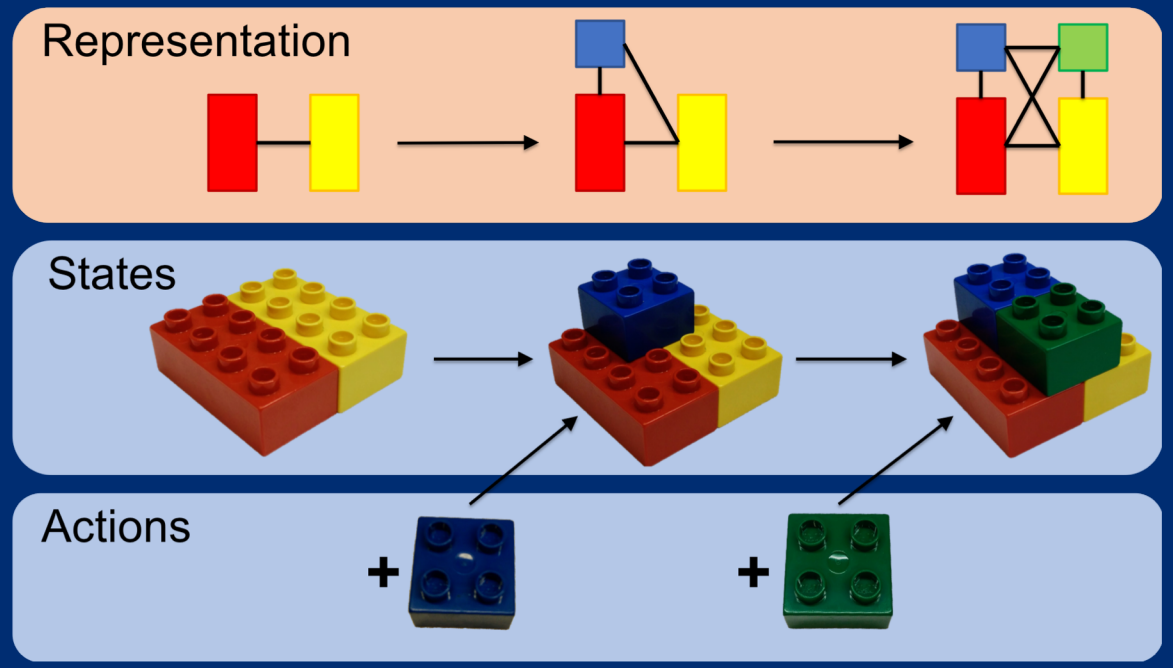
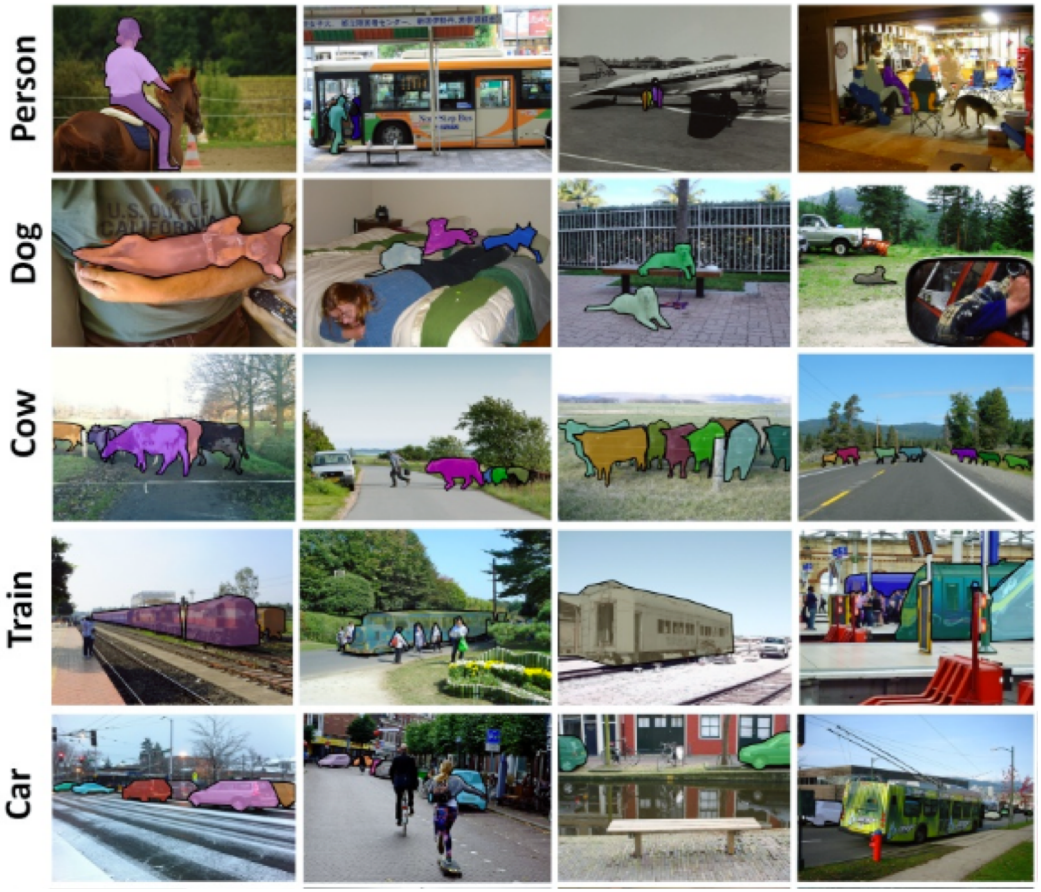
JOHNS HOPKINS UNIVERSITY

