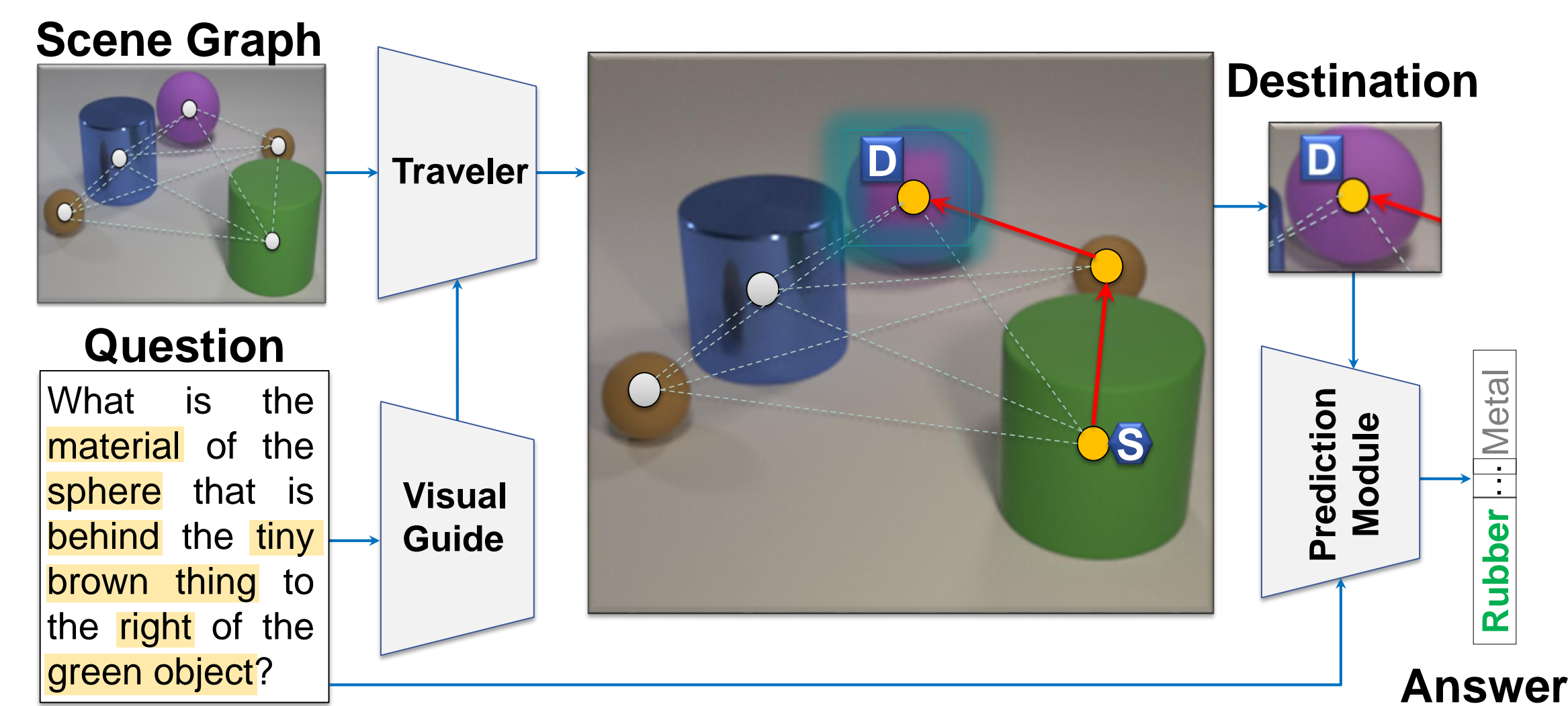


## Motivation

- For most VQA networks the reasoning is not clear
- Long range relations and reasoning steps are often not considered
- Compositional models usually need ground truth programs for training
- We propose a **highly-interpretable graph-based** model for VQA, which operates directly on the **structure of the image**

## Overview of the Approach



## Soft-Paths Components

### Visual Guide

- Takes as input the question and represents it using 1D CNNs
- Produces direction embeddings

