Neuro-Symbolic Embodied AI

Chuang Gan

MIT-IBM Watson AI Lab

My 15-Month-Old Daughter (Carolyn)

Recognize
objects and substances

Plate

Hole

Reason from
abstract knowledge

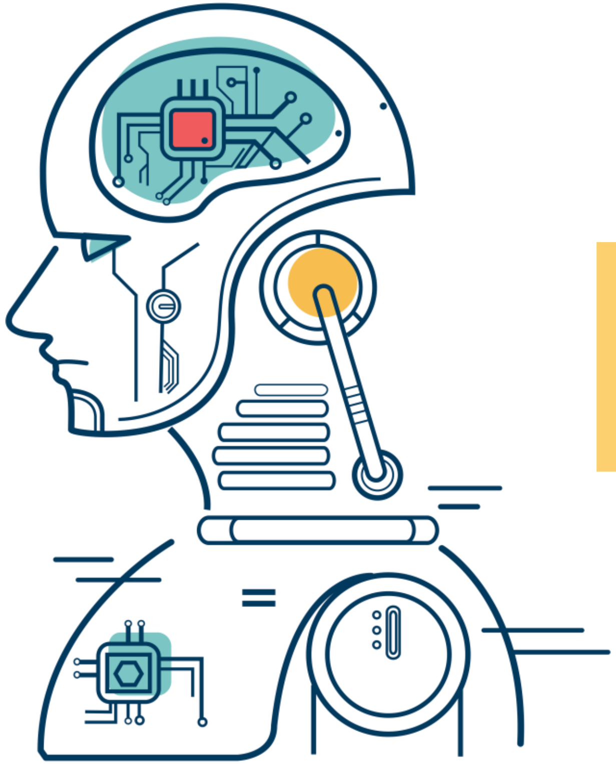
Fit

Not Fit

Use world models for
inference and planning

Try to put the
plate into the
correct hole

How About Machine Intelligence Today?



Machine

Does machine have the intelligence at the the level of Carolyn?

Carolyn

Most Visual Intelligence Now

Recognize
objects and substances

Plate

Hole

Lack Common Sense

Reason from
abstract knowledge

Fit

Not Fit

Use world models for
inference and planning

Try to put the
plate into the
correct hole

Symbolic Program

Recognize
objects and substances

Neural Network

Lack Common Sense

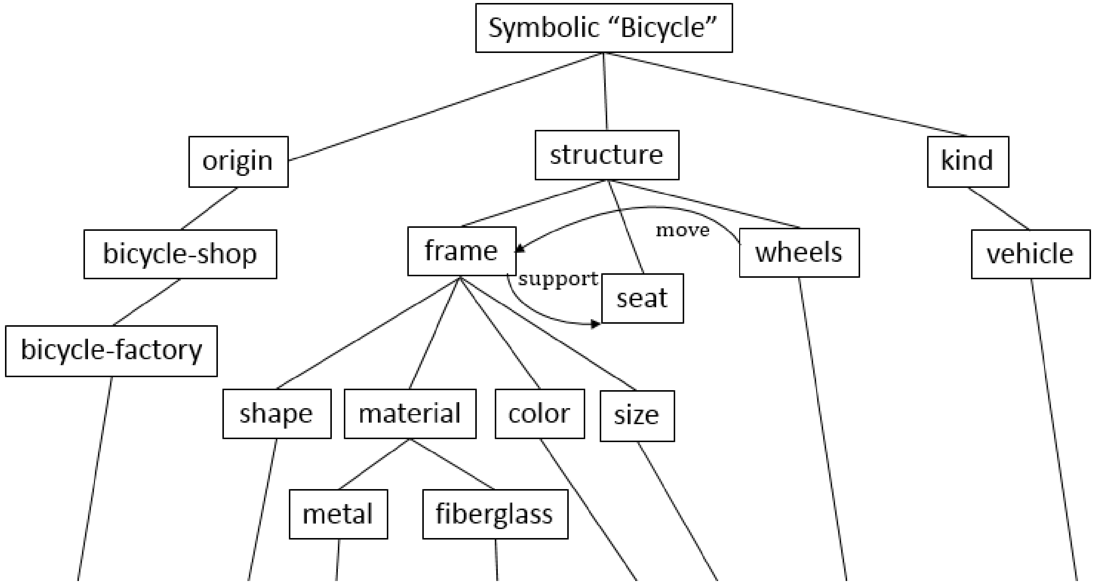
Reason from
abstract knowledge

Symbolic Program

Use world models for
inference and planning

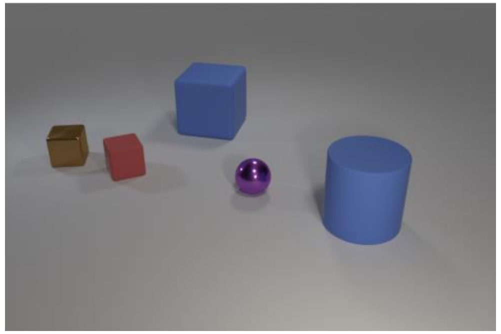
Try to put the plate into the correct hole

Related Work: Symbolic AI



A symbolic representation of bicycle [1]

Q: How many cubes are behind the cylinder?
A: 3



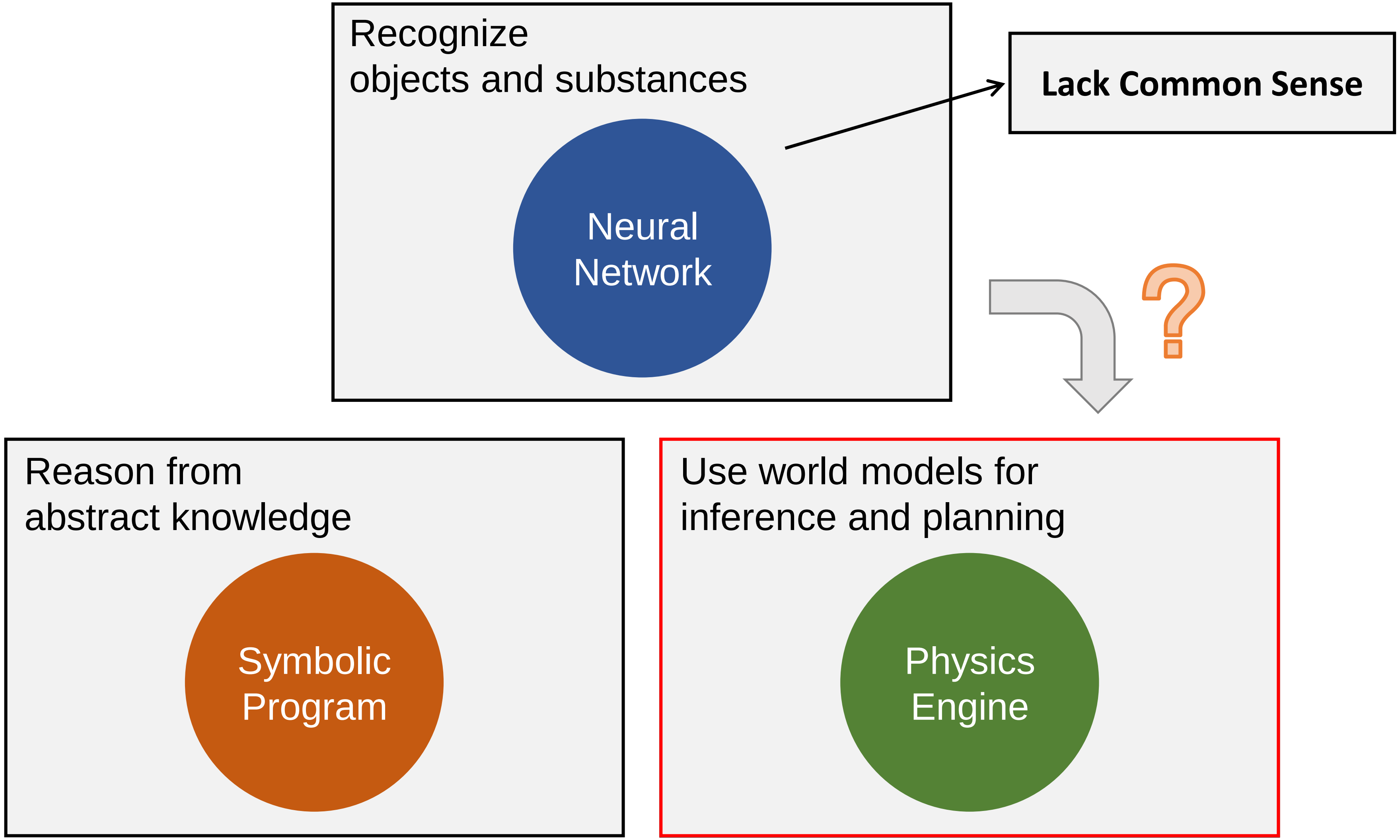
1. filter_shape(scene, cylinder)
2. relate(behind)
3. filter_shape(scene, cube)
4. count(scene)

Program

A symbolic program of CLEVR [2]

[1] Minsky, M., & Winston, P. H. (1990). Logical vs. Analogical or Symbolic vs. Connectionist or Neat vs. Scruffy” and “Excerpts from the Society of Mind. In Artificial Intelligence at MIT, Expanding Frontiers (Vol. 1).
[2] Johnson J, Hariharan B, Van Der Maaten L, et al. Clevr: A diagnostic dataset for compositional language and elementary visual reasoning. Proceedings of the IEEE conference on computer vision and pattern recognition.

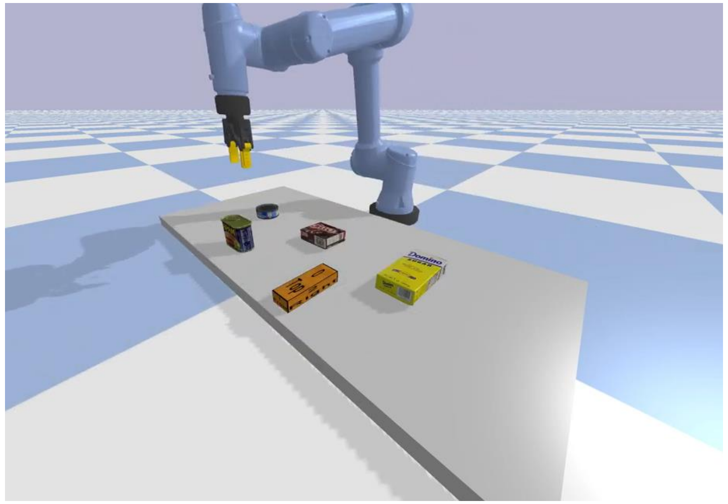
Physics Engines



Related Work: Physics Engines

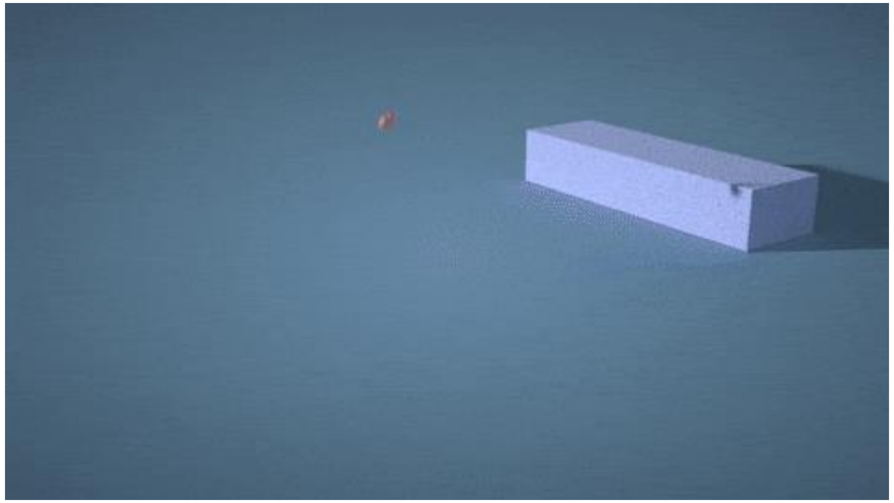
State s_t → **Physics Engine** → State s_{t+1} → **Physics Engine** → State s_{t+2}

PyBullet



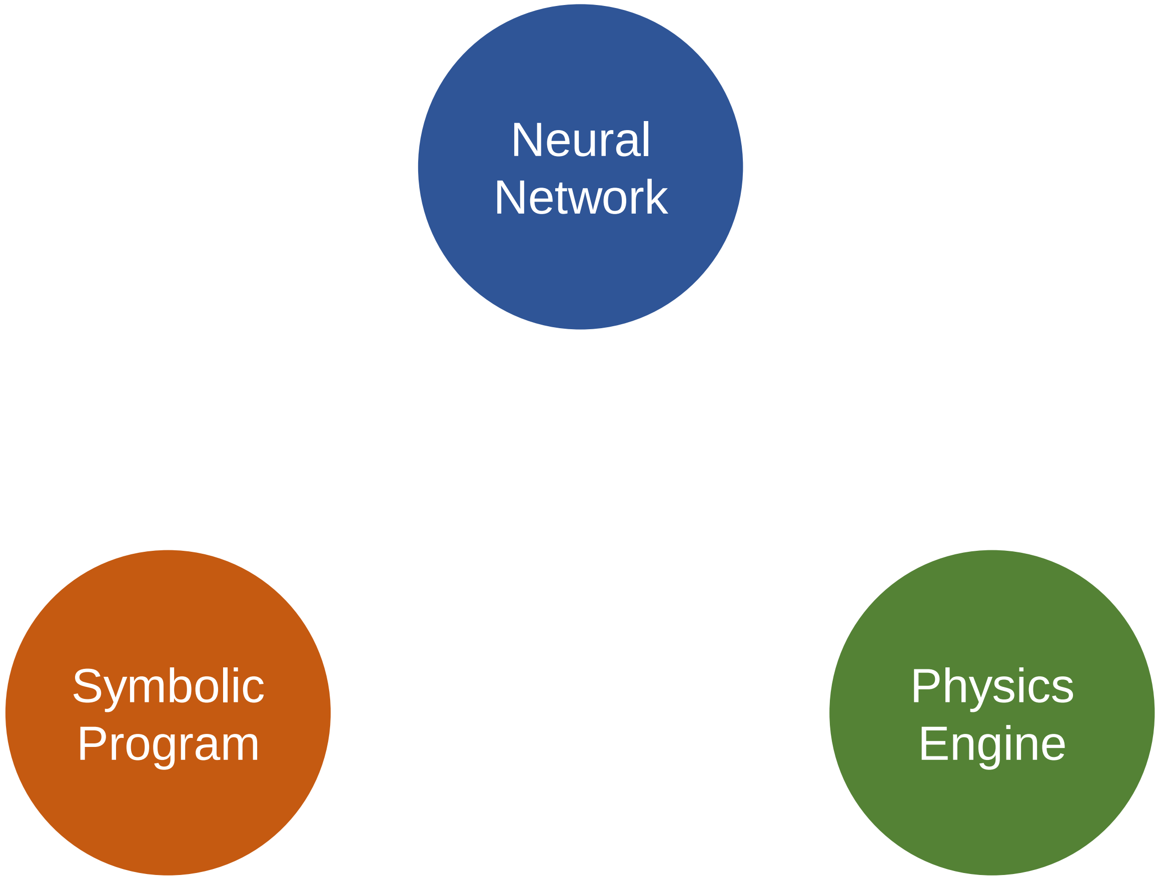
NVIDIA-Flex

Taichi Programming Language



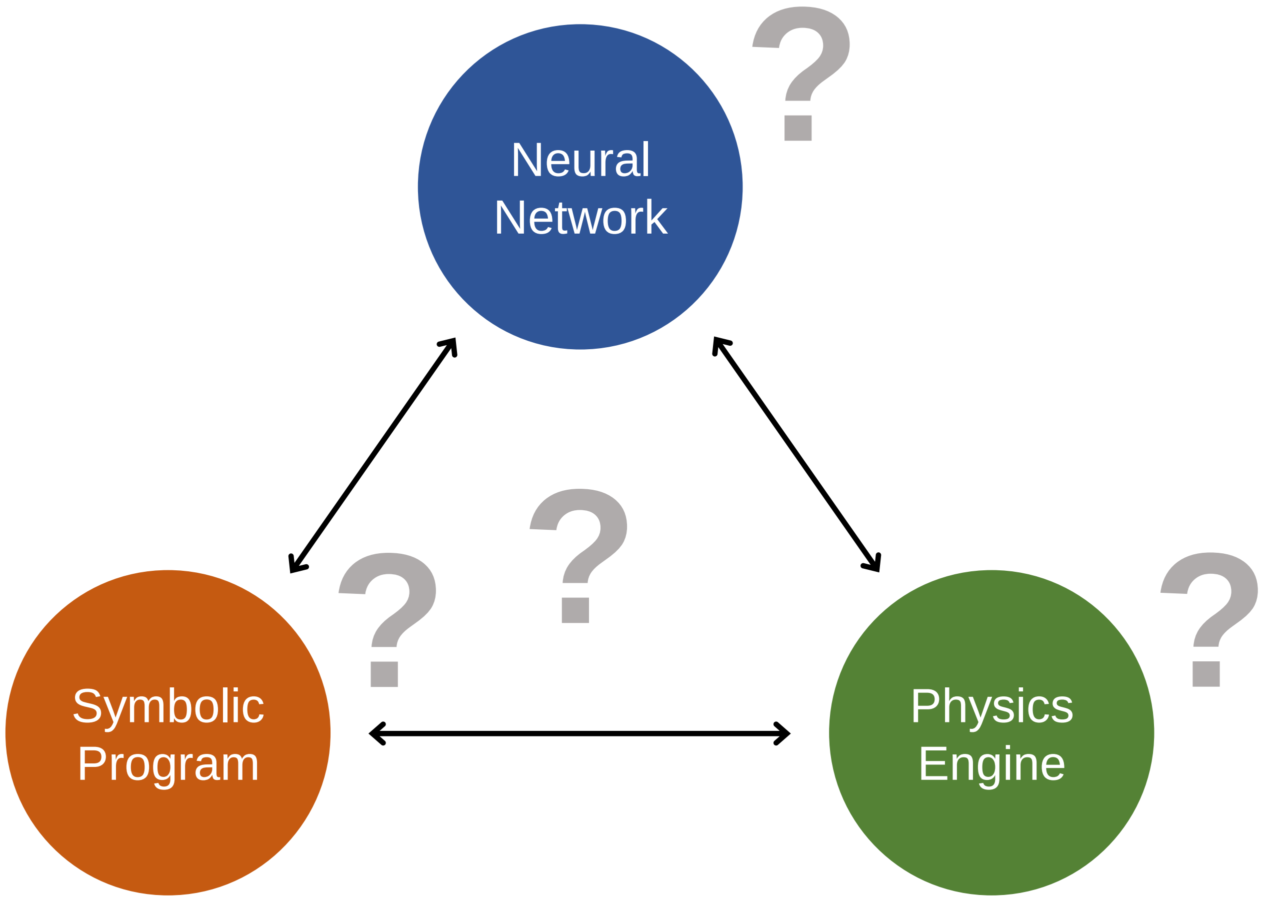
PyBullet: https://www.youtube.com/watch?v=8e-KjBUakqY
NVIDIA-Flex: https://youtu.be/1o0Nuq71gI4
Taichi: https://github.com/taichi-dev/quantaichi

How to Integrate Them to Advance Visual Intelligence?



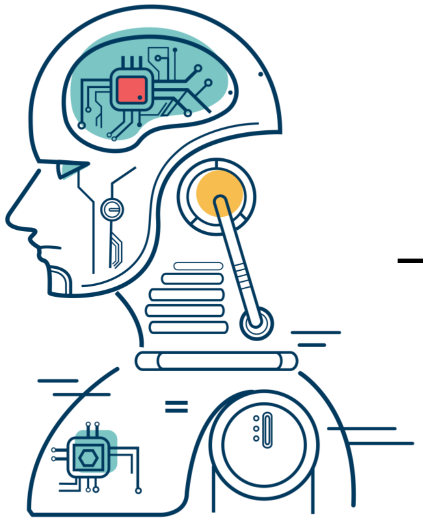
How to Integrate Them to Advance Visual Intelligence?

What is the role of each one?

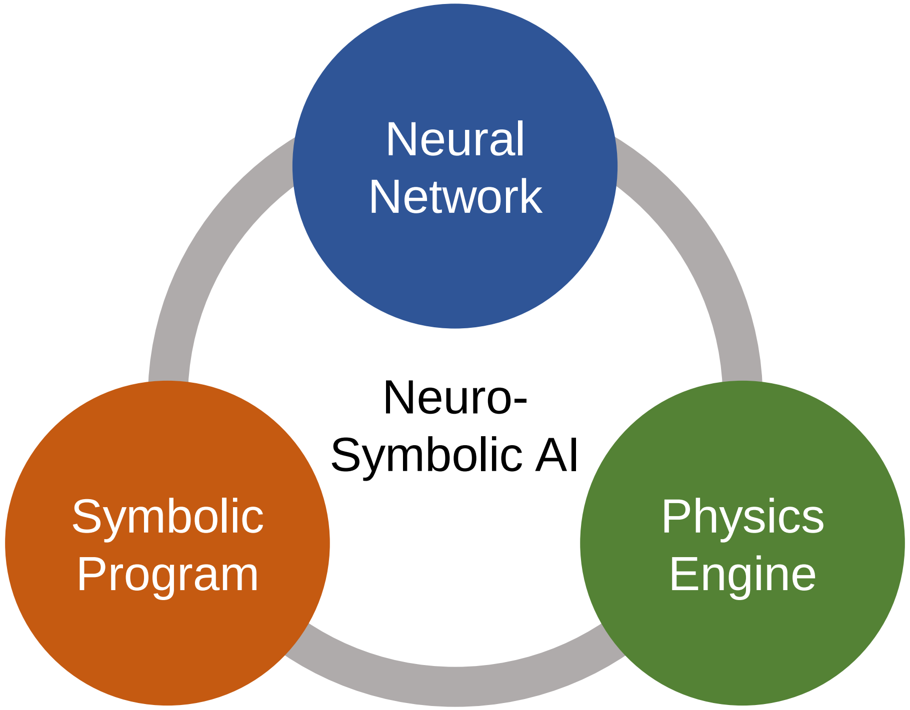


How to combine each strength?

My Research: Neuro-Symbolic AI for Visual Intelligence



Machine

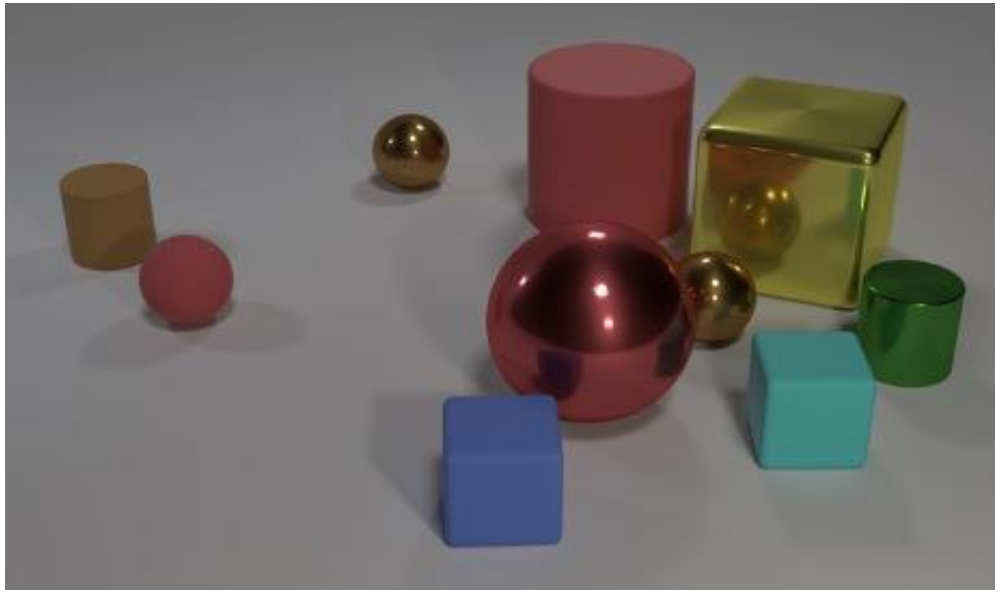


Carolyn

Task: Visual Reasoning

Q: What color is the fire hydrant?

A: Yellow

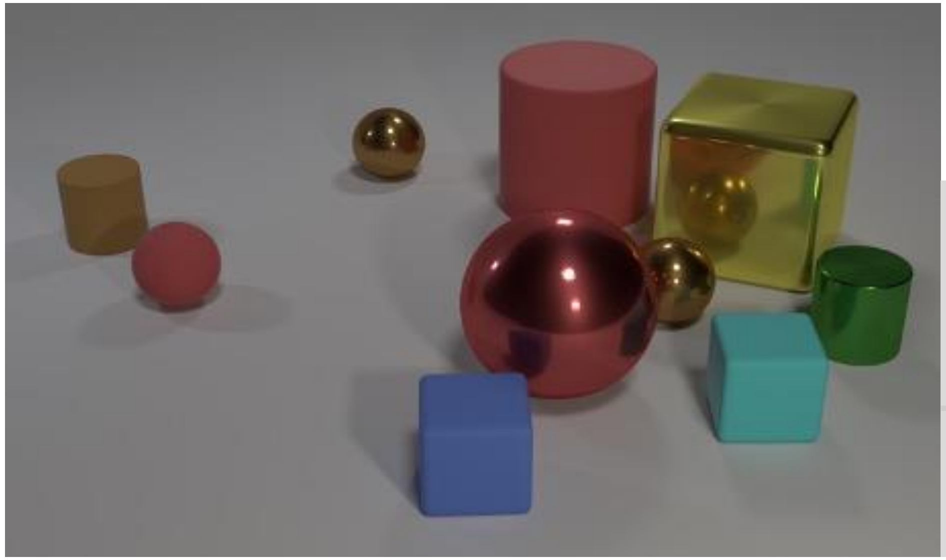


Q: Are there an equal number of large things and metal spheres?

A: Yes

Agrawal et al. VQA, 2015. Johnson et al. CLEVR 2017.