Toward Computer Vision Systems that Understand Real-World Assembly Processes

Jonathan D. Jones, Gregory D. Hager, Sanjeev Khudanpur



Computer vision in dynamic environments



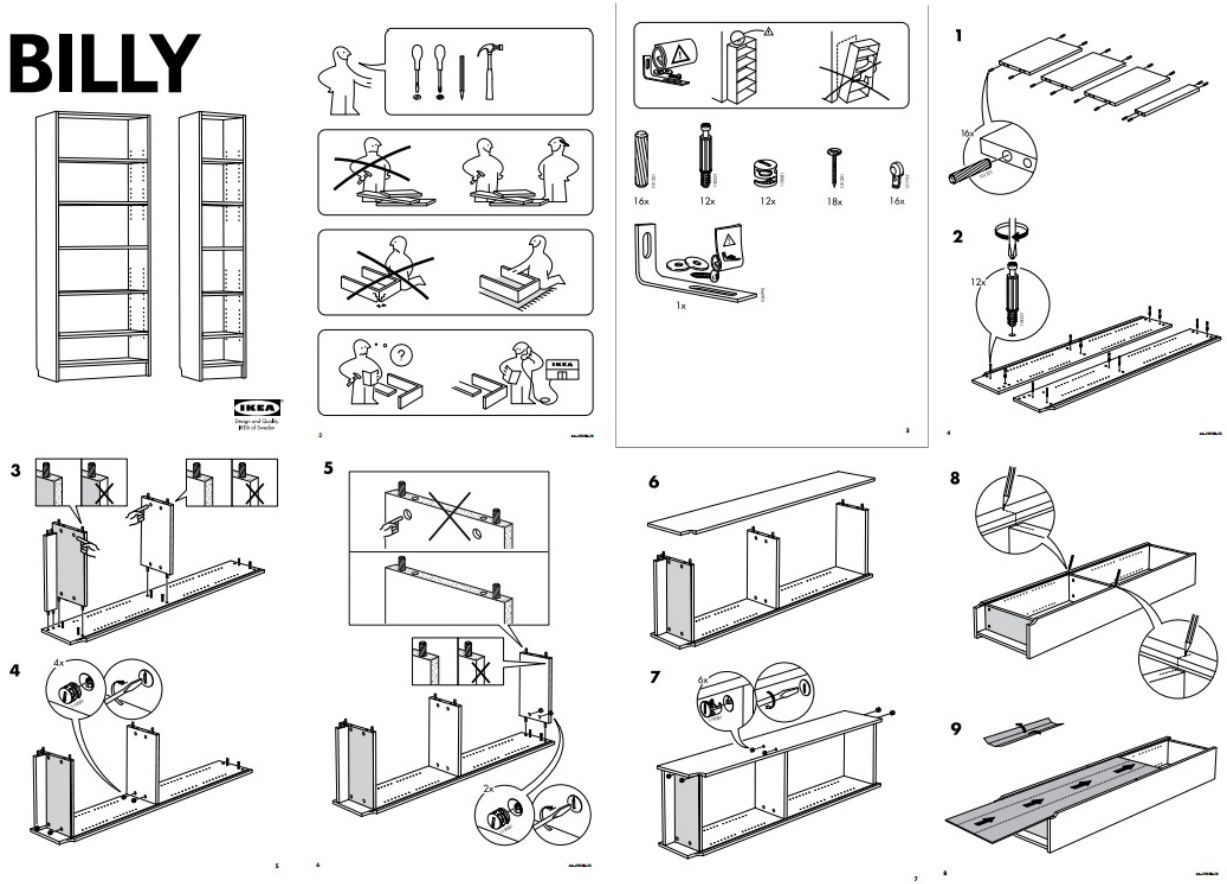
Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." *European conference on computer vision*. Springer, Cham, 2014.

By Steve Jurvetson - Flickr: Tesla Autobots, CC BY 2.0, https://commons.wikimedia.org/w/index.php?curid=24819289

Computer vision in dynamic environments

- Environments not always *static*
  - Objects interact with each other
  - Can experience state changes

- Example: *Assembly processes*
  - Collaborative robots
  - Industrial monitoring



Billy bookcase instructions, IKEA

Application: Parsing Block-Building Videos

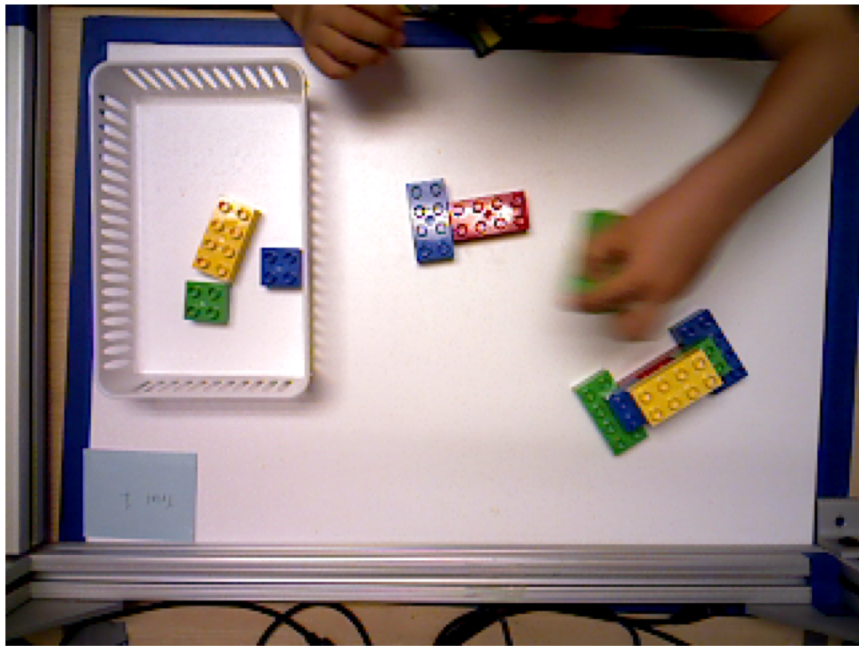
JOHNS HOPKINS UNIVERSITY

Collaborators

Barbara Landau
Cognitive Science

Amy Shelton
Education

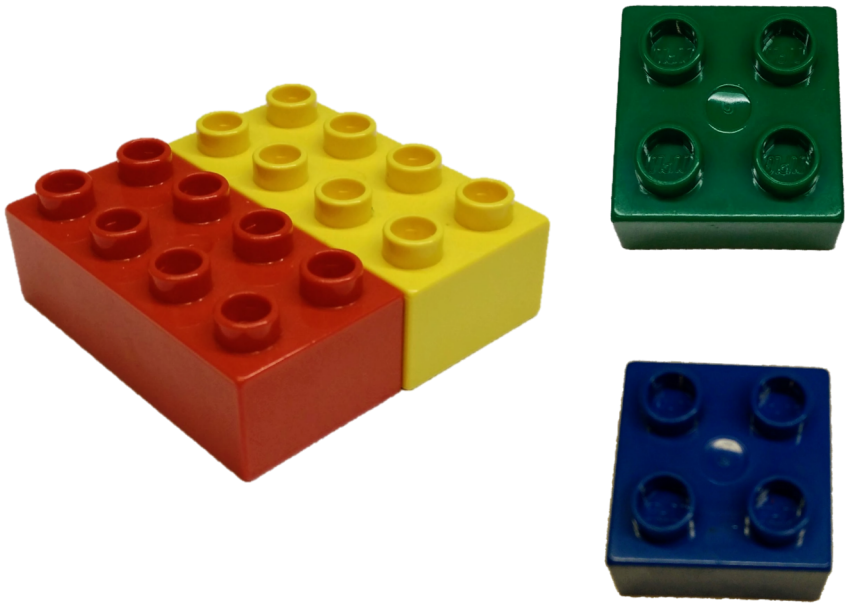
Cathryn Cortesa
Cognitive Science

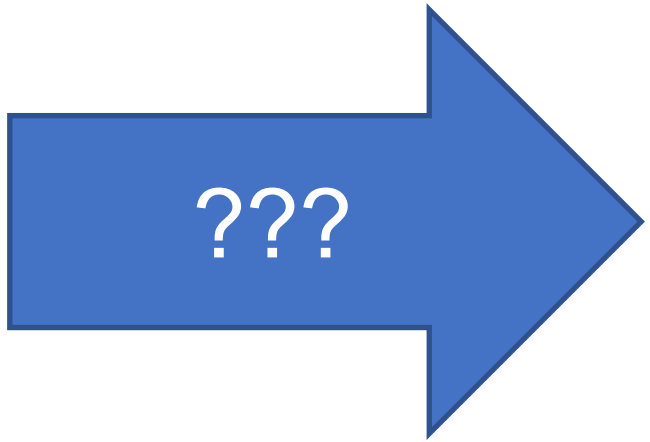


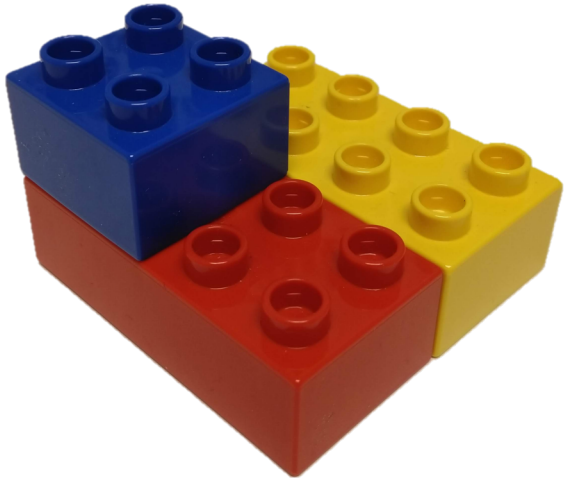
Cortesa, C. S., Jones, J. D., Hager, G. D., Khudanpur, S., Landau, B., & Shelton, A. L. (2018) Constraints and Development in Children's Block Construction. *CogSci 2018 Proceedings*, 246-251.

