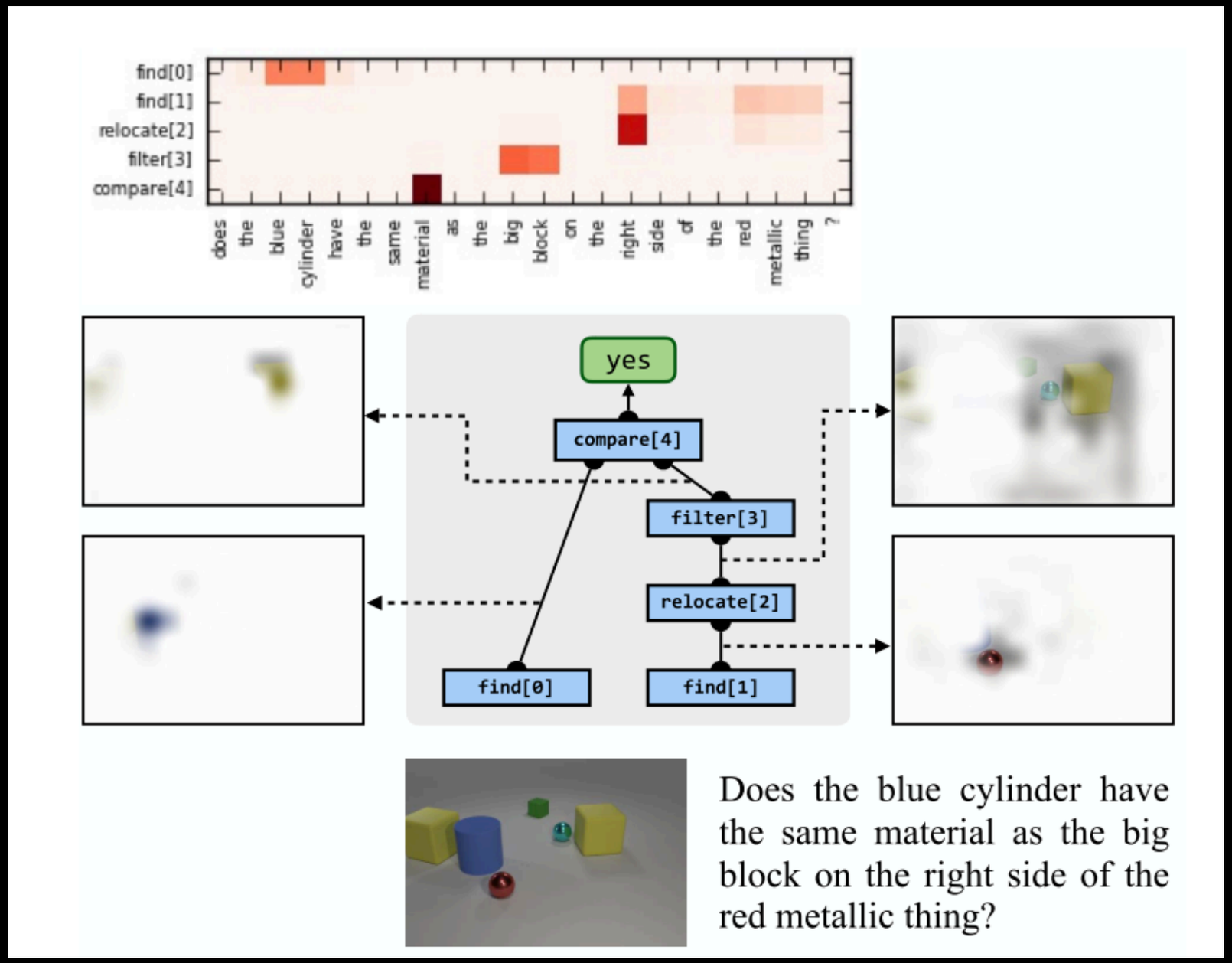


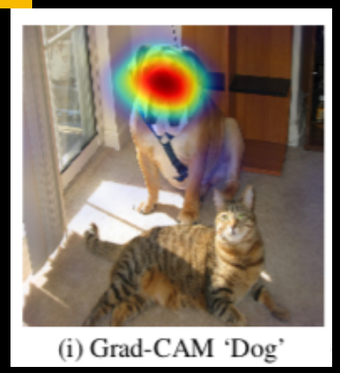
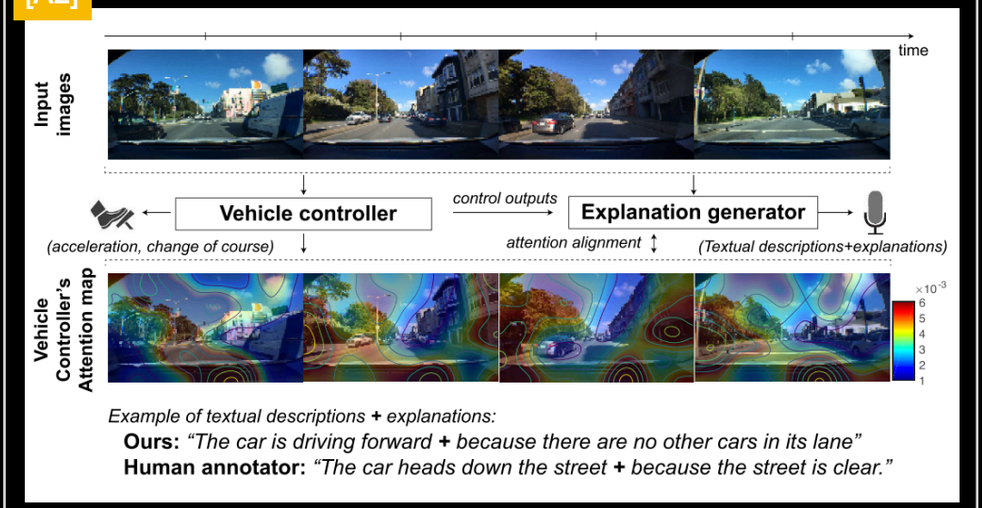
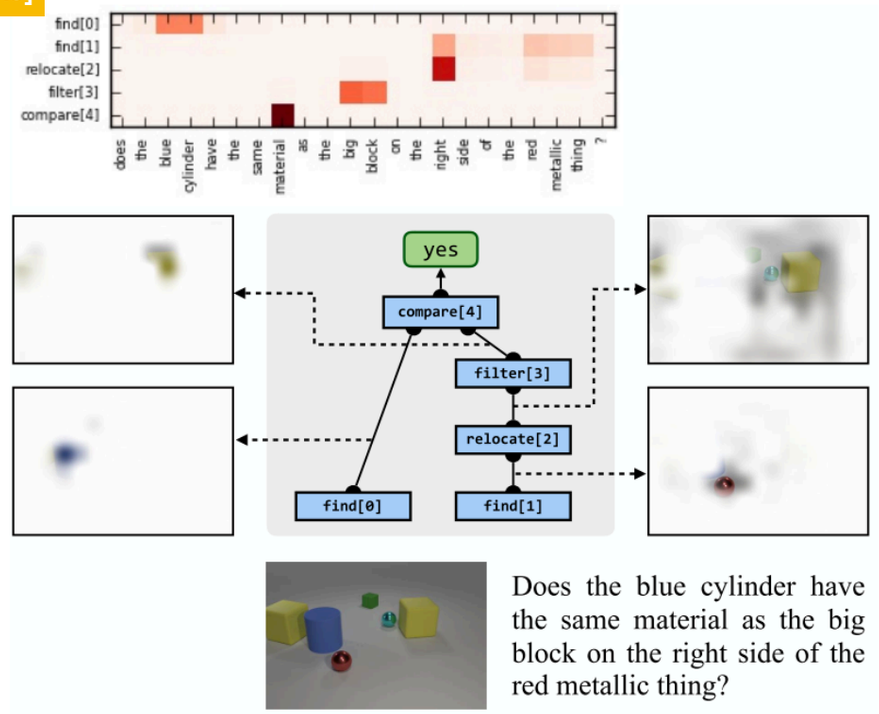
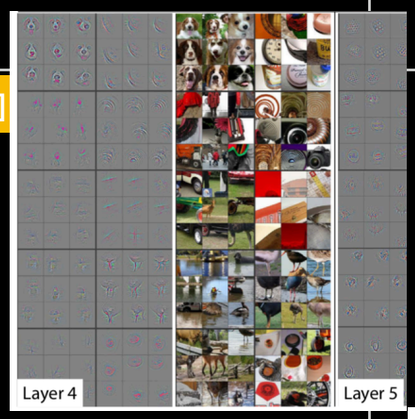
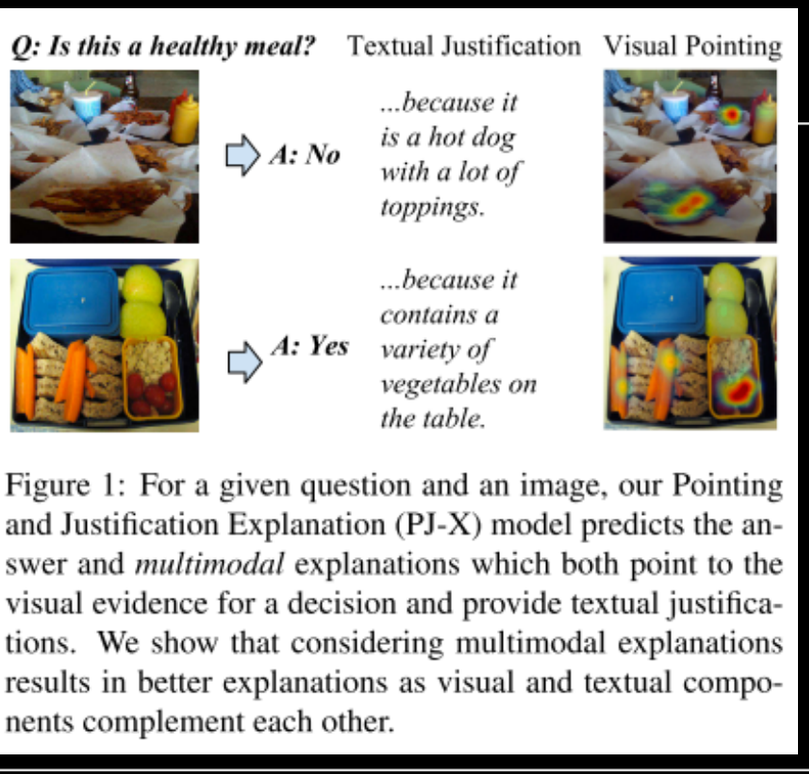
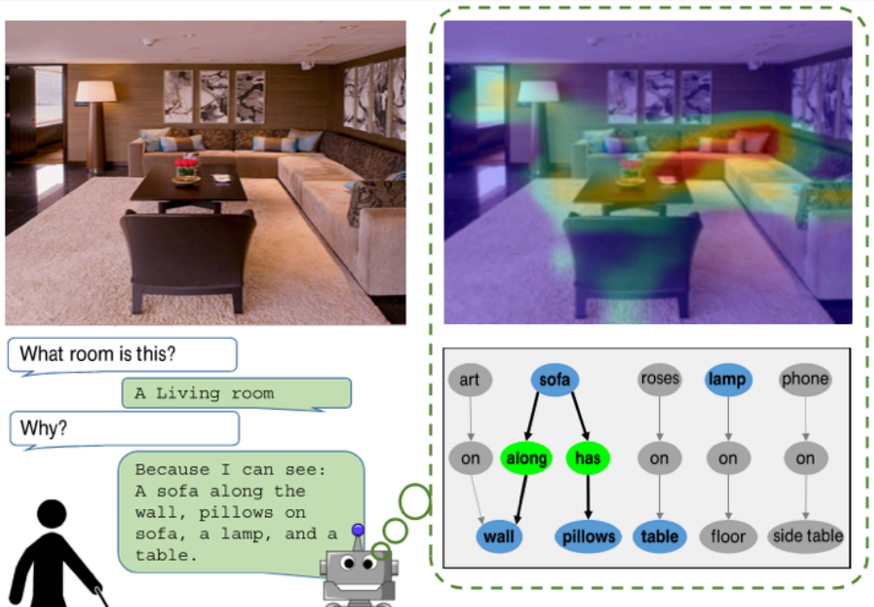
Figure 1: Example Natural Language explanation for an answer to a visually-grounded question. Our approach uses the attention map generated by a visual question-answering model (top right) to identify the relevant components of a scene graph (bottom right) that can be used to provide a justification for the answer given by the model.

