

Neurosymbolic AI

David Cox, Ph.D.
IBM Director, MIT-IBM Watson AI Lab
IBM Research



“Artificial Intelligence”

The evolution of AI

Narrow AI
Emerging

Broad AI
Disruptive and
Pervasive

General AI
Revolutionary

The evolution of AI

Narrow AI

Single task, single domain
Superhuman accuracy and speed for certain tasks



IBM Research AI © 2018 IBM Corporation

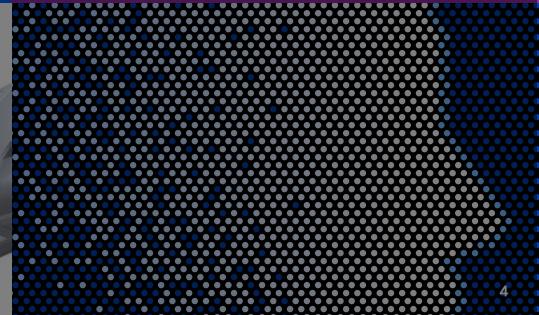
Broad AI

Multi-task, multi-domain
Multi-modal
Distributed AI
Explainable



General AI

Cross-domain learning and reasoning
Broad autonomy



The evolution of AI

Narrow AI

Single task, single domain
Superhuman accuracy and speed for certain tasks



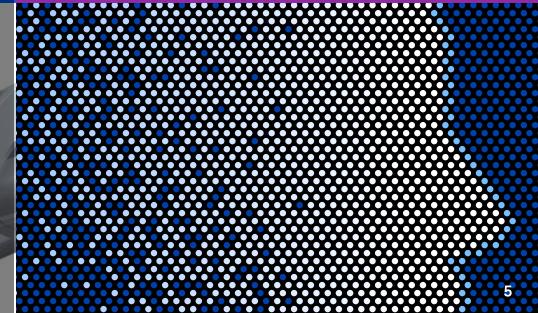
Broad AI

Multi-task, multi-domain
Multi-modal
Distributed AI
Explainable



General AI

Cross-domain learning and reasoning
Broad autonomy



Elon Musk

Elon Musk Compares Building Artificial Intelligence To “Summoning The Demon”

Posted Oct 26, 2014 by [Greg Kumparak \(@grg\)](#)

17.6k
SHARES



Next Story

Technology

Stephen Hawking warns artificial intelligence could end mankind

By Rory Cellan-Jones
Technology correspondent

⌚ 2 December 2014 | [Technology](#) |

The evolution of AI

Narrow AI

Single task, single domain
Superhuman accuracy and speed for certain tasks



IBM Research AI © 2018 IBM Corporation

Broad AI

Multi-task, multi-domain
Multi-modal
Distributed AI
Explainable



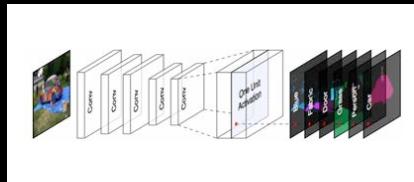
General AI

Cross-domain learning and reasoning
Broad autonomy



The path to a “Broad AI” toolbox

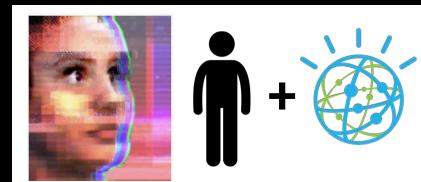
Explainability



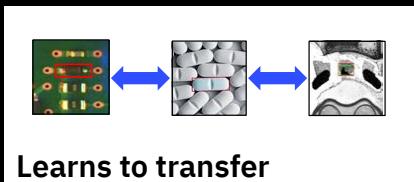
Security



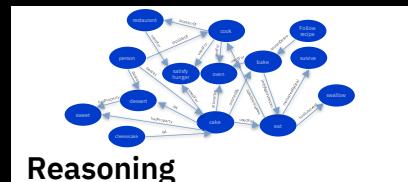
Ethics



Learn more from small data

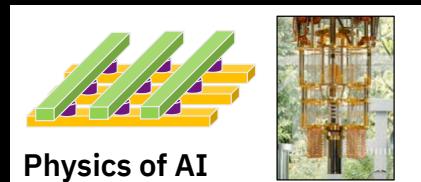


Learns to transfer



Reasoning

Infrastructure



Physics of AI

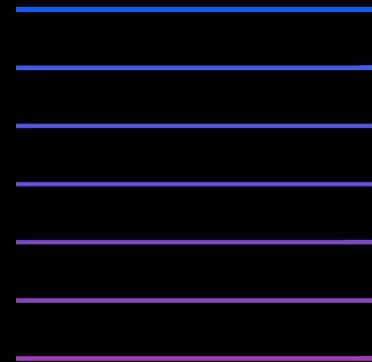
Platform for AI Lifecycle

Compute

Data & Models

Applications

Workflow



**MIT-IBM
WATSON
AI LAB**

The evolution of AI

General AI
Revolutionary

Broad AI
Disruptive and
Pervasive

Narrow AI
Emerging

So what's “narrow” about today's AI toolbox?

DEC 29, 2014 @ 11:37 AM 115,776 

Sell In May & Walk Away: 6 Stocks to Dump

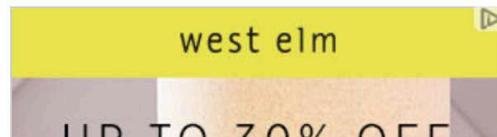
Tech 2015: Deep Learning And Machine Intelligence Will Eat The World



Anthony Wing Kosner, CONTRIBUTOR

Quantum of Content and innovations in user experience [FULL BIO](#) ▾

Opinions expressed by Forbes Contributors are their own.



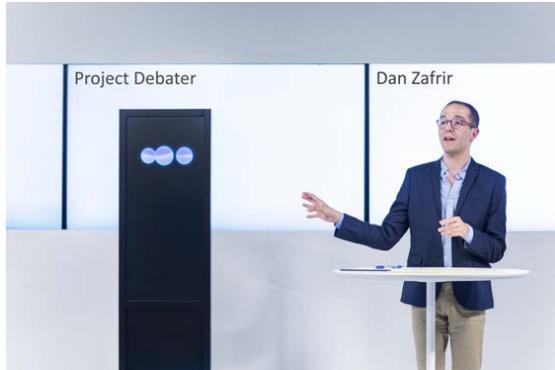


man in black shirt is playing guitar.



construction worker in orange safety vest is working on road.

Karpathy and Li, 2015





Gatys et al. 2015

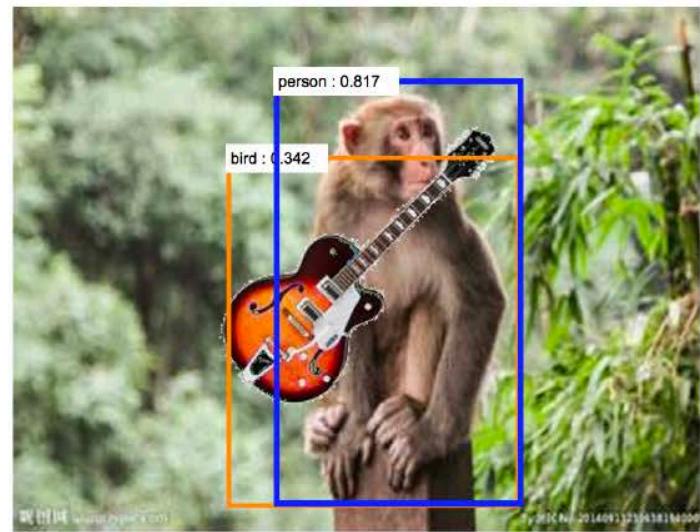
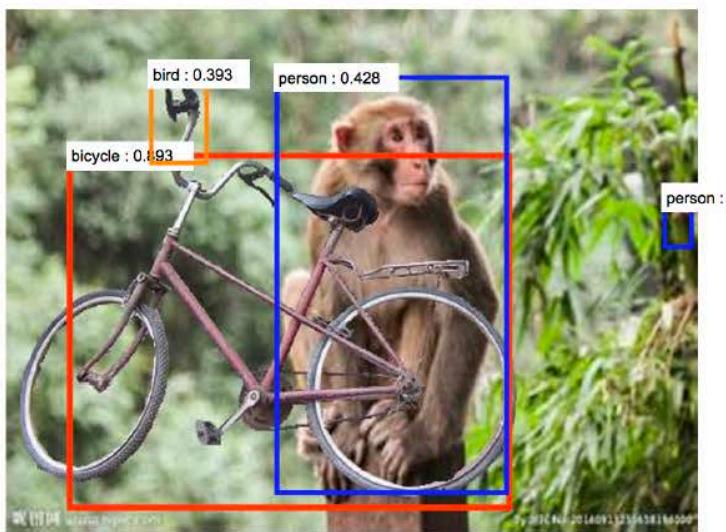


Brock et al. 2018

“Teddy Bear”



Meret Oppenheim, *Le Déjeuner en fourrure*



Wang et al. 2018



man in black shirt is playing guitar.



construction worker in orange safety vest is working on road.

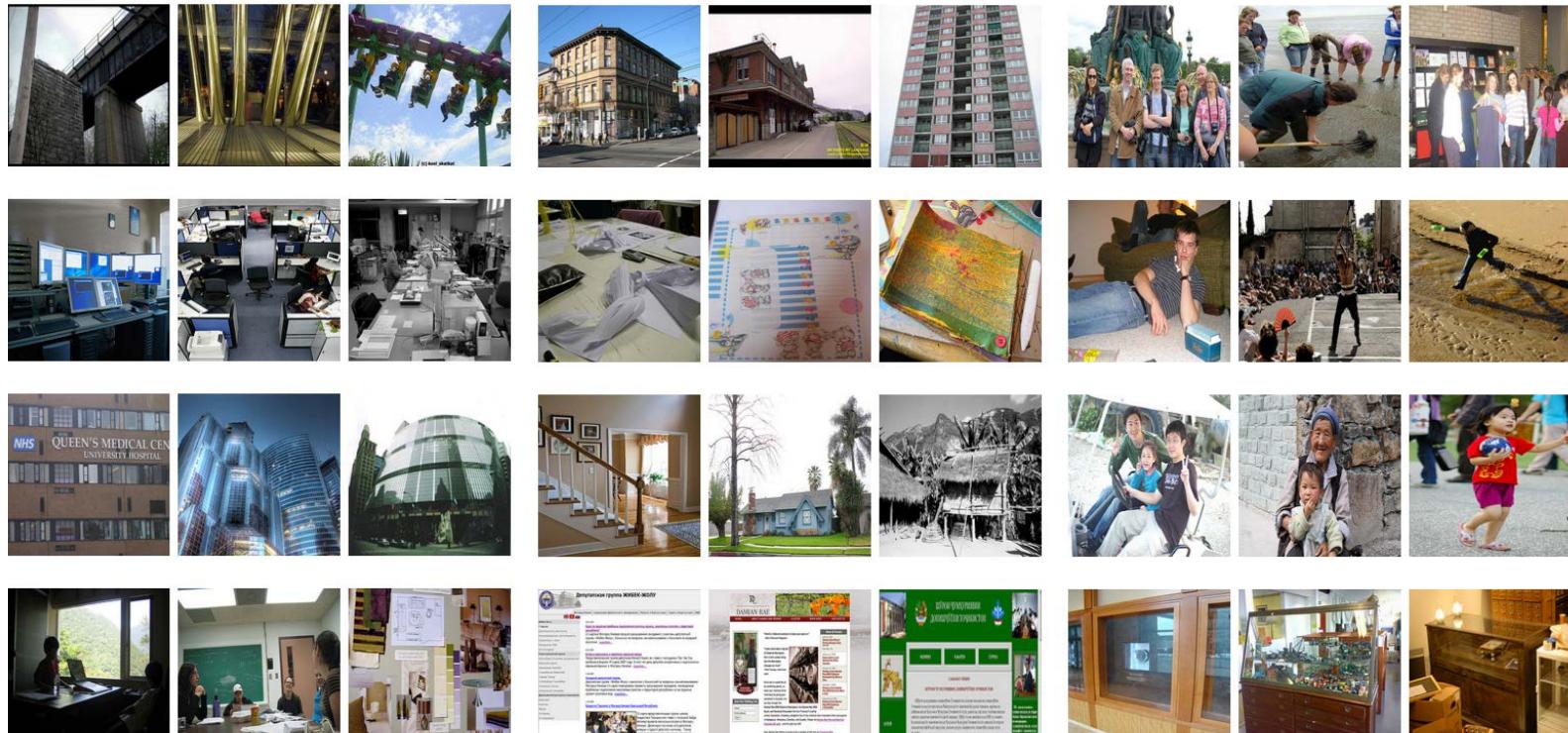
Karpathy and Li, 2015



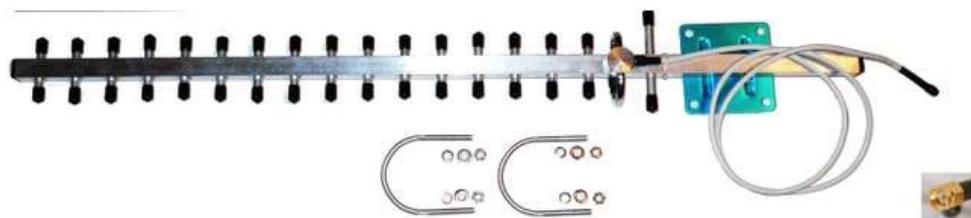
a man riding a
motorcycle on a beach

Lake, Ullman, Tenenbaum & Gershman, 2016

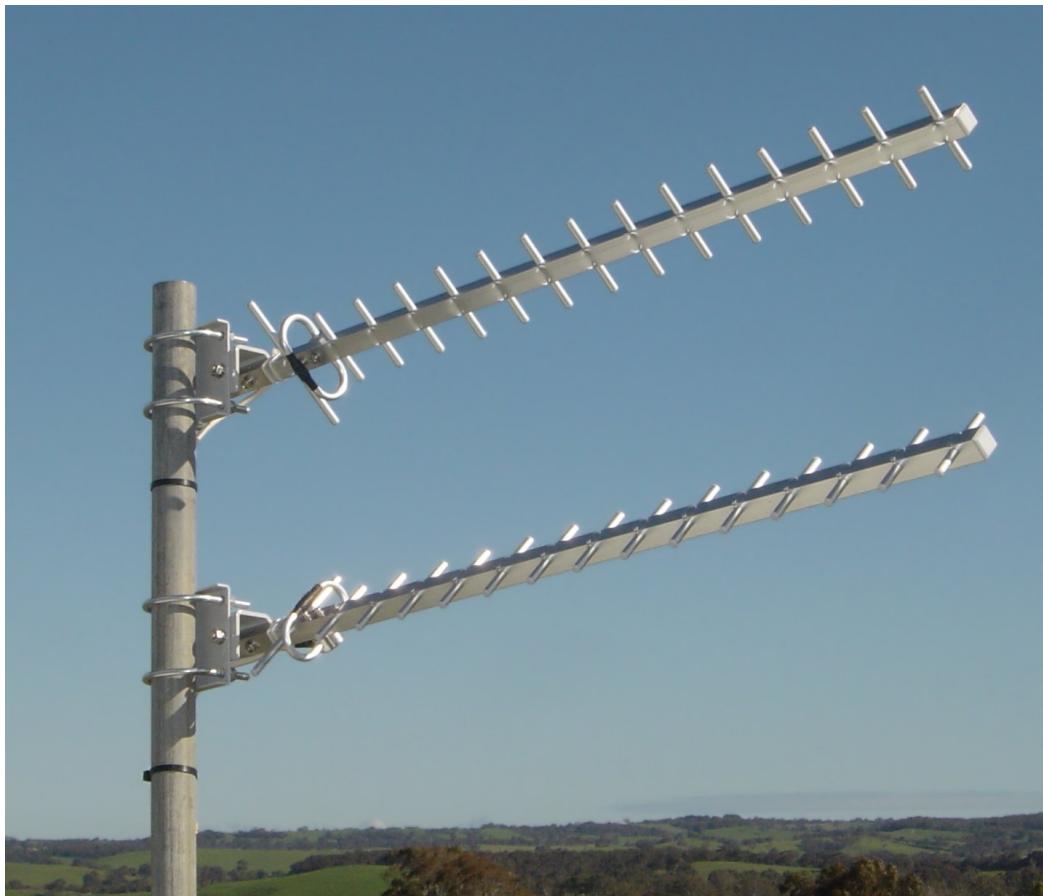
IMAGENET



What's this?









IM[■]GENET



ObjectNet



Boris Katz
MIT

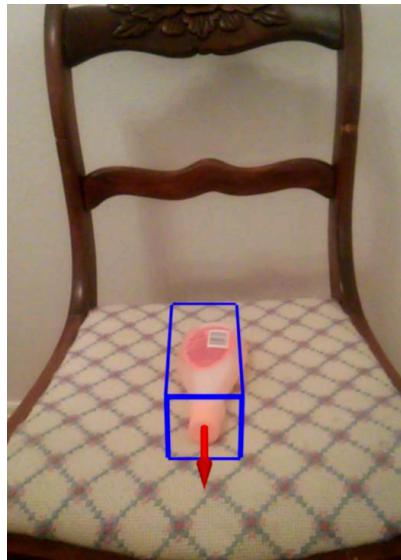
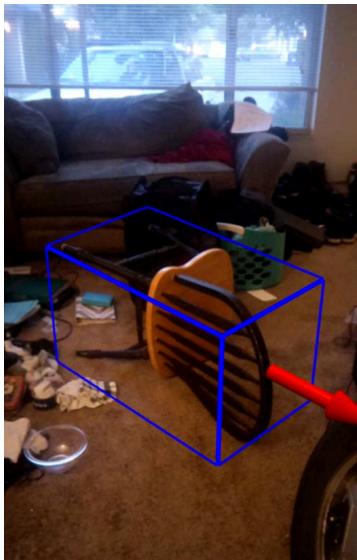


Andrei Barbu
MIT



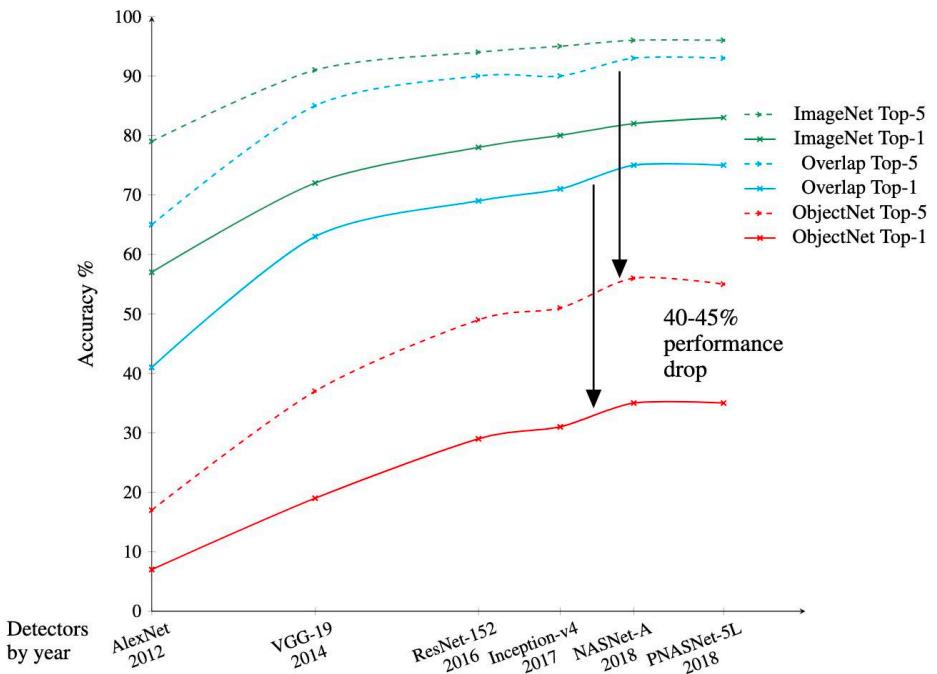
Dan Gutfreund
IBM

ObjectNet



- ~50K images
- ~300 object classes
- 4 different room types

Testing ImageNet-trained models on ObjectNet





Original Top-3 inferred captions:

1. A red stop sign sitting on the side of a road.
2. A stop sign on the corner of a street.
3. A red stop sign sitting on the side of a street.



Pin-yu Chen
IBM



Adversarial Top-3 captions:

1. A brown teddy bear laying on top of a bed.
2. A brown teddy bear sitting on top of a bed.
3. A large brown teddy bear laying on top of a bed.



Xu et al. 2019



How many blocks are on the right of the three-level tower?



Will the block tower fall if the top block is removed?



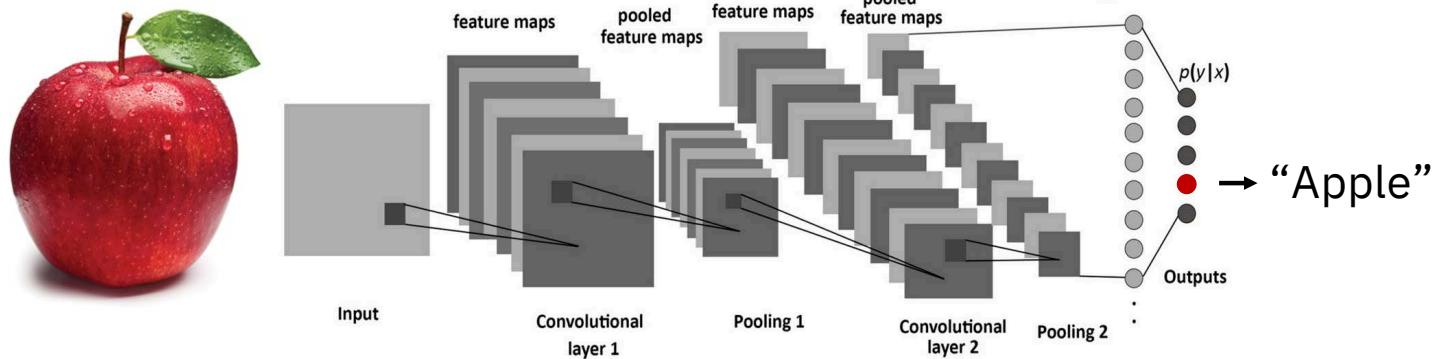
Are there more trees than animals?



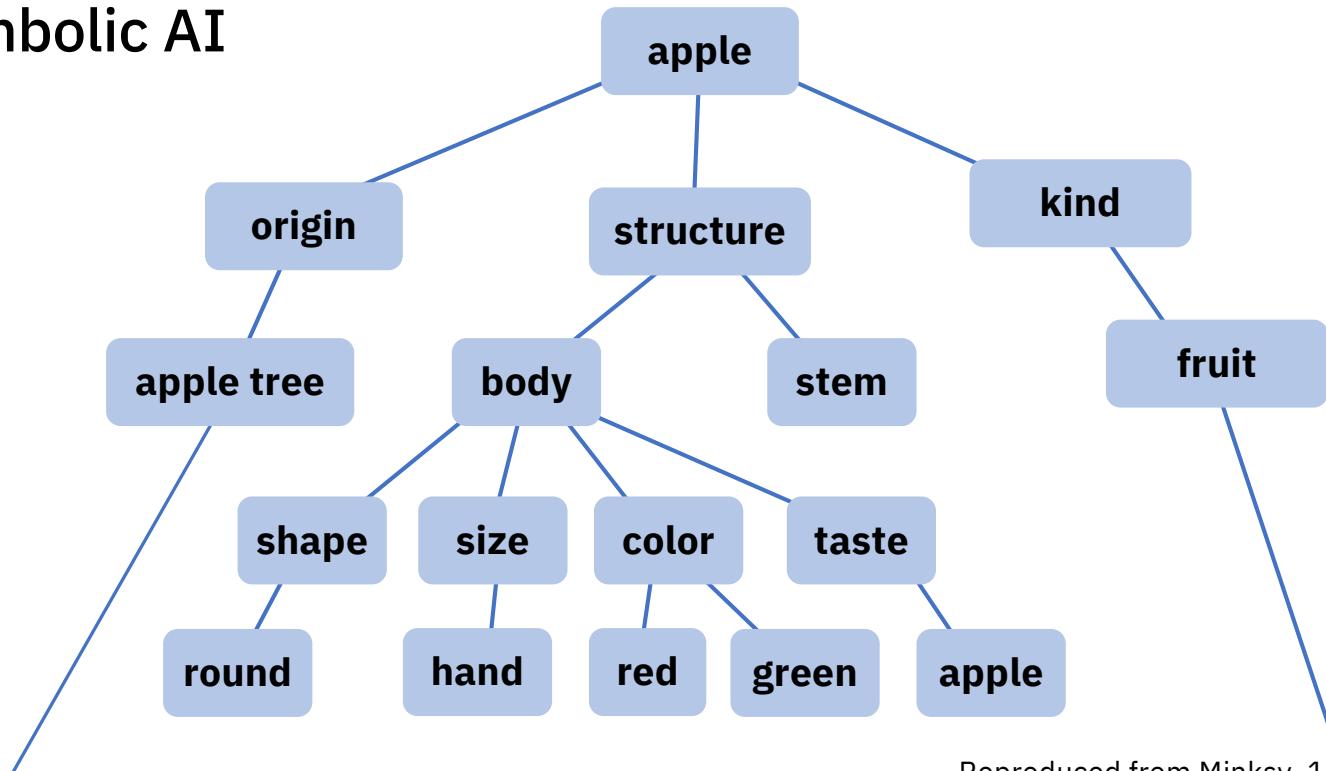
What is the shape of the object closest to the large cylinder?



Neural Networks / Deep Learning



Symbolic AI



Reproduced from Minsky, 1991

Neural-symbolic AI

Disentangling reasoning from vision and language understanding



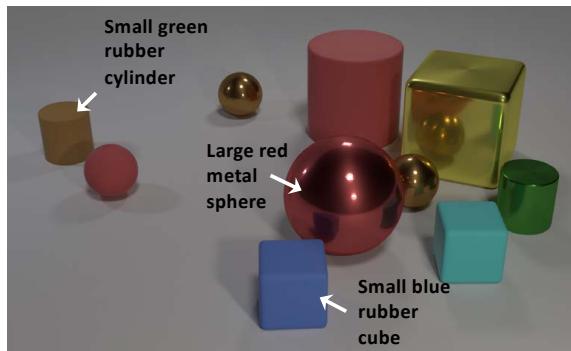
Jiajun Wu



Chuang Gan



Joshua Tenenbaum

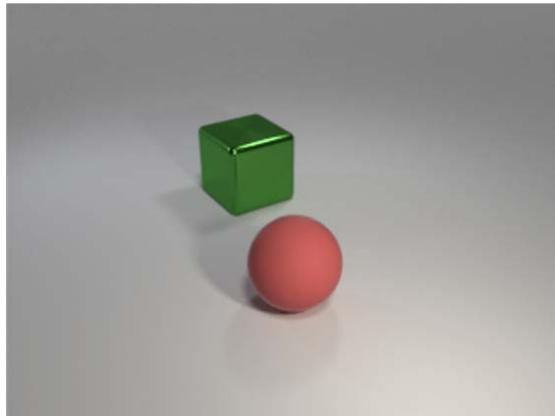


Question: *Are there an equal number of large things and metal spheres?*

Program: `equal_number(count(filter_size(Scene, Large)), count(filter_material(filter_shape(Scene, Sphere), Metal)))`

Answer: Yes

End-to-End Visual Reasoning



Visual Question Answering

Q: What's the shape of the **red** object?



End-to-End
Neural Network

→ A: Sphere.

NMN [Andreas et al., 2016]

IEP [Johnson et al., 2017]

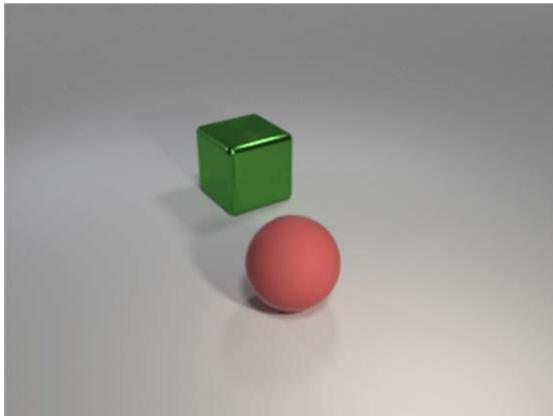
FiLM [Perez et al., 2018],

MAC [Hudson & Manning, 2018]

Stack-NMN [Hu et al., 2018]

TbD [Mascharka et al. 2018]

End-to-End Visual Reasoning



Visual Question Answering

Q: What's the shape of the **red** object?

Concept

(e.g., colors, shapes)

Reasoning

(e.g., count)

NMN [Andreas et al., 2016]

IEP [Johnson et al., 2017]

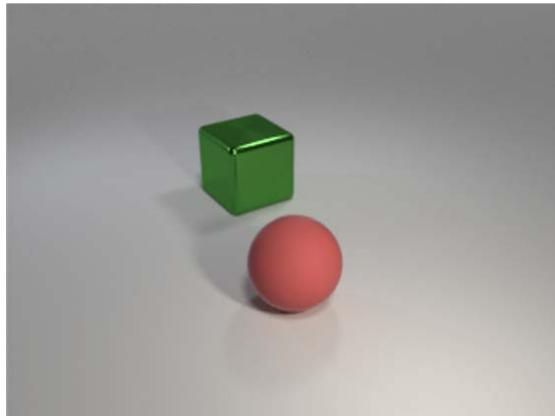
FiLM [Perez et al., 2018],

MAC [Hudson & Manning, 2018]

Stack-NMN [Hu et al., 2018]

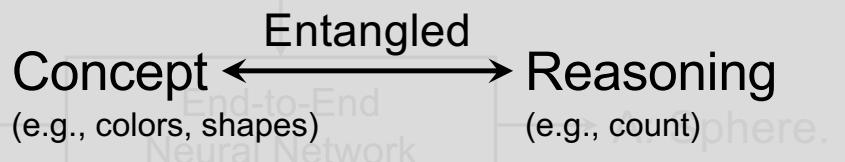
TbD [Mascharka et al. 2018]

End-to-End Visual Reasoning



Visual Question Answering

Q: What's the shape of the **red** object?



NMN [Andreas et al., 2016]

IEP [Johnson et al., 2017]

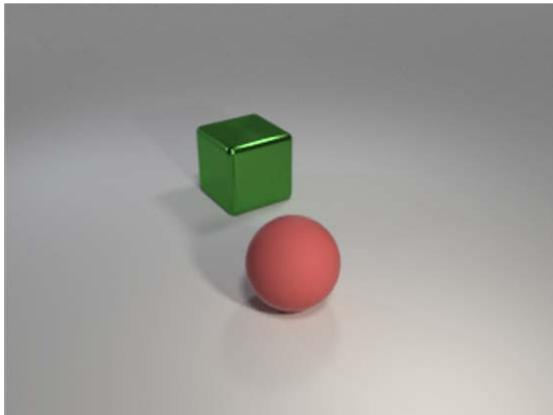
FiLM [Perez et al., 2018],

MAC [Hudson & Manning, 2018]

Stack-NMN [Hu et al., 2018]

TbD [Mascharka et al. 2018]

End-to-End Visual Reasoning



NMN [Andreas et al., 2016]
IEP [Johnson et al., 2017]
FiLM [Perez et al., 2018],
MAC [Hudson & Manning, 2018]
Stack-NMN [Hu et al., 2018]
TbD [Mascharka et al. 2018]

Visual Question Answering

Q: What's the shape of the **red** object?

Concept \longleftrightarrow Reasoning
(e.g., colors, shapes) \longleftrightarrow (e.g., count)

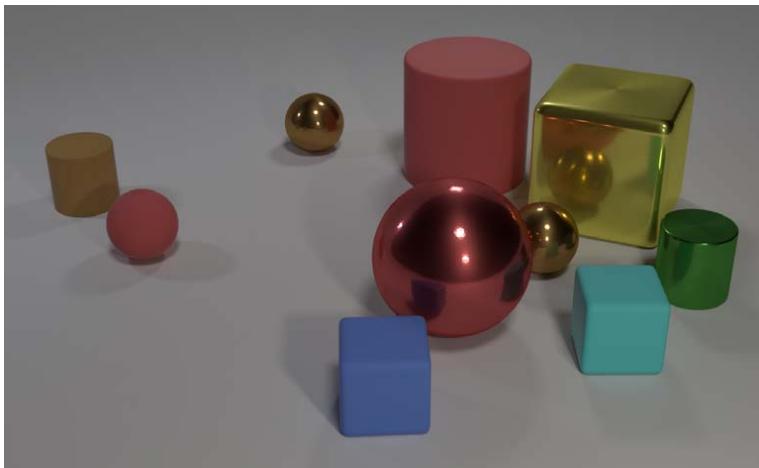
End-to-End Neural Network

Hard to transfer

Image Captioning

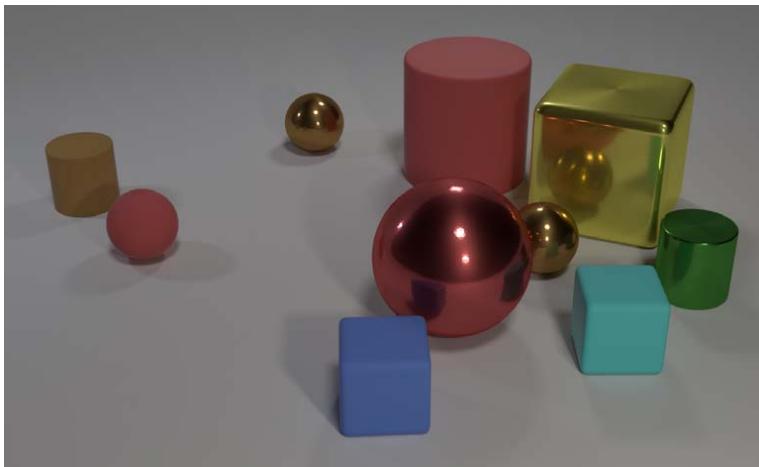
Instance Retrieval

Task: Visual Reasoning



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

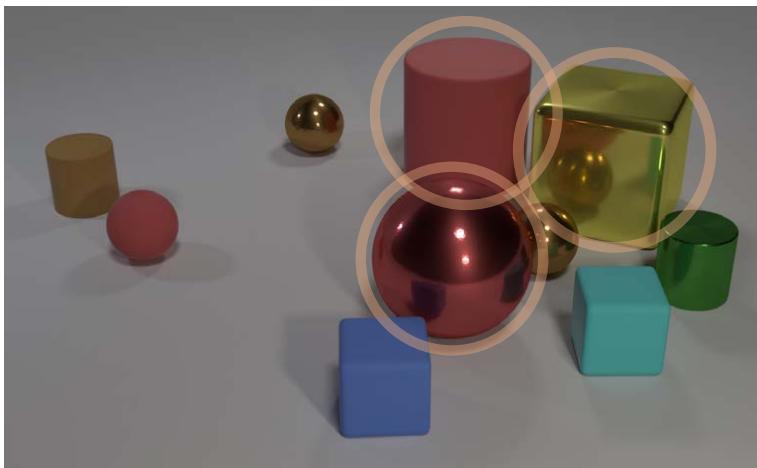
Task: Visual Reasoning



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



Task: Visual Reasoning

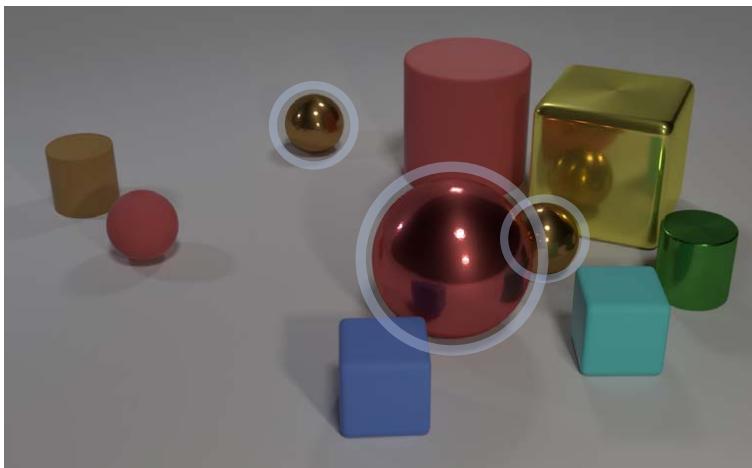


Question: *Are there an equal number of large things and metal spheres?*

3 large
things!



Task: Visual Reasoning



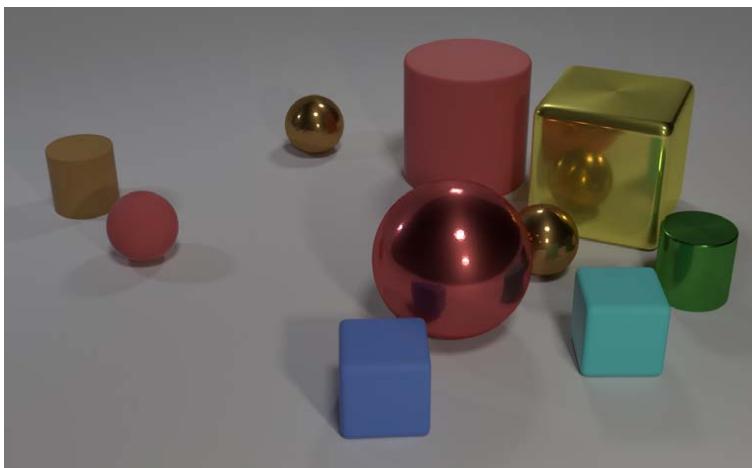
Question: Are there an equal number of large things and metal spheres?

3 large things!

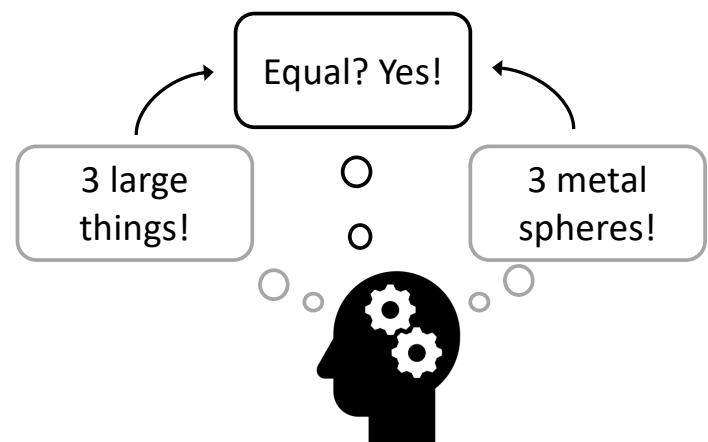
3 metal spheres!



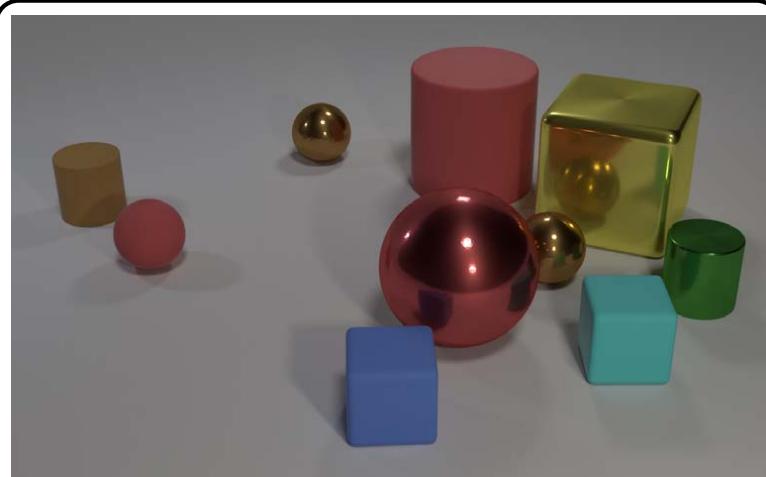
Task: Visual Reasoning



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



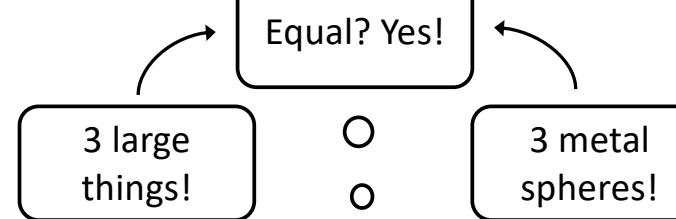
Task: Visual Reasoning



Visual Perception

Question Understanding

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



Logic Reasoning



(a) Input Image



Vision (CNN)

(b) Object Segments



CNN

I. Scene Parsing (de-rendering)

(c) Structural Scene Representation

| ID | Size | Shape | Material | Color | x | y | z |
|----|-------|----------|----------|--------|-------|-------|------|
| 1 | Small | Cube | Metal | Purple | -0.45 | -1.10 | 0.35 |
| 2 | Large | Cube | Metal | Blue | 0.83 | -0.04 | 0.70 |
| 3 | Large | Cylinder | Rubber | Yellow | 0.20 | 0.63 | 0.70 |
| 4 | Small | Cylinder | Rubber | Purple | 0.75 | 1.31 | 0.35 |
| 5 | Large | Cube | Metal | Green | 1.58 | -1.60 | 0.70 |

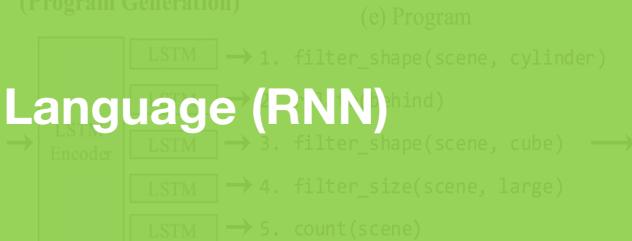
Structured Representation

(d) Question

How many cubes that
are behind the cylinder
are large?

II. Question Parsing
(Program Generation)

Language (RNN)



III. Program Execution

| | |
|-----------------|-----------------|
| 1. filter_shape | 3. filter_shape |
| 2. relate | 4. filter_size |
| | |
| 5. count | |
| | |

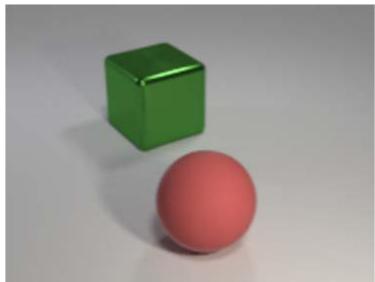
Symantic Program

Answer: 3

NS-VQA [Yi et al. 2018]

Incorporate Concepts in Visual Reasoning

Vision



Scene
Parsing
→

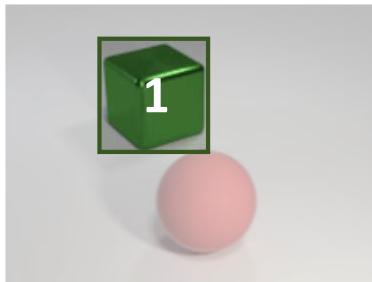
Language

Q: What's the shape of
the red object?

NS-VQA [Yi et al. 2018]

Incorporate Concepts in Visual Reasoning

Vision



Scene
Parsing
→

| ID | Color | Shape | Material |
|----|-------|-------|----------|
| 1 | Green | Cube | Metal |

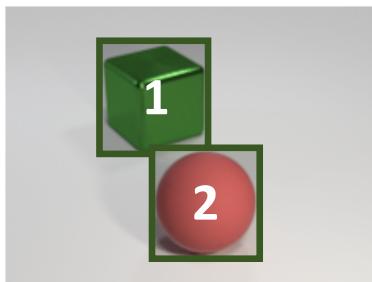
Language

Q: What's the shape of
the red object?

NS-VQA [Yi et al. 2018]

Incorporate Concepts in Visual Reasoning

Vision



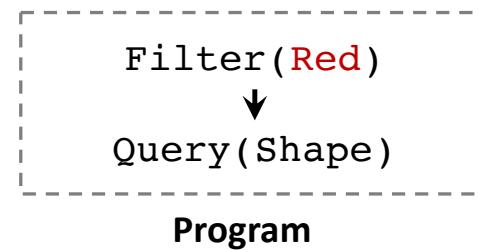
Scene
Parsing
→

| ID | Color | Shape | Material |
|----|-------|--------|----------|
| 1 | Green | Cube | Metal |
| 2 | Red | Sphere | Rubber |

Language

Q: What's the shape of
the red object?

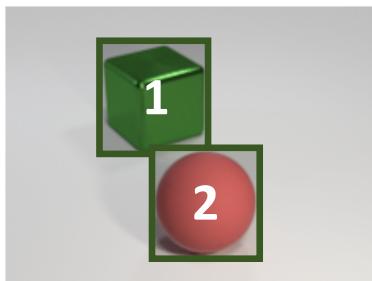
Semantic
Parsing
→



NS-VQA [Yi et al. 2018]

Incorporate Concepts in Visual Reasoning

Vision



Scene
Parsing
→

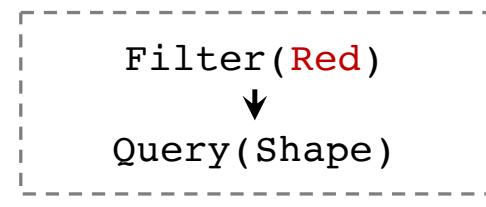
| ID | Color | Shape | Material |
|----|-------|--------|----------|
| 1 | Green | Cube | Metal |
| 2 | Red | Sphere | Rubber |



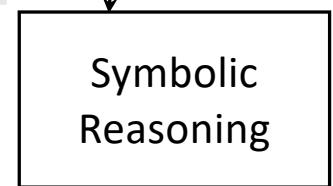
Language

Q: What's the shape of
the red object?

Semantic
Parsing
→



Program



NS-VQA [Yi et al. 2018]

Incorporate Concepts in Visual Reasoning

Vision



Scene
Parsing

| ID | Color | Shape | Material |
|----|-------|--------|----------|
| 1 | Green | Cube | Metal |
| 2 | Red | Sphere | Rubber |

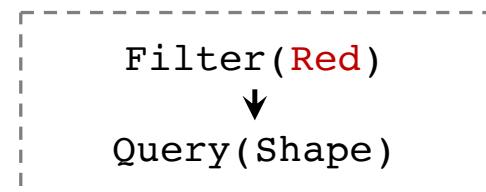


Symbolic
Reasoning

Language

Q: What's the shape of
the red object?

Semantic
Parsing



Program



NS-VQA [Yi et al. 2018]

Incorporate Concepts in Visual Reasoning

Vision



Scene
Parsing

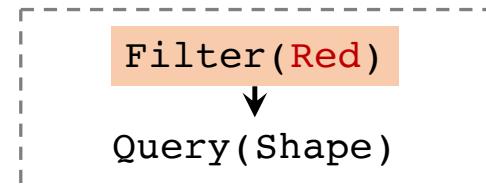
| ID | Color | Shape | Material |
|----|------------|--------|----------|
| 1 | Green | Cube | Metal |
| 2 | <i>Red</i> | Sphere | Rubber |



Language

Q: What's the shape of
the red object?

Semantic
Parsing



Program

Symbolic
Reasoning



NS-VQA [Yi et al. 2018]

Incorporate Concepts in Visual Reasoning

Vision



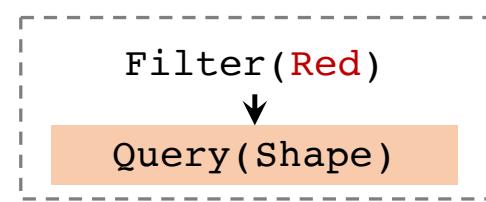
Scene
Parsing

| ID | Color | Shape | Material |
|----|-------|---------------|----------|
| 1 | Green | Cube | Metal |
| 2 | Red | <i>Sphere</i> | Rubber |

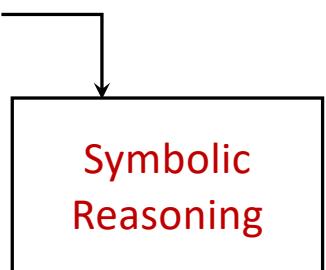
Language

Q: What's the shape of
the red object?

Semantic
Parsing



Program



Sphere

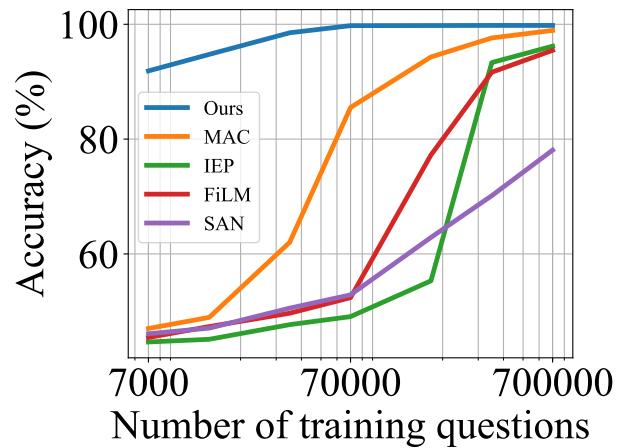
Advantage 1: High Accuracy

| Method | Accuracy (%) |
|------------------|--------------|
| Human | 92.6 |
| RN | 95.5 |
| IEP | 96.9 |
| FiLM | 97.6 |
| MAC | 98.9 |
| TbD | 99.1 |
| NS-VQA (Ours) | 99.8 |

Effectively perfect!

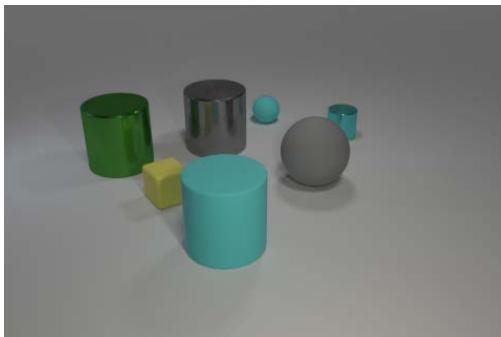
Advantage 2: Data Efficiency

High accuracy when trained with just 1% the of the data that other methods require



[Yi et al. NeurIPS 2018]

Advantage 3: Transparency and Interpretability



Question: Are there more yellow matte things that are right of the gray ball than cyan metallic objects?

```
scene
filter_cyan
filter_metal
count
... (4 modules)
scene
filter_yellow
filter_rubber
count
greater_than
```

Answer: no

[Yi et al. NeurIPS 2018, Johnson et al. ICCV 2017]

NeurIPS 2018: Neurosymbolic VQA:
Properties (e.g. “color”) and values (“red”) predefined

ICLR 2019: Neurosymbolic Concept Learner:
Properties predefined, can learn new values autonomously

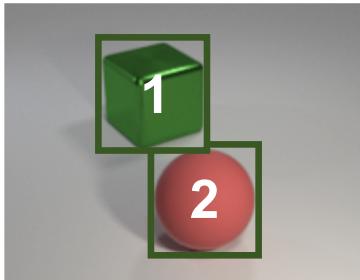
NeurIPS 2019: Neurosymbolic Metaconcept Learner:
Autonomously learns new concepts

ICML 2020 (target submission):
Real world images

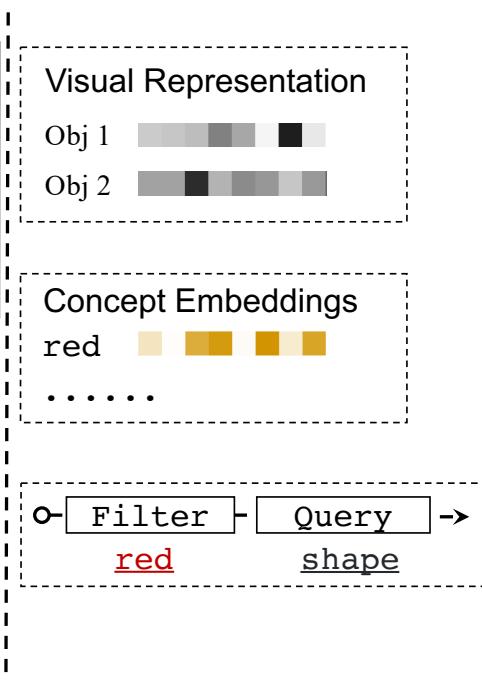


less predefined, more autonomous →

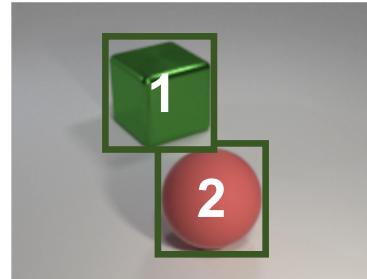
Neuro-Symbolic Concept Learning



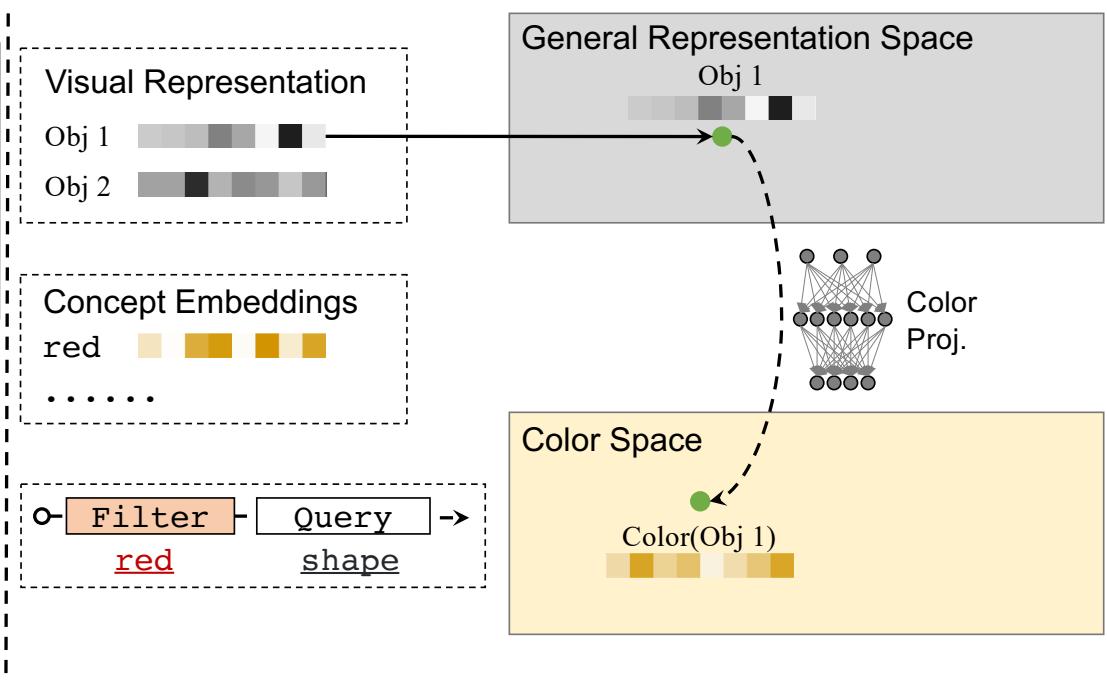
Q: What's the shape of the red object?



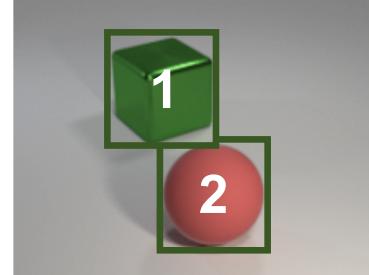
Neuro-Symbolic Concept Learning



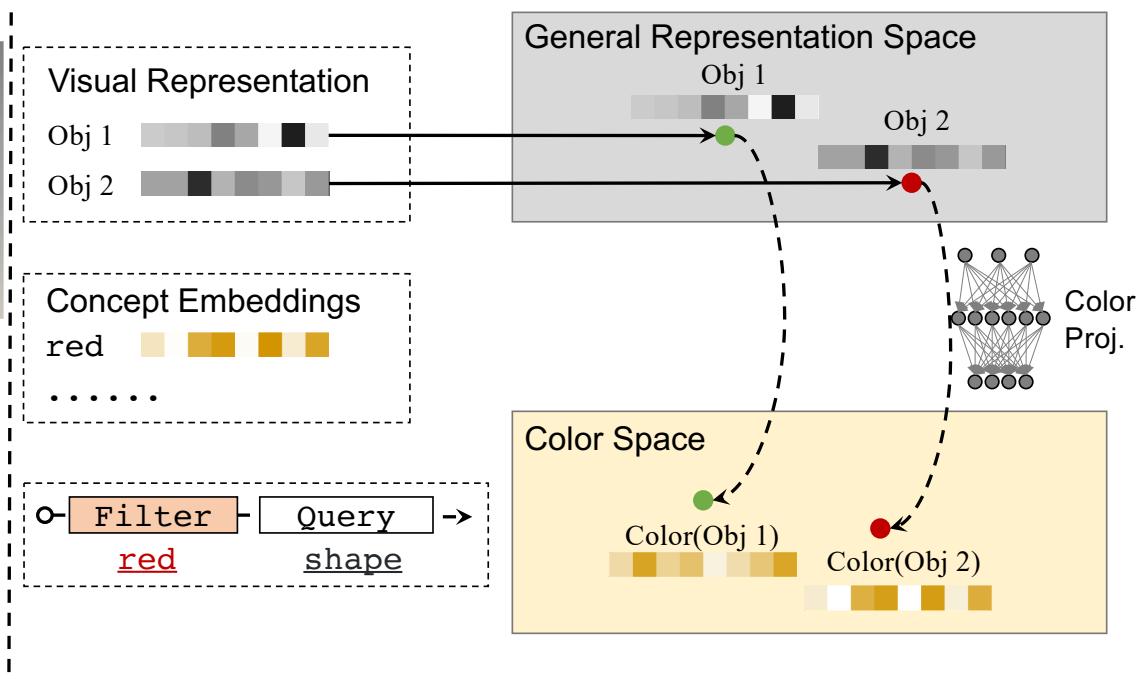
Q: What's the shape of the red object?



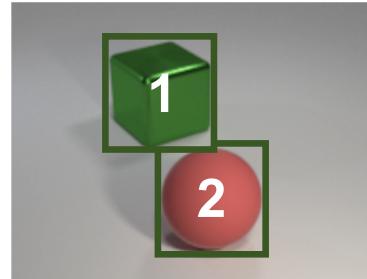
Neuro-Symbolic Concept Learning



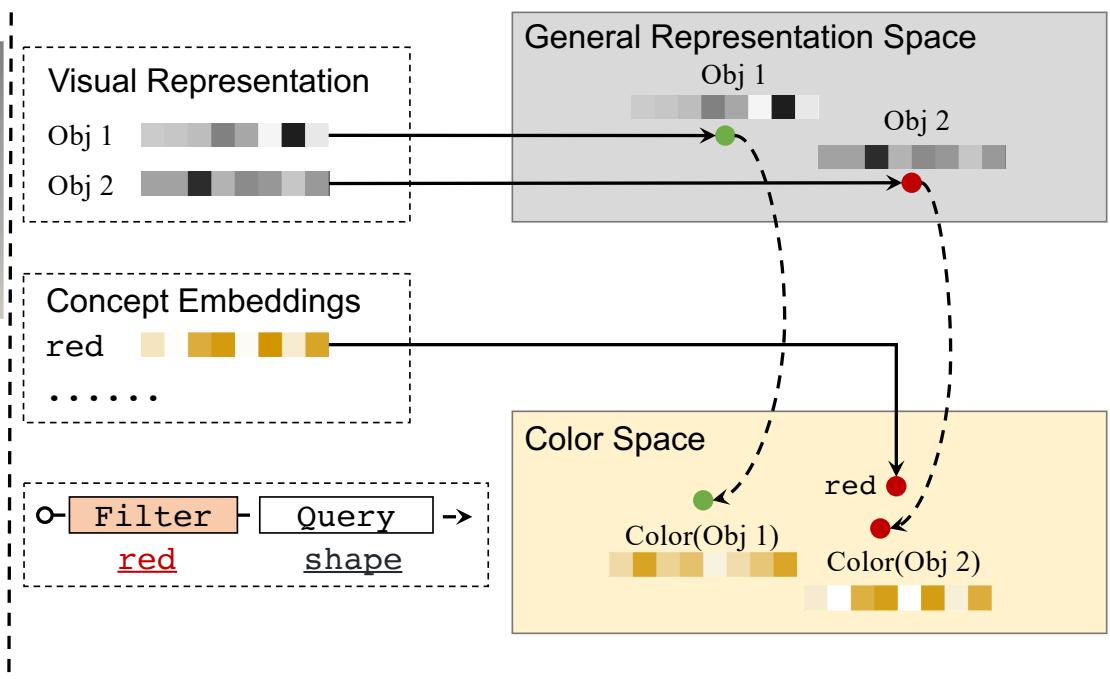
Q: What's the shape of the red object?



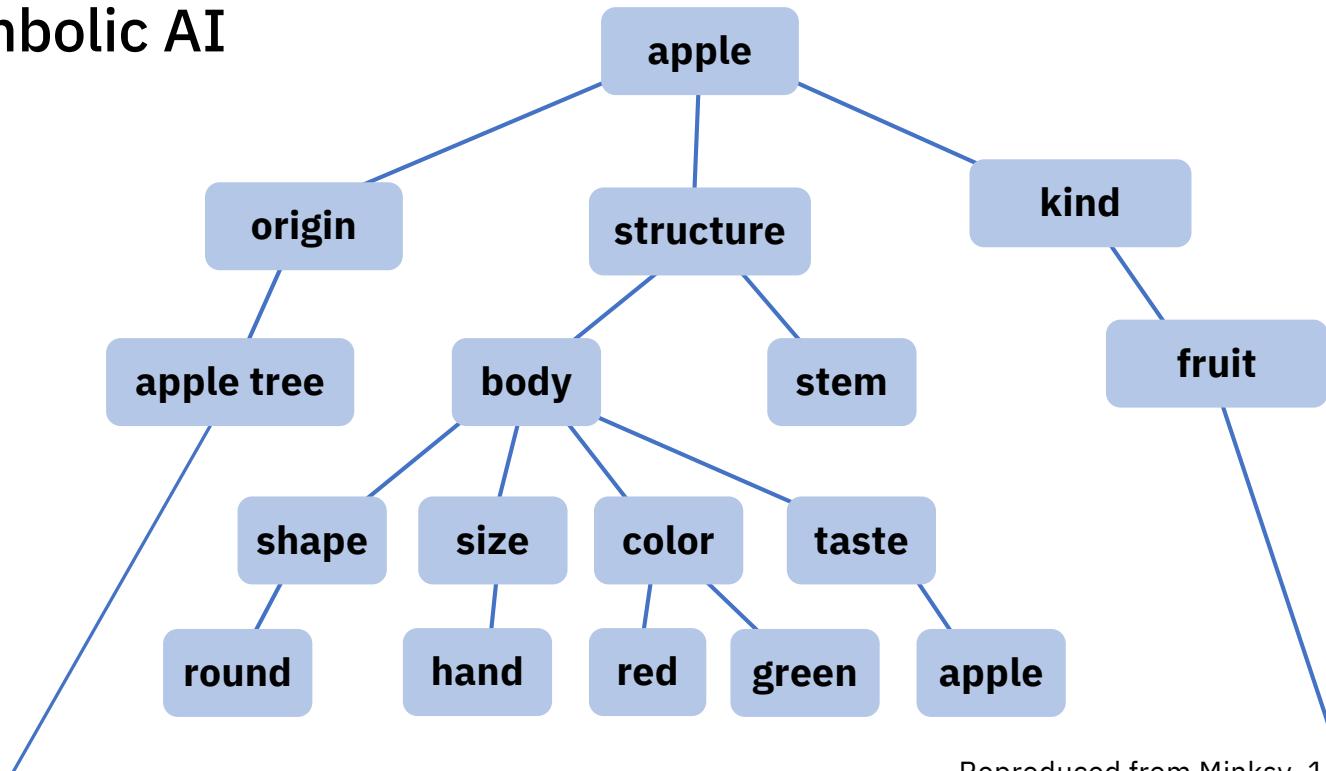
Neuro-Symbolic Concept Learning



Q: What's the shape of the red object?



Symbolic AI



Reproduced from Minsky, 1991

Meta-concept Learning

Han et al. NeurIPS 2019

Visual reasoning questions

color:
red



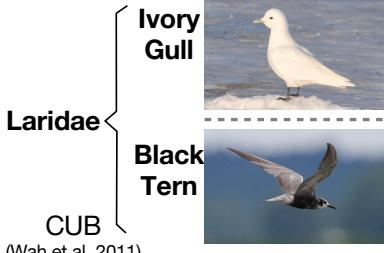
Q: Is there any **red cube**?
A: Yes.

color:
green



Q: Is there any **green block**?
A: Yes

CLEVR
(Johnson et al. 2017)



+ Metaconcept questions

Q: Is red a **same kind** of concept as green?
A: Yes.

Q: Is cube a **synonym** of block?
A: Yes.

Q: Is Laridae a **hypernym** of Ivory gull?
A: Yes.

Augmenting VQA with Metaconcepts

Visual reasoning questions

color:
red



Q: Is there any **red cube**?
A: Yes.

color:
green



Q: Is there any **green block**?
A: Yes

CLEVR
(Johnson et al. 2017)

Ivory
Gull



Q: Is there any **Ivory Gull**?
A: Yes.

Q: Is there any **Laridae**?
A: Yes.

Black
Tern



Q: Is there any **Black Tern**?
A: Yes.

Q: Is there any **Laridae**?
A: Yes.

CUB
(Wah et al. 2011)

+ Metaconcept questions

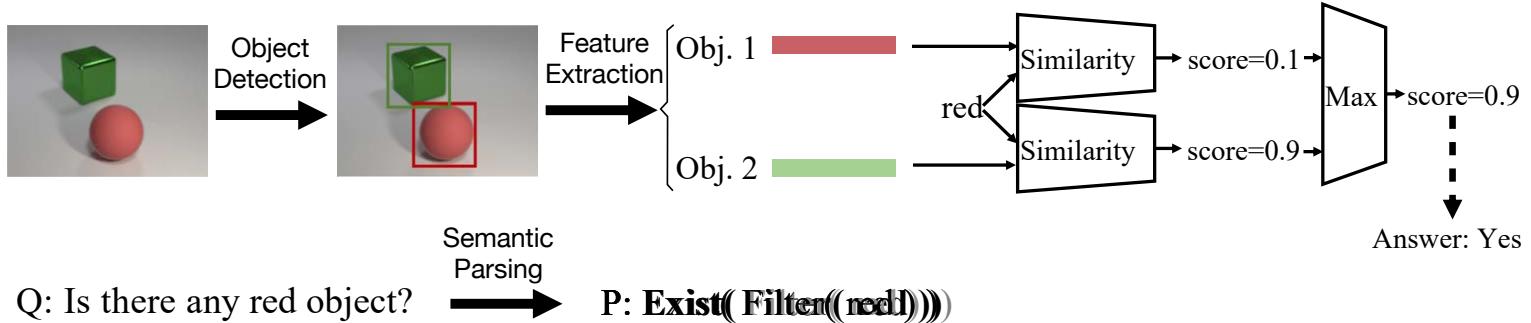
Q: Is red a **same kind** of concept as green?
A: Yes.

Q: Is cube a **synonym** of block?
A: Yes.

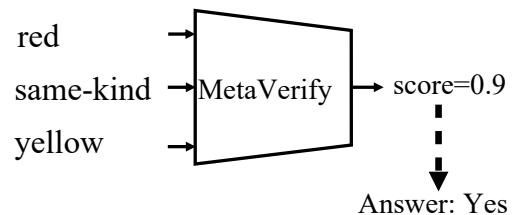
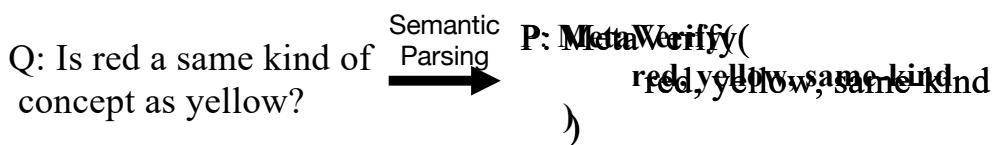
Q: Is Laridae a **hypernym** of Ivory gull?
A: Yes.