Prior Work: End-to-End Visual Reasoning



Visual Question Answering

Q: ***Are there an equal number of large things and metal spheres?***

Concept
(e.g., colors, shapes)

Entangled

Reasoning
(e.g., count)

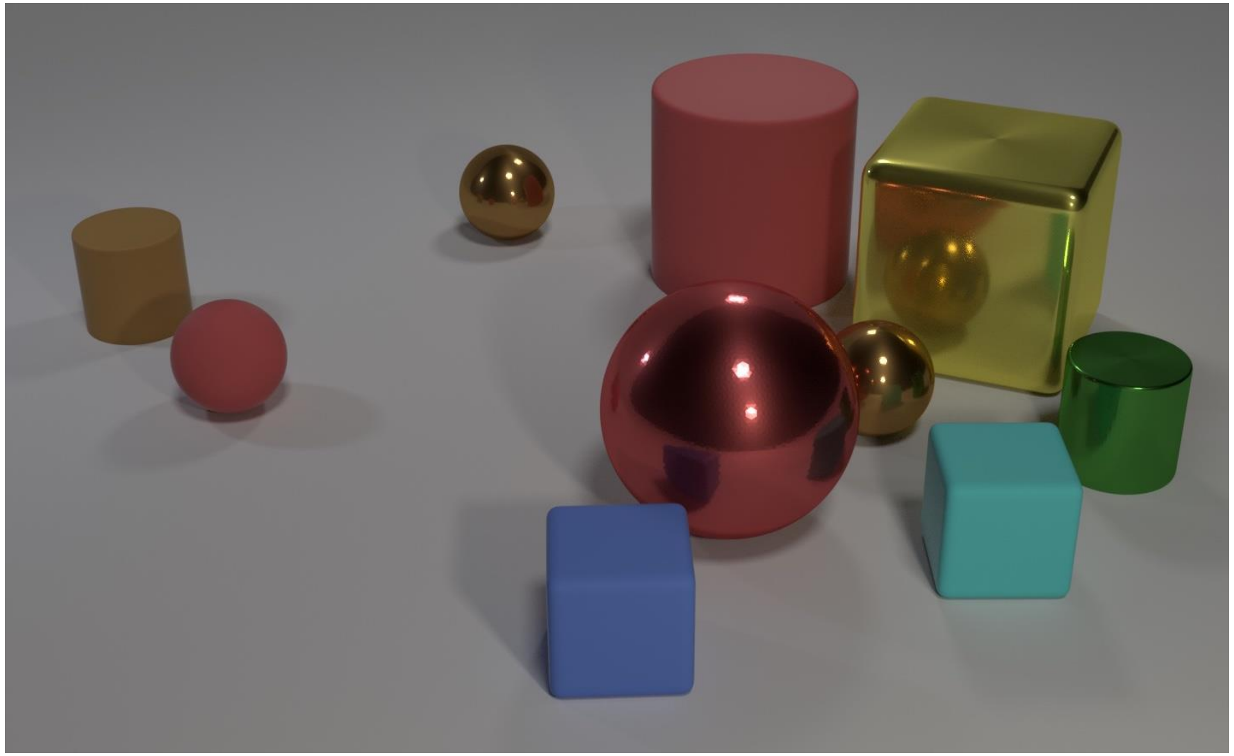
Hard to transfer

Image Captioning

Instance Retrieval

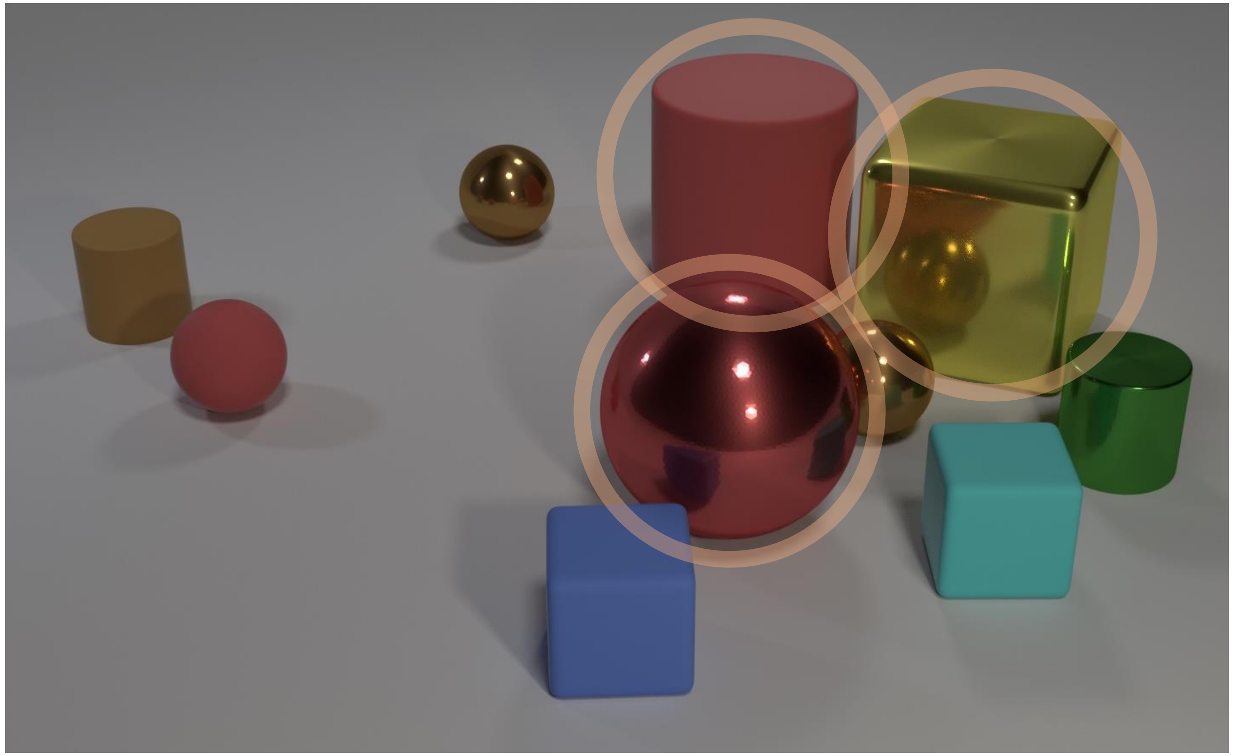
Agrawal et al. VQA, 2015. Johnson et al. CLEVR 2017.
Johnson et al. CLEVR 2017 , Andreas et al. NMN, 2016. Johnson et al. IEP, 2017. Perez et al. FiLM, 2018. Hudson & Manning. MAC, 2018. Hu et al. Stack-NMN, 2018. Mascharka et al. TbD, 2018.

How Do Human Reason From A Visual Scene?

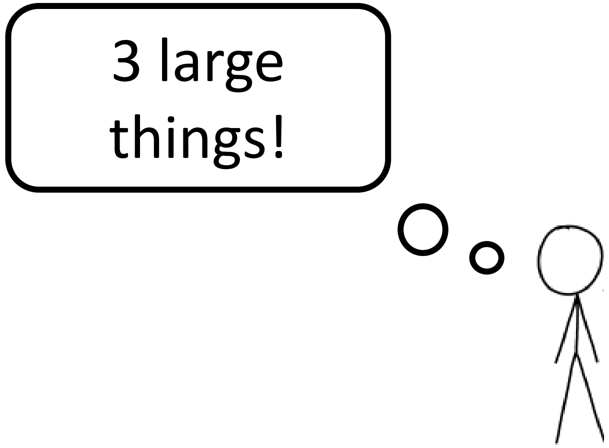


Question: *Are there an equal number of large things and metal spheres?*

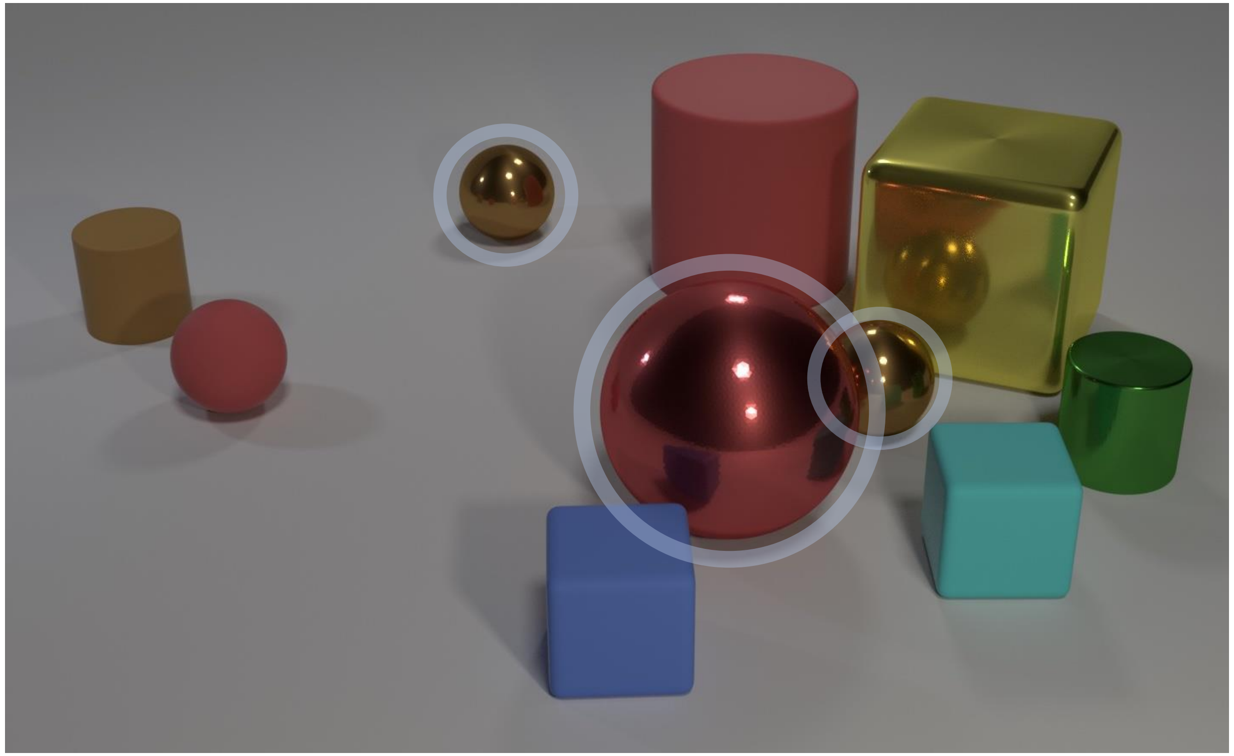
How Do Human Reason From A Visual Scene?



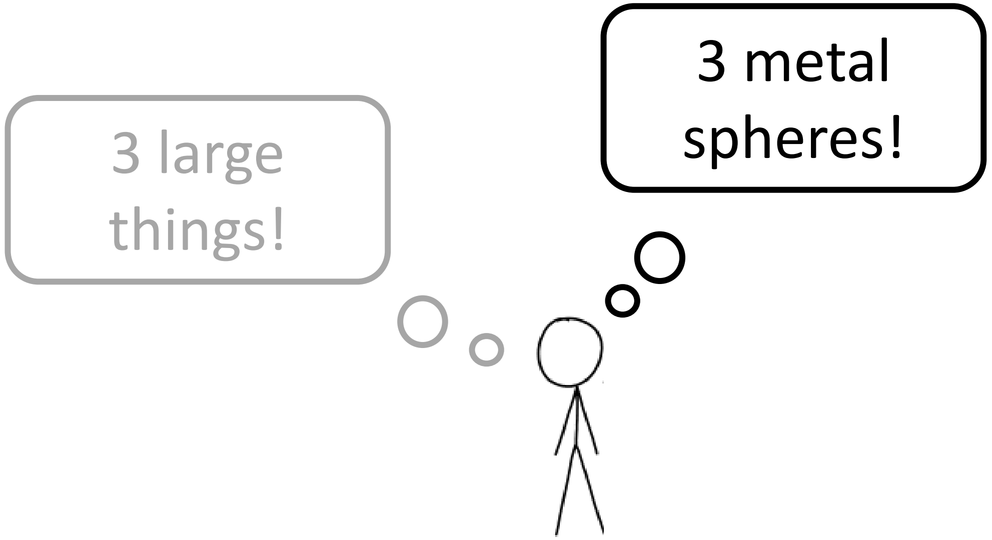
Question: *Are there an equal number of large things and metal spheres?*



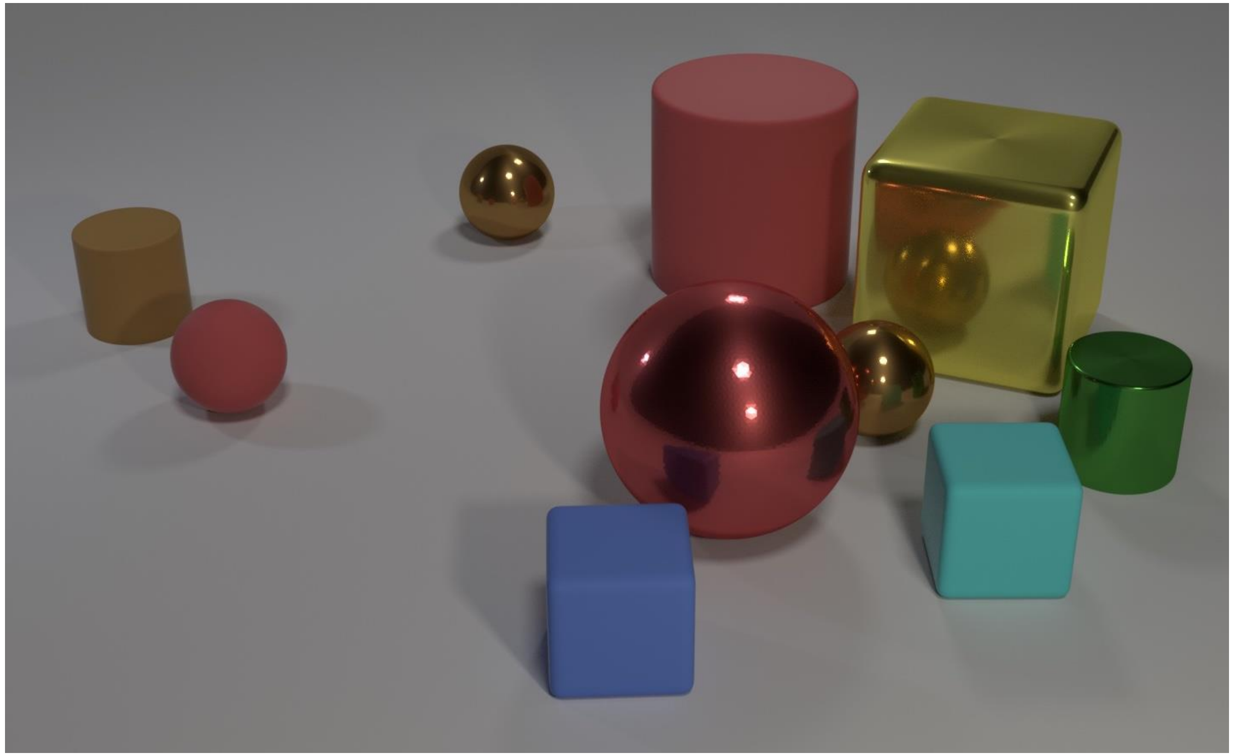
How Do Human Reason From A Visual Scene?



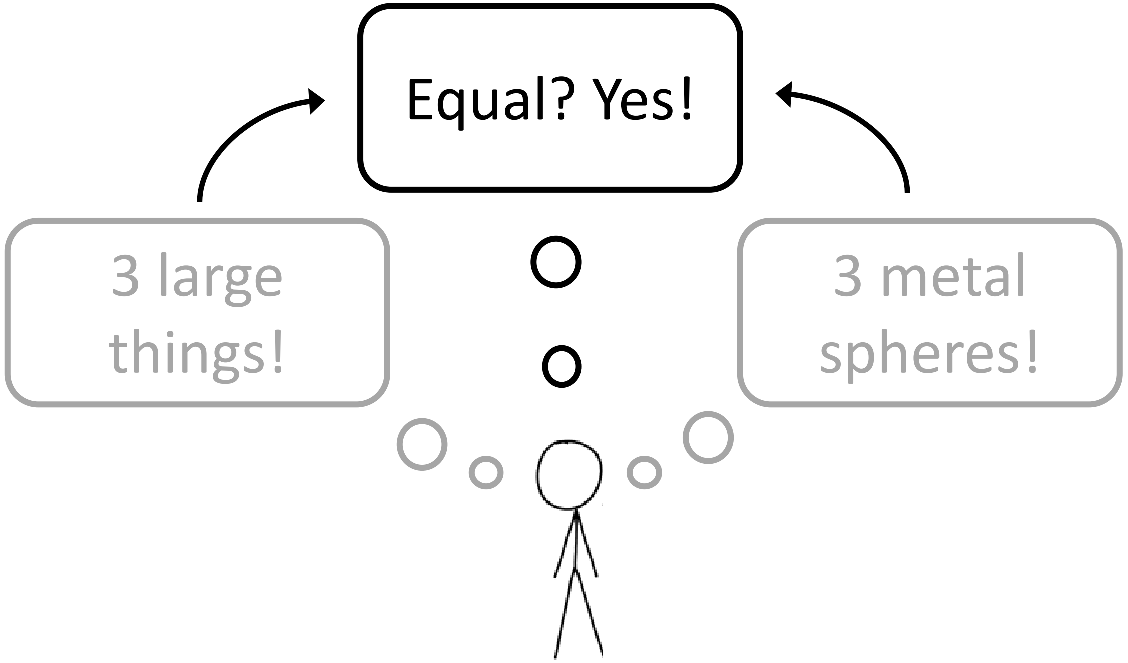
Question: *Are there an equal number of large things and metal spheres?*



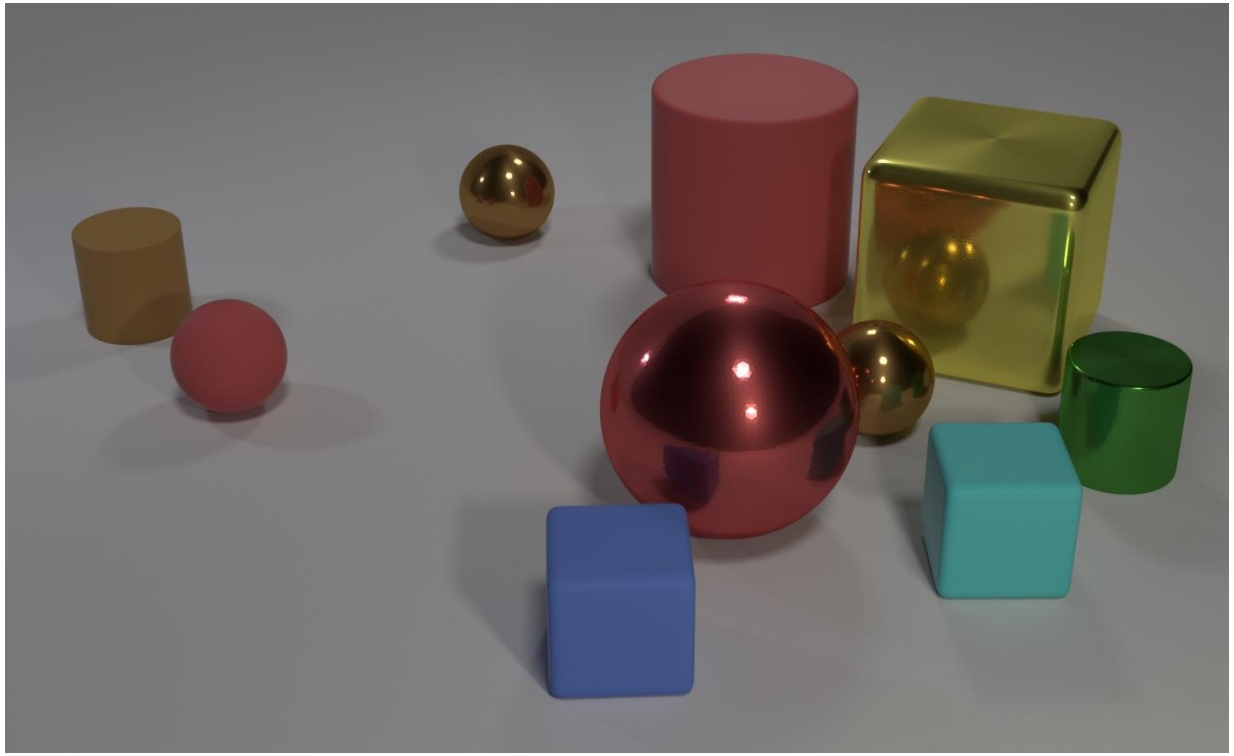
How do human reason from a visual scene?



Question: *Are there an equal number of large things and metal spheres?*



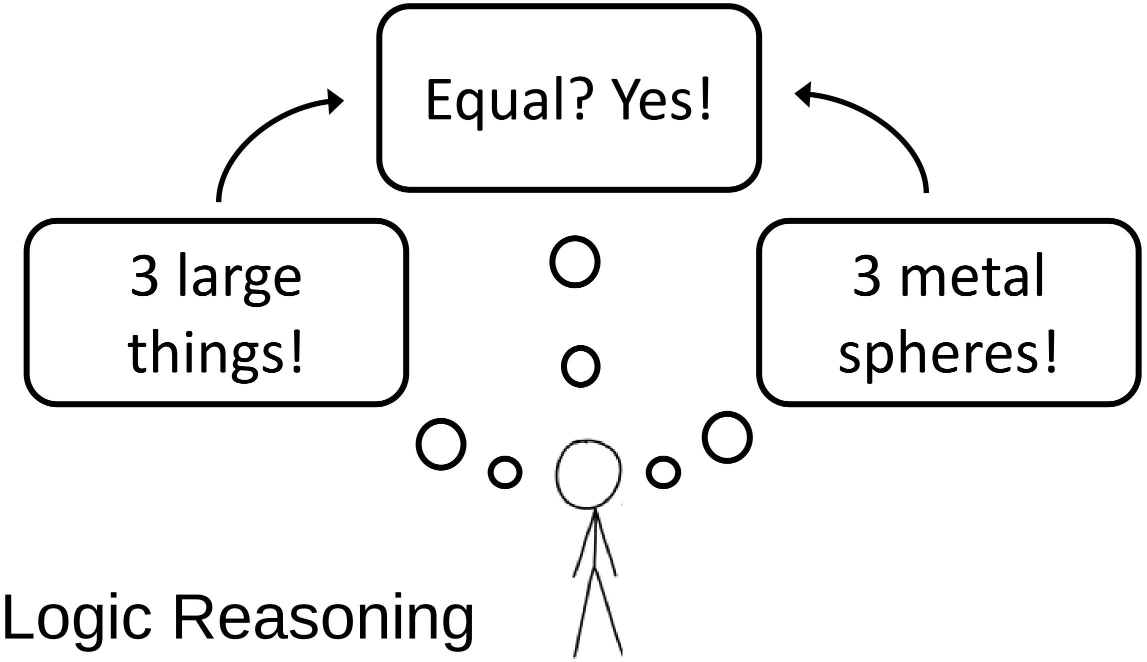
How Do Human Reason From A Visual Scene?



Visual Perception

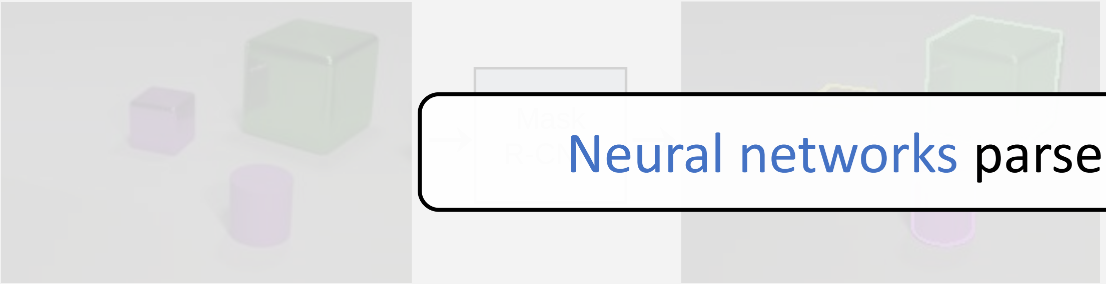
Question Understanding

Question: *Are there an equal number of large things and metal spheres?*



Logic Reasoning

Incorporate Concepts and Symbolic Programs



Neural networks parse images in symbolic concepts

I. *Neural* Scene Parsing

II. *Neural* Question Parsing

Neural networks parse questions into symbolic functional program.

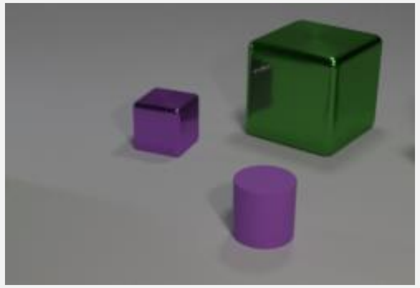
III. *Symbolic* Program Execution

Executing programs on the symbolic space.

VQS: Linking Segmentations to Questions and Answers for VQA. **Gan** et al. ICCV'17

Neural-Symbolic VQA: Disentangling Reasoning from Vision and Language Understanding. Yi, Wu, **Gan**, et al. NeurIPS'18

Neural Scene Parsing



I. *Neural* Scene Parsing

II. *Neural* Question Parsing

How many cubes that are behind the cylinder are large? →

LSTM Encoder | LSTM Decoder

→ 1. filter_shape(scene, cylinder)

→ 2. relate(behind)

→ 3. filter_shape(scene, cube)

→ 4. filter_size(scene, large)

→ 5. count(scene)

III. *Symbolic* Program Execution

1. filter_cylinder
2. relate_behind

ID	Size	Shape	…
1	Small	Cube	…
2	Small	Cylinder	…
3	Large	Cube	…

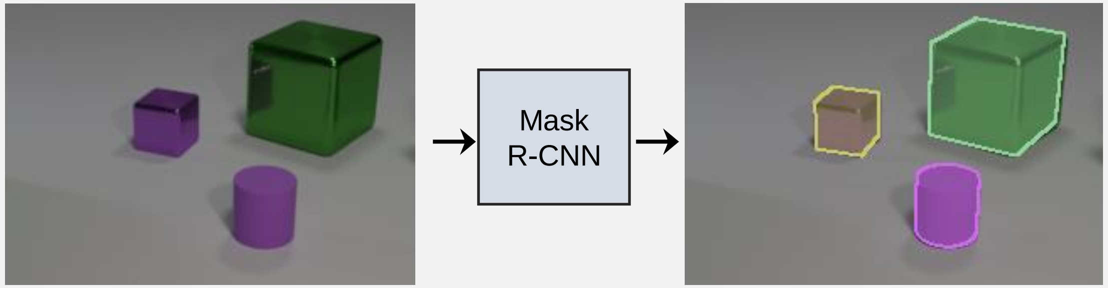
3. filter_cube
4. filter_large

ID	Size	…
3	Large	…

5. count

Answer: 1

Neural Scene Parsing



I. *Neural* Scene Parsing

II. *Neural* Question Parsing

How many cubes that are behind the cylinder are large? → LSTM Encoder | LSTM Decoder

→ 1. filter_shape(scene, cylinder)
→ 2. relate(behind)
→ 3. filter_shape(scene, cube)
→ 4. filter_size(scene, large)
→ 5. count(scene)

III. *Symbolic* Program Execution

1. filter_cylinder
2. relate_behind

ID	Size	Shape	…
1	Small	Cube	…
2	Small	Cylinder	…
3	Large	Cube	…

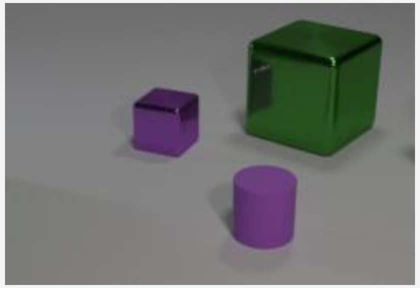
3. filter_cube
4. filter_large

ID	Size	…
3	Large	…

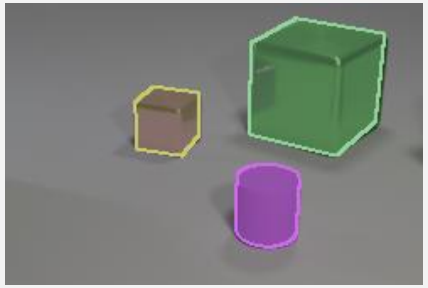
5. count

Answer: 1

Neural Scene Parsing



→ Mask R-CNN →



→ CNN →

ID	Size	Shape	Material	Color	x	y	z
1	Small	Cube	Metal	Purple	-0.45	-1.10	0.35
2	Small	Cylinder	Rubber	Purple	0.75	1.31	0.35
3	Large	Cube	Metal	Green	1.58	-1.60	0.70

I. *Neural* Scene Parsing

II. *Neural* Question Parsing

How many cubes that are behind the cylinder are large? → LSTM Encoder | LSTM Decoder

→ 1. filter_shape(scene, cylinder)
→ 2. relate(behind)
→ 3. filter_shape(scene, cube)
→ 4. filter_size(scene, large)
→ 5. count(scene)

III. *Symbolic* Program Execution

1. filter_cylinder
2. relate_behind

ID	Size	Shape	…
1	Small	Cube	…
2	Small	Cylinder	…
3	Large	Cube	…

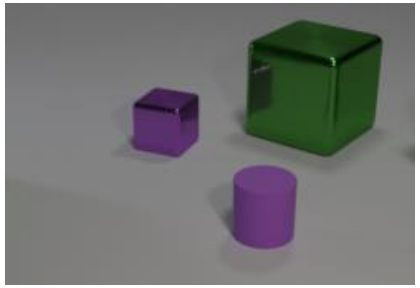
3. filter_cube
4. filter_large

ID	Size	…
3	Large	…

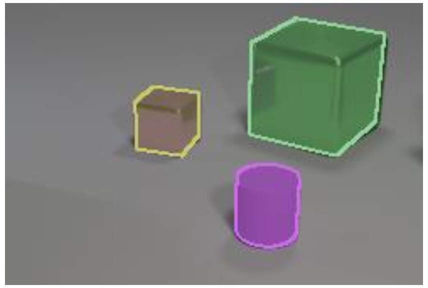
5. count

Answer: 1

Neural Question Parsing



→ Mask R-CNN →



→ CNN →

ID	Size	Shape	Material	Color	x	y	z
1	Small	Cube	Metal	Purple	-0.45	-1.10	0.35
2	Small	Cylinder	Rubber	Purple	0.75	1.31	0.35
3	Large	Cube	Metal	Green	1.58	-1.60	0.70