Ghosh et. al 2018, Generating Natural Language Explanations for Visual QA Using Scene Graphs and Visual Attention

Structure: execution programs



Hu et. al , 2017, Learning to Reason: End-to-End Module Networks for Visual Question Answering

	Attention	Natural Language	Structure
Justification	[A1]  (i) Grad-CAM 'Dog'	[A2]  *Example of textual descriptions + explanations:* **Ours:** *"The car is driving forward + because there are no other cars in its lane"* **Human annotator:** *"The car heads down the street + because the street is clear."*	[A3]  Does the blue cylinder have the same material as the big block on the right side of the red metallic thing?
Introspective	[A4] 	[A6]	
J & I	[A8]  *Q: Is this a healthy meal?* Textual Justification Visual Pointing *A: No* *...because it is a hot dog with a lot of toppings.* *A: Yes* *...because it contains a variety of vegetables on the table.* Figure 1: For a given question and an image, our Pointing and Justification Explanation (PJ-X) model predicts the answer and *multimodal* explanations which both point to the visual evidence for a decision and provide textual justifications. We show that considering multimodal explanations results in better explanations as visual and textual components complement each other.	[A9] 

[A6]

		Participants:				
Paragraph (seq. of steps):		water	light	CO2	mixture	sugar
	state0	soil	sun	?	-	-
Roots absorb water from soil						
	state1	roots	sun	?	-	-
The water flows to the leaf.						
	state2	leaf	sun	?	-	-
Light from the sun and CO2 enter the leaf.						
	state3	leaf	leaf	leaf	-	-
The light, water, and CO2 combine into a mixture.						
	state4	-	-	-	leaf	-
Mixture forms sugar.						
	state5	-	-	-	-	leaf

Evaluation of explanation

	Natural language	**Structure**
Exact match	Not desirable for language generation tasks	Good, but schema is still predefined
METEOR	Better than exact match	- N/A
Human (e.g., show only explanation and ask human to guess model's prediction)	suffers from human-bias	very costly
Automated (e.g., hide attended pixels e.g., [2.2])	Promising direction but limited to vision as is	- N/A
Learned (e.g., [2.1] mentioned next)	Training data generation may be tricky	- N/A

Other issues: BERT could directly learn to explain without any understanding given Multiple choices e.g. due to annotation artifacts