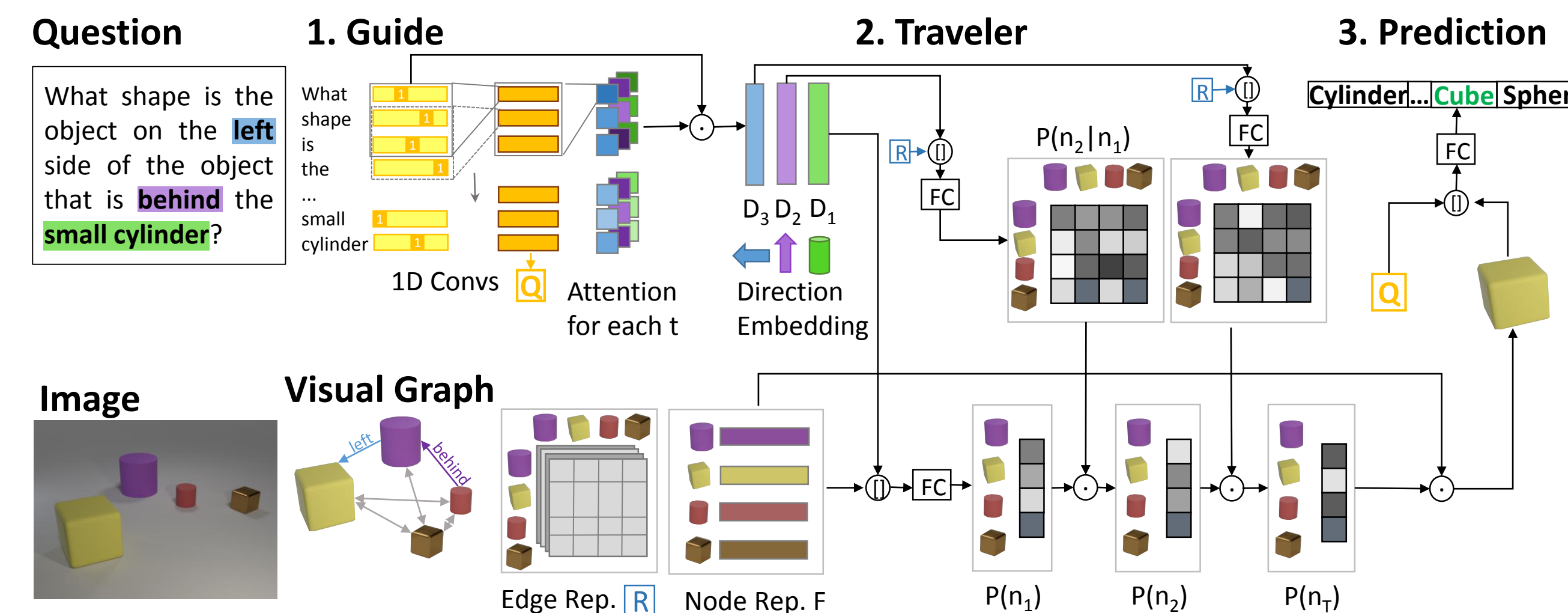
### Traveler

- Traverses the graph using the directions from the guide
- Final destinations are used for the prediction module

### Prediction Module

- Produces the answer using the destination and the question

## The Soft Paths Architecture



## Visual Reasoning Results on CLEVR, AI2D and COG

### Performance on the CLEVR Test Set

Approach	Reference	Count	Exist	Comp. Nrs.	Query Attrs.	Comp. Attrs.	All
<b>Classical VQA Methods</b>							
CNN+SA Santoro et al.	ECCV'16	64.4	82.7	77.4	82.6	75.4	76.6
QGHC Gao et al.	ECCV'18	91.2	78.1	79.2	89.7	86.8	86.3
FiLM Perez et al. et al.	AAAI'18	94.3	99.1	96.8	99.1	99.1	97.7
<b>Compositional Models</b>							
N2NMN* Hu et al.	ICCV'17	68.5	85.7	84.9	90.0	88.7	83.7
PG(700K)* Johnson et al.	ICCV'17	92.7	97.1	98.7	98.1	98.9	96.9
<b>Memory Networks</b>							
Work. Mem. Yang et al.	ECCV'18	91.7	99.0	95.5	98.5	98.8	96.8
MAC <sup>†</sup> Hudson et al.	ICLR'18	97.1	99.3	96.8	99.1	99.1	98.9
<b>Graph Neural Networks</b>							
CNN+RN <sup>‡</sup> Santoro et al.	NIPS'17	90.1	97.8	93.6	97.9	97.1	95.5
Ours	—	91.3	98.6	99.6	99.5	99.8	97.5

Approach	Atts.	Condit.	Point	Yes/No	All
<b>Baselines</b>					
Random	1.9	8.4	17.5	50.0	26.6
Question-only	1.6	2.3	19.4	49.7	27.4
<b>Memory Networks</b>					
Work. Mem. <sup>†</sup> Yang et al.	—	—	—	—	93.7
<b>Graph-based Methods</b>					
Question+Nodes	73.7	63.5	92.5	57.9	63.3
Ours	99.2	98.4	100.0	95.0	97.2