Program Execution Animated

Visual reasoning questions



Metaconcept questions



Generalization

Metaconcept Generalization



Q: Is there any *airplane*?
A: Yes



Q: Is there any *kid*?
A: Yes

Q: Is airplane a *synonym* of plane?
A: Yes



Q: Is there any *plane*?
A: Yes



Q: Is there any *child*?
A: Yes

Q: Is kid a *synonym* of child?
A: Yes

Training



airplane
↑
synonym
↓
plane



kid
↑
synonym?
↓
child

Testing: metaconcepts on unseen pairs of concepts

Generalization

Metaconcept Generalization: Results



Q: Is there any *airplane*?
A: Yes



Q: Is there any *kid*?
A: Yes

Q: Is airplane a *synonym* of plane?
A: Yes



Q: Is there any *plane*?
A: Yes



Q: Is there any *child*?
A: Yes

Q: Is kid a *synonym* of child?
A: Yes

Training

Testing

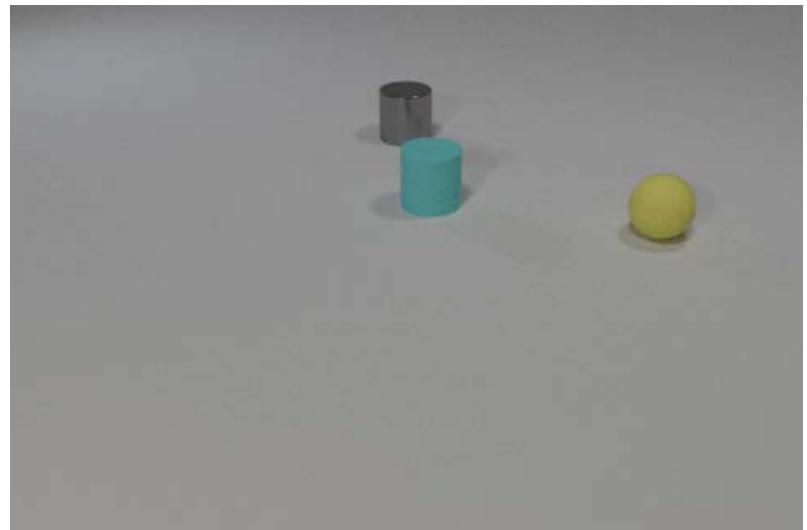
| | | Q.Type | GRU (Lang. Only) [Cho et al., 2014] | GRU-CNN [Zhou et al., 2015] | BERT (question ; concept) [Jacob Devlin, 2018] | NS-CL [Mao et al. 2019] | VCML |
|--------------|-----------|--------|--|--------------------------------|---|-----------------------------------|-----------------------------------|
| CLEVR | Synonym | 50.0 | 66.3 ± 1.4 | 60.9 ± 10.6 | $76.2 \pm 10.2 ; 80.2 \pm 16.1$ | 100.0 ± 0.0 | 100.0 ± 0.0 |
| | Same-kind | 50.0 | 64.7 ± 5.1 | 61.5 ± 6.6 | $75.4 \pm 5.4 ; 80.1 \pm 10.0$ | 92.3 ± 4.9 | 99.3 ± 1.0 |
| GQA | Synonym | 50.0 | 80.8 ± 1.0 | 76.2 ± 0.8 | $76.2 \pm 2.4 ; 83.1 \pm 1.5$ | 81.2 ± 2.8 | 91.1 ± 1.7 |
| | Same-kind | 50.0 | 56.3 ± 2.3 | 57.3 ± 5.3 | $59.5 \pm 2.7 ; 68.2 \pm 4.0$ | 66.8 ± 4.1 | 69.1 ± 1.7 |
| CUB | Hypernym | 50.0 | 74.3 ± 5.2 | 76.7 ± 8.8 | $75.6 \pm 1.2 ; 61.7 \pm 10.3$ | 80.1 ± 7.3 | 94.8 ± 1.3 |
| | Meronym | 50.0 | 80.1 ± 5.9 | 78.1 ± 4.8 | $63.1 \pm 3.2 ; 72.9 \pm 9.9$ | 97.7 ± 1.1 | 92.5 ± 1.0 |

CLEVERER: CoLlision Events for Video REpresentation and Reasoning

- Descriptive

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

A: Metal

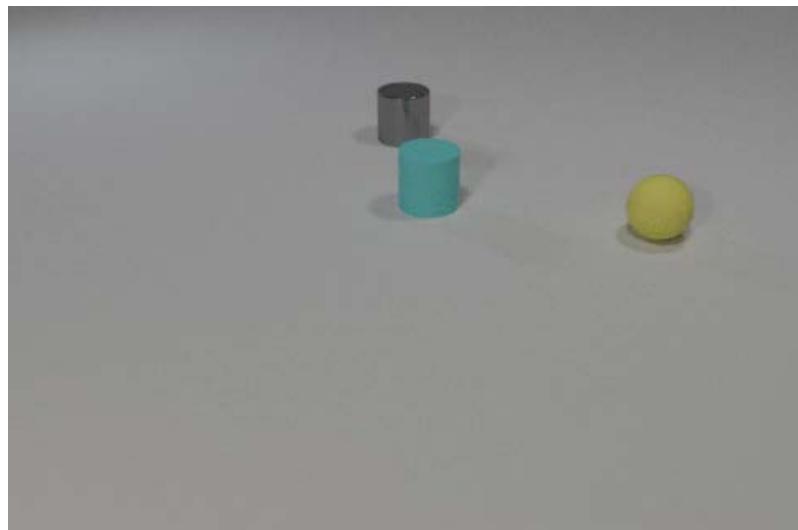


Chuang Gan w/ Kevin Xi, Yunzhu Li, Pushmeet Kohli, Jiajun Wu, Antonio Torralba & Josh Tenenbaum

- Explanatory

Q: What is responsible for the collision between the rubber and metal cylinder?

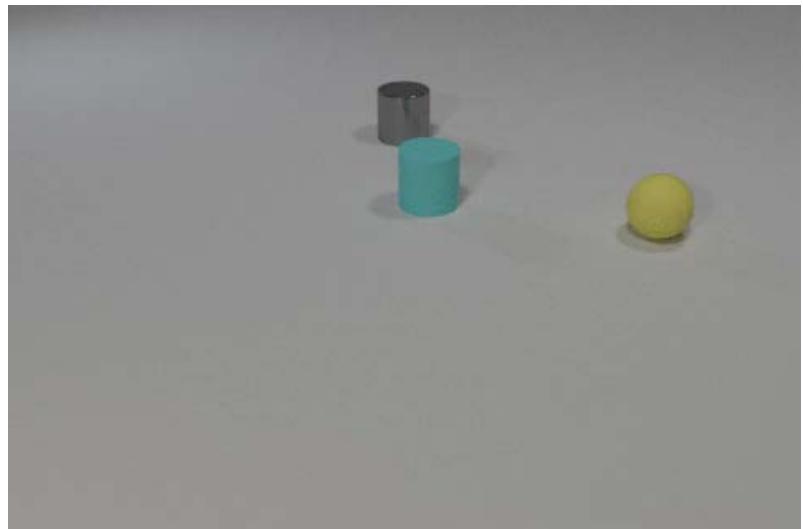
- A. The presence of the yellow sphere*
- B. The collision between the rubber cylinder and the red rubber sphere*



- Counterfactual

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

- A. *The red rubber sphere and the metal sphere collide*
- B. *The red rubber sphere and the gray object collide*



Looking Ahead

How many employees have over 10 years experience but have moved location in the last year?

What factors might contribute to better output from Factory A vs. Factory B?

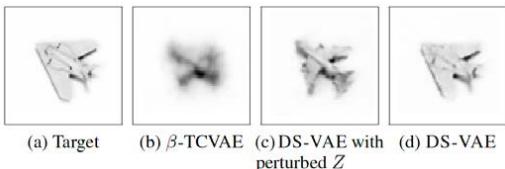
Why is our database down?

With regard to my credentials, I completed my Ph.D. in the Department of Brain and Cognitive Sciences at MIT in 2007 with a specialization in computational neuroscience. Prior to joining IBM Research, I was a postdoctoral researcher at the Department of Engineering and Computer Science at Harvard University; I was a computational neuroscientist and machine learning researcher at IBM Research, where I am now a Director of Research. In addition, I was a visiting scientist at the University of Washington, where I became aware of Dr. Wei's outstanding research. Having worked in a collaborative relationship on a computational project with Dr. Wei, I am well qualified to discuss the impact of his work in this area.

My academic lab's alumni include several professors who have started their own labs and founded several startups that have raised venture funding and gone on to be acquired. I also have extensive experience teaching. I am the author of the course *Neuroscience*, one of the earliest Massive Open Courses at Harvard. Over half a million people in 192 countries have visited the course. On campus, I used this online course as a platform for my research group to teach neuroscience concepts, which colleagues in education research on using our course as a platform for exploring best practices in online education.

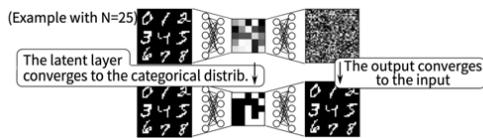
To provide a brief introduction to Dr. Wei's work in AI, I believe it is necessary to understand one of the central problems we face in the world of AI. While recent years

Neurosymbolic Generative Models



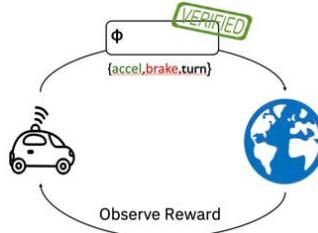
Srivastava et al. 2020 (submitted)

Neurosymbolic Planning



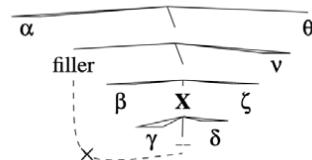
Asai et al. AAAI 2018

Neurosymbolic Safe ML/RL



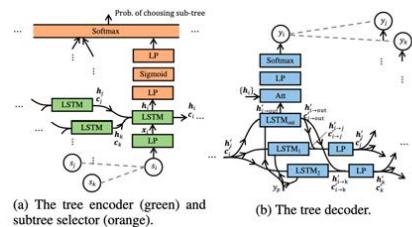
Fulton et al AAAI 2018

Neurosymbolic NLU



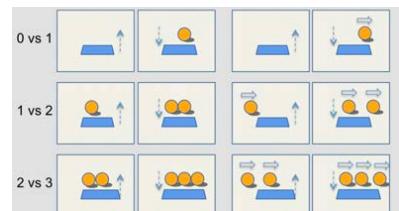
Wilcox et al. NAACL 2019

Neurosymbolic Code Optimization



Shi et al. ICLR 2019

Neurosymbolic Machine Common Sense



Smith et al. NeurIPS 2019

Inducing Behavioral Insight

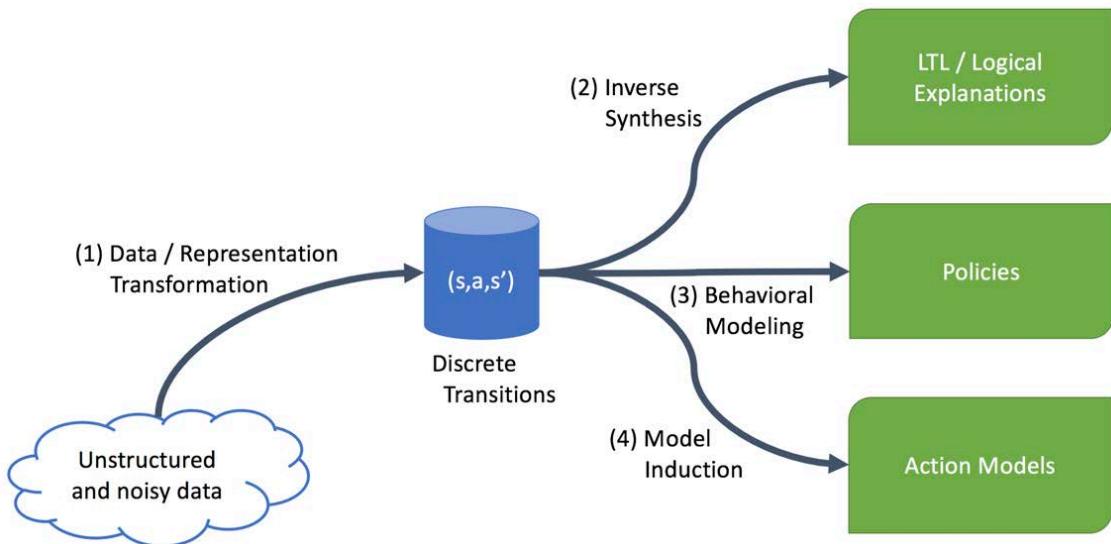
Inferring flexible behavioral plans/policies from temporal observation data



Julie Shah
MIT



Christian Muise
IBM





```
(:action pickup  
:parameters (?b1 ?b2 - block)  
:precondition (and (on ?b1 ?b2)  
                   (hand-clear))  
:effect (and (not (hand-clear))  
           (not (on ?b1 ?b2))  
           (holding ?b1))  
)
```

Task: Induce the action theory of an environment through observations

LatPlan

Mixing symbolic planning with neural networks



Masataro Asai
IBM

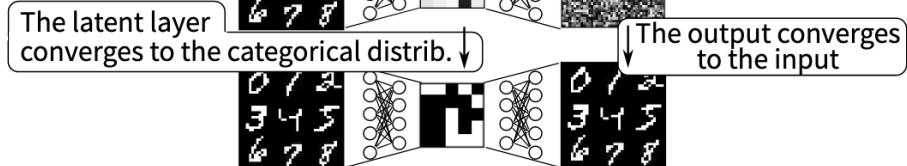
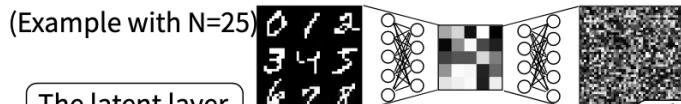


LatPlan

Mixing symbolic planning with neural networks



Masataro Asai
IBM



```

When ;; Translates to a PDDL model below:
Empty(x,yold) ∧ (:action slide-up ...
at(x,ynew,p) ∧ :precondition
up(ynew,yold); (and (empty ?x ?y-old)
then :effects
¬Empty(x,yold) ∧ (at ?x ?y-new ?p) ...)
Empty(x,ynew) ∧ (not (empty ?x ?y-old))
at(x,ynew,p) ∧ (empty ?x ?y-new)
at(x,yold,p) (not (at ?x ?y-new ?p))
(at ?x ?y-old ?p)))

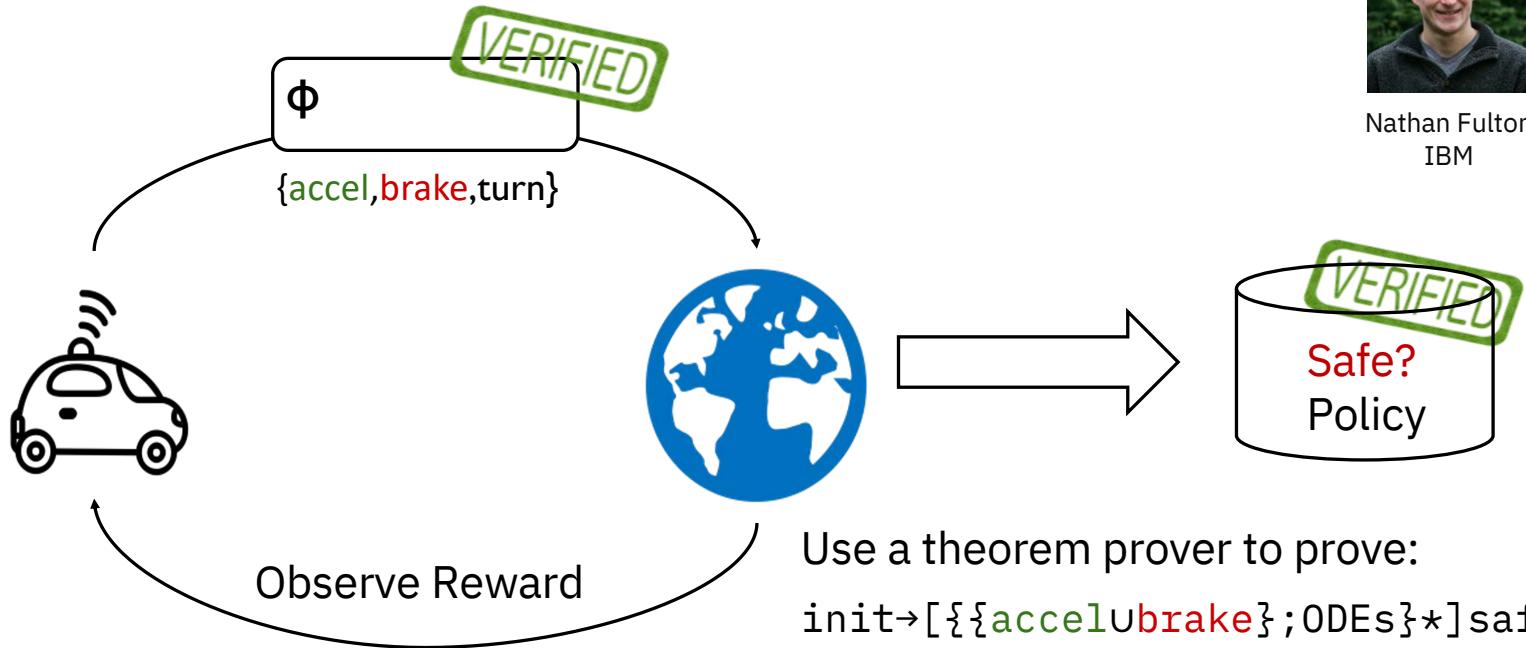
```

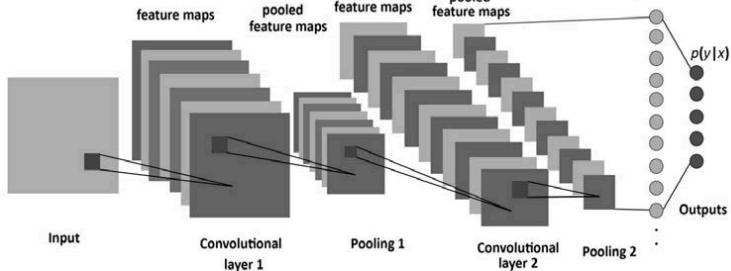
| | | |
|---|---|---|
| | 6 | 8 |
| 7 | 3 | 2 |
| 5 | 1 | 4 |

Verifiably Safe Reinforcement Learning

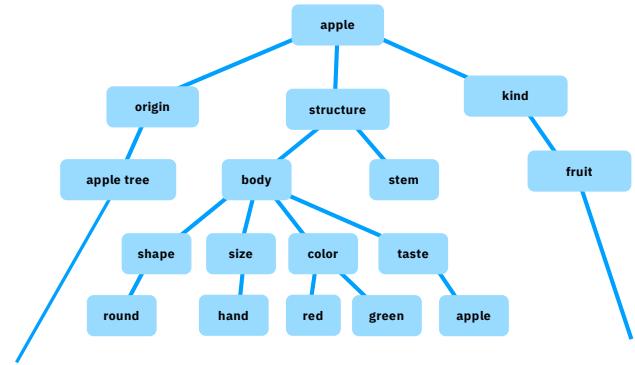


Nathan Fulton
IBM





+



(:action pickup

:parameters (?b1 ?b2 - block)

**:precondition (and (on ?b1 ?b2)
(hand-clear))**

**:effect (and (not (hand-clear))
(not (on ?b1 ?b2))
(holding ?b1))**

)

NEURAL NETWORKS

SYMBOLIC AI

Causal Inference

Beyond Correlation—inferring and testing for causal relationships in complex systems



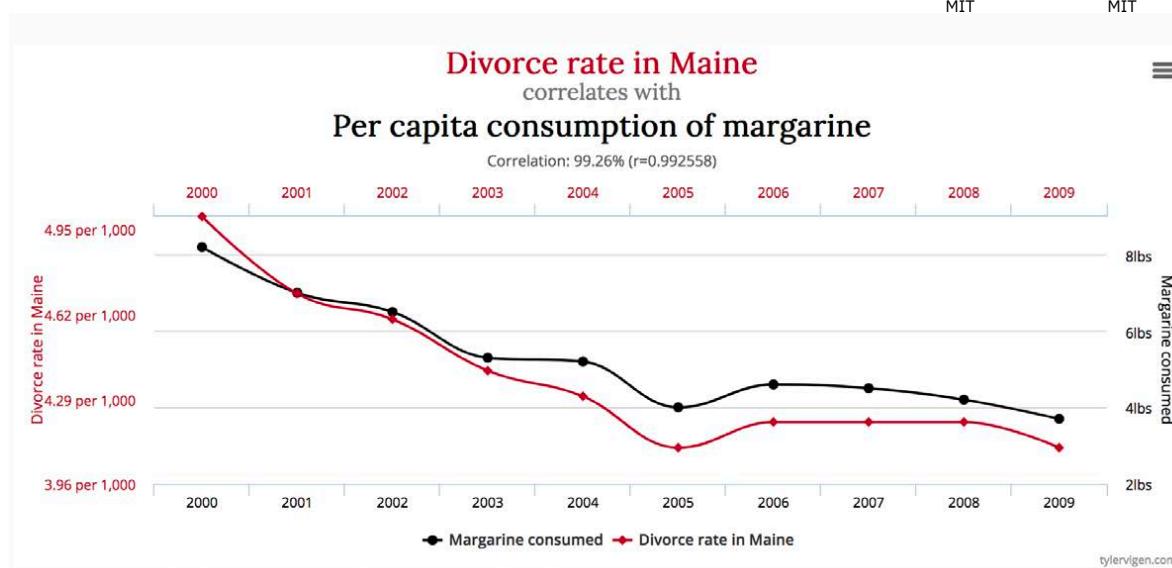
Caroline Uhler
MIT



Guy Bresler
MIT



Karthikeyan
Shanmugam
IBM



Data sources: National Vital Statistics Reports and U.S. Department of Agriculture

tylervigen.com

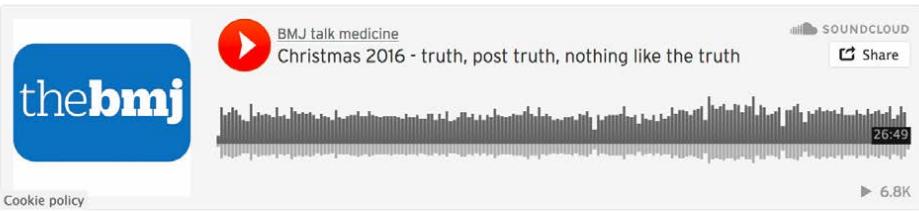
<http://tylervigen.com/spurious-correlations>

thebmj Research ▾ Education ▾ News & Views ▾ Campaigns ▾ Archive

Feature » Christmas 2016: Food for Thought

Is caviar a risk factor for being a millionaire?

BMJ 2016 ; 355 doi: <https://doi.org/10.1136/bmj.i6536> (Published 09 December 2016)
Cite this as: BMJ 2016;355:i6536



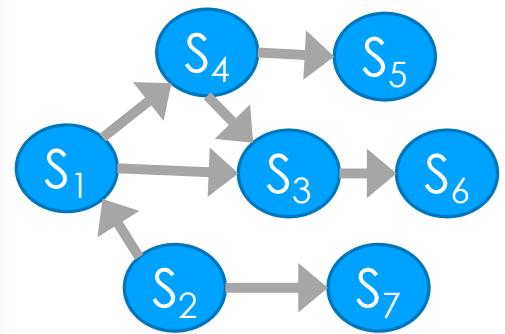
The screenshot shows a blue header bar with the "thebmj" logo. Below it, a main content area features a red play button icon next to the text "BMJ talk medicine". To its right, the title "Christmas 2016 - truth, post truth, nothing like the truth" is displayed above a waveform audio player. The waveform has a timestamp of "26:49" at the bottom right. Above the waveform, there's a "SOUNDCLLOUD" logo and a "Share" button. At the bottom of the main content area, there are four tabs: "Article" (which is selected and highlighted in blue), "Related content", "Metrics", and "Responses".

Cookie policy ► 6.8K

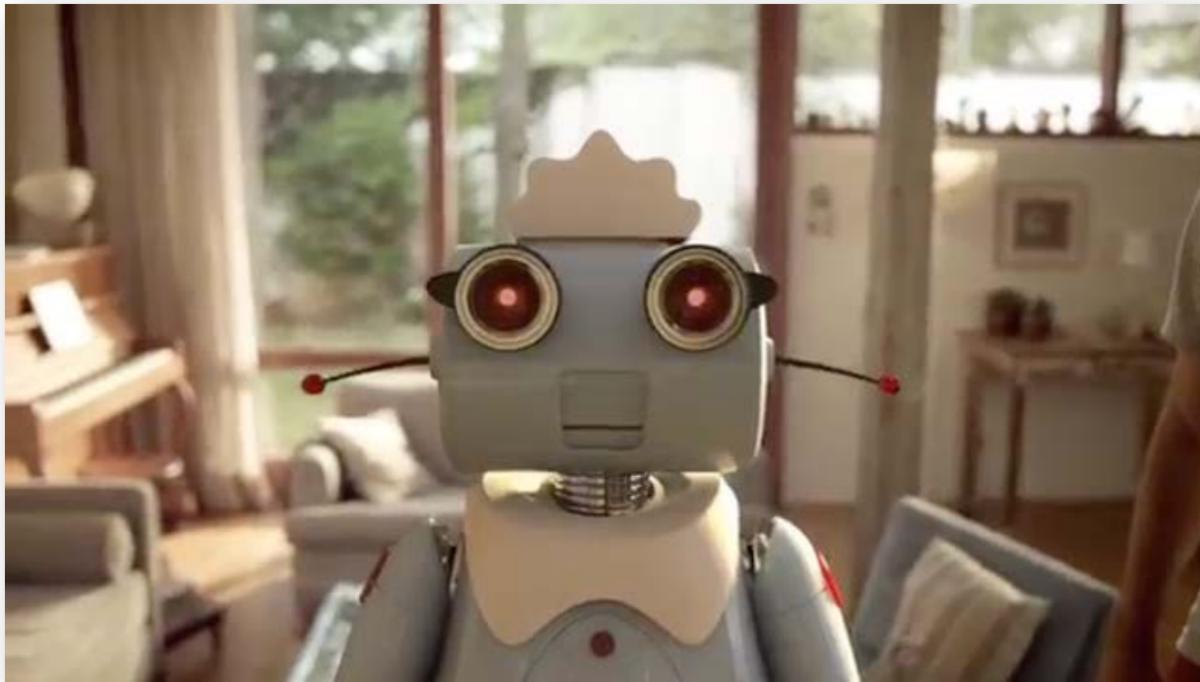
Anders Huitfeldt, postdoctoral scholar

Author affiliations ▾

Correspondence to: A Huitfeldt ahuitfel@stanford.edu



Generalizable Autonomy in Robot Manipulation



Animesh Garg



UNIVERSITY OF
TORONTO



VECTOR
INSTITUTE



NVIDIA®

Generalizable Autonomy in Robot Manipulation



Vacuuming



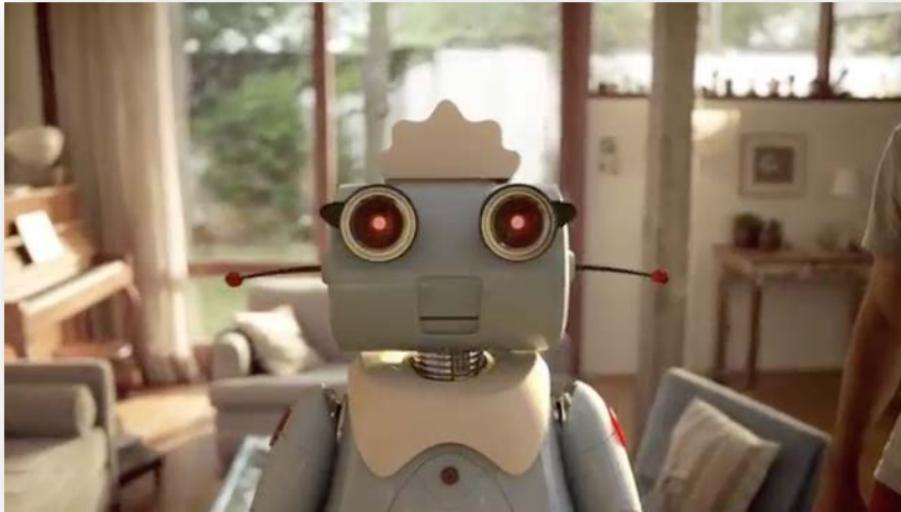
Sweeping/Mopping



Cooking



Laundry



Generalizable Autonomy in Robot Manipulation



Vacuuming



Sweeping/Mopping



Cooking



Laundry



Diversity:
New Scenes,
Tools,...



Complexity:
Long-term
Settings

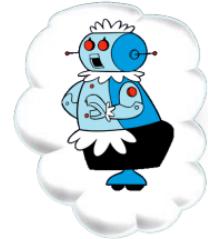


Generalizable Autonomy in Robot Manipulation

Vision: Build Intelligent Robotic Companions
towards Human Enrichment and Augmentation



Generalizable Autonomy in Robot Manipulation

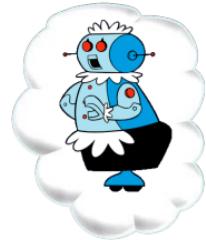


1956 Dartmouth AI Project



1956

Generalizable Autonomy in Robot Manipulation

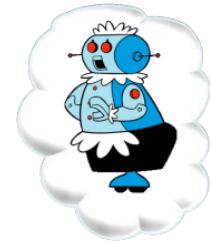


Dartmouth AI Meeting

UNIMATE
1st Industrial robot

1956 '61 1968

Generalizable Autonomy in Robot Manipulation



Dartmouth Al Me

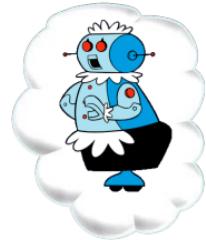
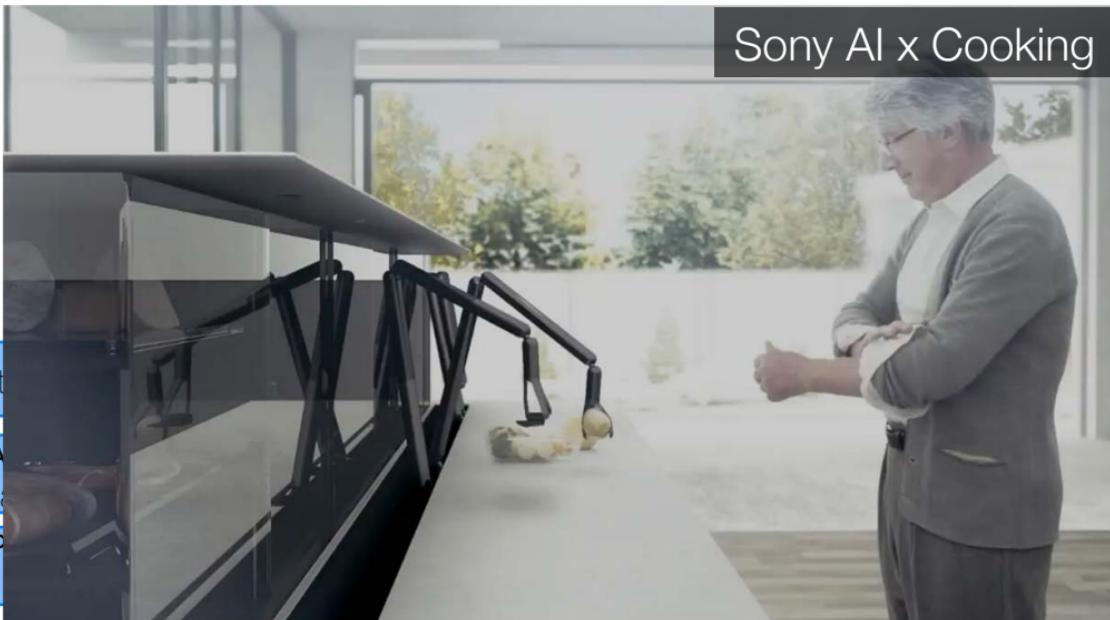
UNIN
1st Indust

ATLAS CAN WALK IN
TOUGH CONDITIONS,

1956 '61 1968

2013

Generalizable Autonomy in Robot Manipulation



Dartmouth AI Meet

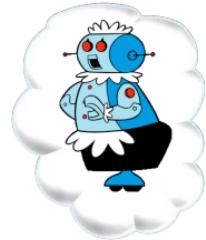
UNIMA
1st Industrial
Robot

P

1956 '61 1968

2013 2018

Generalizable Autonomy in Robot Manipulation



Dartmouth AI Me

UNIN
1st Indust

1956 '61 1968

2013 2018 2019

Generalizable Autonomy in Robot Manipulation

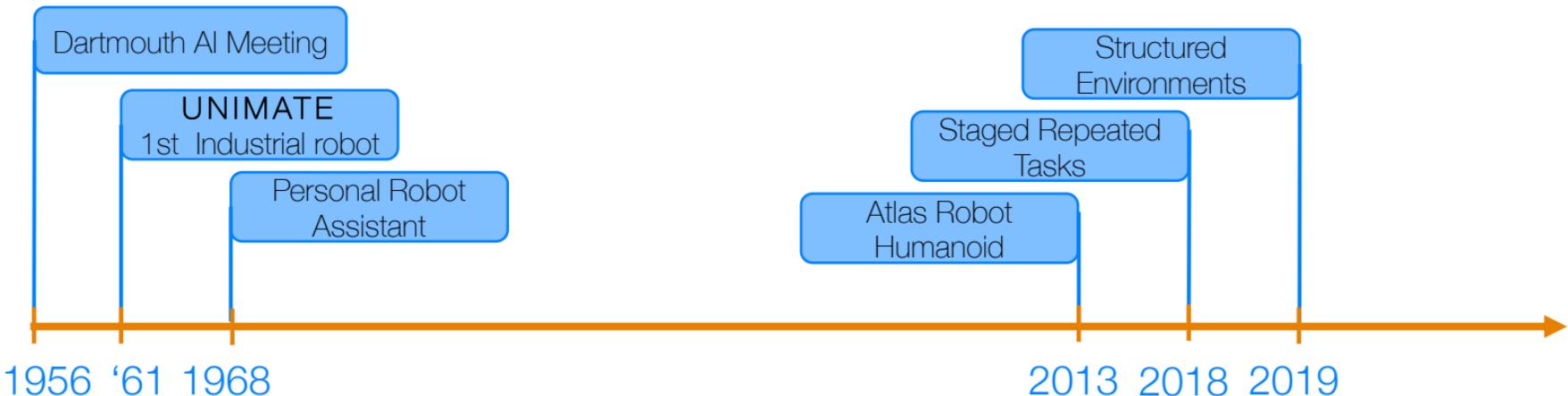
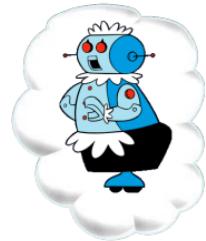


Then



Now

How to Generalize to
Unstructured Scenarios?



Generalizable Autonomy in Robot Manipulation

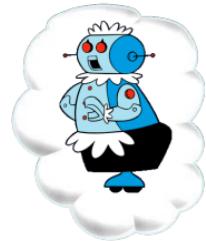


Then

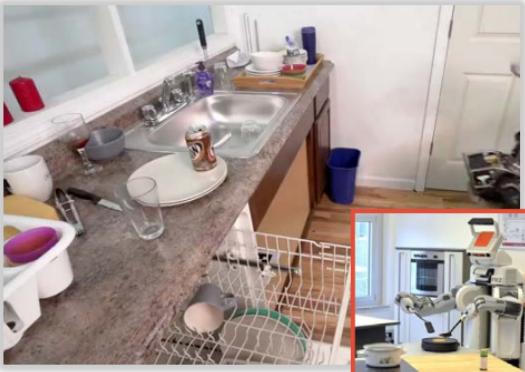


Now

How to Generalize to
Unstructured Scenarios?



Manufacturing/Retail



Personal/Service



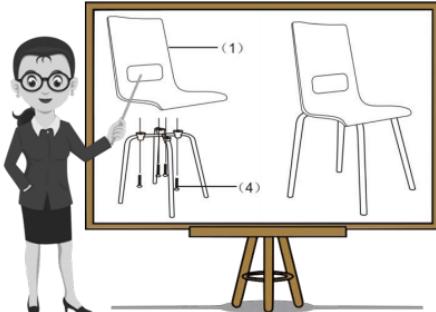
Healthcare/Medicine

Generalizable Autonomy in Robot Manipulation

Vision: Build Intelligent Robotic Companions

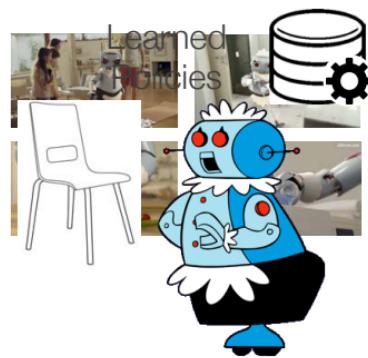
Approach: Learning with **Structured** Inductive Bias and Priors

Demonstration



Instructional Input
(Teleoperation, Video, Language)

Task Imitation



Learn to do the task in
Same Environment

Generalization



New Task Variations
in Novel Environments

Layers of Imitation

Movement
Skills

Control

Skill
Sequencing

Planning

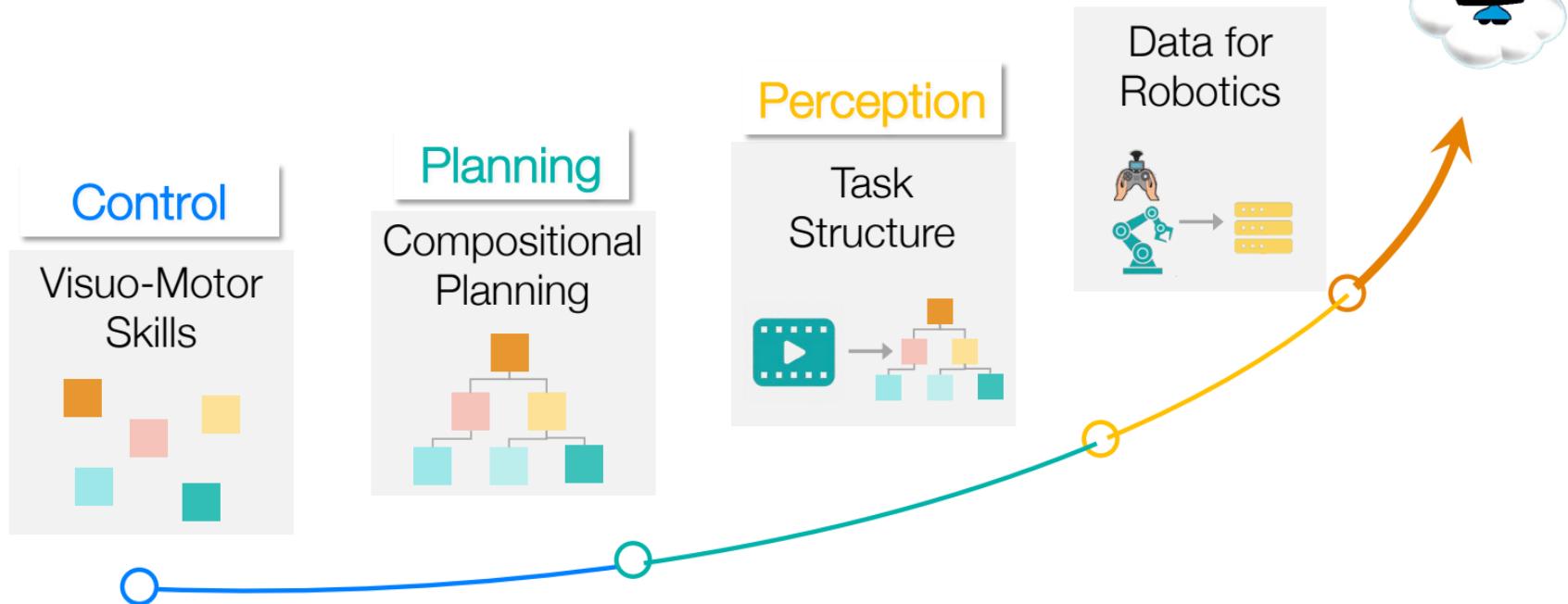
Semantic
Purpose

Perception

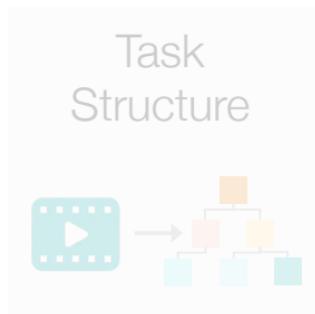
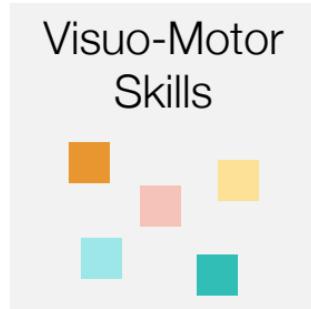


Task Specification

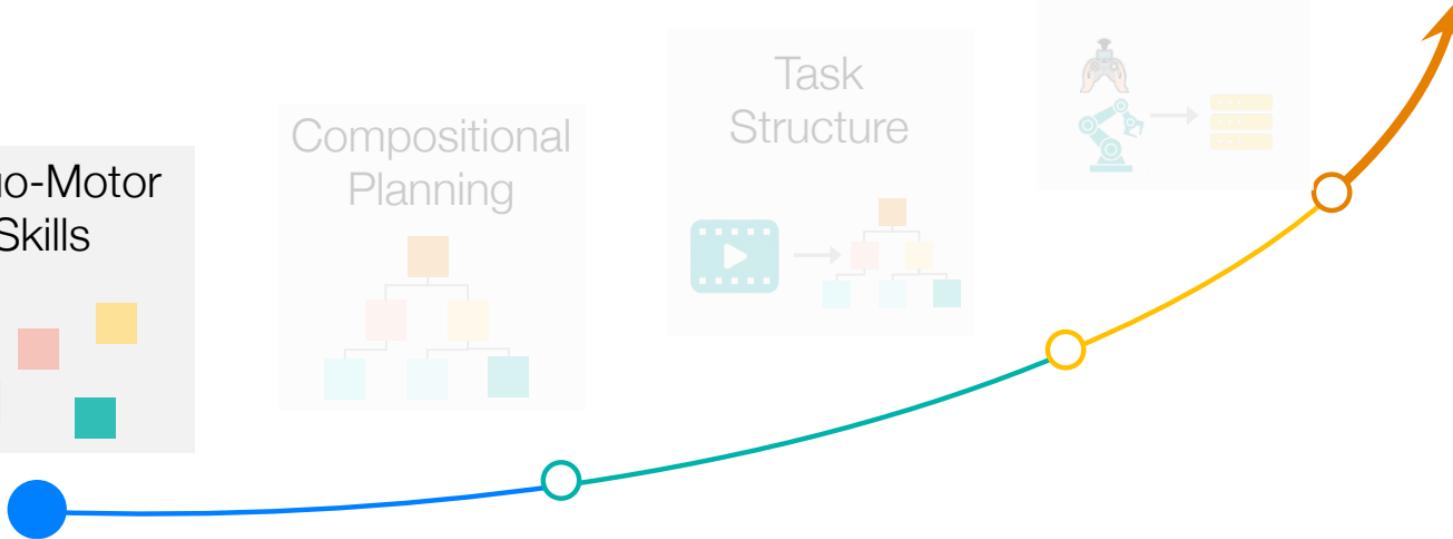
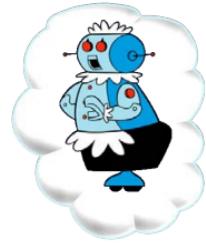
Generalizable Autonomy in Robot Manipulation



Generalizable Autonomy in Robot Manipulation



Data for Robotics



Visuo-Motor Skills

Challenge: Algorithmic frameworks to learn a **diversity** of skills

Approach: Close the **Visuo-Motor Loop** with Learning based **Control**



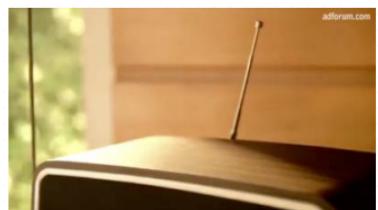
Vacuuming



Sweeping/Mopping



Cooking



Cleaning

Visuo-Motor Skills: Generalization



Cleaning



Skills: Surface Wiping



Hard Stains – Push Harder?



Different Surfaces – Be Gentle?



Generalization



Visuo-Motor Skills: Current Paradigm

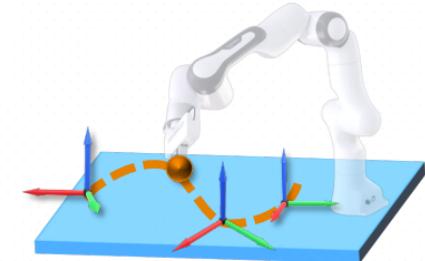
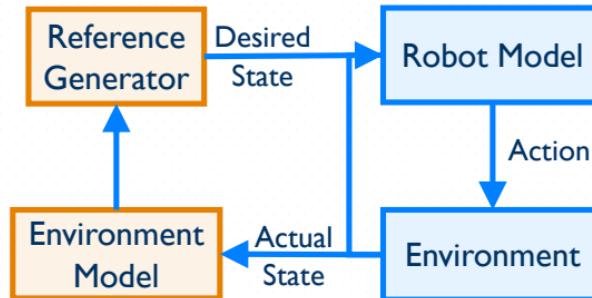
Model Based Task (Operational) Space Control

Actual State: Image, Force, Joint Enc.

Desired State: x_d

Robot Model Parameters: M, J

Action: τ



Robot Model

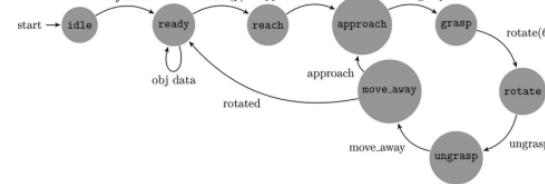
$$\ddot{x}_{ref} = K_p(x_d - x) + K_v(\dot{x}_d - \dot{x}) + \ddot{x}_d$$

$$M(q, \dot{q}) + C(q, \dot{q}) + G(q) + \varepsilon(q, \dot{q}) = \tau$$

$$\tau = J^T (JM^{-1}J^T)^{-1} (\ddot{x}_{ref} - J\dot{q} + JM^{-1}F)$$

- + Leverages Robot Model
- + Compliant Control

Environment Model + Reference Generator



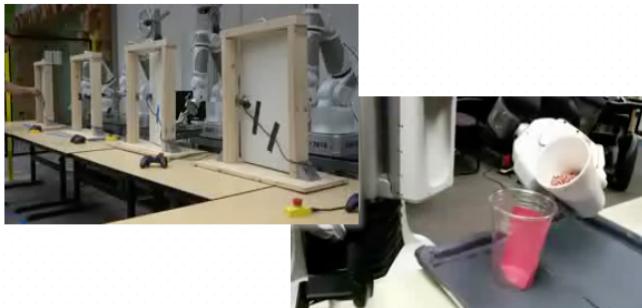
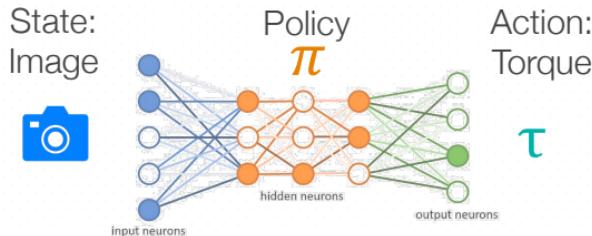
- Needs Environment (Task) Model



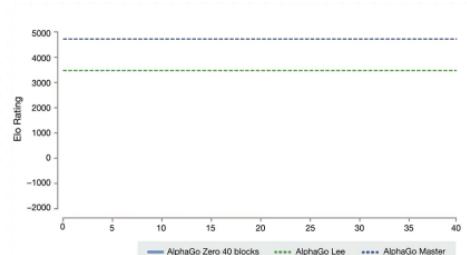
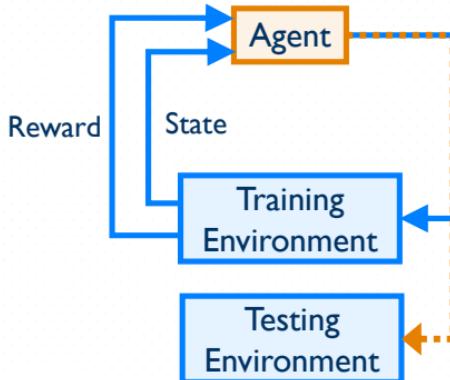
- Task Dependent State
- Explicit State Estimation

Visuo-Motor Skills: Current Paradigm

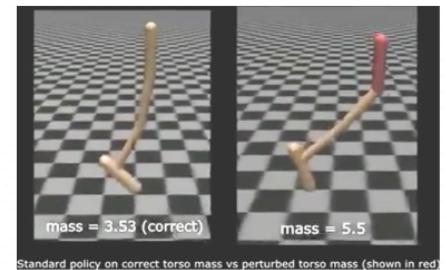
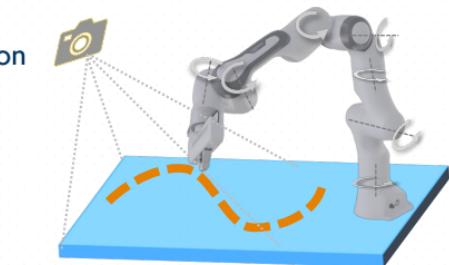
Deep Reinforcement Learning



- + Model Free: No Environment Model
- + State is Image



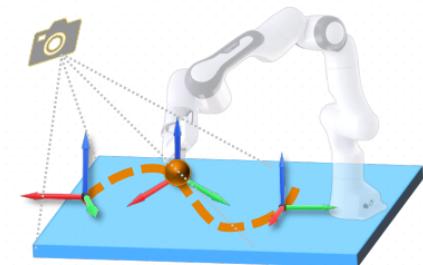
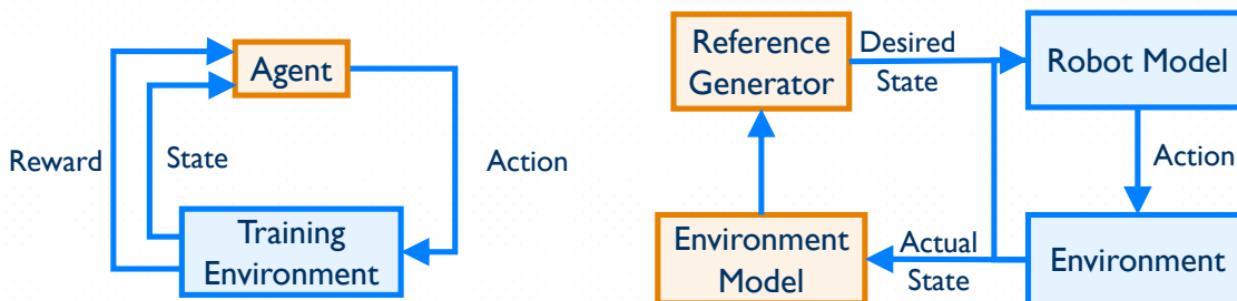
- Sample Inefficient
- Learn robot model (implicitly)



- If Training \neq Testing: Policy Fails!

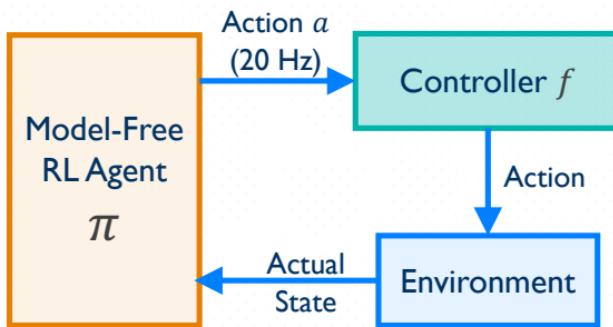
Visuo-Motor Skills: Our Approach

RL with Variable Impedance Task-Space



Visuo-Motor Skills: Our Approach

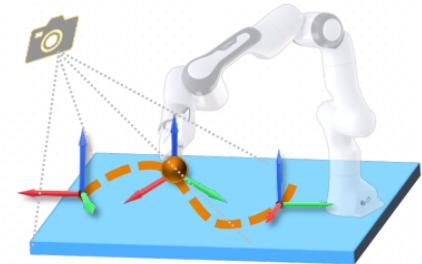
RL with Variable Impedance Task-Space



Reference Generator
(learned)

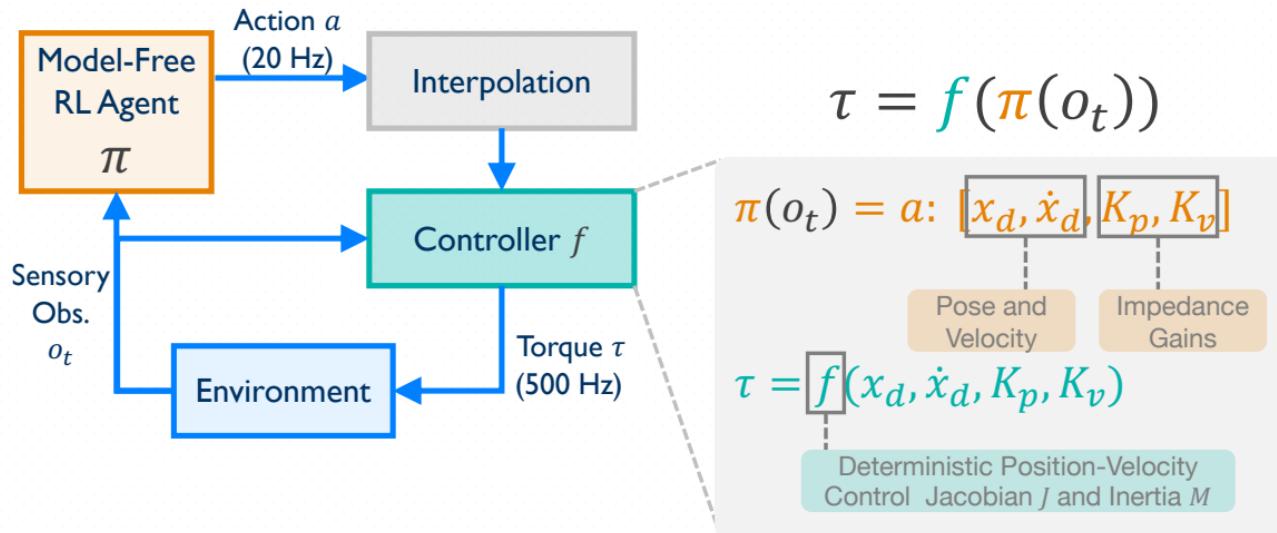
$$\tau = f(\pi(o_t))$$

Robot Model
(Deterministic)



Visuo-Motor Skills: Our Approach

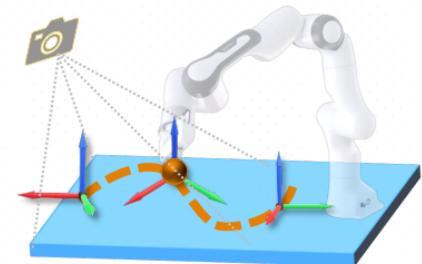
RL with Variable Impedance Task-Space



- + Model Free: No Environment Model
- + State is Image

- + Leverages Robot Model
- + Compliant Control

- + Sample Efficient
- + Transferable



Visuo-Motor Skills: Action Representation

Surface Wiping

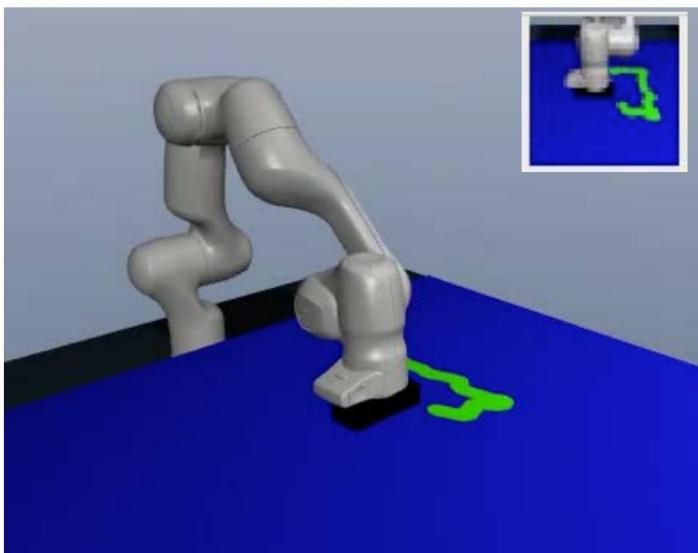
Input: Image (48x48)

Minimize the number
of Dirty Tiles

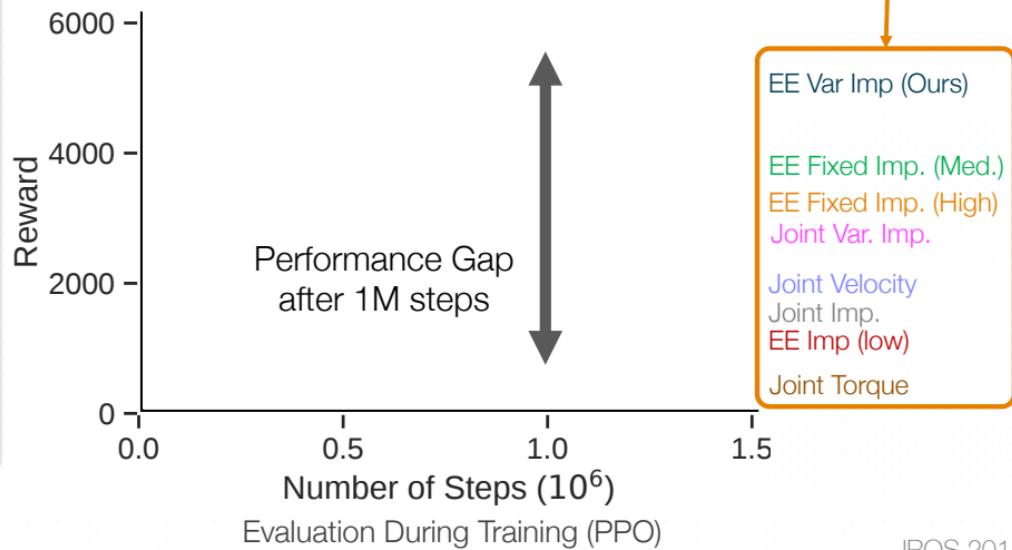
Maintain Contact
with the Table

Don't push with more
than Robot Payload

$$\text{Reward: } \lambda_1 \sum(\text{dirt_on_table}) + \lambda_2 (\text{distance_to_table}) - \lambda_3 \mathbb{I}(F \geq 40N)$$



Trained Policy Rollout (Ours)



Visuo-Motor Skills: Action Representation



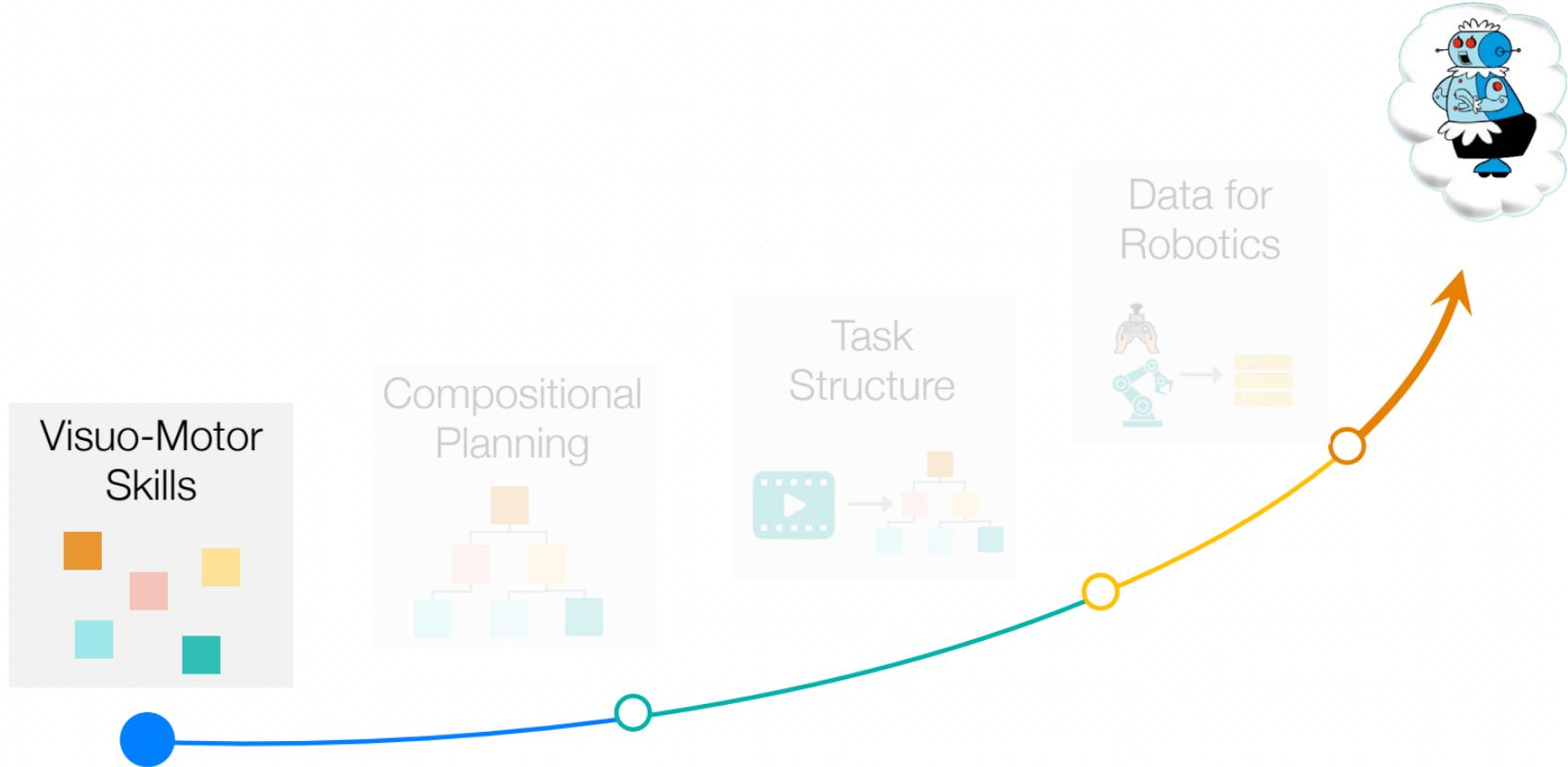
$$\tau = f_{Sim}(\pi(o_t))$$



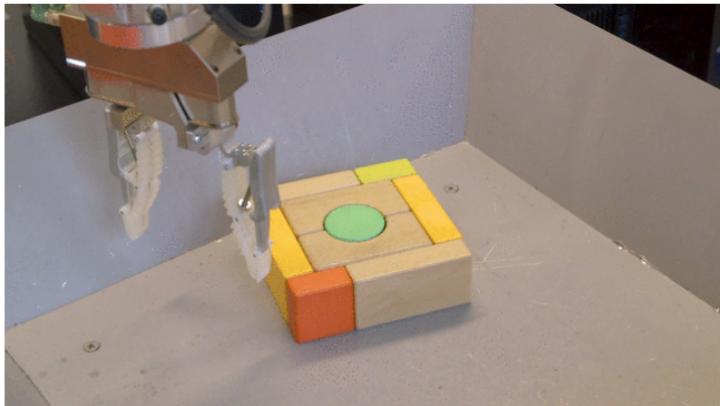
$$\tau = f_{Real}(\pi(o_t))$$

Success 80% (10 Trials)

Generalizable Autonomy in Robot Manipulation



Skills: Imitation from Heuristics



Promise of Deep RL
closed loop-control with images



...albeit, with a lot of training

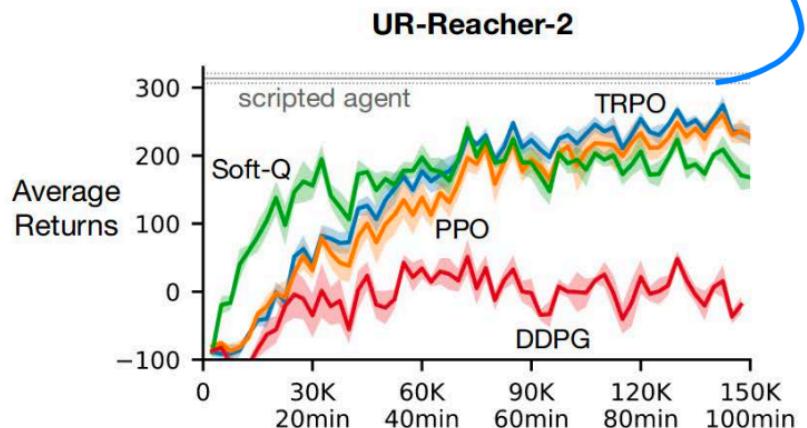
[Kalashnikov et al (2018), Levine et al. (2016), Pinto et al. (2016), Kalashnikov et al. (2018),
Yu et al. (2016), Haarnoja et al. (2018), Lee et al. (2019), Vecerik et al. (2017)]

Skills: Heuristics often beat RL

RL struggles with structured, multi-step skills



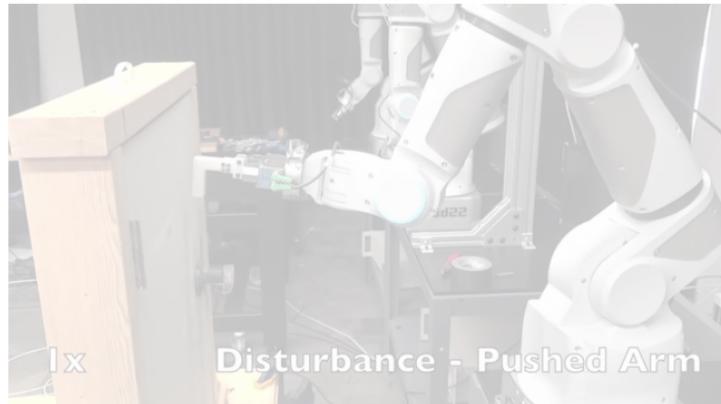
Even simple heuristics beat RL



Skills: Exploration without Guidance



Random Exploration is slow



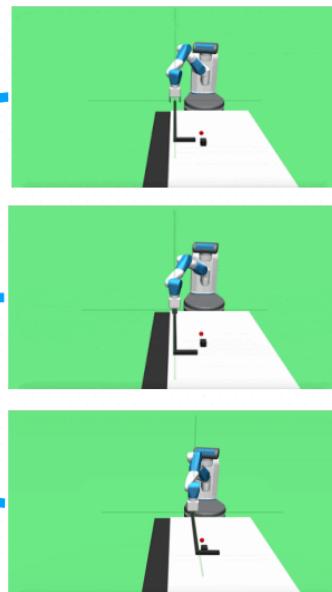
...even when first steps are obvious

Can Human Intuition **Guide** Exploration?