Evaluating Explanations

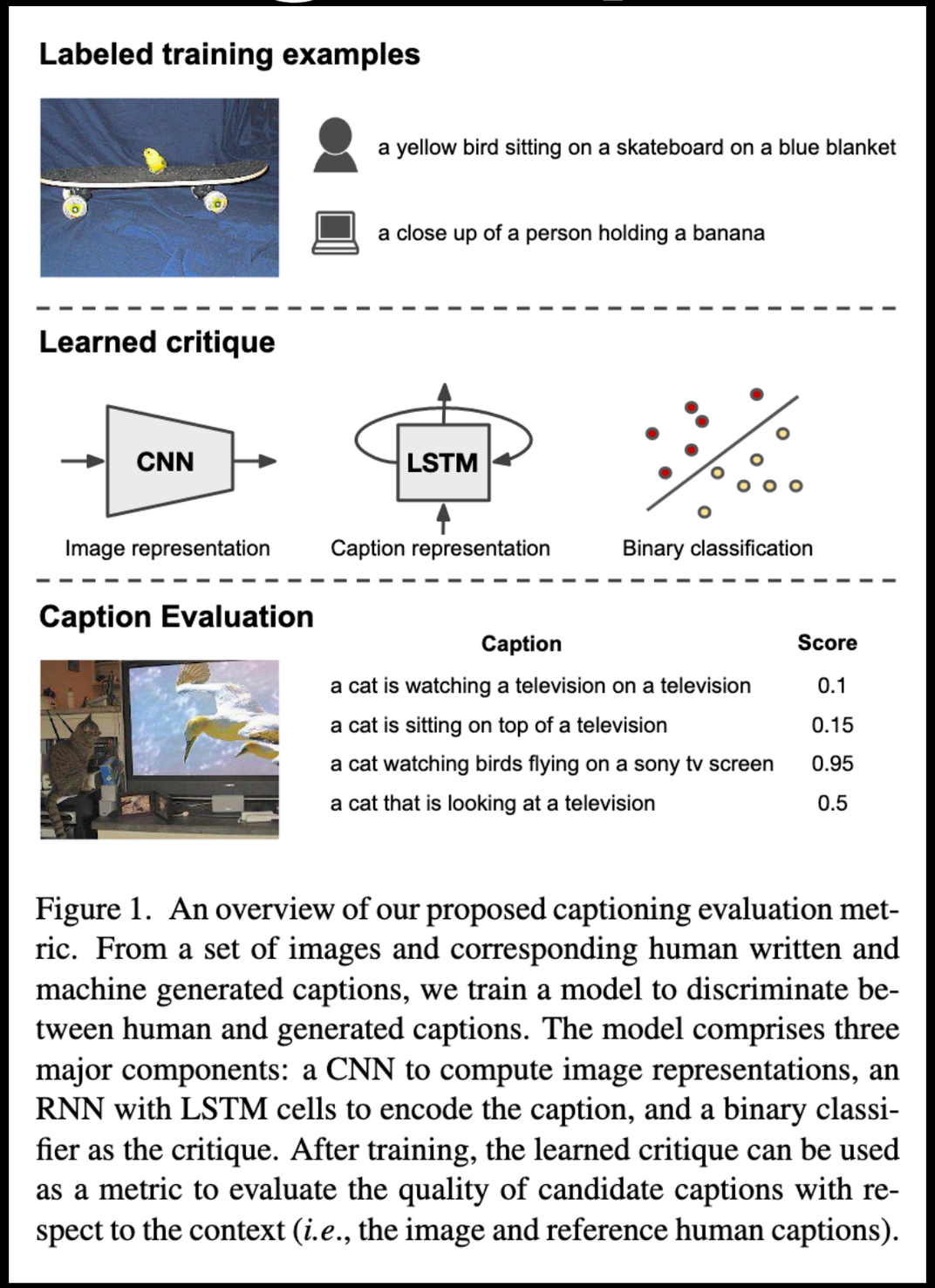


Figure 1. An overview of our proposed captioning evaluation metric. From a set of images and corresponding human written and machine generated captions, we train a model to discriminate between human and generated captions. The model comprises three major components: a CNN to compute image representations, an