I. *Neural* Scene Parsing

II. *Neural* Question Parsing

How many cubes that are behind the cylinder are large? → LSTM Encoder | LSTM Decoder

→ 1. filter_shape(scene, cylinder)
→ 2. relate(behind)
→ 3. filter_shape(scene, cube)
→ 4. filter_size(scene, large)
→ 5. count(scene)

III. *Symbolic* Program Execution

1. filter_cylinder
2. relate_behind

ID	Size	Shape	…
1	Small	Cube	…
2	Small	Cylinder	…
3	Large	Cube	…

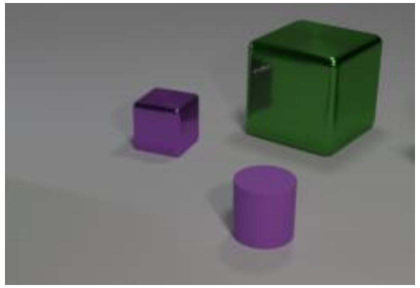
3. filter_cube
4. filter_large

ID	Size	…
3	Large	…

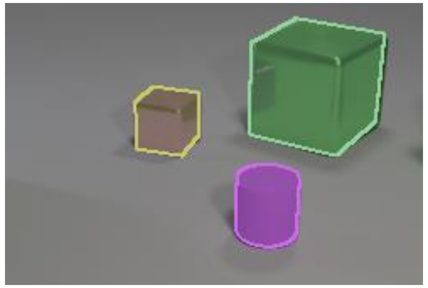
5. count

Answer: 1

Symbolic Reasoning



→ Mask R-CNN →



→ CNN →

ID	Size	Shape	Material	Color	x	y	z
1	Small	Cube	Metal	Purple	-0.45	-1.10	0.35
2	Small	Cylinder	Rubber	Purple	0.75	1.31	0.35
3	Large	Cube	Metal	Green	1.58	-1.60	0.70

I. *Neural* Scene Parsing

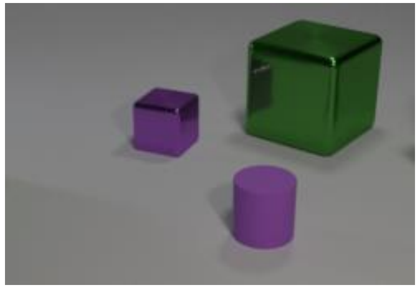
II. *Neural* Question Parsing

How many cubes that are behind the cylinder are large? → LSTM Encoder | LSTM Decoder

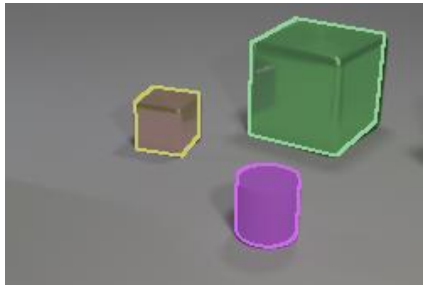
→ 1. filter_shape(scene, cylinder)
→ 2. relate(behind)
→ 3. filter_shape(scene, cube)
→ 4. filter_size(scene, large)
→ 5. count(scene)

III. *Symbolic* Program Execution

Symbolic Reasoning



→ Mask R-CNN →



→ CNN →

ID	Size	Shape	Material	Color	x	y	z
1	Small	Cube	Metal	Purple	-0.45	-1.10	0.35
2	Small	Cylinder	Rubber	Purple	0.75	1.31	0.35
3	Large	Cube	Metal	Green	1.58	-1.60	0.70

I. *Neural* Scene Parsing

II. *Neural* Question Parsing

How many cubes that are behind the cylinder are large? → LSTM Encoder | LSTM Decoder

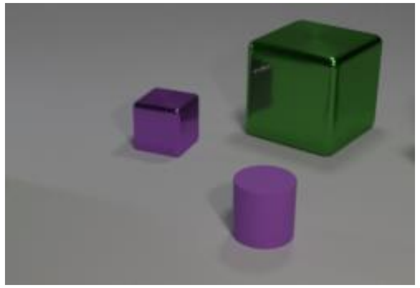
→ 1. filter_shape(scene, cylinder)
→ 2. relate(behind)
→ 3. filter_shape(scene, cube)
→ 4. filter_size(scene, large)
→ 5. count(scene)

→

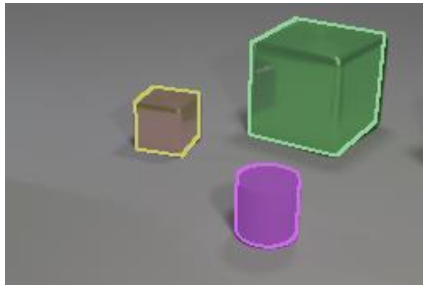
III. *Symbolic* Program Execution

1. filter_cylinder

Symbolic Reasoning



→ Mask R-CNN →



→ CNN →

ID	Size	Shape	Material	Color	x	y	z
1	Small	Cube	Metal	Purple	-0.45	-1.10	0.35
2	Small	Cylinder	Rubber	Purple	0.75	1.31	0.35
3	Large	Cube	Metal	Green	1.58	-1.60	0.70

I. *Neural* Scene Parsing

II. *Neural* Question Parsing

How many cubes that are behind the cylinder are large? → LSTM Encoder | LSTM Decoder

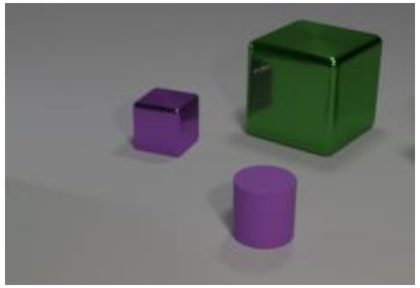
→ 1. filter_shape(scene, cylinder)
→ 2. relate(behind)
→ 3. filter_shape(scene, cube)
→ 4. filter_size(scene, large)
→ 5. count(scene)

III. *Symbolic* Program Execution

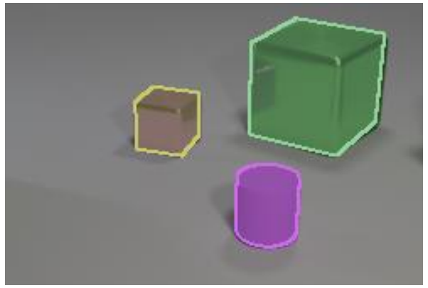
1. filter_cylinder
2. relate_behind

ID	Size	Shape	…
1	Small	Cube	…
3	Large	Cube	…

Symbolic Reasoning



→ Mask R-CNN →



→ CNN →

ID	Size	Shape	Material	Color	x	y	z
1	Small	Cube	Metal	Purple	-0.45	-1.10	0.35
2	Small	Cylinder	Rubber	Purple	0.75	1.31	0.35
3	Large	Cube	Metal	Green	1.58	-1.60	0.70

I. *Neural* Scene Parsing

II. *Neural* Question Parsing

How many cubes that are behind the cylinder are large? → LSTM Encoder | LSTM Decoder

→ 1. filter_shape(scene, cylinder)

→ 2. relate(behind)

→ 3. filter_shape(scene, cube)

→ 4. filter_size(scene, large)

→ 5. count(scene)

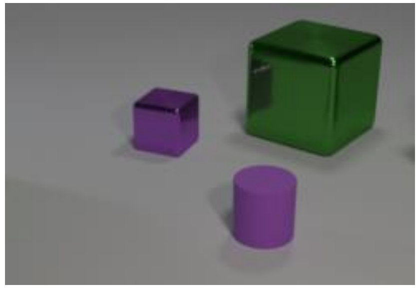
III. *Symbolic* Program Execution

1. filter_cylinder
2. relate_behind

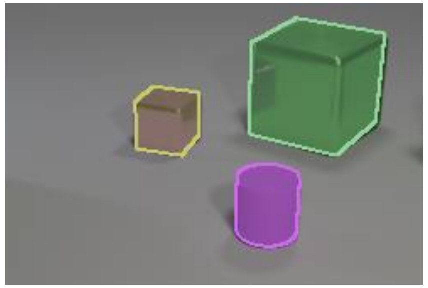
3. filter_cube

ID	Size	Shape	…
1	Small	Cube	…
3	Large	Cube	…

Symbolic Reasoning



→ Mask R-CNN →



→ CNN →

ID	Size	Shape	Material	Color	x	y	z
1	Small	Cube	Metal	Purple	-0.45	-1.10	0.35
2	Small	Cylinder	Rubber	Purple	0.75	1.31	0.35
3	Large	Cube	Metal	Green	1.58	-1.60	0.70

I. *Neural* Scene Parsing

II. *Neural* Question Parsing

How many cubes that are behind the cylinder are large? → LSTM Encoder | LSTM Decoder

→ 1. filter_shape(scene, cylinder)
→ 2. relate(behind)
→ 3. filter_shape(scene, cube)
→ 4. filter_size(scene, large)
→ 5. count(scene)

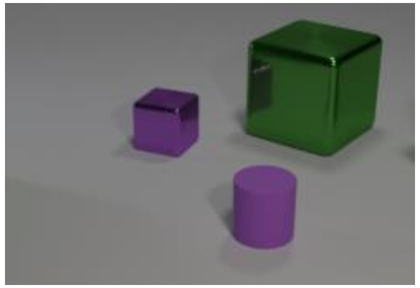
III. *Symbolic* Program Execution

1. filter_cylinder
2. relate_behind
3. filter_cube
4. filter_large

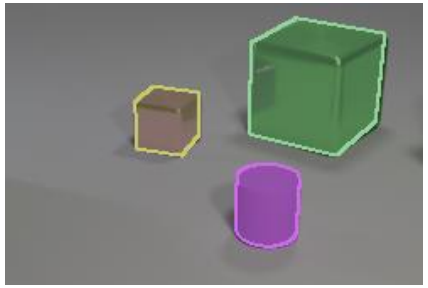
ID	Size	Shape	…
1	Small	Cube	…
3	Large	Cube	…

ID	Size	…
3	Large	…

Symbolic Reasoning



→ Mask R-CNN →



→ CNN →

ID	Size	Shape	Material	Color	x	y	z
1	Small	Cube	Metal	Purple	-0.45	-1.10	0.35
2	Small	Cylinder	Rubber	Purple	0.75	1.31	0.35
3	Large	Cube	Metal	Green	1.58	-1.60	0.70

I. *Neural* Scene Parsing

II. *Neural* Question Parsing

How many cubes that are behind the cylinder are large? → LSTM Encoder | LSTM Decoder

→ 1. filter_shape(scene, cylinder)
→ 2. relate(behind)
→ 3. filter_shape(scene, cube)
→ 4. filter_size(scene, large)
→ 5. count(scene)

III. *Symbolic* Program Execution

1. filter_cylinder
2. relate_behind

ID	Size	Shape	…
1	Small	Cube	…
3	Large	Cube	…

3. filter_cube
4. filter_large

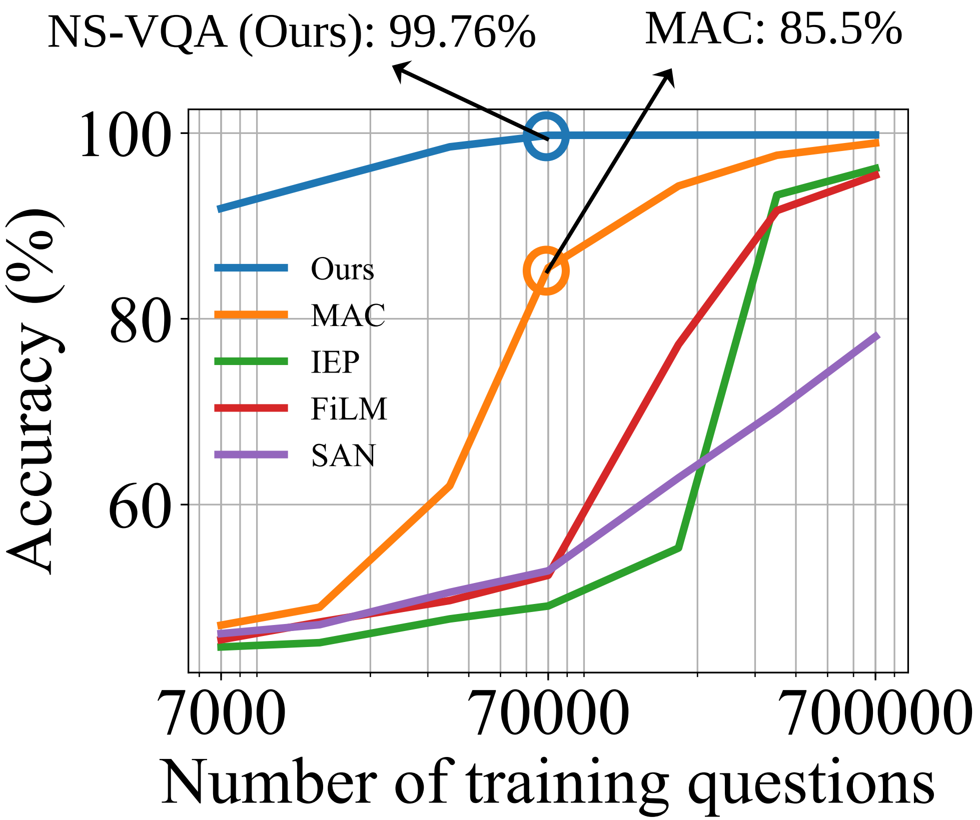
ID	Size	…
3	Large	…

5. count

Answer: 1

Evaluation on CLEVR

Method	Accuracy (%)
Human	92.6
RN	95.5
IEP	96.9
FiLM	97.6
MAC	98.9
NS-VQA (Ours)	**99.8**



[Johnson et al. ICCV 2017, Santoro et al. NIPS 2017, Perez et al. AAAI 2018, Hudson et al. ICLR 2018, Mascharka et al. CVPR 2018]

Part I: Summary

Incorporate symbolic programs for reasoning

Compositional Reasoning

[GLLSG] VQS, ICCV, 17.
[YWGTKT] NS-VQA, NeurIPS, 18.

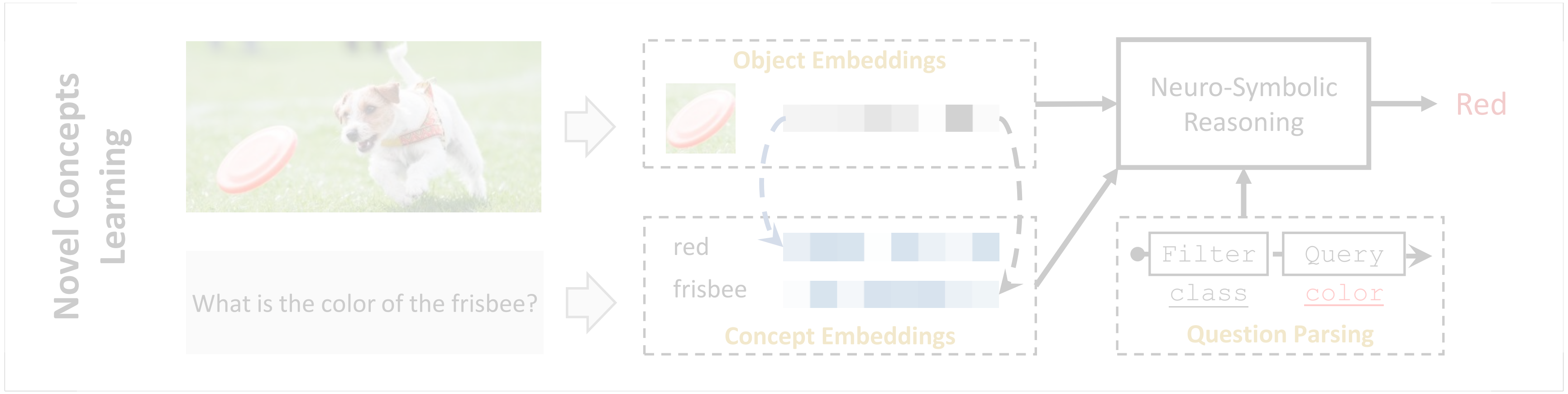
Novel Concepts Learning

[MGKTW] NS-CL, ICLR, 19.
[HMGTW] Meta-NL, NeurIPS, 19.

Compositional Reasoning

Take-Home Message

- Neural Network: extracts symbols from images and language.
- Symbolic Program: reasons in the symbolic space.



Part I: Incorporate Symbolic Programs

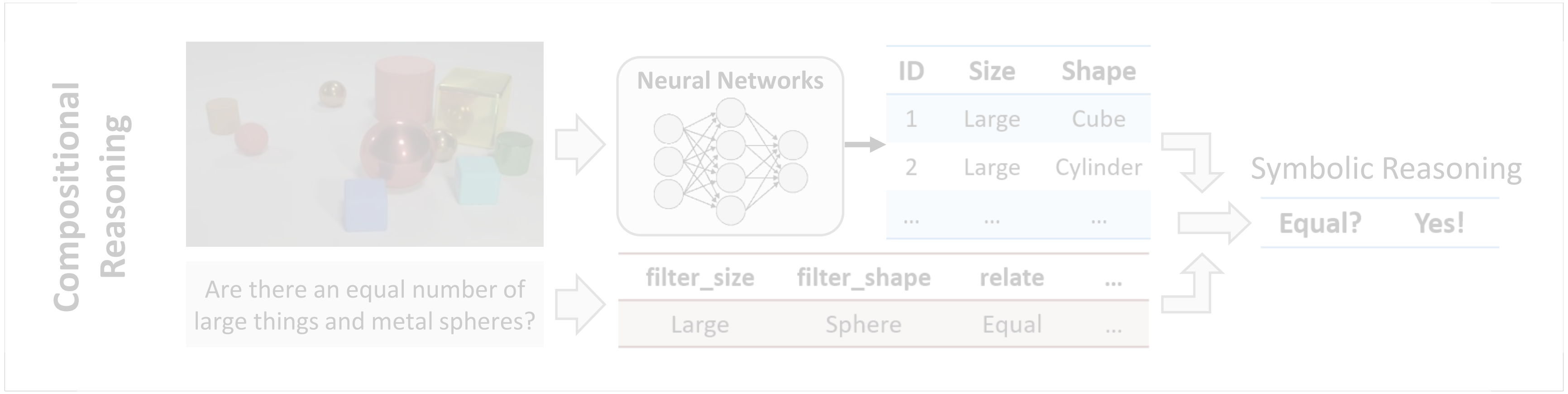
Incorporate symbolic programs for reasoning

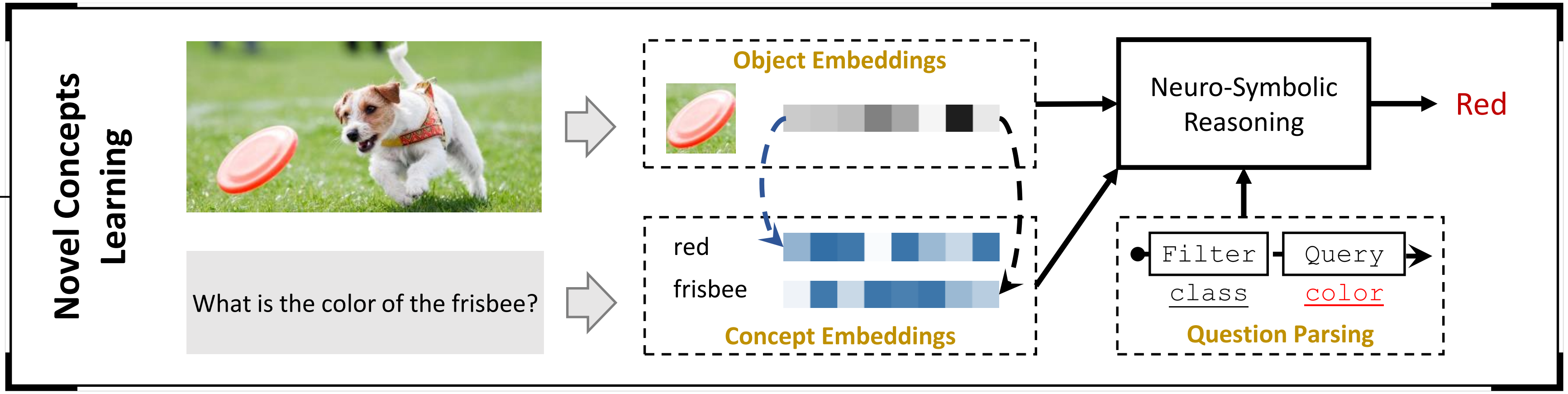
Compositional Reasoning

[GLLSG] VQS, ICCV, 17.
[YWGTKT] NS-VQA, NeurIPS, 18.

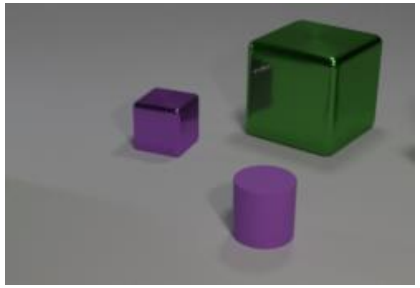
Novel Concepts Learning

[MGKTW] NS-CL, ICLR, 19.
[HMGTW] Meta-NL, NeurIPS, 19.

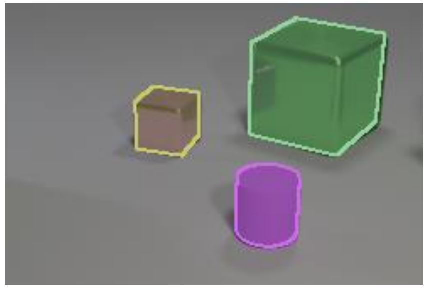




Limitation: Strong Requirement for Labeled Images



→ Mask R-CNN →



→ CNN →

ID	Size	Shape	Material	Color	x	y	z
1	Small	Cube	Metal	Purple	-0.45	-1.10	0.35
2	Small	Cylinder	Rubber	Purple	0.75	1.31	0.35
3	Large	Cube	Metal	Green	1.58	-1.60	0.70

I. *Neural* Scene Parsing

II. *Neural* Question Parsing

How many cubes that are behind the cylinder are large? → LSTM Encoder

LSTM → 1. filter_shape(scene, cylinder)

LSTM → 2. relate(behind)

LSTM → 3. filter_shape(scene, cube)

LSTM → 4. filter_size(scene, large)

LSTM → 5. count(scene)

III. *Symbolic* Program Execution

1. filter_cylinder
2. relate_behind

ID	Size	Shape	…
1	Small	Cube	…
2	Small	Cylinder	…
3	Large	Cube	…

3. filter_cube
4. filter_large

ID	Size	…
3	Large	…

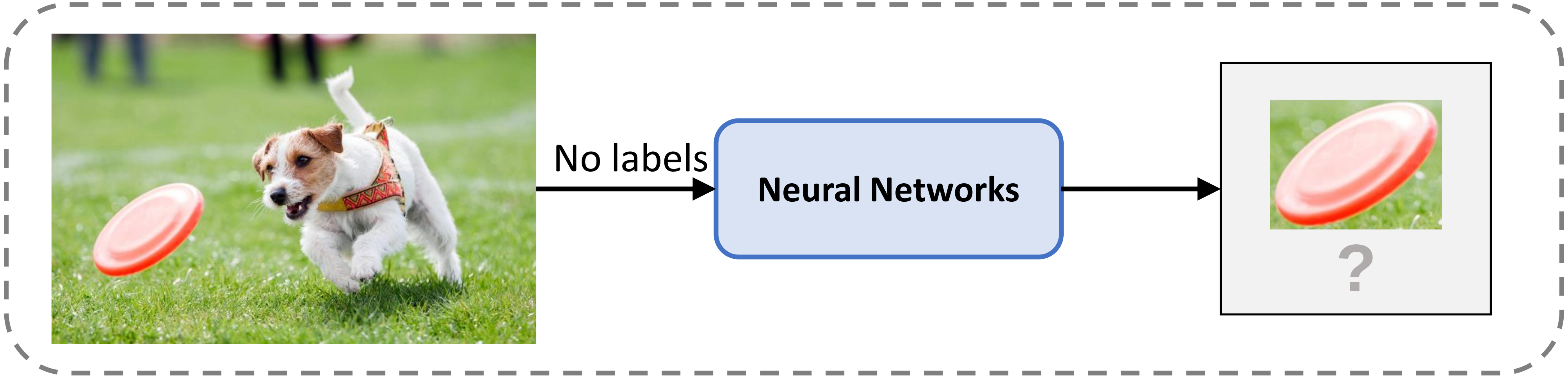
5. count

Answer: 1

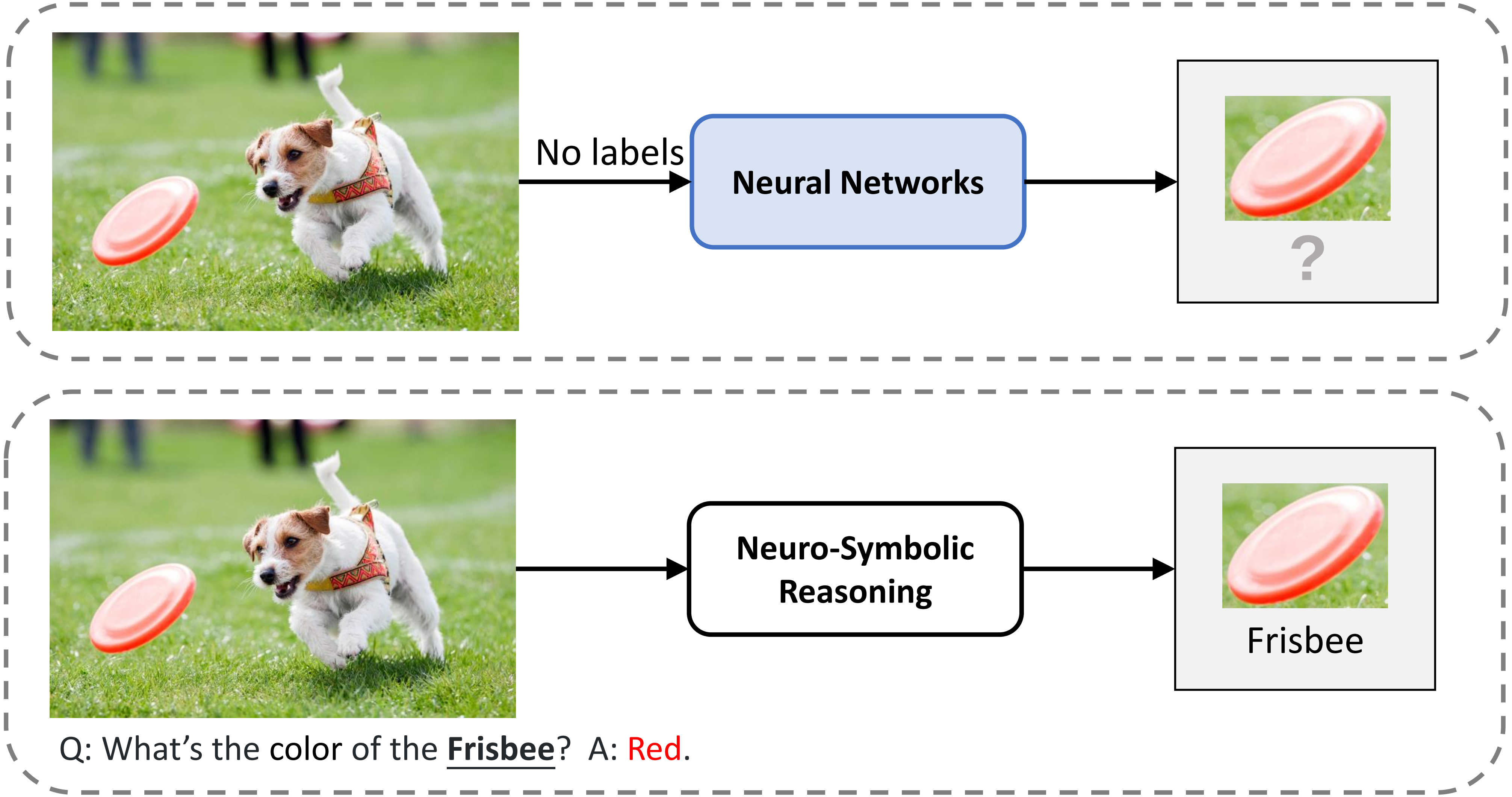
VQS: Linking Segmentations to Questions and Answers for VQA. **Gan** et al. ICCV’17

Neural-Symbolic VQA: Disentangling Reasoning from Vision and Language Understanding. Yi, Wu, **Gan**, et al. NeurIPS’18

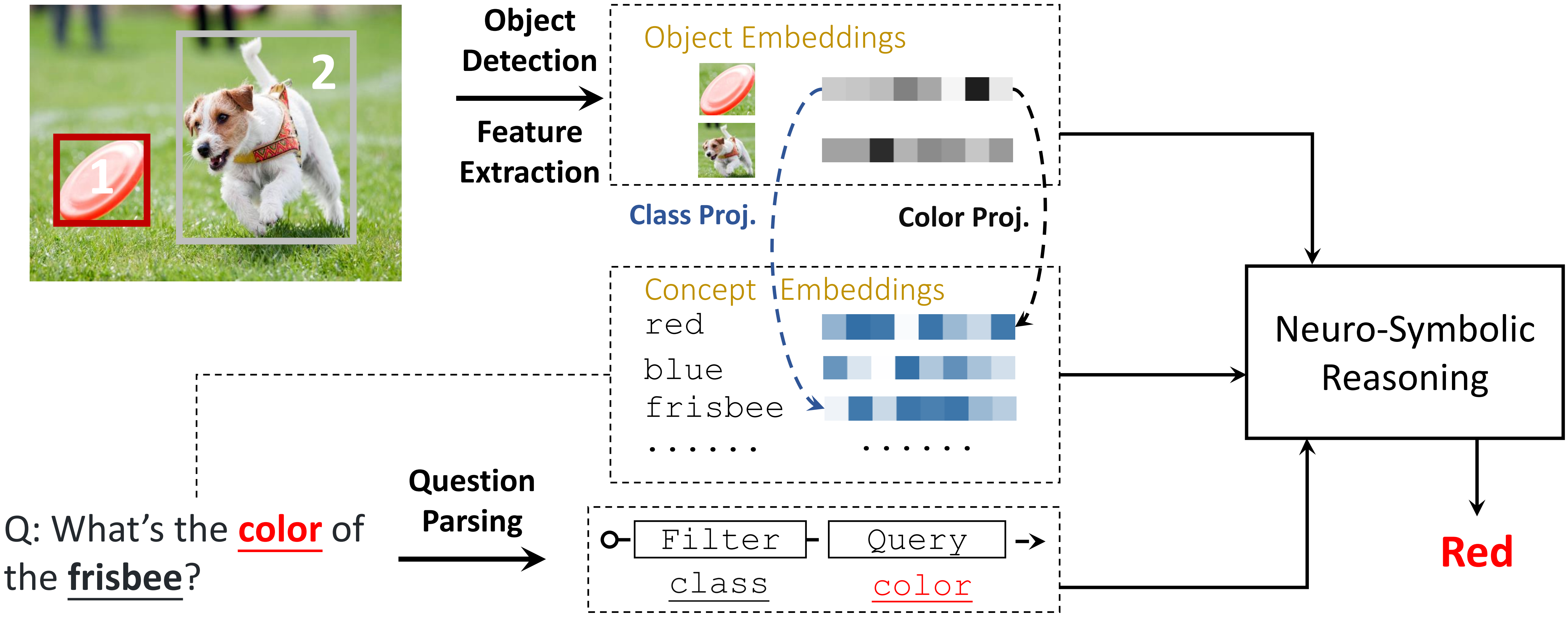
How About Images Without Concept Labels?