Skills: Imitation from Heuristics



Teachers



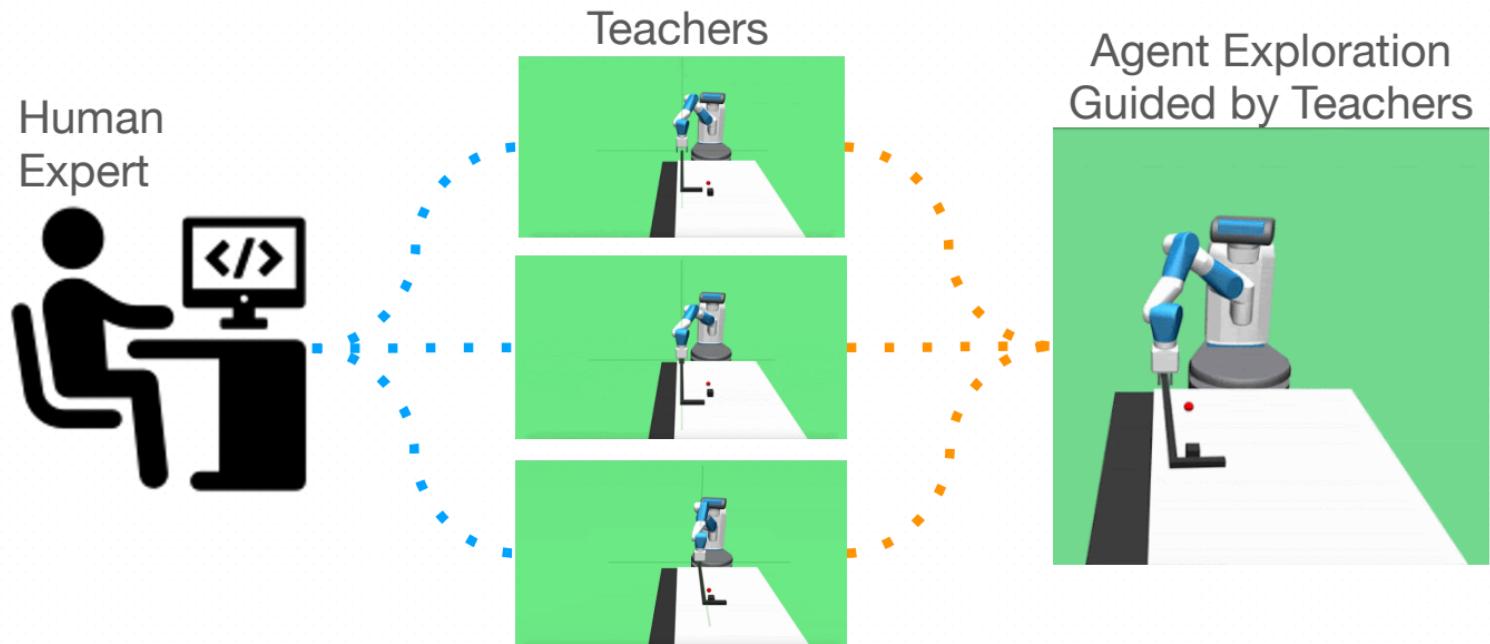
Intuition

Implement Useful Skills
...but not full solution

Teachers

Black-box controllers
solving parts of the task

Skills: Imitation from Heuristics

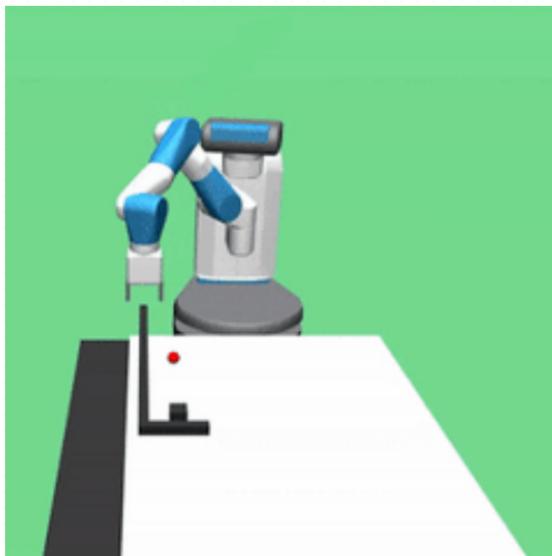


Goals: A) faster agent training B) optimal test-time agent performance

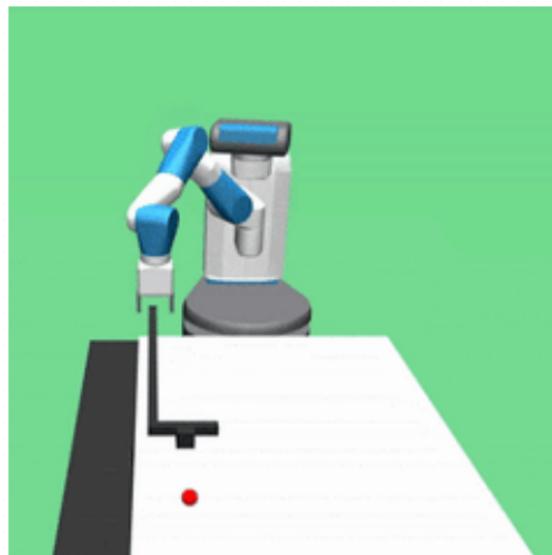
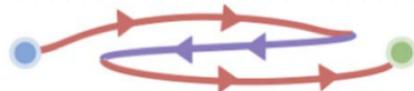
Skills: Imitation from Heuristics

Naive action choice might not work well!

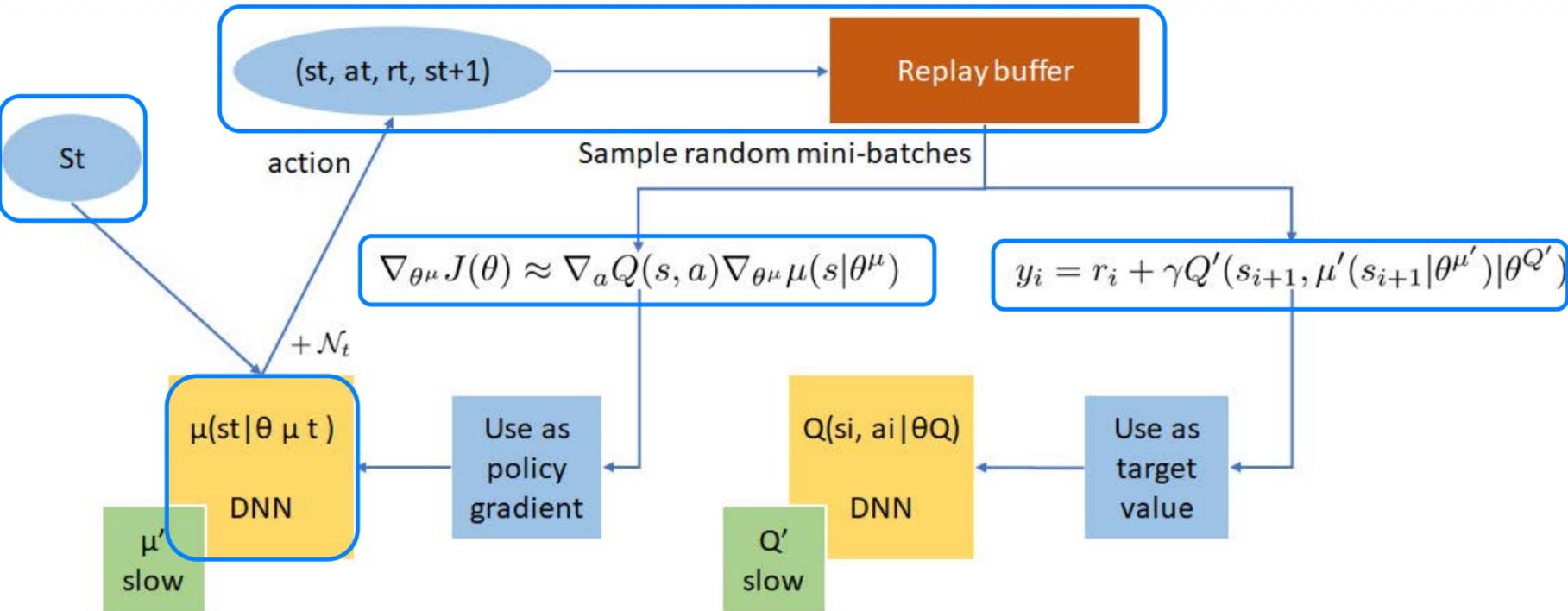
Partial



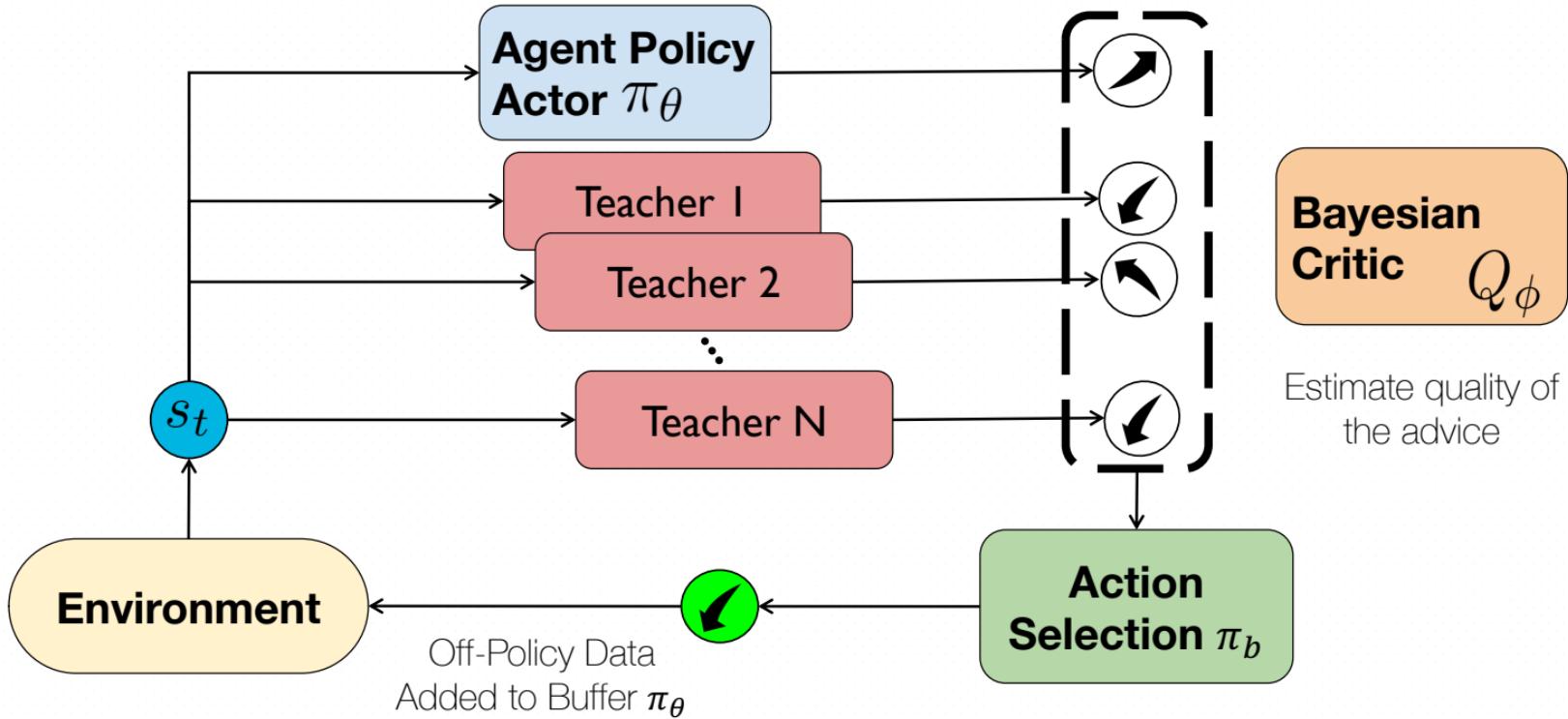
Contradictory



Off-Policy RL: DDPG Review

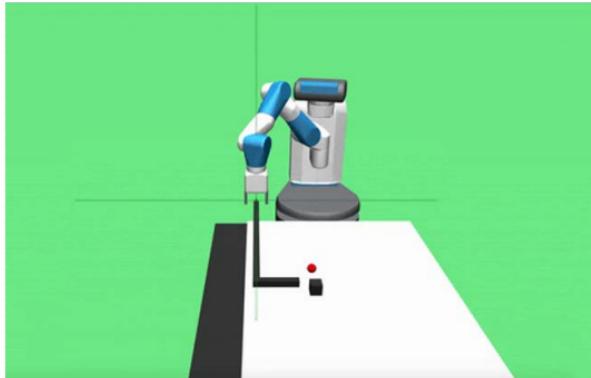


AC-Teach: Actor-Critic with Teachers



Experiments

Task:



Teachers:



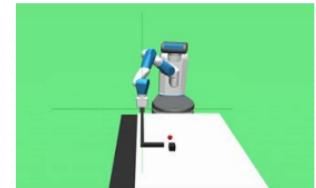
grab hook



position hook



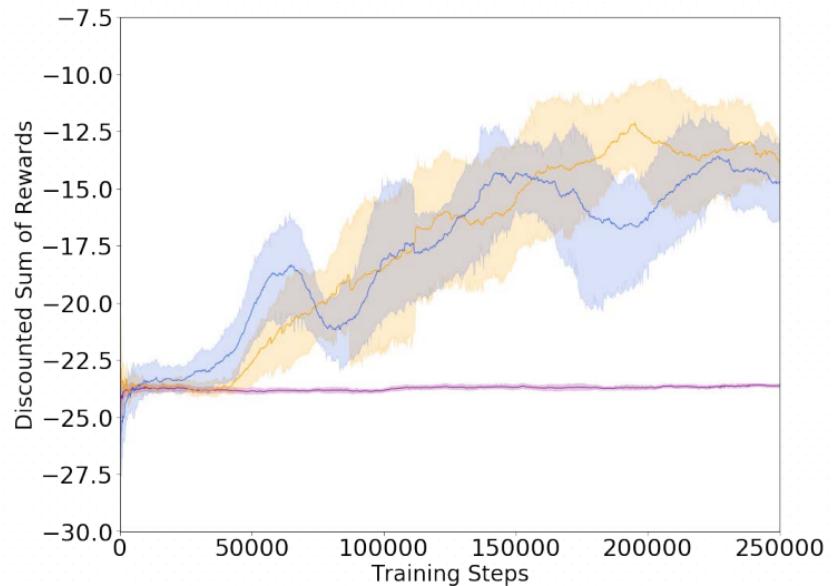
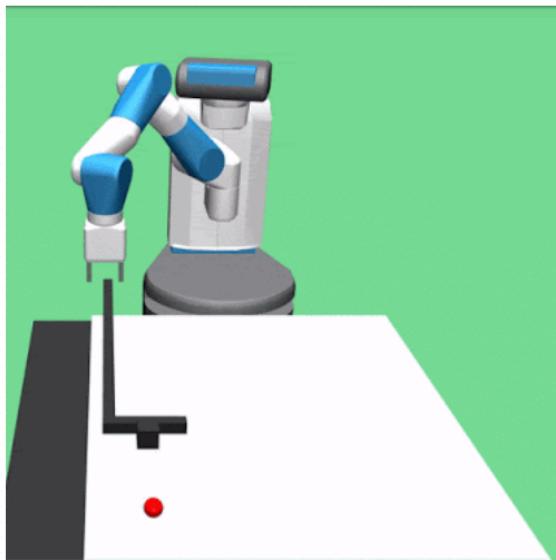
pull



push

Results

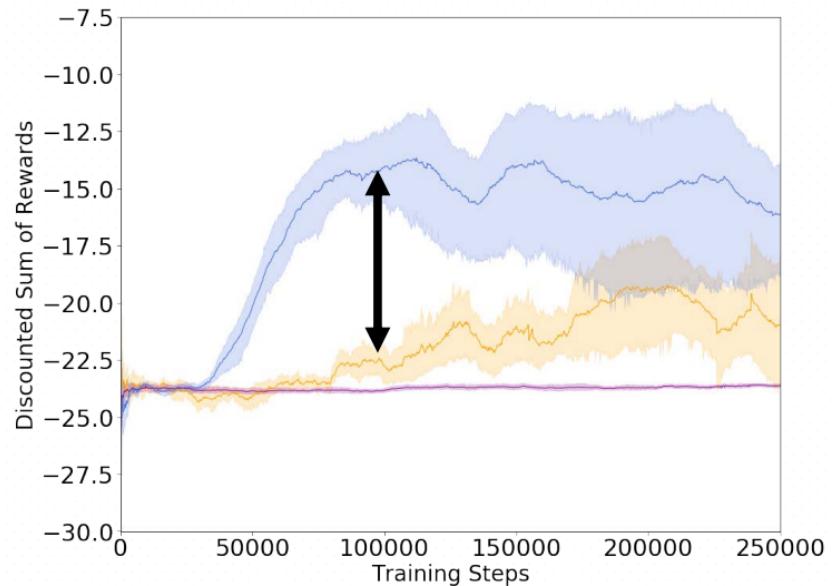
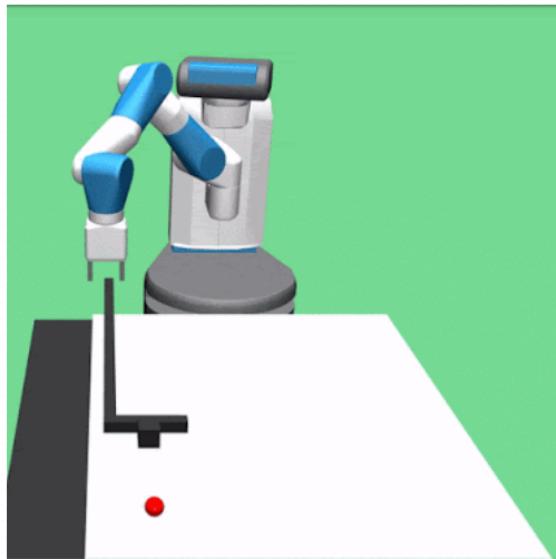
— B-DDPG + AC-Teach (ours) — B-DDPG + DQN — B-DDPG (no teachers)



AC-Teach is able to leverage a single teacher well

Results

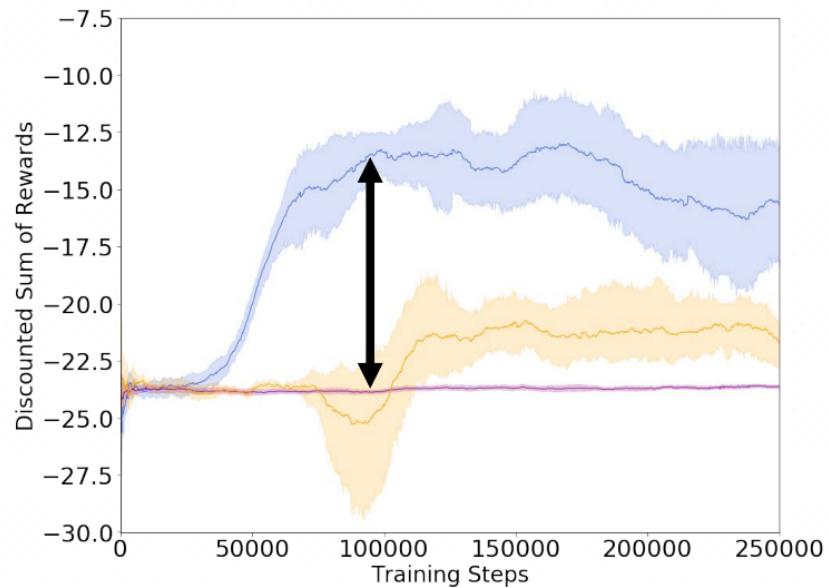
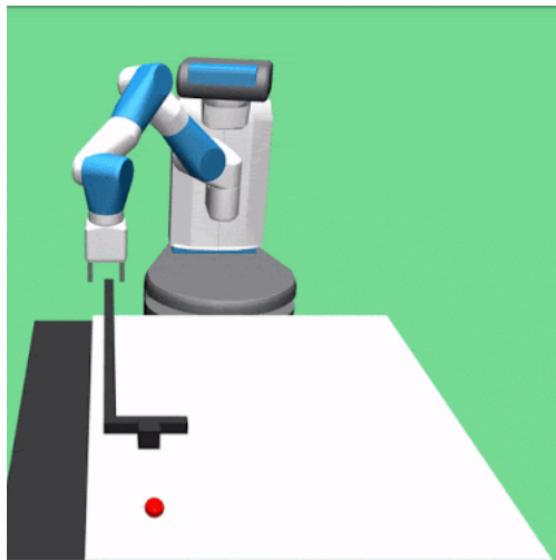
— B-DDPG + AC-Teach (ours) — B-DDPG + DQN — B-DDPG (no teachers)



AC-Teach speeds up training given multiple teachers

Results

— B-DDPG + AC-Teach (ours) — B-DDPG + DQN — B-DDPG (no teachers)



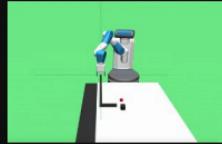
AC-Teach has agent learn behaviors not in teacher set

Visuo-Motor Skills

- Grasping
- Picking
- Wiping
- Pushing
- Open door



IROS 2019

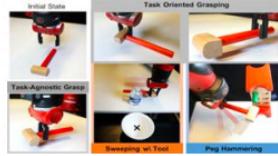
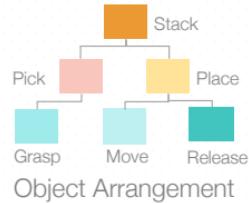


CoRL 2019

Data for
Robotics

Action Representations and Weak-Supervision provide
Visuo-Motor structure to enable learning efficiency and generalization

Generalizable Autonomy in Robot Manipulation



RSS 2018, IJRR 2019

Visuo-Motor Skills



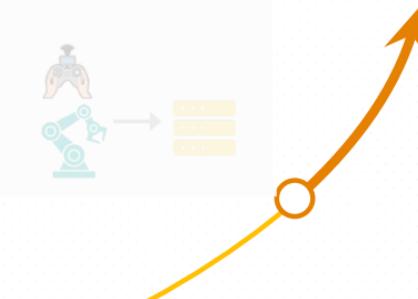
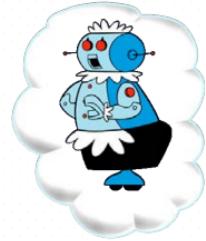
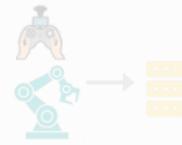
Compositional Planning



Task Structure



Data for Robotics



Sequential Skills



Skills: Surface Wiping

Primitive Skills

Grasping

Pushing

Picking

Wiping

Open door



Skills: Tool Use

Sequential Skills

— Hammering (with unknown objects)

— Cutting (with new knife)

— Sweeping (with new broom)

Sequential Skills: Manipulation with Tools

Tool-Use

Initial State



Unknown Object

Task-Agnostic
Grasping¹



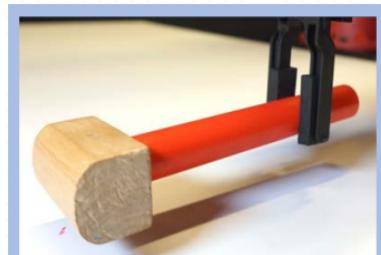
Optimizes for Grasp
Success Only

Suboptimal for Task!

Task-Oriented Grasping



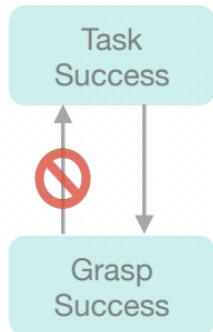
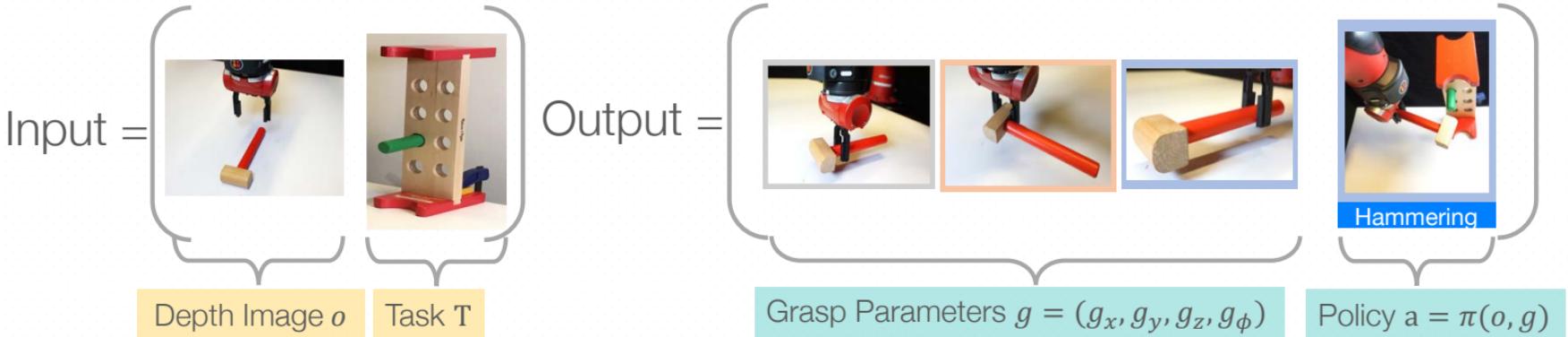
Sweeping



Hammering

¹ Pinto et al. '16, Levine et al. '16, Mahler et al. '18, Kalashnikov et al. '18

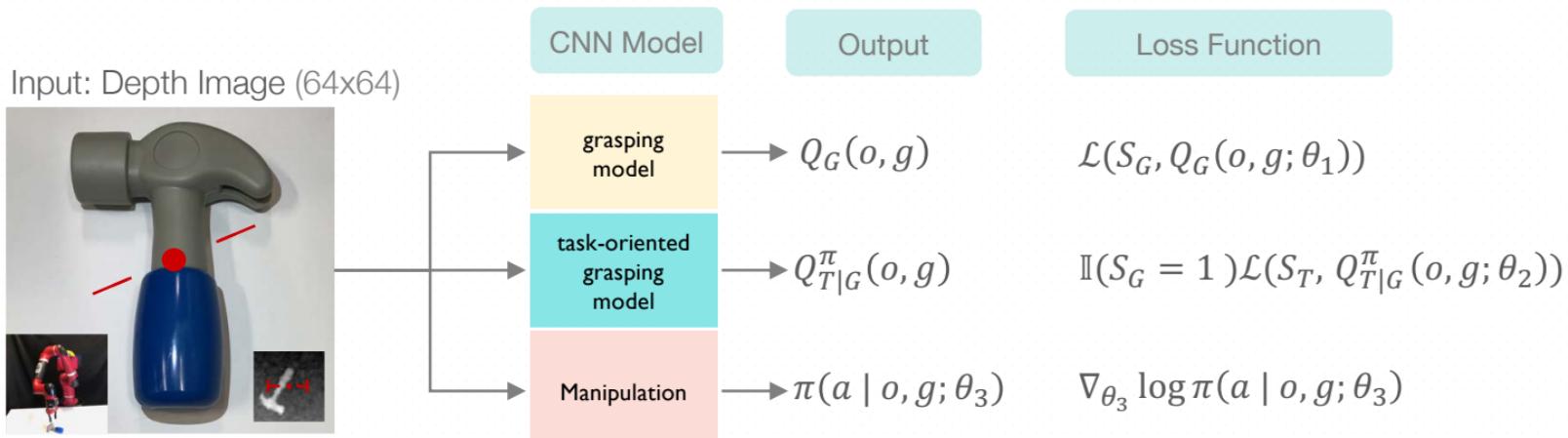
Visuo-Motor Skills: Task-Oriented Grasping



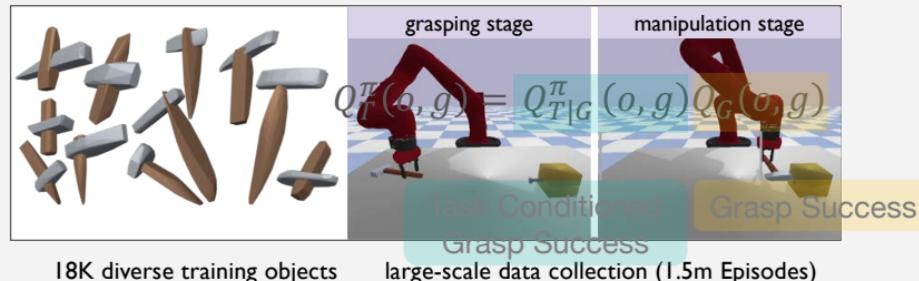
$$g^*, \pi^* = \underset{g, \pi}{\operatorname{argmax}} Q_T^\pi(o, g) \quad \text{Score Function}$$
$$Q_T^\pi(o, g) = P_\pi(S_T = 1 | S_G, \pi) \mathbf{1}_{\text{pogg}} P(S_G = 1 | o, g)$$
$$Q_T^\pi(o, g) = Q_{\text{task success}}^\pi(o, g) Q_{S_G}(o, g)$$

Task Conditioned Grasp Success Grasp Success

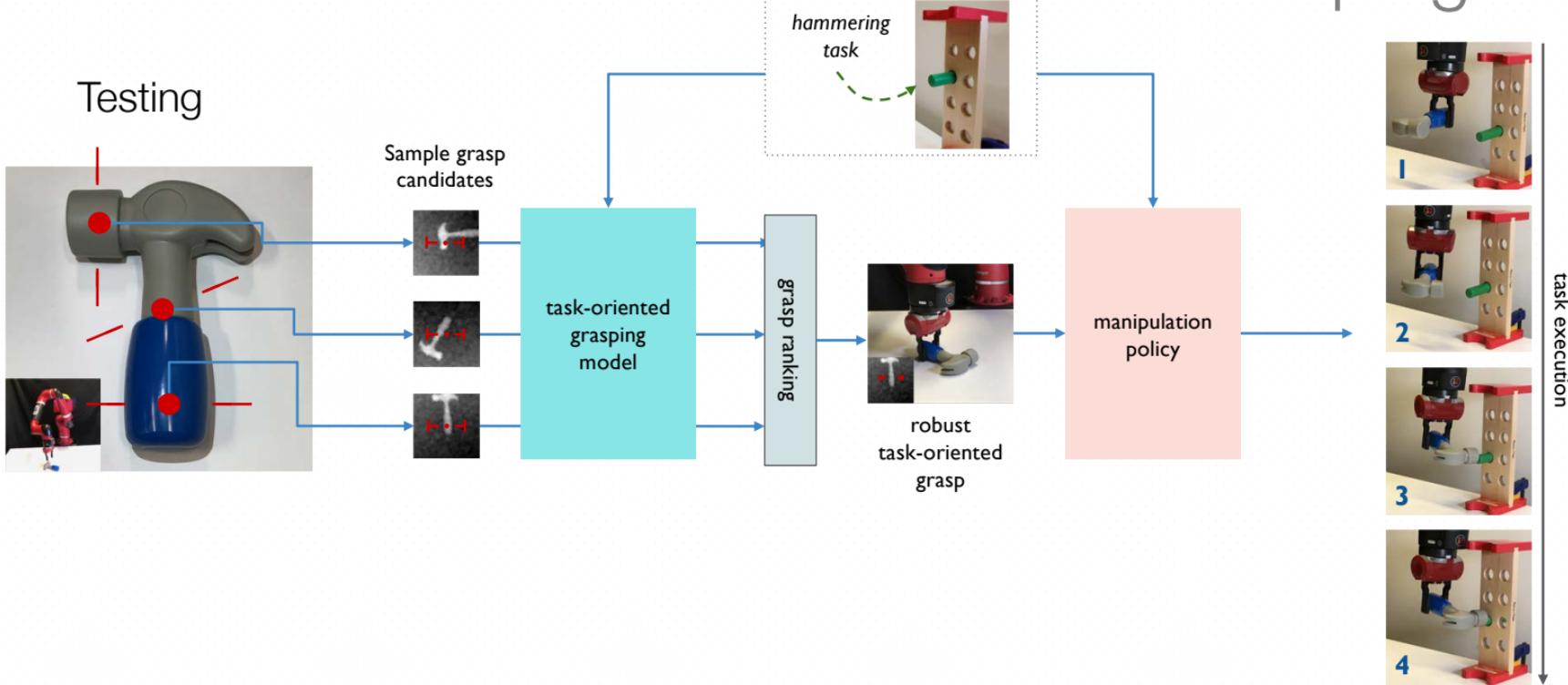
Visuo-Motor Skills: Task-Oriented Grasping



Training



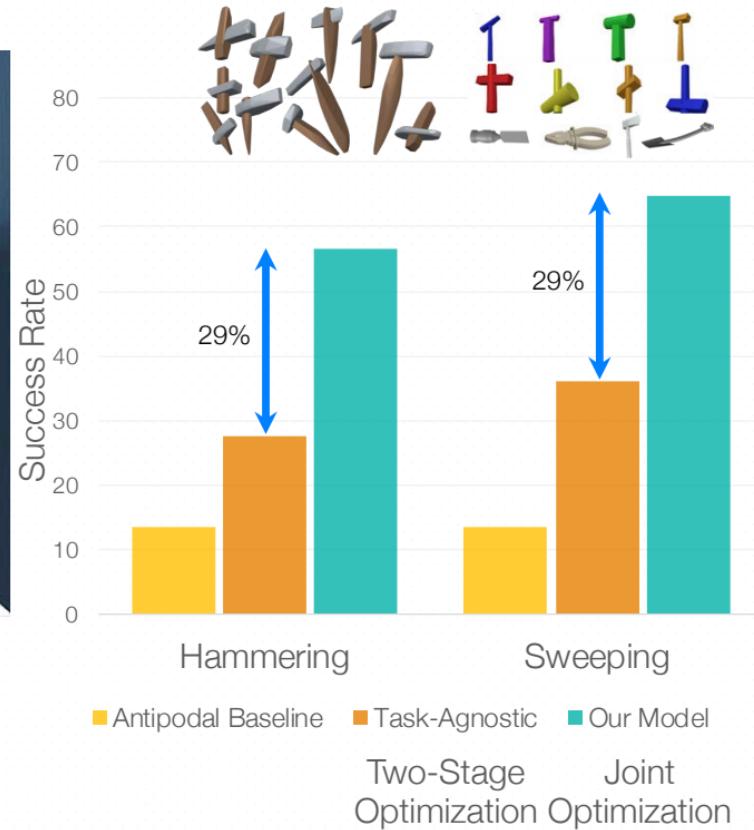
Visuo-Motor Skills: Task-Oriented Grasping



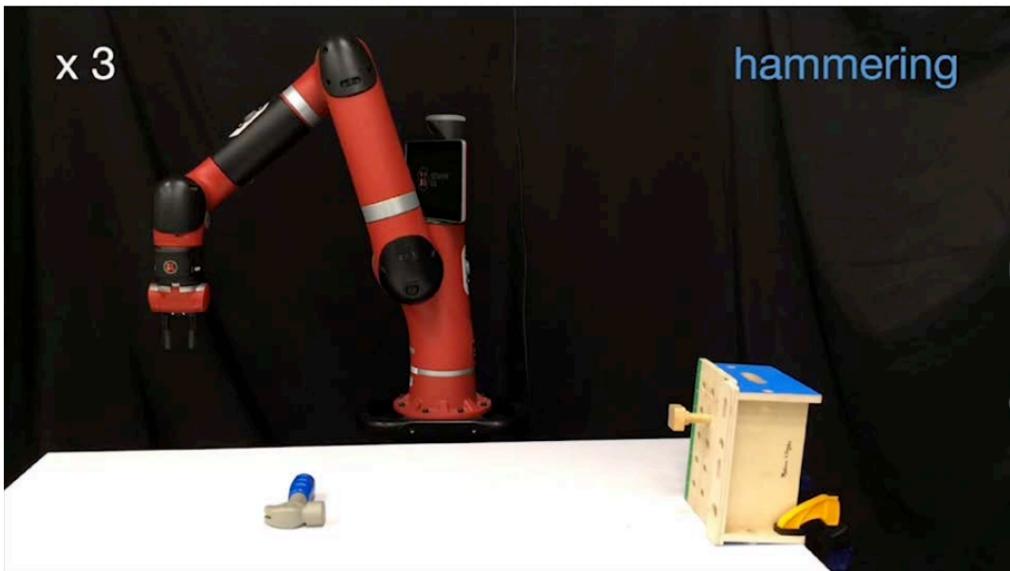
Sequential Skills: Task-Oriented Grasping



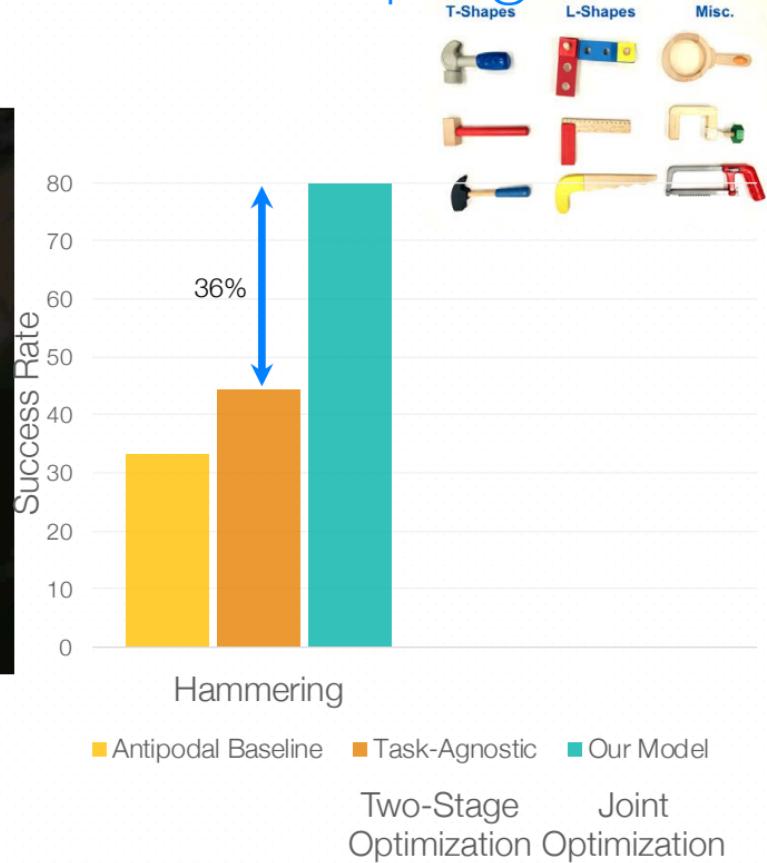
Trained Policy Rollout (Ours)
Unseen Test Objects



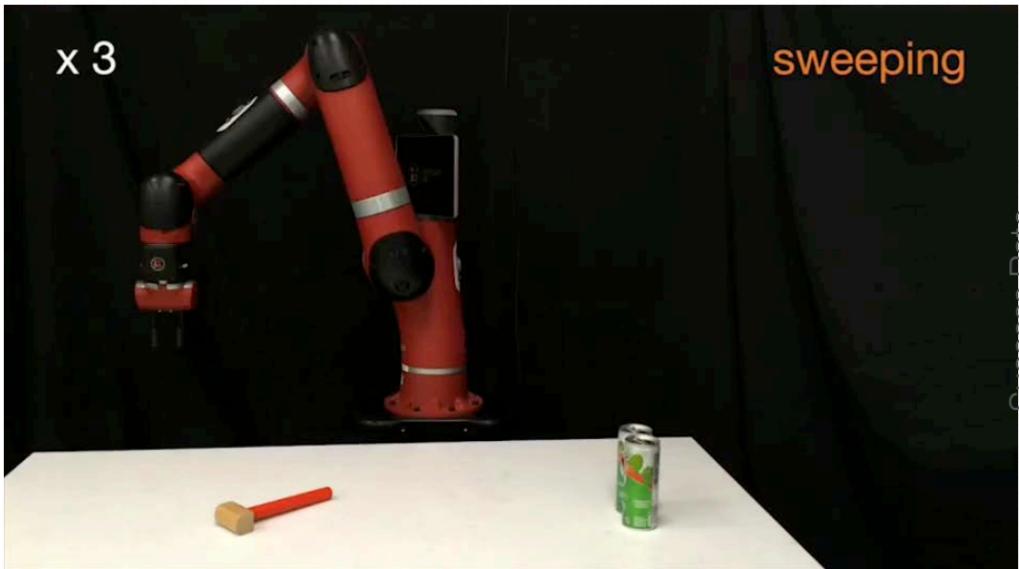
Sequential Skills: Task-Oriented Grasping



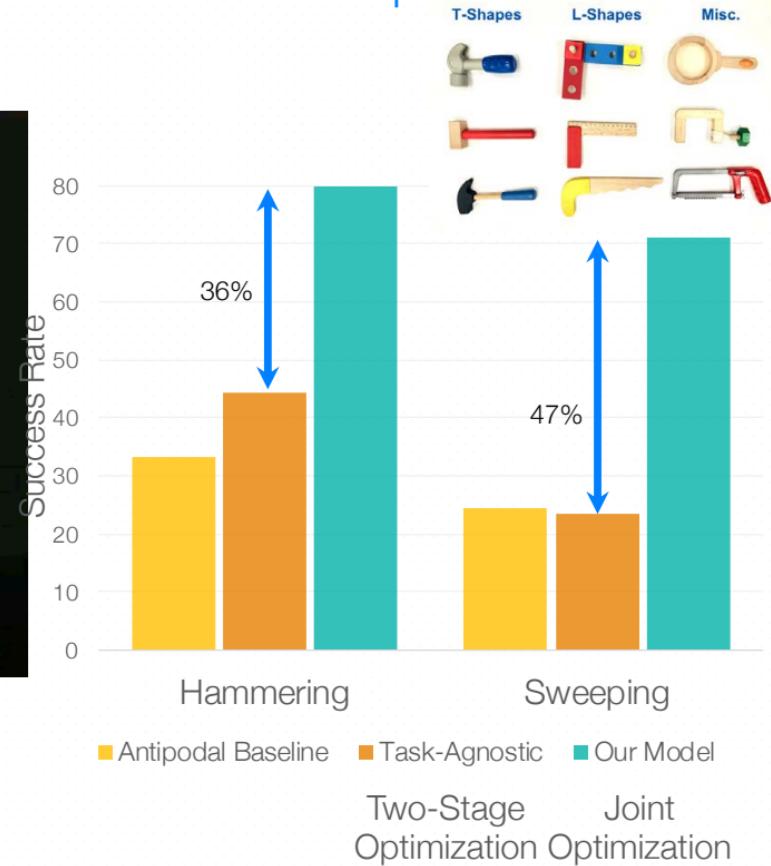
Trained Policy Rollout (Ours)
Unseen Test Objects



Sequential Skills: Task-Oriented Grasping



Trained Policy Rollout (Ours)
Unseen Test Objects



Sequential Skills



Skills: Surface Wiping

Primitive Skills

Grasping

Pushing

Picking

Wiping

Open door



Skills: Tool Use

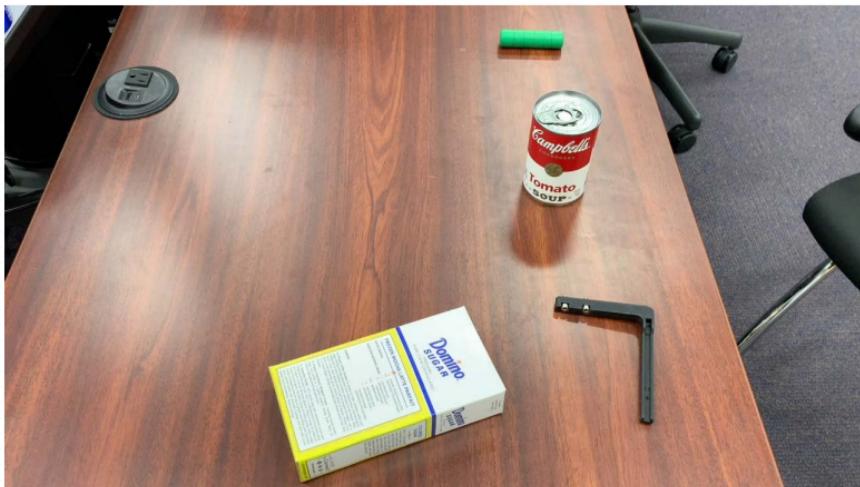
Sequential Skills

— Hammering (with unknown objects)

— Cutting (with new knife)

— Sweeping (with new broom)

Sequential Skills: Multi-Step Reasoning



Skills: Multi-Step Reasoning

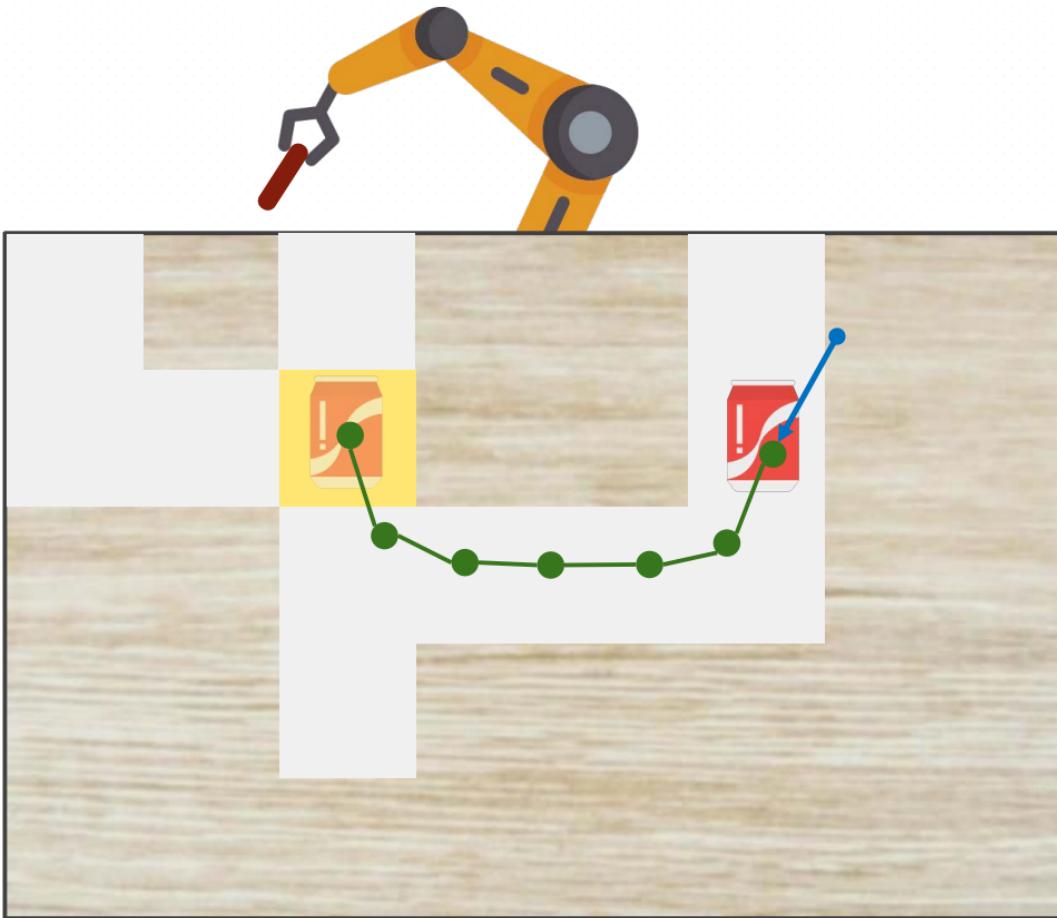


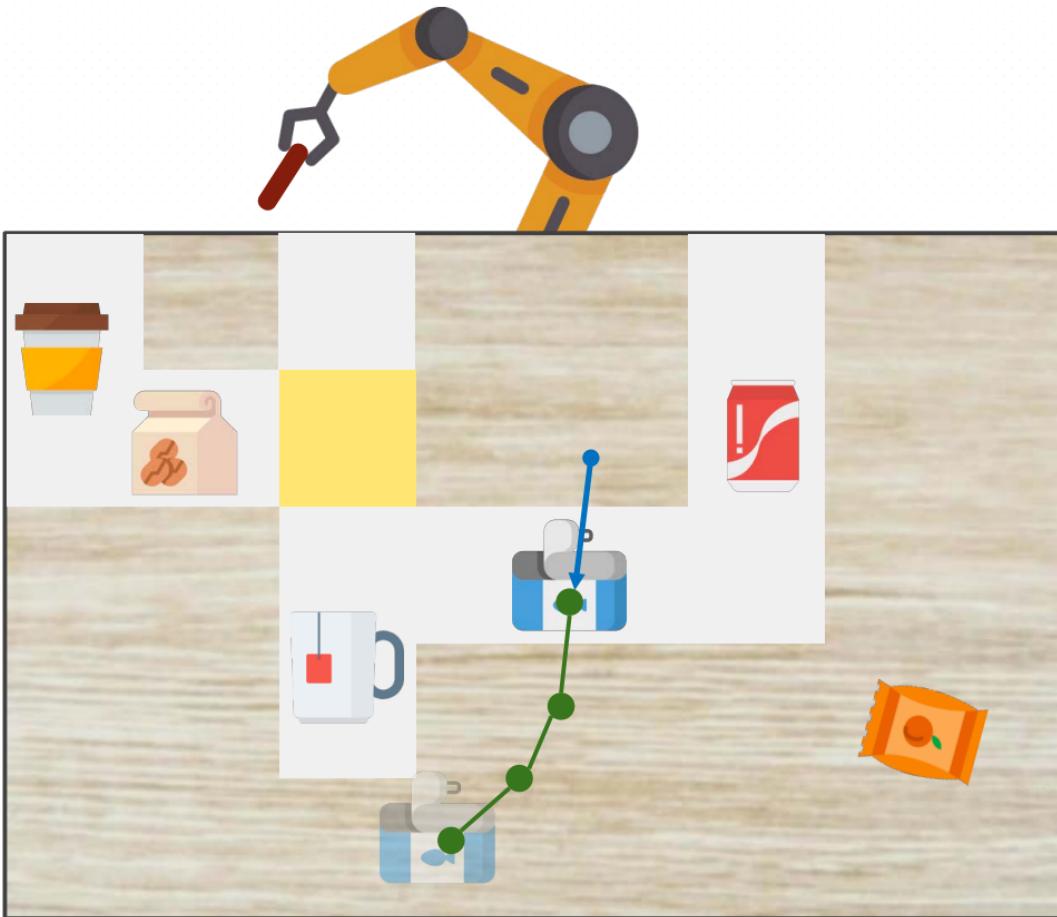
Generalization

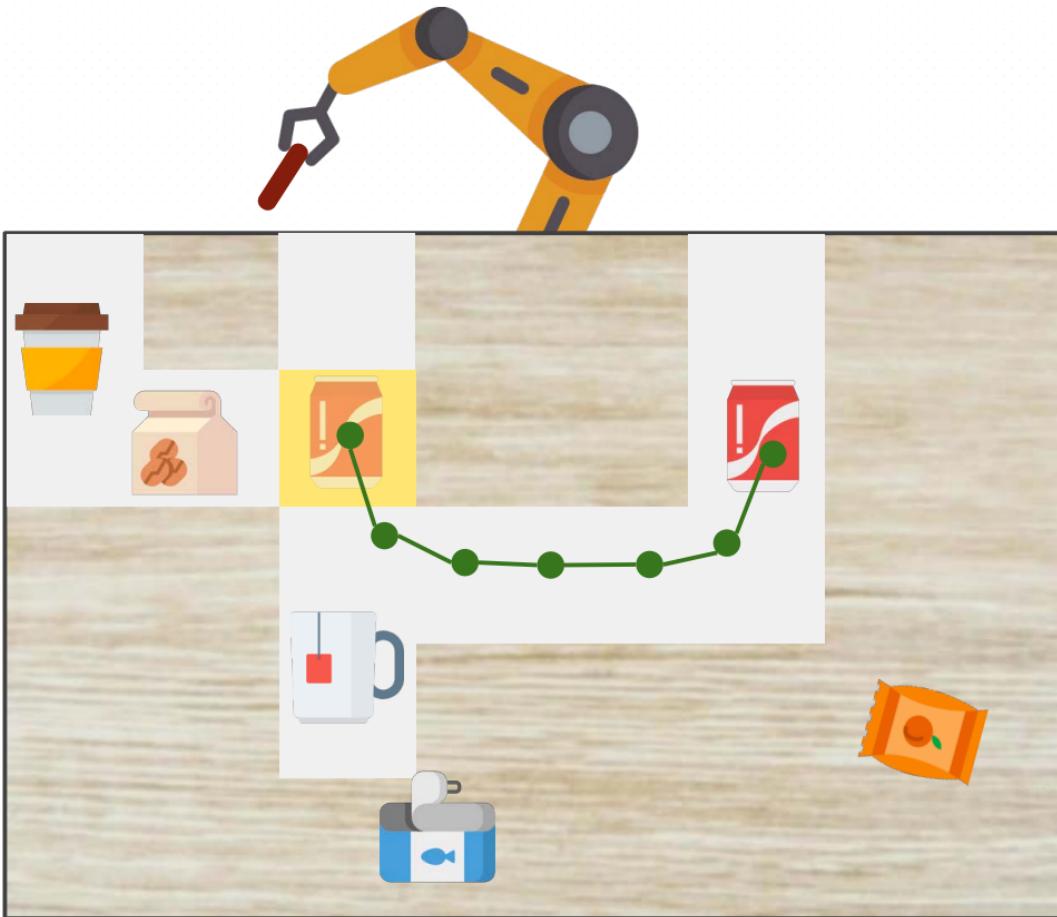




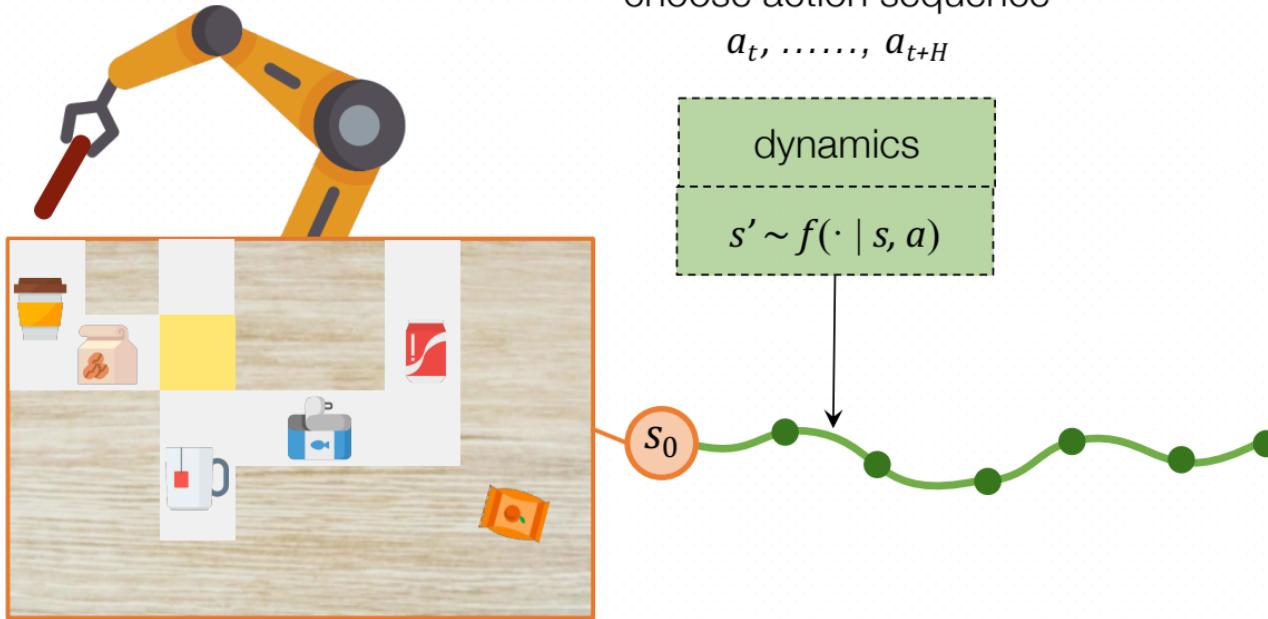
Can we learn multi-step reasoning in robotics
under physical and semantic constraints







Model-based learning



[Deisenroth et al, RSS'07], [Guo et al, NeurIPS'14], [Watter et al, NeurIPS'15], [Finn et al, ICRA'17],

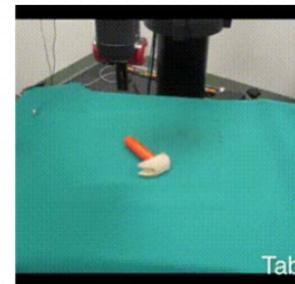
Model-based learning



data ↑
learning ↑



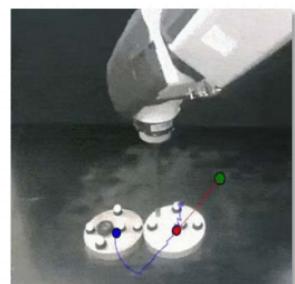
[Deisenroth et al. RSS'07]



[Agrawal et al. ICRA'16]

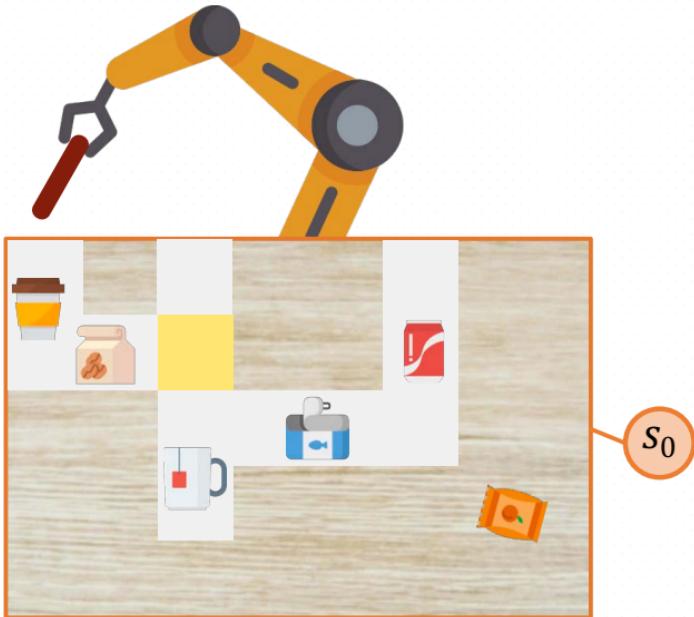


[Ebert et al. CoRL'17]



[Janer et al. ICRA'19]

CAVIN: Hierarchical planning in learned latent spaces



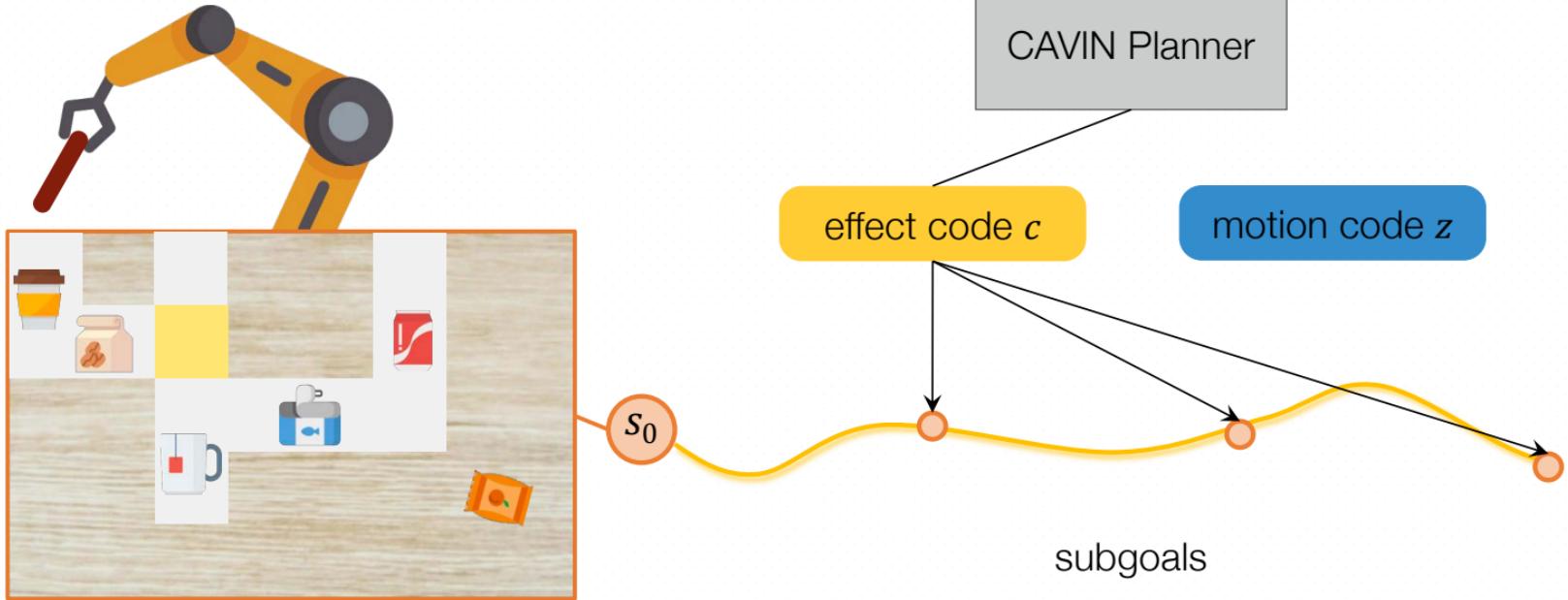
CAVIN Planner

effect code c

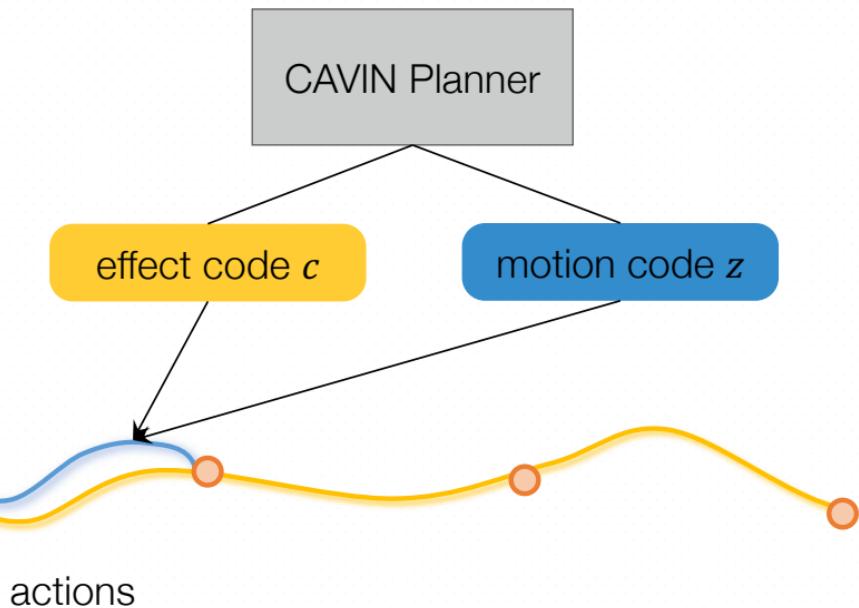
motion code z

Leverage **Hierarchical Abstraction** in Action Space
Without **Hierarchical Supervision**

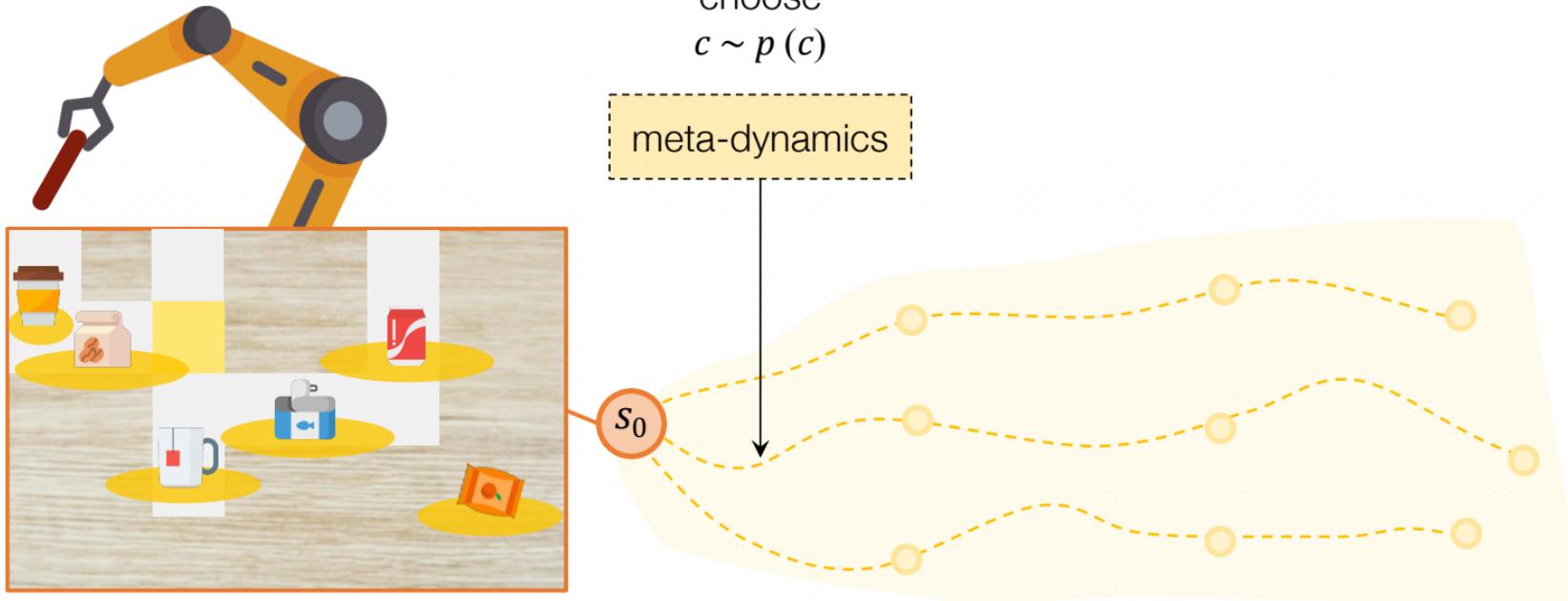
CAVIN: Hierarchical planning in learned latent spaces



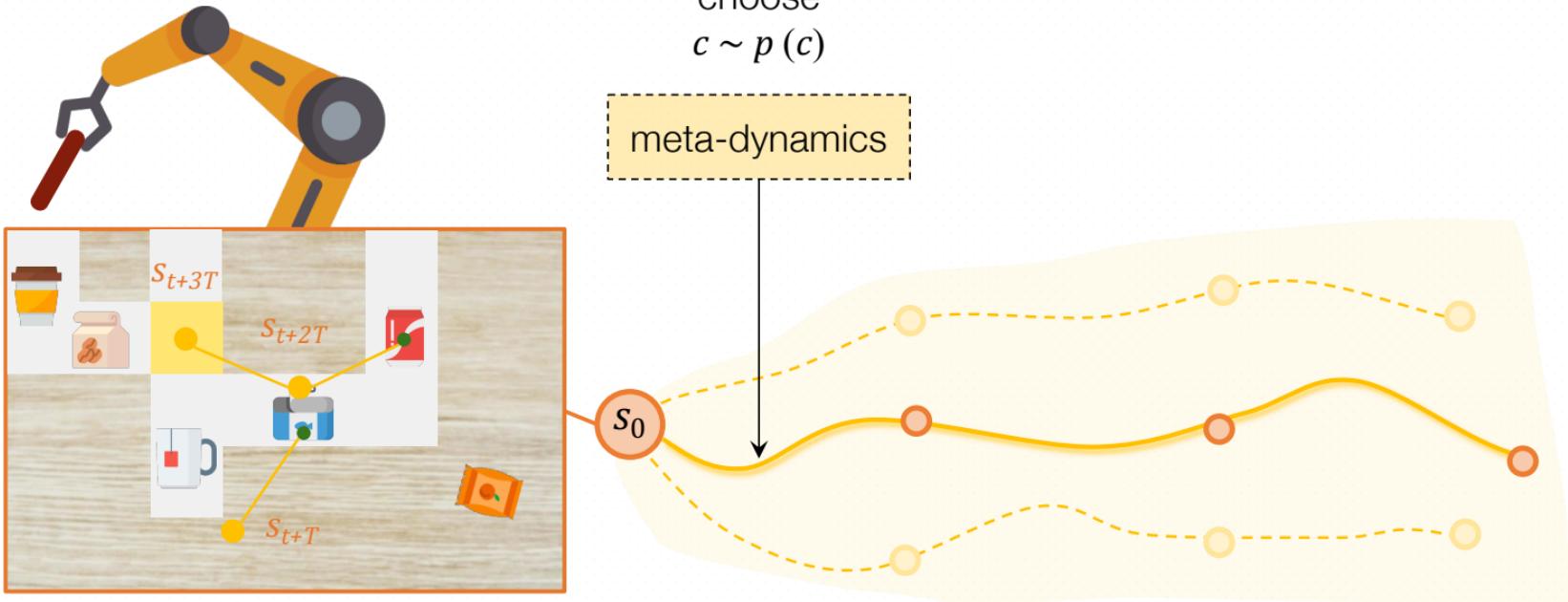
CAVIN: Hierarchical planning in learned latent spaces



