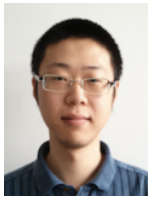
max planck institut informatik

UNIVERSITY OF BIRMINGHAM

Visual Stability Prediction and Its Application to Manipulation

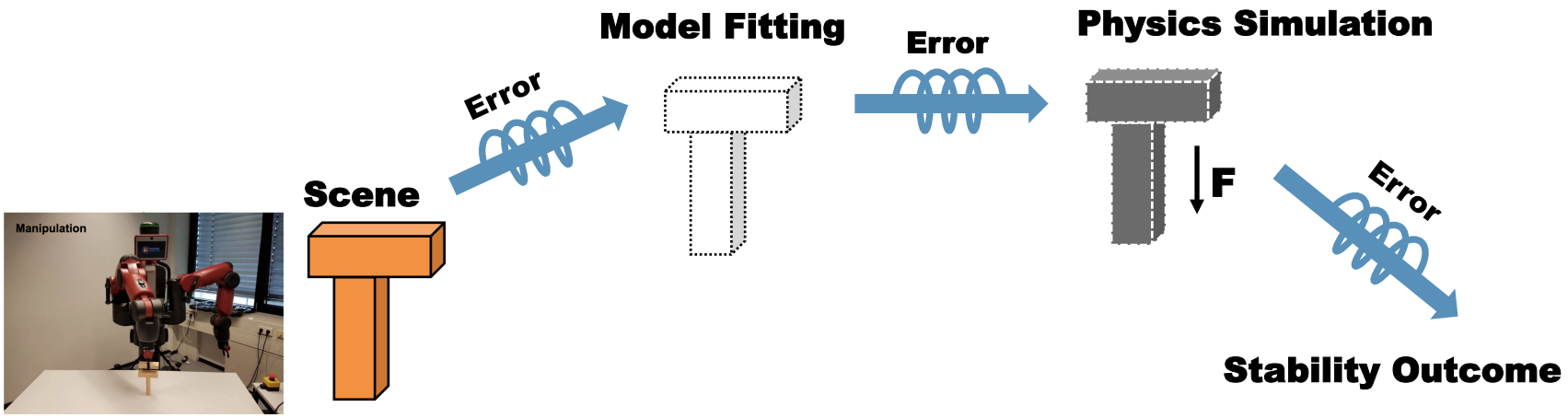


Wenbin Li[1] Aleš Leonardis[2] Mario Fritz[1]
Max Planck Institute for Informatics, Saarland Informatics Campus, Germany[1]
School of Computer Science University of Birmingham, United Kingdom[2]

Motivation

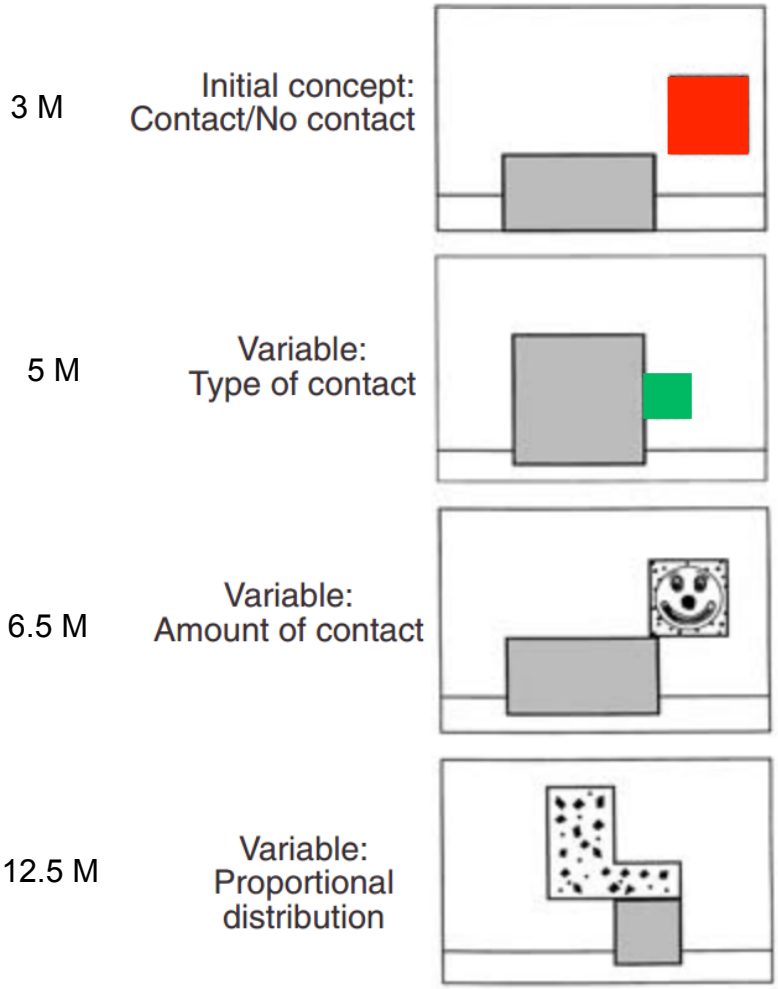
- Understanding physics empowers human's capability for manipulation