Our Idea: Learning Concepts From Weak Supervisions

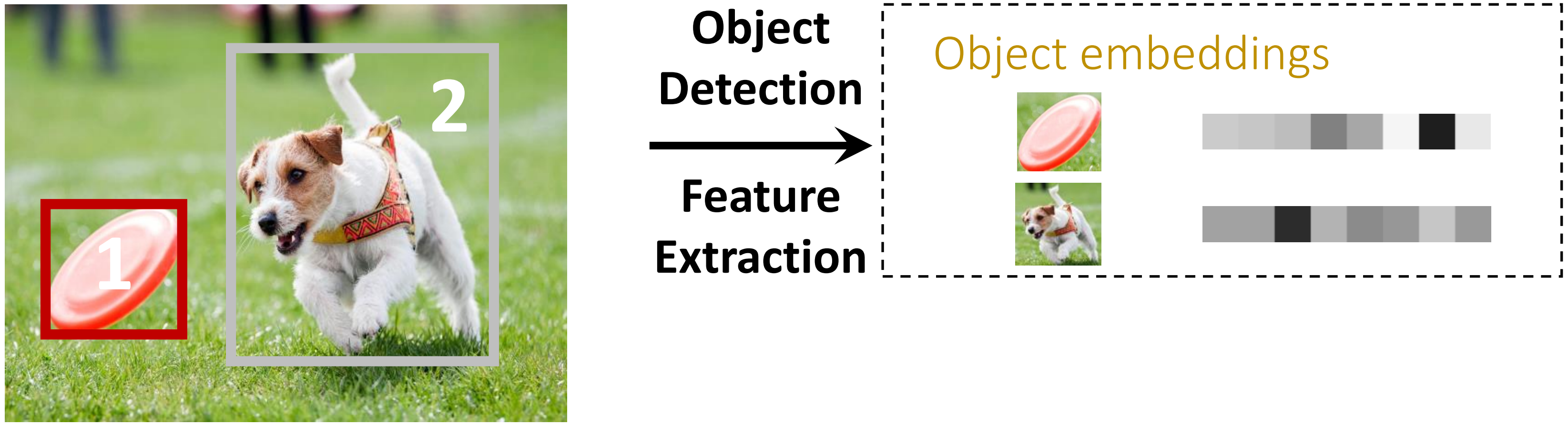


Neuro-Symbolic Concepts Learner



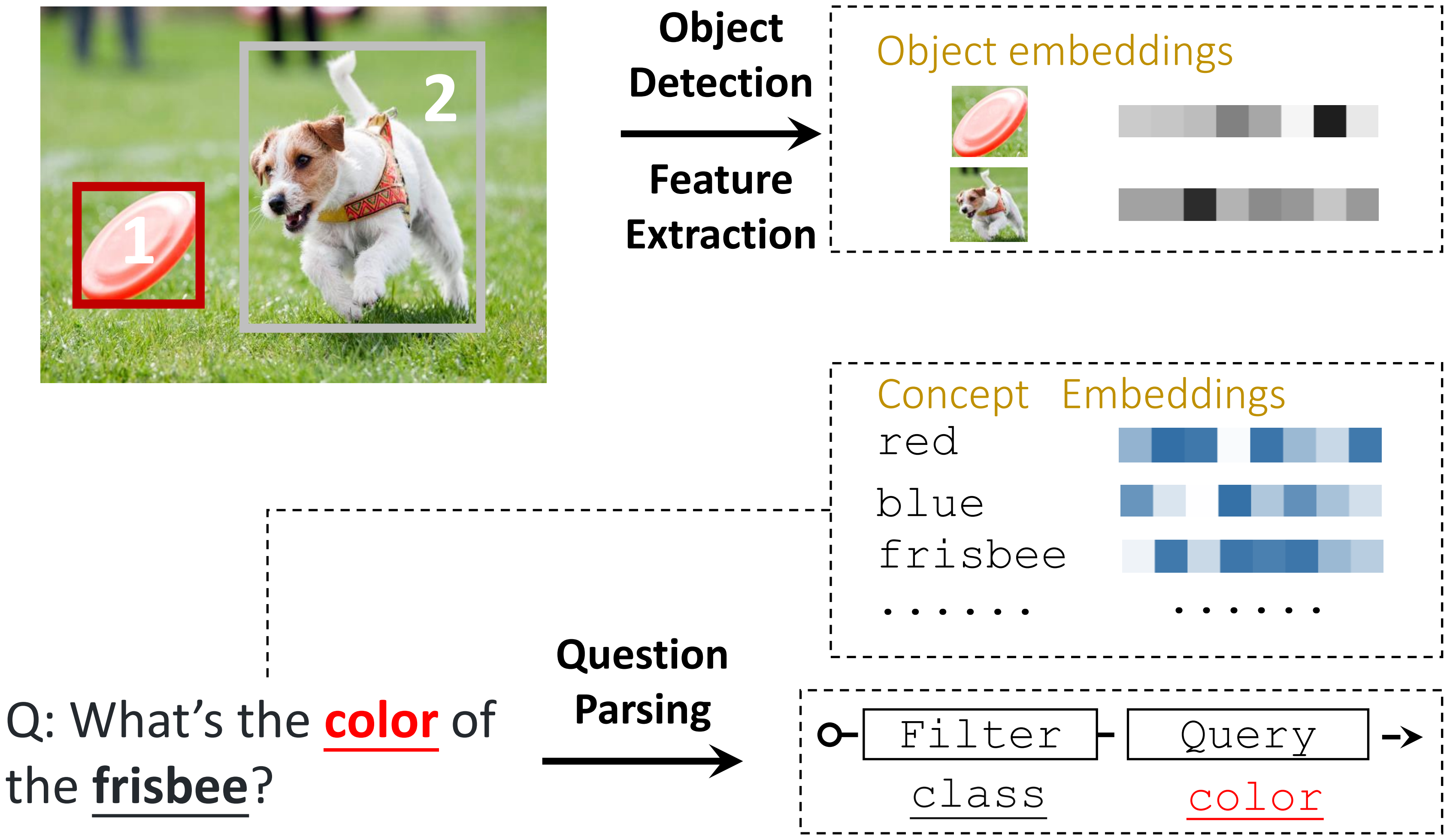
The neuro-symbolic concept learner: Interpreting scenes, words, and sentences from natural supervision. Mao, **Gan**, et al. ICLR'19

Object Embeddings

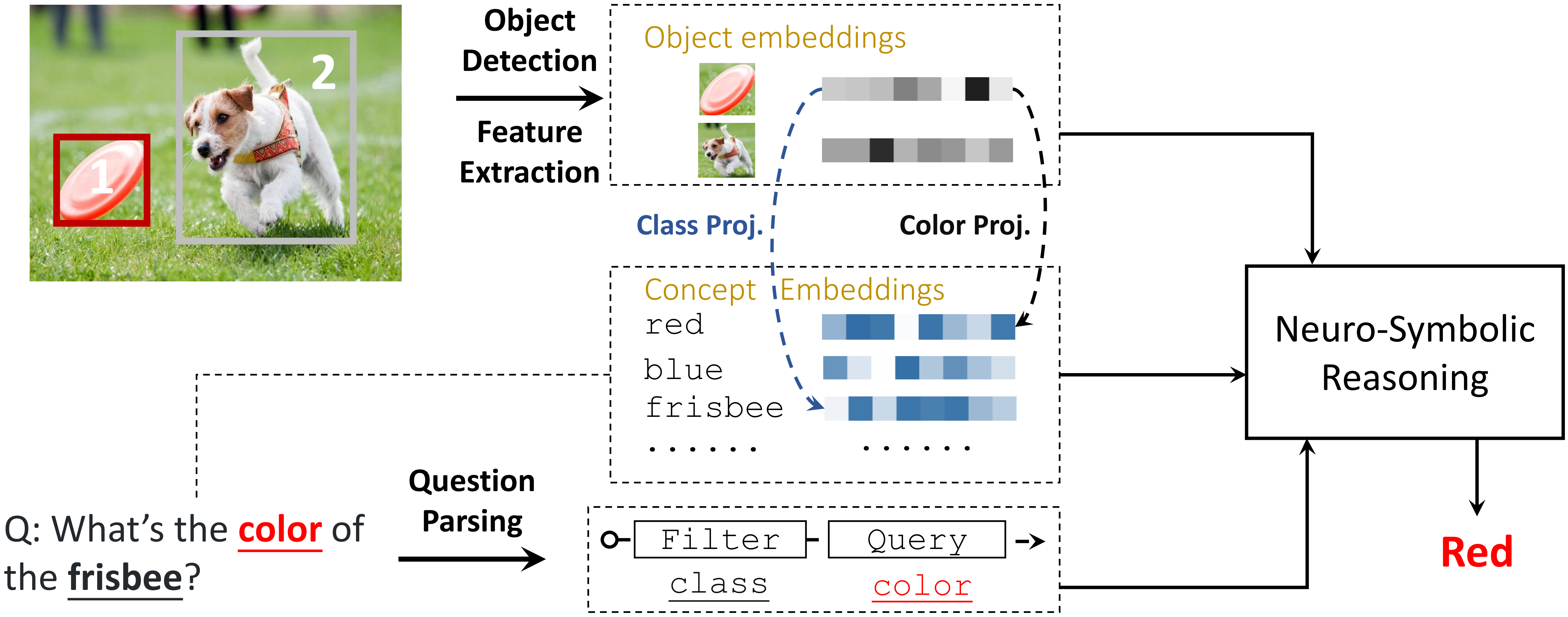


Q: What’s the **color** of the **frisbee**?

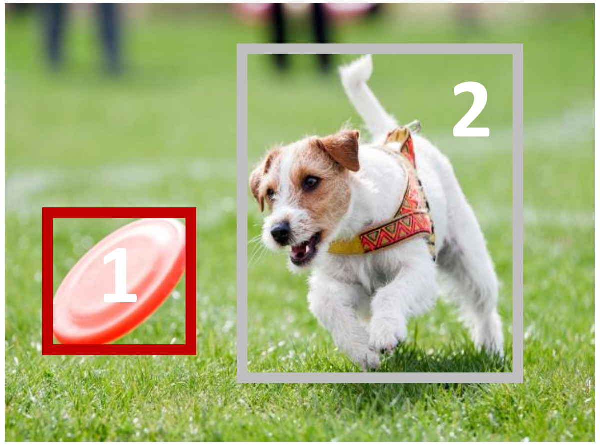
Concept Embeddings



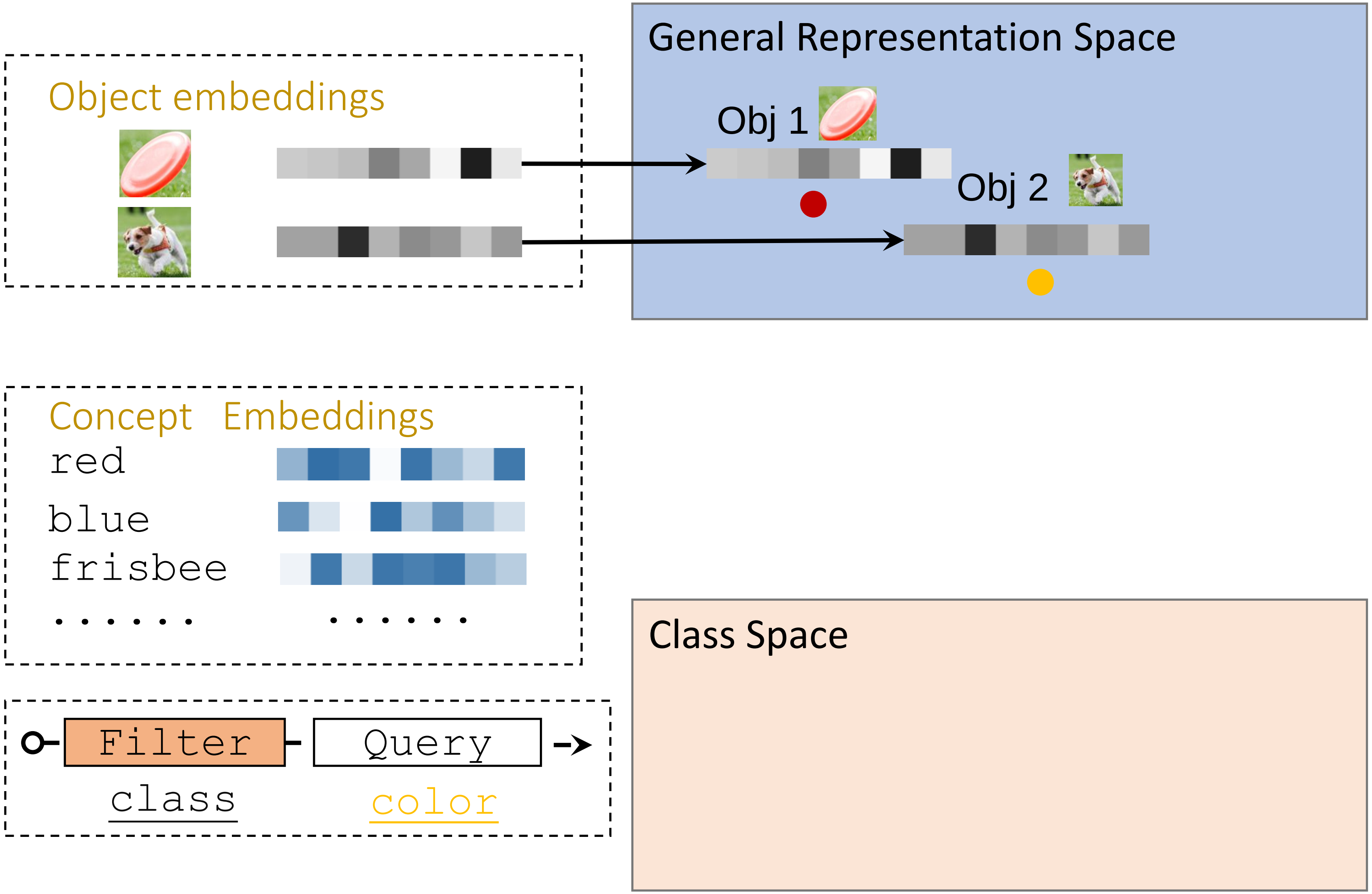
Neuro-symbolic Reasoning



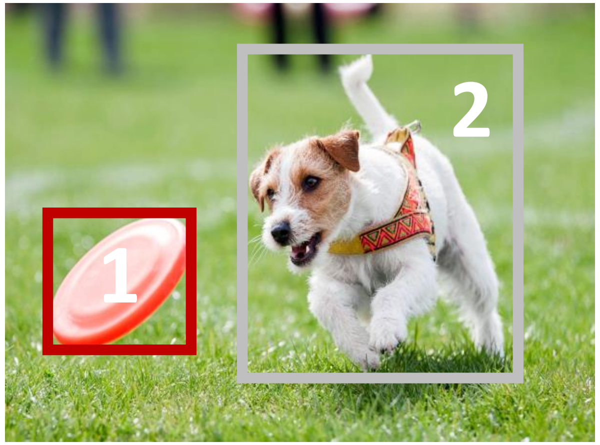
Concept Grounding (Class)

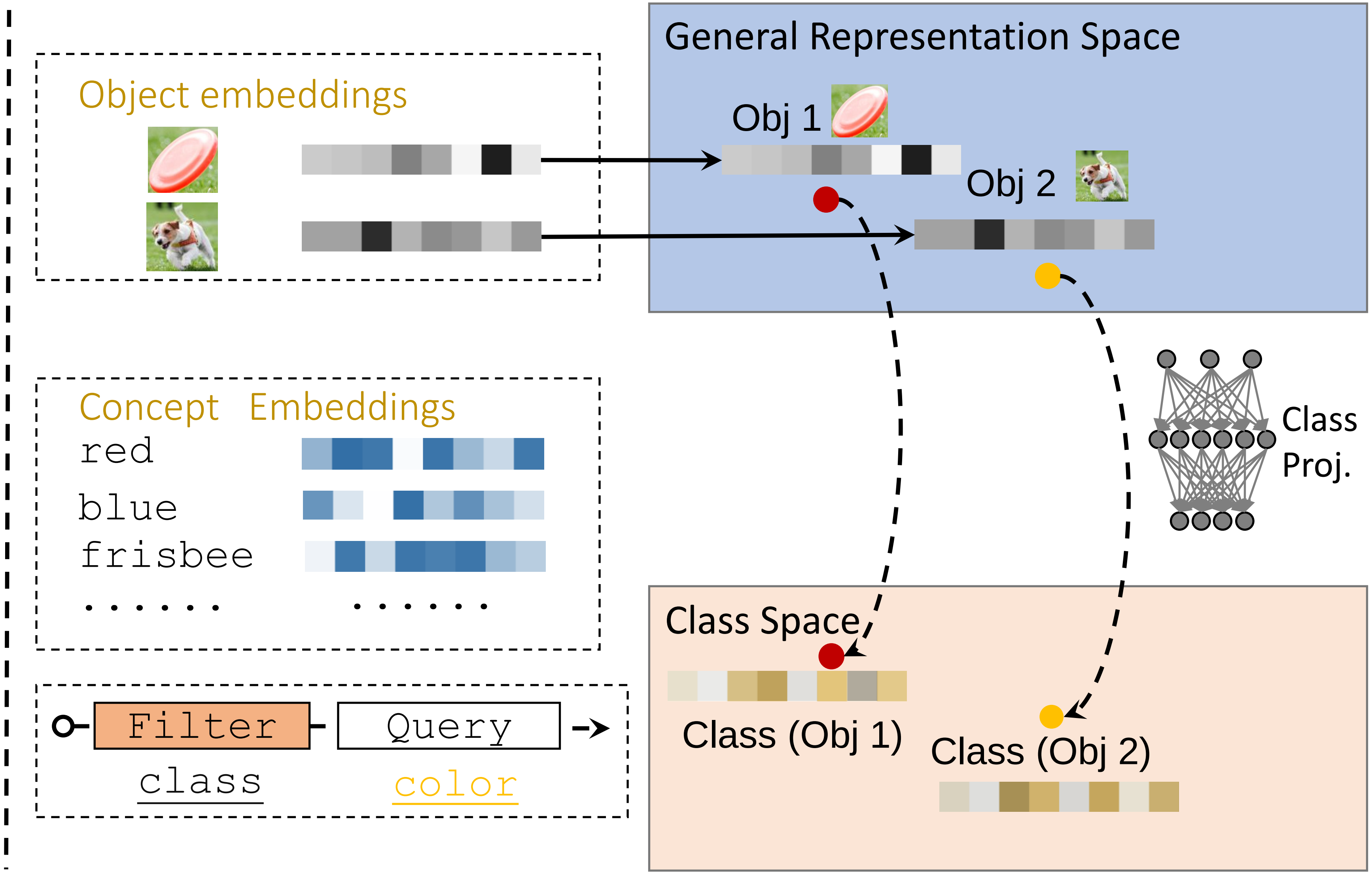


Q: What's the **color** of the **frisbee**?



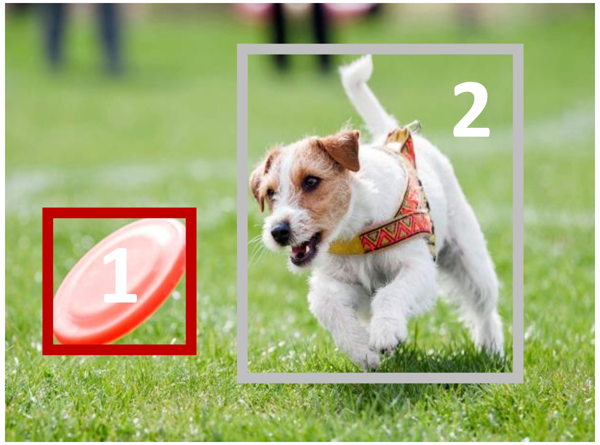
Concept Grounding (Class)



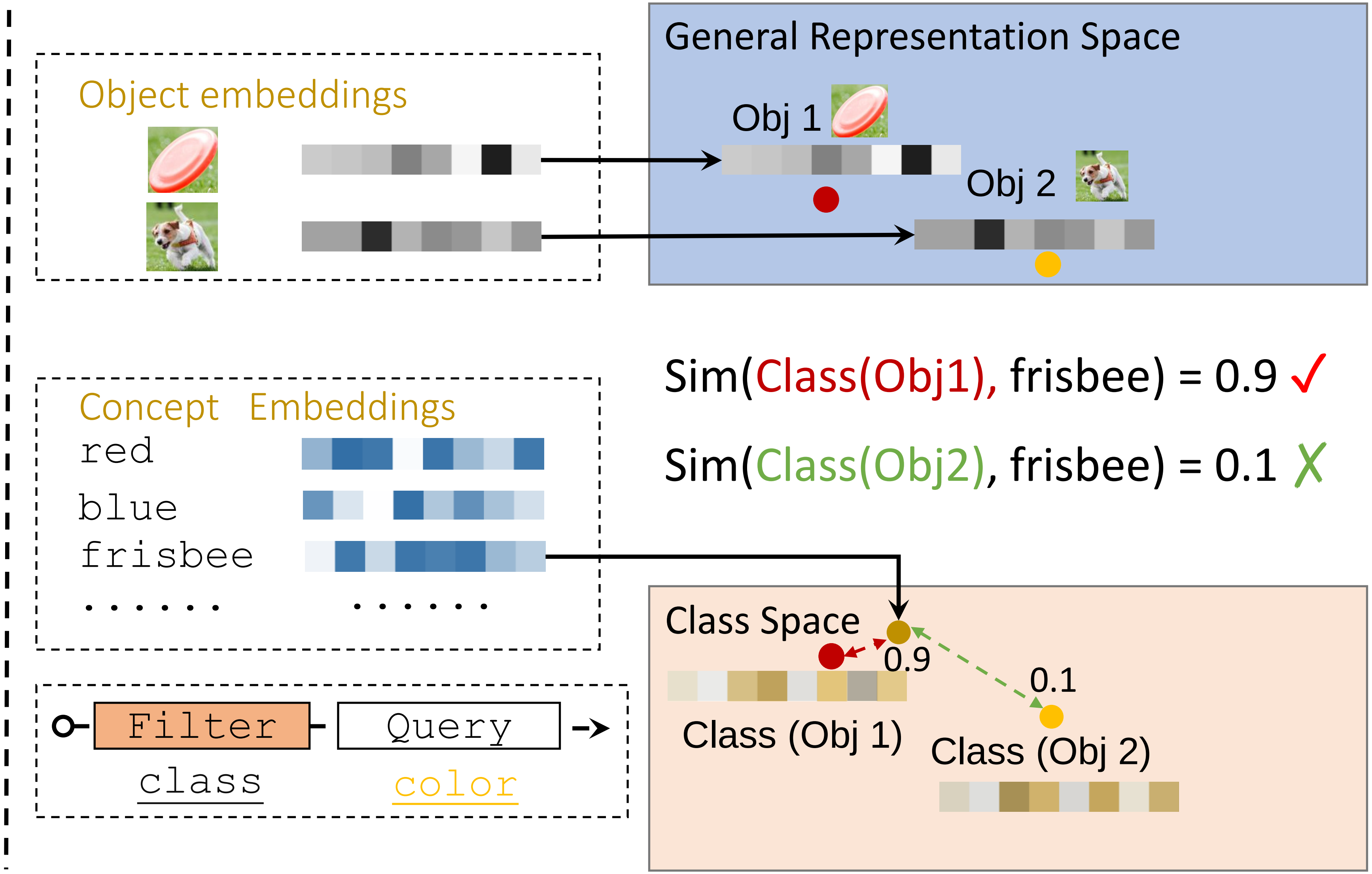


Q: What's the **color** of the **frisbee**?