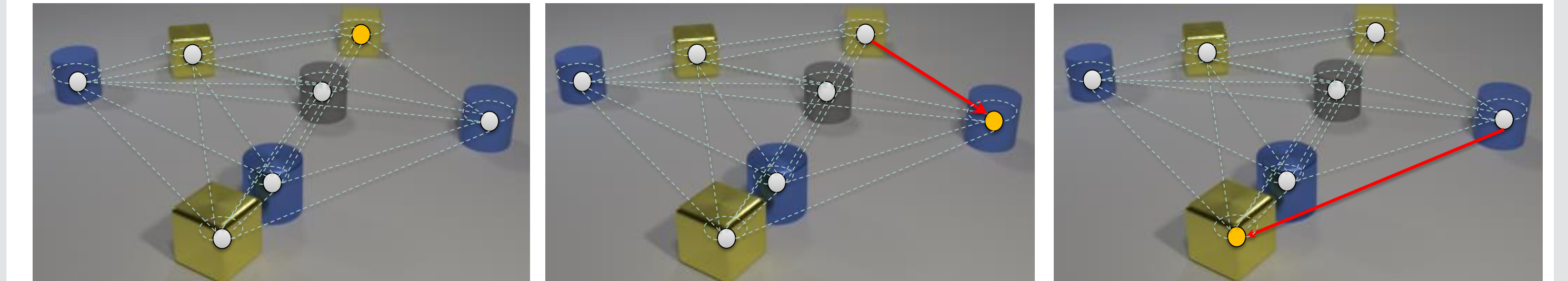
### Accuracy on the AI2D Diagram Dataset

Approach	All
<b>Baselines</b>	
Random	25.00
<b>Classical VQA Methods</b>	
VQA Agrawal et al.	32.90
<b>Graph Neural Networks</b>	
DQA-Net [DSDP] Kembhavi et al.	38.47
DQA-Net [DGGN] Kim et al.	39.73
DQA-Net Kim et al.	41.55
Ours	43.45

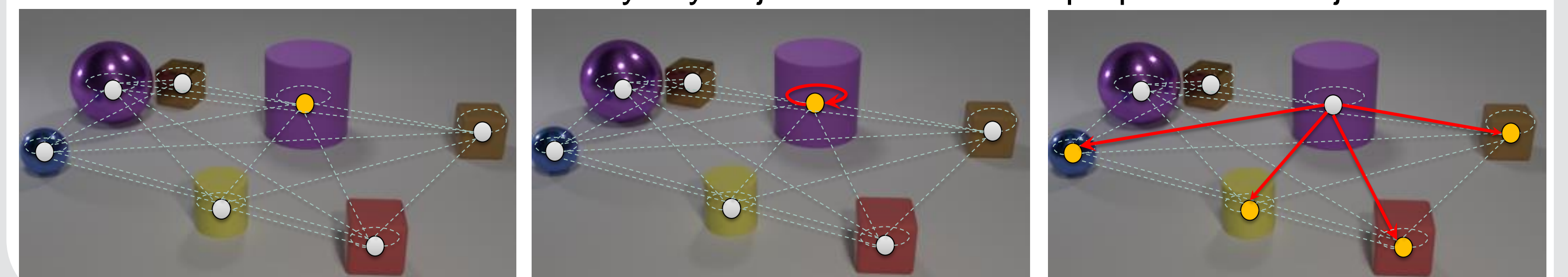
### Test Performance on the COG Video Dataset