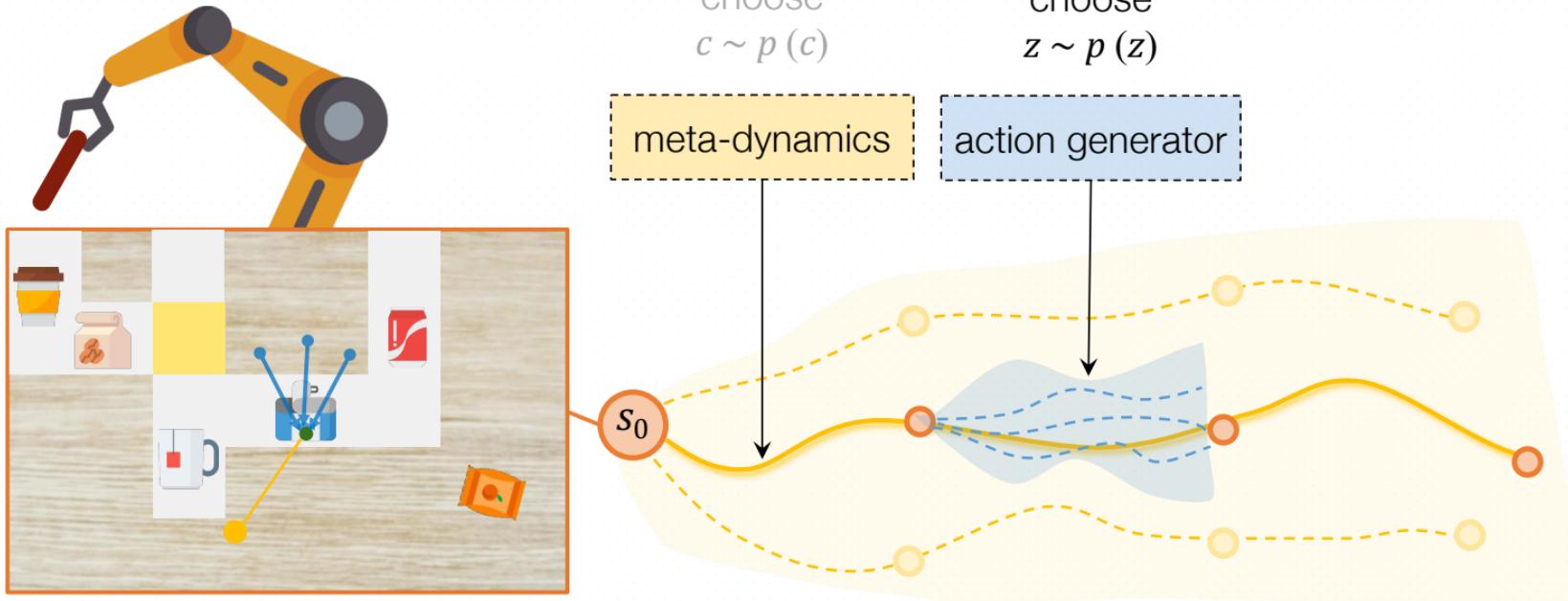
CAVIN: Hierarchical planning in learned latent spaces



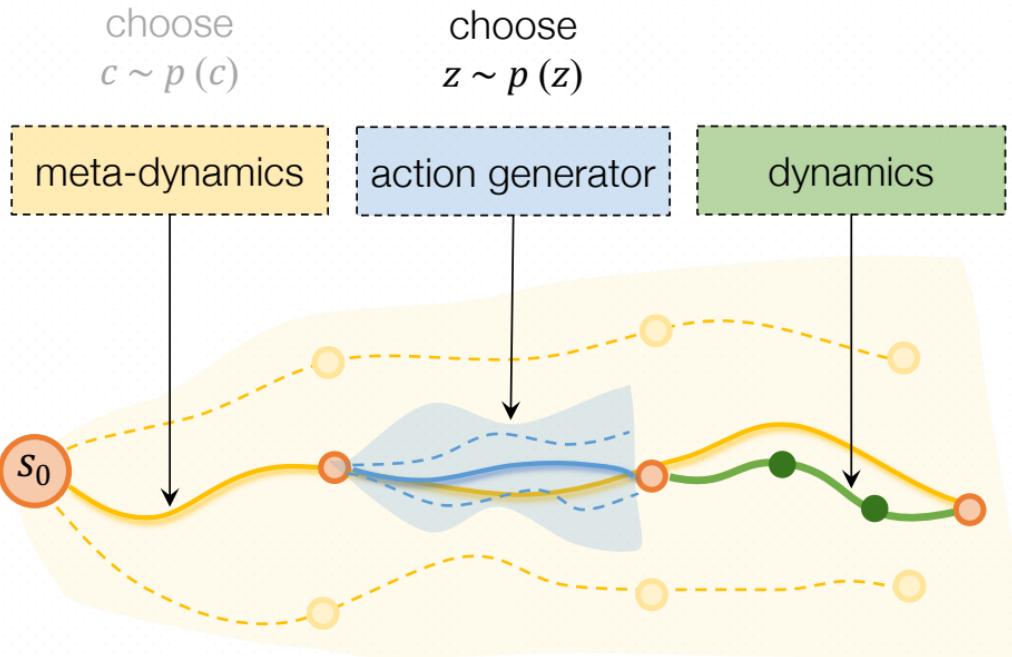
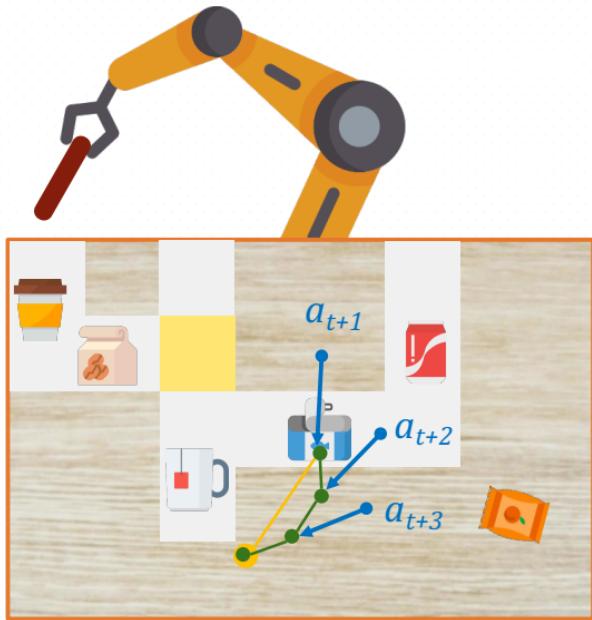
CAVIN: Hierarchical planning in learned latent spaces



Hierarchical planning in learned latent spaces

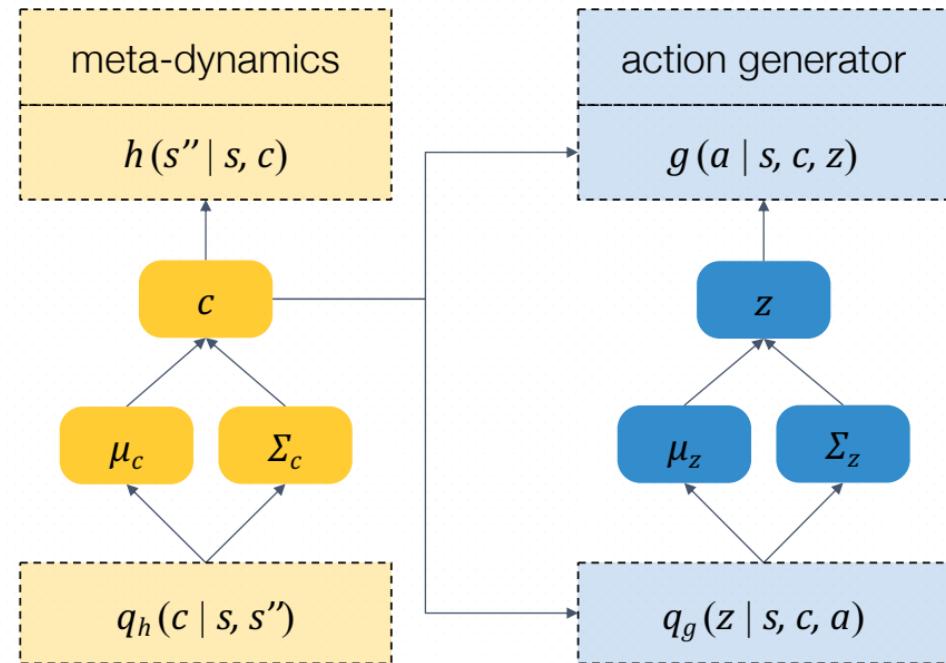
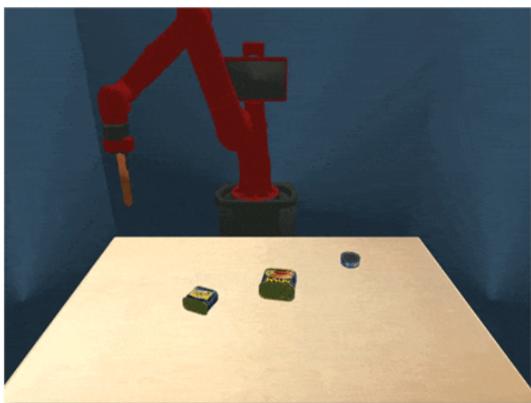


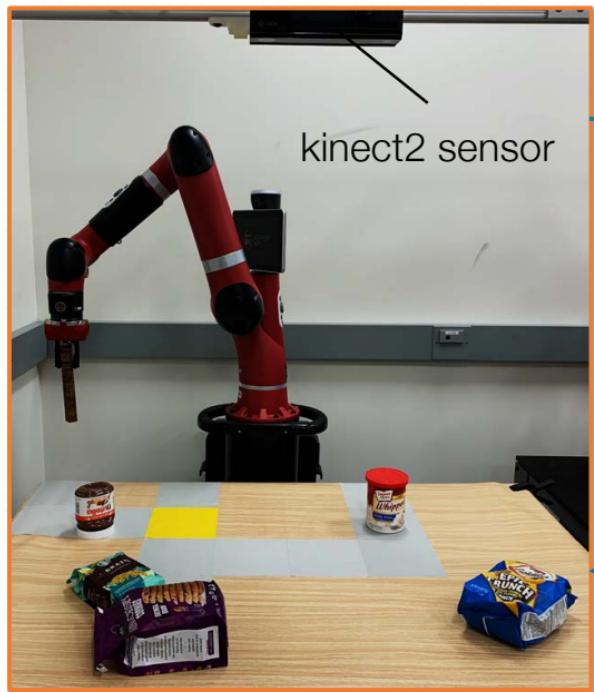
CAVIN: Hierarchical planning in learned latent spaces



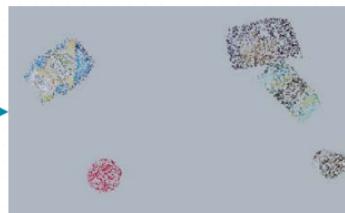
Learning with cascaded variational inference

task-agnostic interaction





visual observation



preprocess



CAVIN Planner

action
[$x, y, \Delta x, \Delta y$]

Tasks

clearing



Clear all objects within the area of **blue tiles**.

insertion



Move the target to the goal without traversing **red tiles**.

crossing



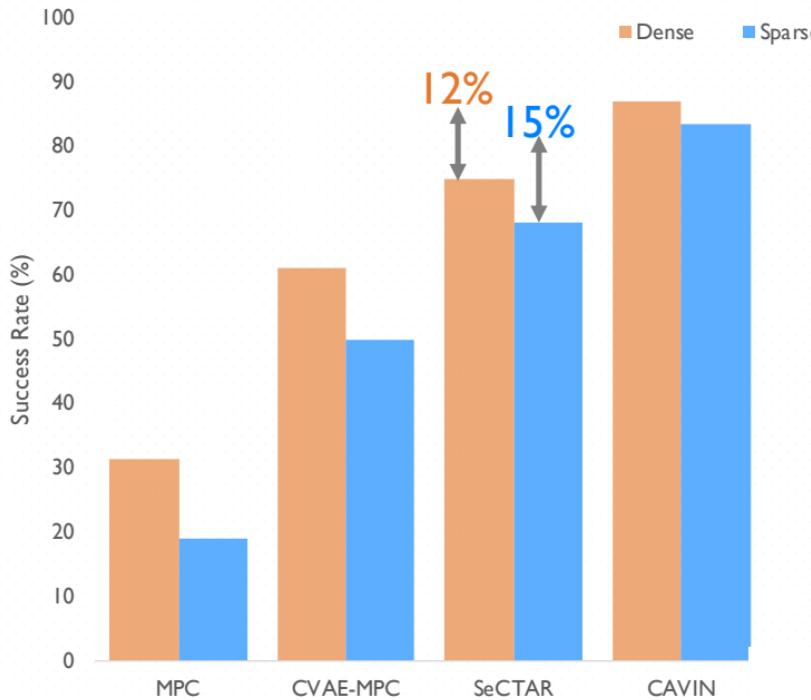
Move the target to the goal across **grey tiles**.

Real

Simulated



Quantitative Evaluation

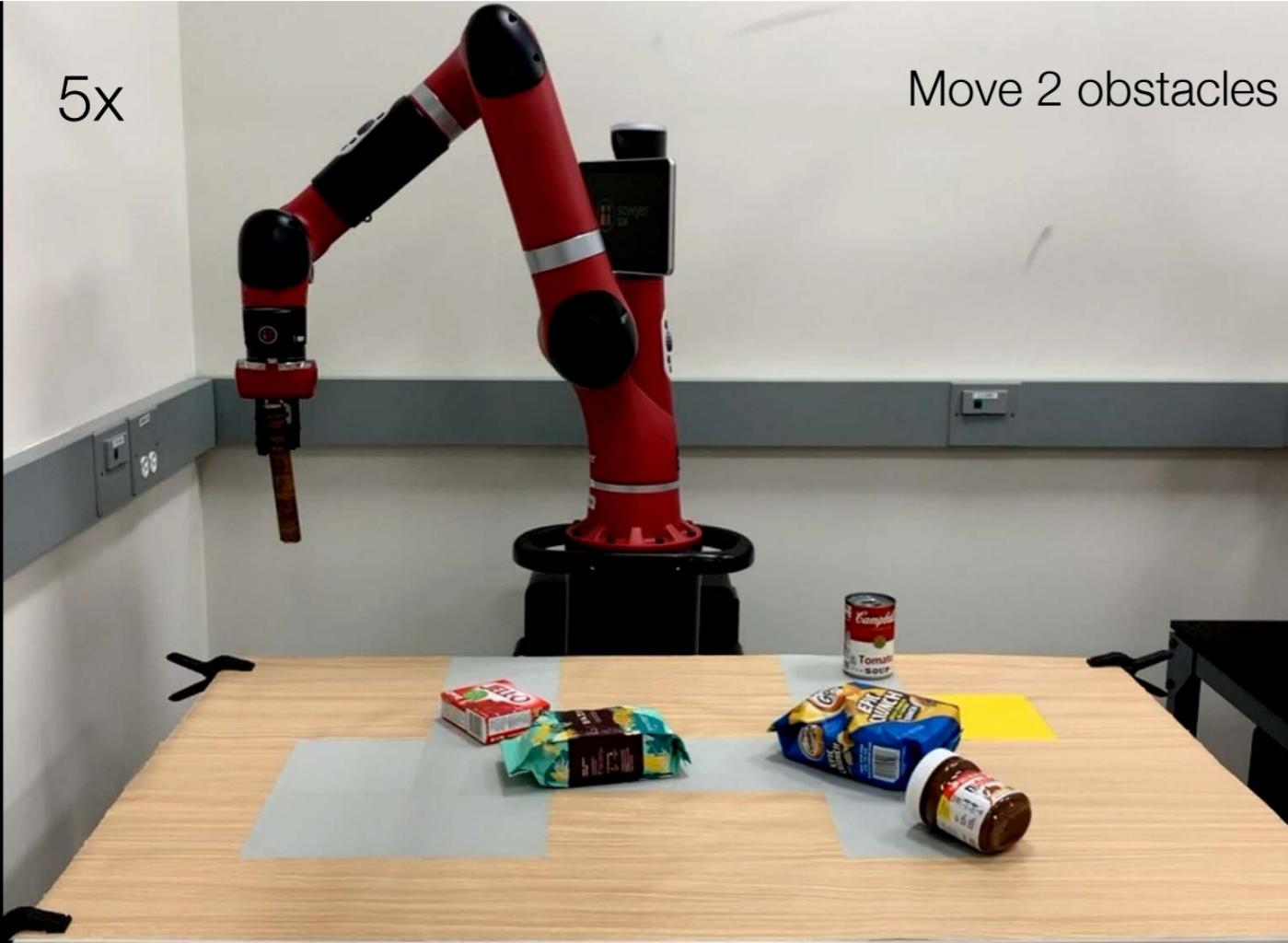


Hierarchical Latent space dyn.
↓
Better performance with sparse
reward signal

Averaged over 3 Tasks
with 1000 test instances each

5x

Move 2 obstacles



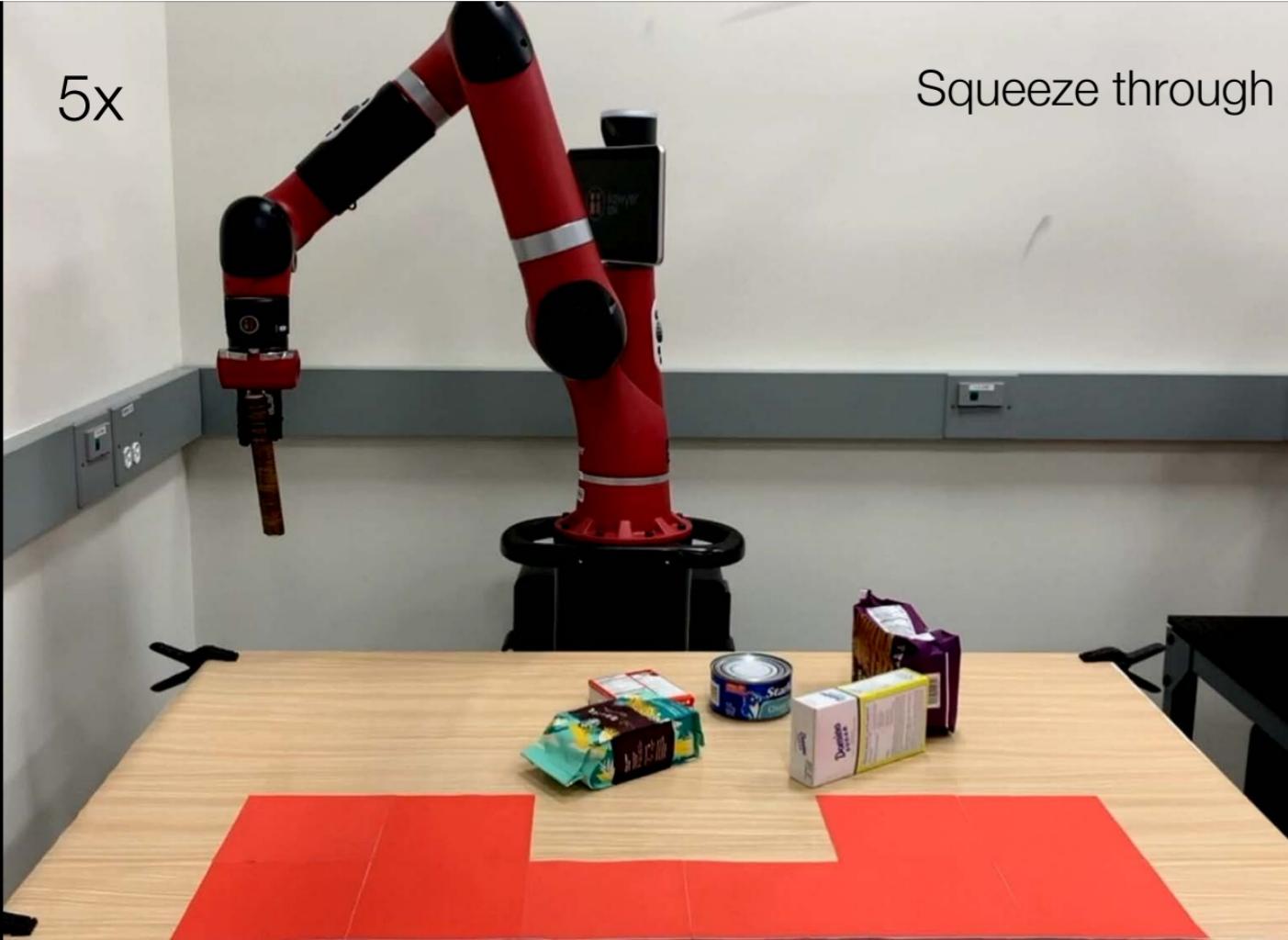
5x

Get around



5x

Squeeze through

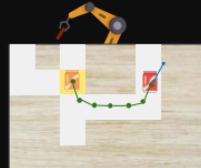
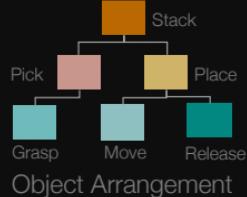


5x

Open path

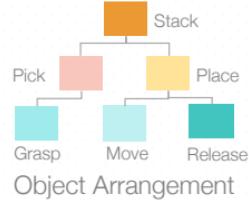


Compositional Planning

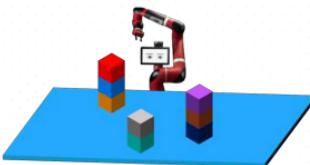


Self-Supervision and Structured Latent Variable Models
lead to good representations that generalize

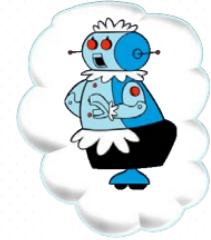
Generalizable Autonomy in Robot Manipulation



ICRA 2018



CVPR 2019 (oral)

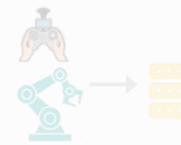
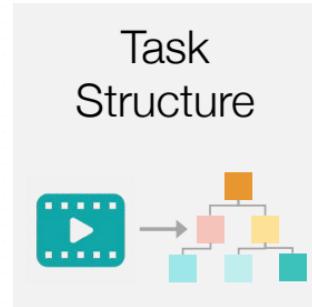


Data for
Robotics

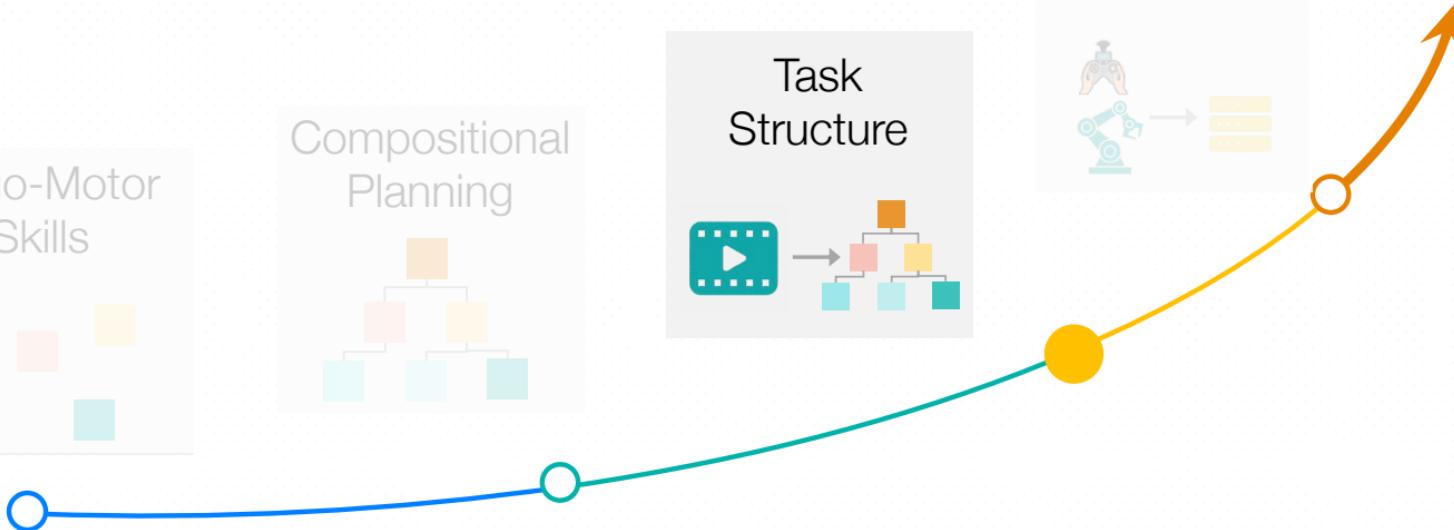
Visuo-Motor
Skills



Compositional
Planning



Robotics Data



Complex Task Structure

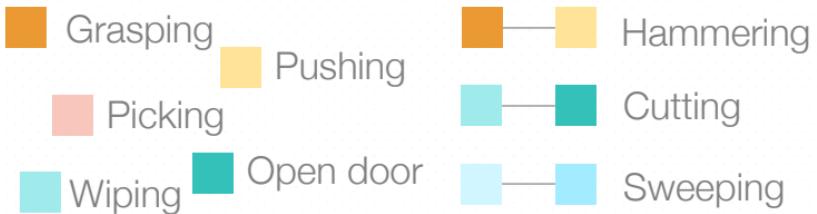


Visuo-Motor Skills

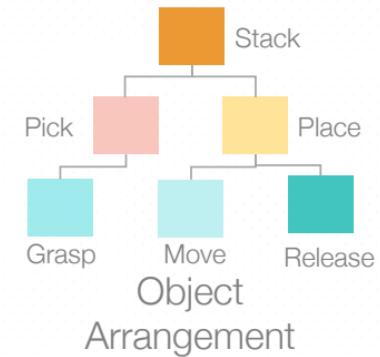
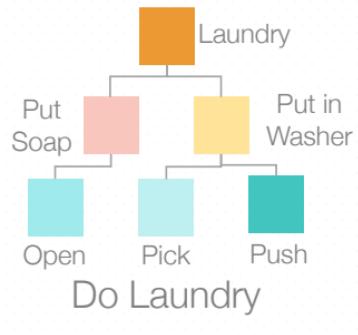


Complex Task Structure

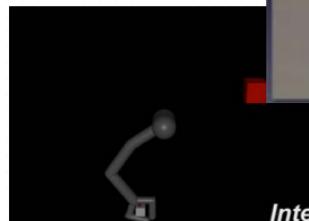
Visuo-Motor Skills



Complex Task Structure

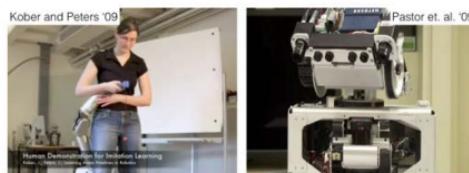


Compositional Planning: Current Paradigm



Reinforcement Learning

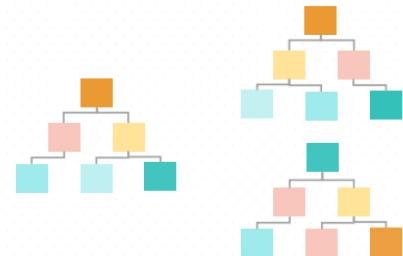
- Sample Inefficient
- Multi-step Structured Tasks
- Needs non-trivial Reward Shaping



Imitation Learning

- Task Segmentation is non-trivial
- Multi-modality of Search Space
- Fixed Permutation of Primitives

Desired



Train \neq Test

Meta Imitation Learning

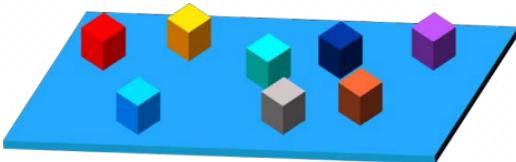
- New Task Structures
- Few-Shot performance
- Input State as Video

RL: [Schaal 1997], [Chebotar et al., '17], [Yahya et al., '16], [James et al., '17], [Popov et al., '17], [Zhu et al. 18], [Hausman et al. 18]

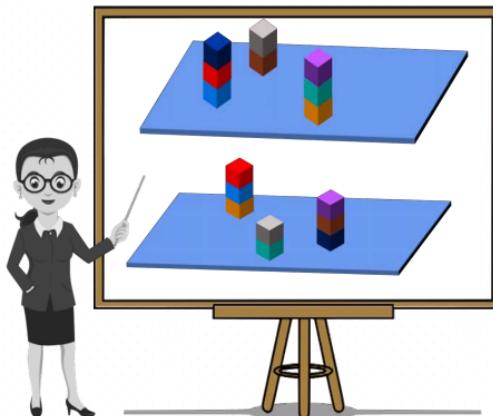
Imitation: [Calinon et al 2008], [Argall et al 2009], [Kober, Peters, et al. 09], [Pastor et al. 09], [Schulman et al. 2013], [Kroemer et al. 15], [Garg et al 2017]

Compositional Planning: Challenge

Task Domain

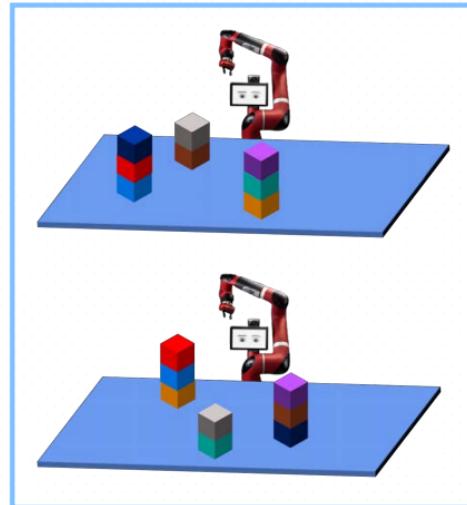


Instructional Demos



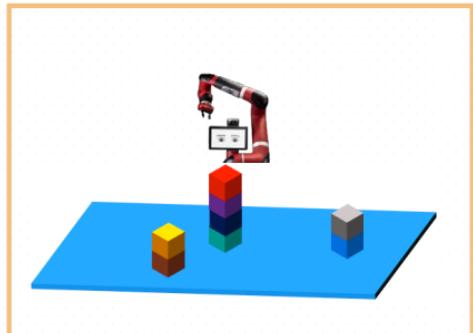
I. Learn **Multiple Tasks** in the Same Domain

Training Tasks

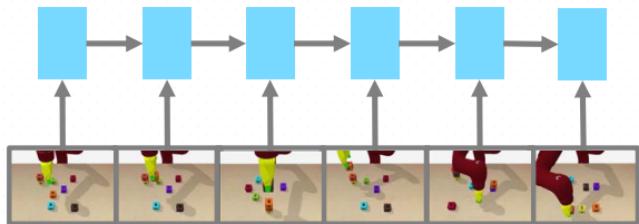


II. Generalize to New Tasks with a **Single Demo**

Test Task

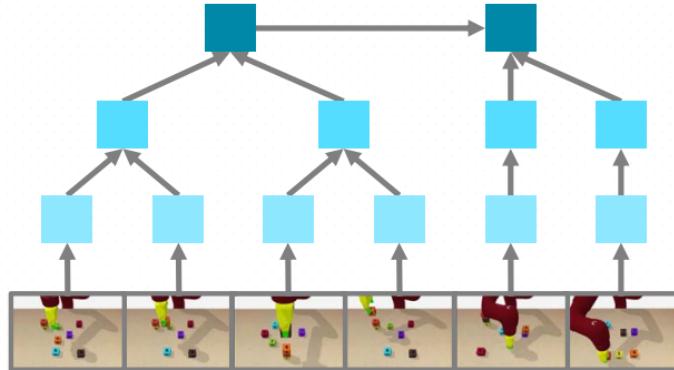


Compositional Planning



[Duan et al. 17; Finn et al. 2017;
Wang et al. 2017; Yu et al. 2018]

Models input demonstration
as a **flat sequence**



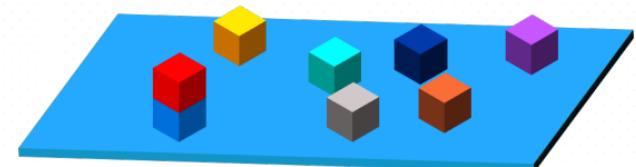
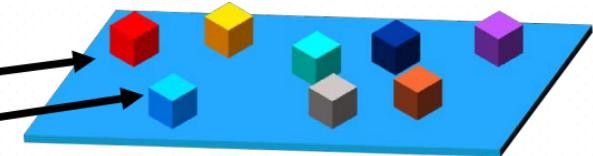
Our Method
[ICRA'18], [CVPR'19], [IROS'19]

Models input demonstration
as a **Compositional Hierarchy**

One Shot Imitation Learning from Videos

Compositional Planning: Task Programming

```
Block Stacking (...):  
while (done):  
    pick_and_place (RED, BLUE):  
        pick (RED):  
            move_to (RED)  
            Grasp (RED)  
        <end> Pop  
        place(BLUE):  
            move_to (BLUE)  
            Release (RED)  
        <end> Pop  
<end> Pop
```



Task 1
Sub-task 1
Move Red-block on top of Blue

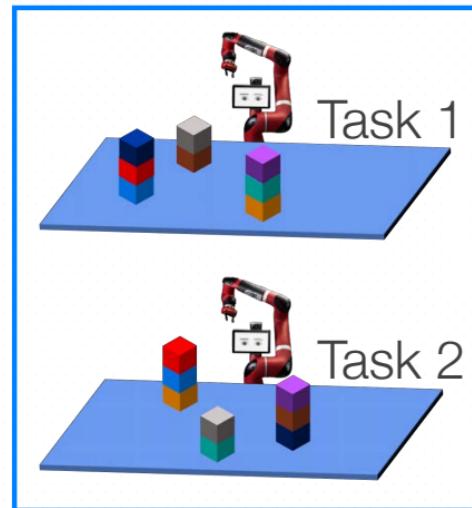
Compositional Planning: Task Programming

Block Stacking (...): Program 1

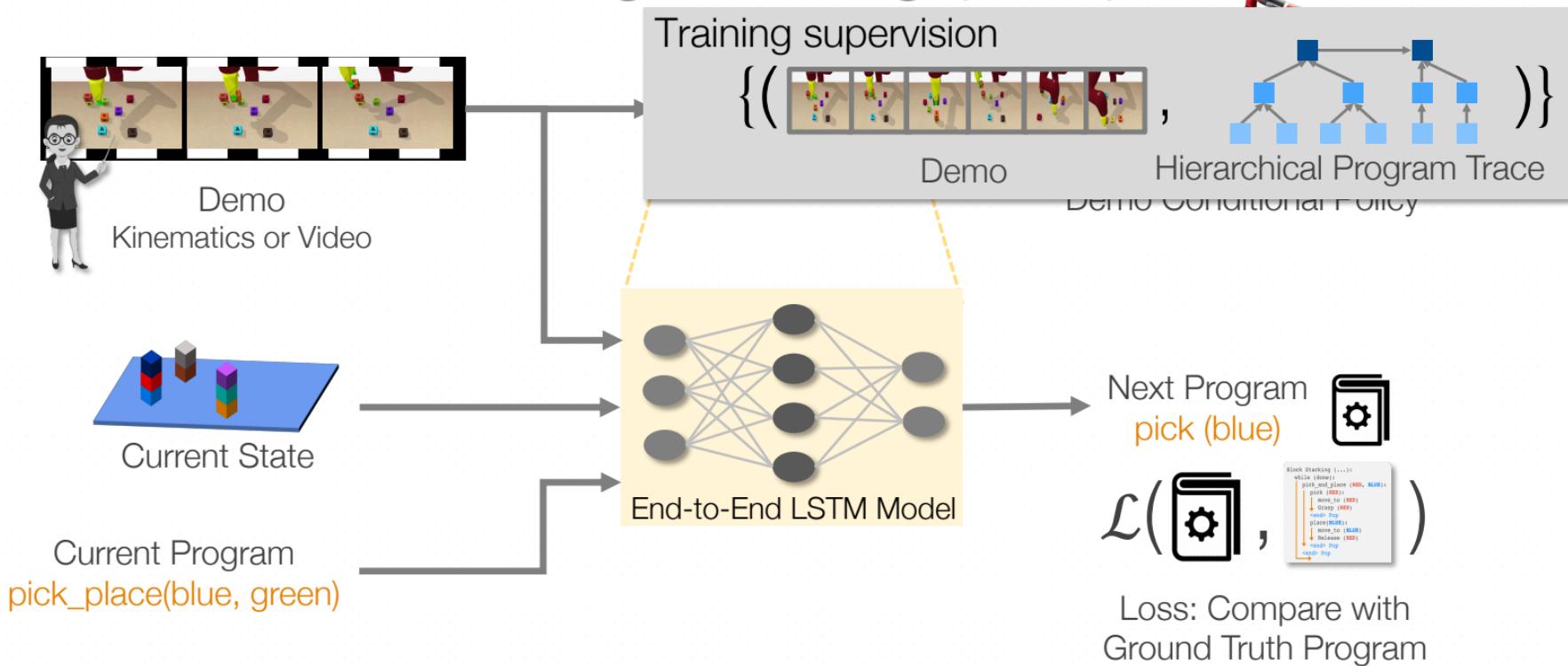
Block Stacking (...): Program 2

```
while (done):
    pick_and_place (RED, BLUE):
        pick (RED):
            move_to (RED)
            Grasp (RED)
            <end> Pop
        place(BLUE):
            move_to (BLUE)
            Release (RED)
            <end> Pop
    <end> Pop
```

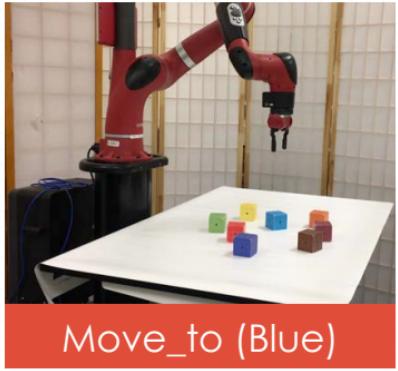
Training Task Structures



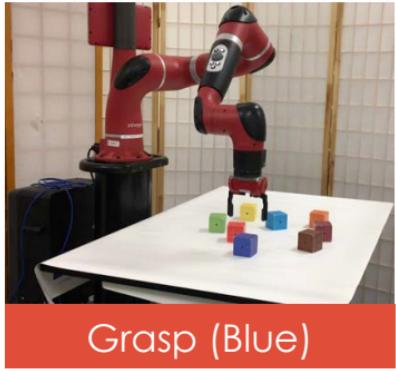
Neural Task Programming (NTP)



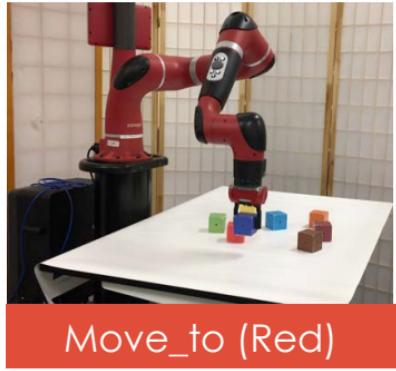
Hierarchical Policy Learning as **Program Induction**



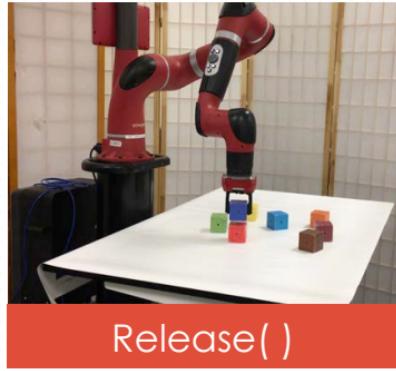
Move_to (Blue)



Grasp (Blue)



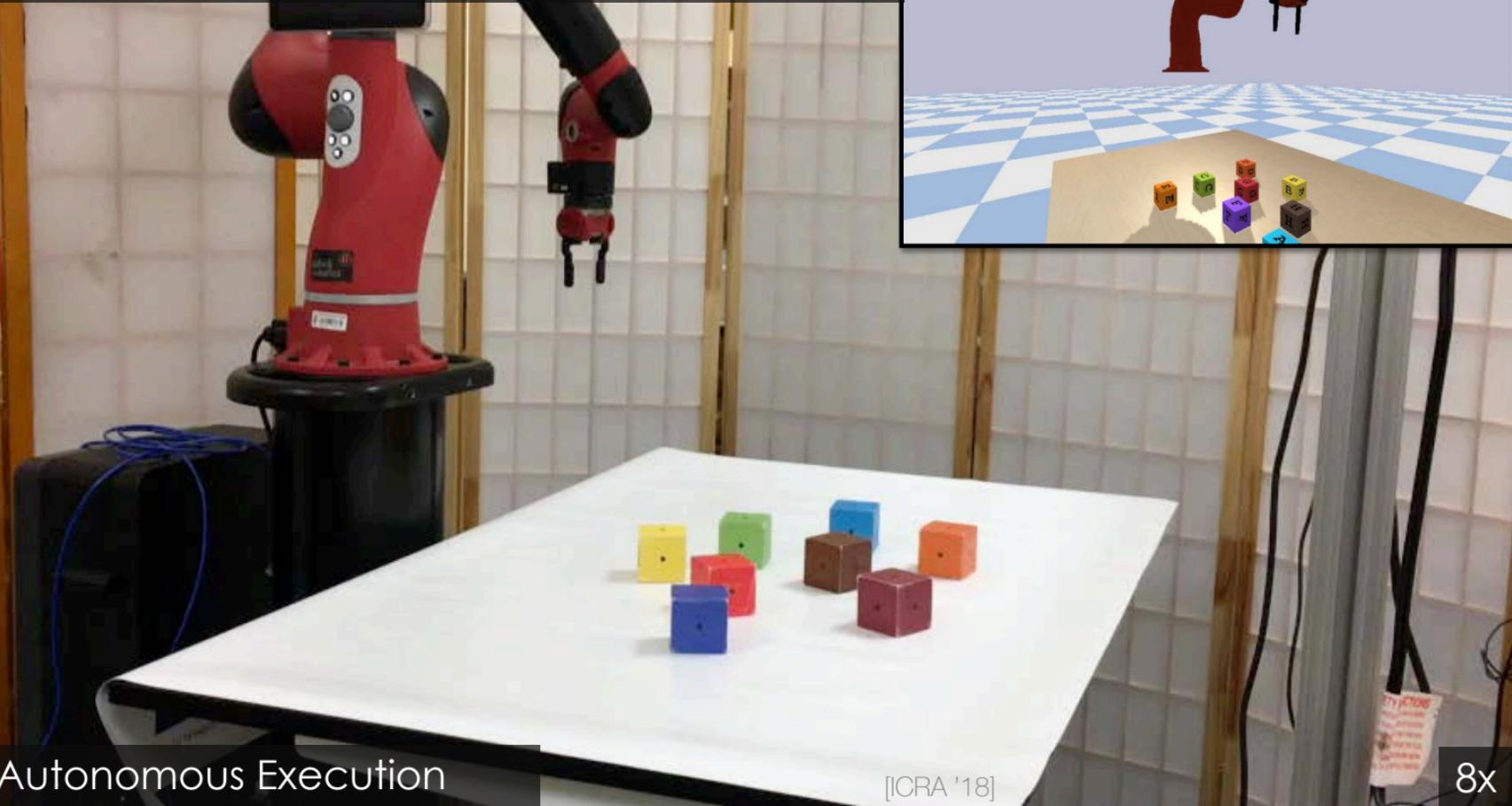
Move_to (Red)



Release()

Neural Task Programming

Demo

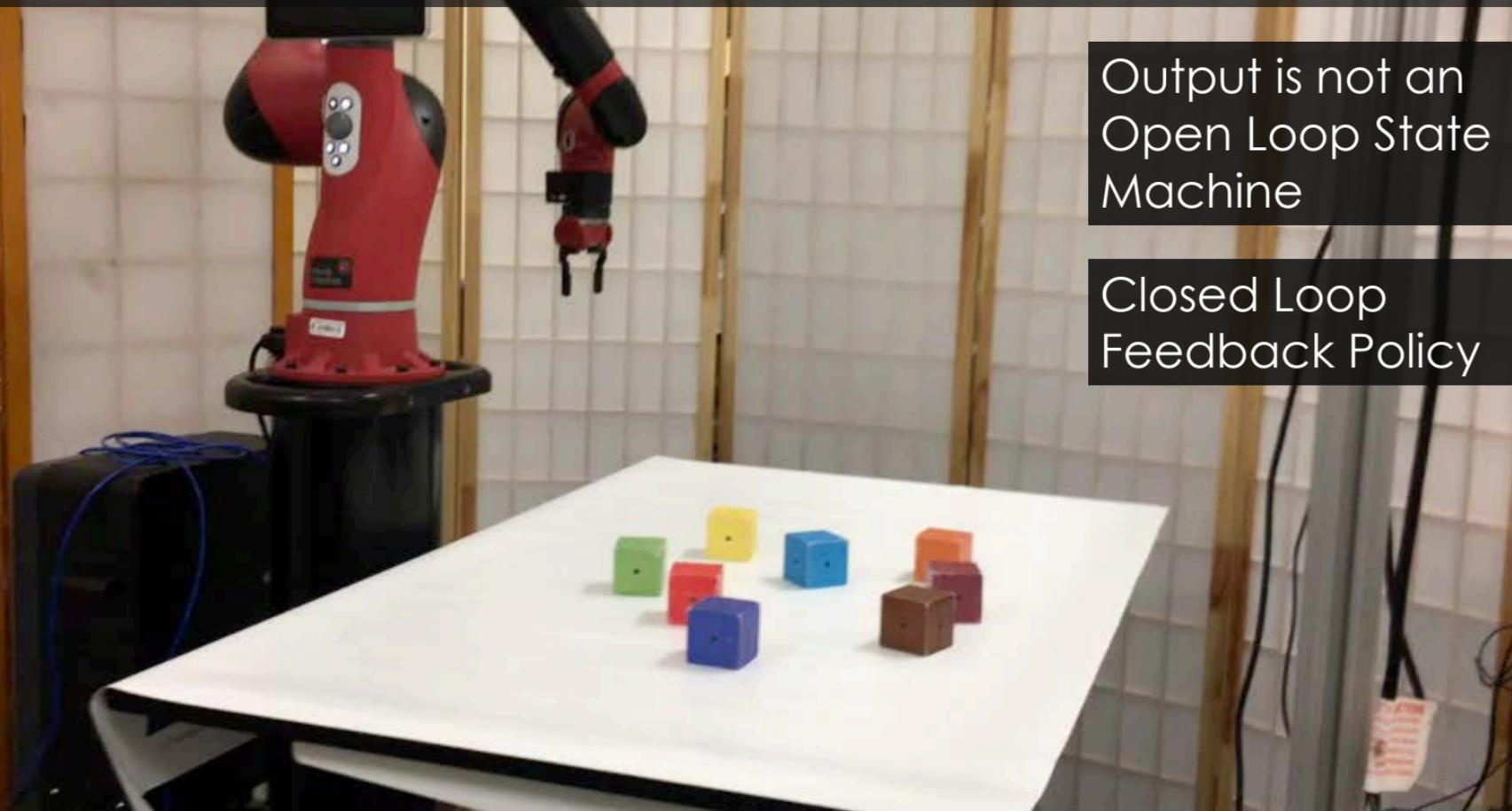


Autonomous Execution

[ICRA '18]

8x

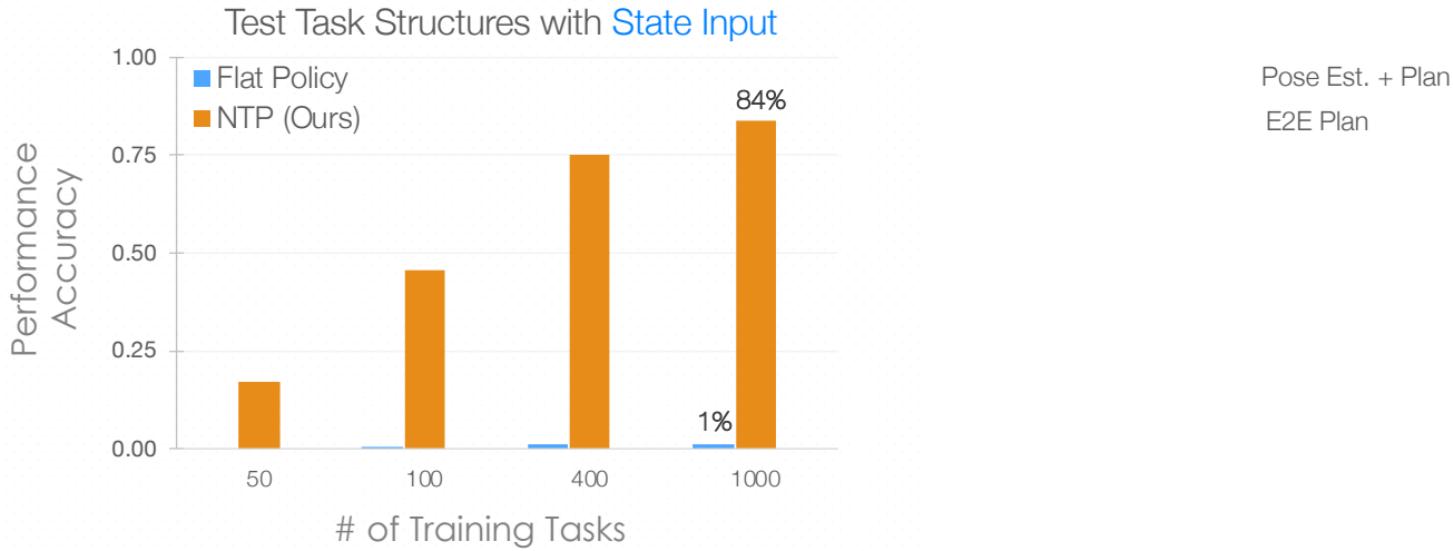
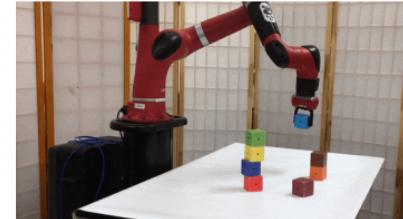
Recovery from Intermediate Failures



Output is not an
Open Loop State
Machine

Closed Loop
Feedback Policy

Neural Task Programming Results



Better Generalization than Flat Policy + Works with Vision

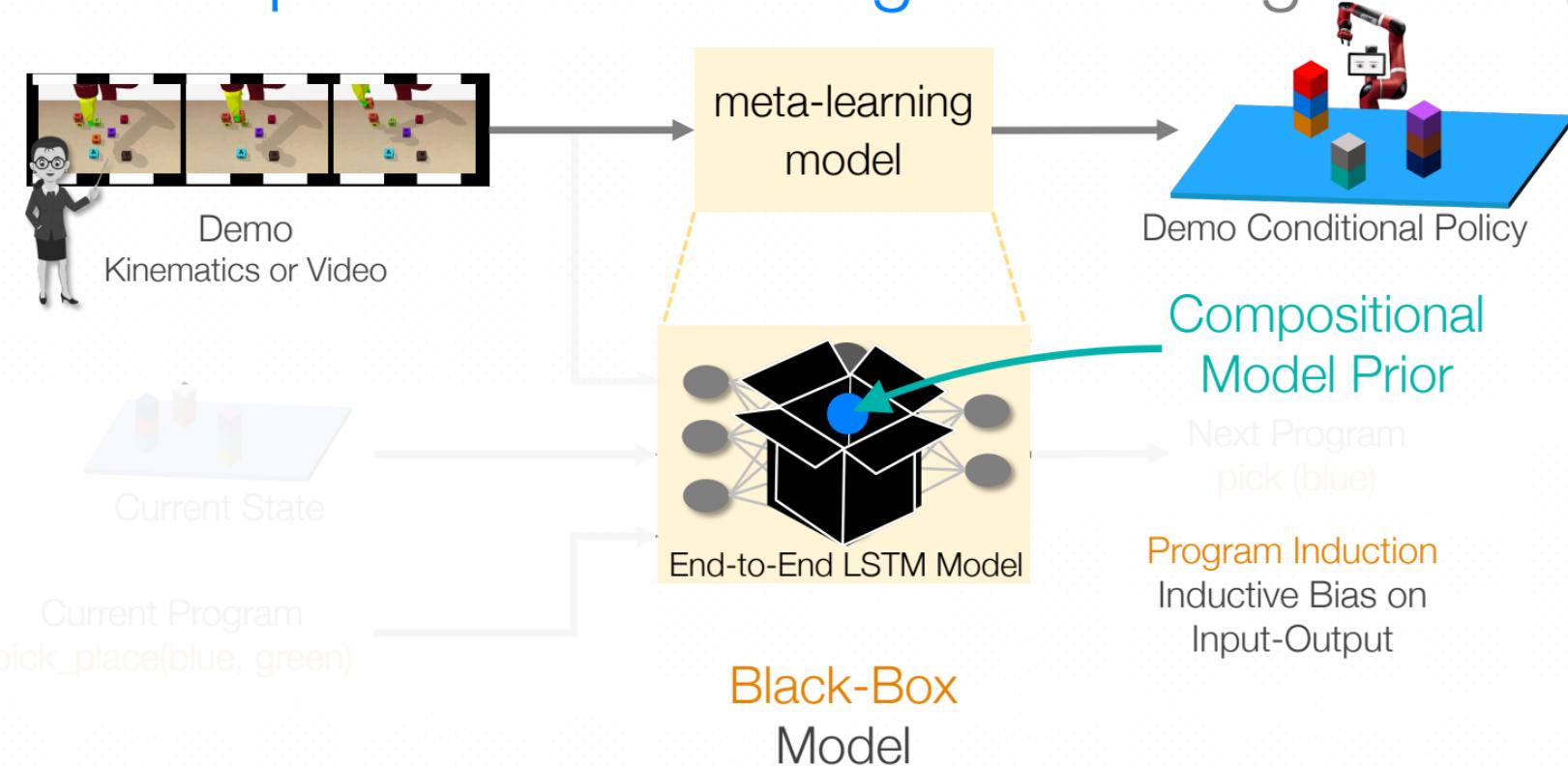
Failure Modes



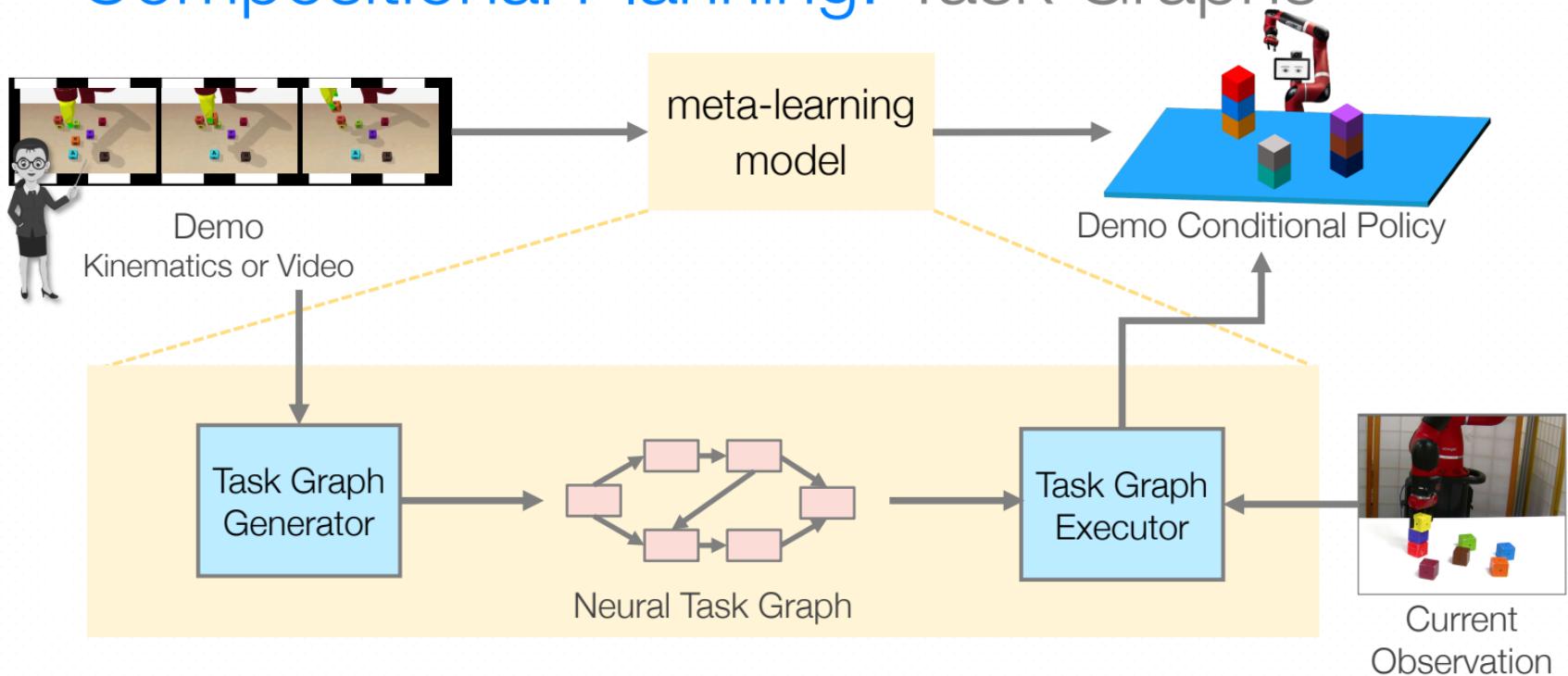
Grasping Failures

2x

Compositional Planning: Task Programming

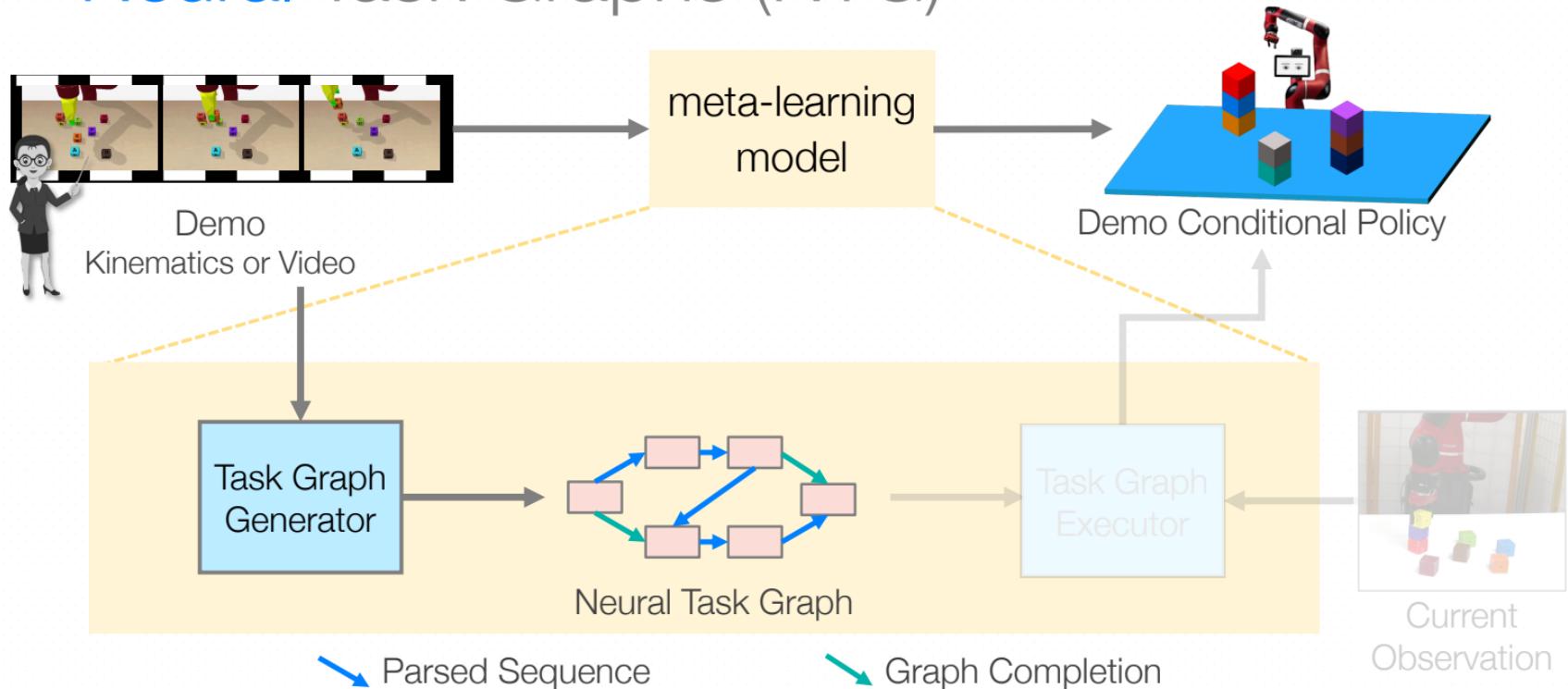


Compositional Planning: Task Graphs



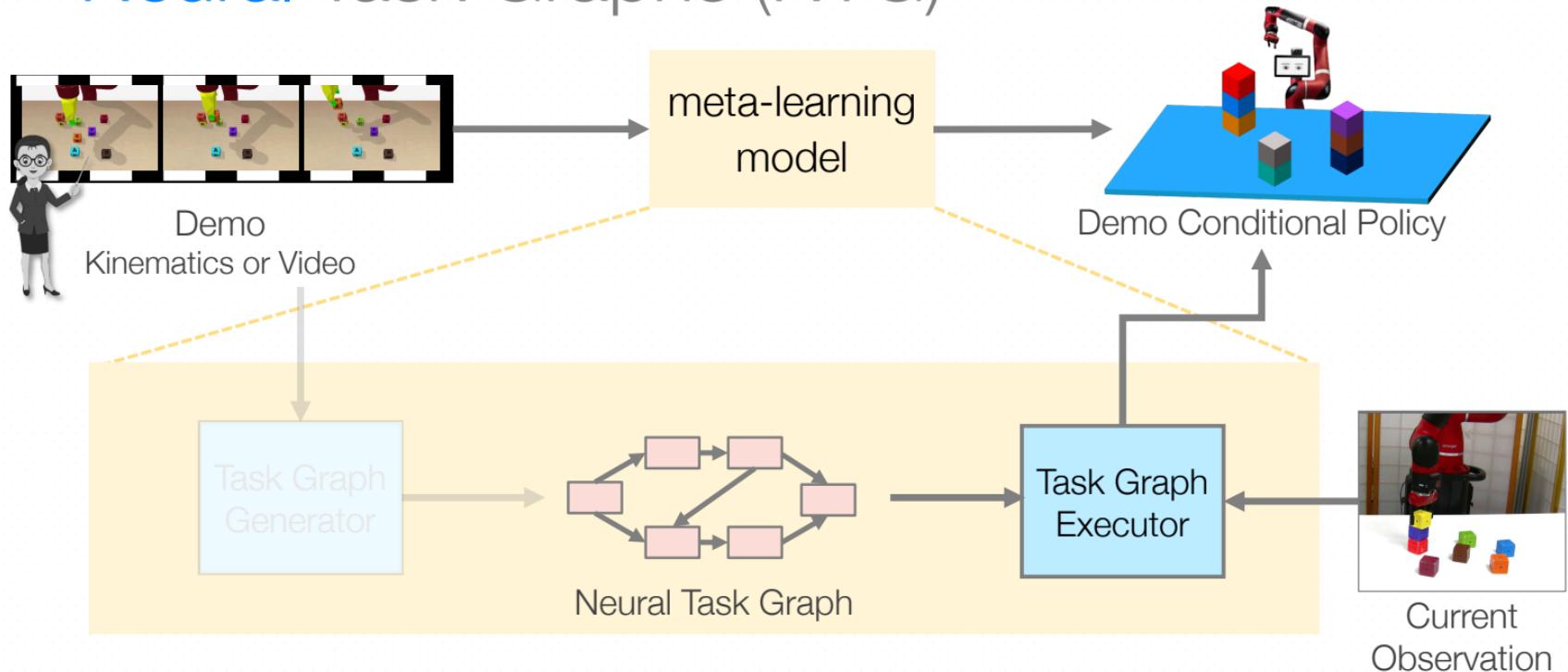
Hierarchical Policy Learning as **Graph Induction**

Neural Task Graphs (NTG)



Hierarchical Policy Learning as **Graph Induction**

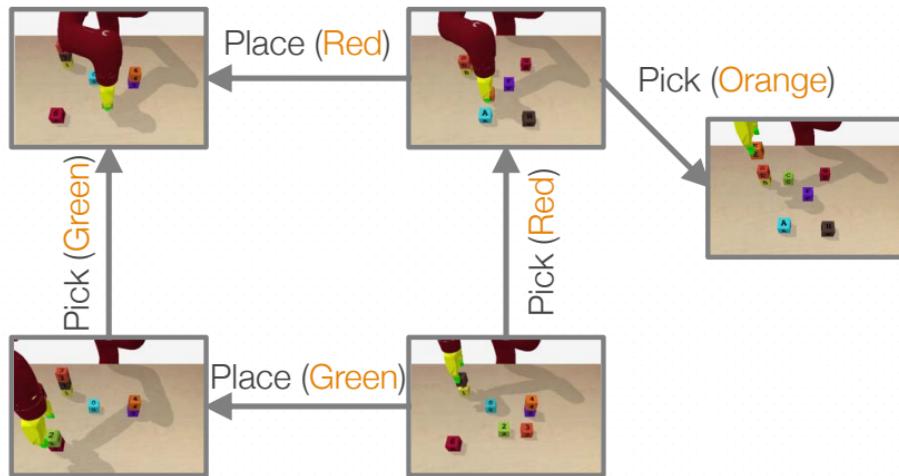
Neural Task Graphs (NTG)