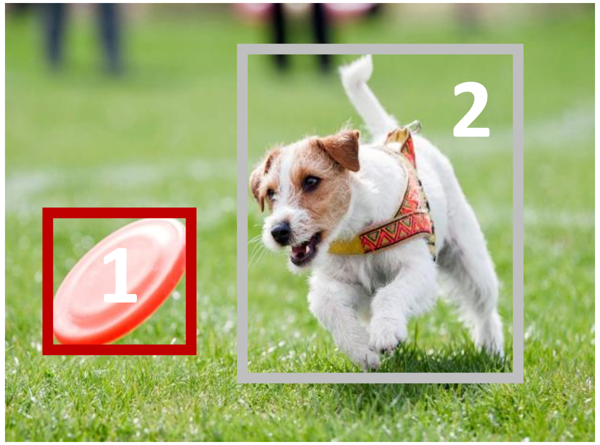
Concept Grounding (Class)



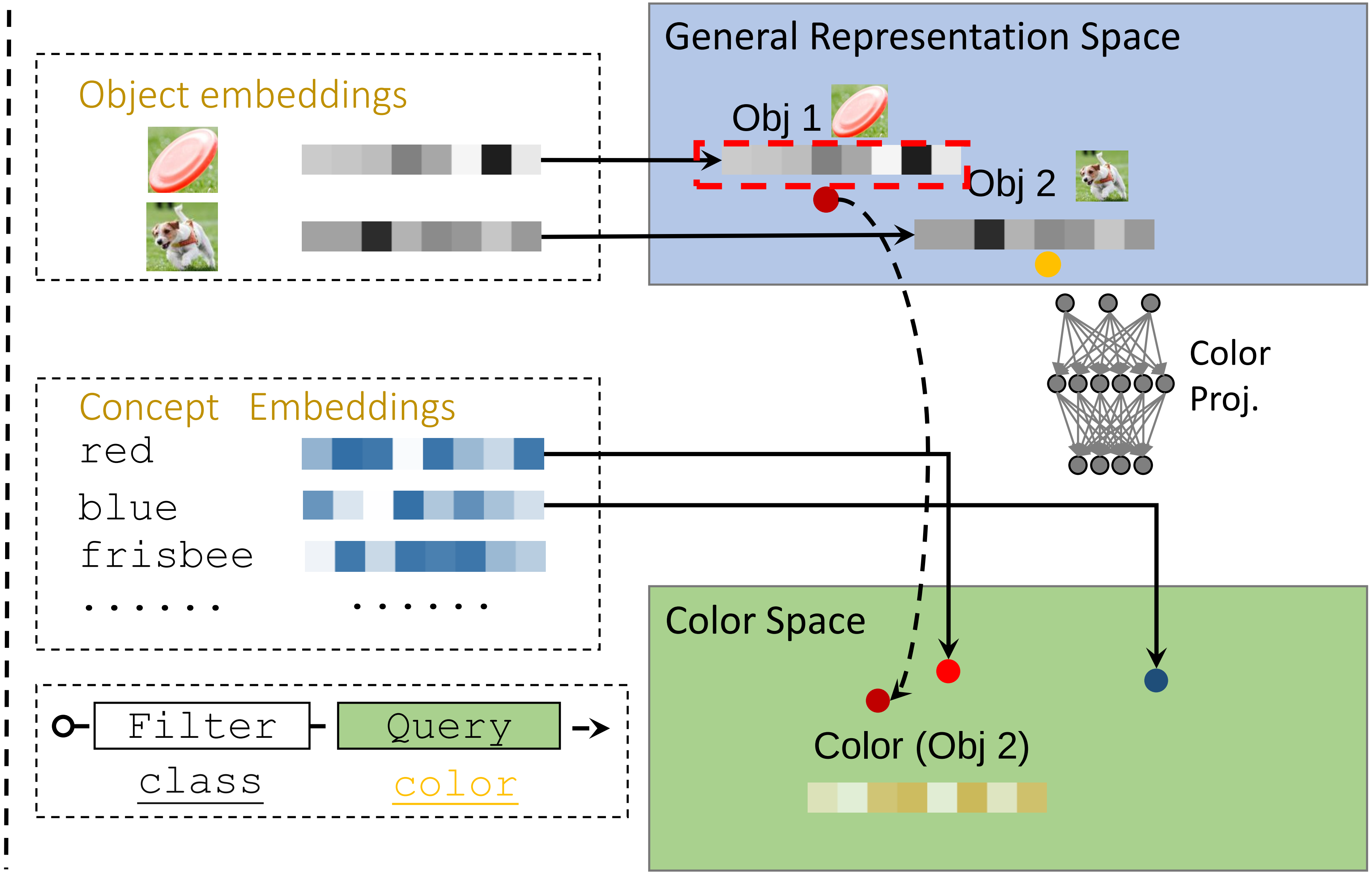
Q: What's the **color** of the **frisbee**?



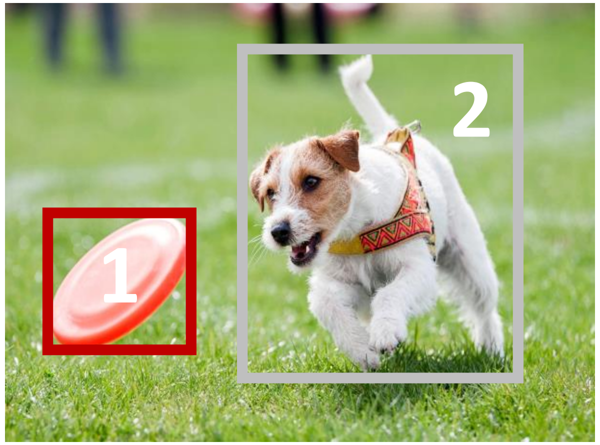
Concept Grounding (Color)



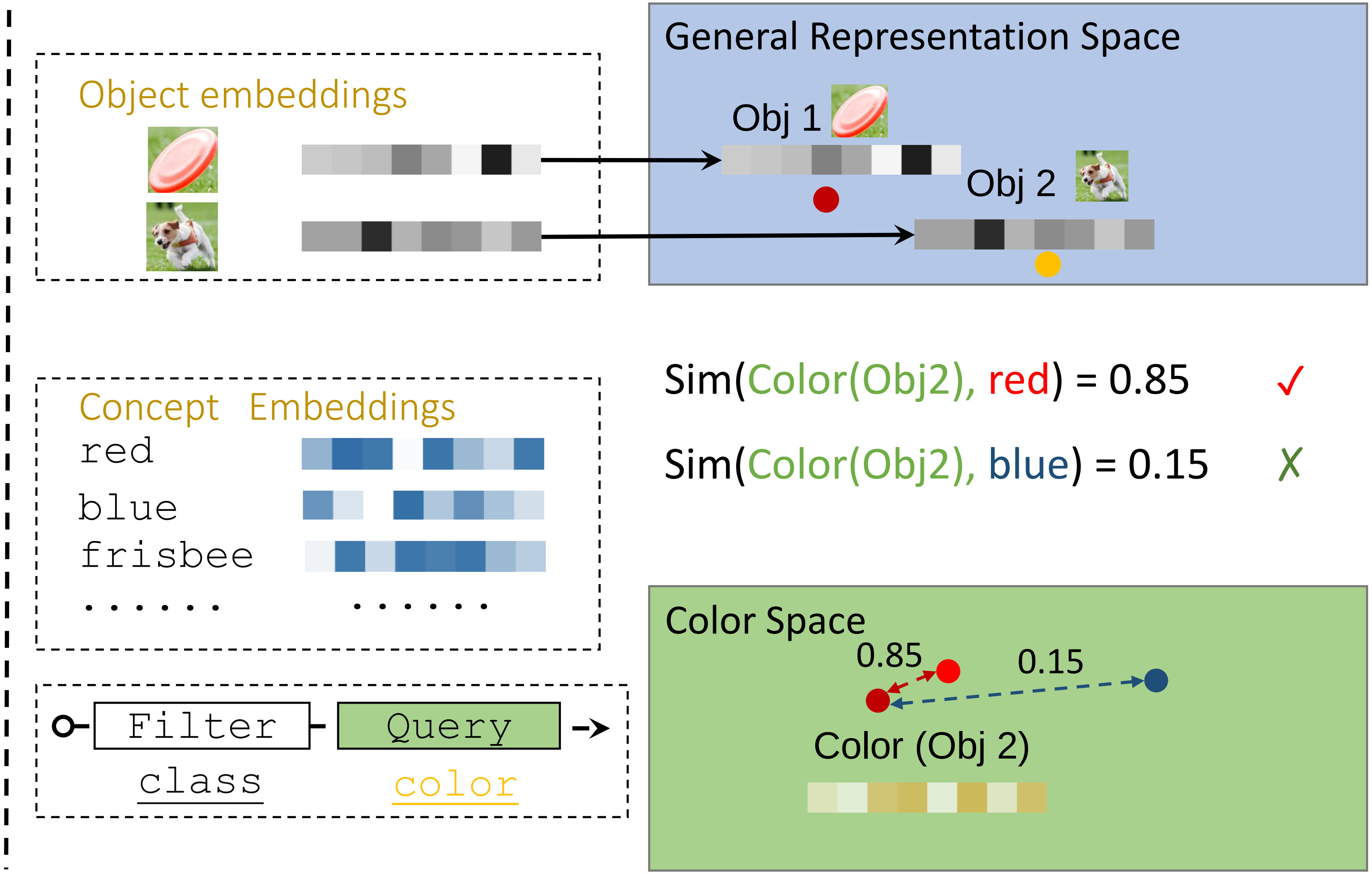
Q: What's the **color** of the **frisbee**?



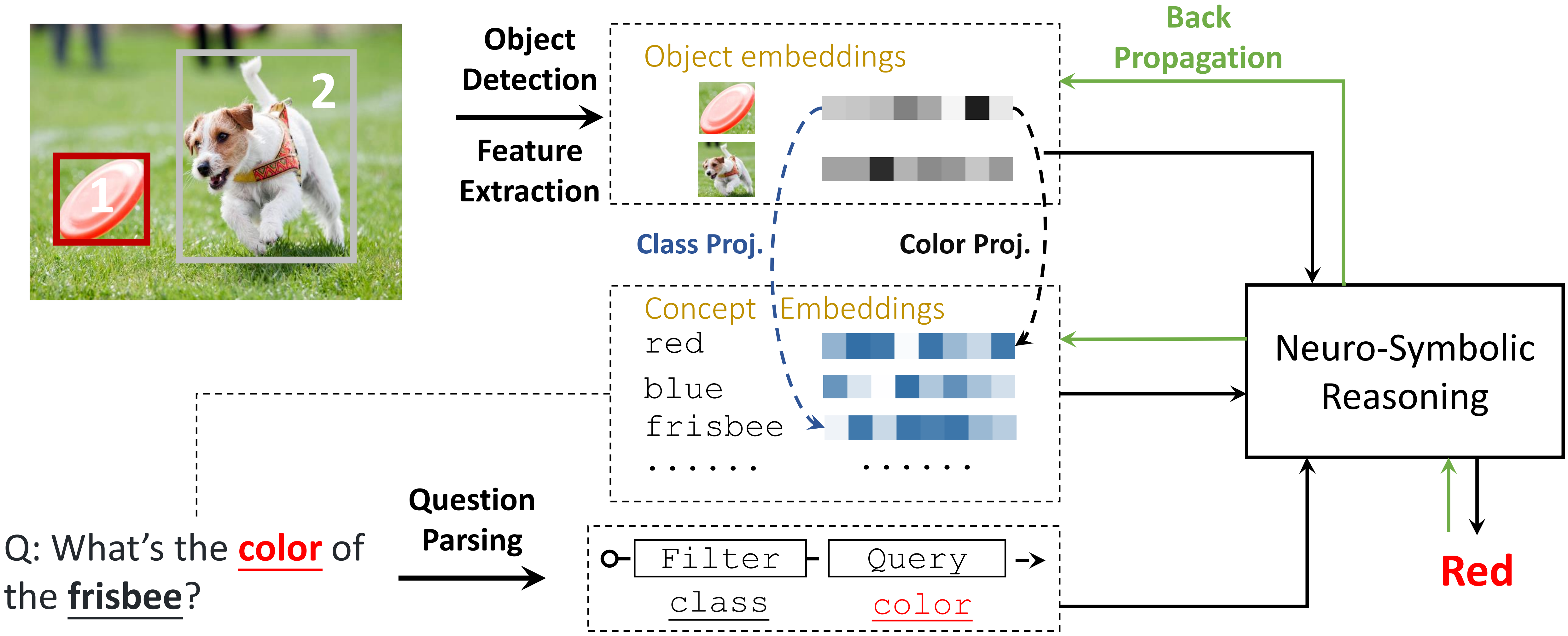
Concept Grounding (Class)



Q: What's the **color** of the **frisbee**?



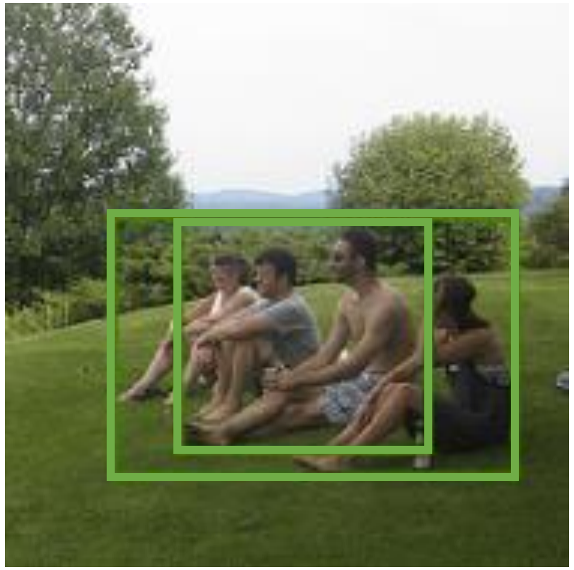
Neuro-symbolic Reasoning



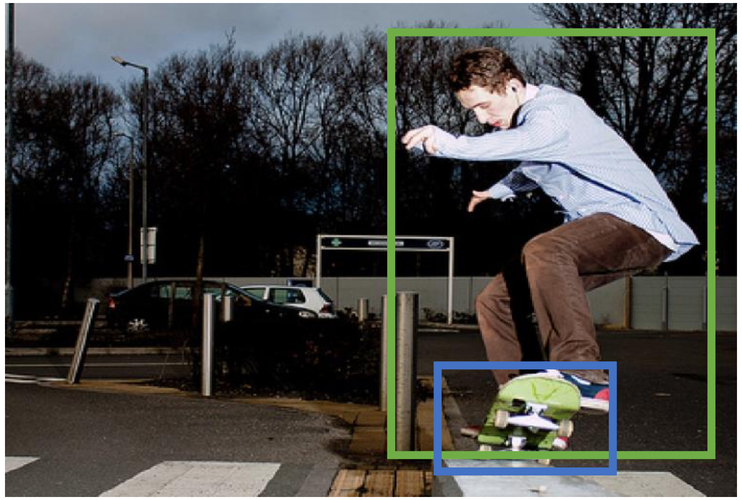
Evaluation Results

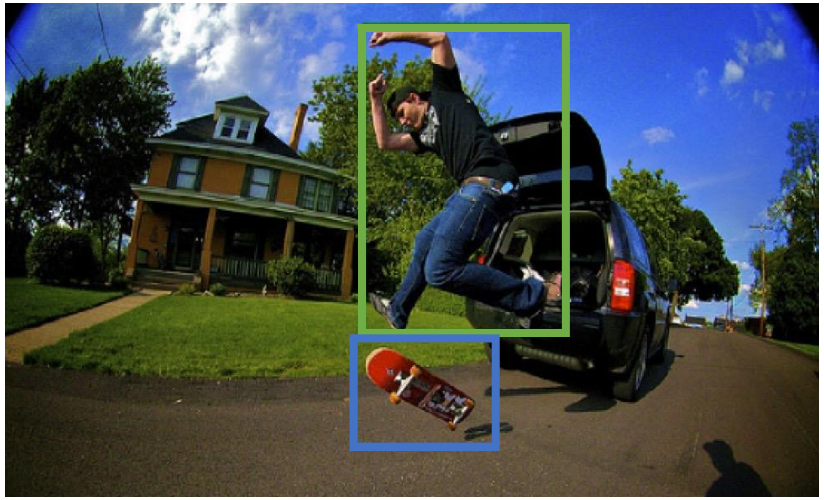
Concept: Horse

Concept: Person



Concept: Person On a Skateboard





My vision: Neuro-Symbolic AI

Neural Network

Extracting structured representations from data

Representing and reasoning from abstract knowledge

Symbolic Program

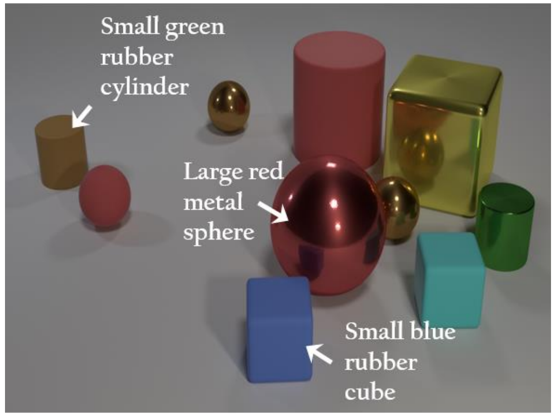
Physics Engine

Existing Visual Reasoning Datasets

Q: *What is the mustache made of?*

A: *Banana*

VQA [Antol et al. ICCV 2015]



Q: *Are there an equal number of large things and metal spheres?*

A: *Yes*

CLEVR [Johnson et al. CVPR 2017]

Q: *What does the cat do three times?*

A: *Put head down*

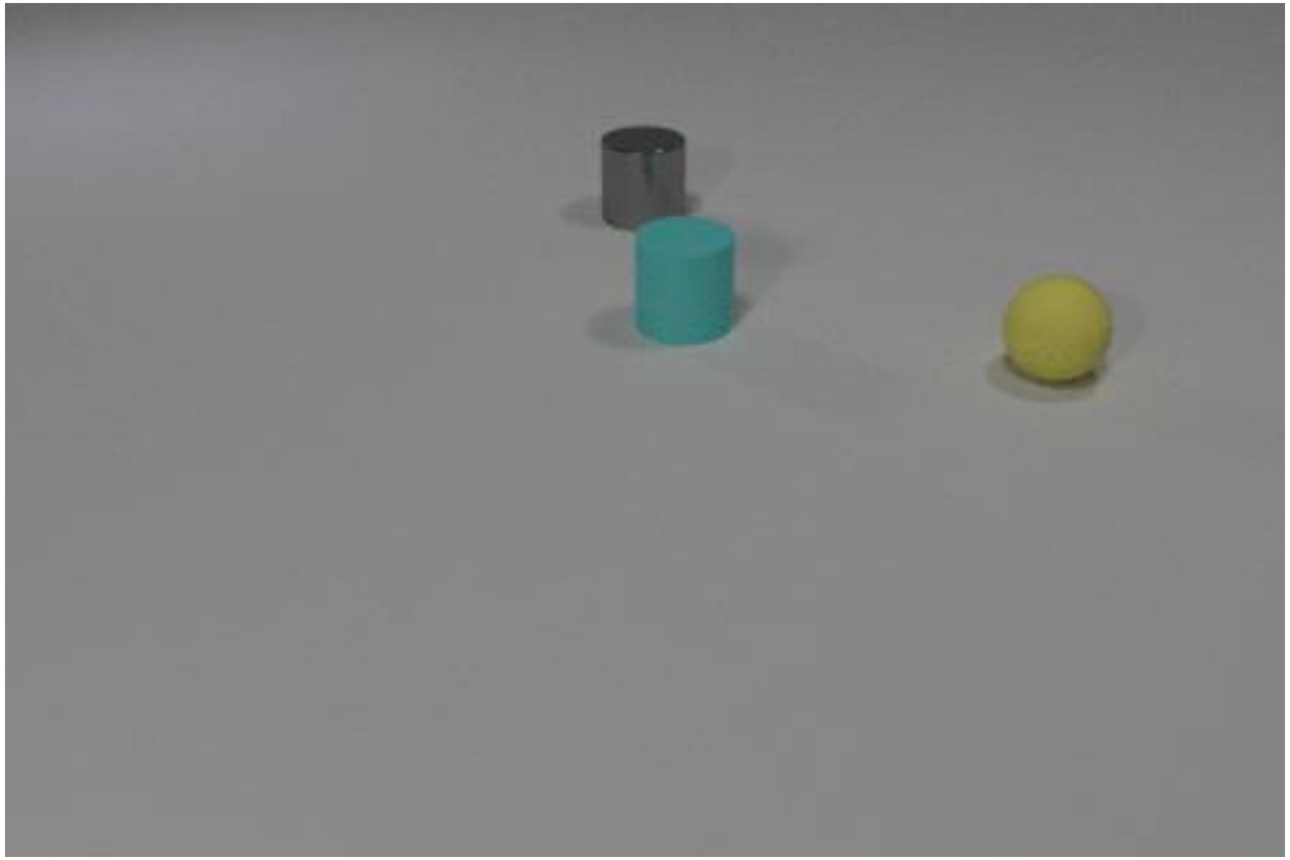
TGIF-QA [Jang et al. CVPR 2017]

We Also Need to Understand Dynamics

- Describe what has happened
- Explain why it has happened
- Predict what is about to happen

CLEVRER Dataset: From Static Scene to Dynamic Scene

- 20,000 Synthetic videos
- 300,000+ questions
- Controlled biases
- Diagnostic annotations
- Dynamics visual reasoning
  - Descriptive
  - Explanatory
  - Predictive
  - Counterfactual



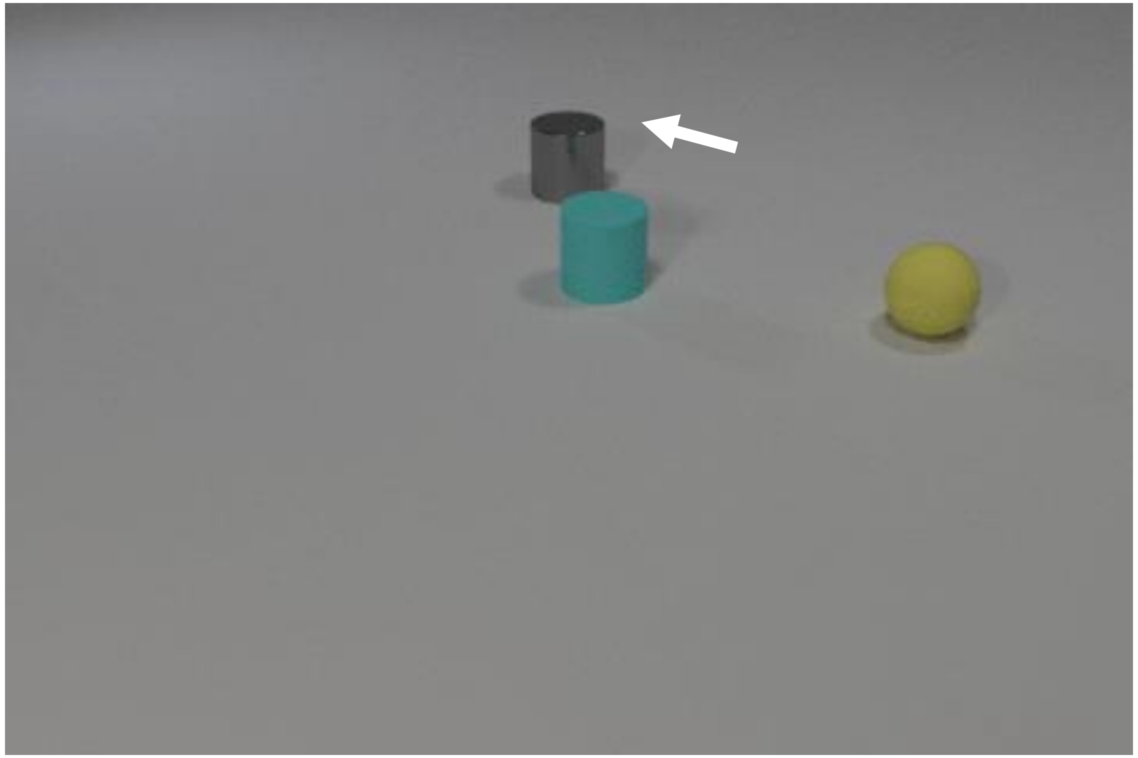
CLEVRER: CoLlision Events for Video REpresentation and Reasoning. Yi*, **Gan***, et al. ICLR’20