Cortesa, C. S., Jones, J. D., Hager, G. D., Khudanpur, S., Shelton, A. L., & Landau, B. (2017). Characterizing spatial construction processes: Toward computational tools to understand cognition. *CogSci 2017 Proceedings*, 246-251.

Representing assembly processes

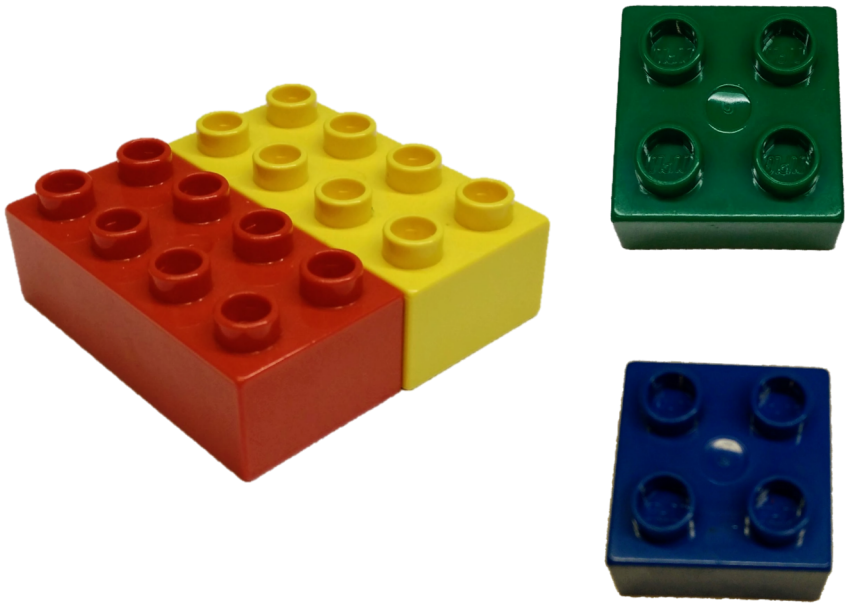




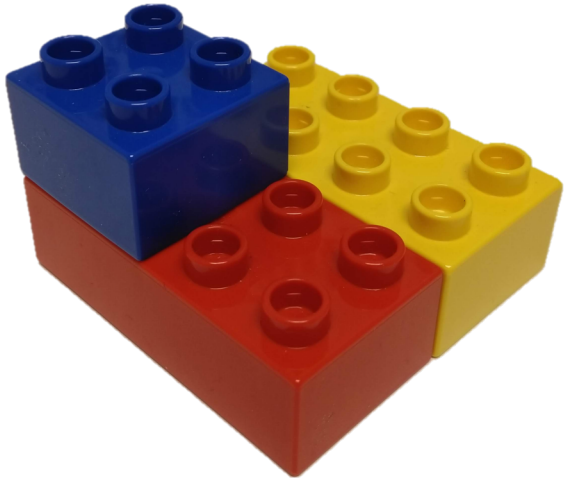


1. Choose the right block
2. Make the right connections
3. Connect in the right way

Representing assembly processes



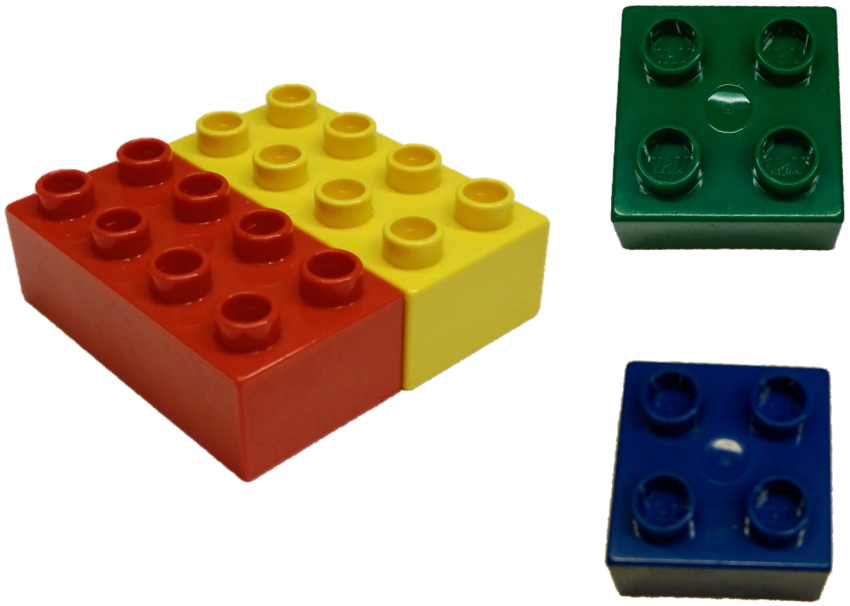
action

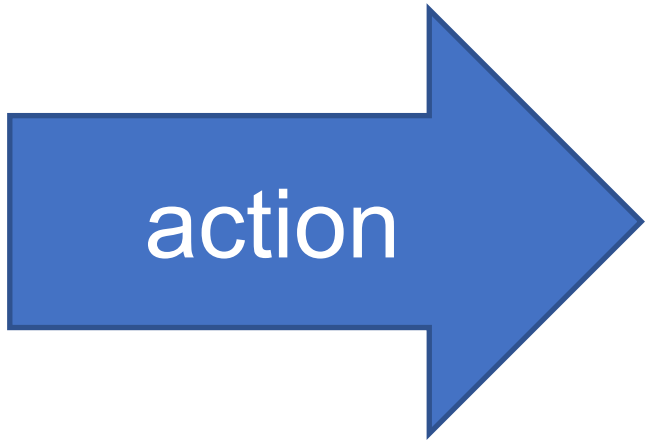


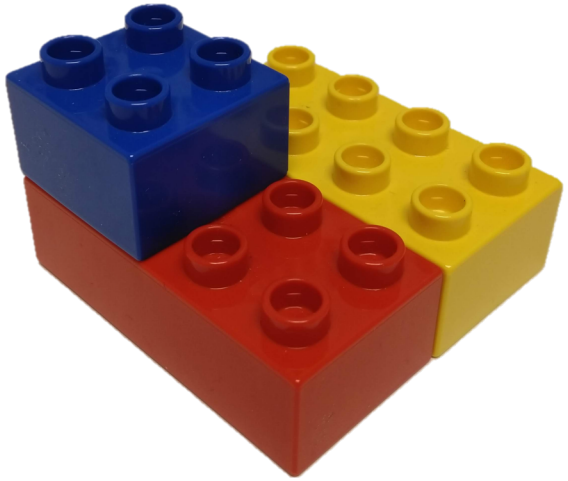
1. Choose the right block
2. Make the right connection
3. Connect in the right way

connect (, , ,)