RNN with LSTM cells to encode the caption, and a binary classifier as the critique. After training, the learned critique can be used as a metric to evaluate the quality of candidate captions with respect to the context (*i.e.*, the image and reference human captions).

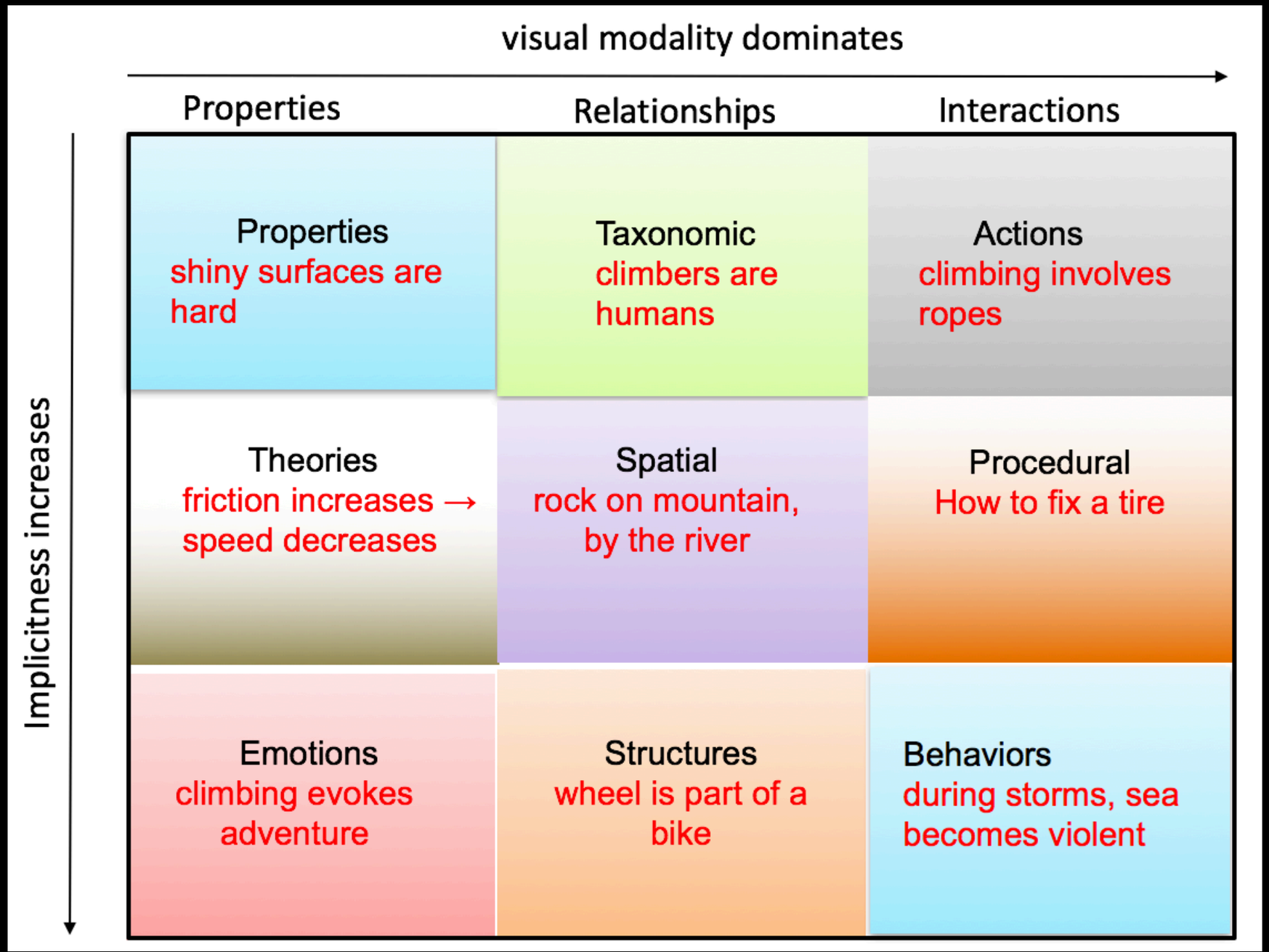
[2.1]: Cui et al 2018, Learning to Evaluate Image Captioning