Question Types

Descriptive | Explanatory | Predictive | Counterfactual

Q: *What is the material of the last object to collide with the cyan cylinder?*

A: *Metal*

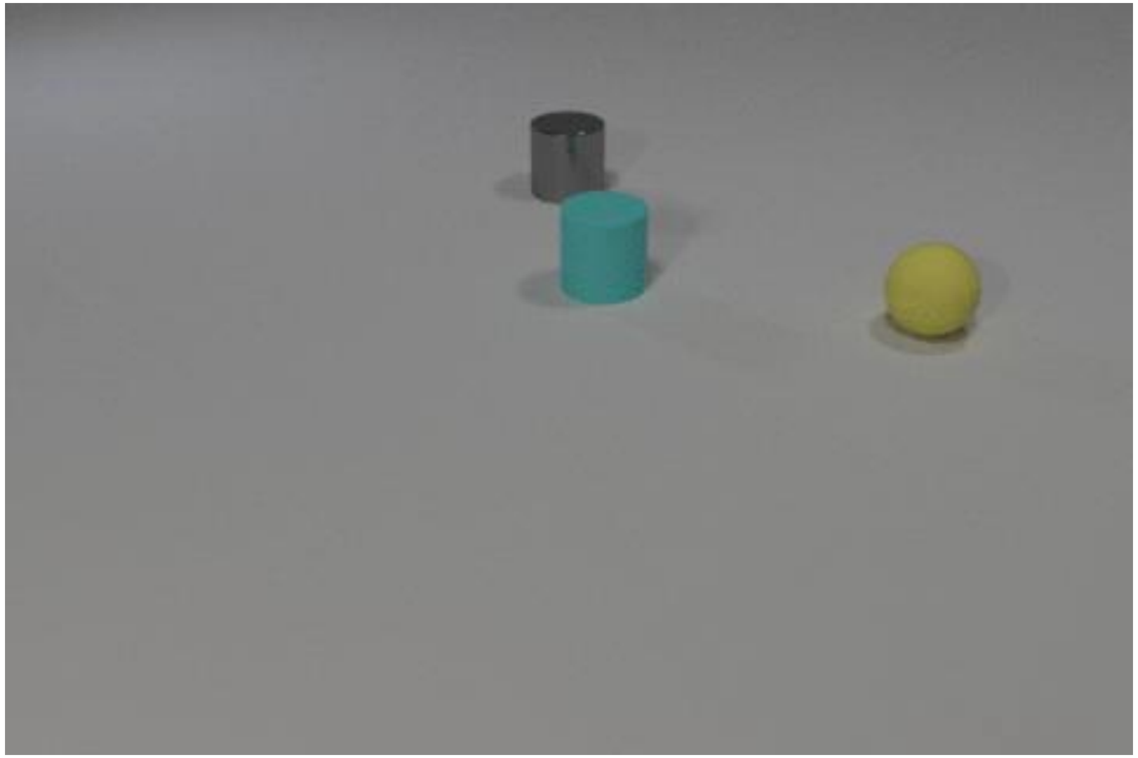


Question Types

Descriptive | **Explanatory** | Predictive | Counterfactual

Q: *What is responsible for the collision between the cyan and gray cylinder?*

A: *The collision between the cyan cylinder and the red rubber sphere.*

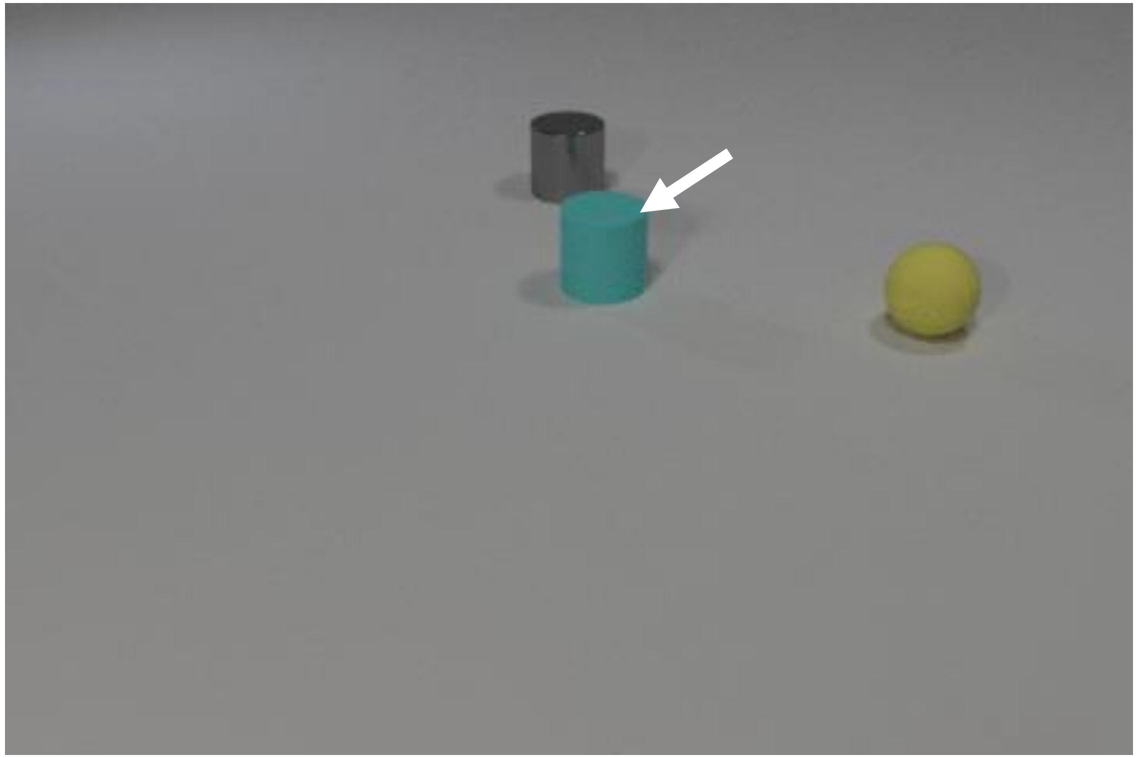


Question Types

Descriptive | Explanatory | Predictive | Counterfactual

Q: *What will happen next?*

A: *The red rubber sphere collides with the metal sphere.*

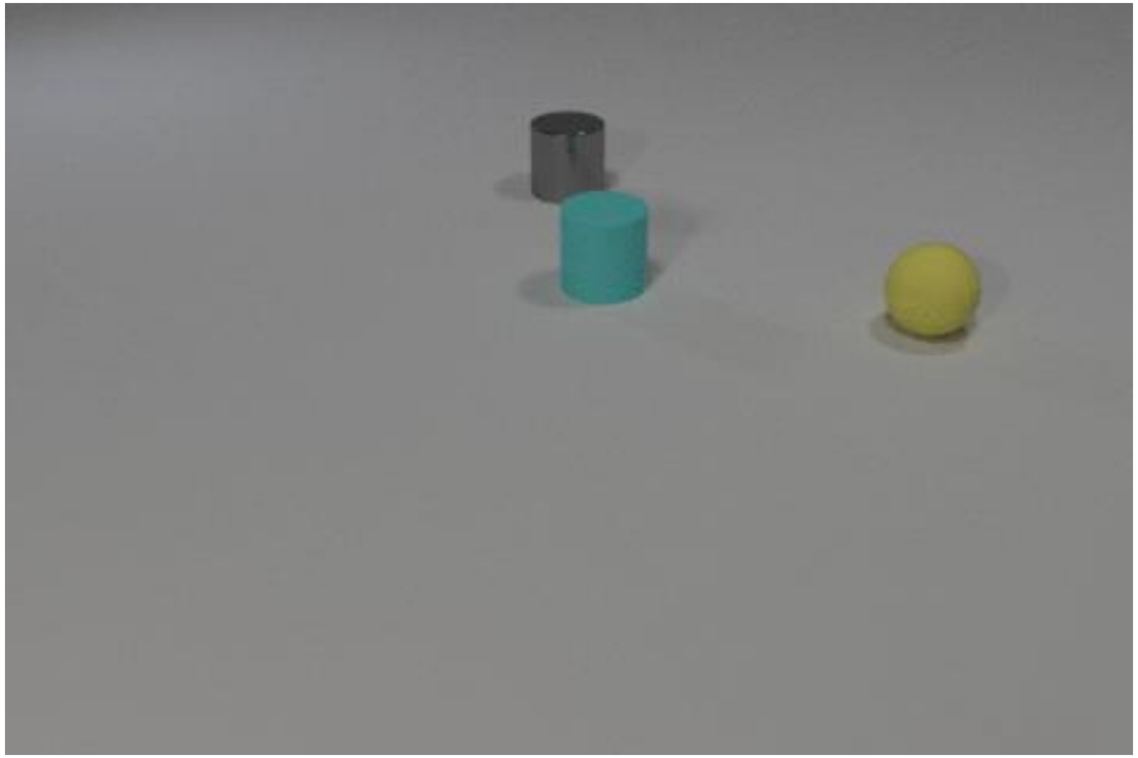


Question Types

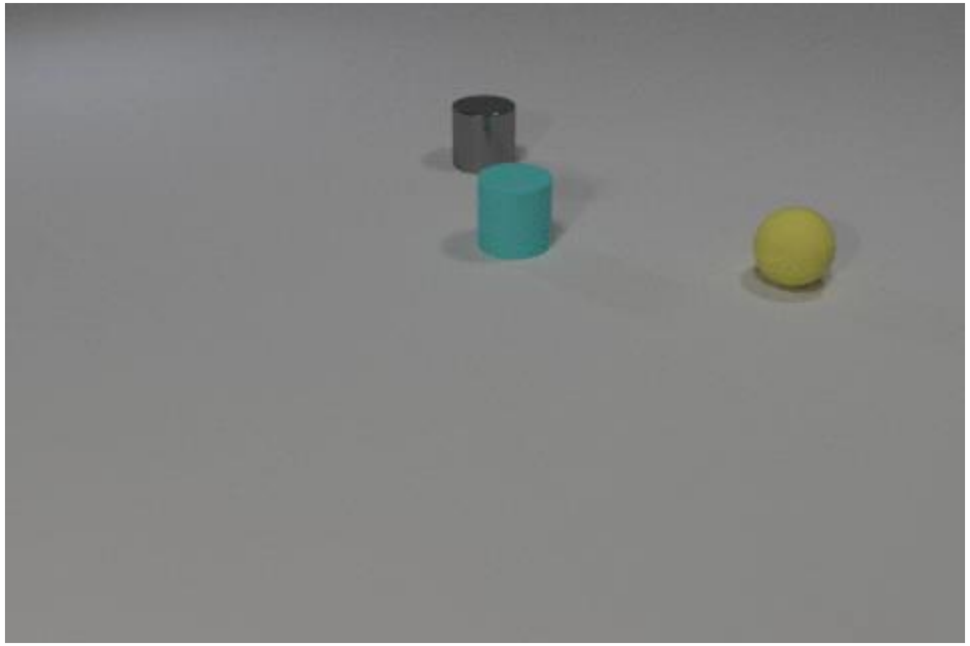
Descriptive | Explanatory | Predictive | **Counterfactual**

Q: *What would happen without the cyan cylinder?*

A: *The red rubber sphere and the gray object collide.*



Prior Work: End-to-End Video Reasoning



Video Question Answering

Question: *What will happen without the cyan cylinder?*

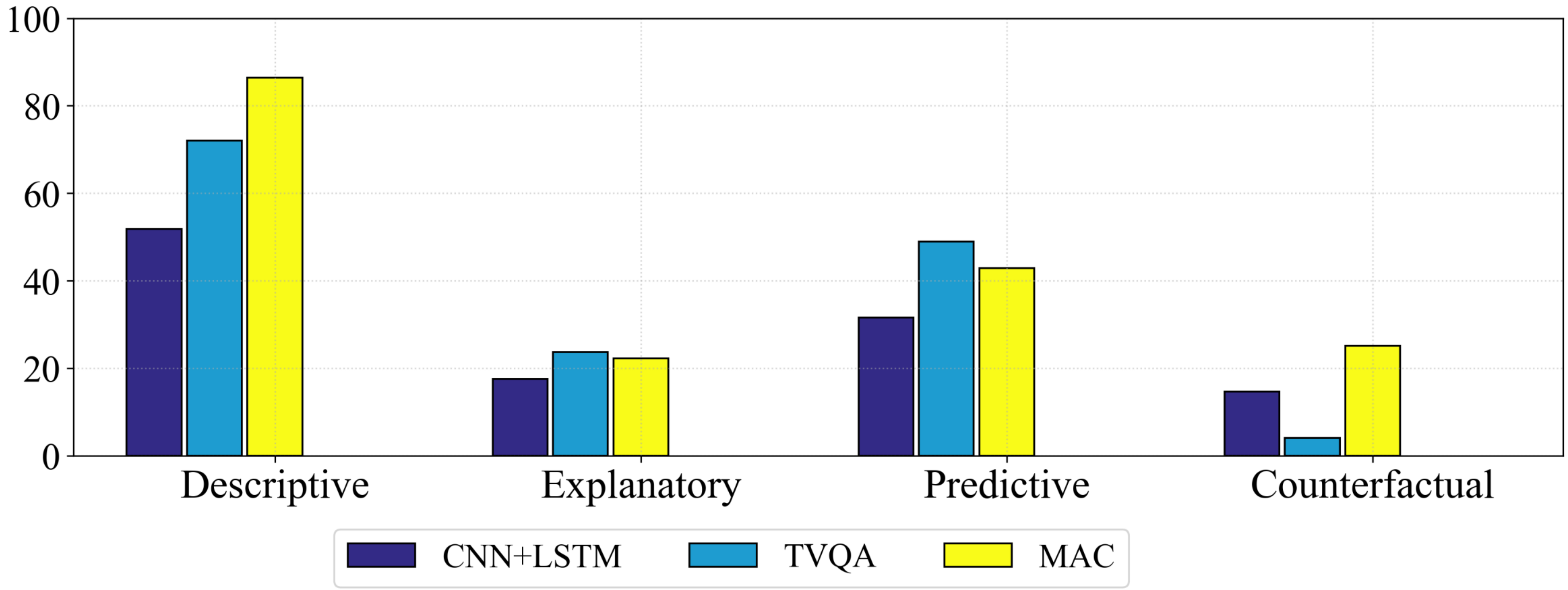
End-to-End
Neural Network

Yes/No

Candidate Answer: *The red rubber sphere and the gray object collide.*

Shi et al. LSTM, 2015. Lei et al. TVQA, 2018. Hudson & Manning. MAC, 2018.

Evaluation of End-to-End Video Reasoning

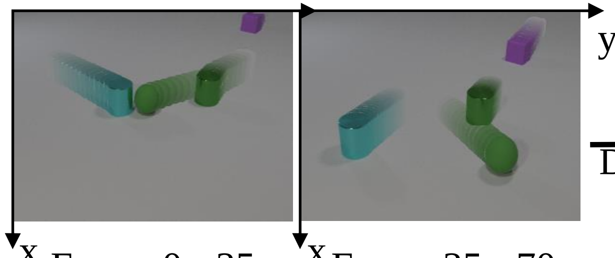


Physical reasoning requires to understand object dynamics.

Shi et al. LSTM, 2015. Lei et al. TVQA, 2018. Hudson & Manning. MAC, 2018.

Neuro-symbolic Dynamics Reasoning

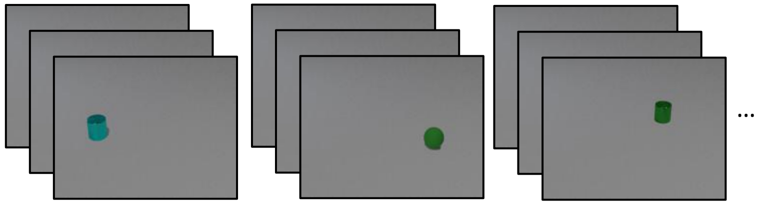
Input Video Frames



Frame 0 - 35

Frame 35 - 70

Object Trajectories



Cyan Cylinder

Green Sphere

Green Cylinder

Concept Learner

Visual Properties

	Shape	Color	Location
Obj1	Cylinder	Cyan	(-2.3,-1.7)
Obj2	Sphere	Green	(-1.2, 1.1)
Obj3	Cylinder	Green	(-2.2, 2.3)
Obj4	Cube	Purple	(-6.7, 4.1)

Collision Events

Time	1.3s
Obj1	Cylinder
	Cyan
Obj2	Sphere
	Green

Physical Parameters

	Mass	Restitution	Velocity	…
Obj1	?	?	?	
Obj2	?	?	?	
Obj3	?	?	?	
Obj4	?	?	?	

Physical Object Representation

Descriptive Q: What is the color of the object that collides with the green sphere?

Parse

Programs

```
1. Filter(collisions)
2. Filter(green, Sphere)
3. Relate_collision()
4. Query(color)
```

Question Parser

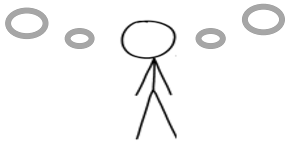
Answer: **Cyan**

Predictive Q: Will the green cylinder collide with the cube?

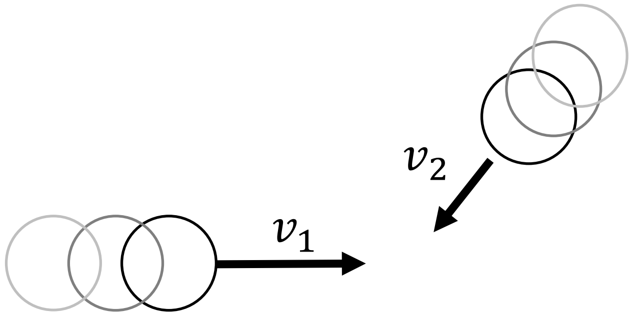
Dynamic Visual Reasoning by Learning Differentiable Physics Models from Video and Language. Ding, Chen, Du, Luo, Tenenbaum, **Gan**. NeurIPS'21

Parameters Estimation from Collision Events

Which object is heavier, the gray sphere or the green sphere?



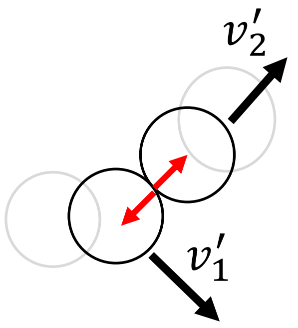
Forward Simulation of Sphere Collisions



	Shape	Location (x)	Velocity (v)	…
Obj1	Sphere	(-1.0,0.0)	(1.0, 0.0)	
Obj2	Sphere	(1.5, 1.5)	(-0.1, -0.1)	
…				

$$x_1^t = x_1^{t-1} + v_1 dt$$

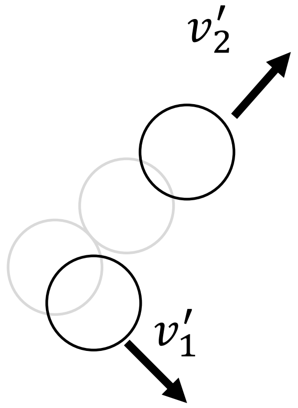
$$x_2^t = x_2^{t-1} + v_2 dt$$



	Shape	Location (x)	Velocity (v)	…
Obj1	Sphere	(0.0,0.0)	(0.5, -0.5)	
Obj2	Sphere	(1.4, 1.4)	(0.6, 0.6)	
…				

$$m_1 v_1 + m_2 v_2 = m_1 v_1' + m_2 v_2'$$

Coefficient of Restitution: $e = \frac{v_2' - v_1'}{v_2 - v_1}$



	Shape	Location (x)	Velocity (v)	…
Obj1	Sphere	(0.5, -0.5)	(0.5, -0.5)	
Obj2	Sphere	(2.0, 2.0)	(0.6, 0.6)	
…				

$$x_1^{t+1} = x_1^t + v_1' dt$$

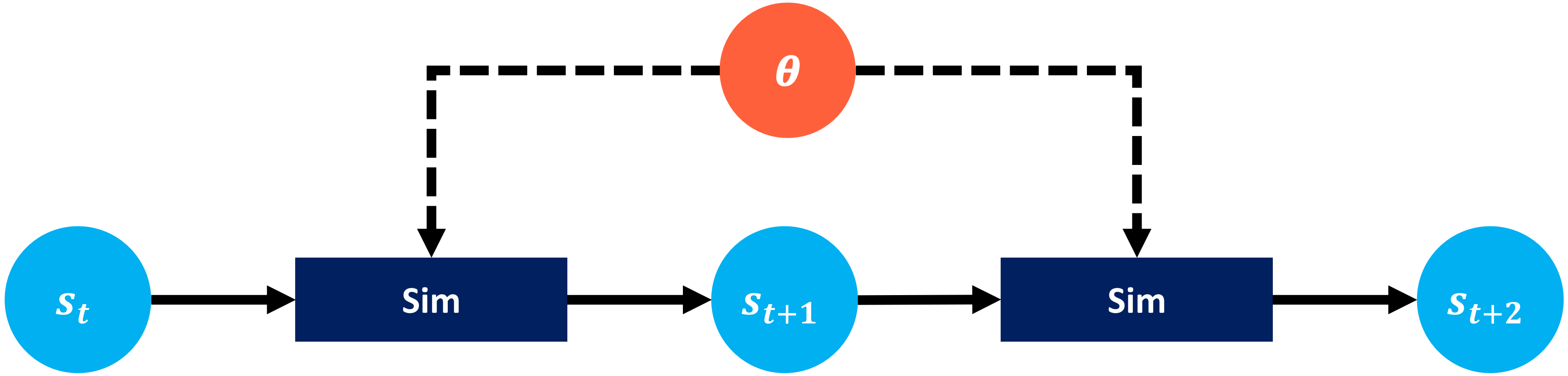
$$x_2^{t+1} = x_2^t + v_2' dt$$

Physical parameters to be optimized: m, e

	Mass (m)	Restitution (e)	…
Obj1	0.94	0.78	
Obj2	0.76	0.95	
…			

Forward Simulations Using Differentiable Physics

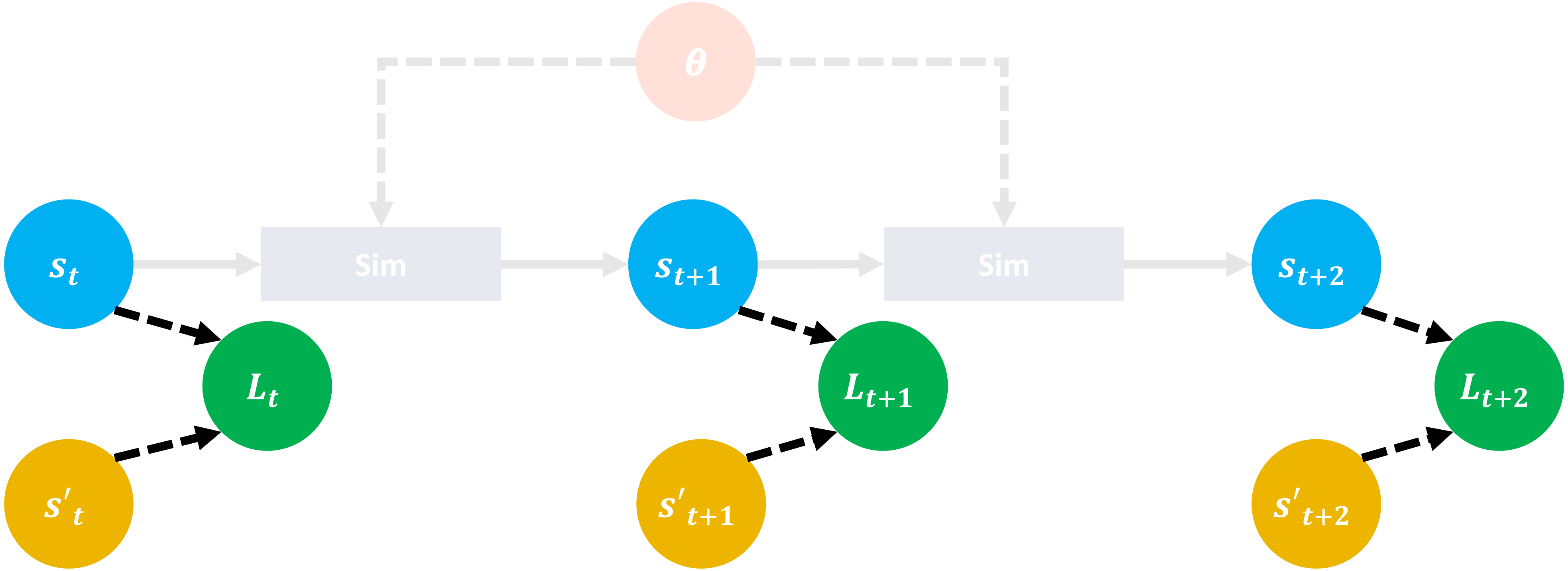
- Forward simulation



Physical parameters $\boldsymbol{\theta}$: e.g., mass and restitution of rigid bodies
State s: e.g., positions and velocities of rigid bodies

Parameters Estimation with Differentiable Physics

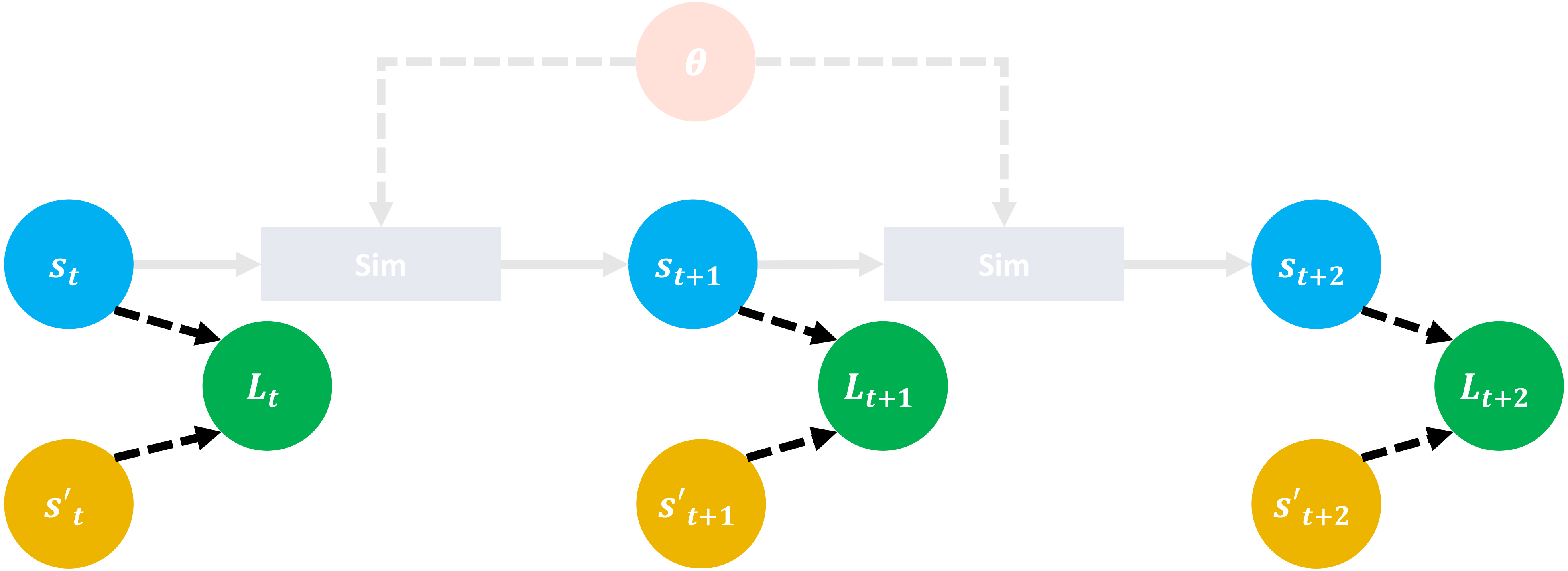
- Loss computation



- Physical parameters $\boldsymbol{\theta}$: e.g., mass and restitution of rigid bodies
- State s: e.g., positions and velocities of rigid bodies
- Observation s': e.g., raw observations (positions) of rigid bodies
- Loss L: the distance between the state and the visual observation

Parameters Estimation with Differentiable Physics

- Minimize the loss by inferring physical paramter $\boldsymbol{\theta}$

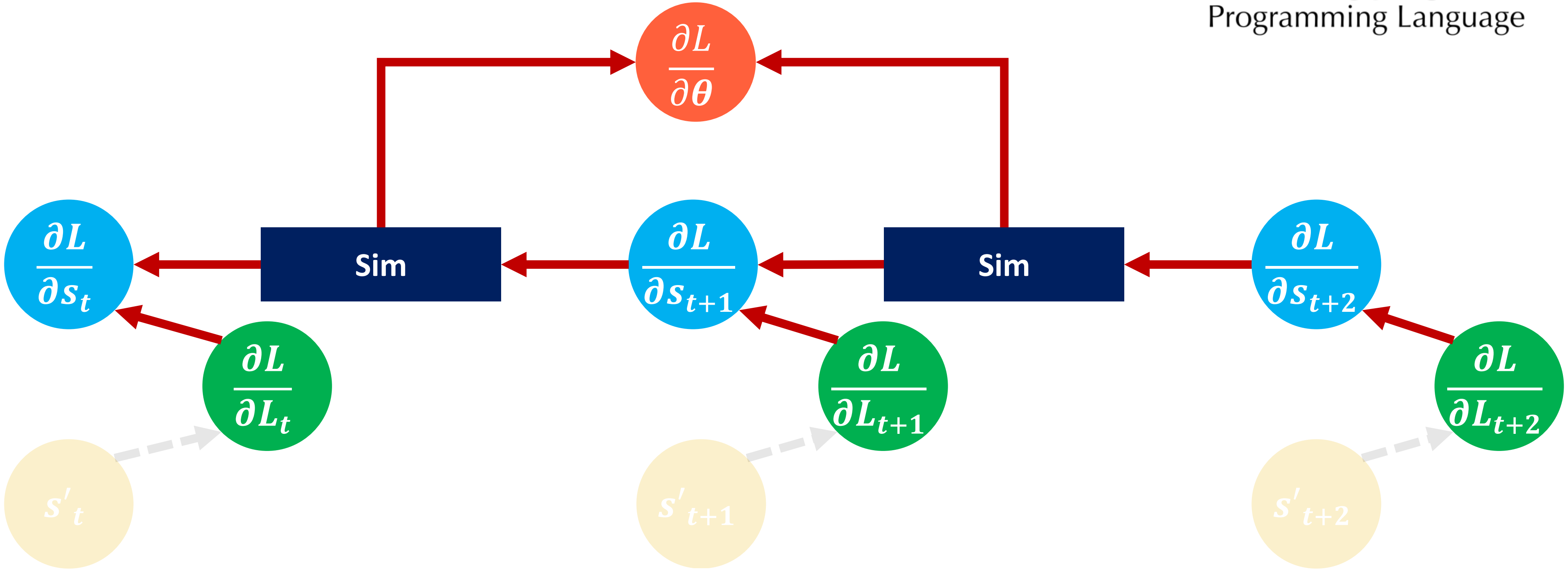


$$\boldsymbol{L = L_1 + L_2 + \cdots + L_T}$$

- Physical parameters $\boldsymbol{\theta}$: e.g., mass and restitution of rigid bodies
- State s: e.g., positions and velocities of rigid bodies
- Observation s': e.g., raw observations (positions) of rigid bodies
- Loss L: the errors between the state and the visual observation

Parameters Estimation with Differentiable Physics

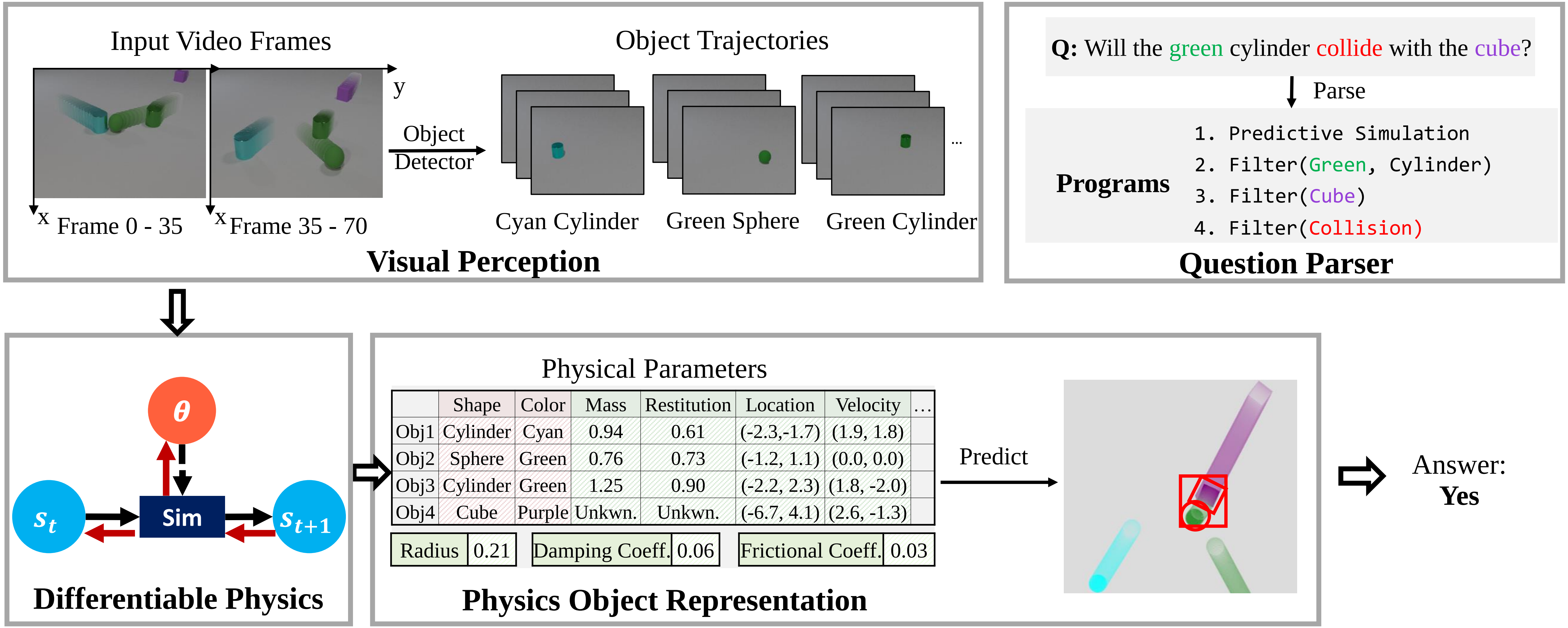
- Backward



$$L = L_1 + L_2 + \cdots + L_T$$

- Physical parameters $\boldsymbol{\theta}$: e.g., mass and restitution of rigid bodies
- State s: e.g., positions and velocities of rigid bodies
- Observation s': e.g., raw observations (positions) of rigid bodies
- Loss L: the error between the state and the visual observation