Hierarchical Policy Learning as **Graph Induction**

Neural Task Graphs (NTG): Representation

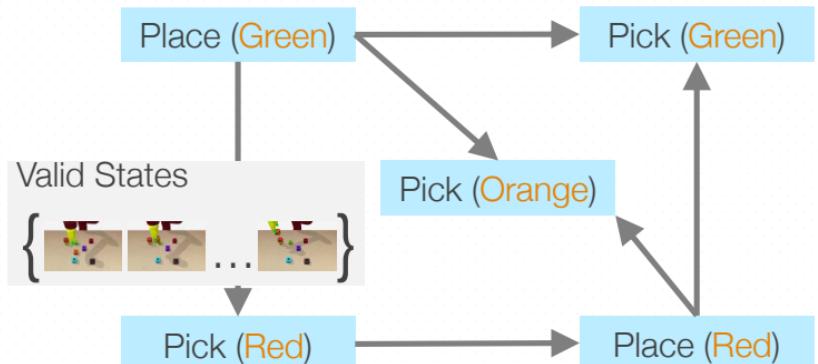
Task Graph



Nodes: States Combinatorial

Edges: Action

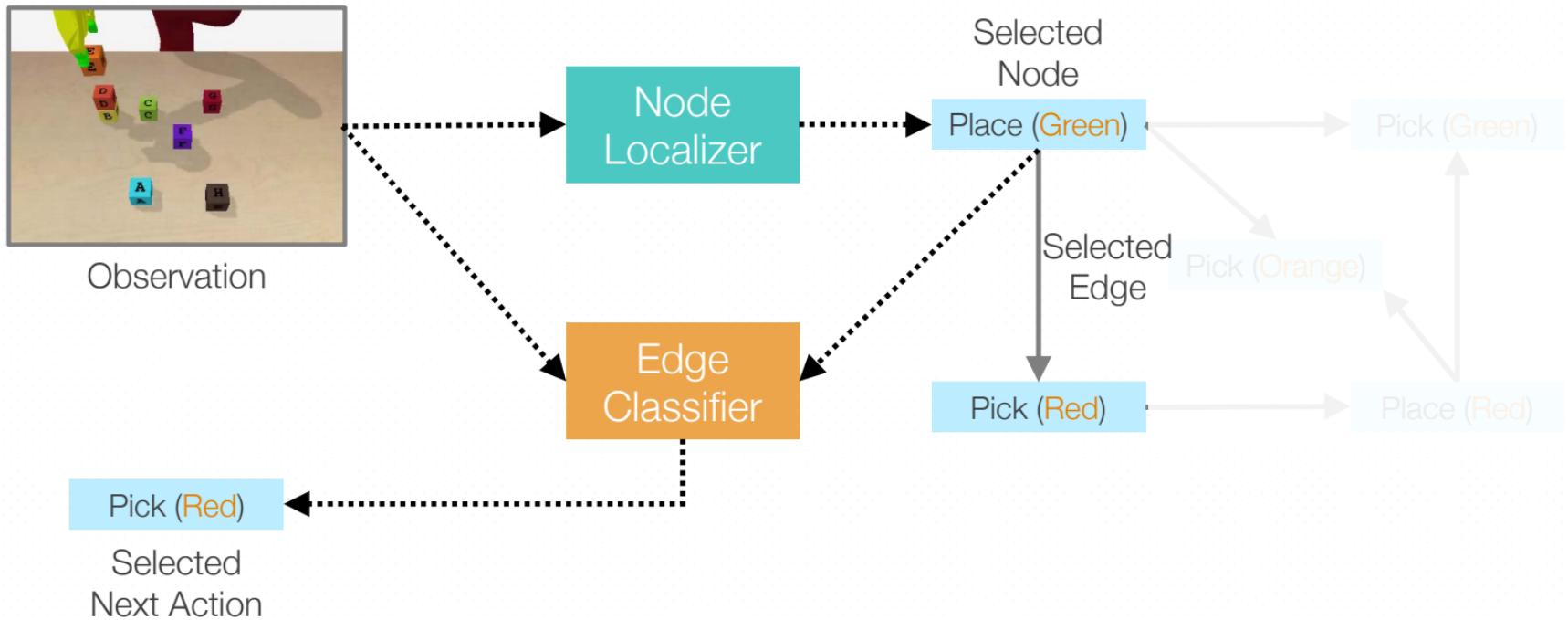
Conjugate Task Graph



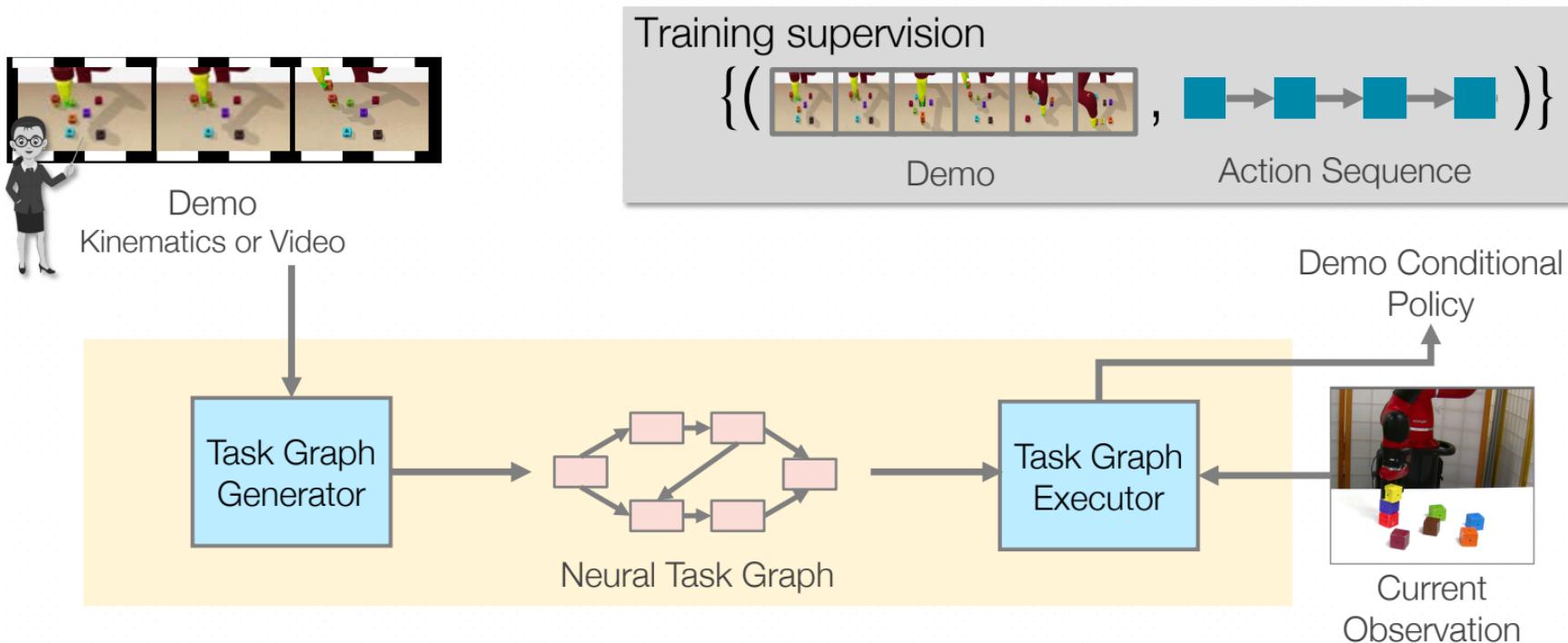
Nodes: Actions Finite

Edges: States (Preconditions)

Neural Task Graphs (NTG): Execution

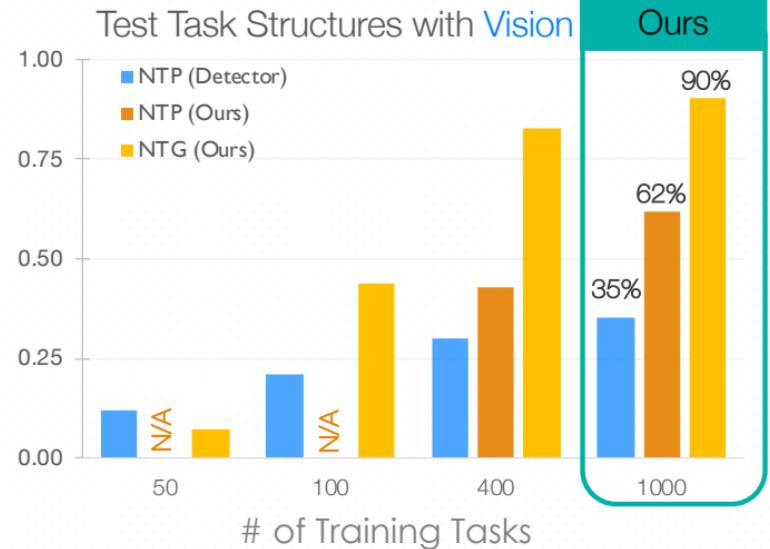
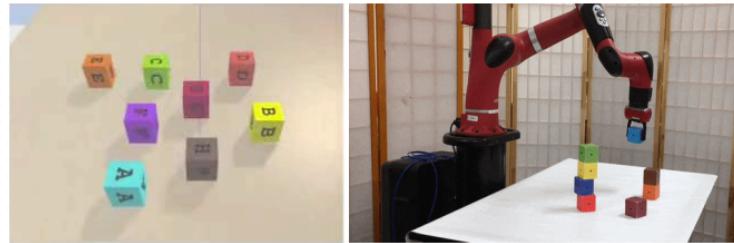
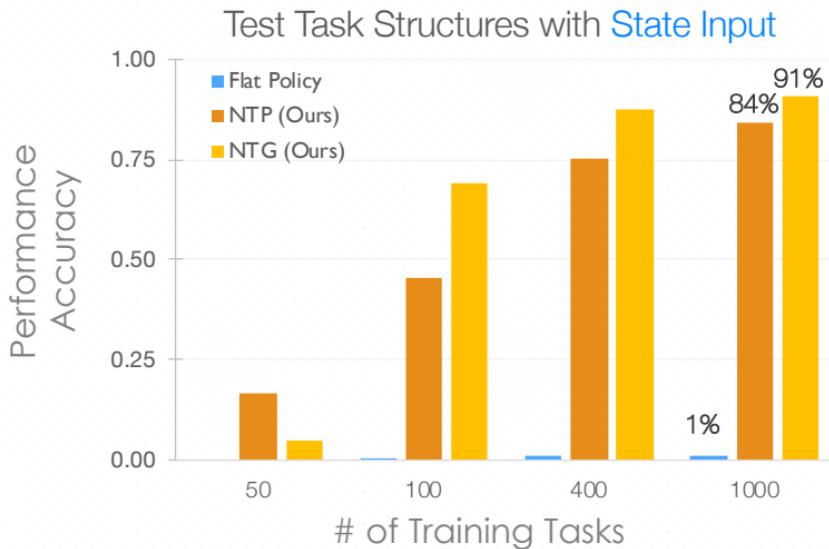


Neural Task Graphs (NTG)



Hierarchical Policy Learning as **Graph Induction**

Neural Task Graph Results



Weaker Supervision and Better Generalization

Compositional Planning: NTP and NTG



Object Sorting
(NTP)

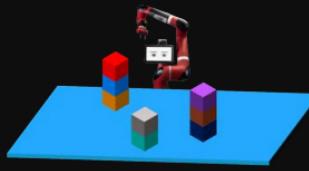
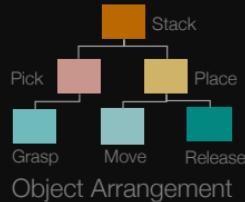


Table Clean Up
(NTP)



Sequential Search and Prediction
AI2 Thor with NTG

Task Structure Learning

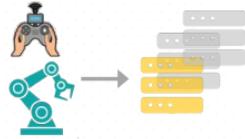


Data for
Robotics



Compositional priors with modular structure enable
generalizable learning in hierarchical domains

Generalizable Autonomy in Robot Manipulation



CoRL 2018, IROS 2019

Visuo-Motor Skills



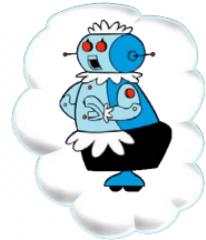
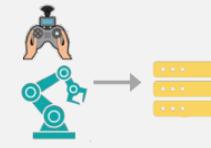
Compositional Planning



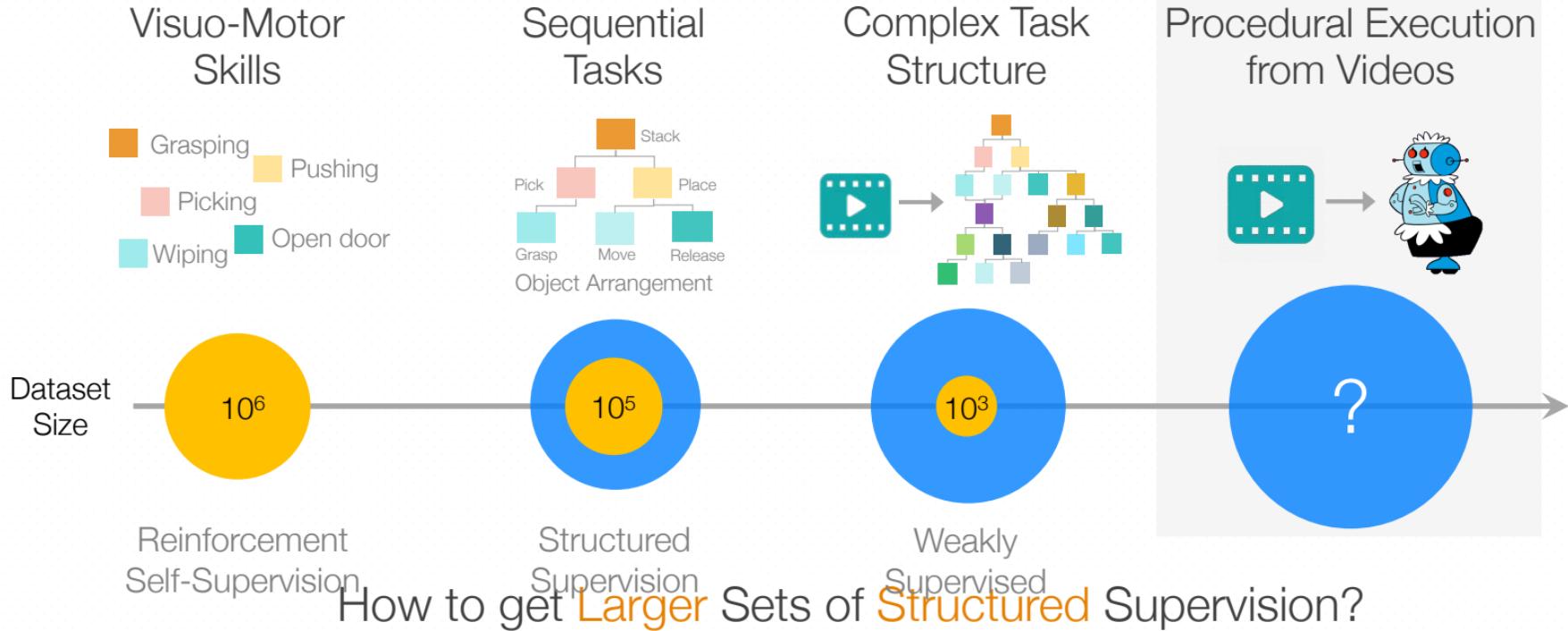
Task Structure



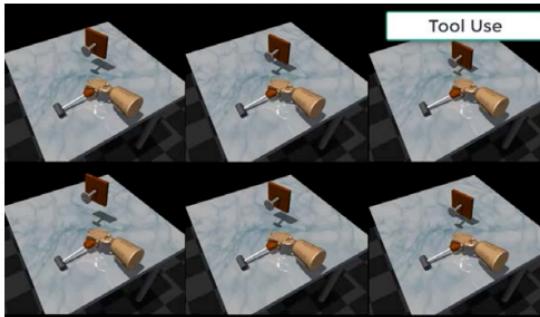
Data for Robotics



Data for Robotics



Data for Robotics: Imitation + RL



Rajeswaran et al. (2018)

25 demonstrations

~ 10 Minutes



Finn et al. (2017)

30 demonstrations

~ 10 Minutes

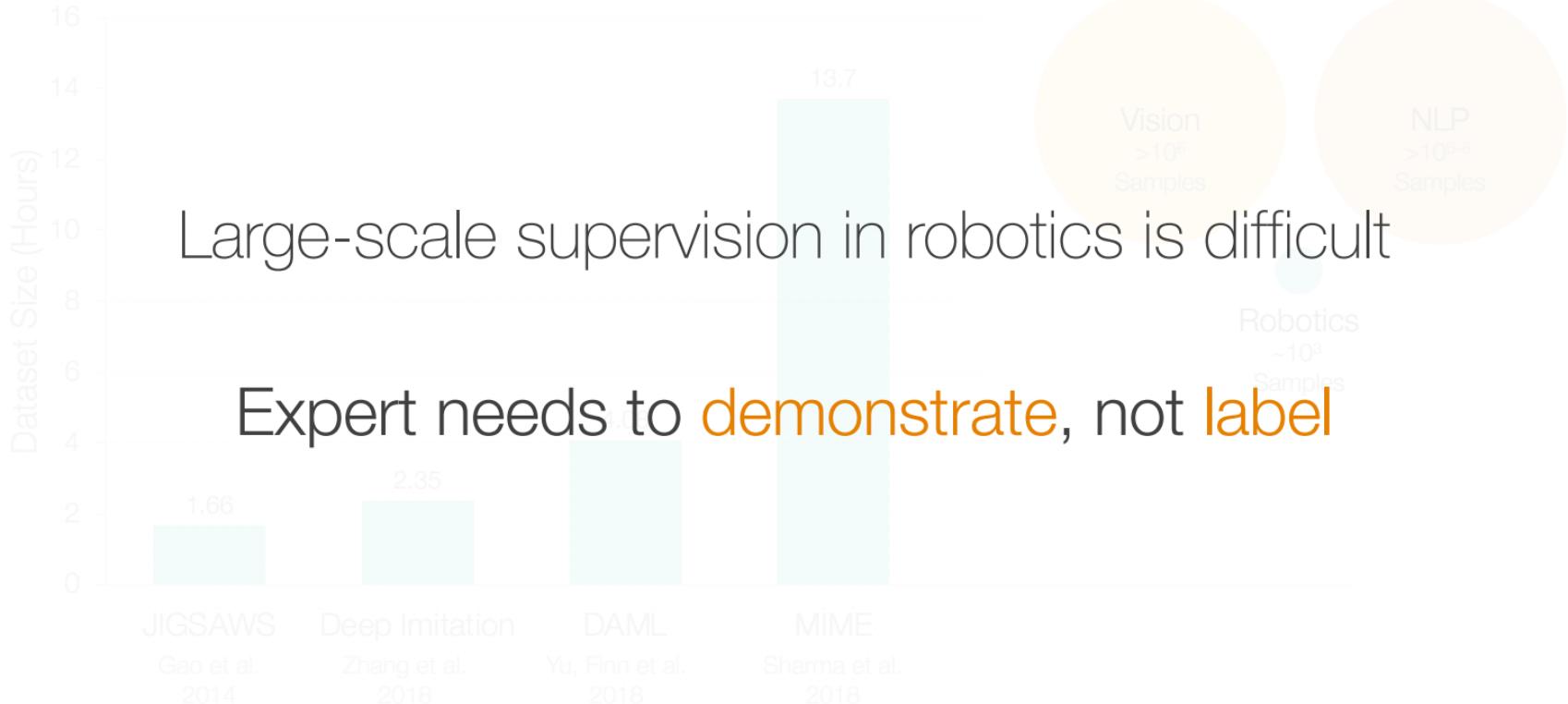


Vecerik et al. (2017)

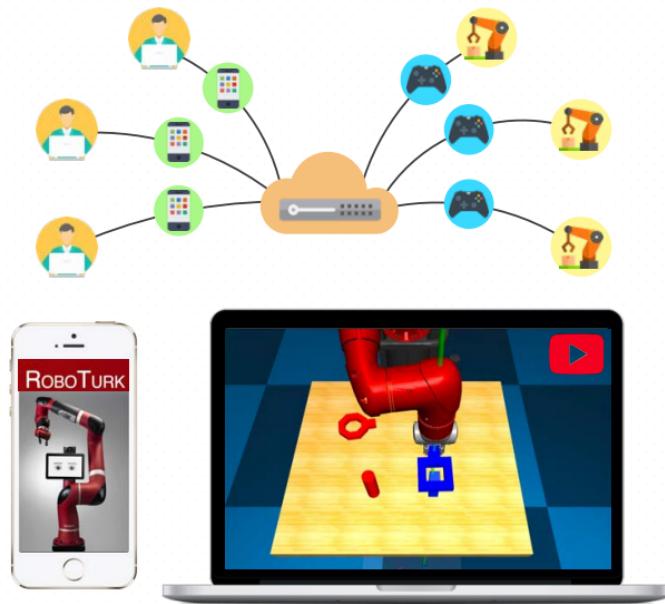
100 demonstrations

~ 30 Minutes

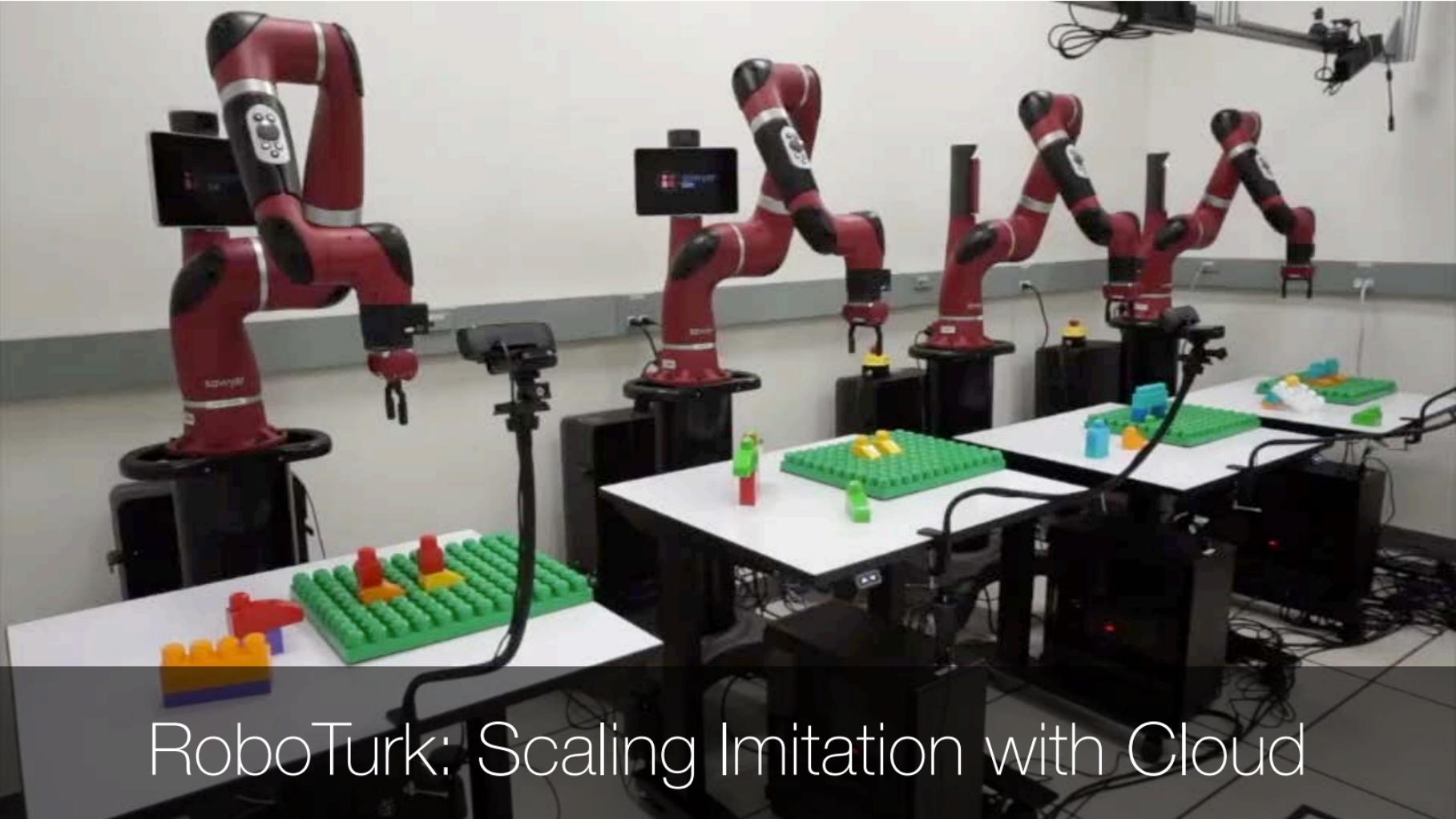
Data for Robotics



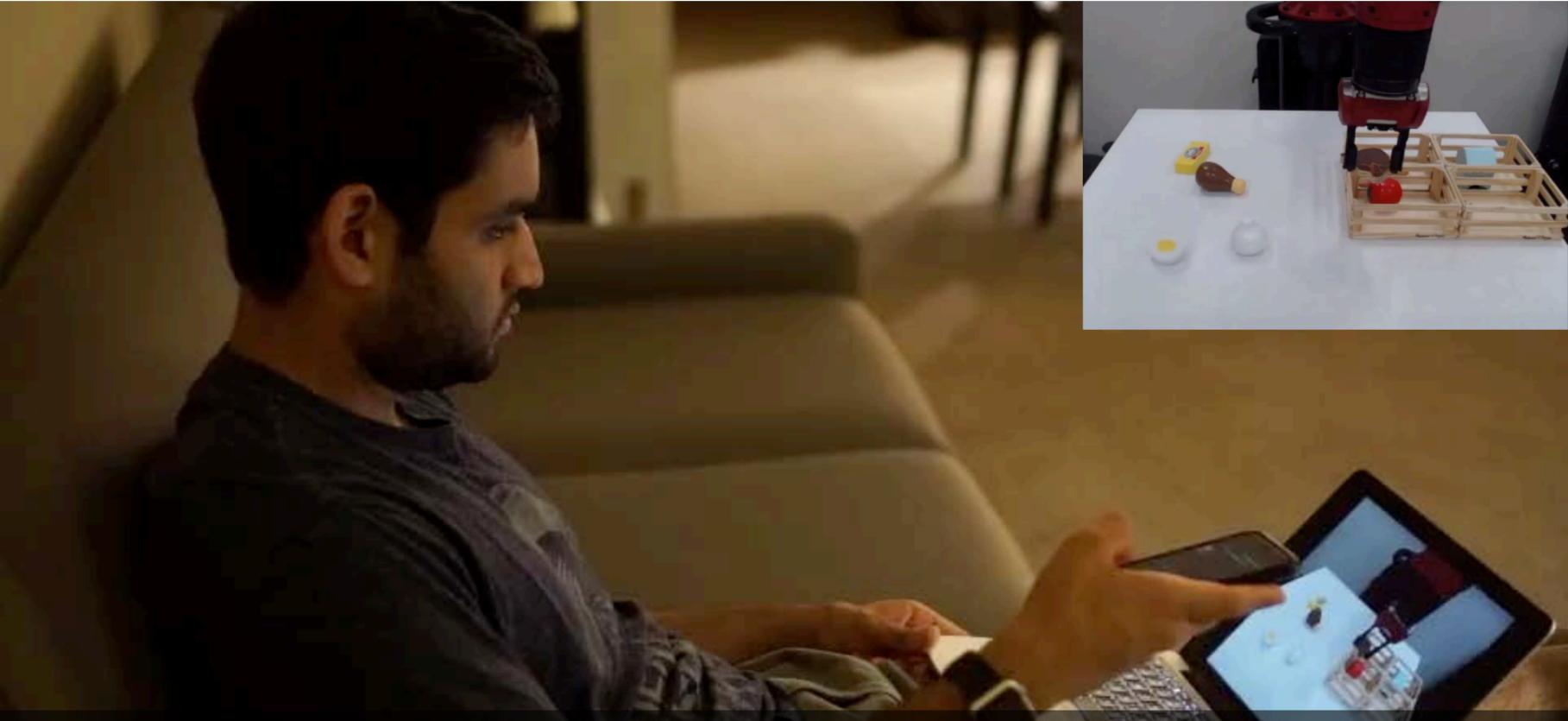
Data for Robotics: RoboTurk



- + Scales easily with commodity hardware
- + Natural 6-DoF Free Space Control



RoboTurk: Scaling Imitation with Cloud



RoboTurk: Imitation for everyone, everywhere

RoboTurk Pilot Datasets

Simulated Data

137.5 hours of demonstrations

22 hours of total platform usage

3 dexterous manipulation tasks

3224 total attempted demos

15 novice, remote users

Real Robot Data

111 hours of robot demos

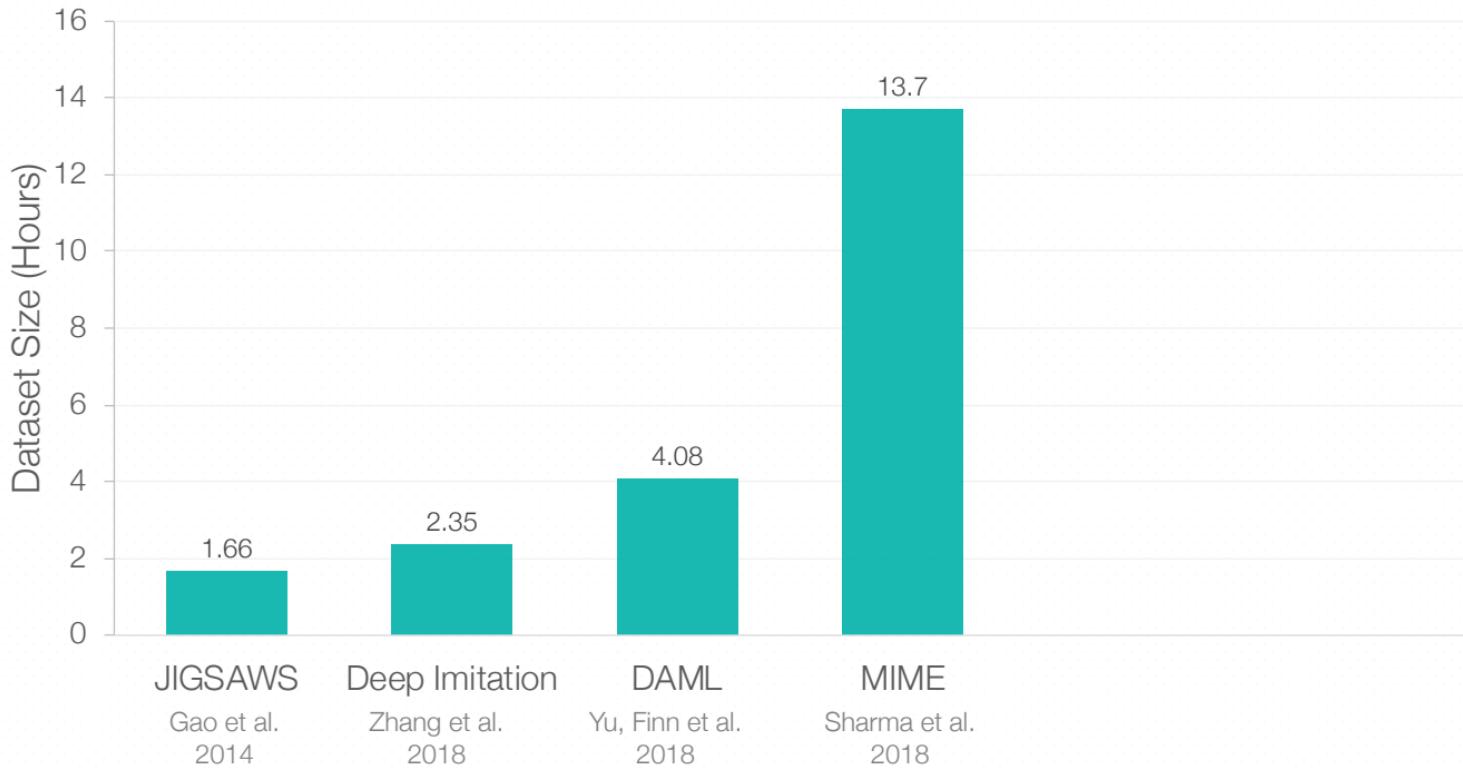
1 week of data collection

3 dexterous manipulation tasks

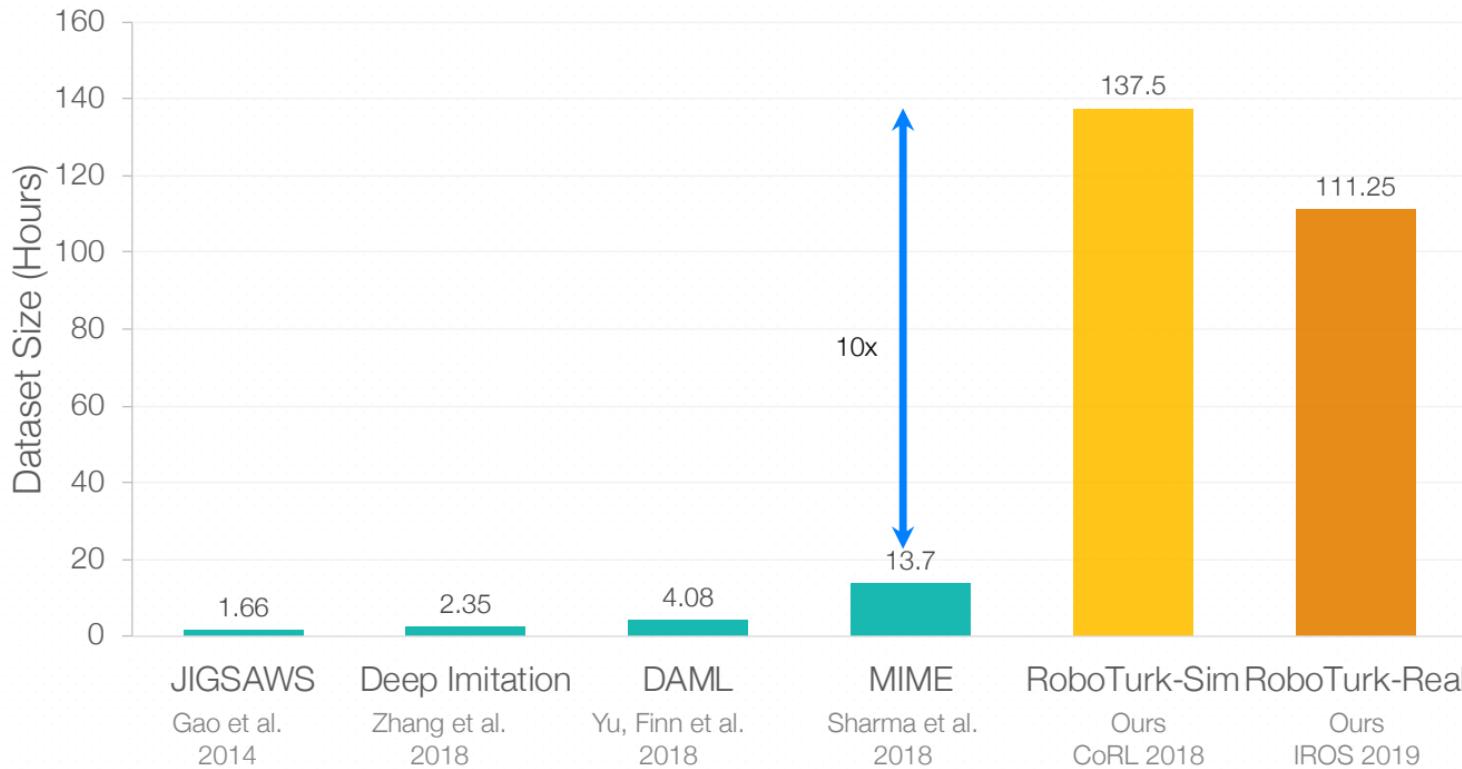
2144 total demonstrations

54 non-expert users

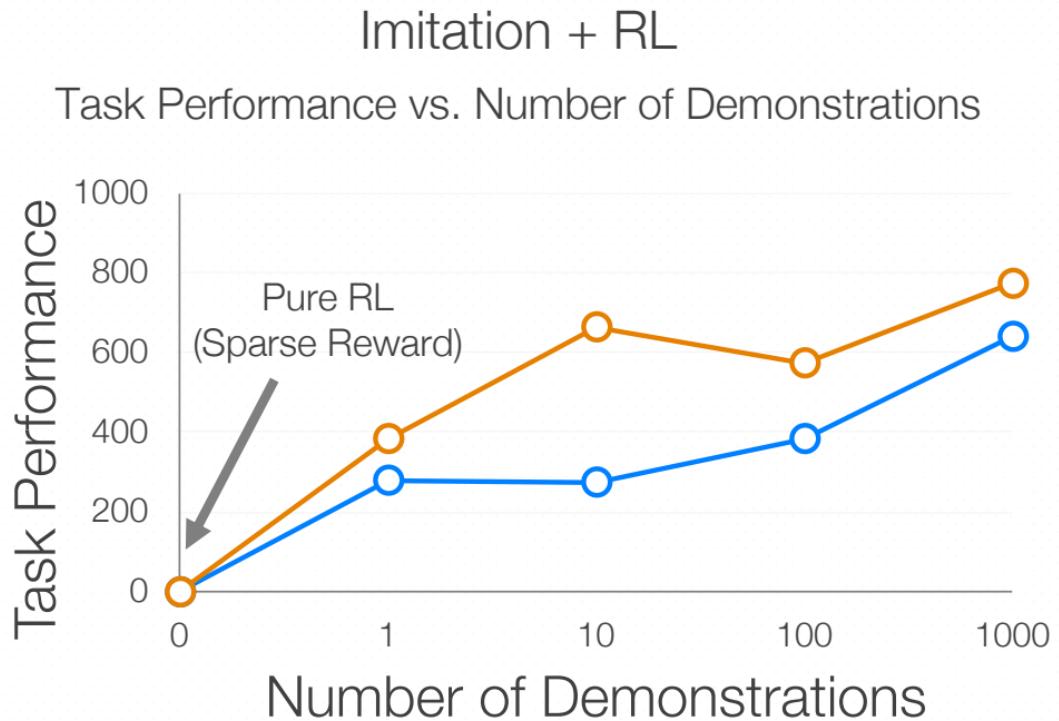
Data for Robotics: RoboTurk



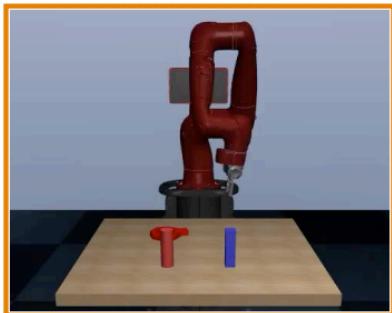
Data for Robotics: RoboTurk



Data for Robotics: RoboTurk



Trained Policy Rollout

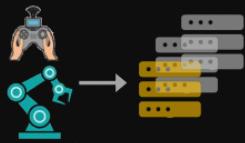


Nut Assembly



Bin Picking

Data for Robotics: RoboTurk

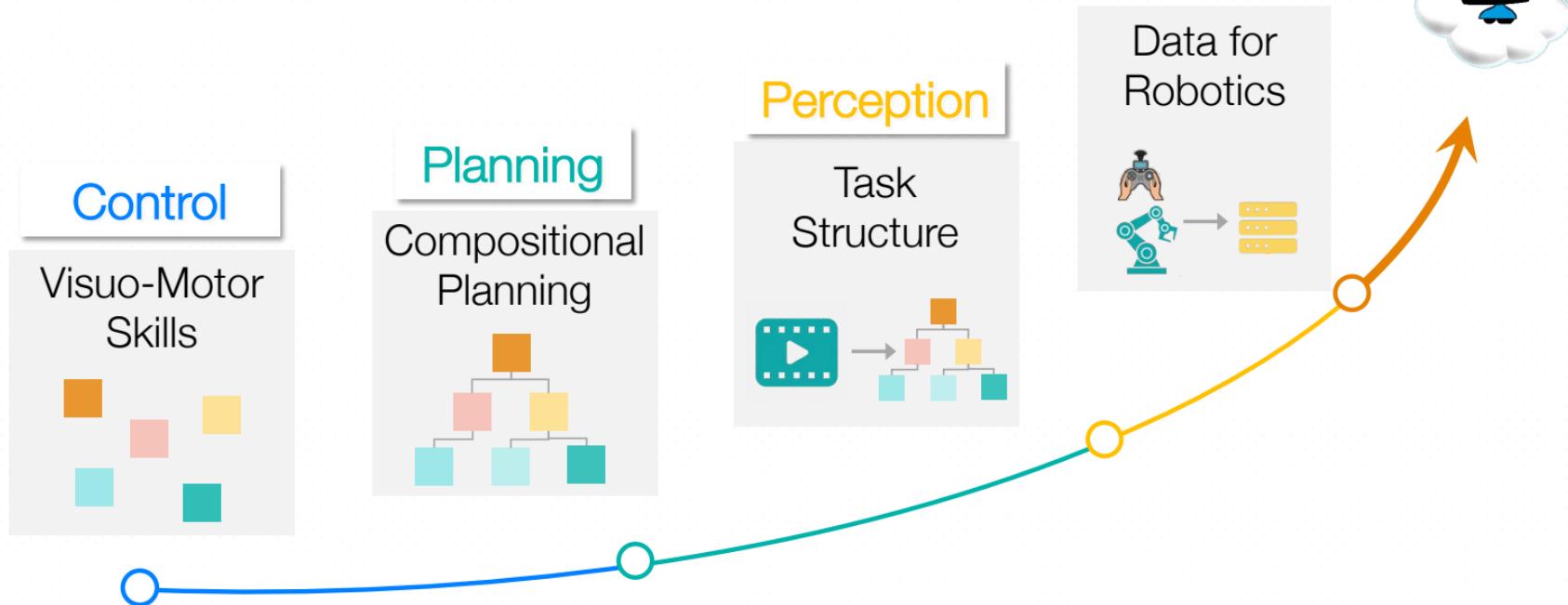


CoRL 2018, IROS 2019

Data for
Robotics

Structured supervision for Robotics through scalable
crowdsourcing can empower robot learning in complex tasks.

Generalizable Autonomy in Robot Manipulation



Opportunity: Personal Robotics



Instructional Youtube Video
How to make Meatball Pasta?



Where / How should Rosie start?
What is the recipe?
How to execute the plan?
How to plan?

Reasoning for Physical Interaction

Understanding Purpose



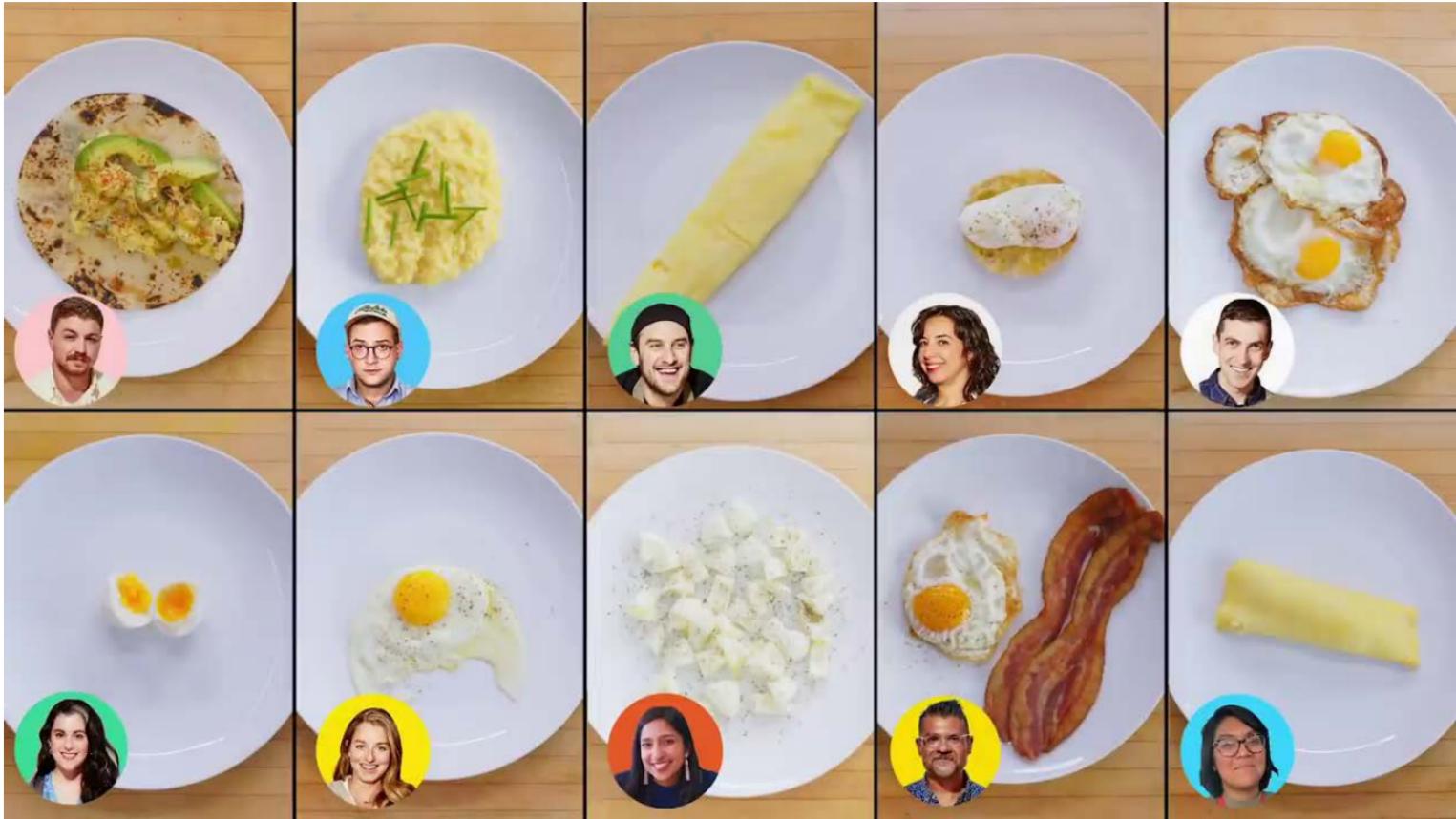
Ideal Tool
During Training



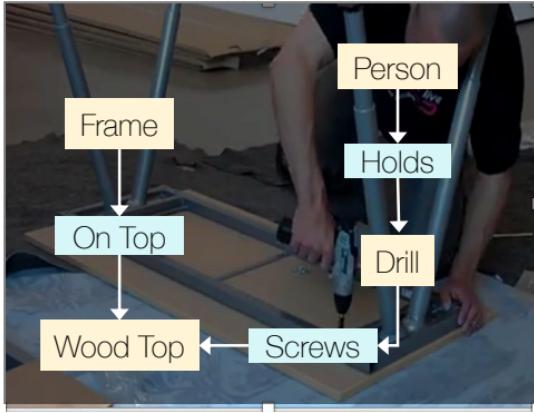
Task-Based Tool Adaptation
During Execution



Grounding: So many ways to “make eggs”



Generalizable Autonomy in Robot Manipulation



Higher-Order **Semantics**



What makes an
object a **hammer**?



State Change: Breaking Eggs

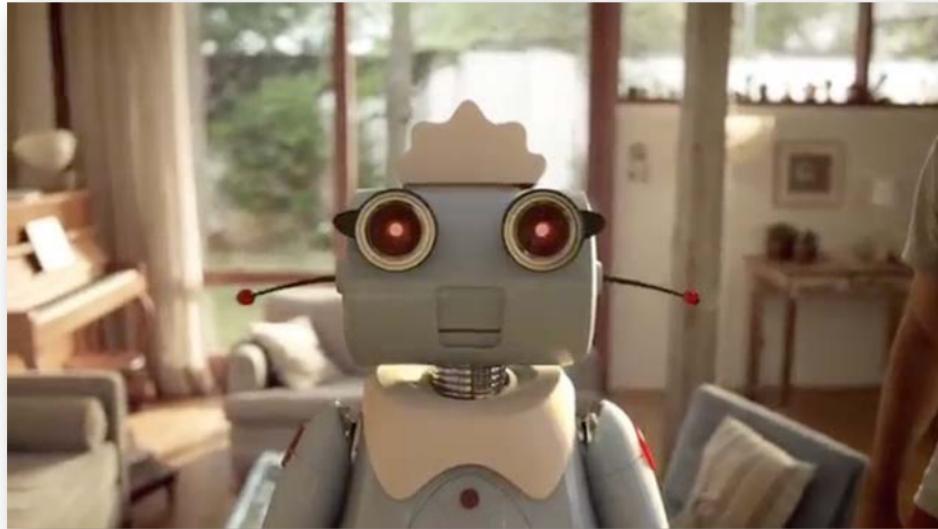
- Perception for Physical Interaction
- Reasoning through Learned Dynamics

- Transfer Learning with Formal Guarantees
- Continual Skill Adaptation & Accumulation

Generalizable Autonomy in Robot Manipulation



Generalizable Autonomy in Robot Manipulation



garg@cs.toronto.edu

@Animesh_Garg

Animesh Garg



Neural Rendering

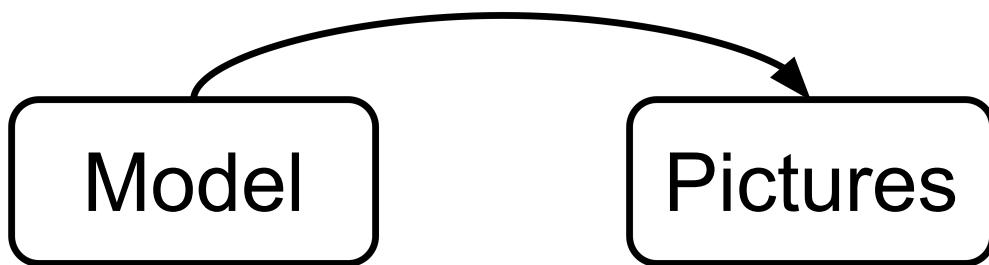
Chuan Li

Lambda Labs

Collaborators: Thu Nguyen-Phuoc, Bing Xu, Yongliang Yang, Stephen Balaban, Lucas Theis, Christian Richardt, Junfei Zhang, Rui Wang, Kun Xu, Rui Tang

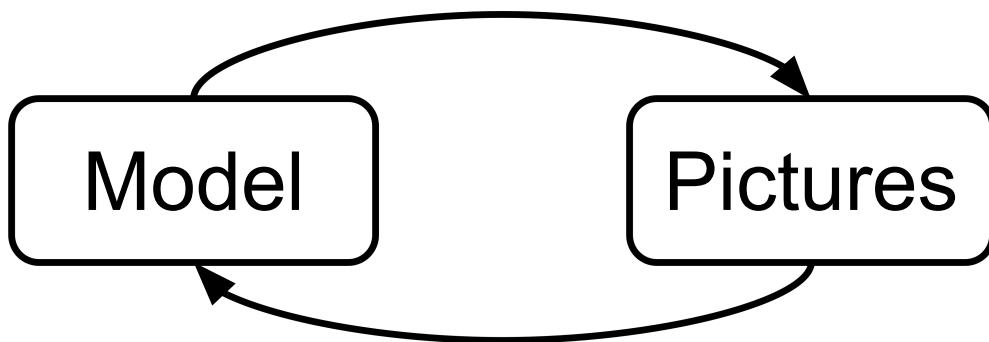


Forward (Computer Graphics)

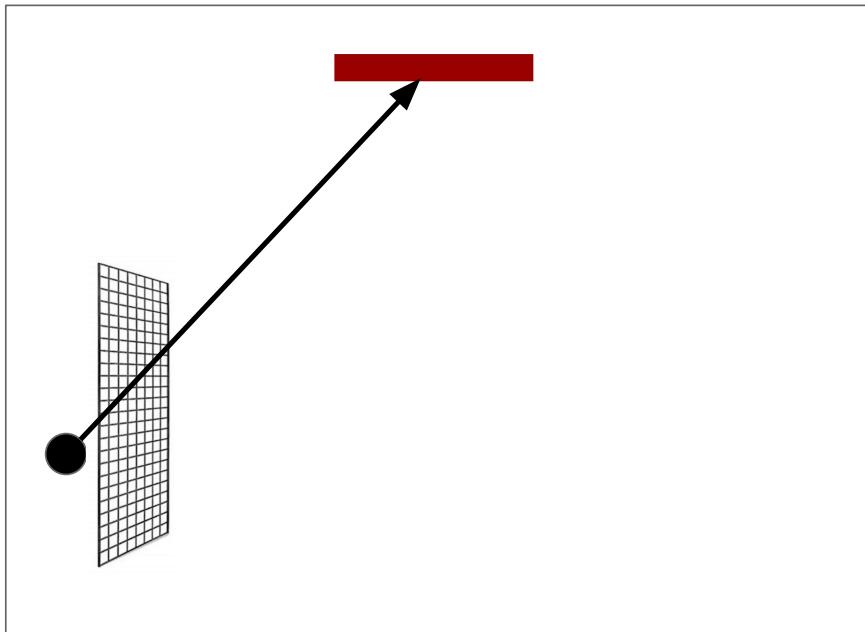


Lambda

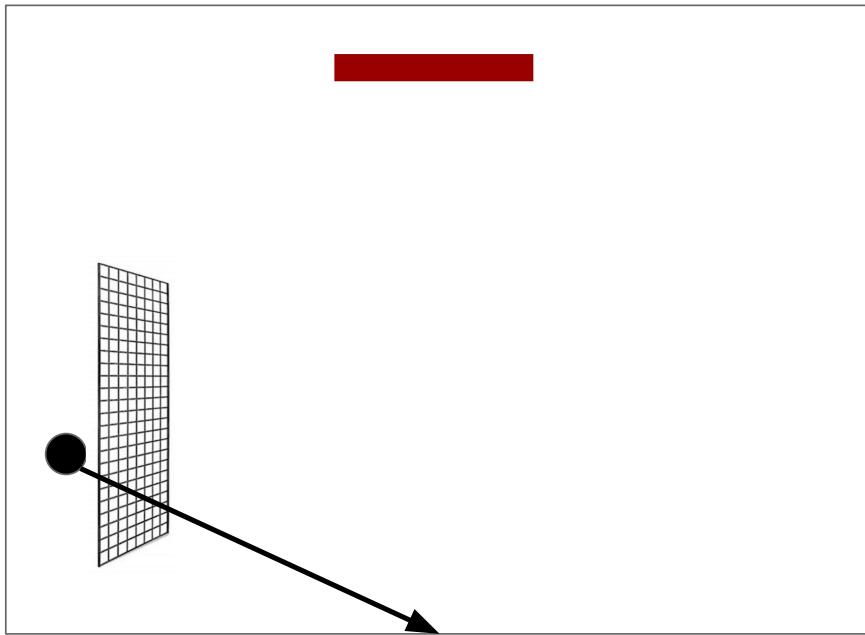
Forward (Computer Graphics)



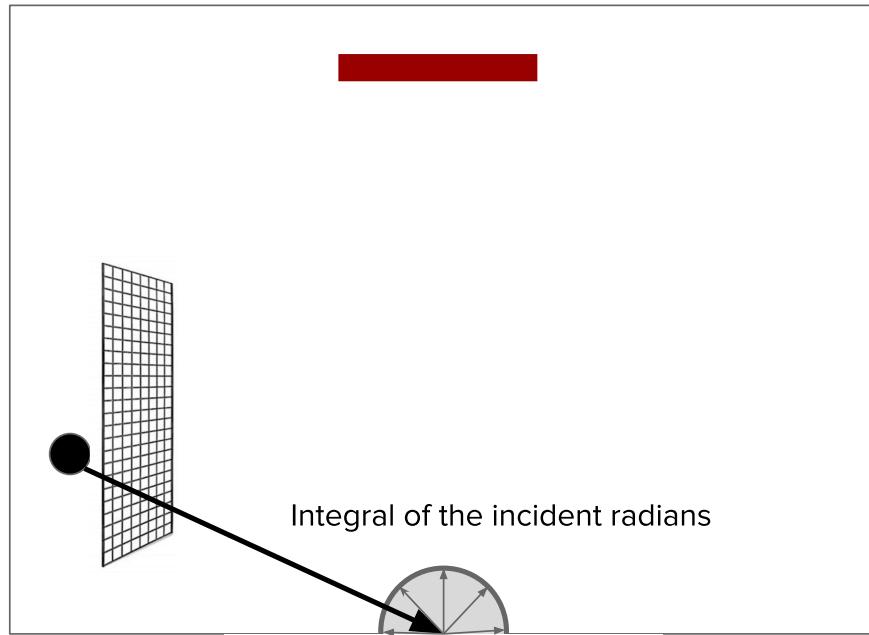
Inverse (Computer Vision)



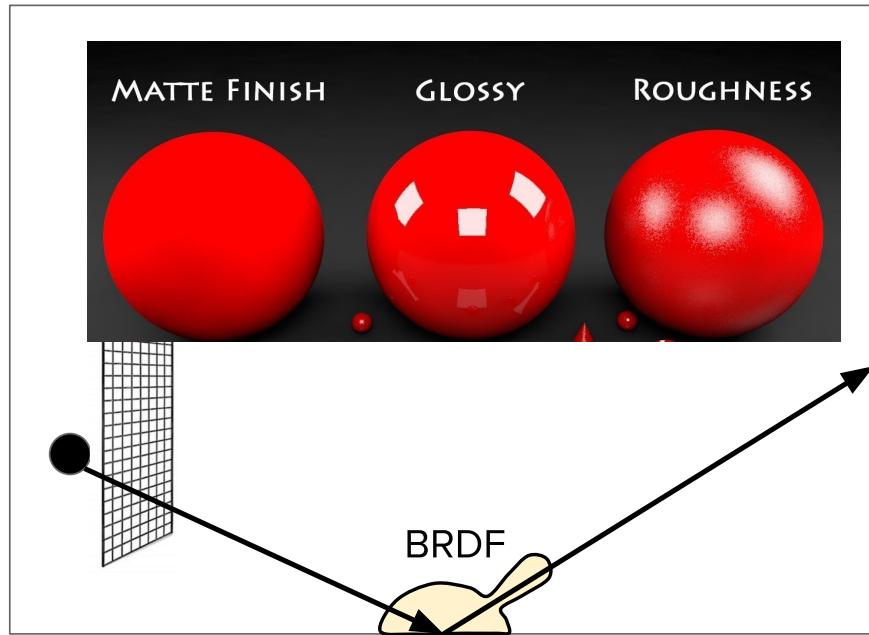
Lambda



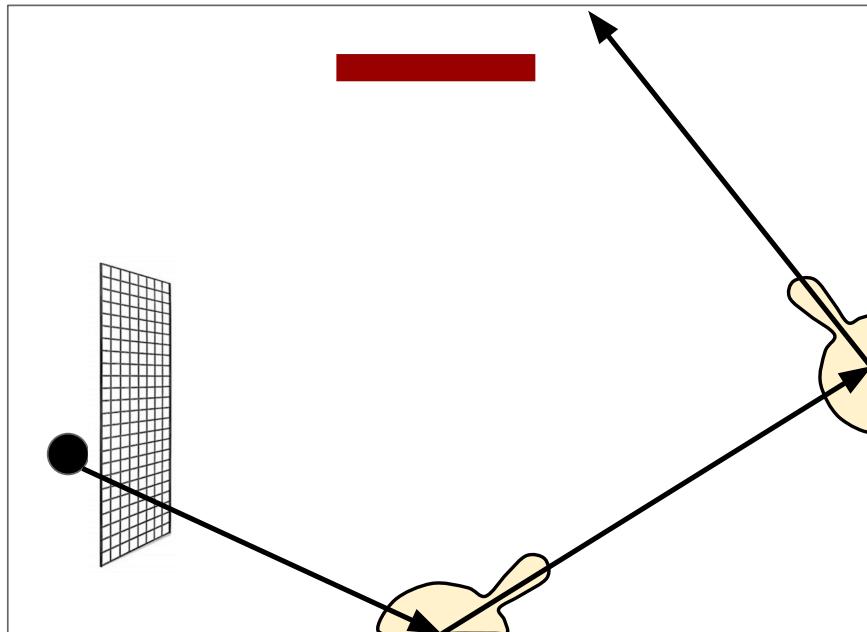
λ Lambda



λ Lambda



λ Lambda

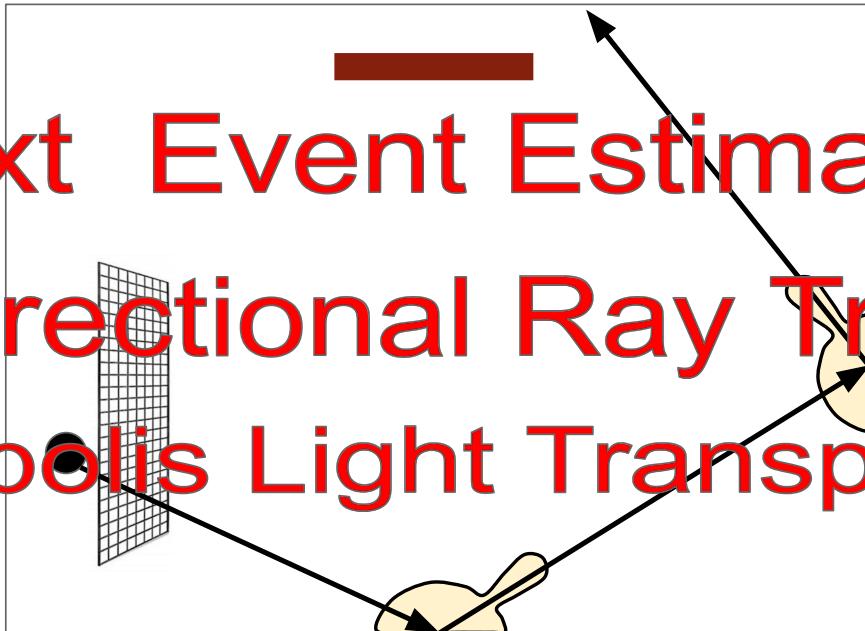


λ Lambda

Next Event Estimation

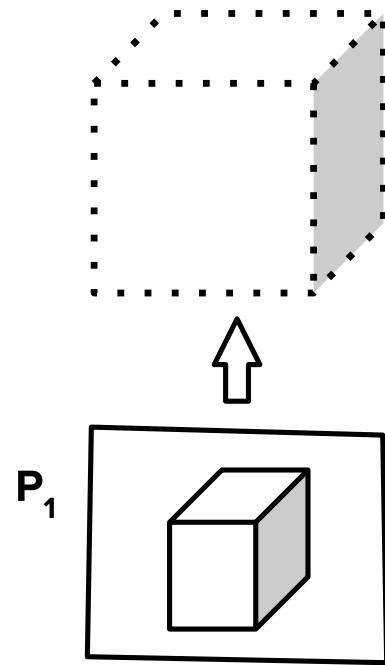
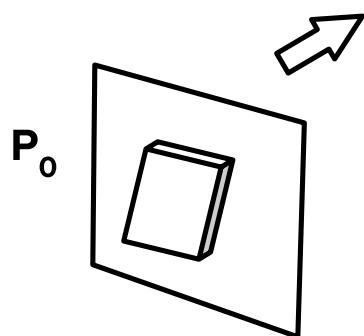
Bidirectional Ray Tracing

Metropolis Light Transportation



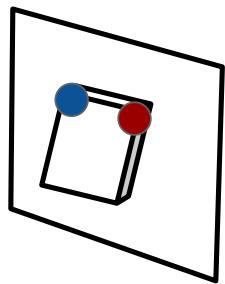


32K SPP Ray Tracing (90 mins 12 CPU Cores)
The Tungsten Renderer

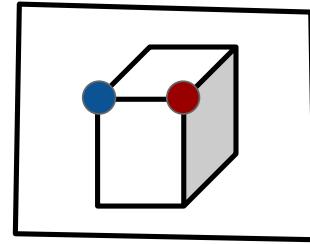


λ Lambda

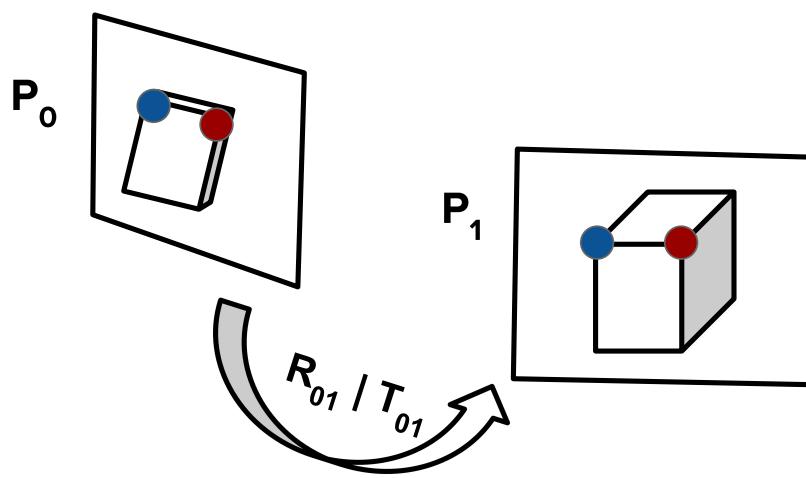
P_0



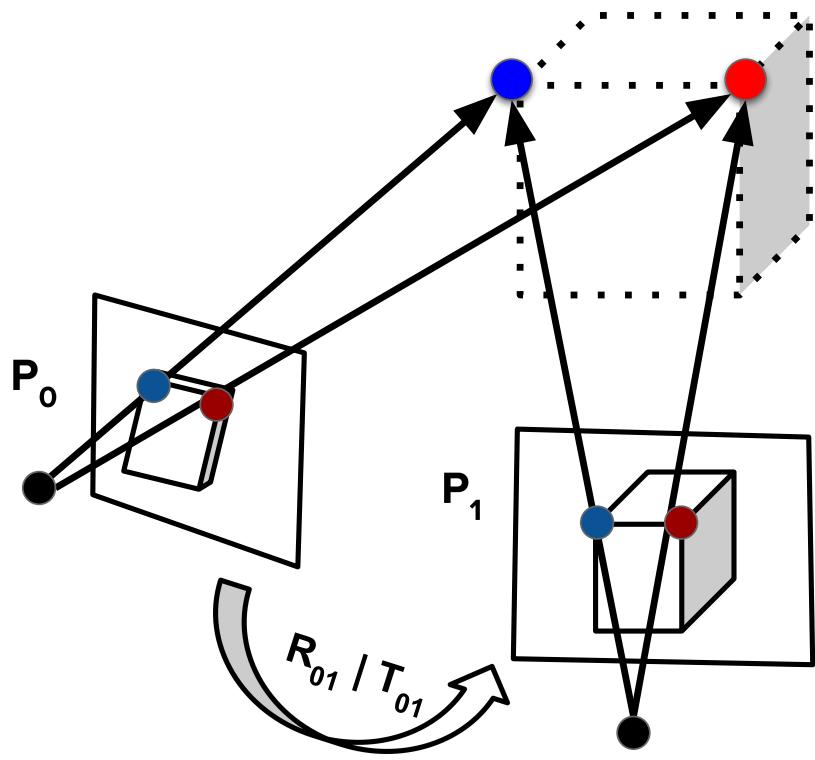
P_1



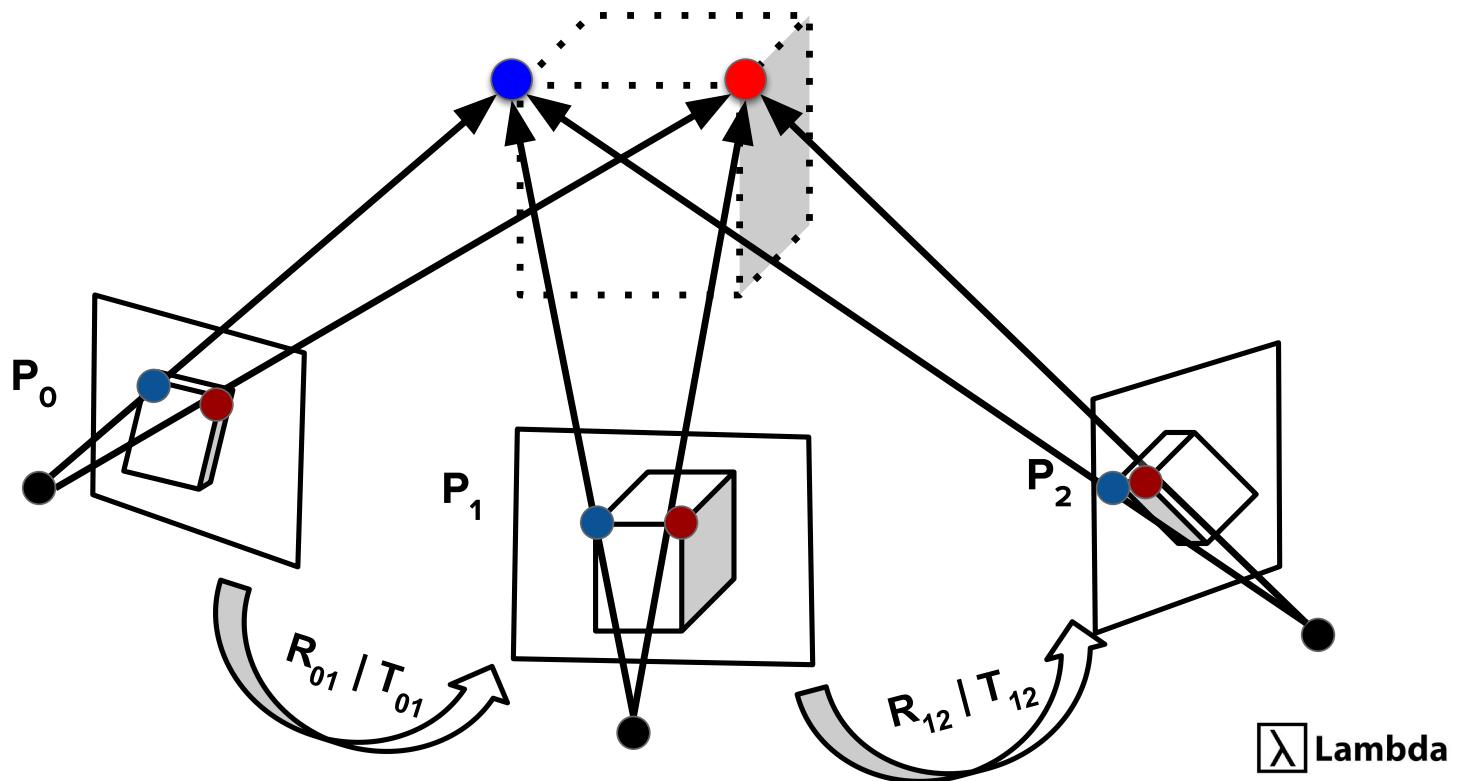
λ Lambda

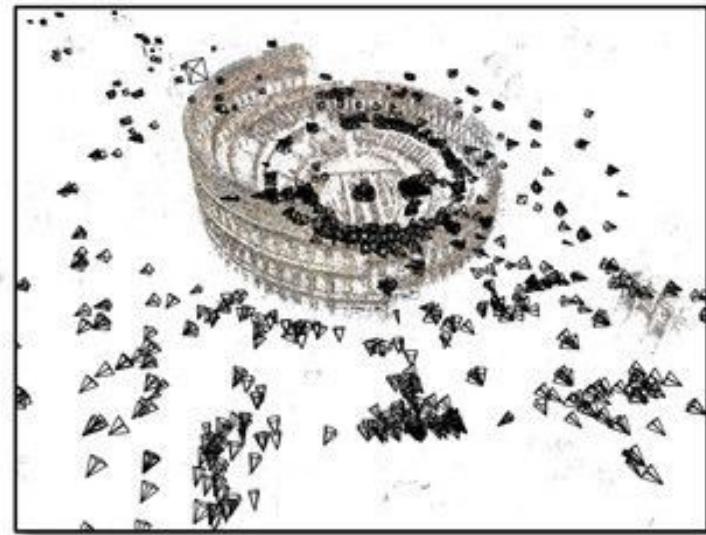


Lambda



λ Lambda

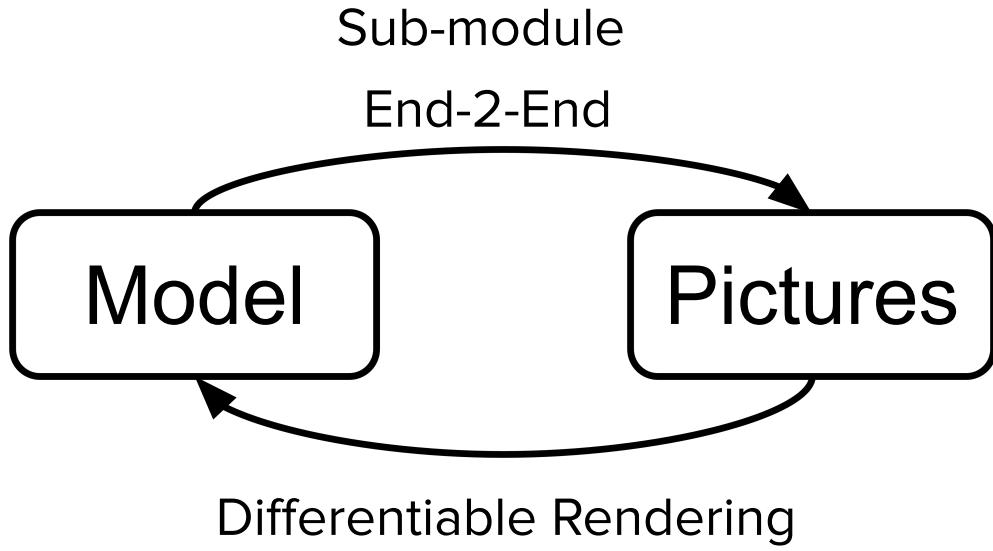




Building Rome in a Day

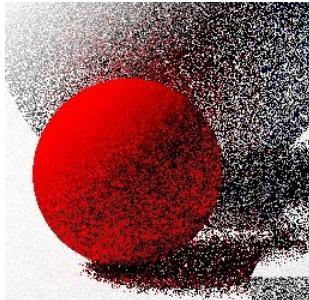
Sameer Agarwal, Noah Snavely, Ian Simon, Steven M. Seitz and Richard Szeliski





λ Lambda

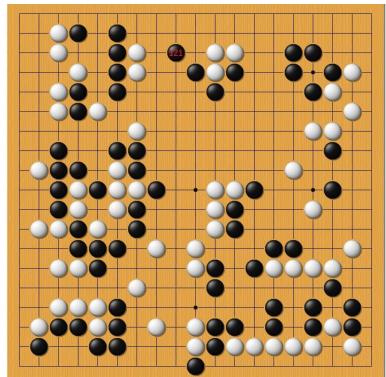
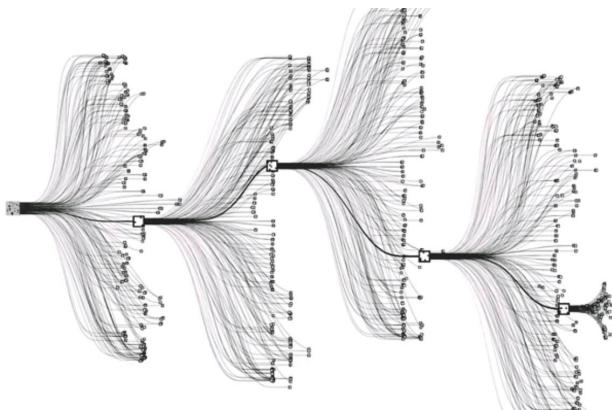
1 SPP



2048 SPP



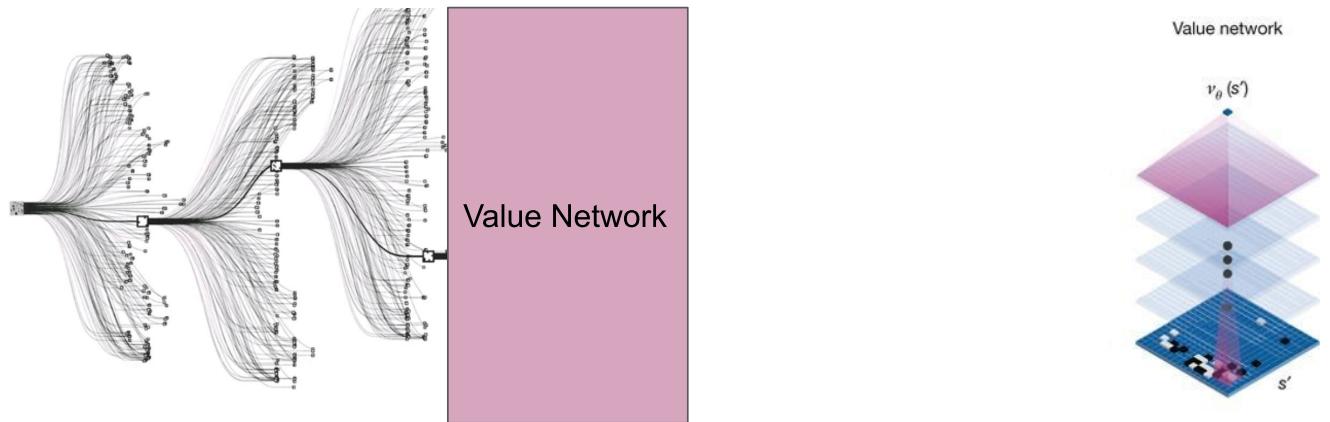
λ Lambda



Mastering the game of Go with deep neural networks and tree search

David Silver et al.