Part I: CSK compensates for limited training data

Part II: Explanations and CSK

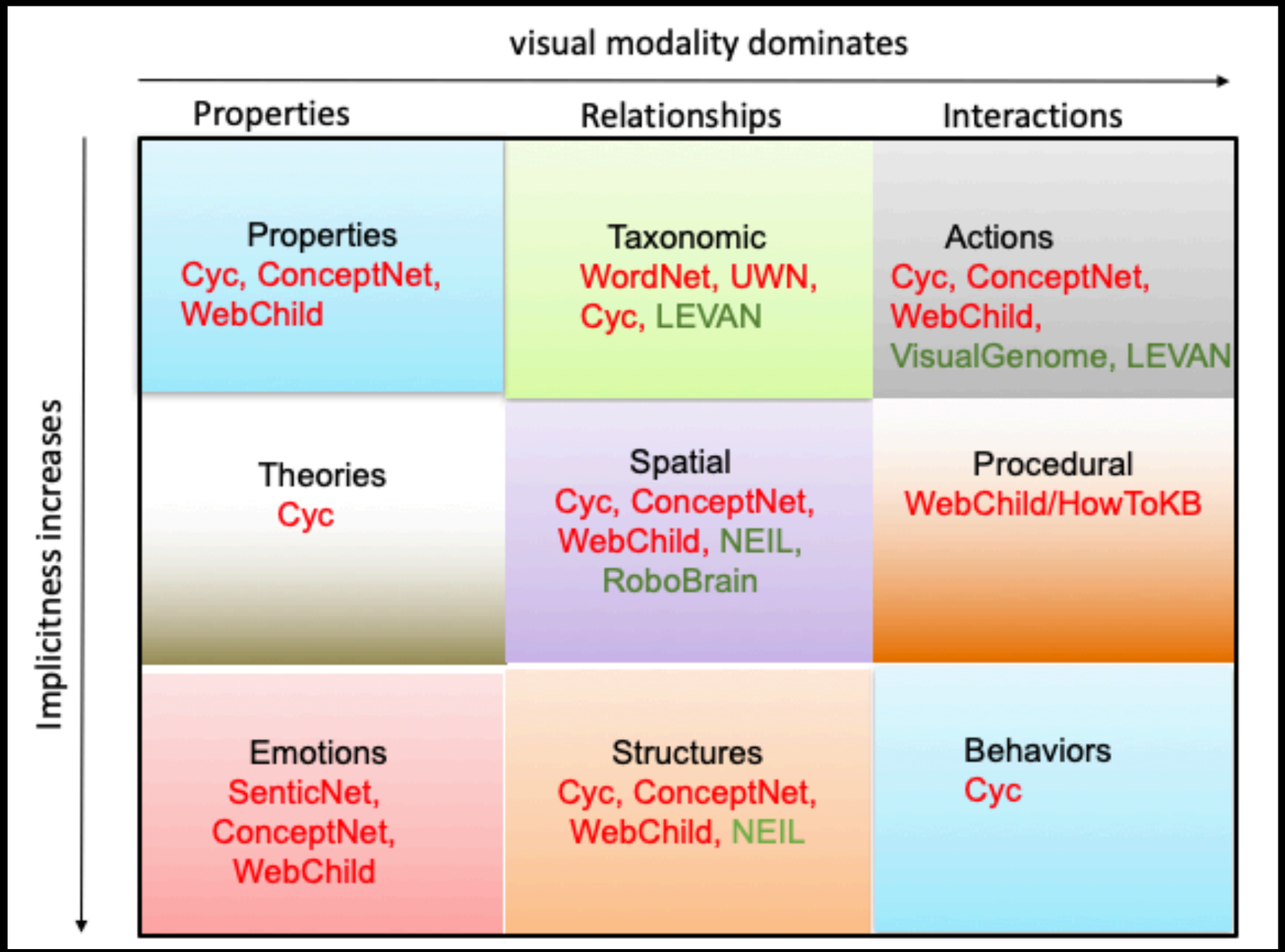
Part III: DL friendly commonsense representations