## Qualitative Results

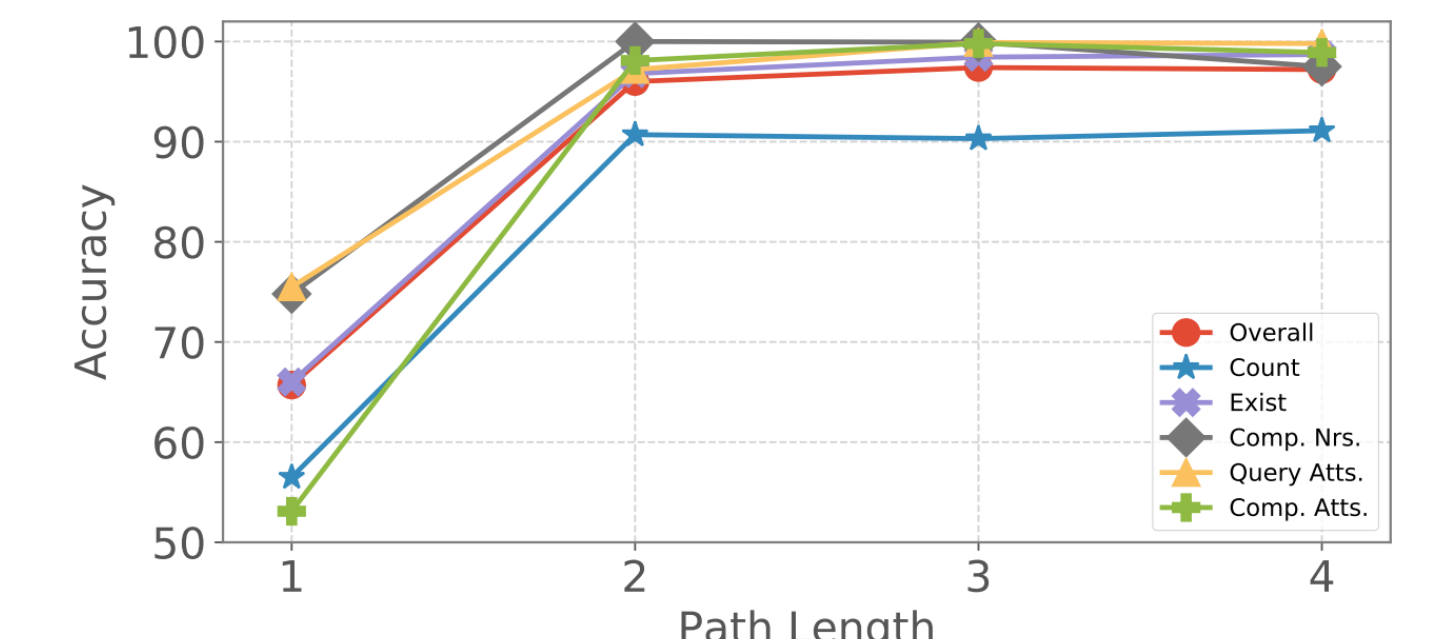
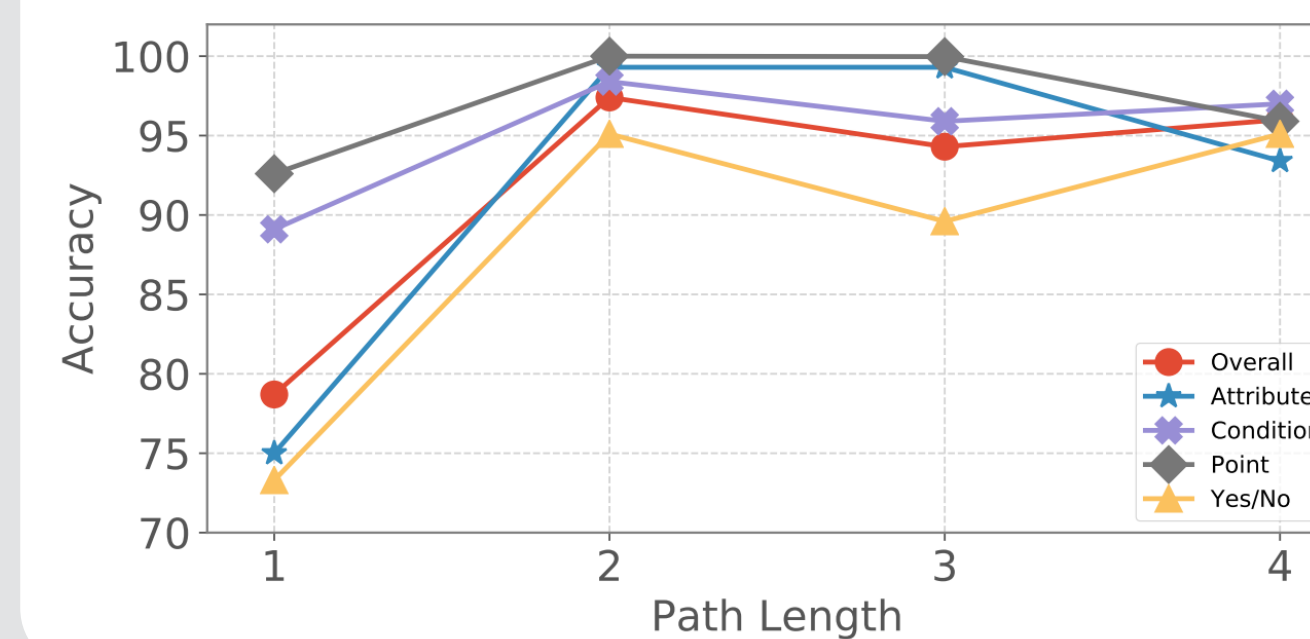
**Question:** There is a tiny rubber thing that is right of the matte cube. Are there any yellow cubes in front of it?



**Question:** Are there any tiny objects in front of the purple rubber object?



## Impact of Path Length on Performance



## Contact

- Email [haurilet@kit.edu](mailto:haurilet@kit.edu)
- Website <https://cvhci.anthropomatik.kit.edu/~mhaurile/>