Representing assembly processes



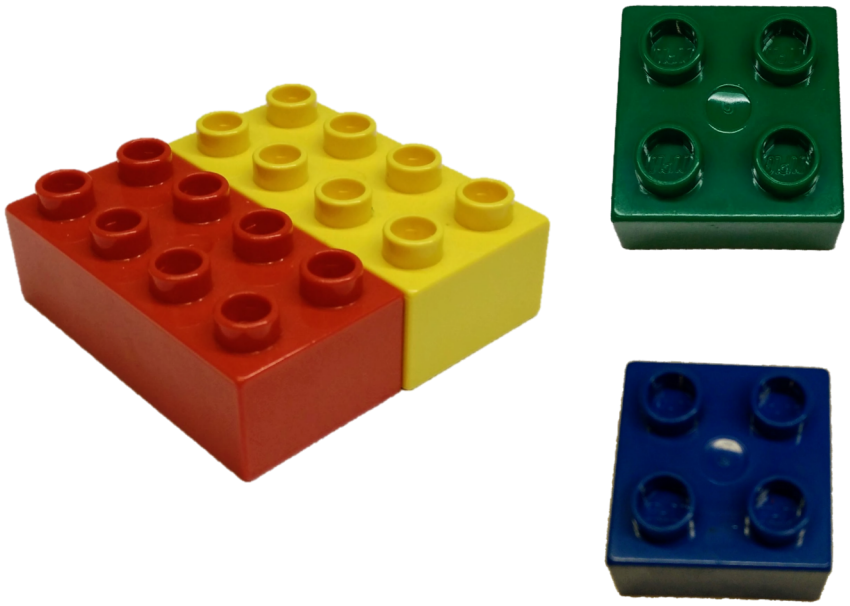


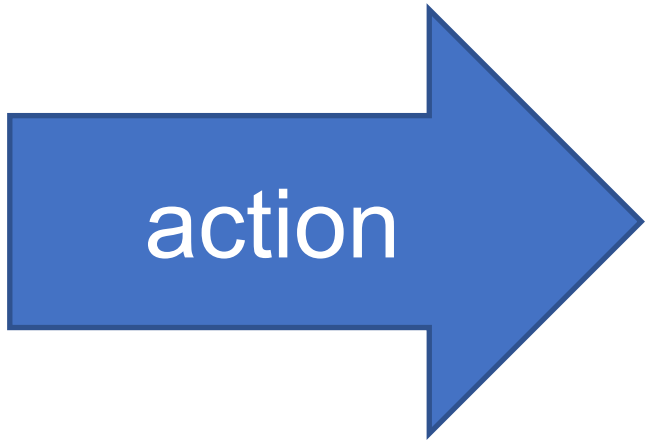


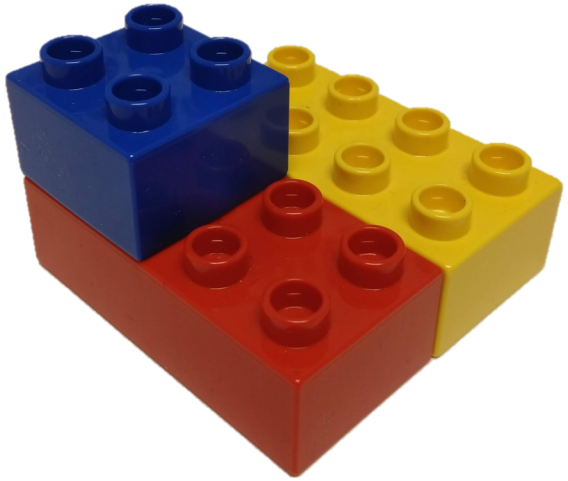
1. Choose the right block
2. Make the right connection
3. Connect in the right way

connect (, , ,)

Representing assembly processes





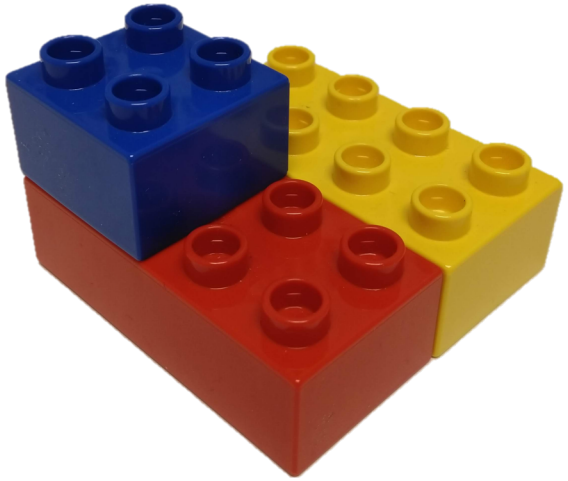


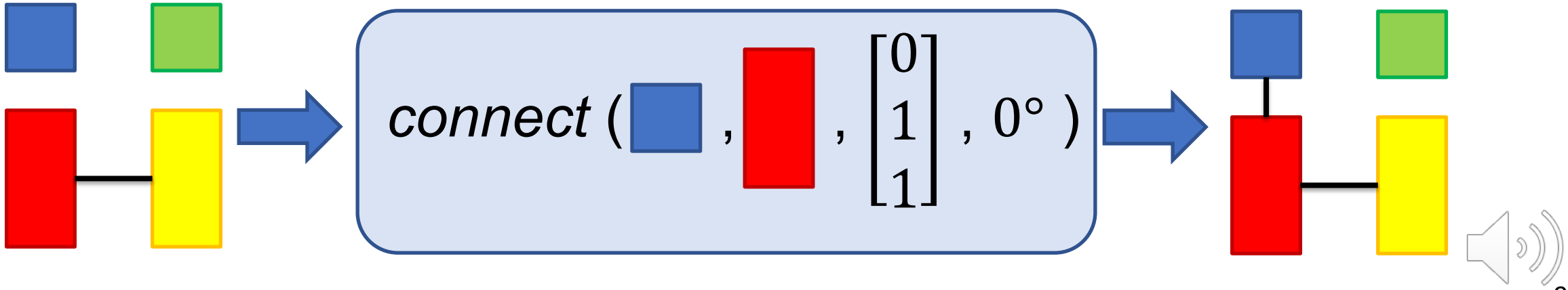
1. Choose the right block
2. Make the right connection
3. Connect in the right way

$$connect\left(\ \square\ ,\ \square\ ,\ \begin{bmatrix}0\\1\\1\end{bmatrix},\ 0°\ \right)$$