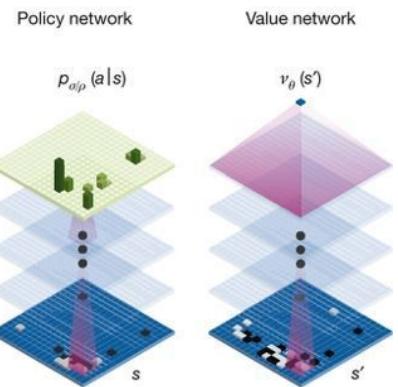
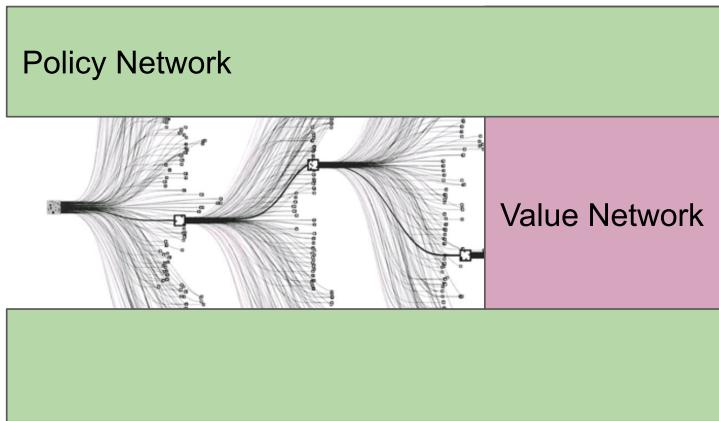




Mastering the game of Go with deep neural networks and tree search

David Silver et al.



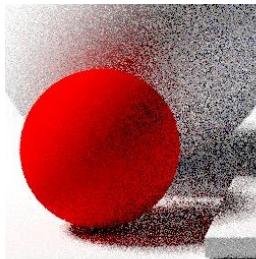


Mastering the game of Go with deep neural networks and tree search

David Silver et al.



4 SPP

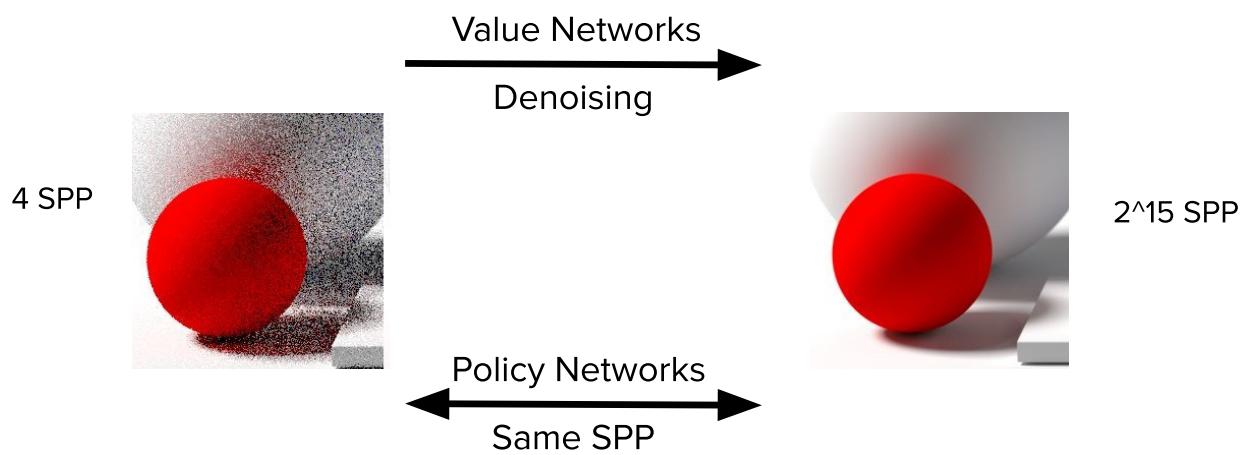


Value Networks
Denoising

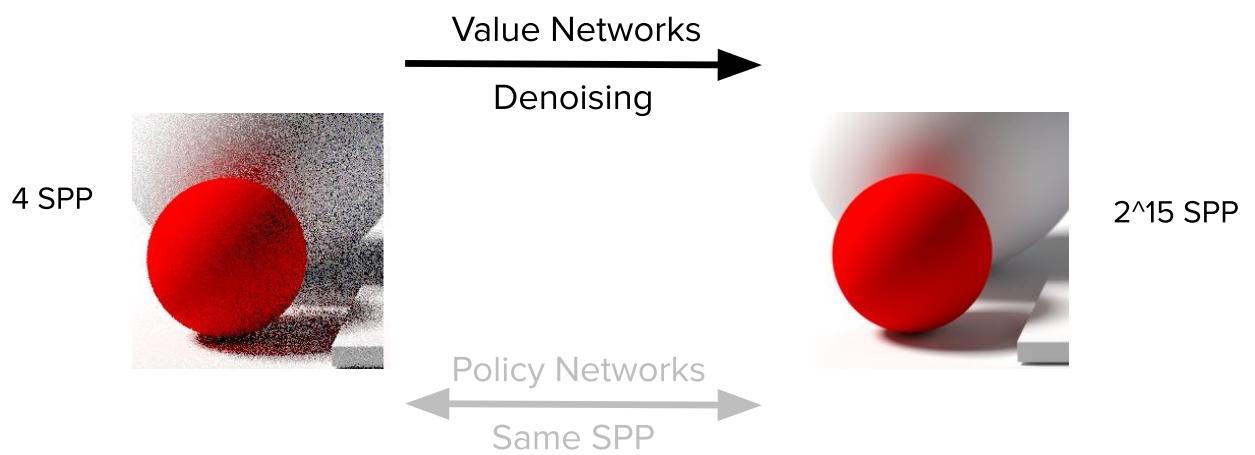


2^{15} SPP

λ Lambda



λ Lambda



λ Lambda



4 SPP



Denoised
1 sec 2080 Ti

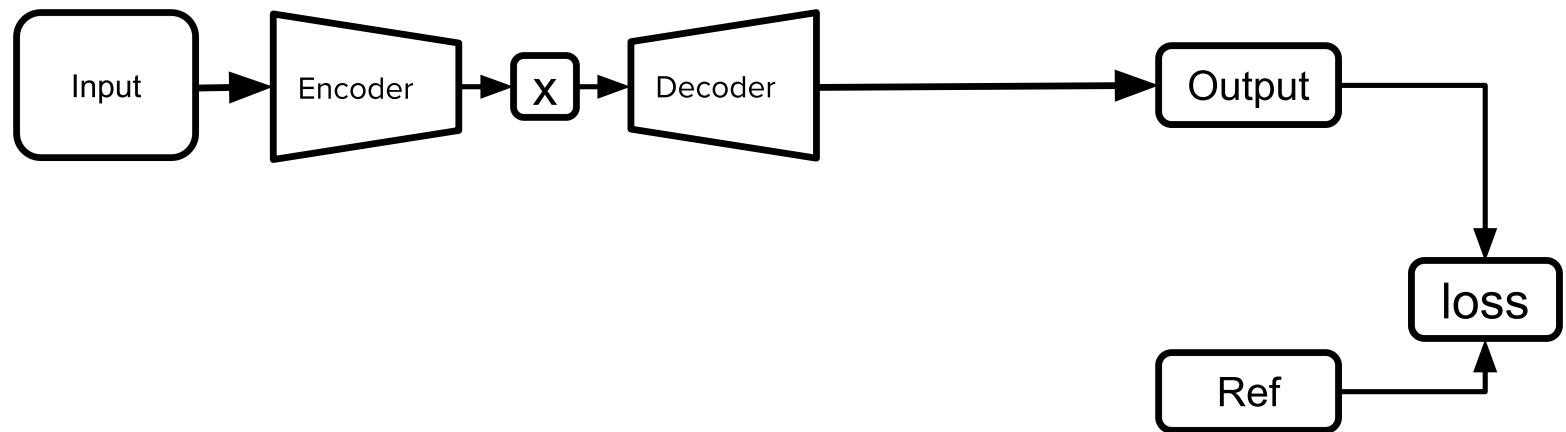


32K SPP Ray Tracing
90 mins 12 cores CPU

Adversarial Monte Carlo denoising with conditioned auxiliary feature modulation

B Xu et al. Siggraph Asia 2019



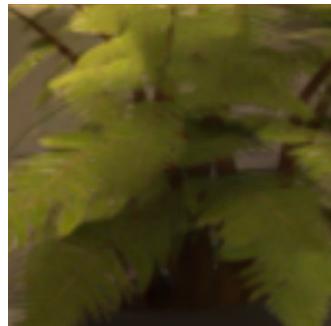


Adversarial Monte Carlo denoising with conditioned auxiliary feature modulation

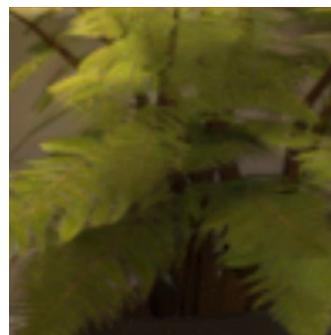
B Xu et al. Siggraph Asia 2019



L1 VGG Loss



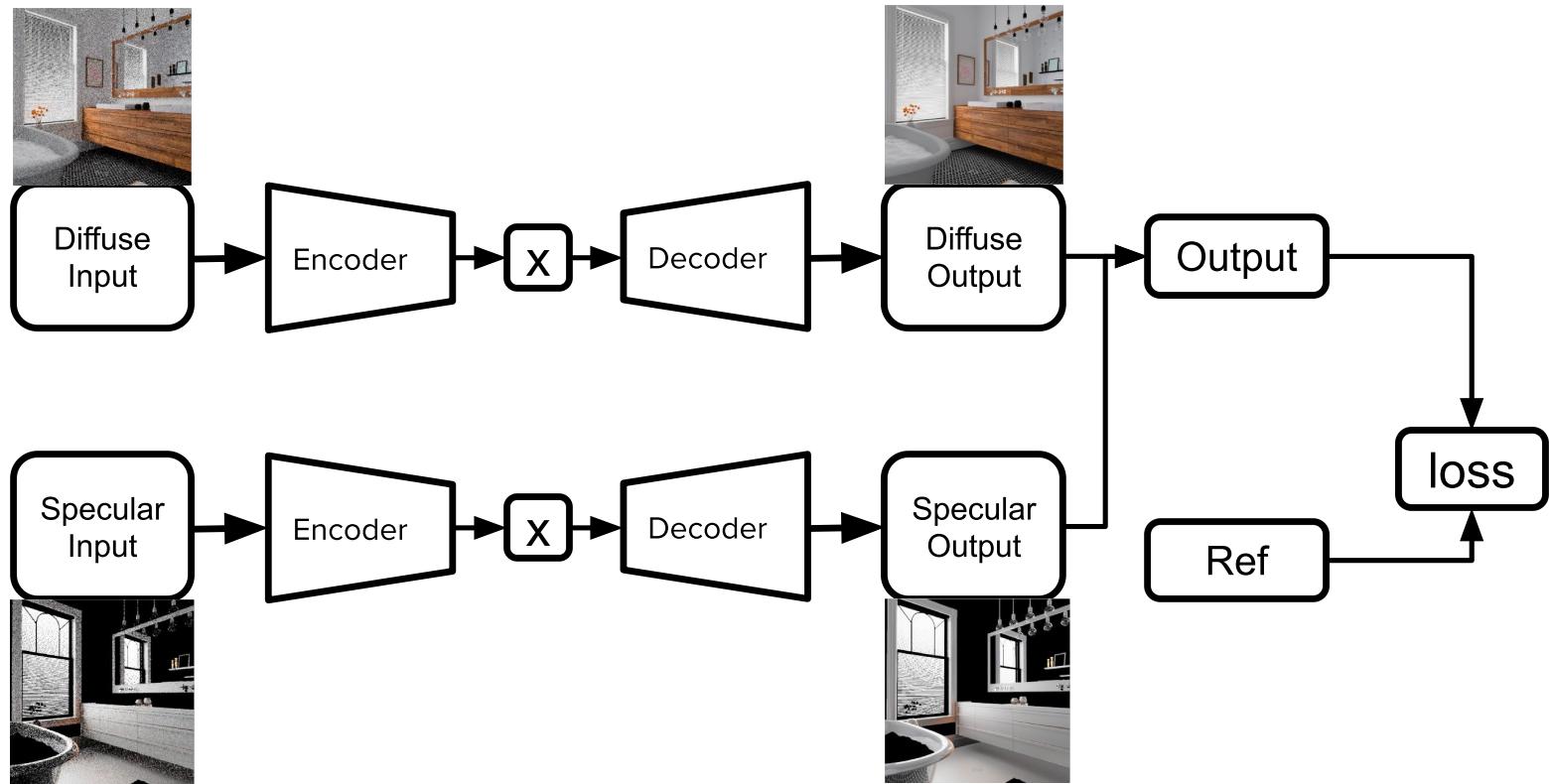
L1 VGG Loss + GAN



Adversarial Monte Carlo denoising with conditioned auxiliary feature modulation

B Xu et al. Siggraph Asia 2019

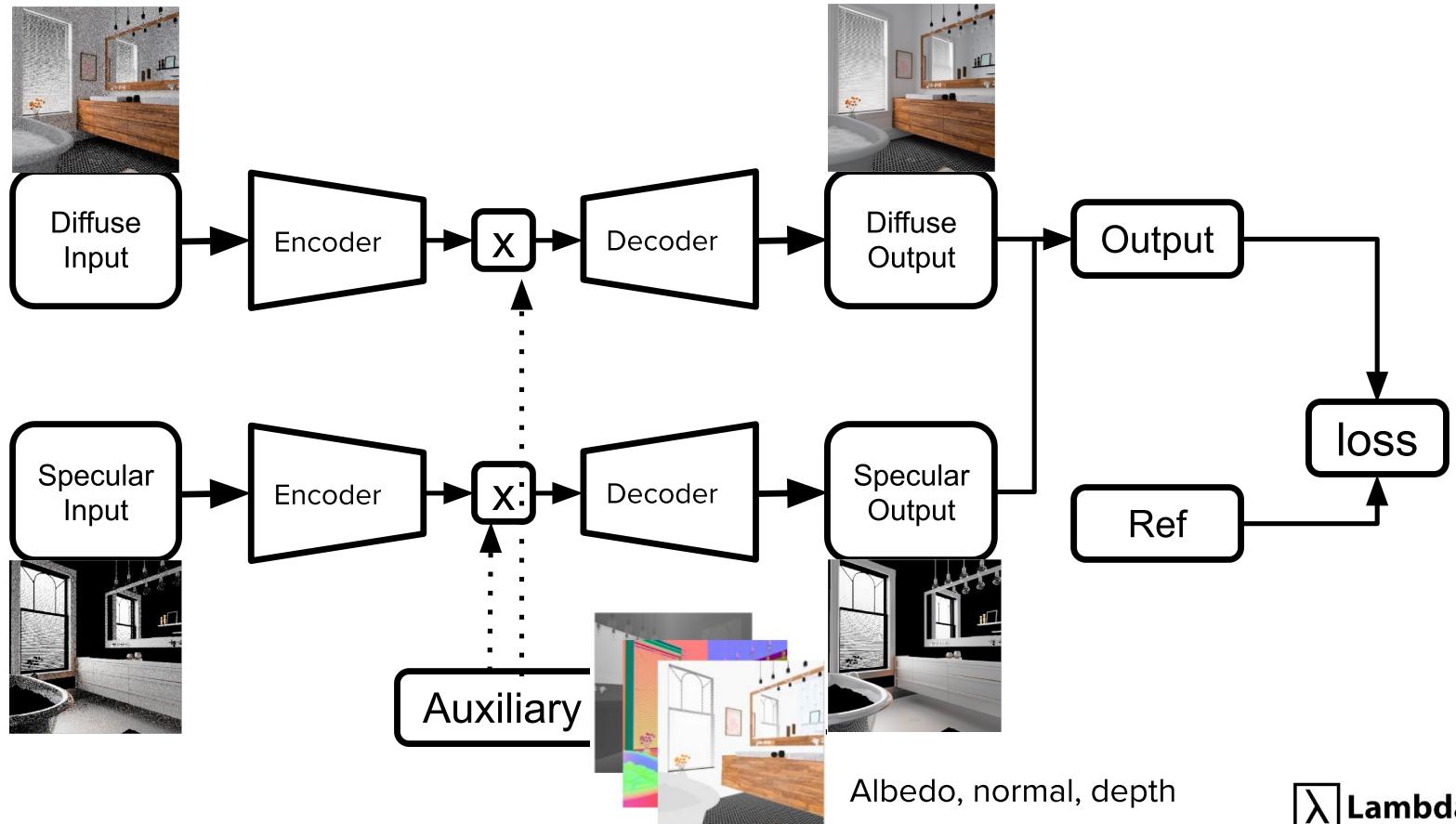


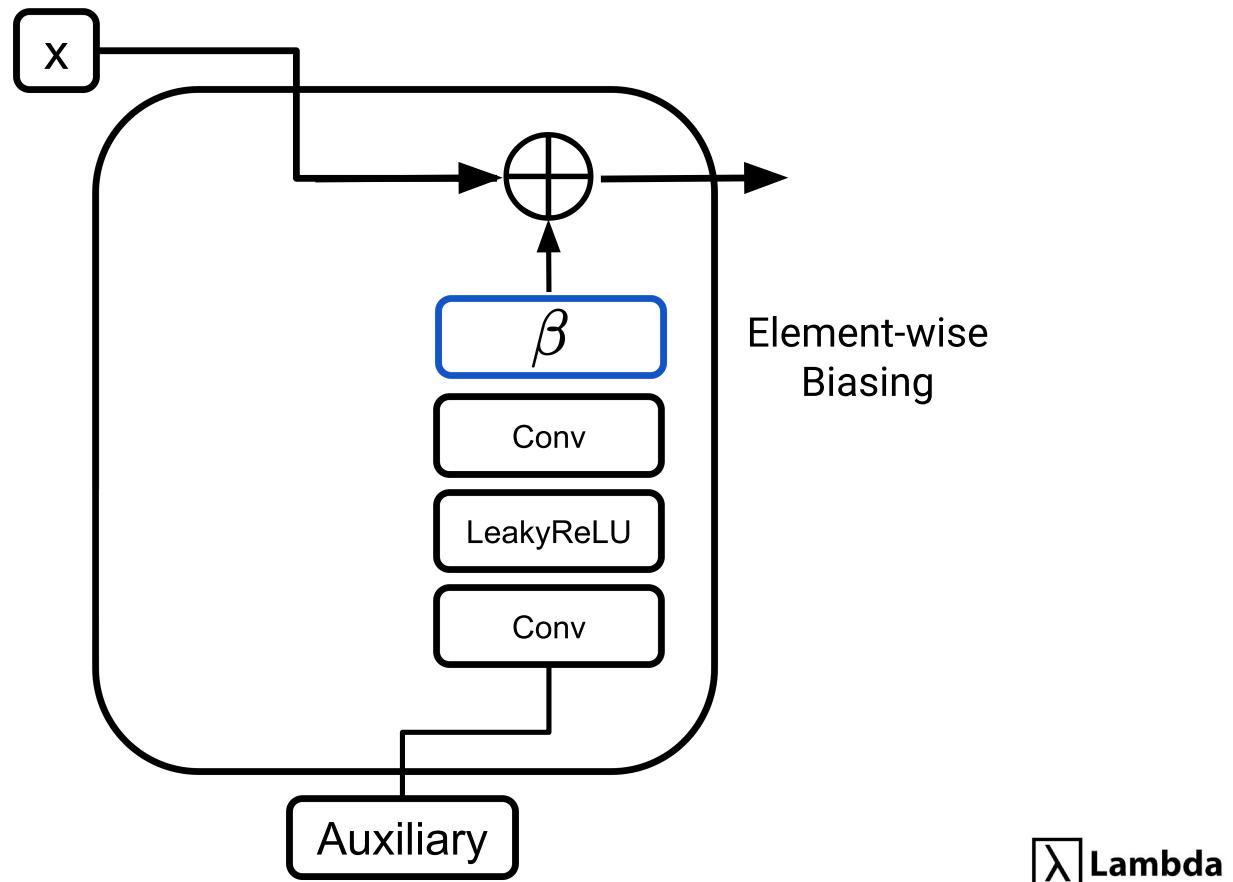


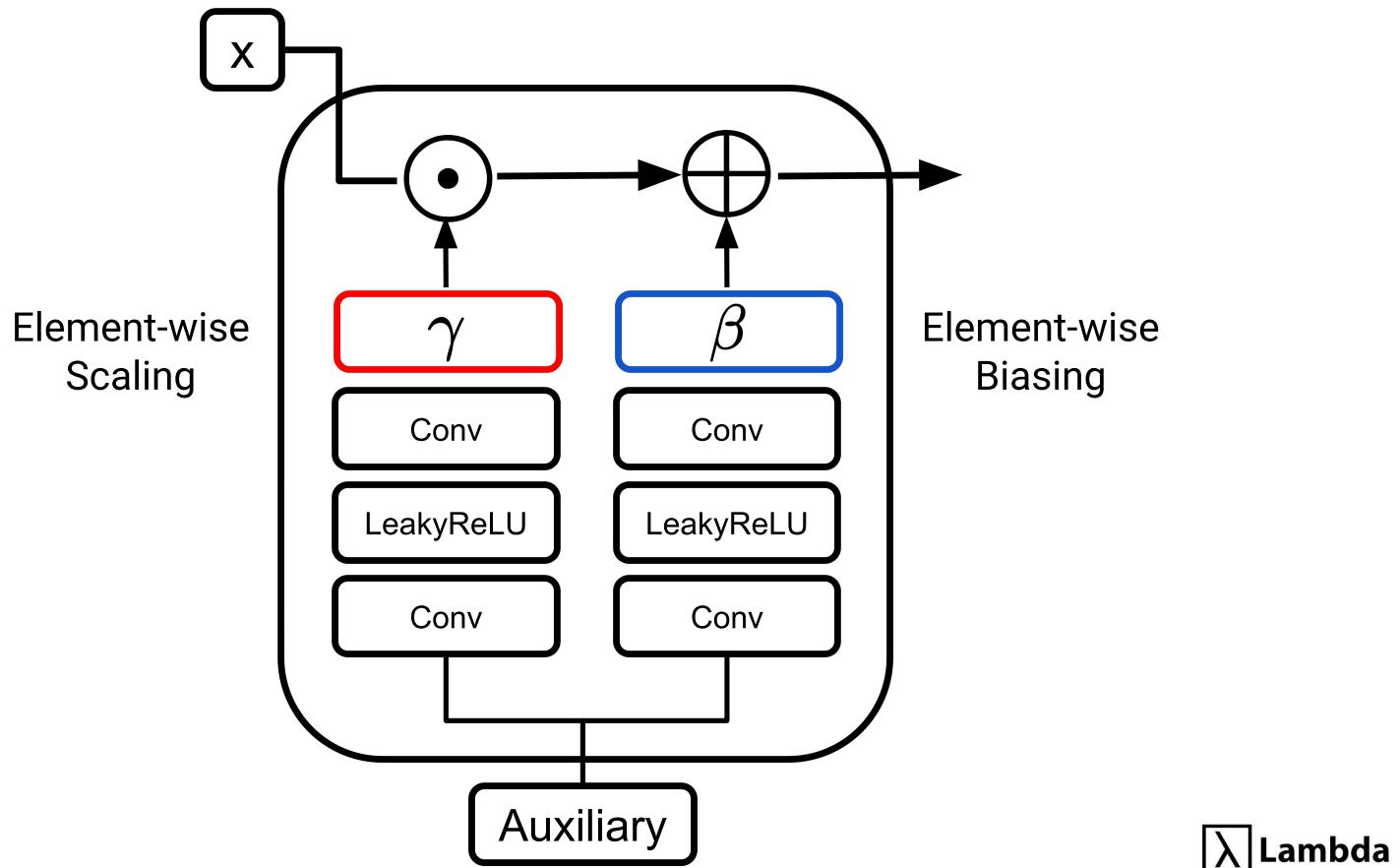
Adversarial Monte Carlo denoising with conditioned auxiliary feature modulation

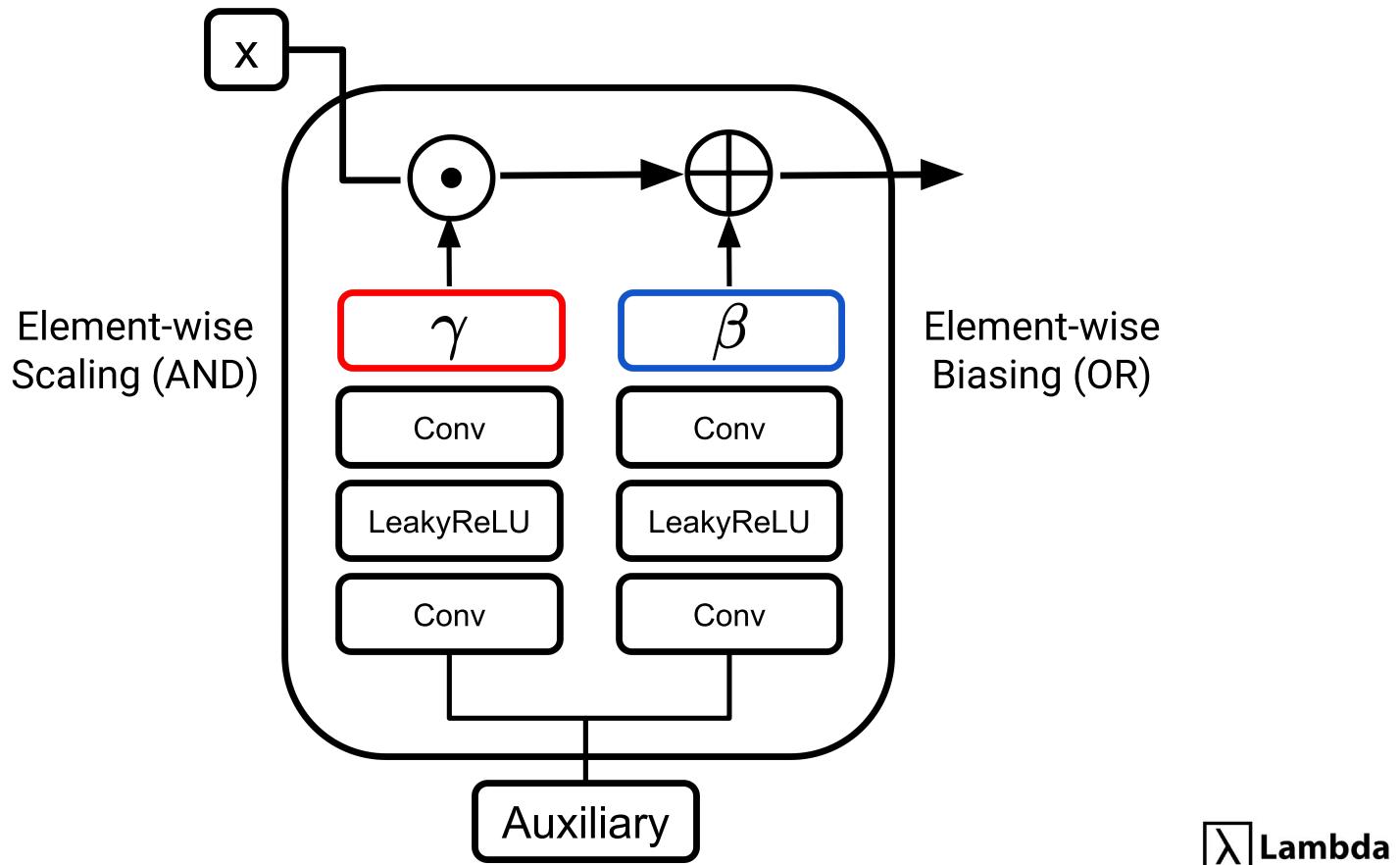
B Xu et al. Siggraph Asia 2019

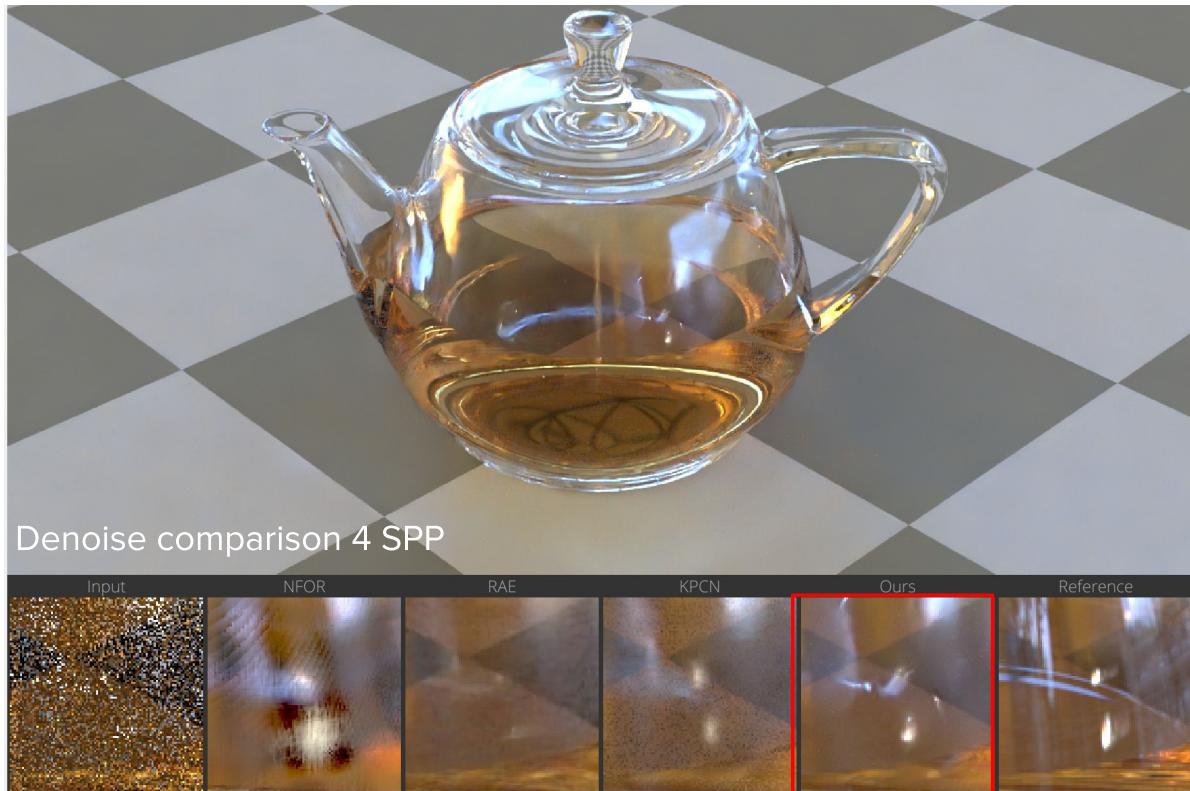








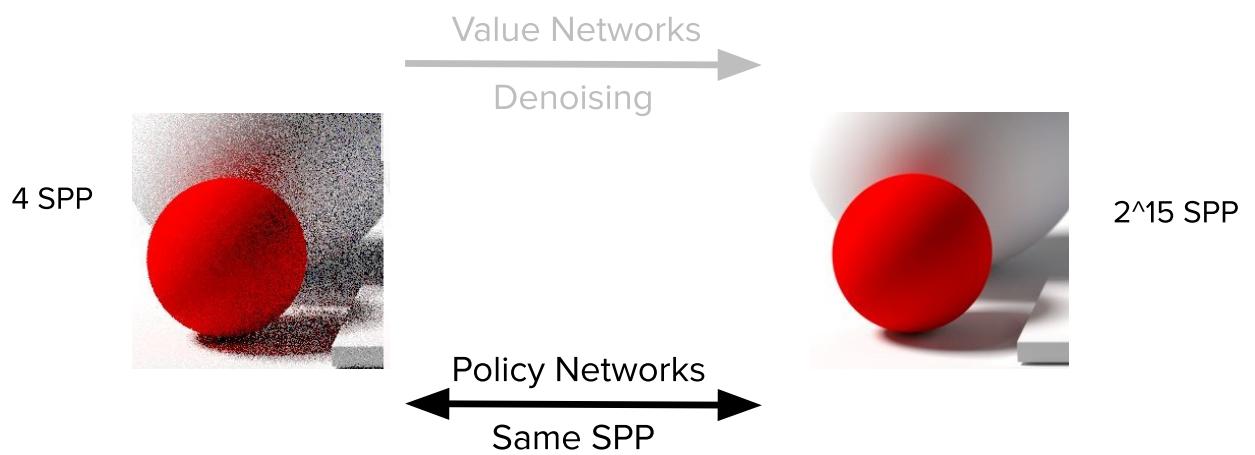




Denoise comparison 4 SPP
Adversarial Monte Carlo denoising with conditioned auxiliary feature modulation

B Xu et al. Siggraph Asia 2019





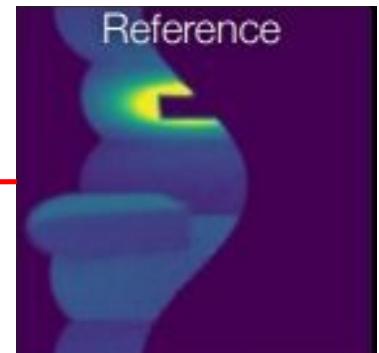
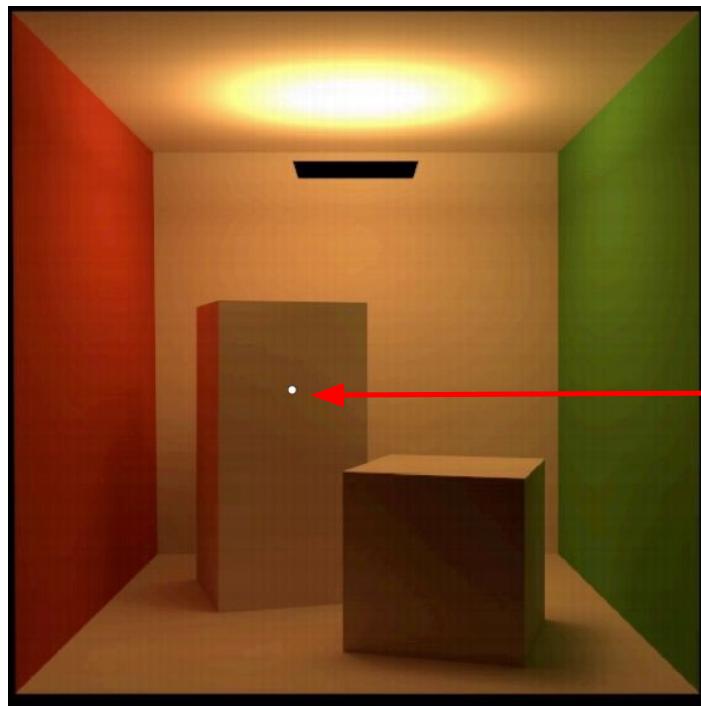
λ Lambda



Neural Importance Sampling

Thomas Müller et al. ACM Transactions on Graphics 2019





incidence radiance map

Neural Importance Sampling

Thomas Müller et al. ACM Transactions on Graphics 2019



$$Z \longrightarrow G(Z)$$

$$\mathcal{N}(0,\sigma^2)$$

3

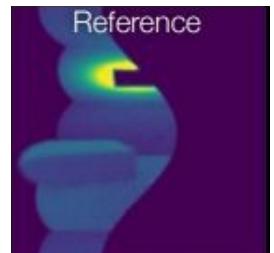
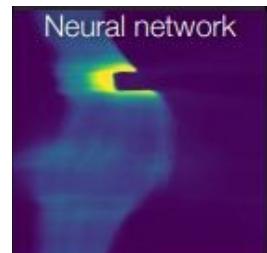
Lambda

$$Z \longrightarrow G(Z)$$

$$\mathcal{N}(0, \sigma^2)$$



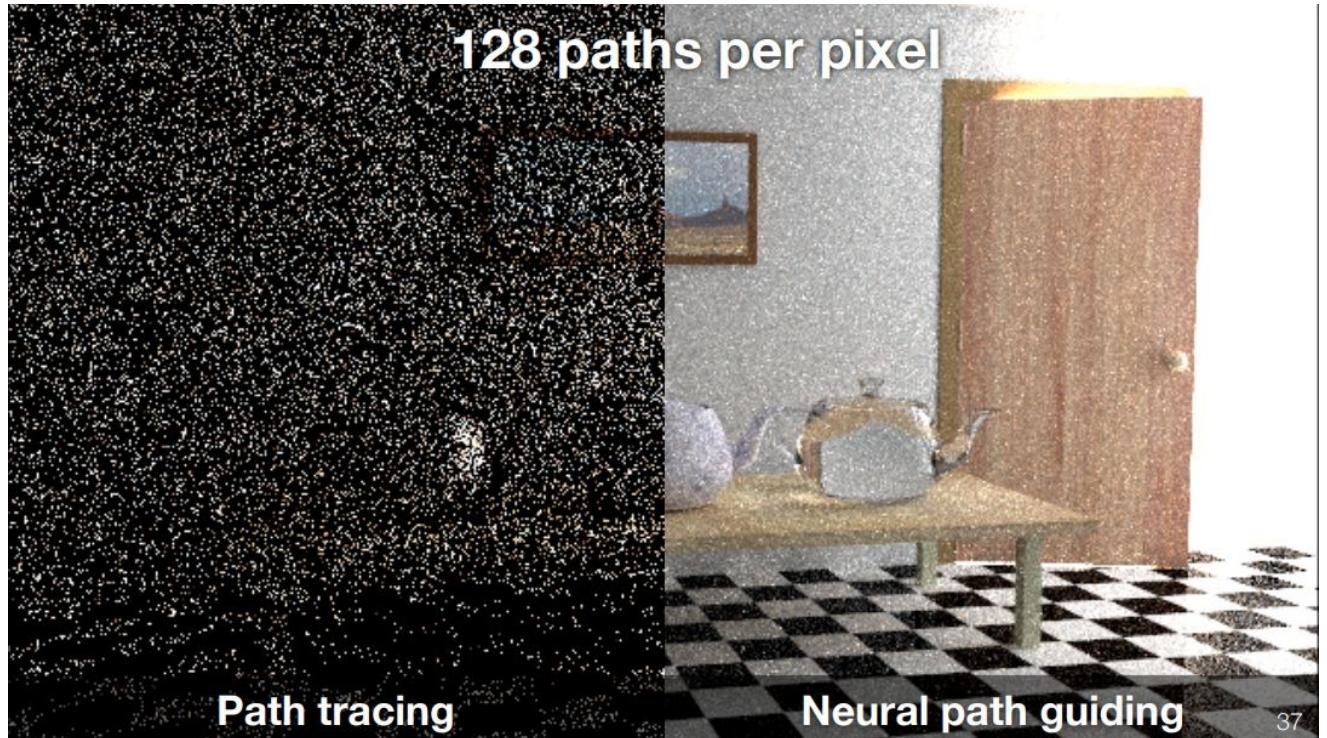
$$[P, \omega_{in}, N]$$



Neural Importance Sampling

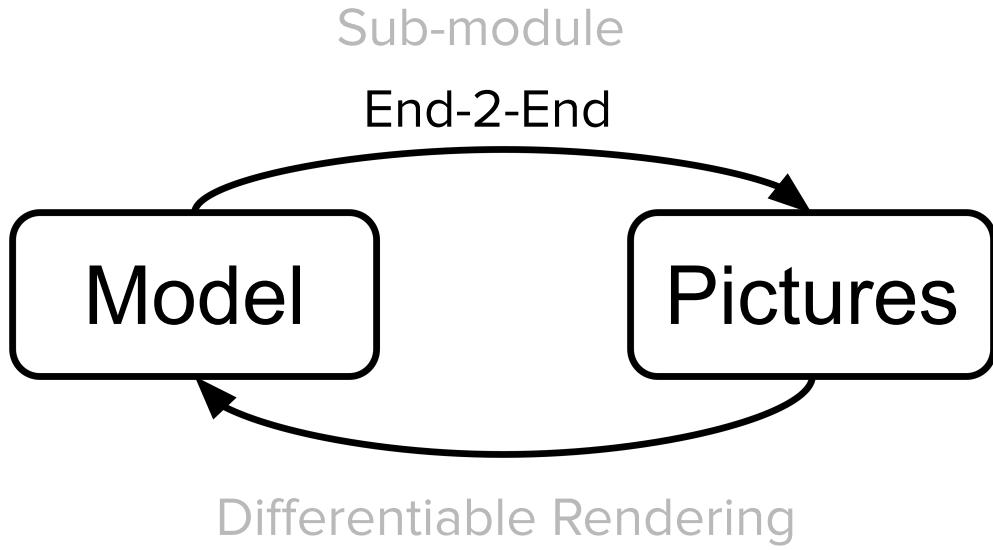
Thomas Müller et al. ACM Transactions on Graphics 2019

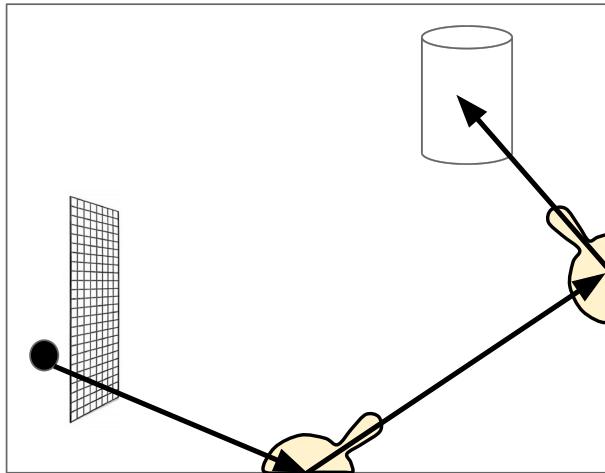




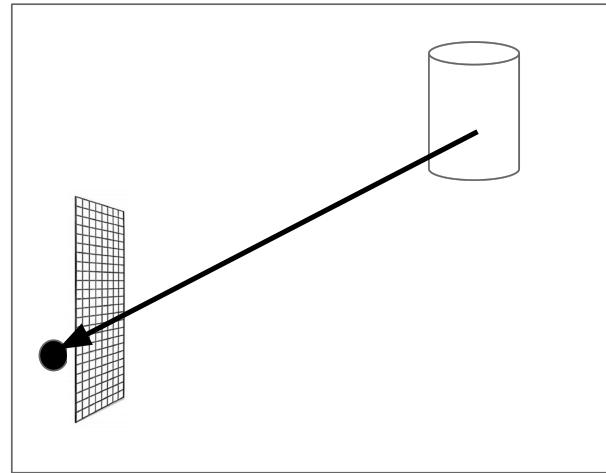
Neural Importance Sampling
Thomas Müller et al. ACM Transactions on Graphics 2019





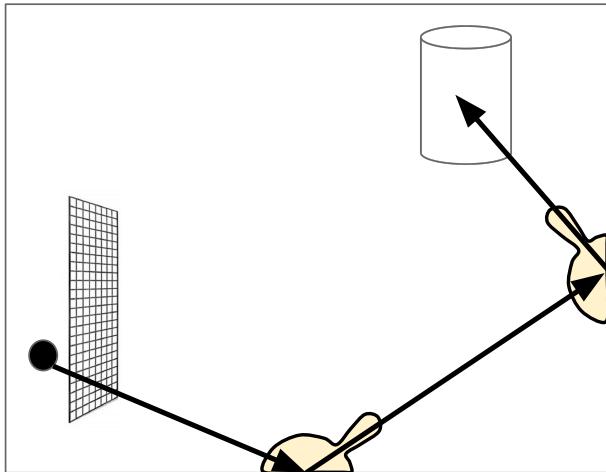


Ray Tracing
Image Centric

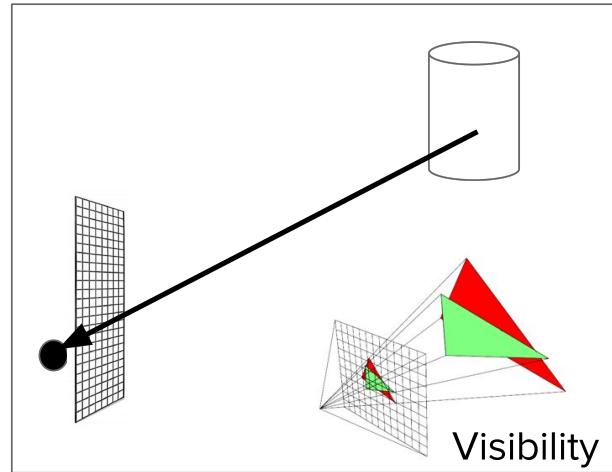


Rasterization
Object Centric

λ Lambda

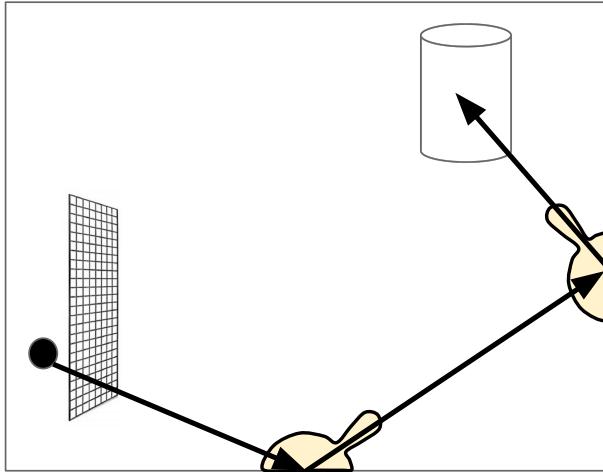


Ray Tracing
Image Centric

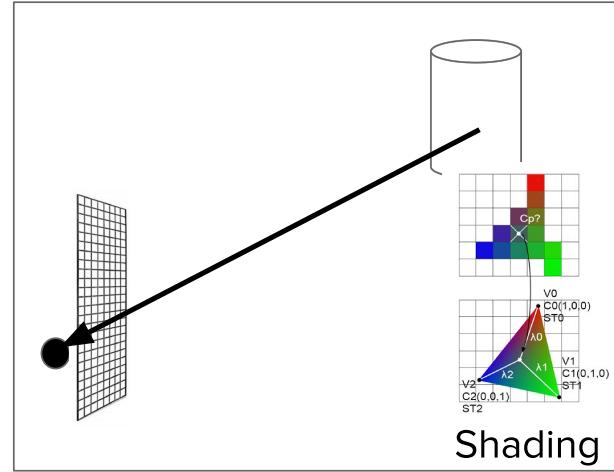


Rasterization
Object Centric

λ Lambda

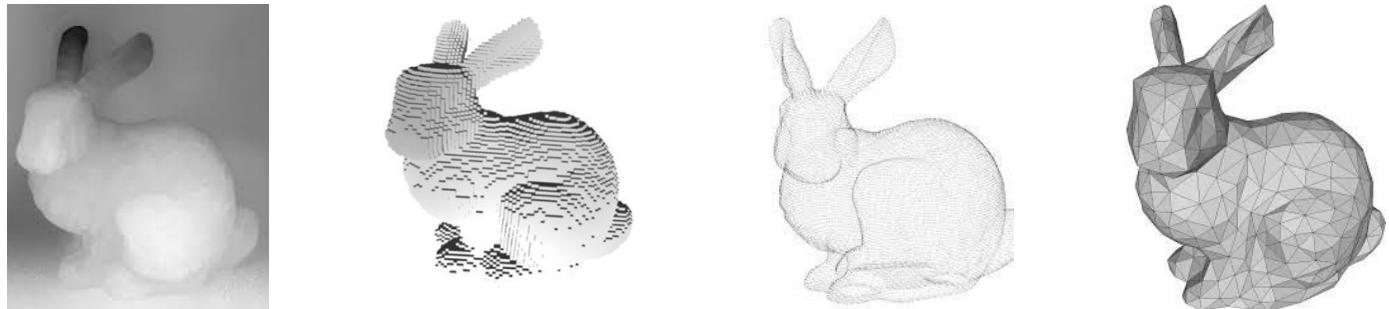


Ray Tracing
Image Centric

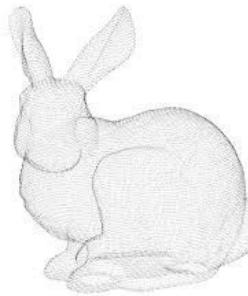


Rasterization
Object Centric

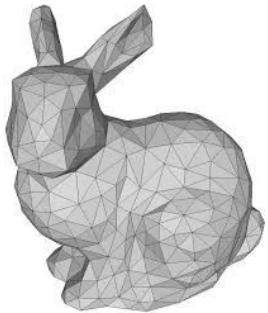
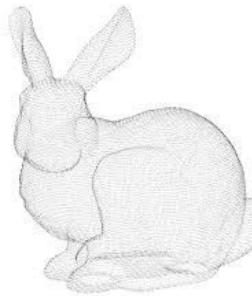
λ Lambda



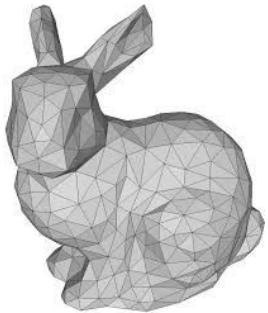
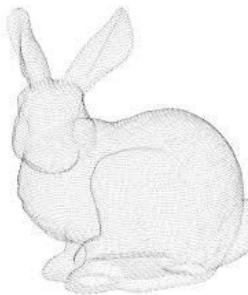
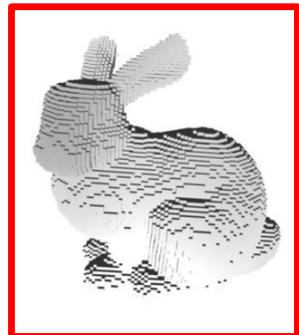
| | Depth Map | Voxel | Point Cloud | Mesh |
|-------------|-----------|-----------|-------------|-----------|
| Memory | Good | Very Poor | Poor | Very Good |
| NN friendly | Great | Yes | No | Enemy |



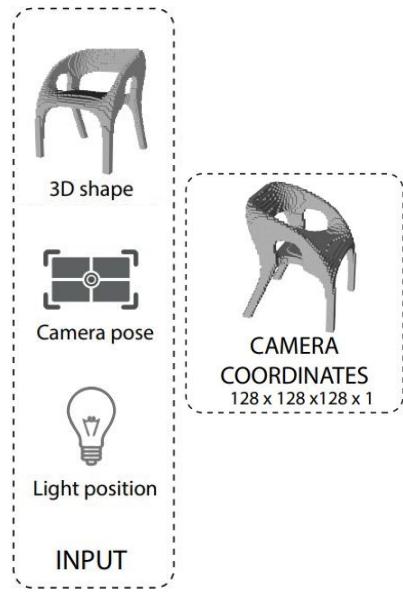
| | Depth Map | Voxel | Point Cloud | Mesh |
|-------------|-----------|-----------|-------------|-----------|
| Memory | Good | Very Poor | Poor | Very Good |
| NN friendly | Great | Yes | No | Enemy |



| | Depth Map | Voxel | Point Cloud | Mesh |
|-------------|-----------|-----------|-------------|-----------|
| Memory | Good | Very Poor | Poor | Very Good |
| NN friendly | Great | Yes | No | Enemy |



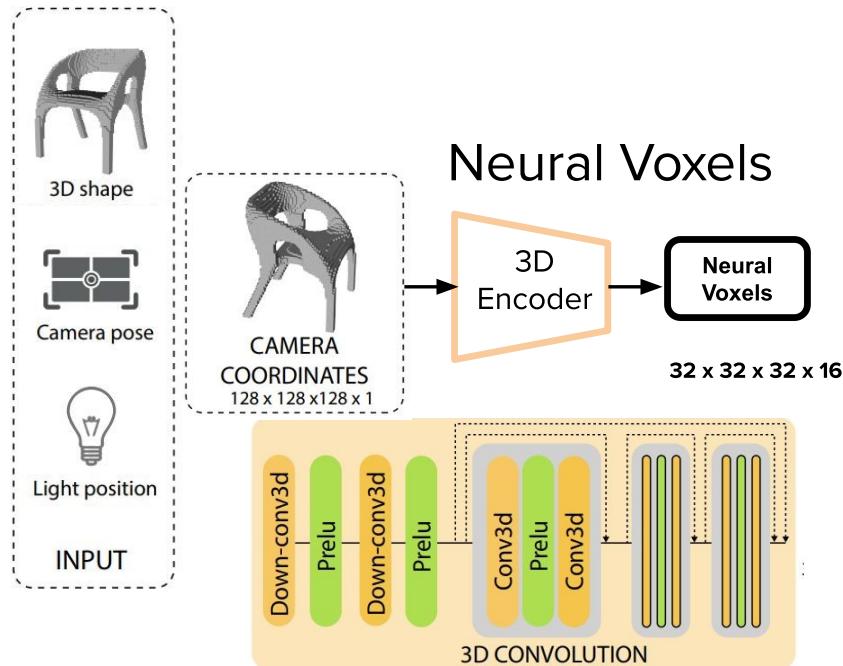
| | Depth Map | Voxel | Point Cloud | Mesh |
|-------------|-----------|-----------|-------------|-----------|
| Memory | Good | Very Poor | Poor | Very Good |
| NN friendly | Great | Yes | No | Enemy |



RenderNet: A deep convolutional network for differentiable rendering from 3D shapes

Thu Nguyen-Phuoc et al. NeurIPS 2018

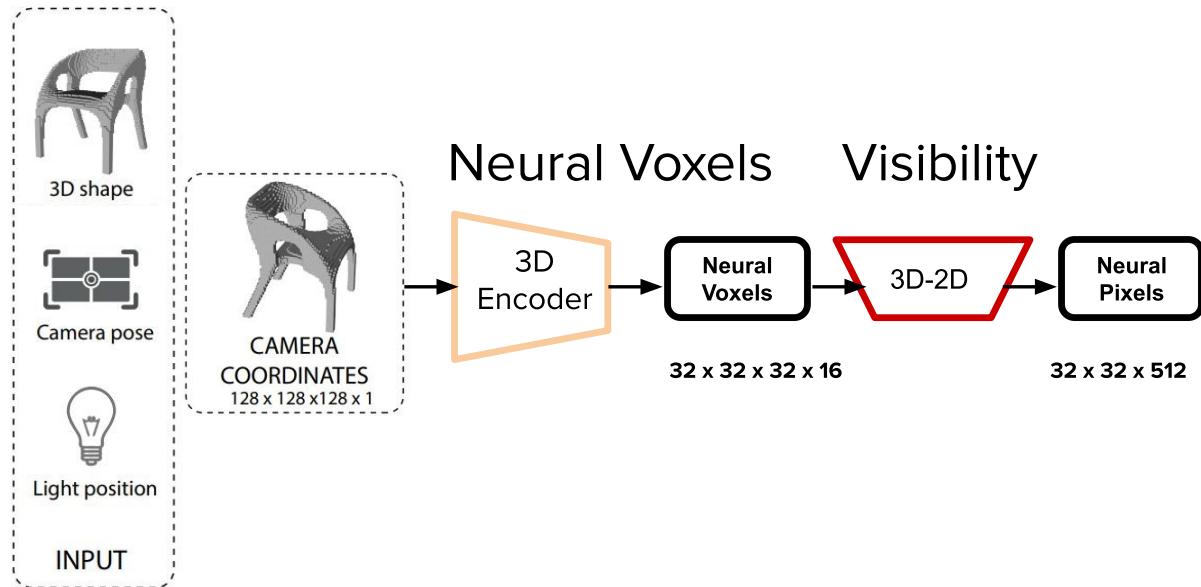




RenderNet: A deep convolutional network for differentiable rendering from 3D shapes

Thu Nguyen-Phuoc et al. NeurIPS 2018

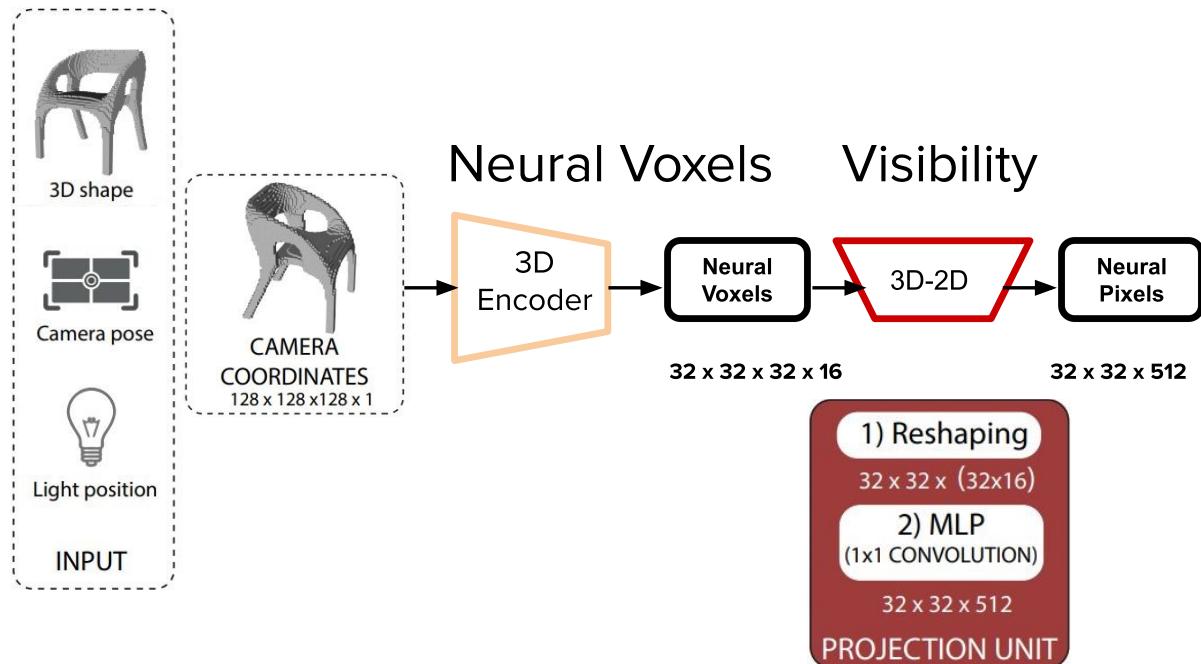




RenderNet: A deep convolutional network for differentiable rendering from 3D shapes

Thu Nguyen-Phuoc et al. NeurIPS 2018

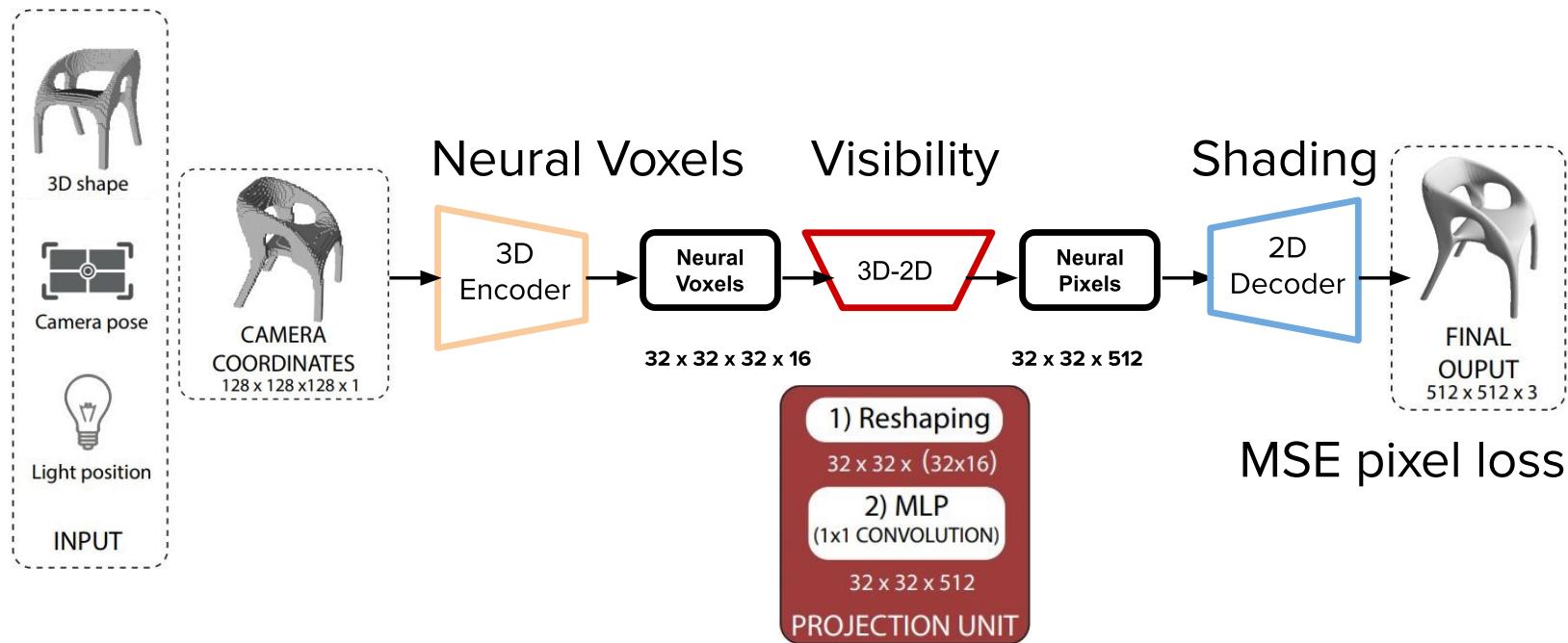




RenderNet: A deep convolutional network for differentiable rendering from 3D shapes

Thu Nguyen-Phuoc et al. NeurIPS 2018





RenderNet: A deep convolutional network for differentiable rendering from 3D shapes

Thu Nguyen-Phuoc et al. NeurIPS 2018





RenderNet: A deep convolutional network for differentiable rendering from 3D shapes