Neuro-symbolic Dynamics Reasoning *with* physics

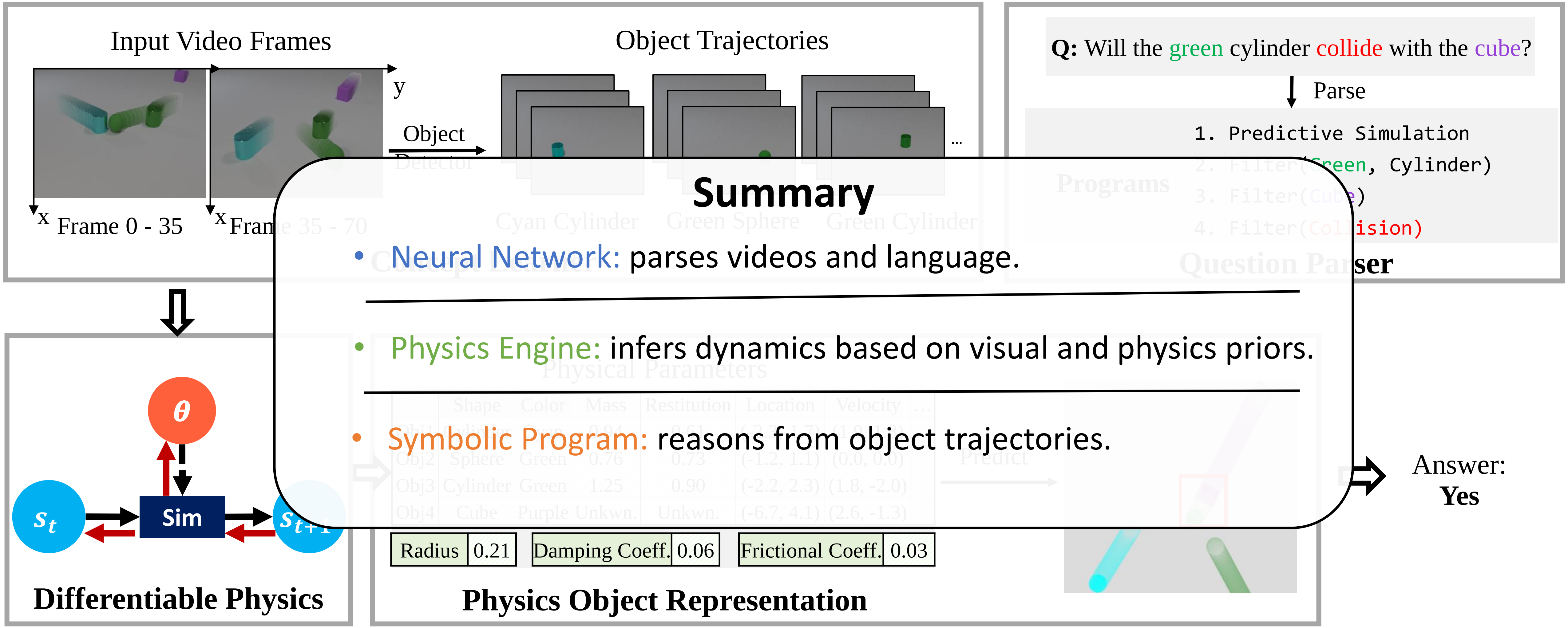


	Shape	Color	Mass	Restitution	Location	Velocity	…
Obj1	Cylinder	Cyan	0.94	0.61	(-2.3,-1.7)	(1.9, 1.8)	
Obj2	Sphere	Green	0.76	0.73	(-1.2, 1.1)	(0.0, 0.0)	
Obj3	Cylinder	Green	1.25	0.90	(-2.2, 2.3)	(1.8, -2.0)	
Obj4	Cube	Purple	Unkwn.	Unkwn.	(-6.7, 4.1)	(2.6, -1.3)	

Radius	0.21	Damping Coeff.	0.06	Frictional Coeff.	0.03

Dynamic Visual Reasoning by Learning Differentiable Physics Models from Video and Language. Ding, Chen, Du, Luo, Tenenbaum, **Gan**. NeurIPS’21

Neuro-symbolic Dynamics Reasoning *with* physics



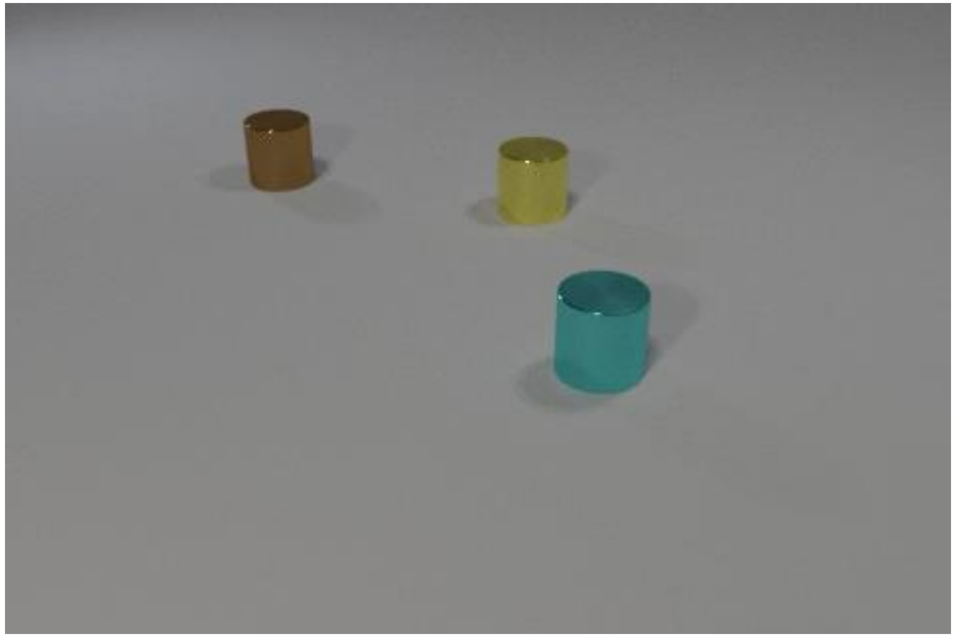
Summary

- Neural Network: parses videos and language.
- Physics Engine: infers dynamics based on visual and physics priors.
- Symbolic Program: reasons from object trajectories.

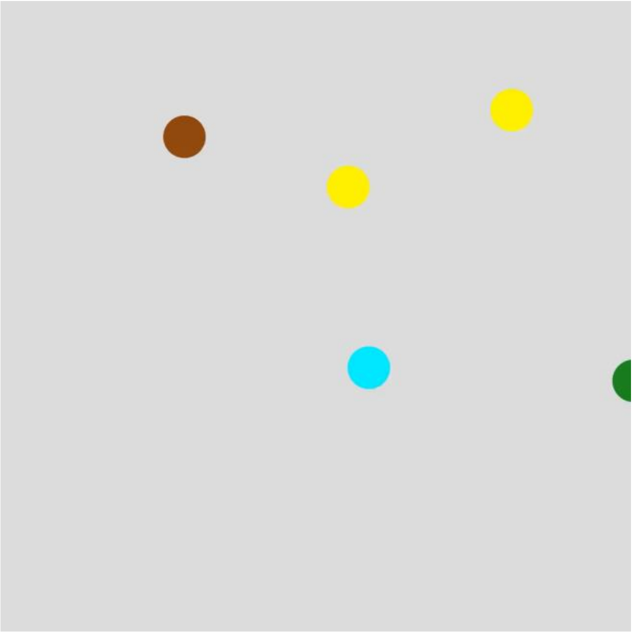
Dynamic Visual Reasoning by Learning Differentiable Physics Models from Video and Language. Ding, Chen, Du, Luo, Tenenbaum, **Gan**. NeurIPS’21

Visualization of the Rconstructed Physics Model

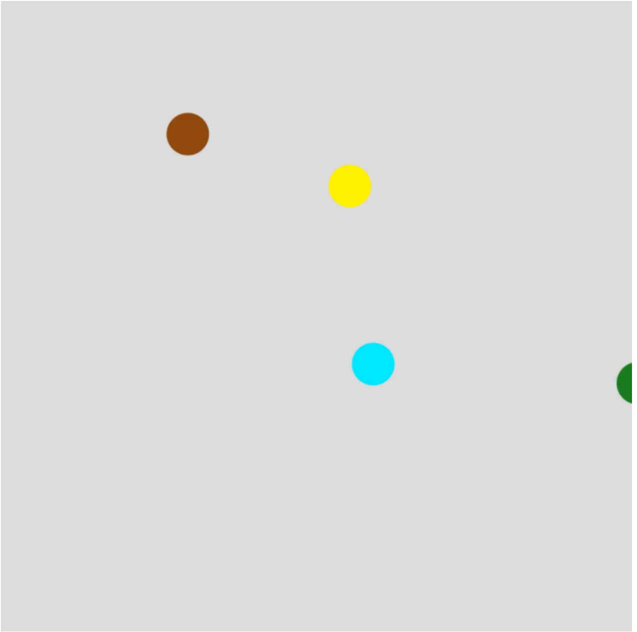
- Results: Fitting the optimized simulation to video observations



Video observations



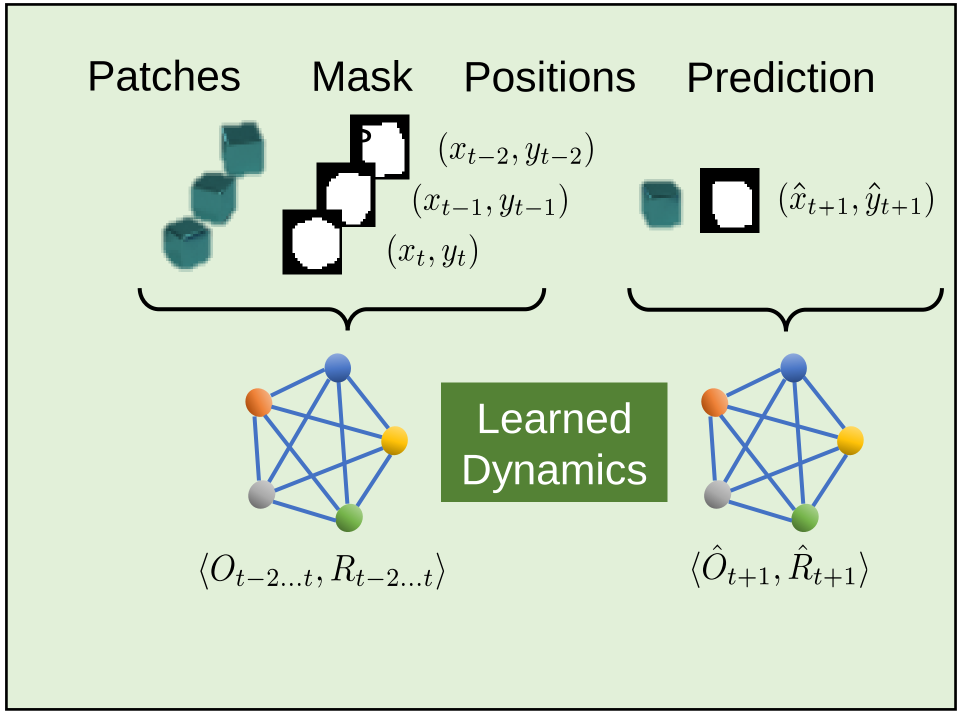
Random initialization



Reconstructed scene

Why Not Learned dynamics Models?

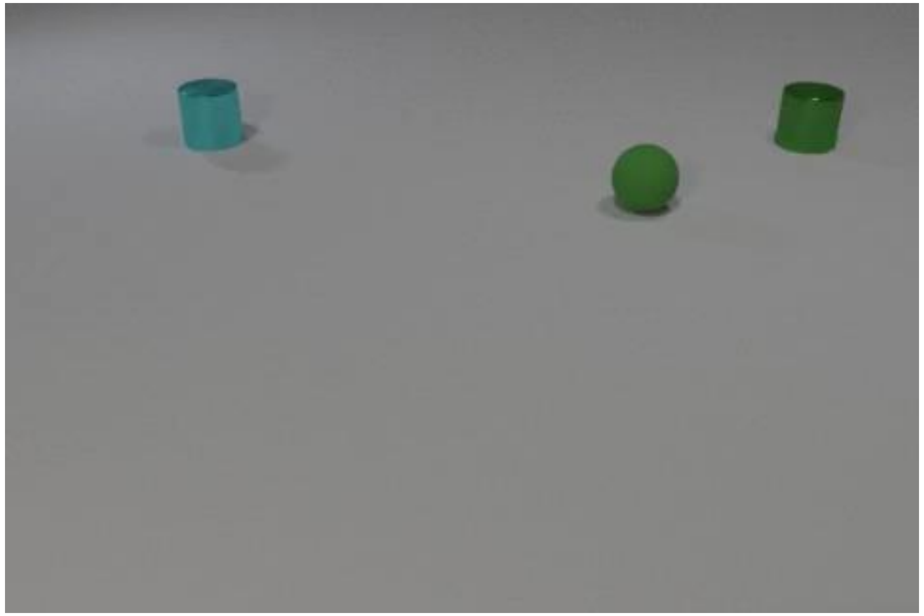
- Input:
  - Nodes: object patches and positions over a time window
  - Edges: collision labels between two objects
- Output:
  - Object positions at the next step
  - Collision labels between two objects



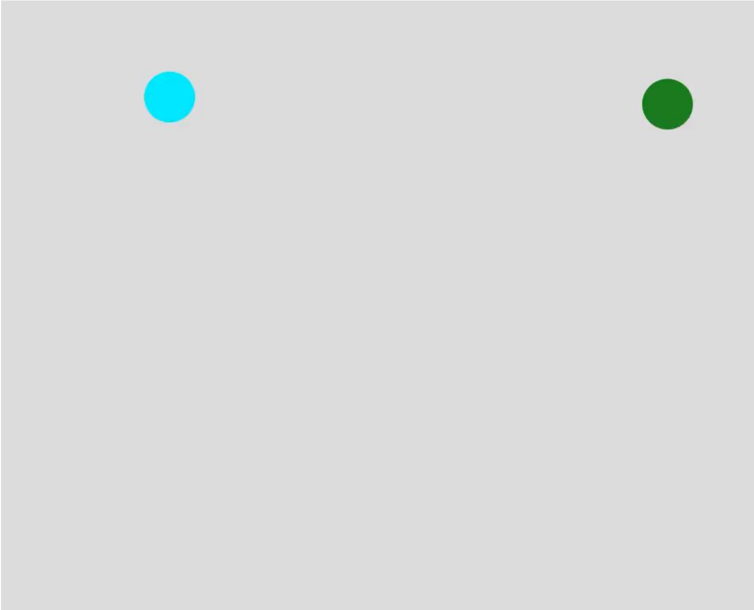
Shi et al. Interaction networks, 2015. Mrowca, et al. HRN, 2018. Li. DPI, 2019.

Counterfactual Dynamics Rollout

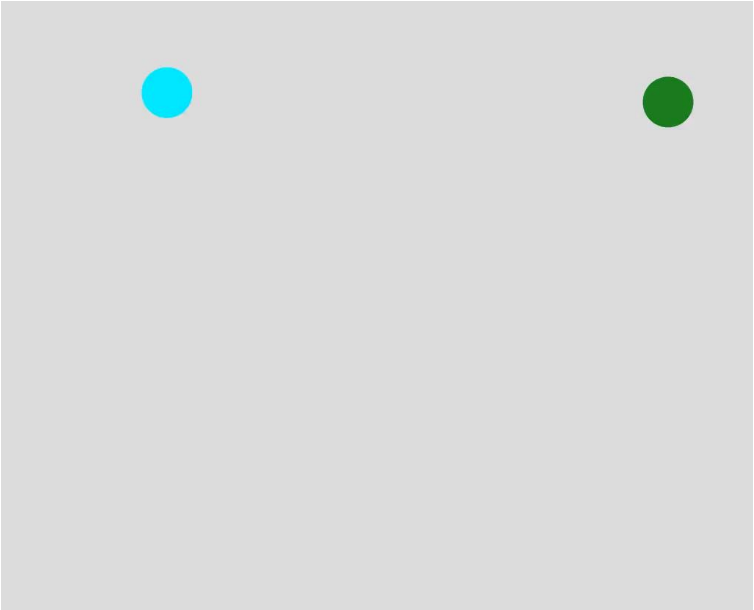
- Remove the green sphere
  - Q: Would the green cylinder collide with the cyan object if the green sphere is removed?
  - A: True



Video observations

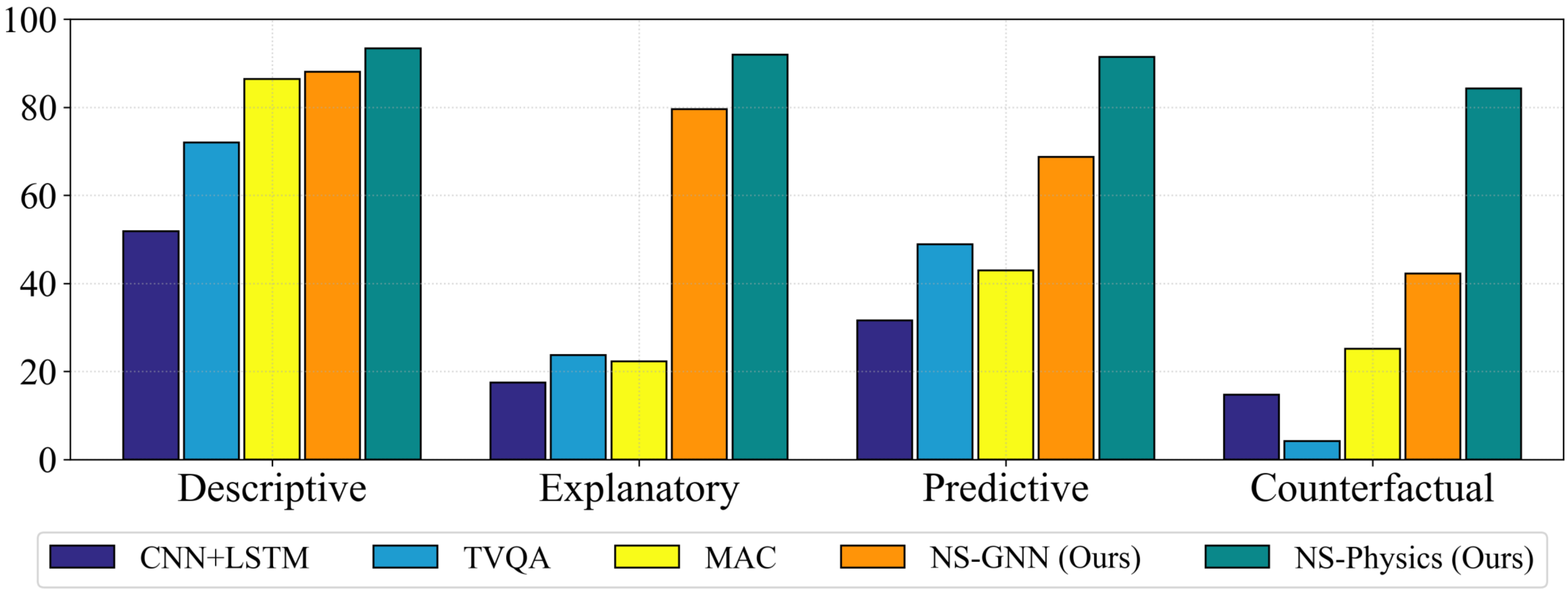


GNNs (×)



Physics Model (√)

CLEVRER Test Accuracy

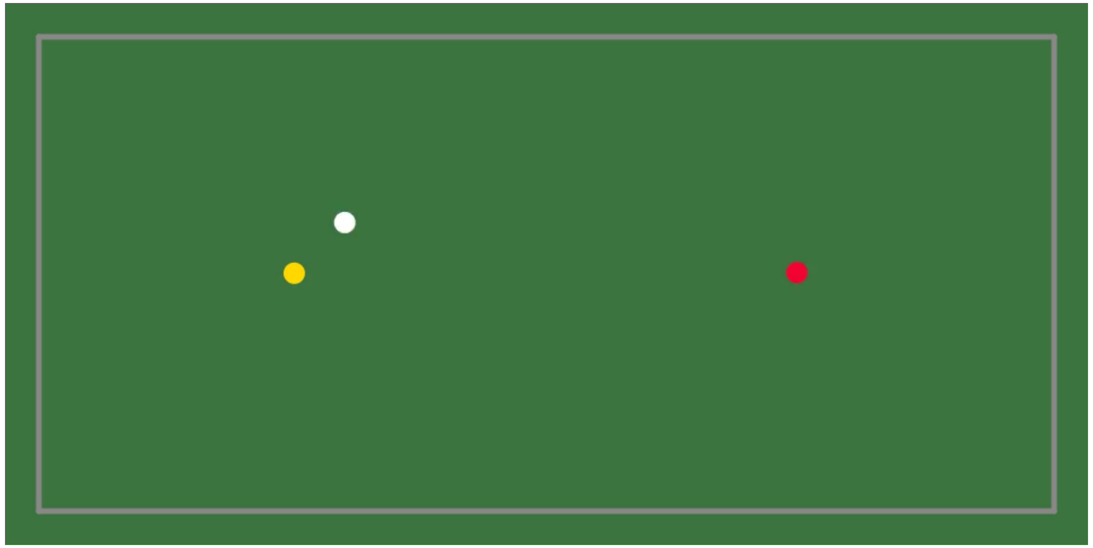


Shi et al. LSTM, 2015. Lei et al. TVQA, 2018. Hudson & Manning. MAC, 2018.
Yi et al. NS-GNN, 2020. Ding et al. NS-Physics, 2021.

Dynamics Visual Reasoning on Real-Billiard

- Estimate the physics models of billiards from **a single video.**

Real world video observations



Predictive Simulation

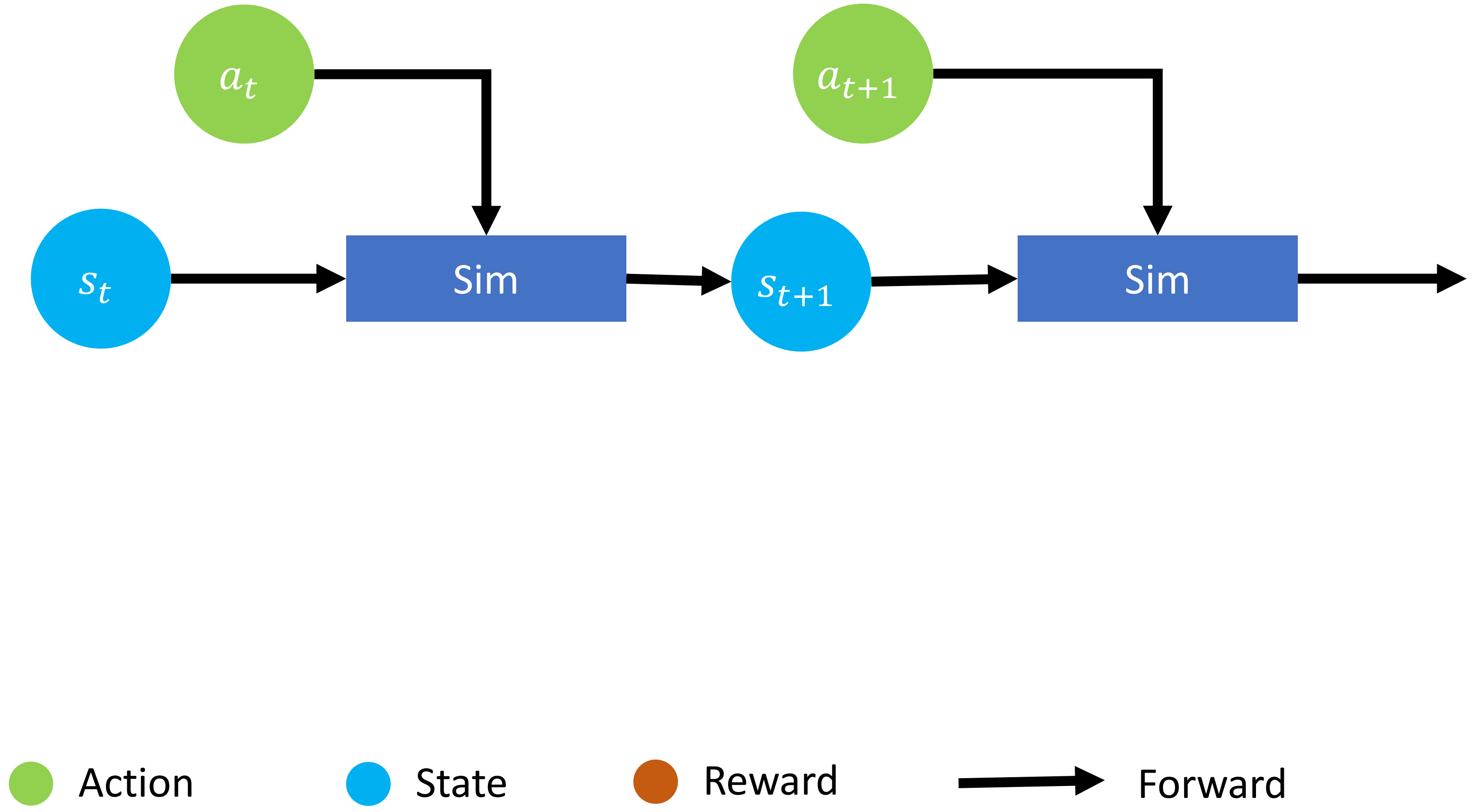
Application: Invert Physics for Real-world Quadrotor

- Imitate the trajectory in a real-world video clip where the **underlying physics are unknown.**

RISP: Rendering-Invariant State Predictor with Differentiable Simulation and Rendering for Cross-Domain Parameter Estimation, Ma*, Du*, Matusik, Tenenbaum, **Gan**. ICLR’22