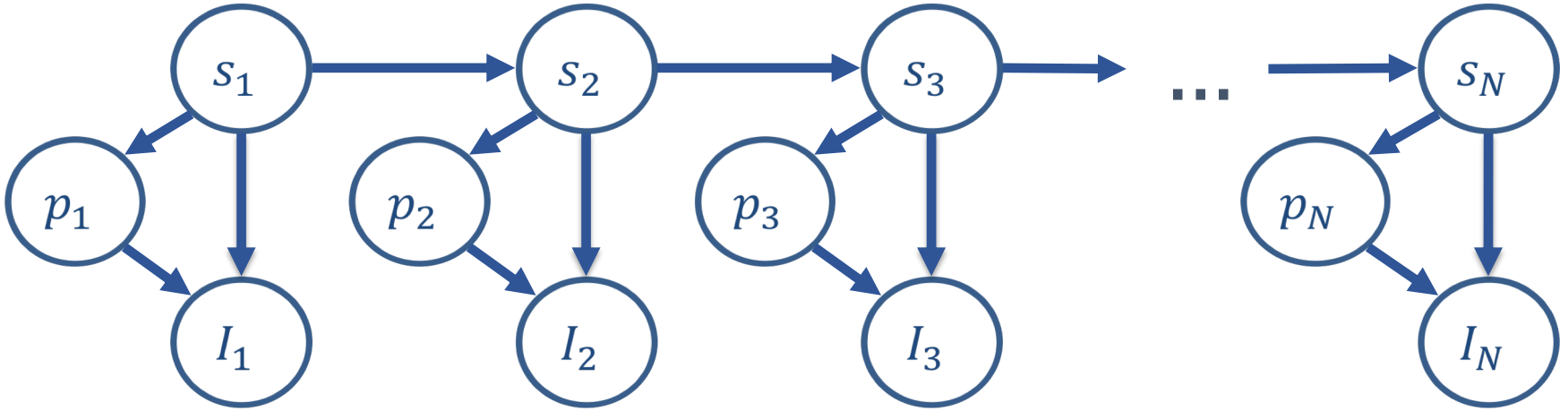
Representing assembly processes

- State is a graph
- Vertices are blocks
- Edges are block connections
- Edges are labeled with a block's relative pose in the coordinate frame of its neighbor





A probabilistic model

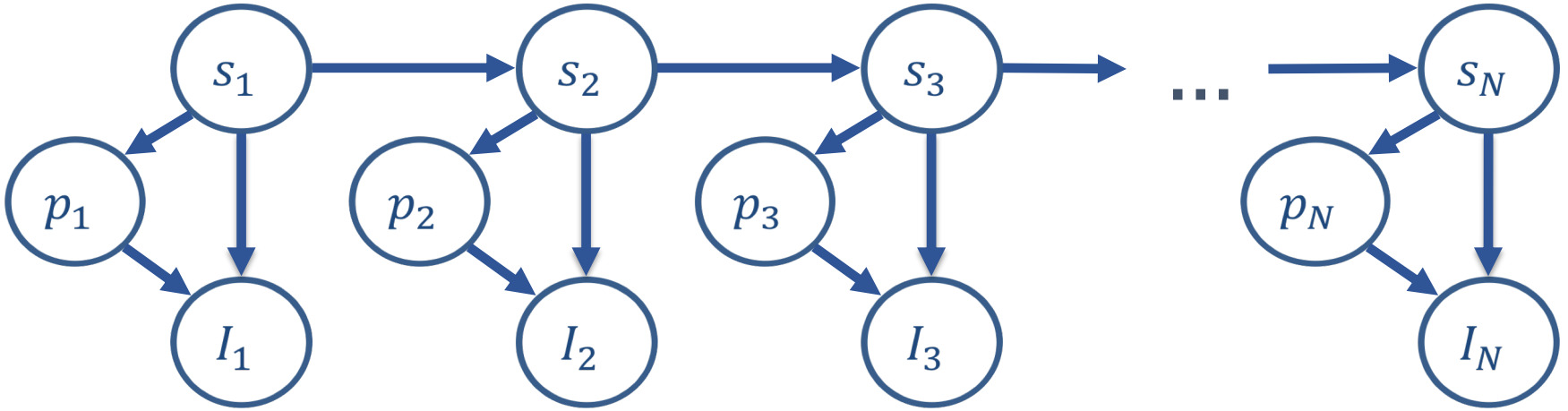


s_t: state of block model

p_t: pose of block model

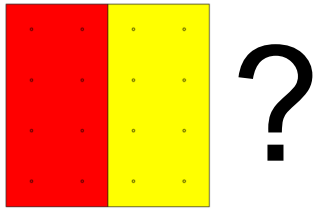
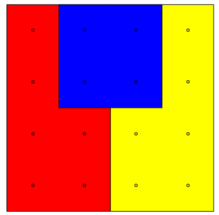
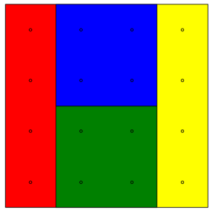
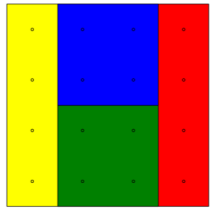
I_t: video keyframe

A probabilistic model



$$p^*_{1:N}, s^*_{1:N} = \underset{p_{1:N},\, s_{1:N}}{\operatorname{argmax}} P\left(p_{1:N}, s_{1:N} \mid I_{1:N} \right)$$