Thu Nguyen-Phuoc et al. NeurIPS 2018





Contour



Toon

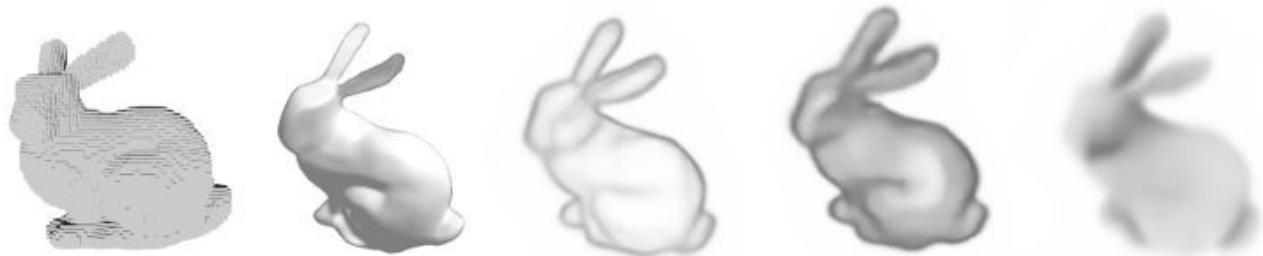


Ambient Occlusion

RenderNet: A deep convolutional network for differentiable rendering from 3D shapes

Thu Nguyen-Phuoc et al. NeurIPS 2018



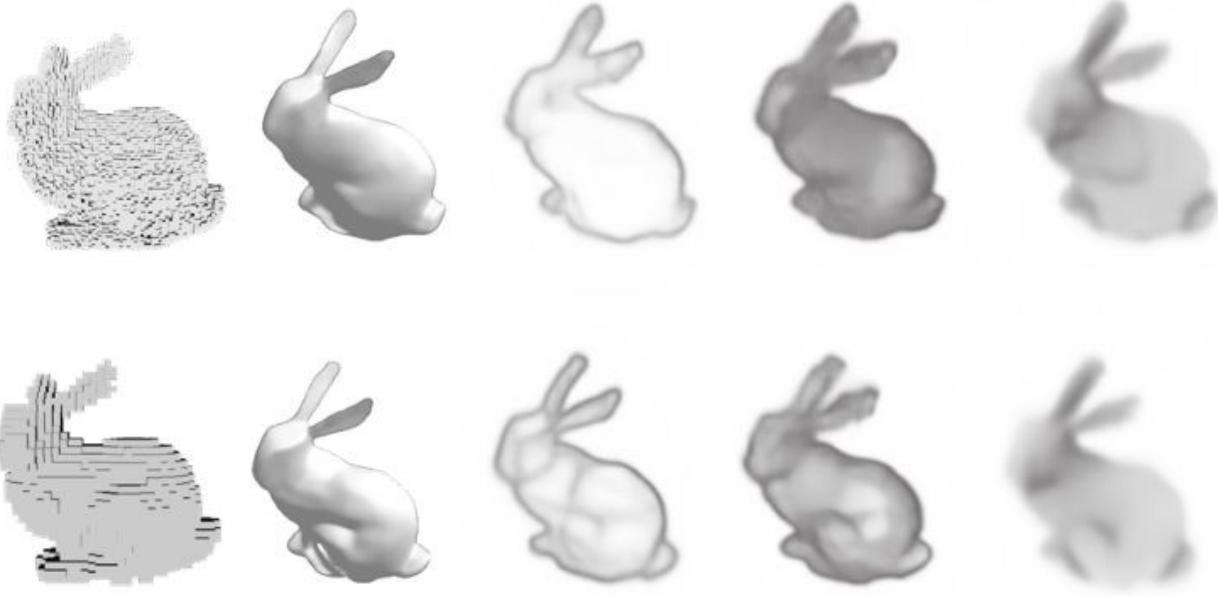


RenderNet: A deep convolutional network for differentiable rendering from 3D shapes

Thu Nguyen-Phuoc et al. NeurIPS 2018



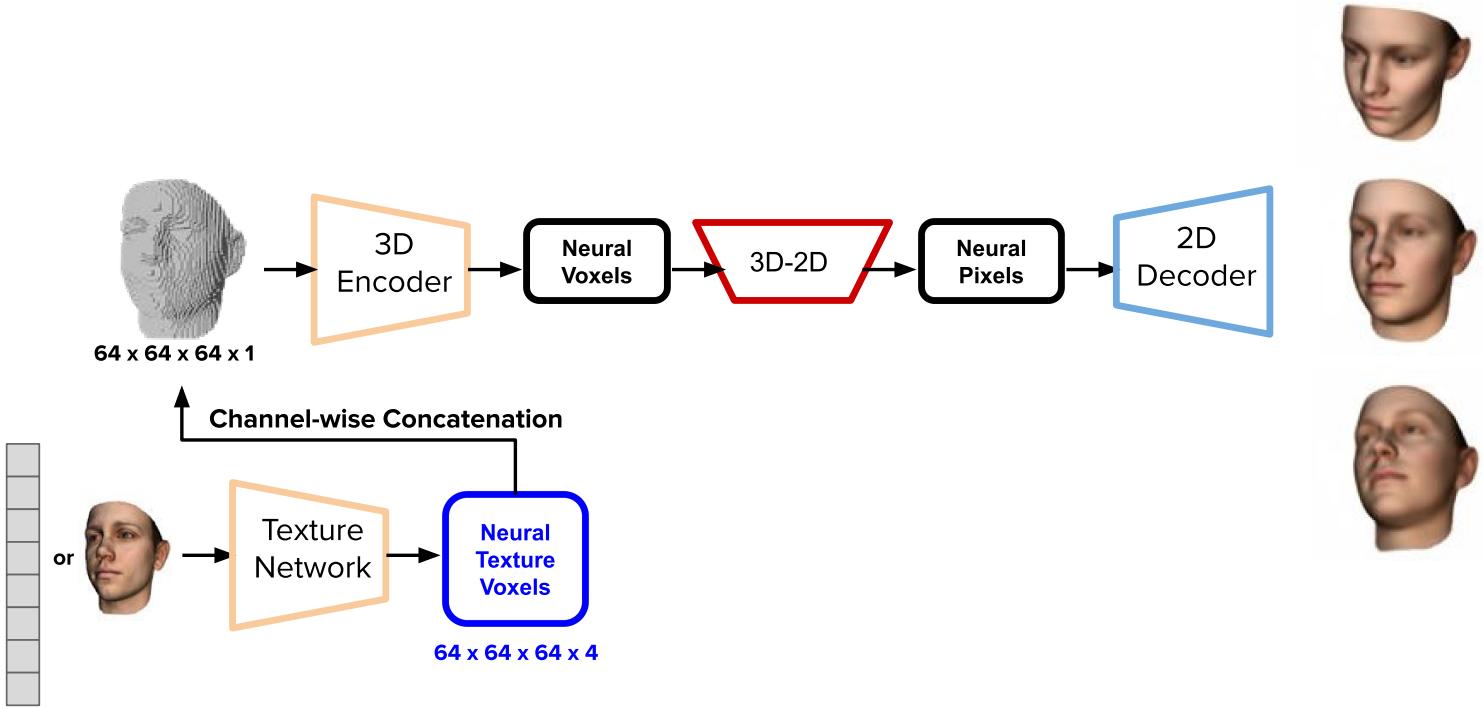
Corrupted
Low-res
(x0.5 original res)



RenderNet: A deep convolutional network for differentiable rendering from 3D shapes

Thu Nguyen-Phuoc et al. NeurIPS 2018

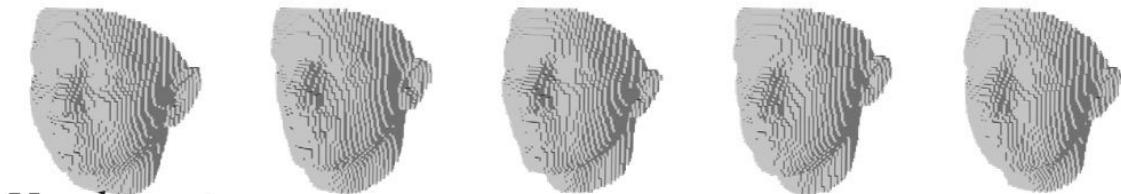
λ Lambda



RenderNet: A deep convolutional network for differentiable rendering from 3D shapes

Thu Nguyen-Phuoc et al. NeurIPS 2018





Voxel input



GT

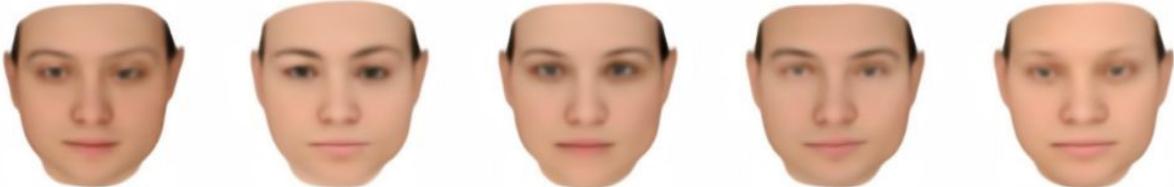


Results

RenderNet: A deep convolutional network for differentiable rendering from 3D shapes

Thu Nguyen-Phuoc et al. NeurIPS 2018





Same shape, different textures

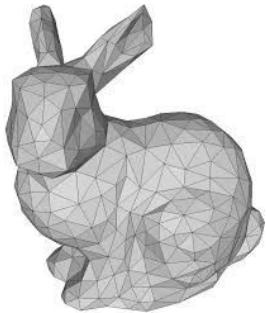
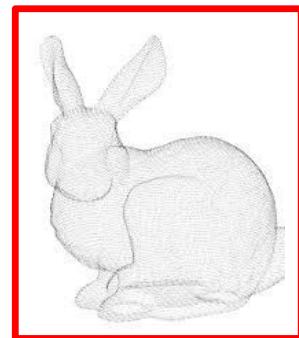


Same texture, different shapes

RenderNet: A deep convolutional network for differentiable rendering from 3D shapes

Thu Nguyen-Phuoc et al. NeurIPS 2018





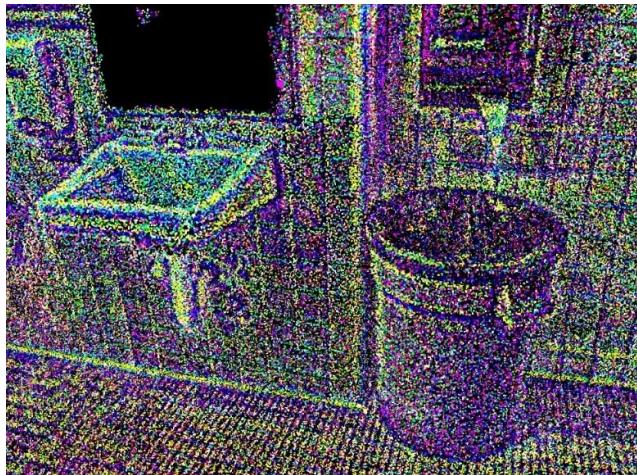
| | Depth Map | Voxel | Point Cloud | Mesh |
|-------------|-----------|-----------|-------------|-----------|
| Memory | Good | Very Poor | Poor | Very Good |
| NN friendly | Great | Yes | No | Enemy |



Rasterization a RGB point cloud

Neural Point-Based Graphics
KA Aliev et al, arxiv 2019





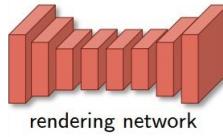
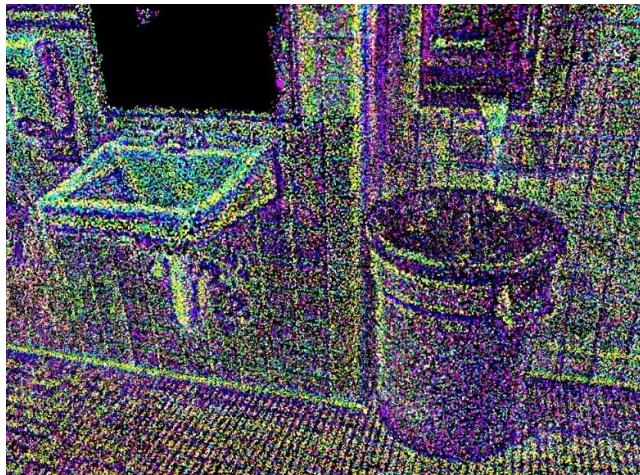
Rasterization a neural point cloud

(First three PCA dimensions of the neural descriptor)

Neural Point-Based Graphics

KA Aliev et al, arxiv 2019





Rasterization a neural point cloud

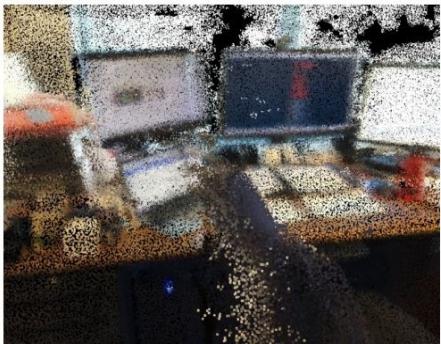
(First three PCA dimensions of the neural descriptor)

Neural Point-Based Graphics

KA Aliev et al, arxiv 2019



RBG rasterization



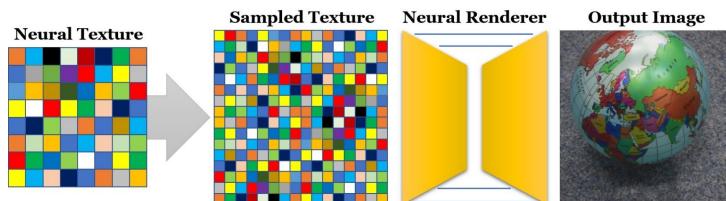
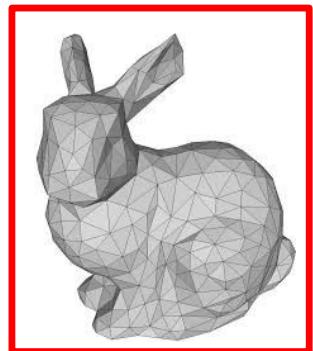
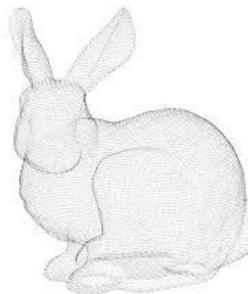
Neural rasterization



Neural Point-Based Graphics

KA Aliev et al, arxiv 2019



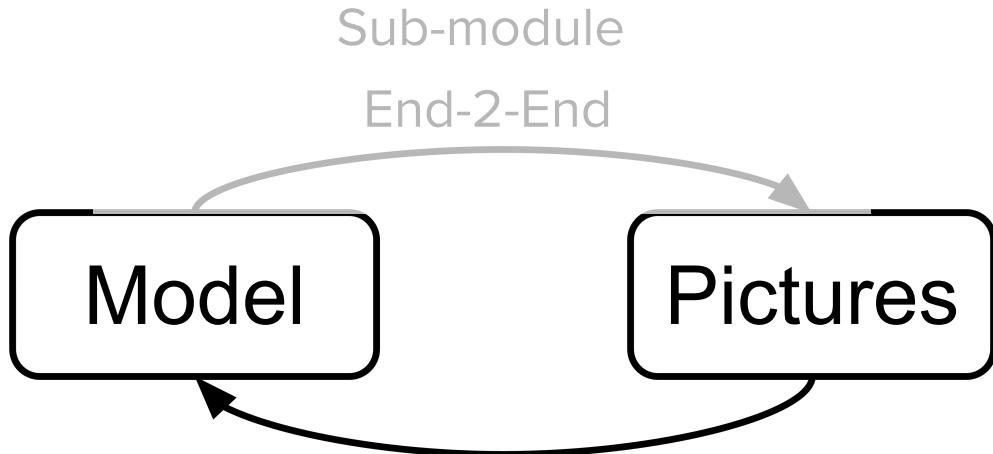


Deferred Neural Rendering:
Image Synthesis using Neural Textures
J Thies et al, Siggraph 2019

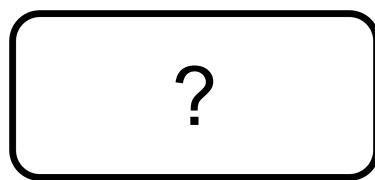


Neural 3D Mesh Renderer
H Kato et al, CVPR 2018



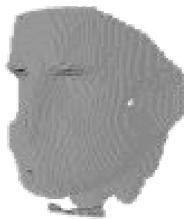


λ Lambda



Lambda

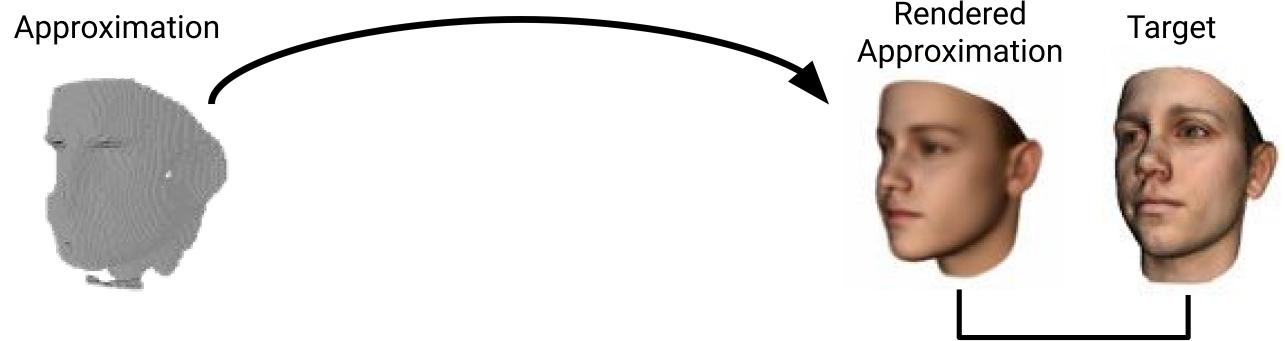
Approximation



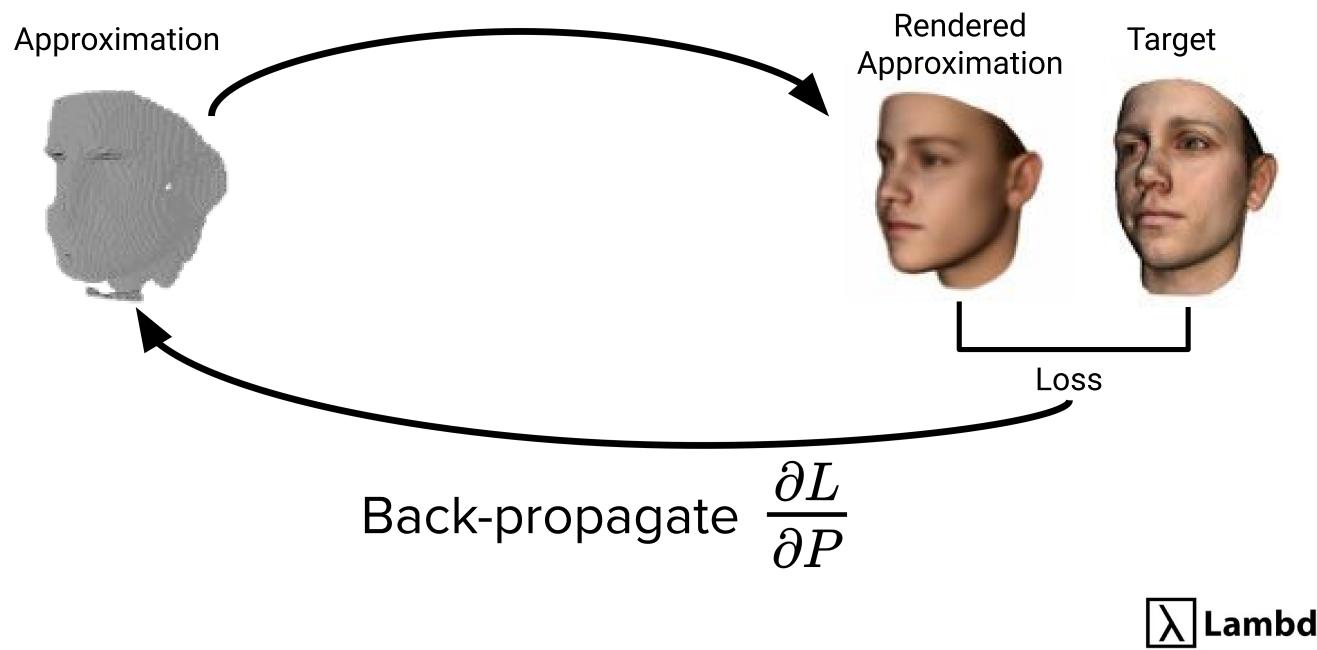
Target

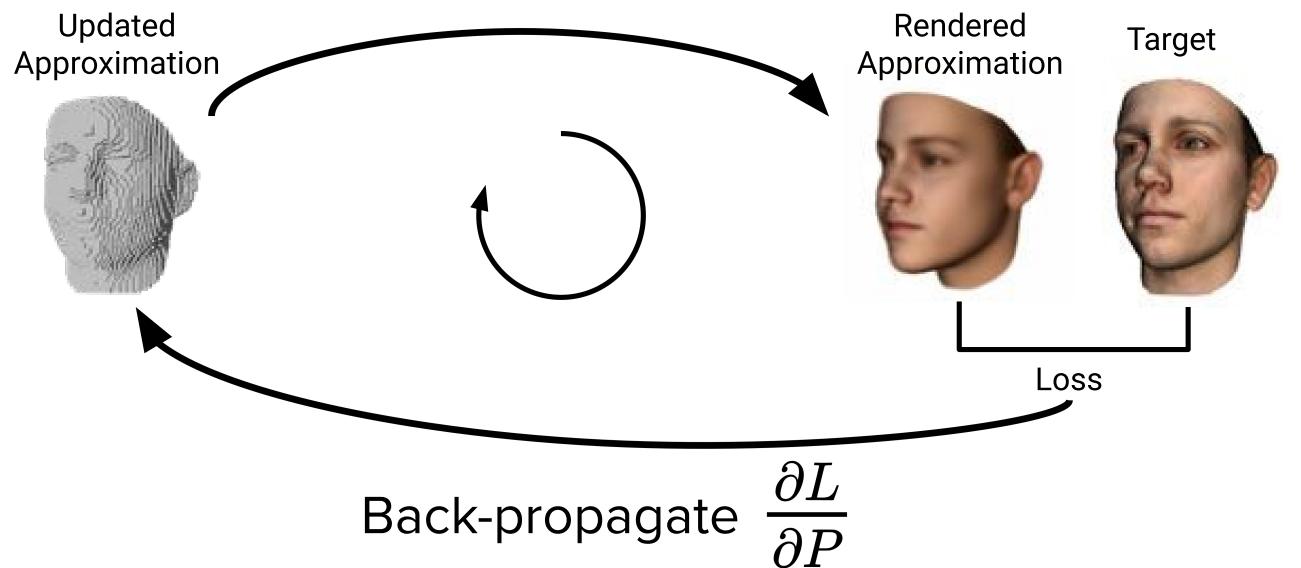


λ Lambda

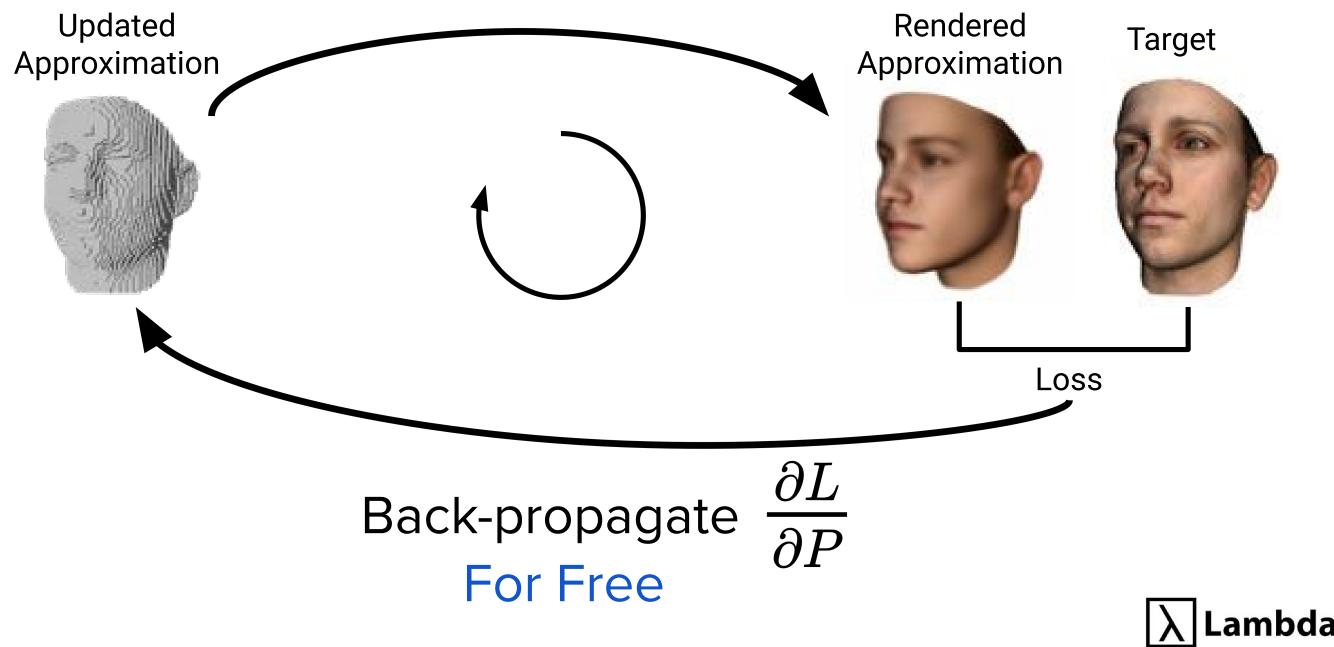


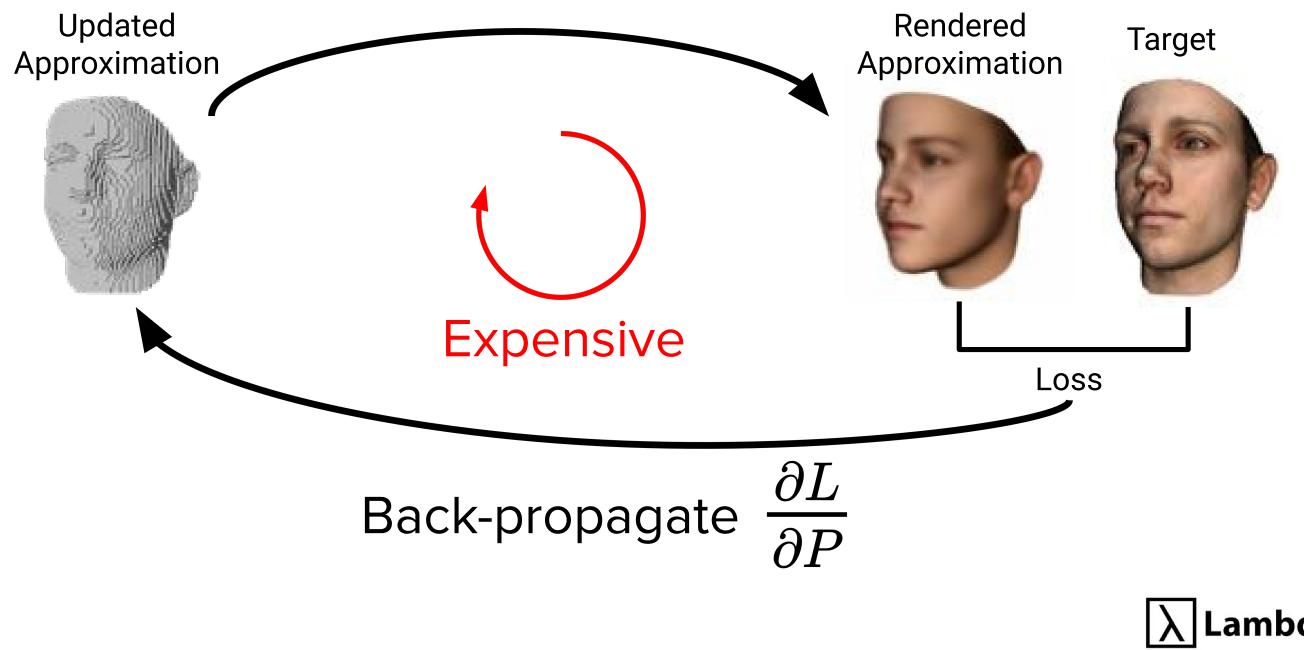
λ Lambda

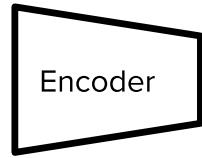




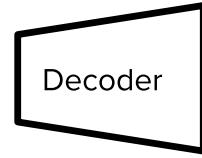
λ Lambda







Reconstruction



Rendering

Rendered
Approximation



Loss

λ Lambda

Inductive Bias: Separate Appearance from Pose



Human perception imposes coordinate frame on objects



Learning 3D representation from natural images without 3D supervision

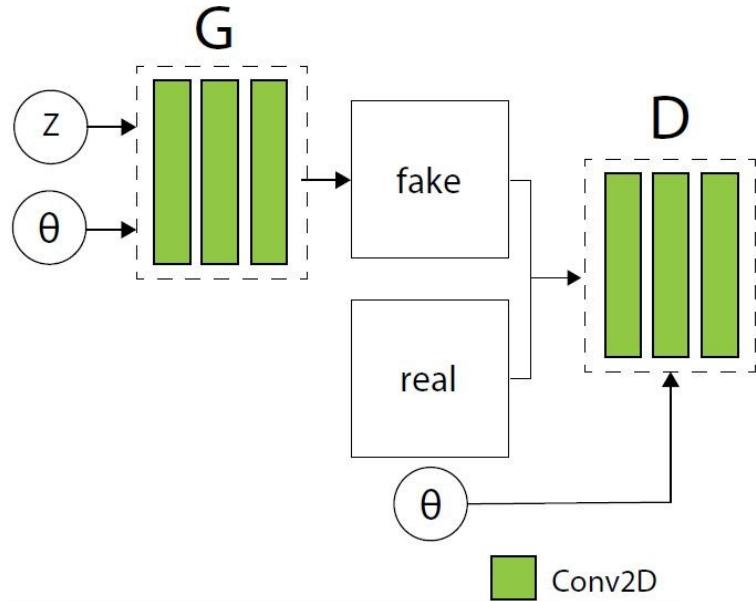


HoloGAN: Unsupervised learning of 3D representations from natural images

Thu Nguyen-Phuoc et al, ICCV 2019



Conditional GANs

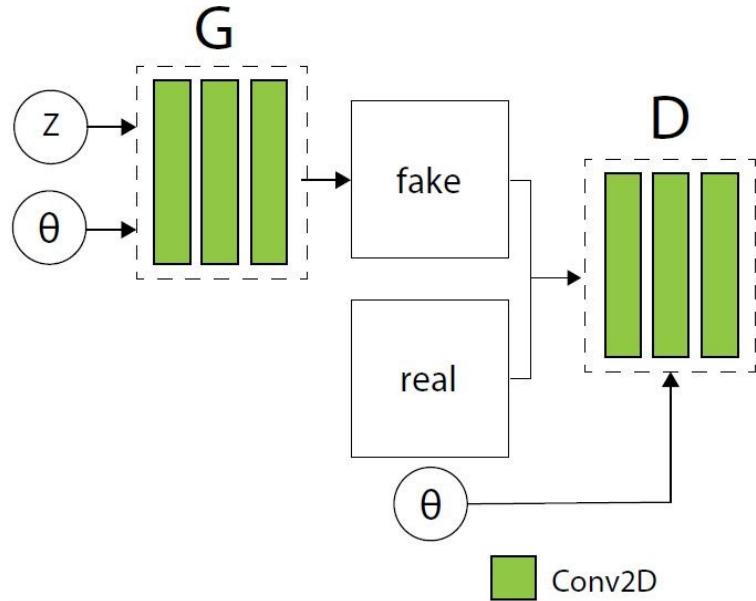


HoloGAN: Unsupervised learning of 3D representations from natural images

Thu Nguyen-Phuoc et al, ICCV 2019



Conditional GANs



Info GANs

HoloGAN: Unsupervised learning of 3D representations from natural images

Thu Nguyen-Phuoc et al, ICCV 2019





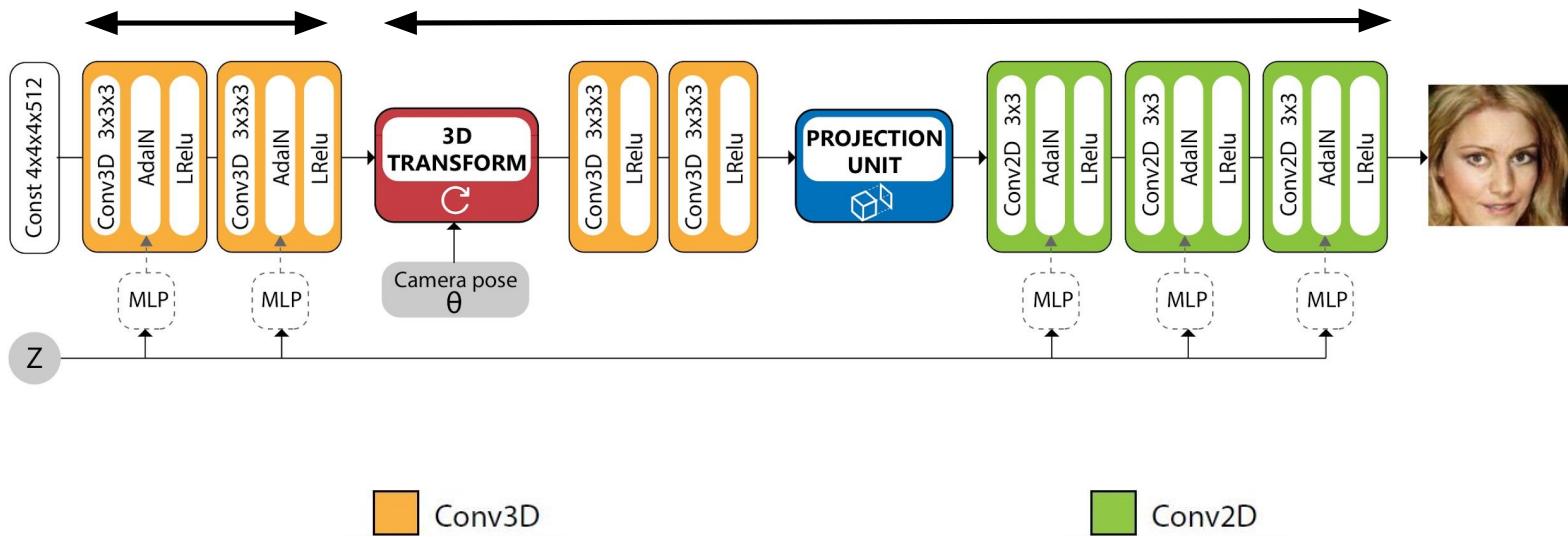
HoloGAN: Unsupervised learning of 3D representations from natural images

Thu Nguyen-Phuoc et al, ICCV 2019



3D Generator

RenderNet

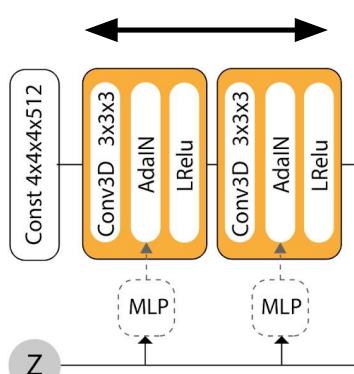


HoloGAN: Unsupervised learning of 3D representations from natural images

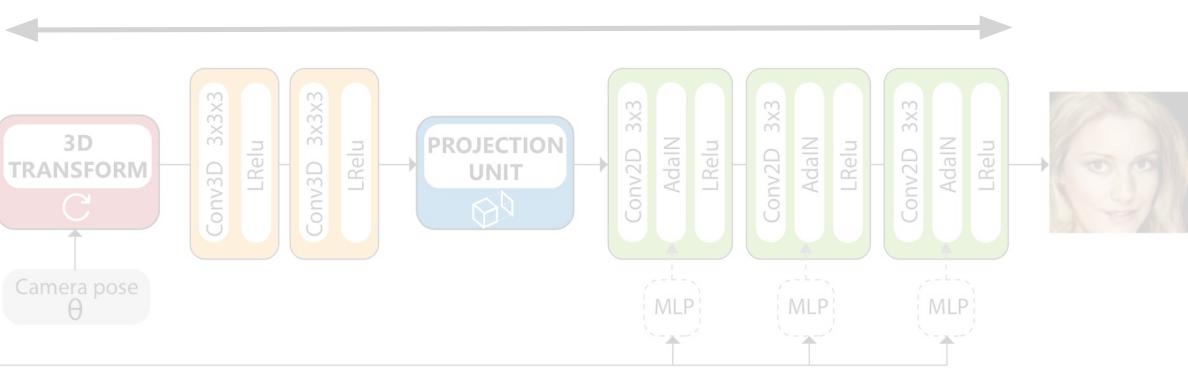
Thu Nguyen-Phuoc et al, ICCV 2019



3D Generator



RenderNet



3D StyleGAN

Conv3D

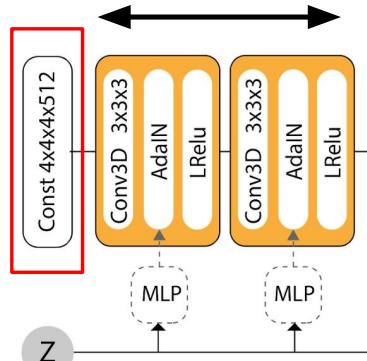
Conv2D

HoloGAN: Unsupervised learning of 3D representations from natural images

Thu Nguyen-Phuoc et al, ICCV 2019

λ Lambda

3D Generator



3D StyleGAN



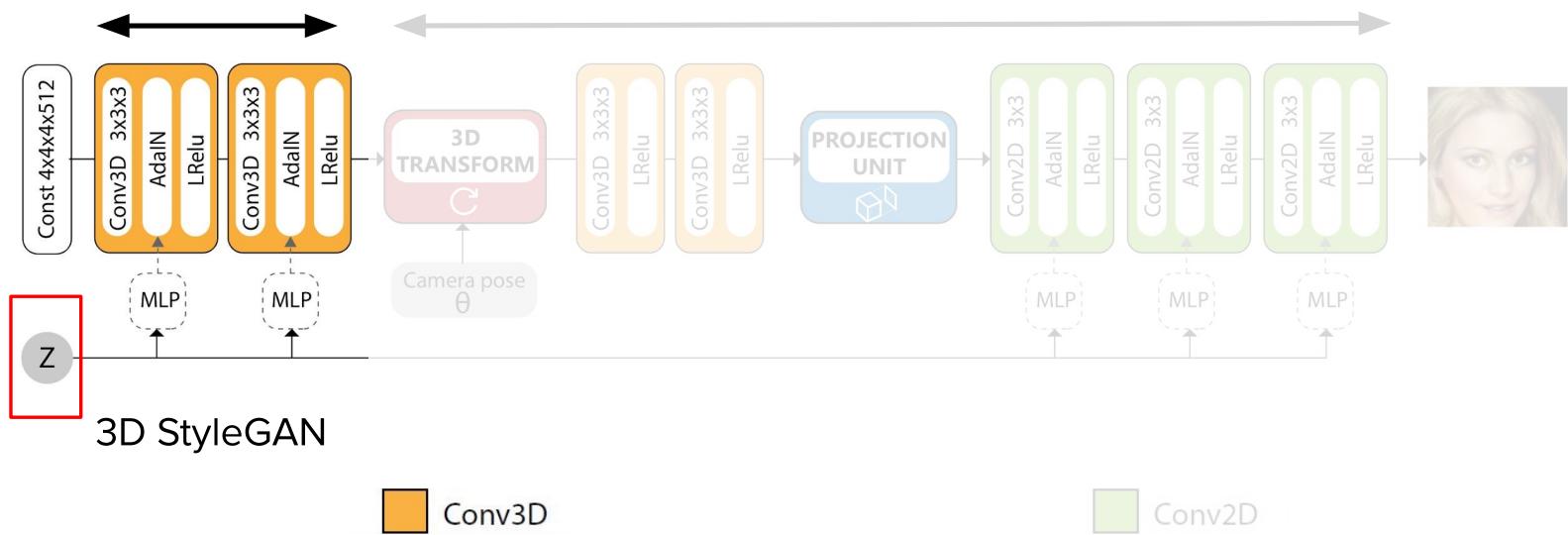
HoloGAN: Unsupervised learning of 3D representations from natural images

Thu Nguyen-Phuoc et al, ICCV 2019



3D Generator

RenderNet



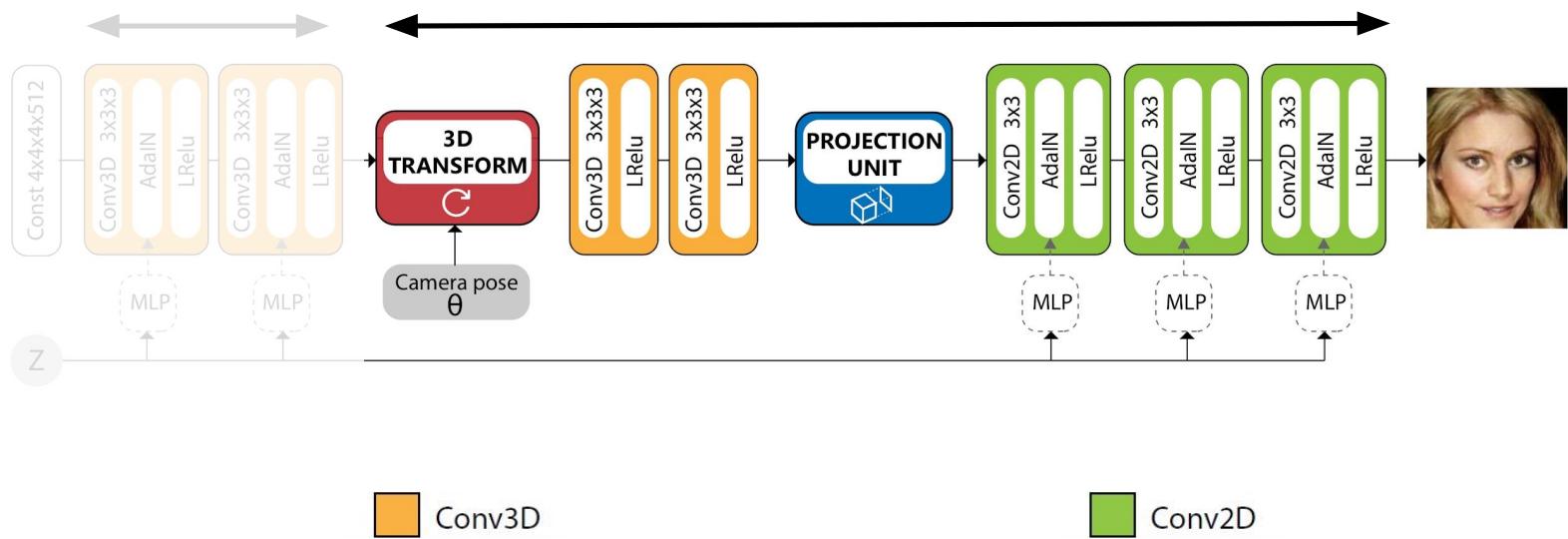
HoloGAN: Unsupervised learning of 3D representations from natural images

Thu Nguyen-Phuoc et al, ICCV 2019



3D Generator

RenderNet



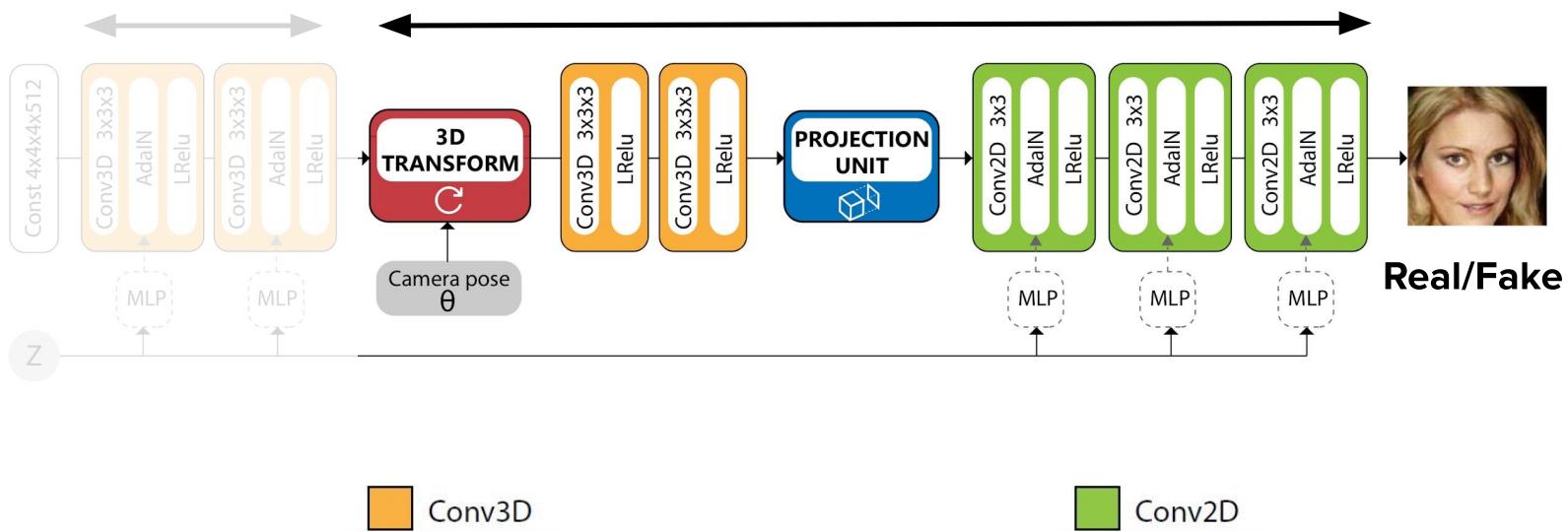
HoloGAN: Unsupervised learning of 3D representations from natural images

Thu Nguyen-Phuoc et al, ICCV 2019



3D Generator

RenderNet



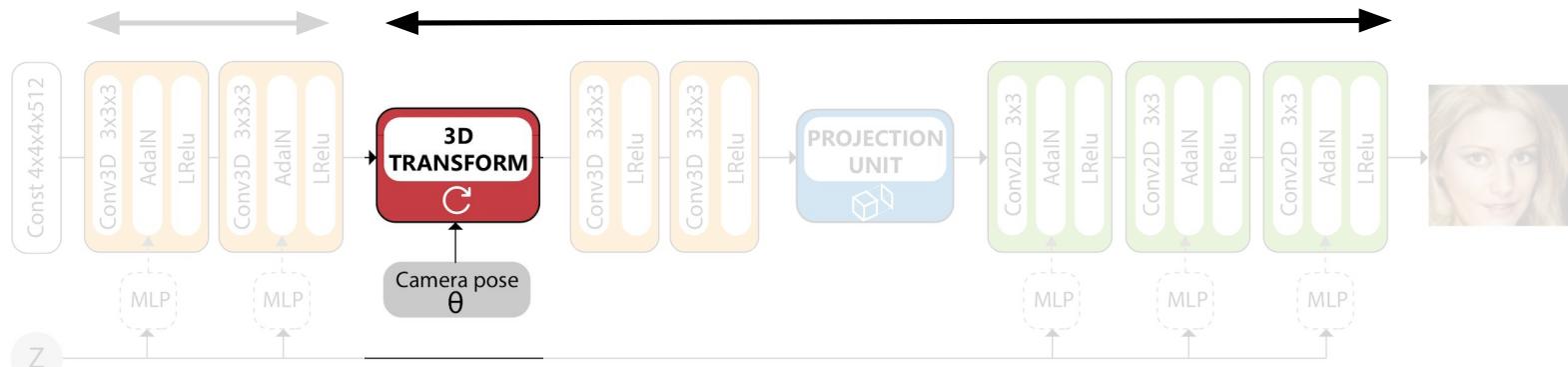
HoloGAN: Unsupervised learning of 3D representations from natural images

Thu Nguyen-Phuoc et al, ICCV 2019

λ Lambda

3D Generator

RenderNet



Conv3D

Conv2D

HoloGAN: Unsupervised learning of 3D representations from natural images

Thu Nguyen-Phuoc et al, ICCV 2019

λ Lambda



HoloGAN: Unsupervised learning of 3D representations from natural images

Thu Nguyen-Phuoc et al, ICCV 2019





HoloGAN: Unsupervised learning of 3D representations from natural images

Thu Nguyen-Phuoc et al, ICCV 2019

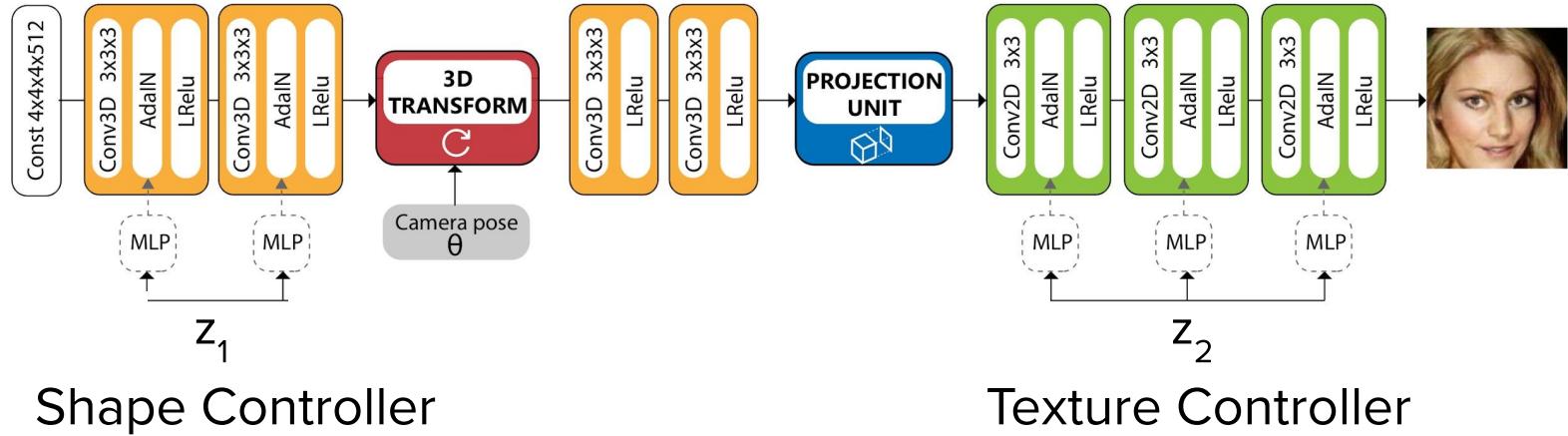




HoloGAN: Unsupervised learning of 3D representations from natural images

Thu Nguyen-Phuoc et al, ICCV 2019

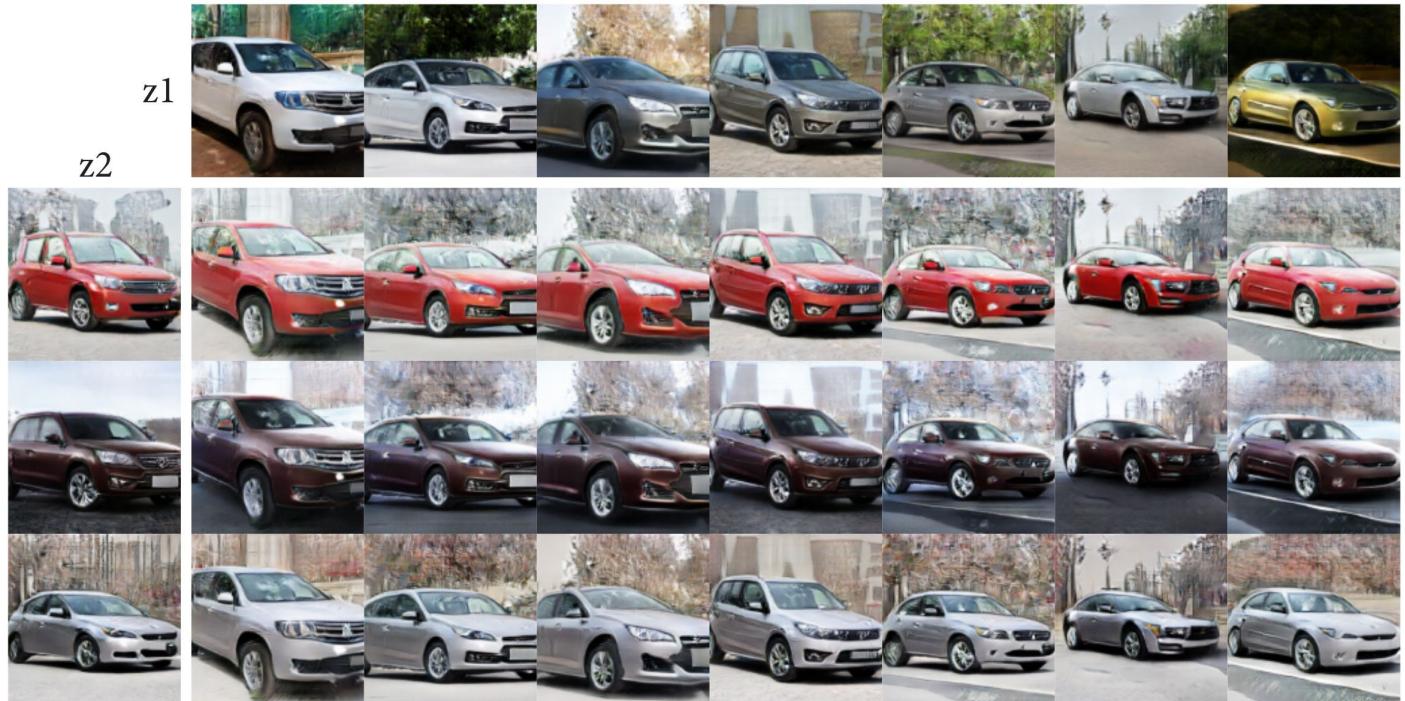




HoloGAN: Unsupervised learning of 3D representations from natural images

Thu Nguyen-Phuoc et al, ICCV 2019



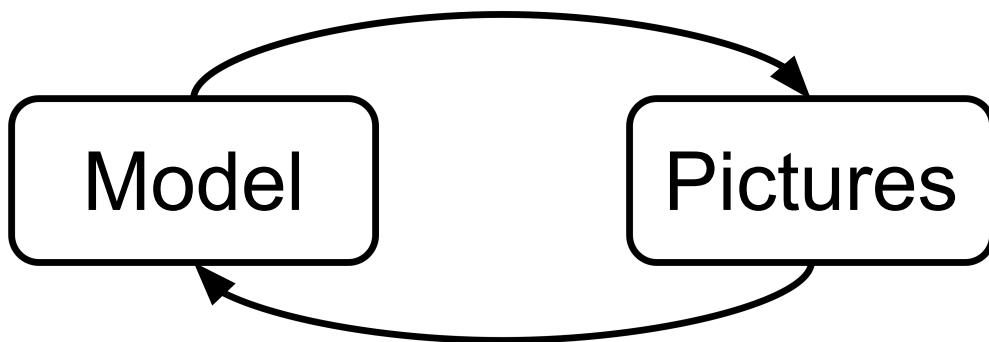


HoloGAN: Unsupervised learning of 3D representations from natural images

Thu Nguyen-Phuoc et al, ICCV 2019



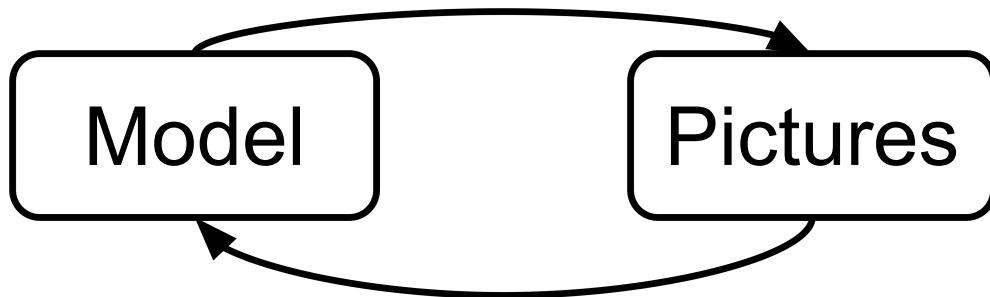
Forward (Computer Graphics)



Inverse (Computer Vision)

Sub-module for Ray Tracing (Value / Policy Networks)

End-2-End Rasterization (Depthmap, Voxel, Point Cloud, Mesh)



Differentiable Rendering (Representation Learning)



Thu Nguyen-Phuoc



Bing Xu



Yongliang Yang



Stephen Balaban

Lucas Theis

Christian Richardt

Junfei Zhang

Rui Wang

Kun Xu

Rui Tang



ML for Scent

Alex Wiltschko, Benjamin Sanchez-Lengeling, Brian Lee, Carey Radebaugh, Emily Reif, Jennifer Wei

Hi!



I'm **Alex Wiltschko**, a scientist at Google Research.
I lead a research group within **Google Brain**
that focuses on **machine learning for olfaction**.

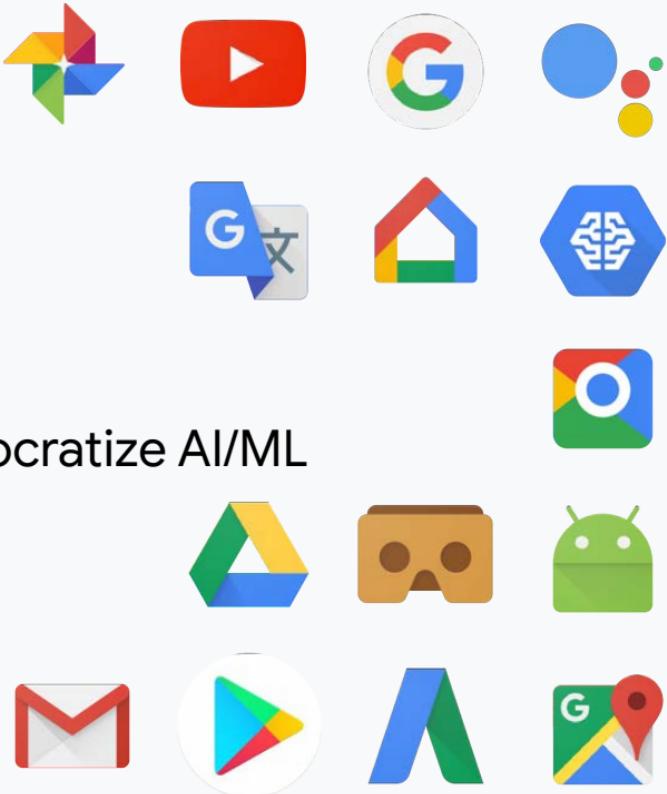
Google Research

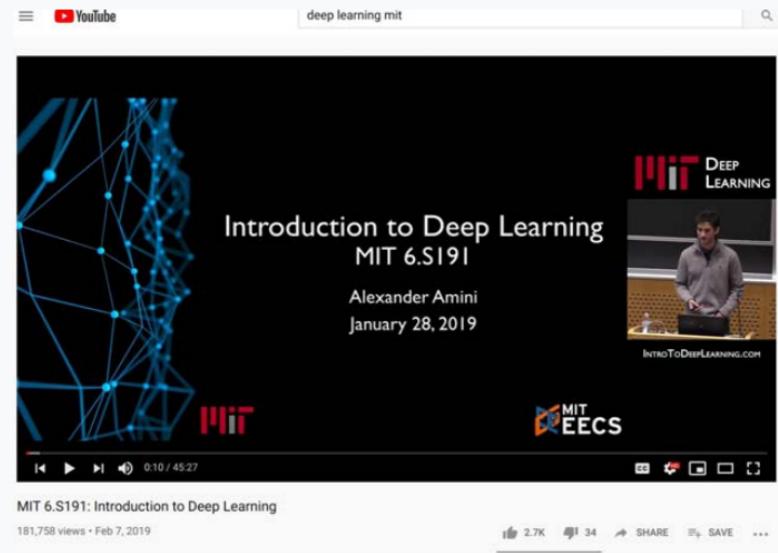
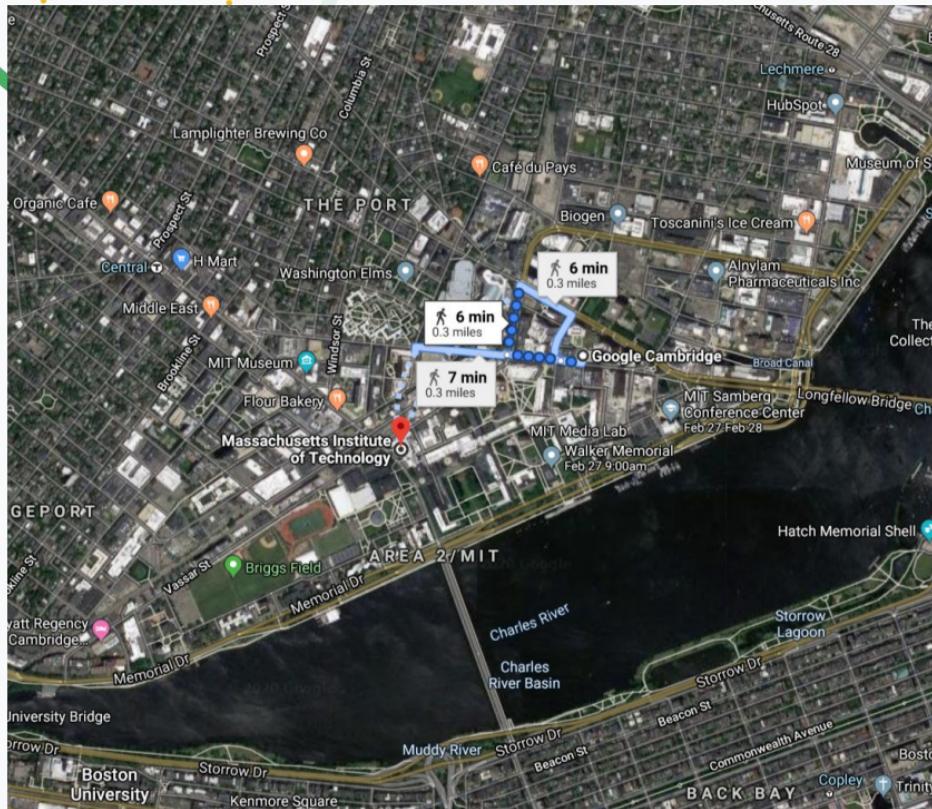


Make machines intelligent. Improve people's lives.

Our Approach

- Foundational research
- Building tools to enable research & democratize AI/ML
- AI-enabling Google products





YouTube

deep learning mit

MIT DEEP LEARNING

Introduction to Deep Learning
MIT 6.S191

Alexander Amini
January 28, 2019

INTROTODEEPLearning.COM

MIT EECS

0:10 / 45:27

MIT 6.S191: Introduction to Deep Learning
181,758 views • Feb 7, 2019

2.7K 34 SHARE SAVE

This image is a screenshot of a YouTube video player. The video is titled "Introduction to Deep Learning" and is part of the MIT 6.S191 course. It features a dark background with a blue neural network diagram on the left. On the right, there's a thumbnail of a man speaking at a podium. The MIT logo and "DEEP LEARNING" text are visible in the top right corner. The video has 181,758 views and was posted on February 7, 2019. The player interface includes standard controls like play/pause, volume, and a progress bar showing 0:10 of 45:27.

What's our goal?

Do for olfaction what machine learning has already done for vision and hearing.

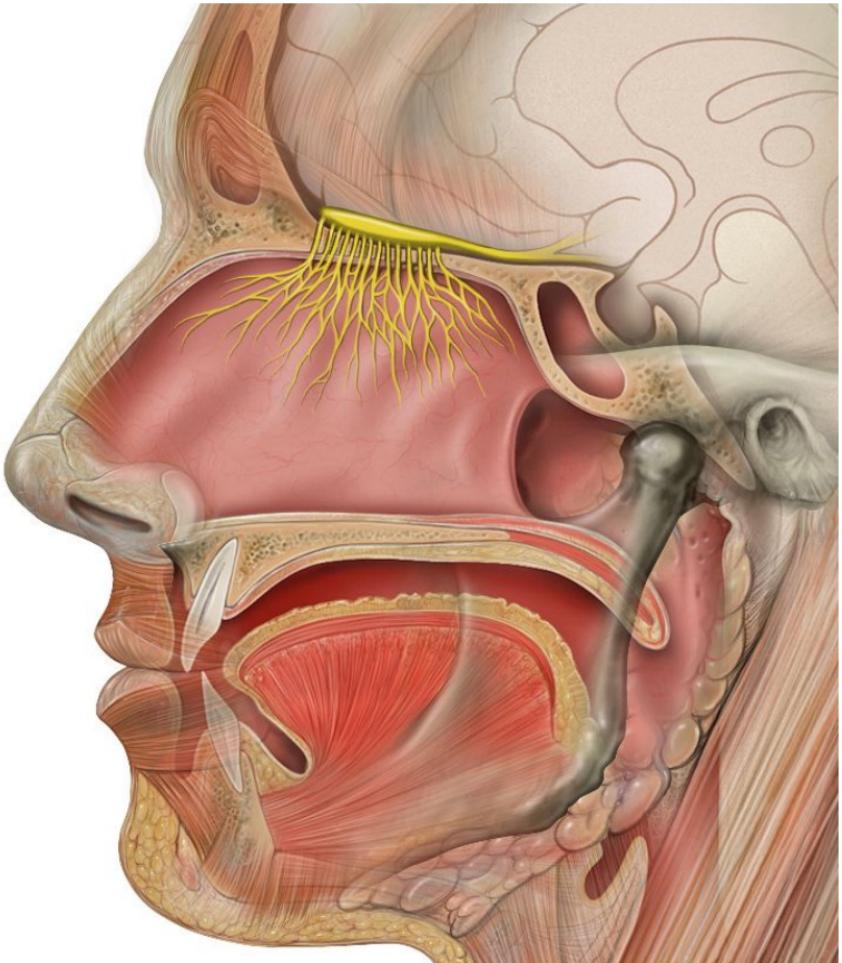
To **digitize the sense of smell**, and make the world's smells and flavors searchable. Every flower patch, every natural gas leak, every item on every menu in every restaurant.

We're starting at the very beginning, with the simplest problem... but first, some olfaction facts!

Most airflow is not smelled. Passes right on through the lower turbinates to your lungs.

The OSNs are one of two parts of your brain that are exposed to the world (the other is the pituitary gland, and that's in blood, so only half-counts).

Taste lives on your tongue. Flavor is both taste and retronasal olfaction, from a "chimney effect".



GPCR: G-protein coupled receptor
OR: GPCR Olfactory Receptor
OSN: Olfactory sensory neuron

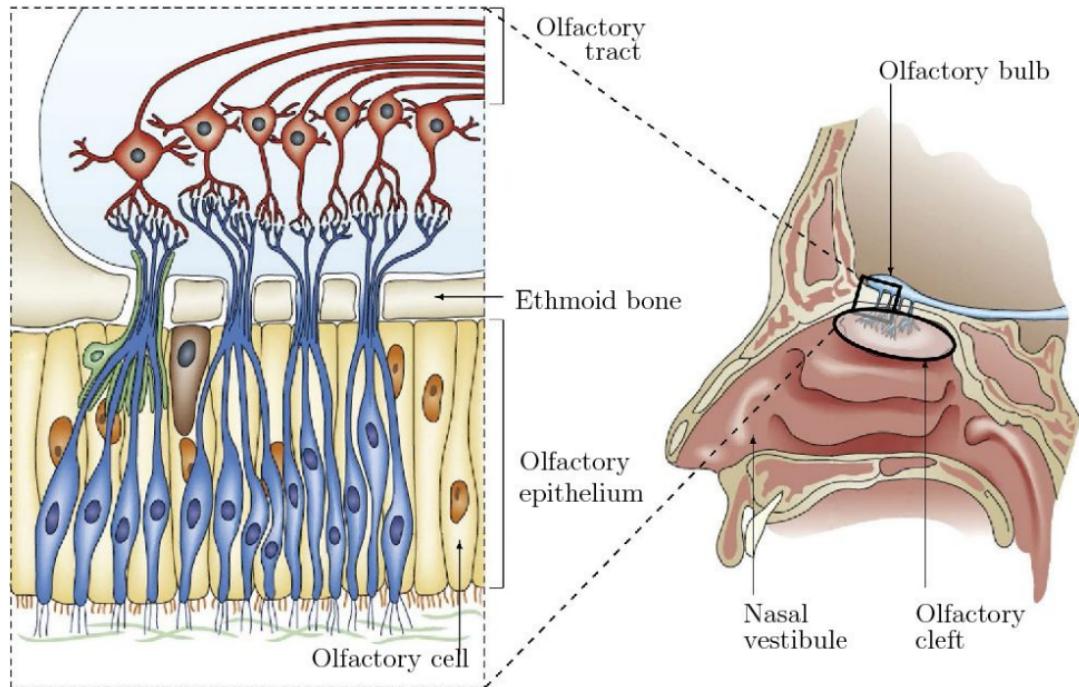
~400 ORs expressed in humans (as opposed to 3 types of cones)
~1000 in mice. ~2000 in elephants!