Some categories of CSK



[3.6] Tandon et. al: Commonsense tutorial at CIKM 2017 (Inspired by Peter Clark's visualization)

What exists?

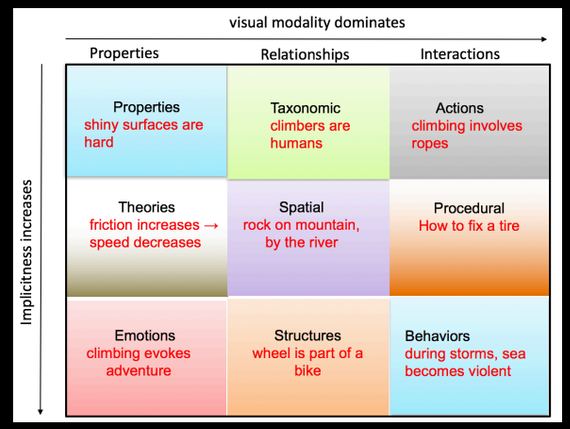

visual modality dominates
Properties
Relationships
Interactions
Implicitness increases
Properties
Cyc, ConceptNet, WebChild
Taxonomic
WordNet, UWN, Cyc, LEVAN
Actions
Cyc, ConceptNet, WebChild, VisualGenome, LEVAN
Theories
Cyc
Spatial
Cyc, ConceptNet, WebChild, NEIL, RoboBrain
Procedural
WebChild/HowToKB
Emotions
SenticNet, ConceptNet, WebChild
Structures
Cyc, ConceptNet, WebChild, NEIL
Behaviors
Cyc

What is needed?

- Challenges

- ACQUISITION:
  - Implicit
  - Reporting bias
  - Multimodal
  - Contextual

- REPRESENT:
  - Logical
  - Graph
  - Unstructured



In the context of DL	Pros	Cons
Logical	Structured and highly interpretable	Need to first connect the noisy space of data (q) to the perfect world of MT (kb). Not scalable, not crowdsourcable
Tuples or graph [3.1]	Structured	Lexical variations: q and kb Construction is noisy when scaling Task specific KBs more effective. Different models for KB and model
embedded tuples or embedded graph [3.2]	Reduces lexical variation problem	embedded graph can still be far from perfect.
adversarial setting [3.3]	directly use as training data	using kb to generate data is non-trivial and task specific
unstructured sentence or paragraph [3.4]	highly scalable, e.g. finding more paragraphs about a topic Same model "creates" KB	Topical retrieval of paragraphs. Reasoning becomes limited.