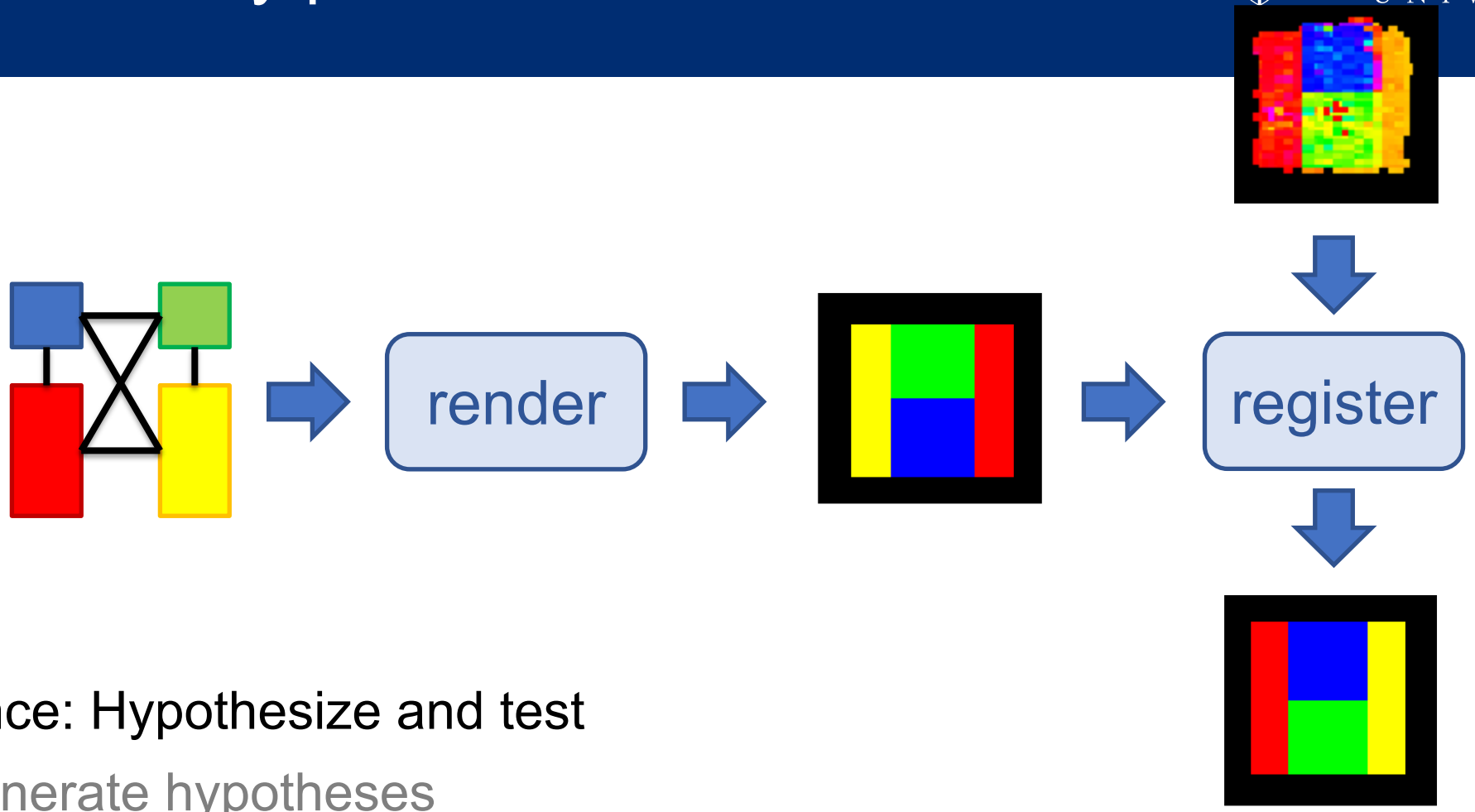
Parsing assembly processes


?

?

?

?