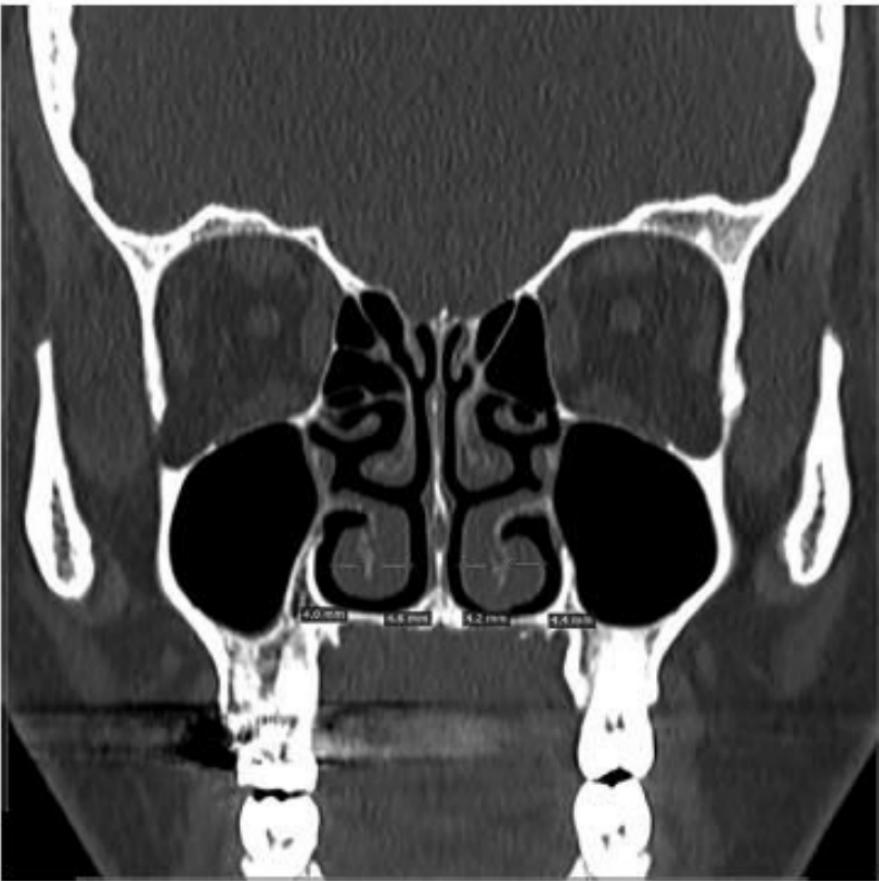
One OR per OSN.

ORs comprise 2% of your genome, but many are pseudogenes.

OR structure is unknown, they are uncrystallized. Further, only ~40 expressed in cell lines.

Their ligand responses are broadly tuned, but many ORs (22/400) are still orphans, with no known ligand.







People do smell different things!

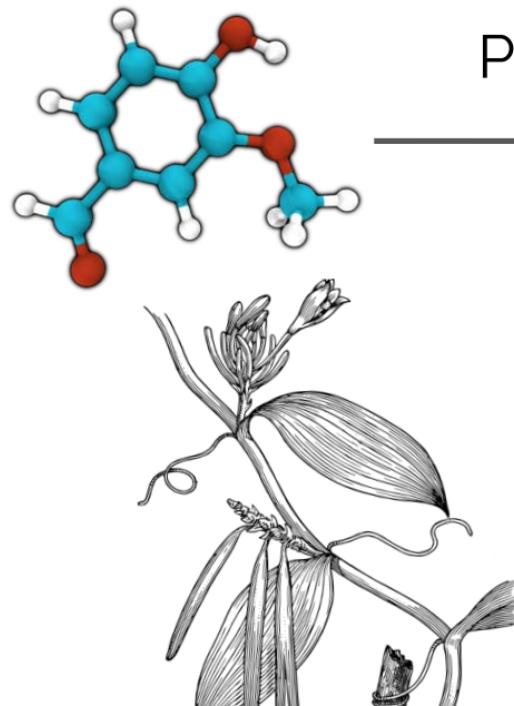
SNPs in single ORs result in sensory dimorphisms. The most famous ones are:

- OR7D4 T113M: normally funky beta-androstenone (boar taint) is rendered pleasant.
- OR5A1 N183D: nearly completely Mendelian. Carriers of the mutation can detect beta-ionine at two orders of magnitude lower concentration
- Olfactory sensory dimorphisms are likely common — humans differ functionally at 30% of OR alleles.
- ~4.5% of the world is colorblind ([CBA](#))
- 13% in the US has selective hearing loss ([NIDCD](#))
- All this to argue — smell is not defacto finicky or illogical.



Right now, we're starting with the ***simplest problem***

"Smells **sweet**, with a hint of **vanilla**, some notes of **creamy** and back note of **chocolate**."



Predict

citrus **creamy**

sweet baked spicy

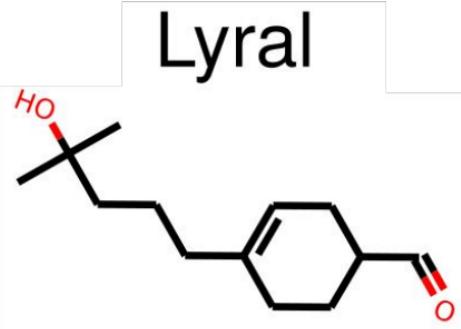
odorless **vanilla**

clean musky beefy

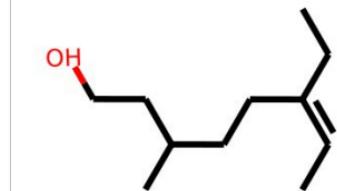
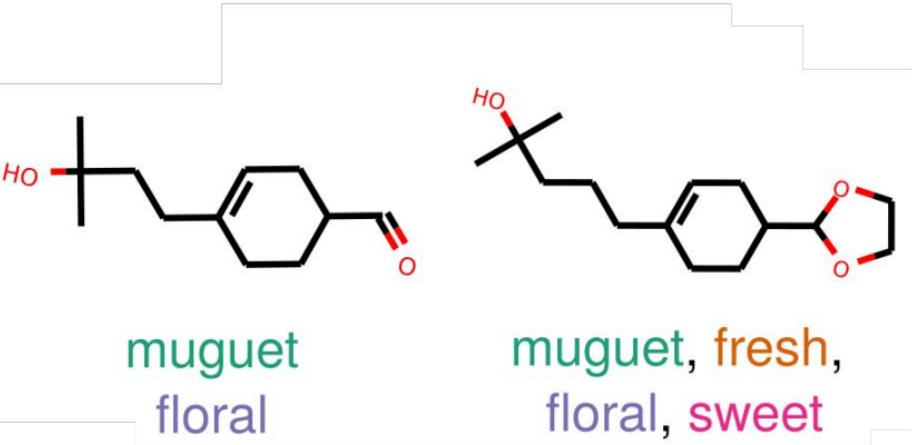
chocolate fruity

Odor
descriptors

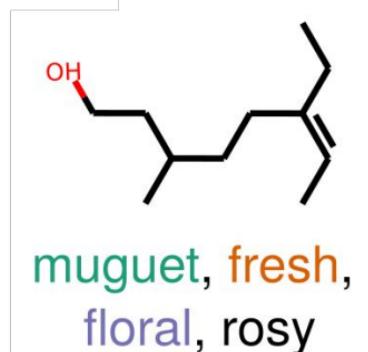
And why is this hard?



muguet, fresh,
floral, sweet

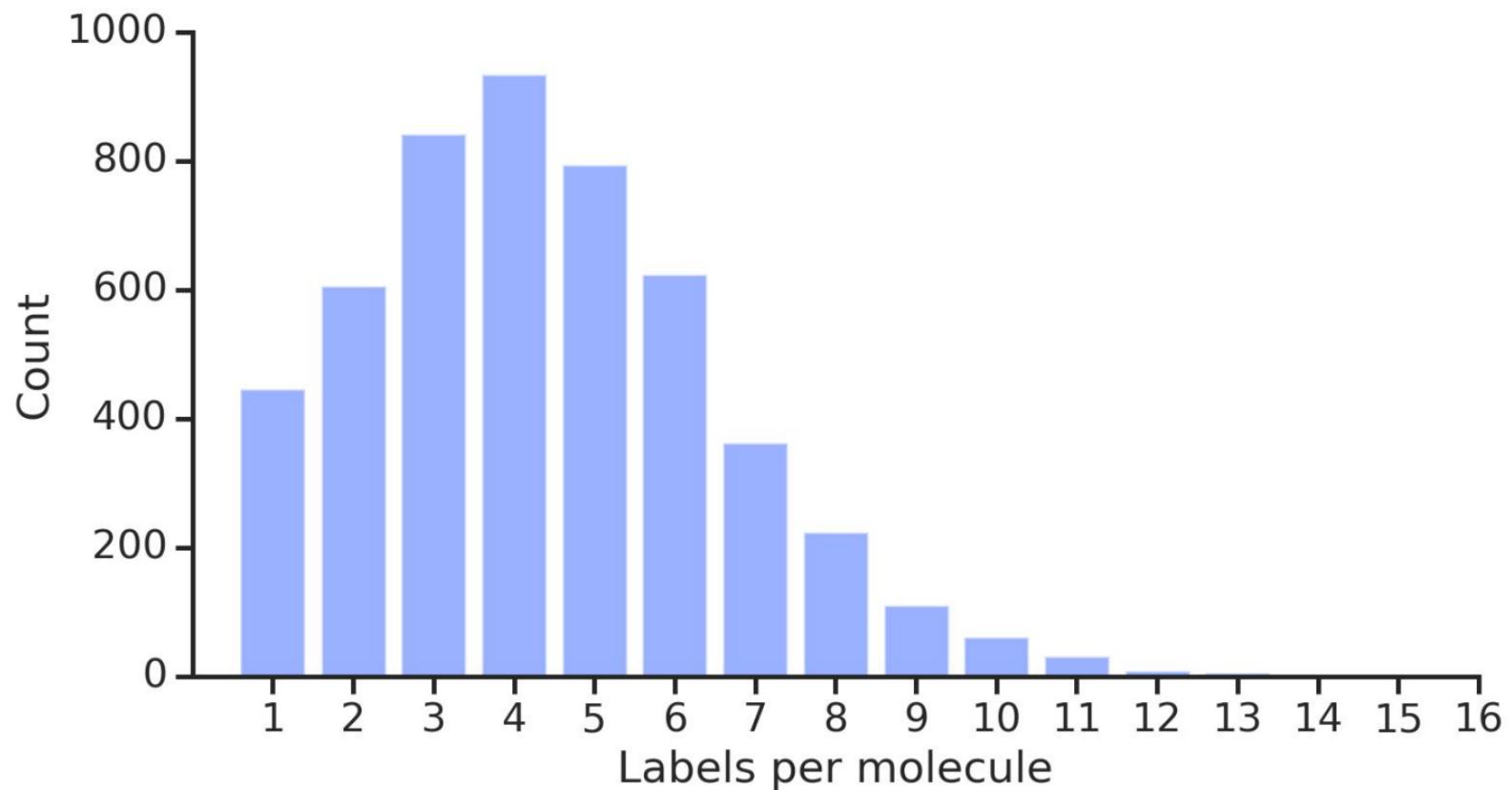


muguet, fresh,
floral, rosy

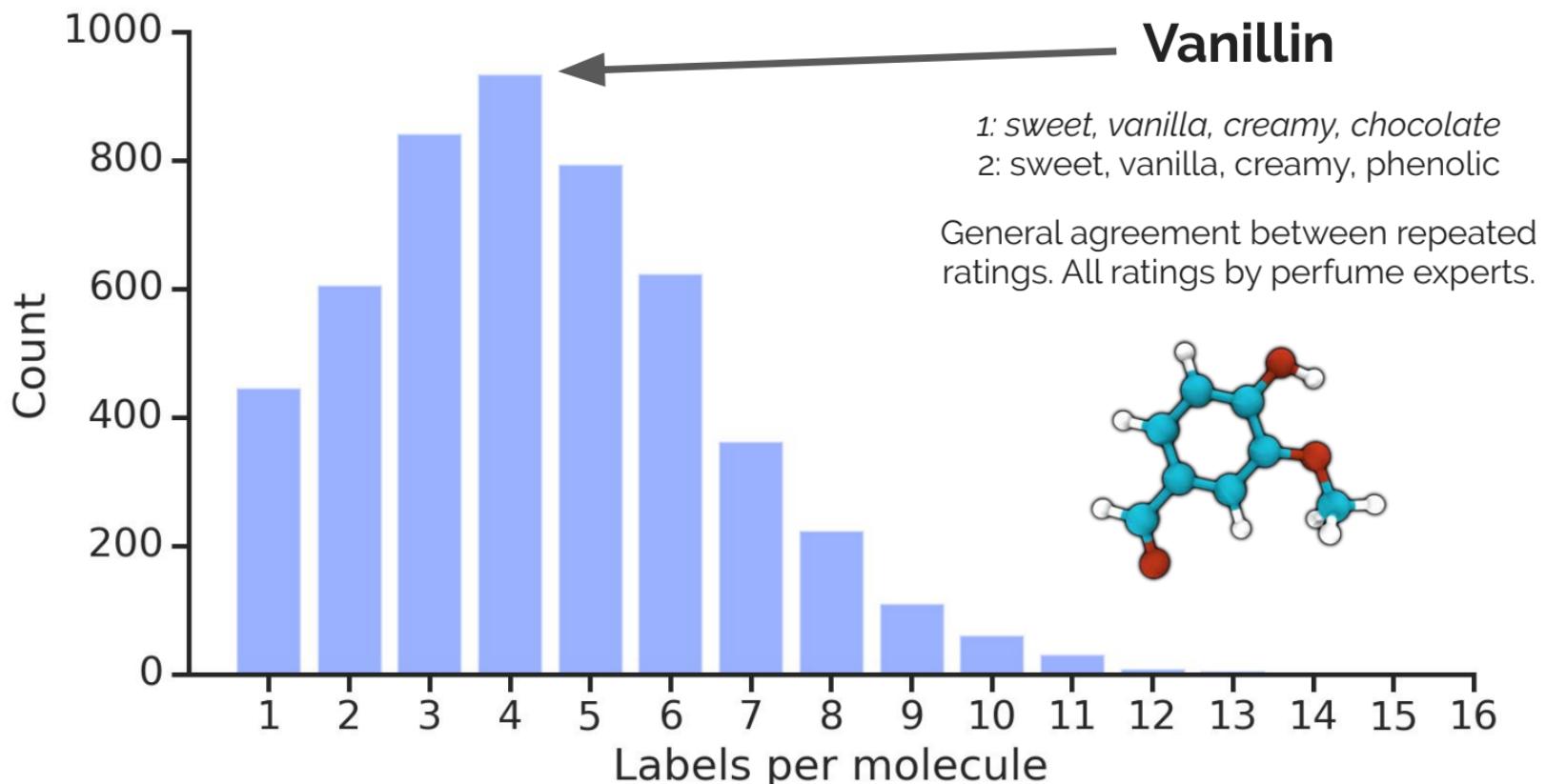


odorless

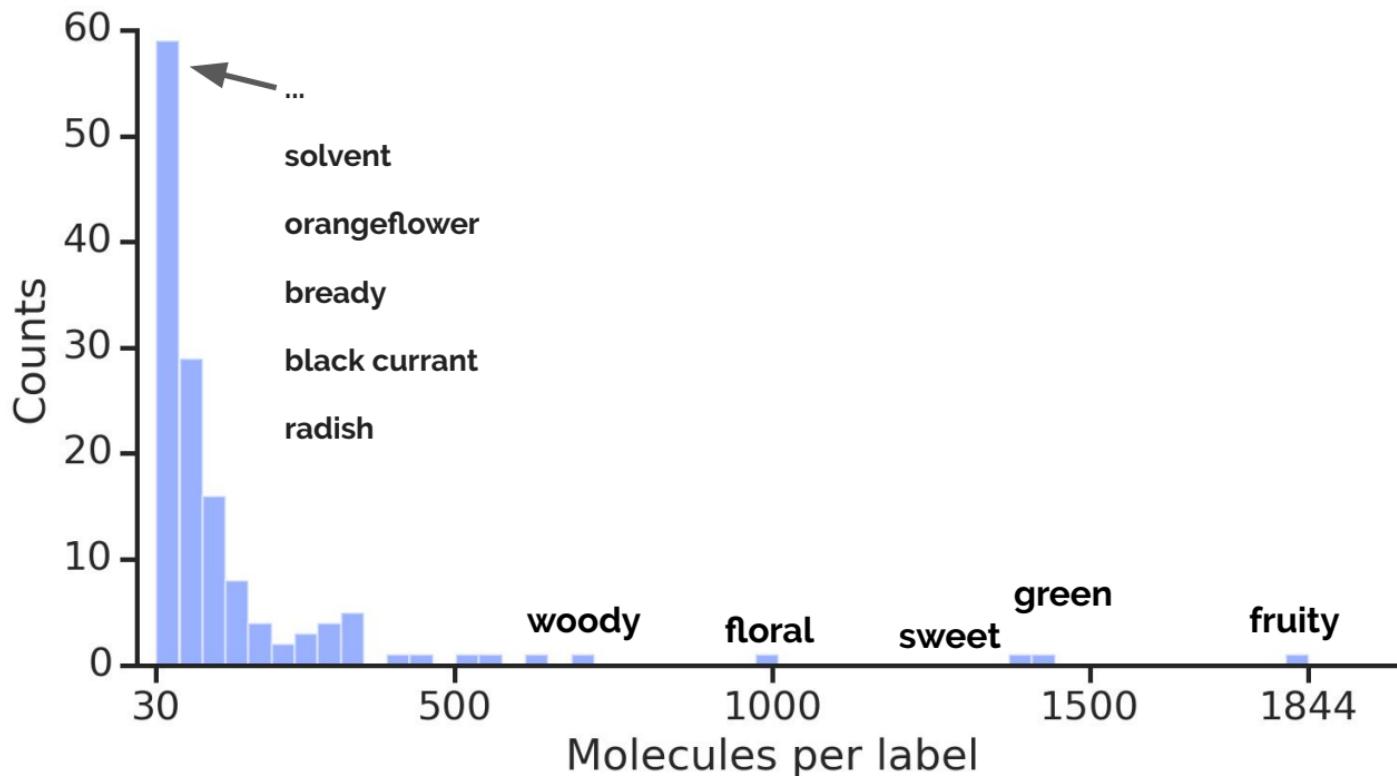
We built a benchmark from perfumery raw materials



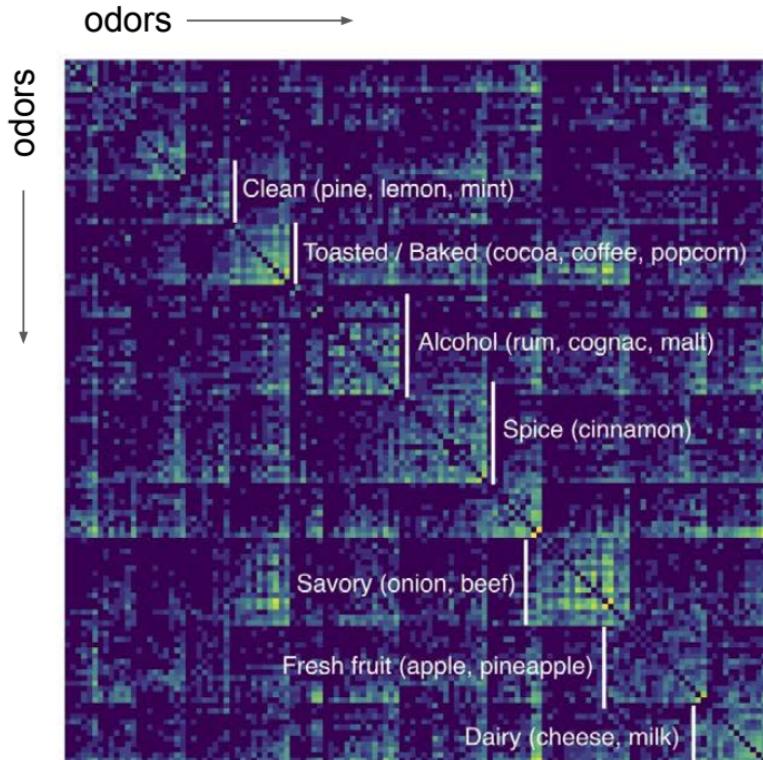
We built a benchmark from perfumery raw materials



We built a benchmark from perfumery raw materials



We built a benchmark from perfumery raw materials



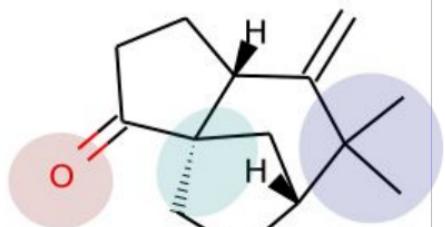
Historical SOR approaches

Pen & Paper

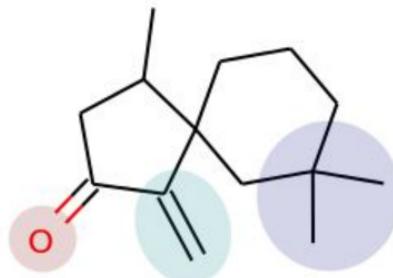
Ohloff's rule

Bajgrowicz and Broger's ambergris
osmophore model
Buchbauer's santalols
Boelens' synthetic muguet

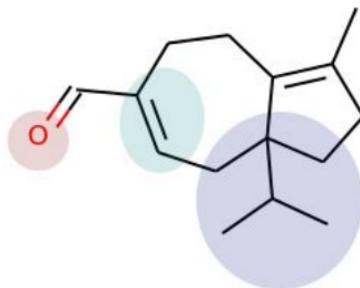
Kraft's vetiver rule



(-) -khusimone



4,7,7-Trimethyl-1-methylidene
spiro[4.5]decan-2-one

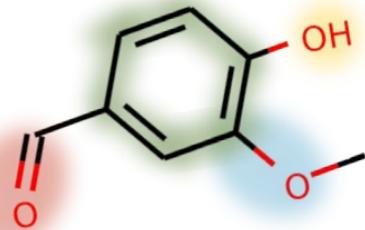
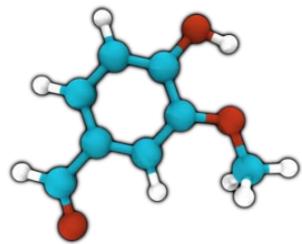


1,7-cyclogermacra-1
(10),4-dien-15-al

Fig 3.22 Scent and Chemistry (Ohloff, Pickenhagen, Kraft)

Rule-based principles for predicting odor. There are as many exceptions as there are rules.

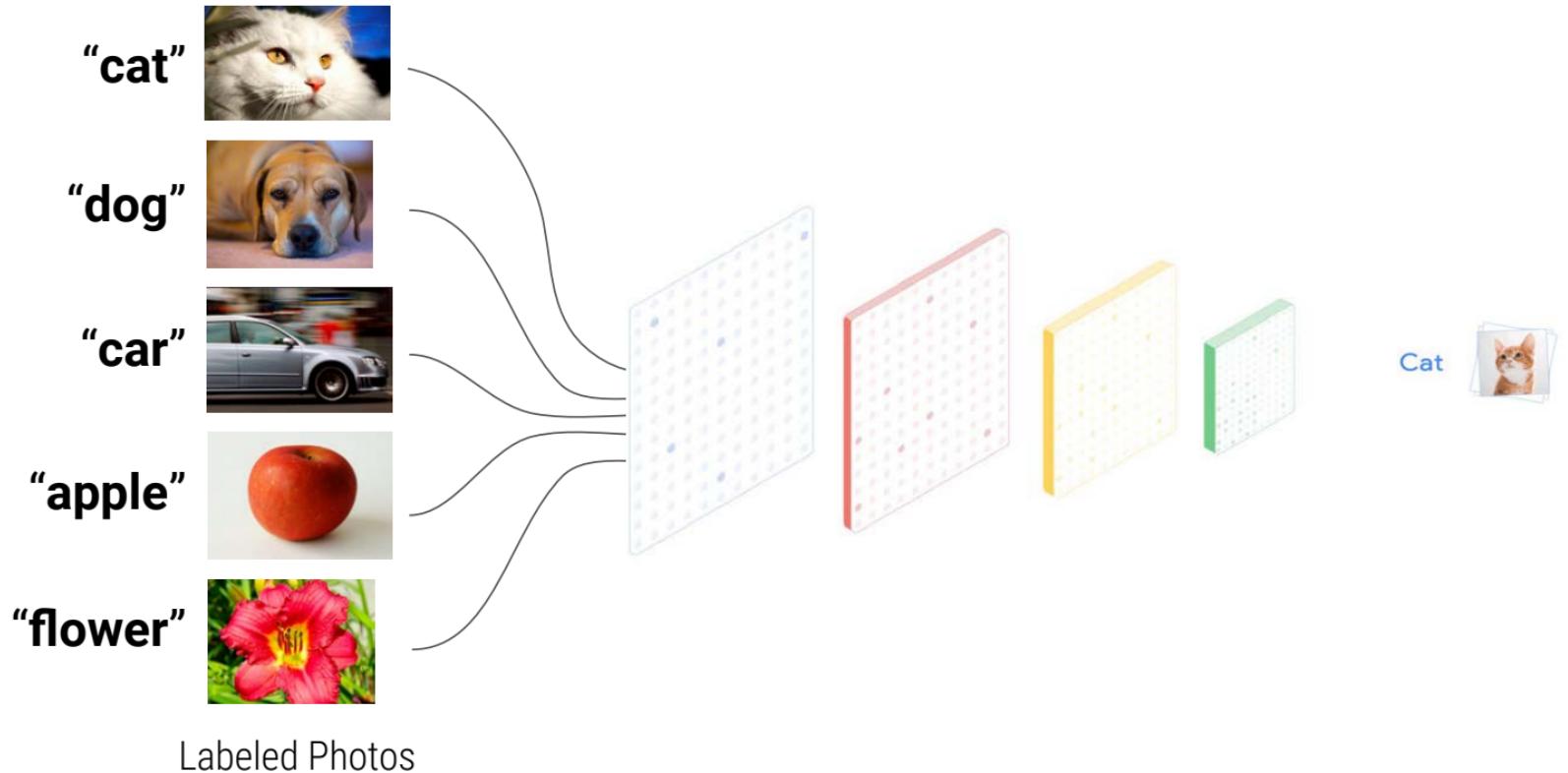
Traditional Computational Approaches



Predict

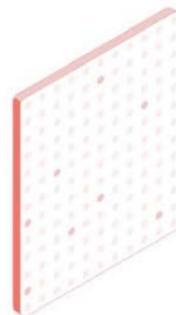
- Toxicity
- Solubility
- Photovoltaic efficiency (solar cell)
- Chemical potential (batteries)
- ...

"bag of sub-graphs" representation
AKA molecular fingerprints





Unlabeled Photo



Dog



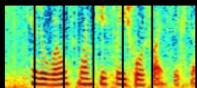
Input

Output



| PIXELS

"lion"



| AUDIO

"How cold is it outside?"

"Hello,
how are
you?"

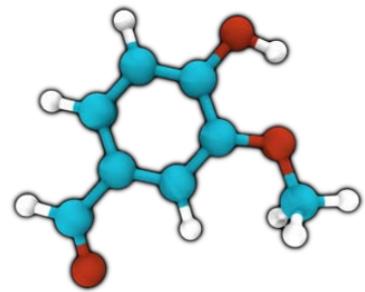
| TEXT

"你好， 你好吗？"



| PIXELS

"A blue and yellow train travelling down the tracks"



Graphs as input to neural networks: not just images, sounds or words

Convolutional Networks on Graphs for Learning Molecular Fingerprints

David Duvenaud[†], Dougal Maclaurin[†], Jorge Aguilera-Iparraguirre
Rafael Gómez-Bombarelli, Timothy Hirzel, Alán Aspuru-Guzik, Ryan P. Adams
Harvard University

Abstract

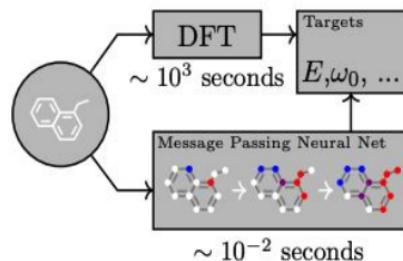
We introduce a convolutional neural network that operates directly on graphs. These networks allow end-to-end learning of prediction pipelines whose inputs are graphs of arbitrary size and shape. The architecture we present generalizes standard molecular feature extraction methods based on circular fingerprints. We show that these data-driven features are more interpretable, and have better predictive performance on a variety of tasks.

Neural Message Passing for Quantum Chemistry

Justin Gilmer¹ Samuel S. Schoenholz¹ Patrick F. Riley² Oriol Vinyals³ George E. Dahl¹

Abstract

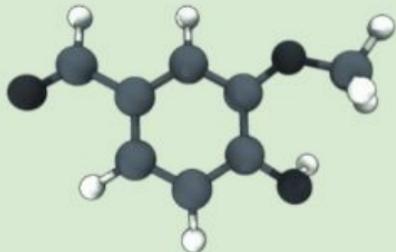
Supervised learning on molecules has incredible potential to be useful in chemistry, drug discovery, and materials science. Luckily, several promising and closely related neural network models invariant to molecular symmetries have already been described in the literature. These models learn a message passing algorithm and aggregation procedure to compute a function of their entire input graph. At this point, the next step is to find a particularly effective variant of



Inside a GNN

Converting a molecule to a graph

Molecule (e.g., vanillin)

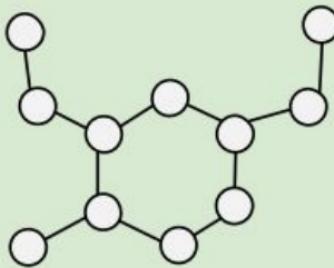
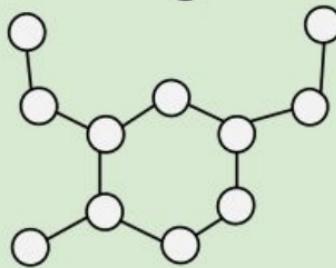


Inside a GNN

Propagating information & transforming a graph

Layer N → Layer N+1

For each  :

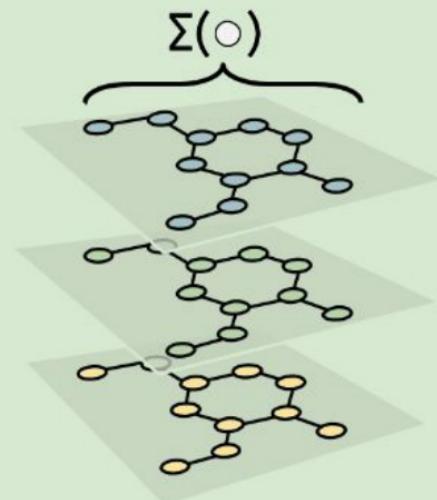
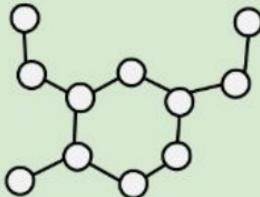
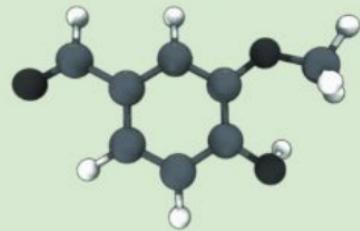


A GNN to predict odor descriptors

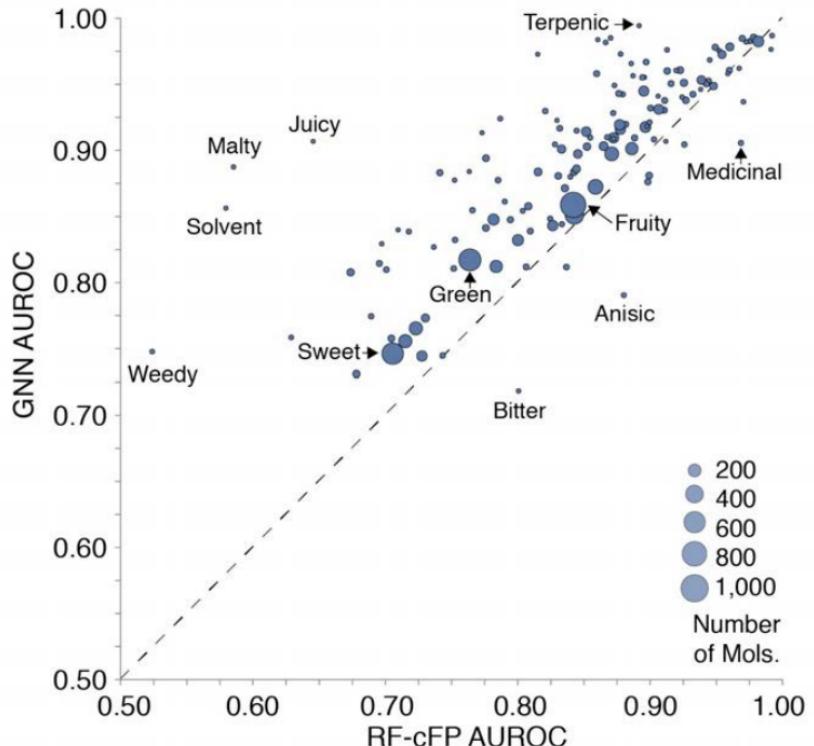
*citrus creamy sweet baked spicy odorless
vanilla clean alcoholic beefy chocolate fruity*



Molecule (e.g., vanillin)



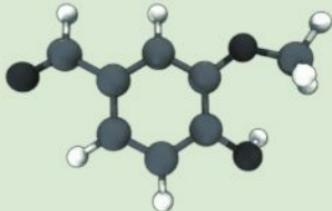
And how well can we predict?



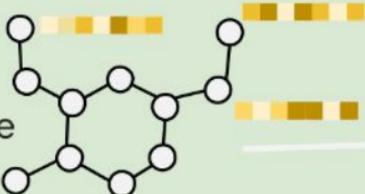
A representation optimized for odor

Last layer embeddings
63 dimension vector

Molecule (e.g., vanillin)



Prepare a graph with node information



citrus creamy sweet baked spicy odorless
vanilla clean alcoholic beefy chocolate fruity

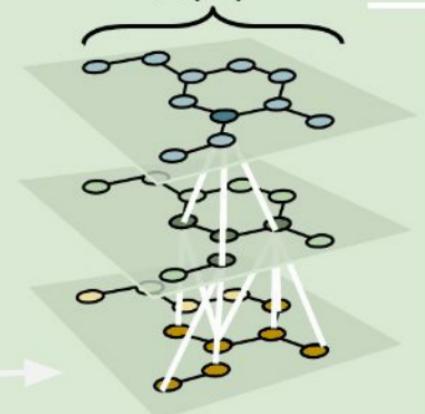


NN for final prediction



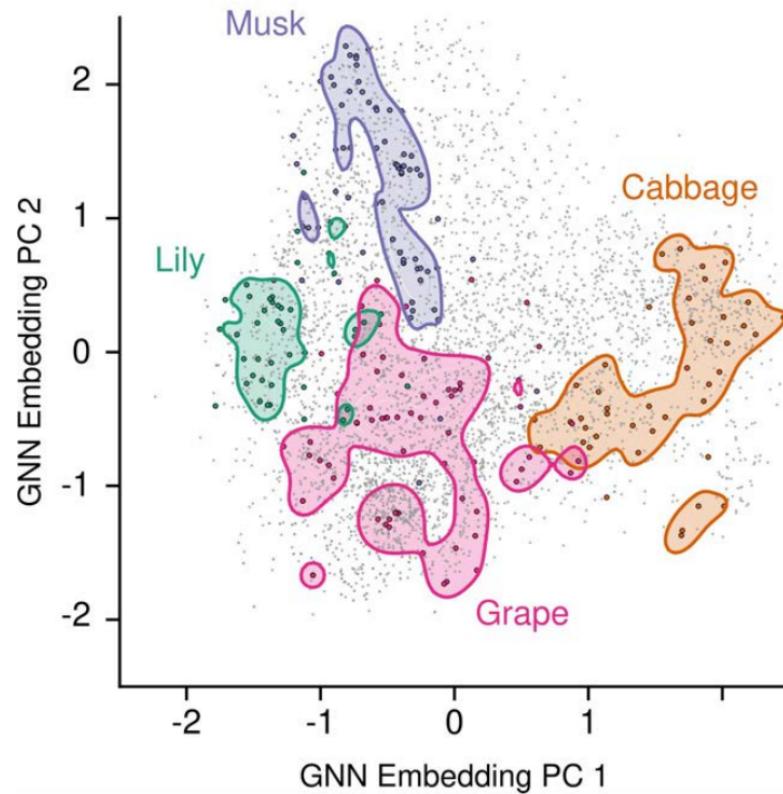
Reduce the graph to a vector (e.g., sum nodes)

$$\Sigma(\circ)$$

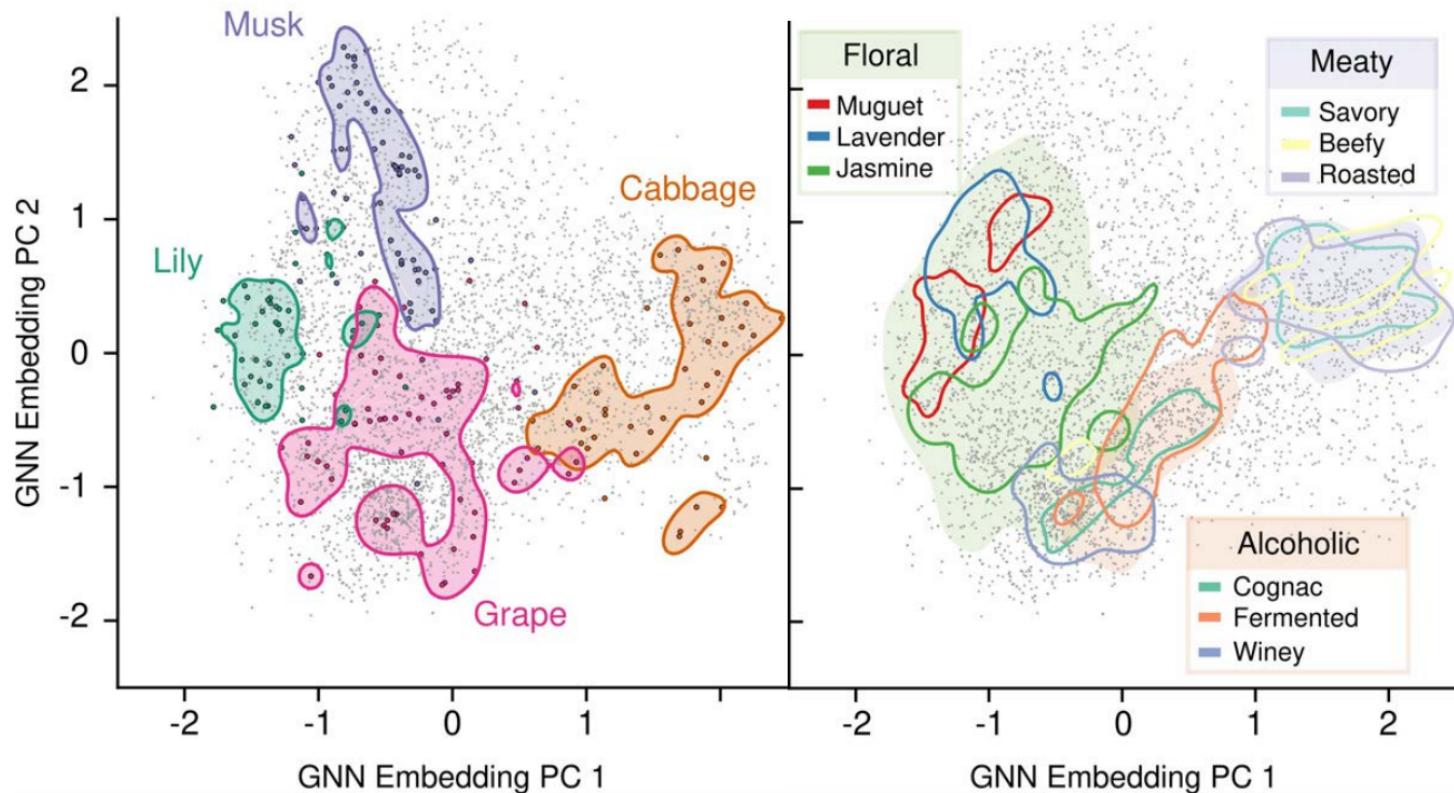


Learn a new representation of nodes

Exploring the geometric space of odor



Exploring the geometric space of odor



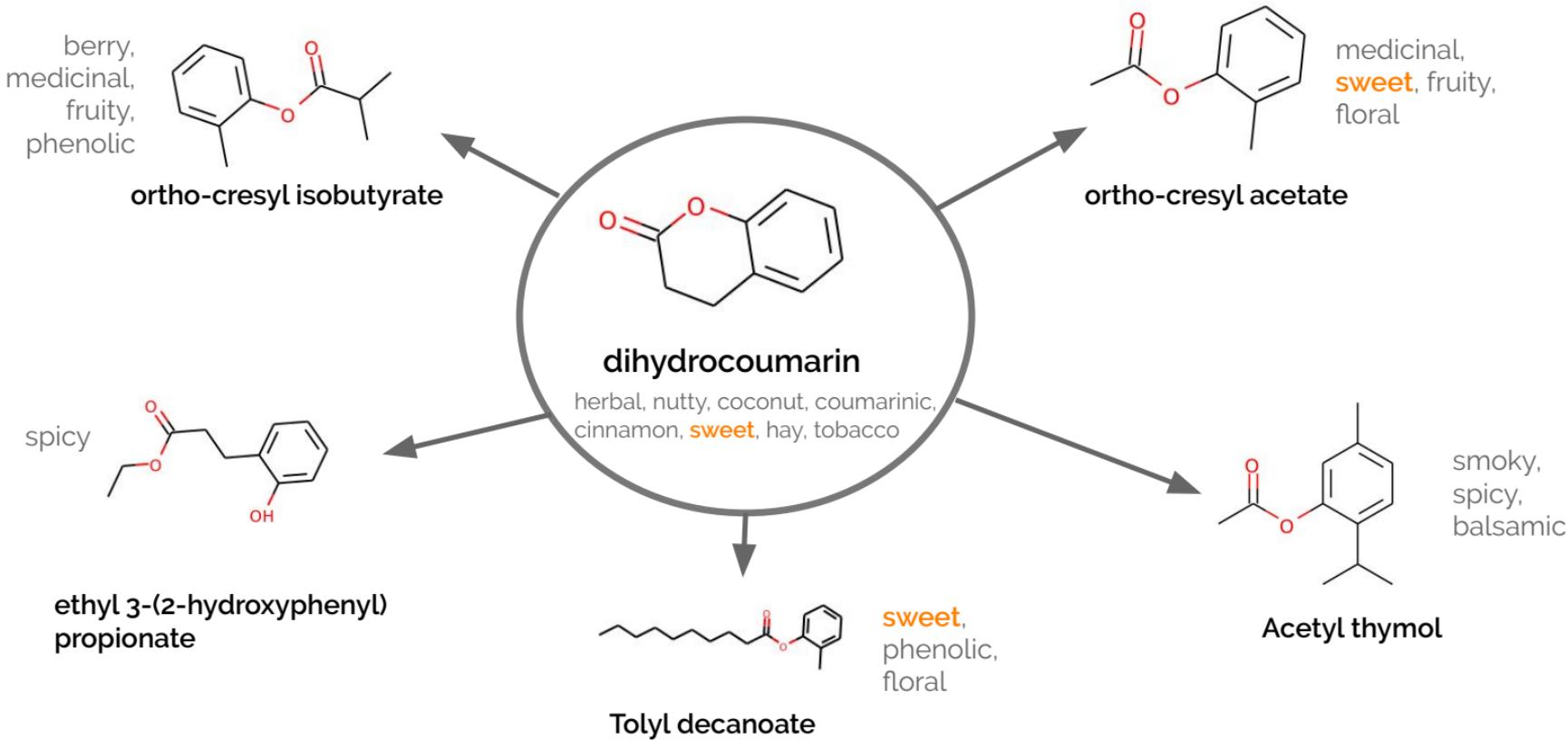
What do nearby molecules look like?

Inspired by word embeddings. Are there “molecular synonyms”?

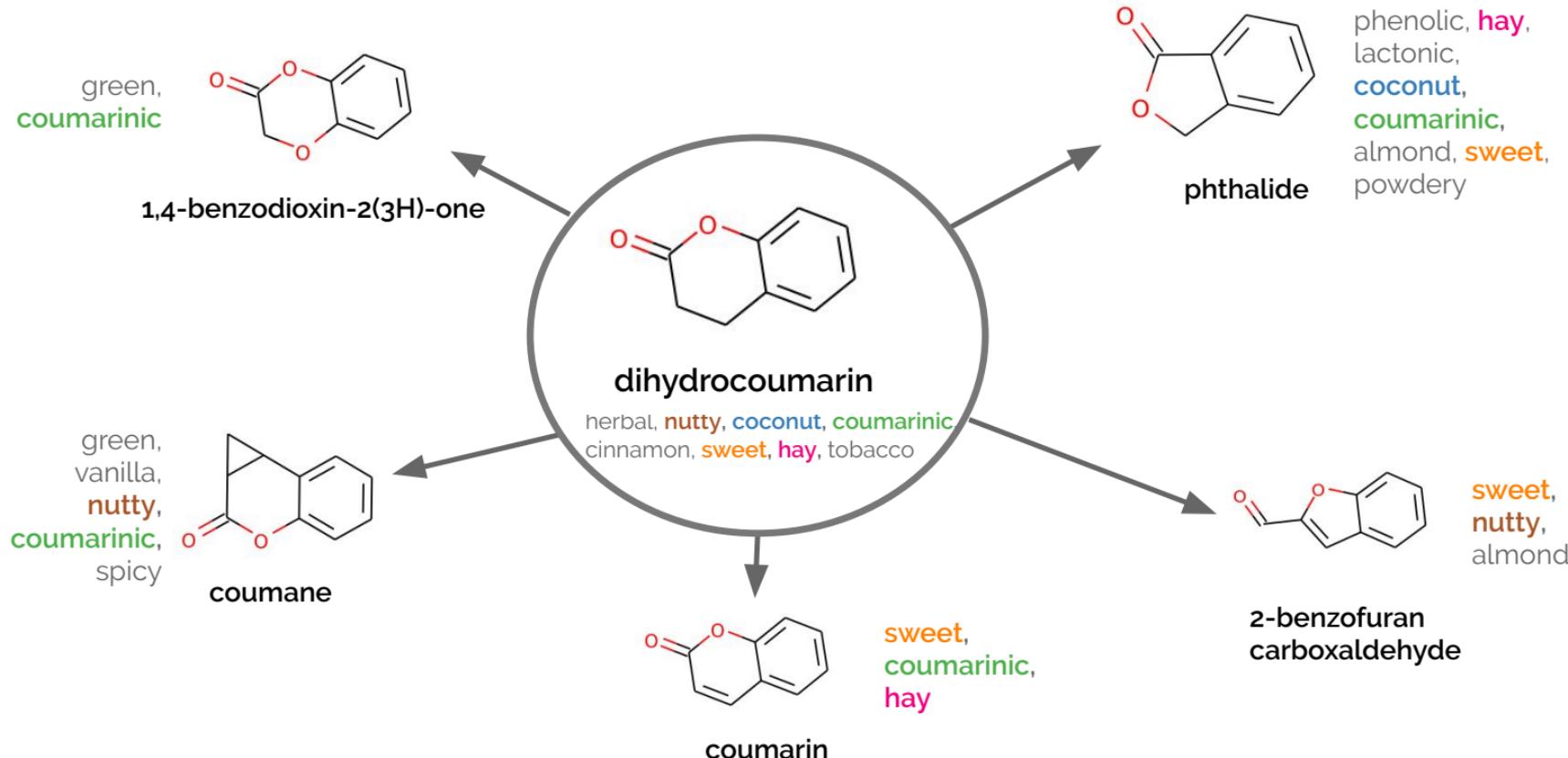
First, what do “nearest neighbors” look like if you use just structure, and ignore our neural network?

Then, what do nearest neighbors look like to our GCN?

Molecular neighbors: using structure



Molecular neighbors: using GCN features



Do these representations generalize?

Using a learned model to make predictions on a new task is ‘transfer learning’

You might hear ‘fine-tuning’ referred to as a strategy for ‘transfer learning’.

Transfer learning in chemistry, today, rarely works. Do our embeddings transfer learn to other tasks?

Do these representations generalize?



IN CS, IT CAN BE HARD TO EXPLAIN
THE DIFFERENCE BETWEEN THE EASY
AND THE VIRTUALLY IMPOSSIBLE.



PARK or BIRD

Want to know if your photo is from a U.S. national park? Want to know if it contains a bird? Just drag it into the box to the left, and we'll tell you. We'll use the GPS embedded in your photo (if it's there) to see whether it's from a park, and we'll use our super-cool computer vision skills to try to see whether it's a bird (which is a hard problem, but we do a pretty good job at it).

To try it out, just drag any photo from your desktop into the upload box, or try dragging any of our example images. We'll give you your answers below:

Want to know more about PARK or BIRD, including why the heck we did this? Just click here for more info → [?](#)

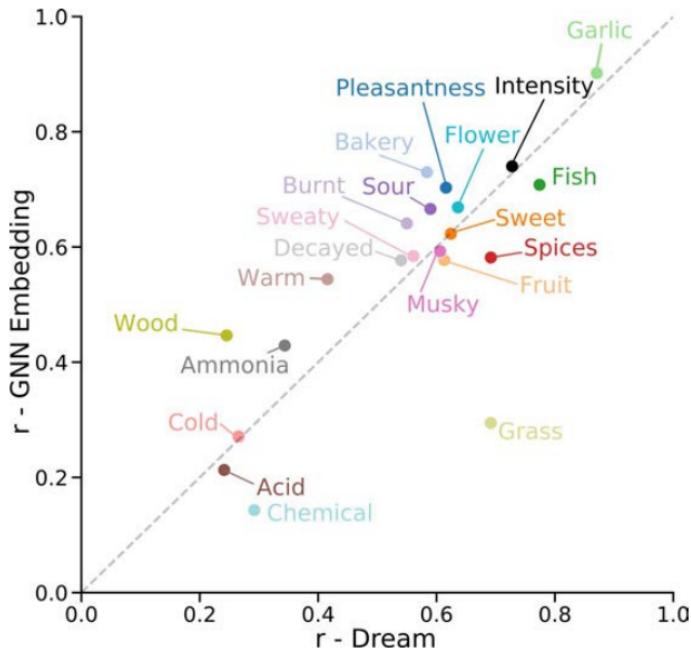
PARK?
???

No idea. There's no GPS info in that photo.

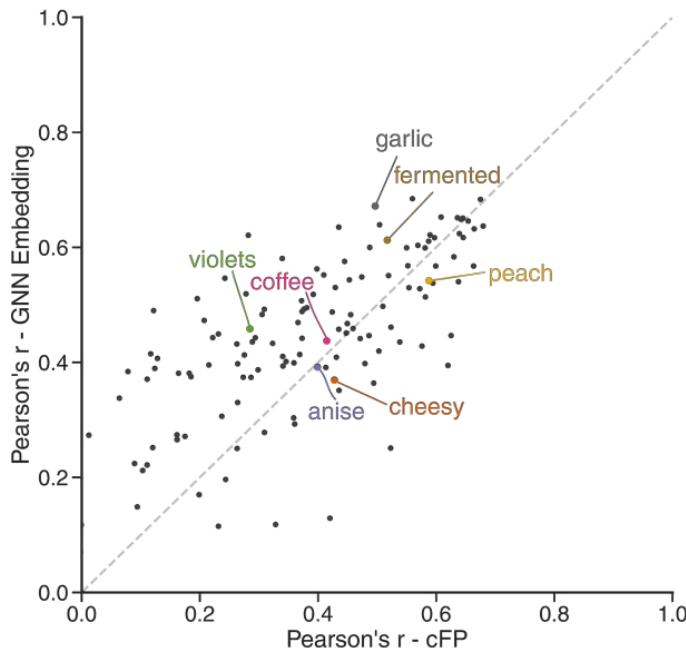
BIRD?
YES

Dude, that is such a bird.

DREAM Olfactory Challenge



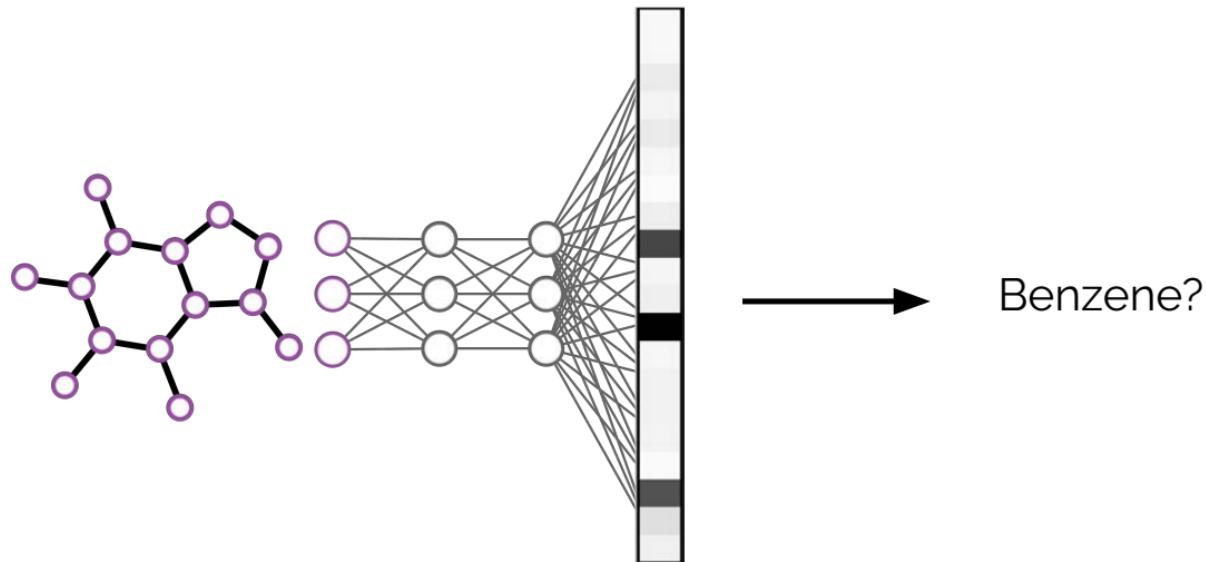
Dravnieks



Transfer-learned to achieve state-of-the-art on the two major olfactory benchmark tasks

But *why* is the neural network making these predictions?

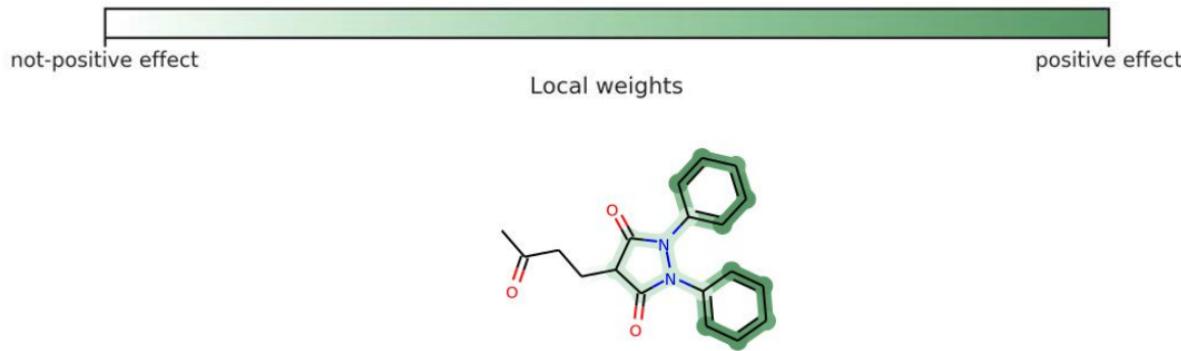
Toy test example: classify whether a molecule has benzene. Which atoms contribute to predictions?



This is just one task of potentially hundreds, of varying complexity.

But why is the neural network making these predictions?

Toy test example: classify whether a molecule has benzene. Which atoms contribute to predictions?



But why is the neural network making these predictions?

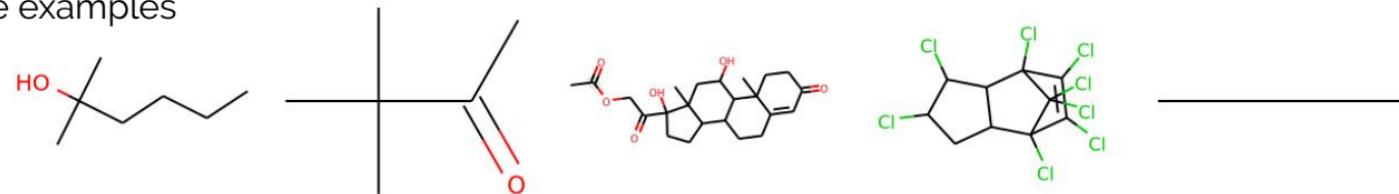
Toy test example: classify whether a molecule has benzene. Which atoms contribute to predictions?



Positive examples

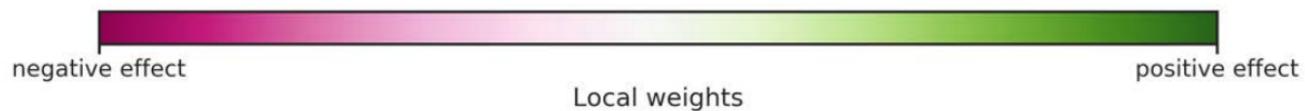


Negative examples



But why is the neural network making these predictions?

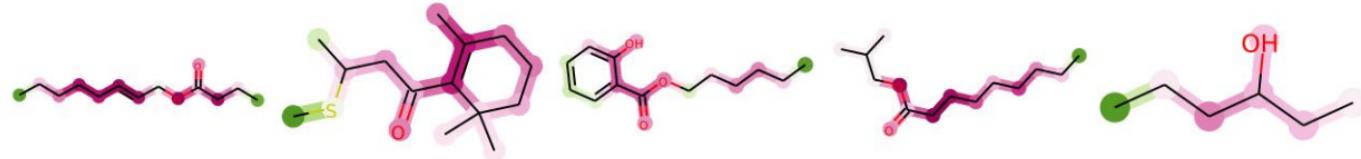
Odor percept – “garlic”



Positive examples



Negative examples



But why is the neural network making these predictions?

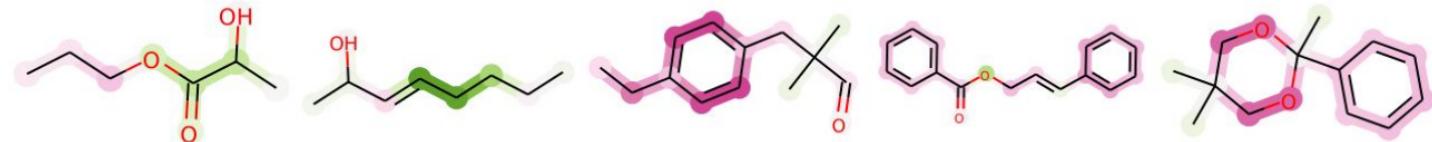
Odor percept – “fatty”



Positive examples

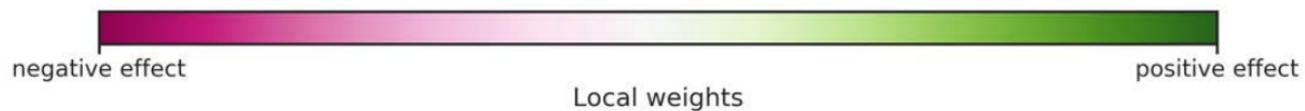


Negative examples

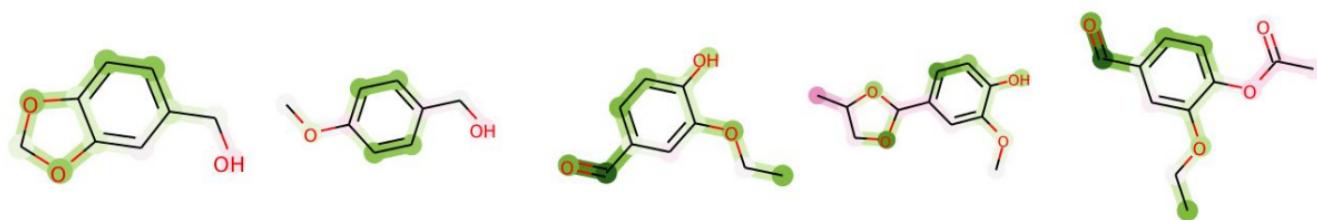


But why is the neural network making these predictions?

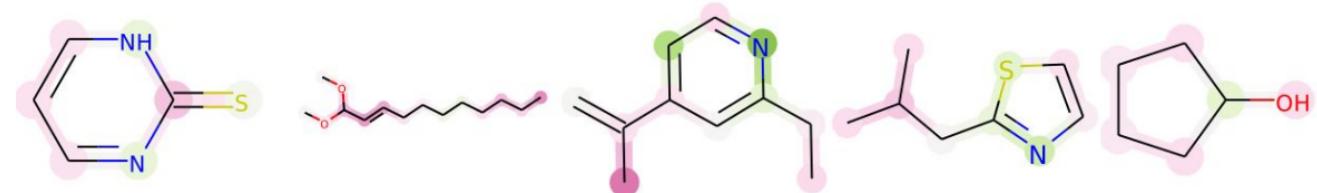
Odor percept — “vanilla”



Positive examples

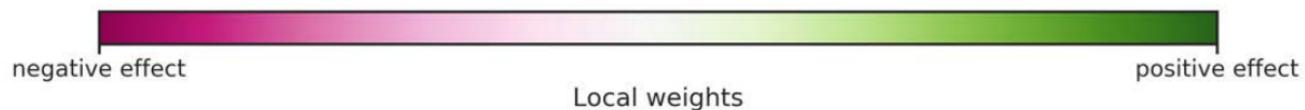


Negative examples

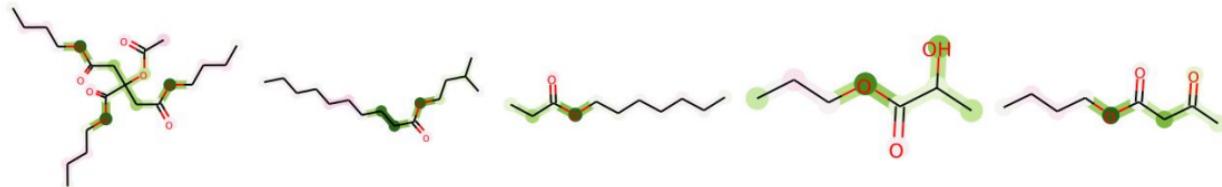


But why is the neural network making these predictions?

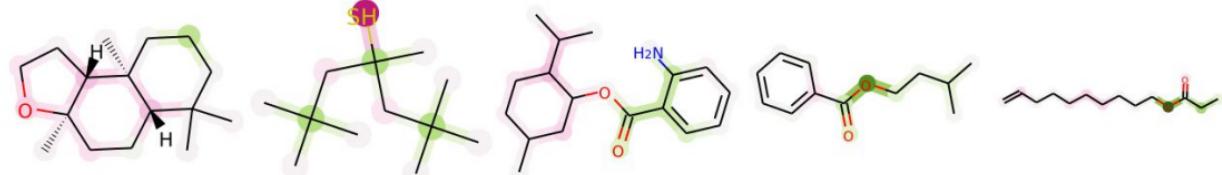
Odor percept — “winey”

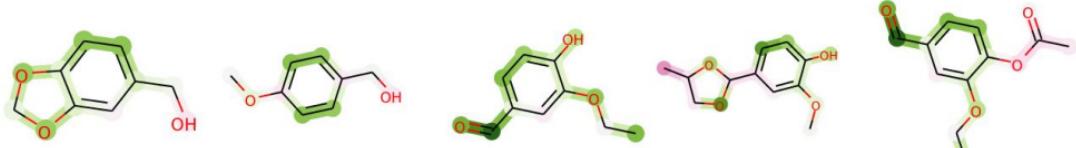
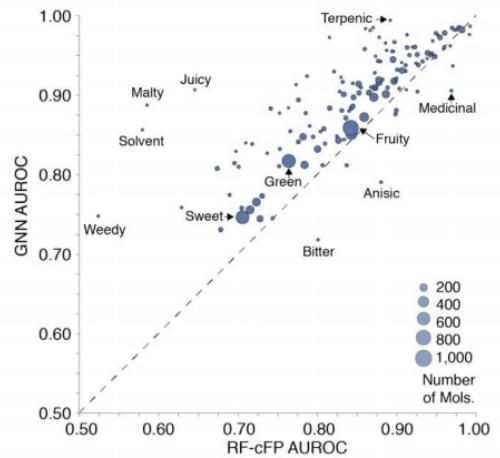
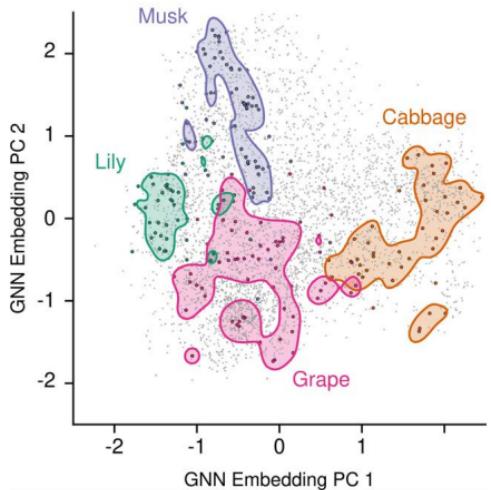
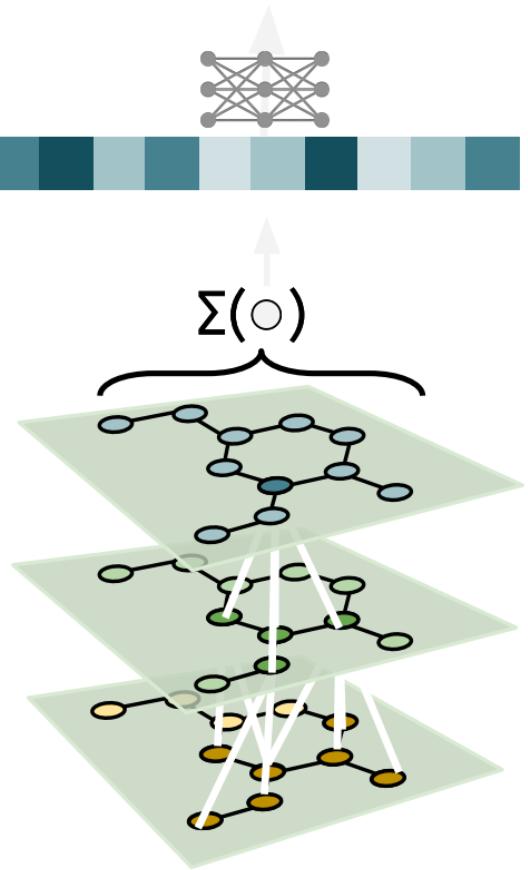


Positive examples



Negative examples





Future Directions

Collecting interest & those interested in collaborating.

- **Test ML-driven molecular design** for humans in a safe context.
- Build bedrock understanding in single-molecules before working on **odor mixtures**
- Build a **foundational dataset** for the ML on molecules community.

Benjamin Sanchez-Lengeling

Brian Lee

Carey Radebaugh

Emily Reif

Jennifer Wei

Alex Wiltschko