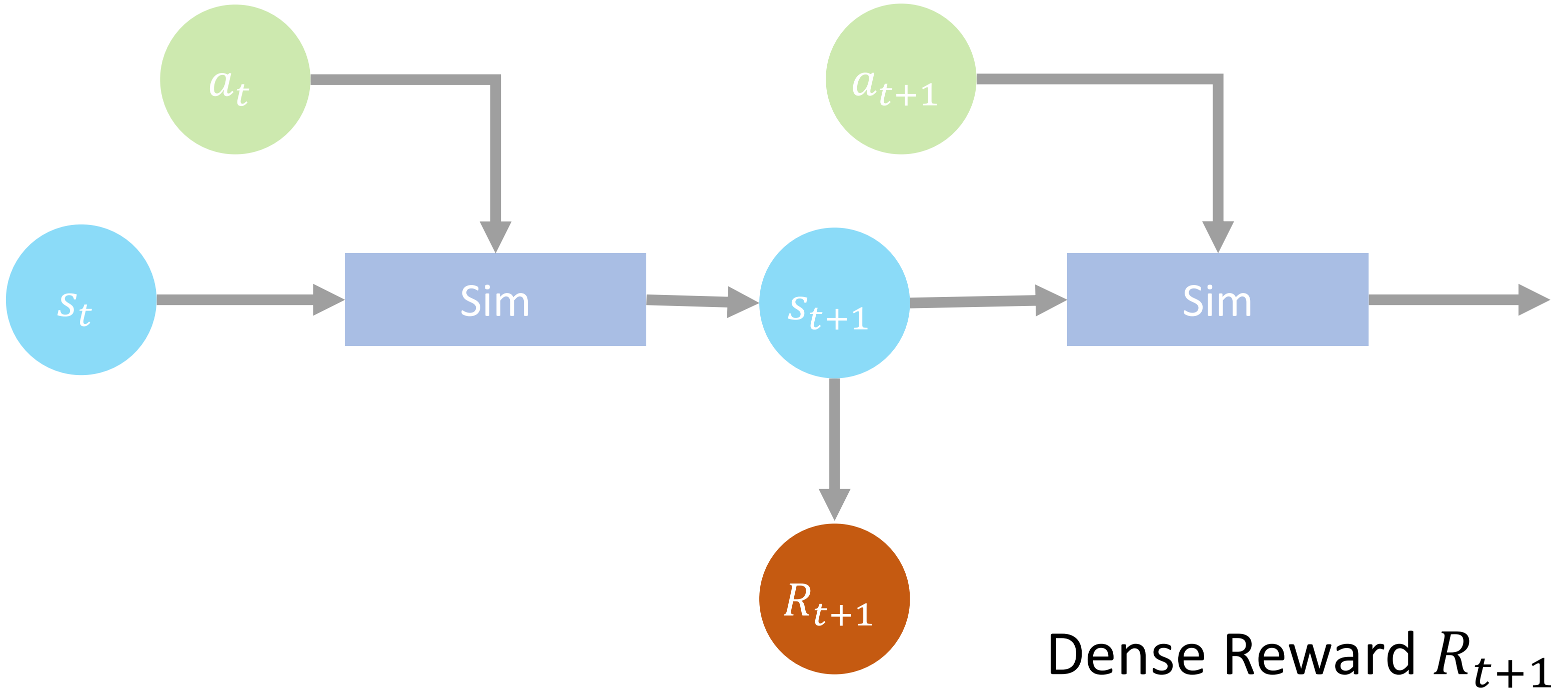
Differentiable Physics for Planning

- Forward simulation with action

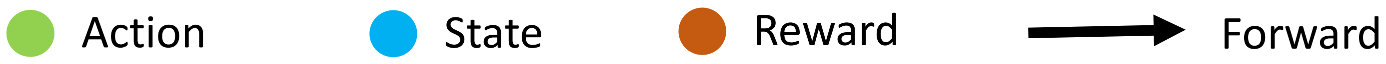


Differentiable Physics for Planning

- Reward

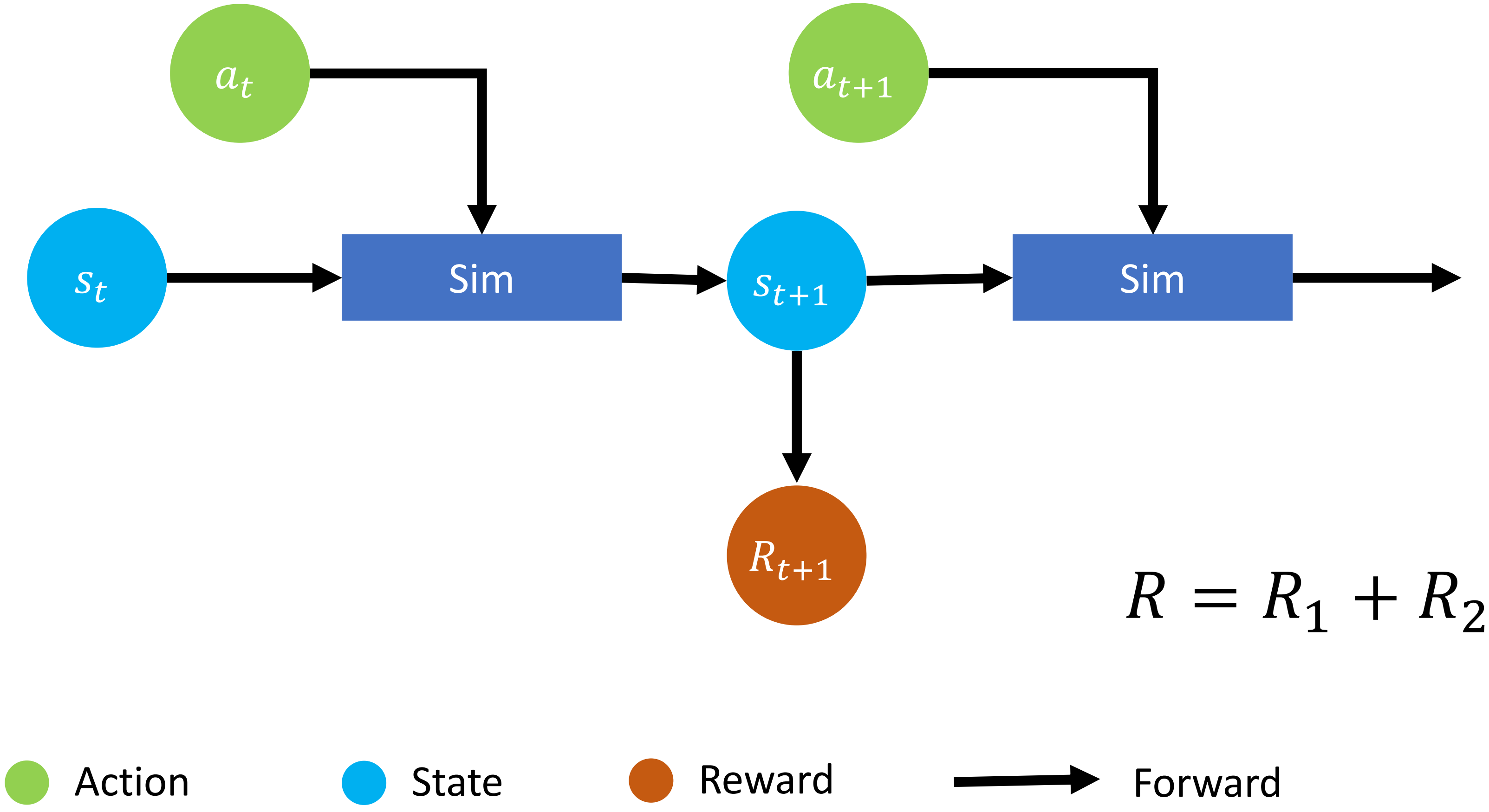


Dense Reward $R_{t+1}(s_{t+1})$



Differentiable Physics for Planning

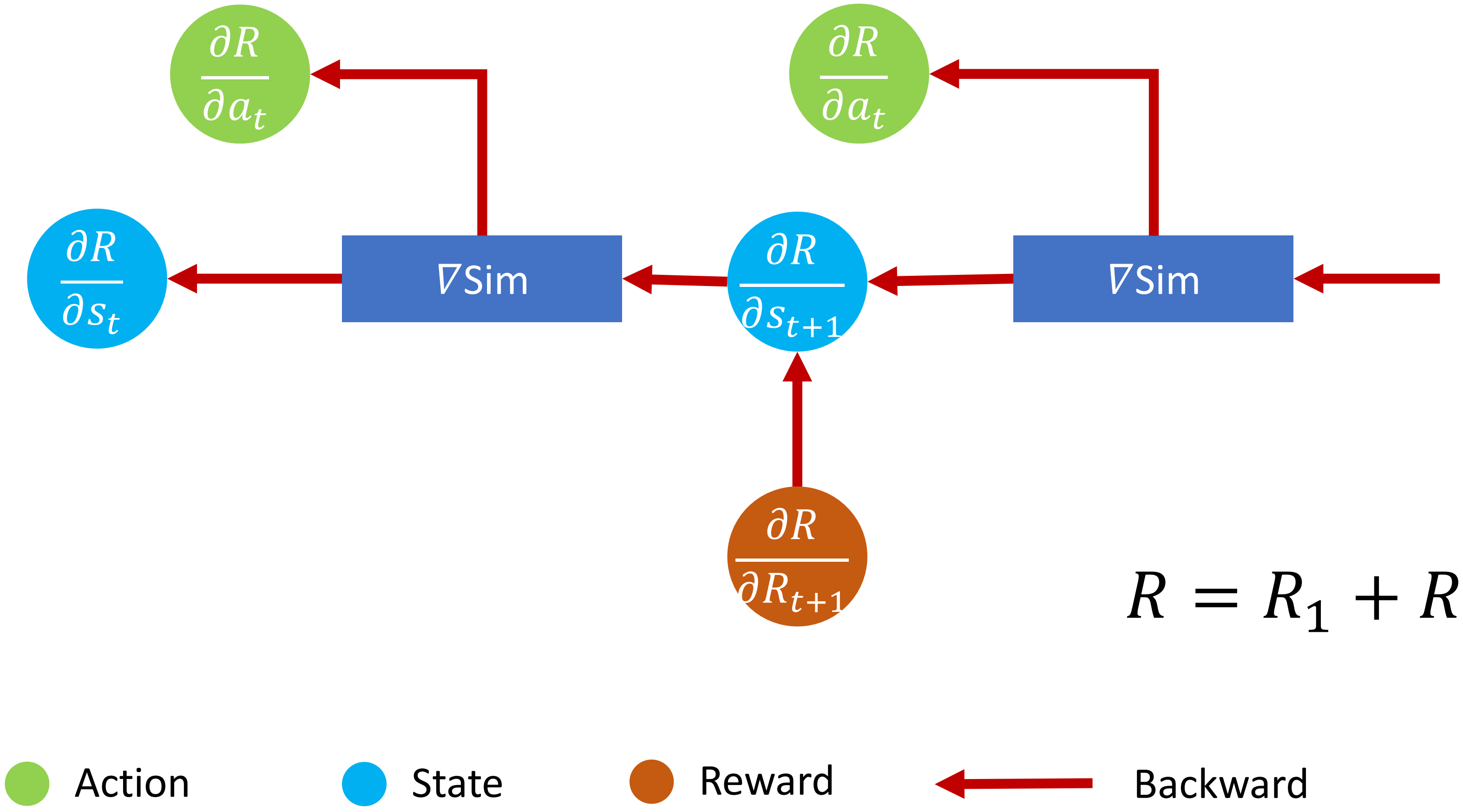
- Maximize total reward R by choosing $a_1, a_2, \dots, a_{T-1}$



$$R = R_1 + R_2 + \cdots + R_T$$

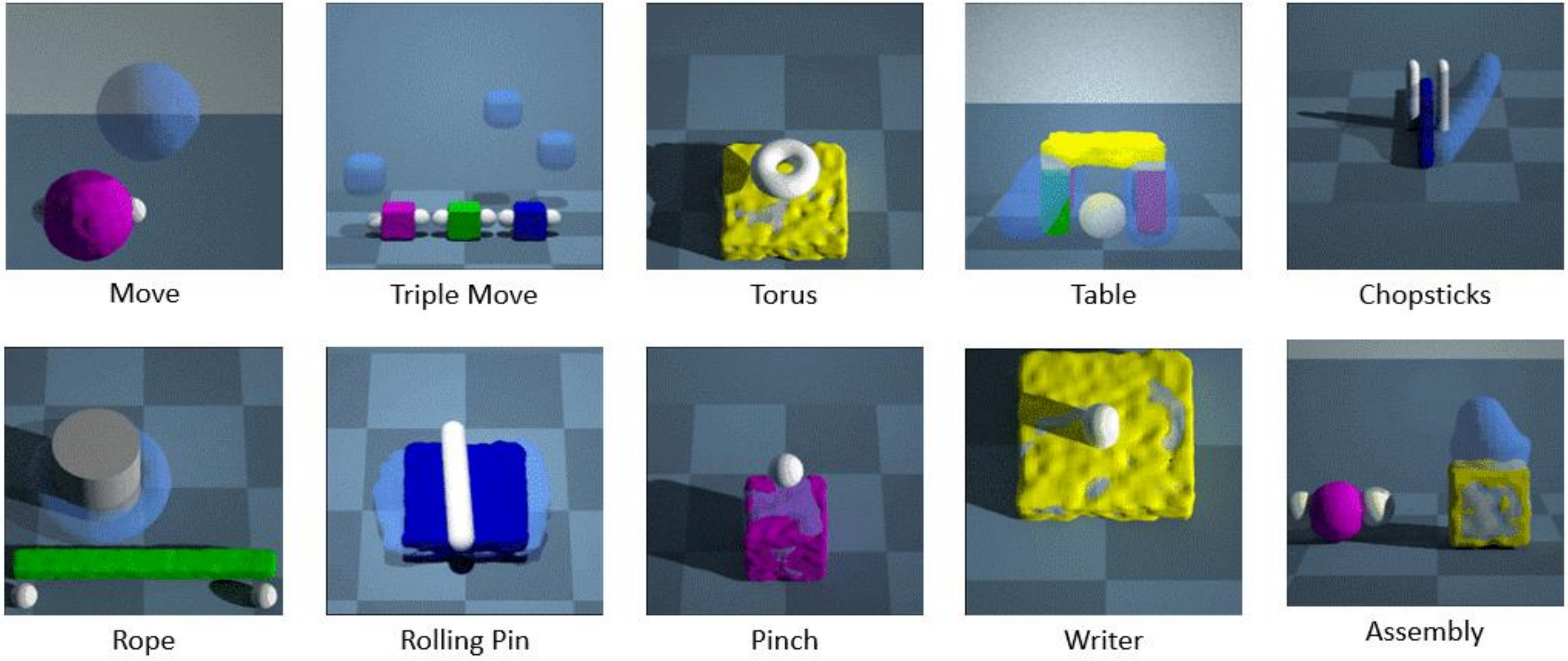
Differentiable Physics for Planning

- Backward



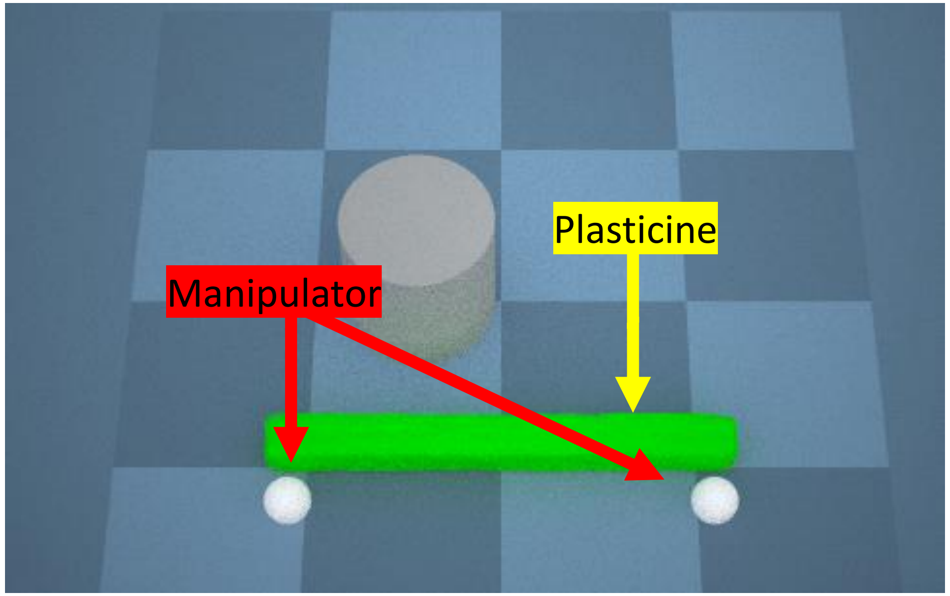
$$R = R_1 + R_2 + \cdots + R_T$$

Applications: Soft-body Manipulation

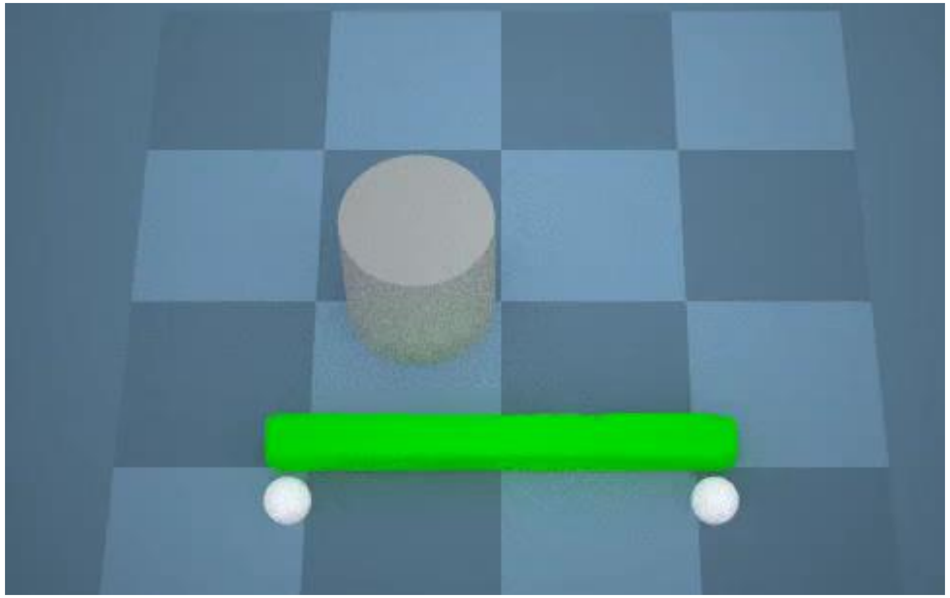


PlasticineLab: A Soft-Body Manipulation Benchmark with Differentiable Physics. Huang ,Hu, Du, Zhou, Su, Tenenbaum, **Gan**. ICLR’21

Example: Rope Task



Initial State

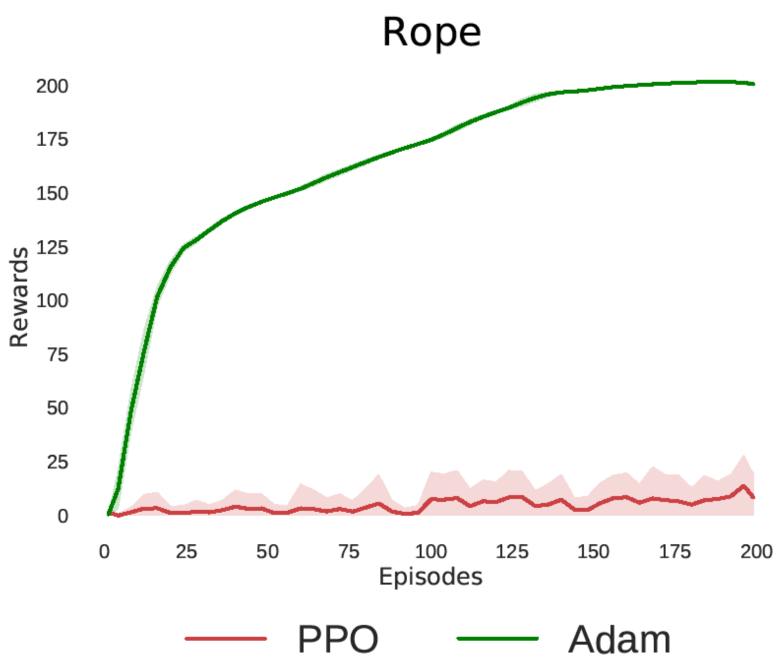


Deform

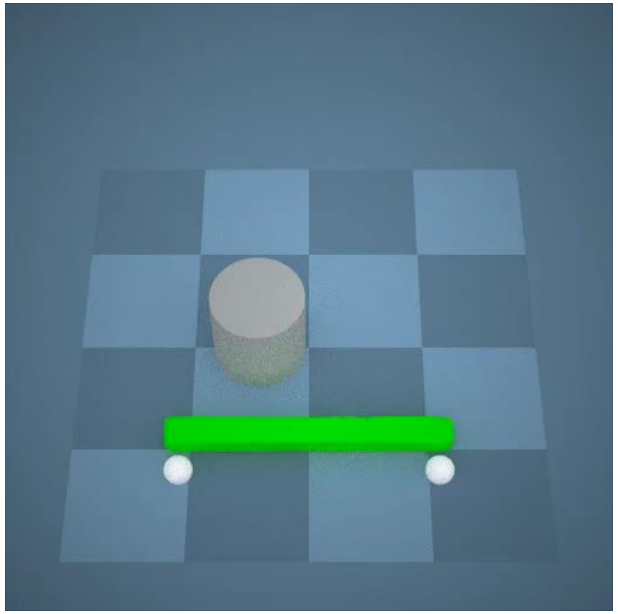
77

Evaluation Results

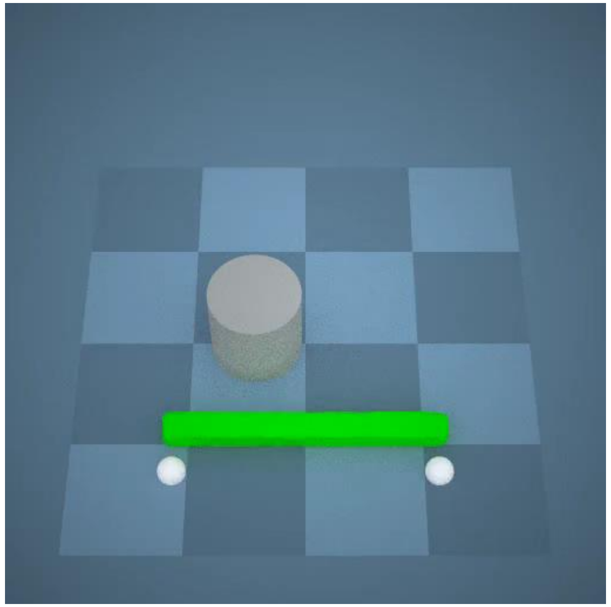
- Gradient-based optimization is much efficient than RL.



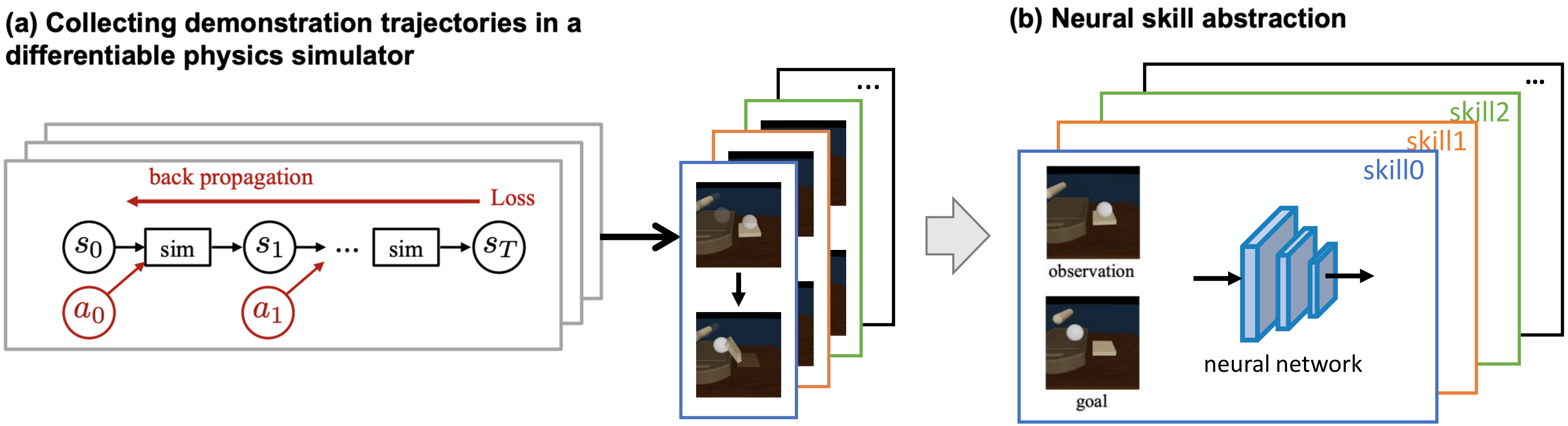
Adam Episode 200



PPO Episode 10K

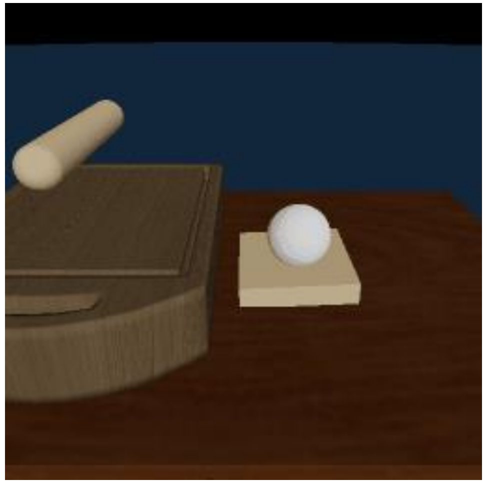


Abstract Skills Using Neural Networks



Rearrange and lift

Lift and spread



DiffSkill: Skill Abstraction from Differentiable Physics for Deformable Object Manipulations with Tools. Li, Huang ,Li, Tenenbaum, **Gan**. ICLR'22

My Vision: Neuro-Symbolic Embodied AI

Neural Network

Extracting complex pattens from raw sensory

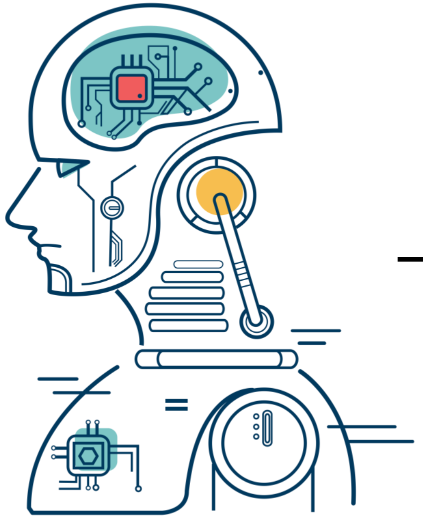
Symbolic Program

Representing and reasoning from abstract knowledge

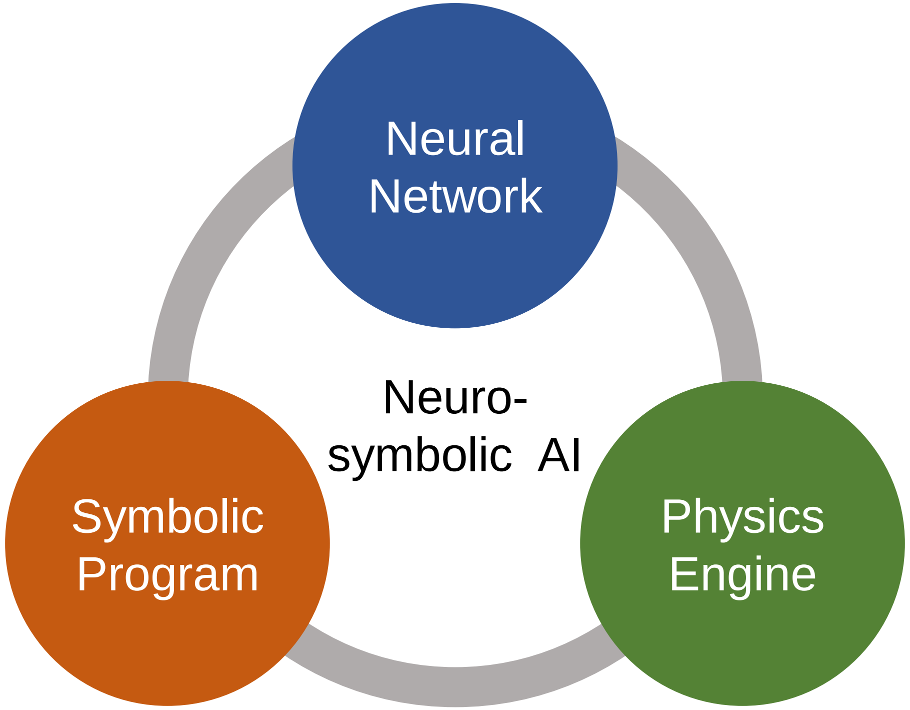
Physics Engine

Physical inference and planning

My Vision: Neuro-Symbolic AI



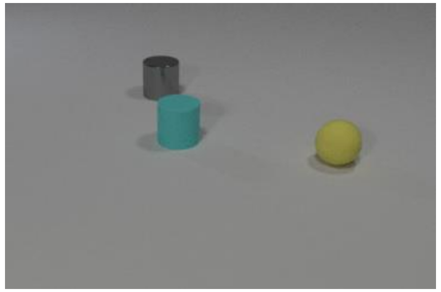
Machine



Carolyn

Scene Understanding

Dynamics Reasoning



Physical Interaction