Inference: Hypothesize and test

1. Generate hypotheses
2. Test hypotheses locally (render & register template)
3. Decode best state sequence globally

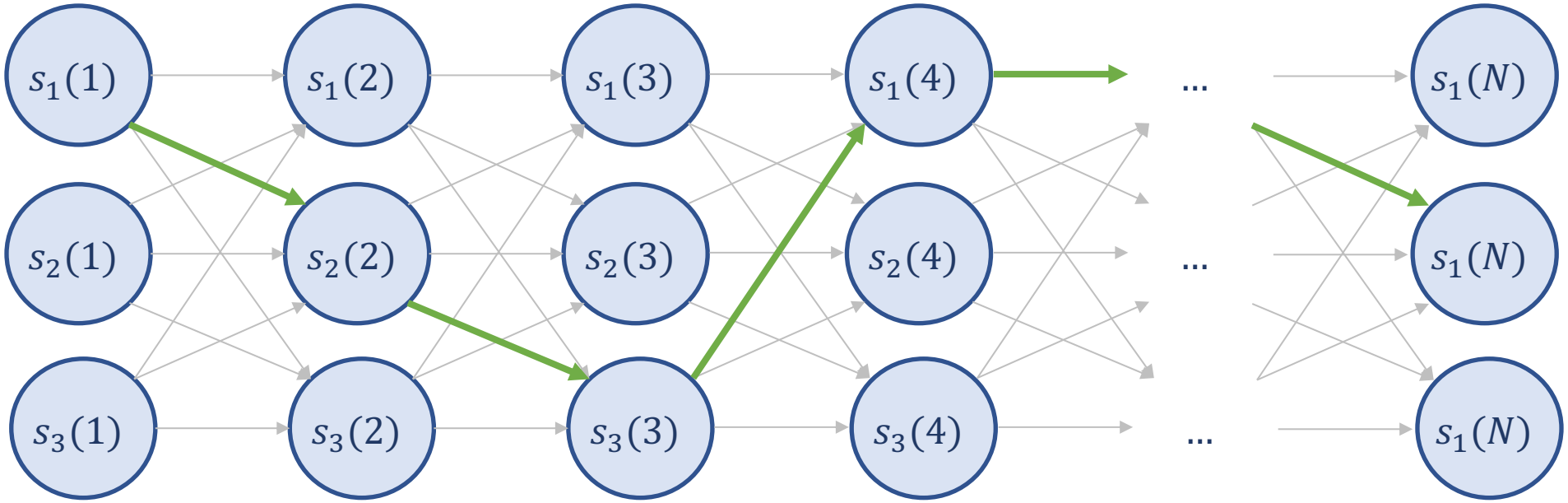
Parsing assembly processes



Inference: Hypothesize and test

1. Generate hypotheses
2. Test hypotheses locally (render & register template)
3. Decode best state sequence globally

Parsing assembly processes



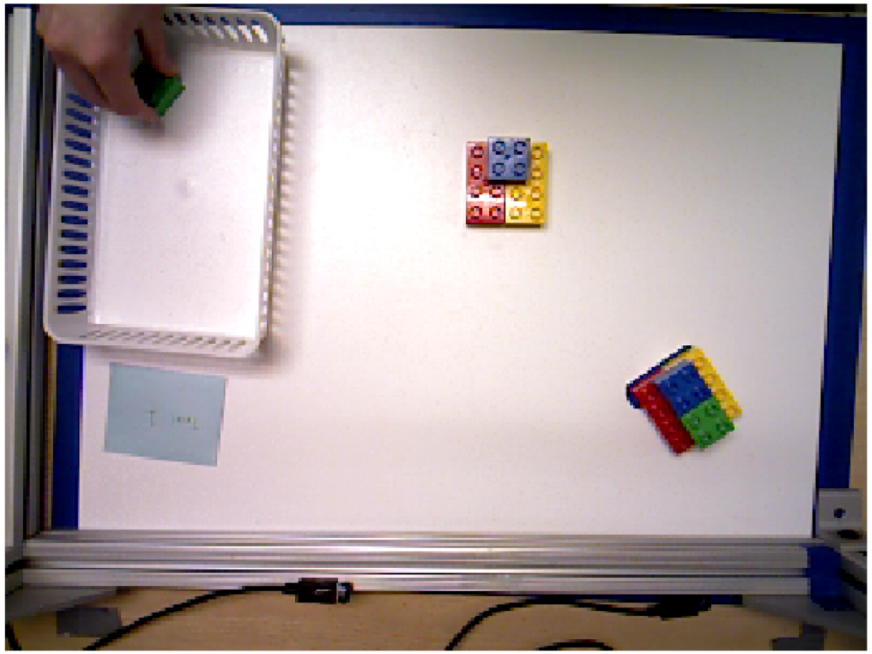
Inference: Hypothesize and test

1. Generate hypotheses
2. Test hypotheses locally (render & register template)
3. Decode best state sequence globally

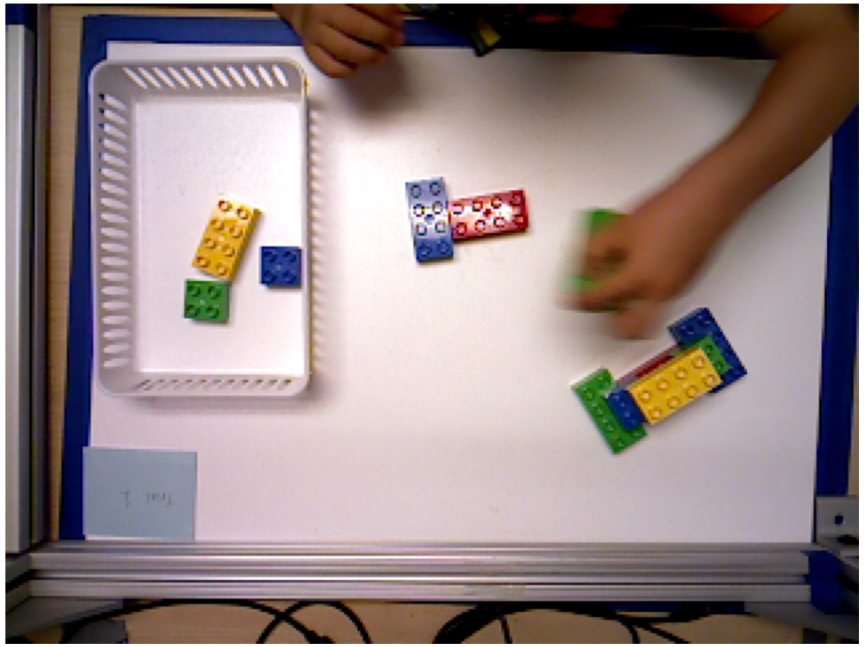
Experiments: Data

JOHNS HOPKINS UNIVERSITY

Controlled dataset

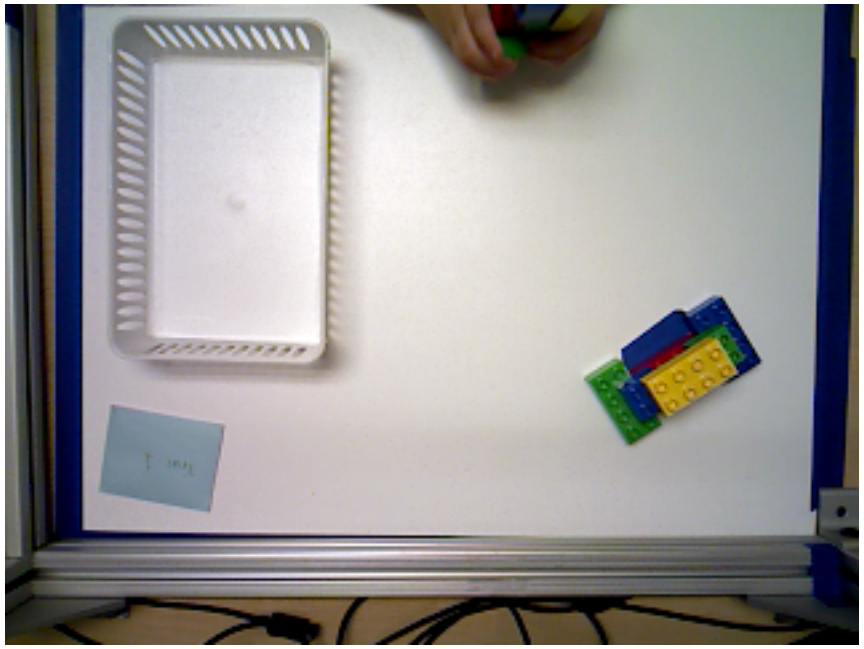


“Child’s play” dataset

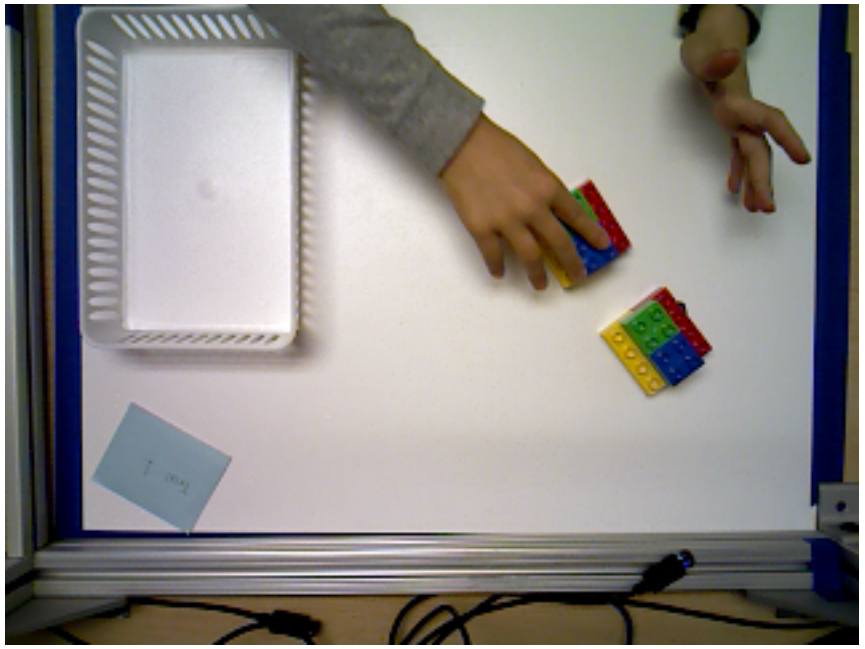