Part I: CSK compensates for limited training data



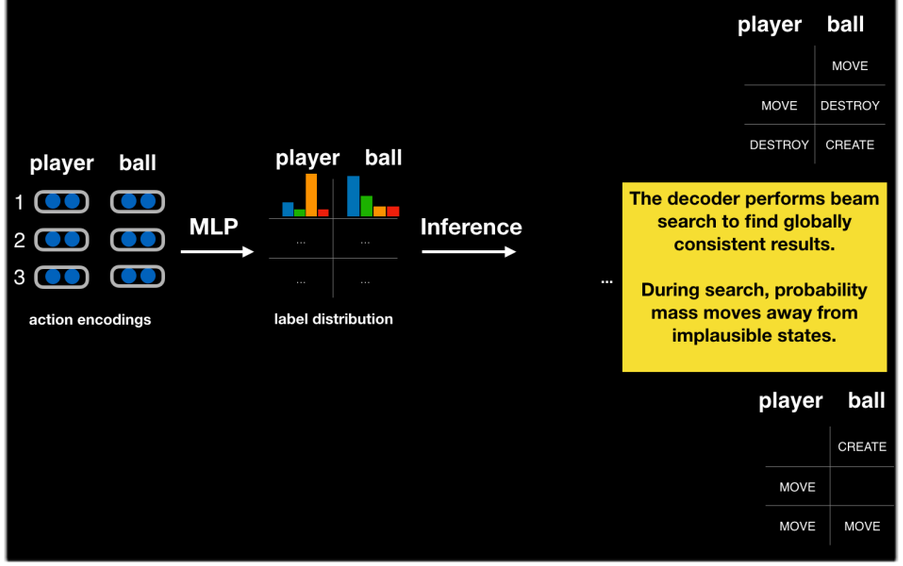
Part II: Explanations and CSK



Part III: DL friendly commonsense representations

Conclusion

Commonsense aware model



Explanation:

- new datasets
- attention + structure based model
- evaluation is challenging

In the context of DL	Pros	Cons
Logical	Structured and highly interpretable	Need to first connect the noisy space of data (q) to the perfect world of MT (kb). Not scalable, not crowdsourcable
Tuples or graph [3.1]	Structured	Lexical variations: q and kb Construction is noisy when scaling Task specific KBs more effective. Different models for KB and model
embedded tuples or embedded graph [3.2]	Reduces lexical variation problem	embedded graph can still be far from perfect.
adversarial setting [3.3]	directly use as training data	using kb to generate data is non-trivial and task specific
unstructured sentence or paragraph [3.4]	highly scalable, e.g. finding more paragraphs about a topic Same model "creates" KB	Topical retrieval of paragraphs. Reasoning becomes limited.

Commonsense for Deep Learning can help overcome challenges making models more robust and more amicable to explanation

Acknowledgements

Collaborators at AI2







Collaborators outside AI2





References (intro part)

- [1]: Shah et. al 2019, Cycle-Consistency for Robust Visual Question Answering

- [2]: Deep neural networks are easily fooled: evolvingai.org/fooling

- [3]: Kang et. al 2018, AdvEntuRe: Adversarial Training for Textual Entailment with Knowledge-Guided Examples

- [4] Peter Clark, Elementary School Science and Math Tests as a Driver for AI: Take the Aristo Challenge!

References (Part I)

- [1.1] Dalvi, Tandon, Clark: ProPara task: data.allenai.org/propara
- [1.2]: Tandon et. al 2018, Reasoning about Actions and State Changes by Injecting Commonsense Knowledge
- [D1]: Antolet et. al 2015, VQA: Visual Question Answering.
- [D2]: Johnson et. al 2016, CLEVR: A Diagnostic Dataset for Compositional Lang. and Elementary Visual Reasoning
- [D3]: Zeller et. al 2019, From Recognition to Cognition: Visual Commonsense Reasoning (VCR)
- [D4]: Camburu et. al 2018L e-SNLI: Natural Language Inference with Natural Language Explanations
- [D5]: Yang et. al , 2018: HOTPOTQA: A Dataset for Diverse, Explainable Multi-hop Question Answering
- [D6]: Jansen et. al , 2018, WorldTree: A Corpus of Explanation Graphs for Elementary Science Questions supporting Multi-Hop Inference
- [A2]: Kim et. al , 2018 Textual Explanations for Self-Driving Vehicles
- [A9]: Ghosh et. al 2018, Generating Natural Lang. Explanations for VQA Using Scene Graphs and Visual Attention
- [A4]: Zeiler 2013, Visualizing and Understanding Convolutional Networks
- [A8]: Park et al 2017, Multimodal Explanations: Justifying Decisions and Pointing to the Evidence
- [A3]: Hu et. al, 2017, Learning to Reason: End-to-End Module Networks for Visual Question Answering
- [A1]: Ramaprasaath 2016, Why did you say that? visual explanations from deep networks via gradient-based localization
- [A6]: Tandon et. al 2018, Reasoning about Actions and State Changes by Injecting Commonsense Knowledge

References (Part II)

- [2.1] Cui et al 2018, Learning to Evaluate Image Captioning

- [2.2] Petsuik et. al , 2018 , RISE: Randomized Input Sampling for Explanation of Black-box Models

References (Part III)

- [3.1] : Tandon et. al. 2018, Reasoning about Actions and State Changes by Injecting Commonsense Knowledge.

- [3.2] : (work under submission, will update once accepted)

- [3.3] : Kang et. al 2018, AdvEntuRe: Adversarial Training for Textual Entailment with Knowledge-Guided Examples.

- [3.4] : Du et. al 2019, Be Consistent! Improving Procedural Text Comprehension using Label Consistency