Experiments: Data

“Child’s play” dataset



assembly out of view

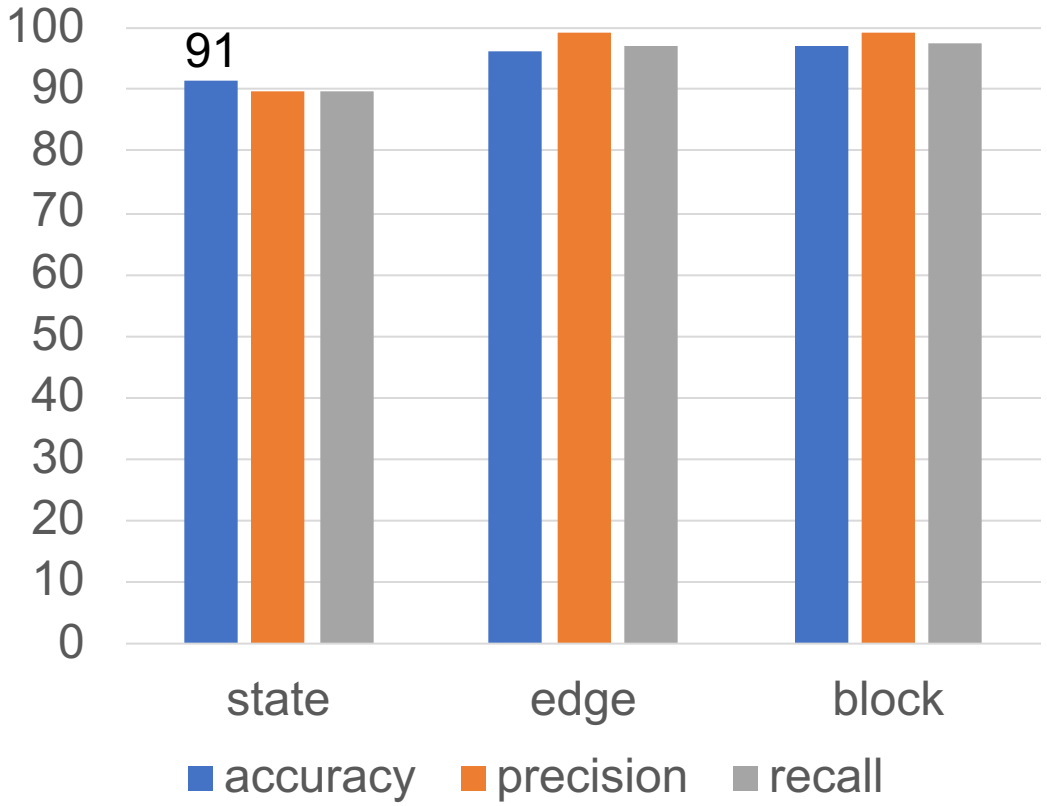


assembly occluded

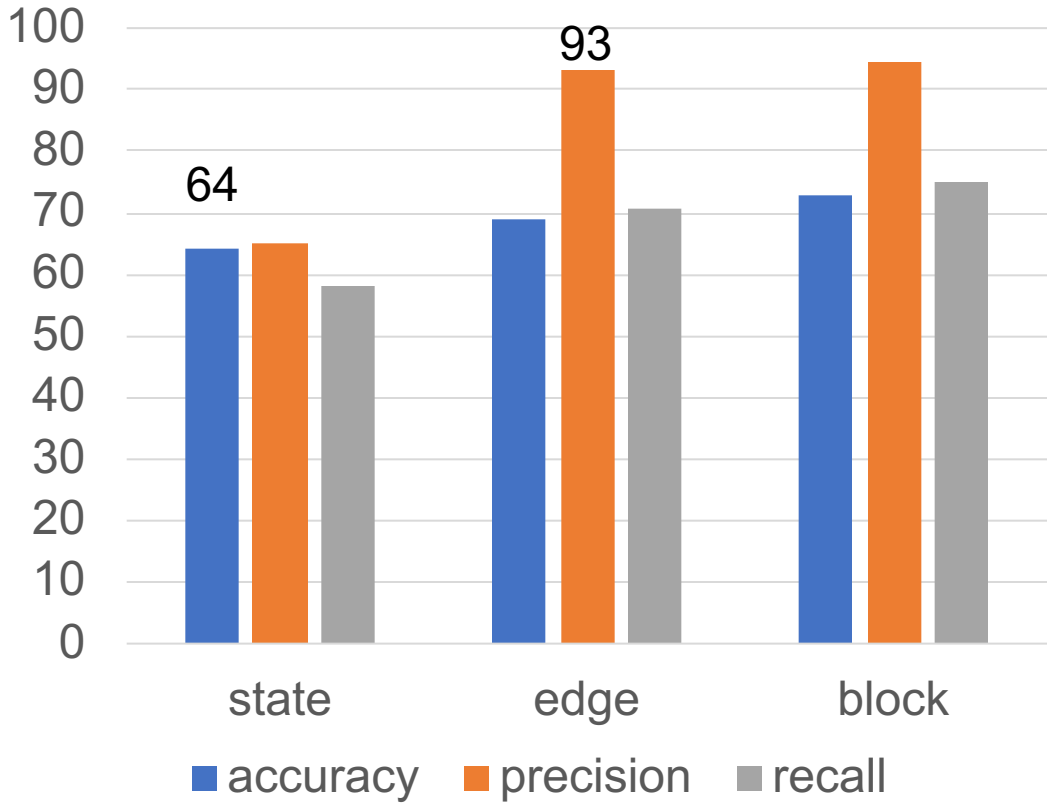
Experiments: Results

JOHNS HOPKINS UNIVERSITY

Controlled dataset

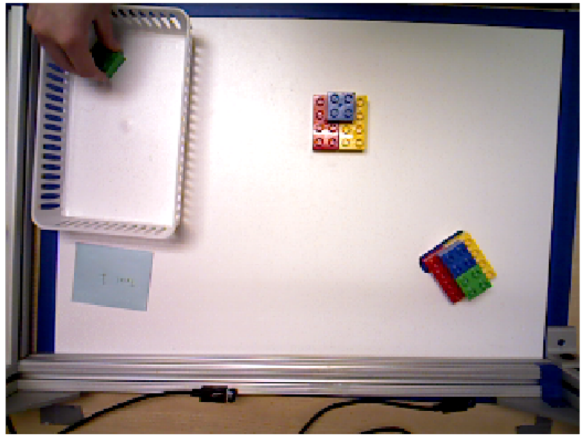


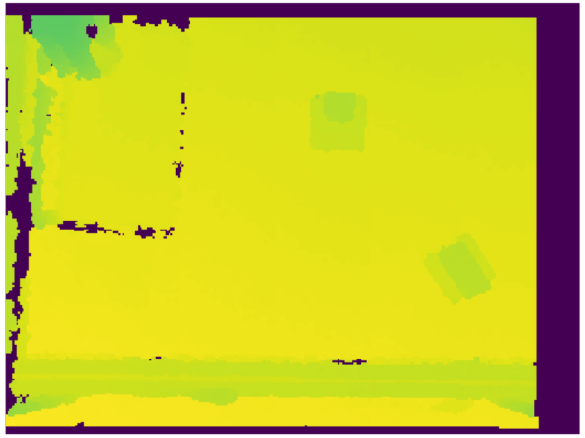
Child's play dataset

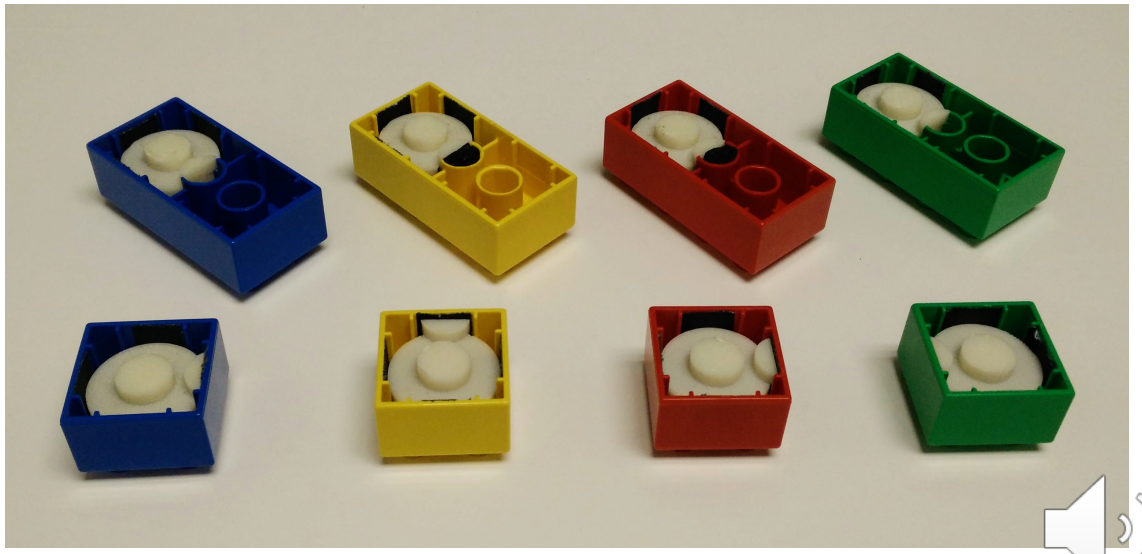


Future work

- Multimodal inference
  - RGB-depth
  - acceleration, angular velocity
- Handling occlusion
- Modeling actions







Acknowledgements

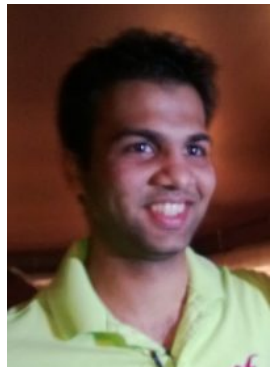
Barbara Landau
Cognitive Science

Amy Shelton
Education

Greg Hager
Computer Science

Sanjeev Khudanpur
Electrical Engineering

Cathryn Cortesa
Cognitive Science



Anand Malpani
Computer Science

Jonathan Jones
Electrical Engineering