Acquiring Intuitive Physics

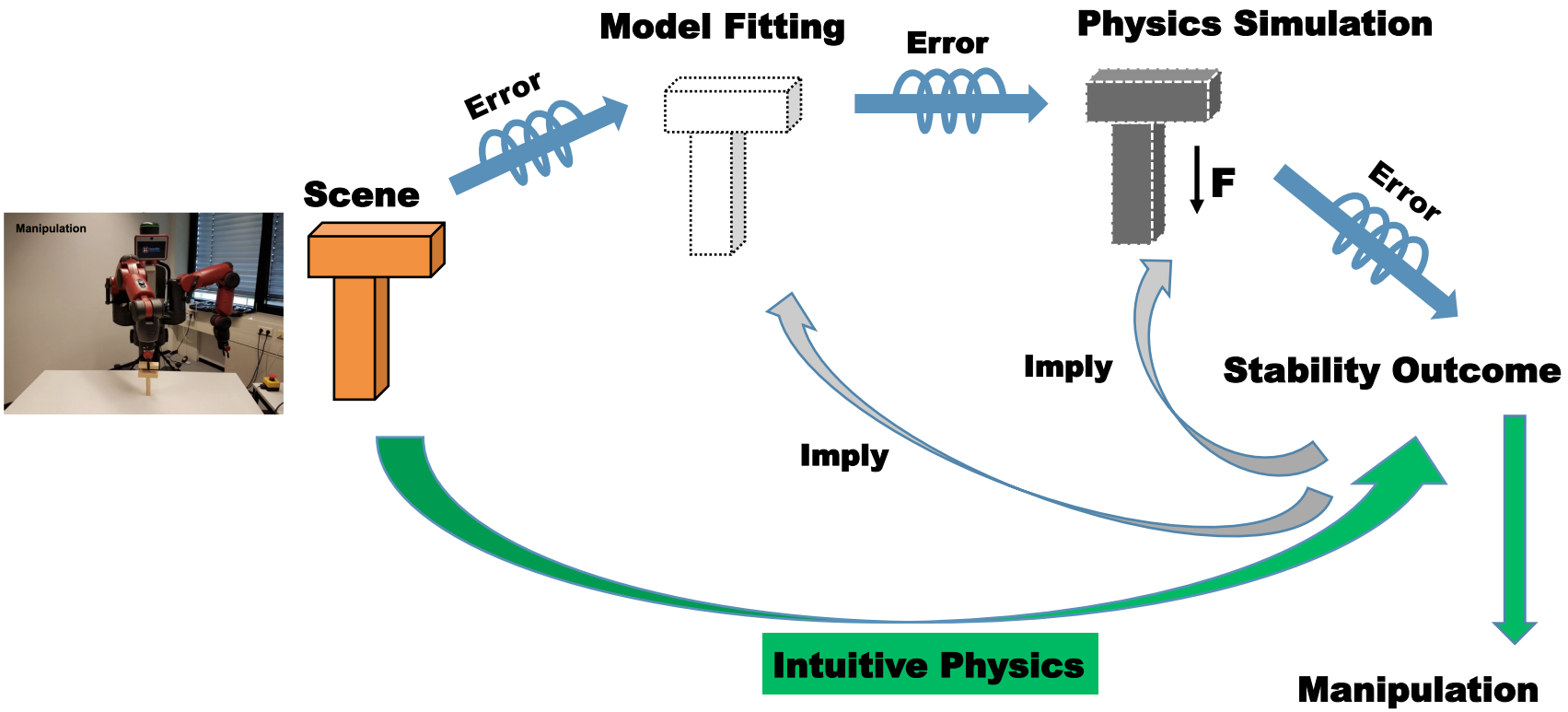


Infant Learn about Support Events

[Baillargeon' 2002]

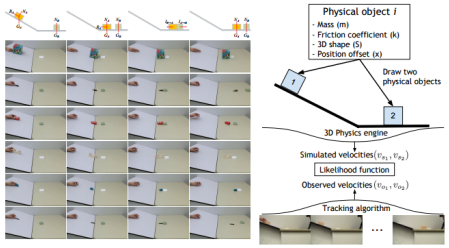


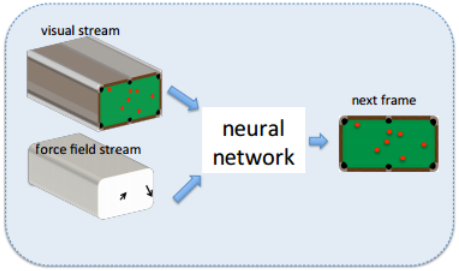
Acquiring Intuitive Physics

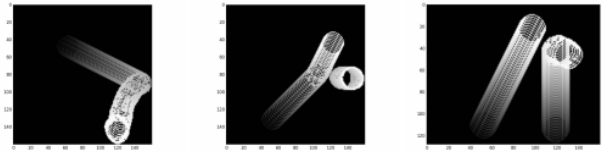


Related Work

- Simulation-based approach
  - [Battaglia'2013]
- Parametric models
  - [Mottaghi'2016, Wu'2015]
- Object-centric
  - [Fragkiadaki'2016]
- Pixel-based
  - [Lerer'2016, Bhattacharyya'2016]

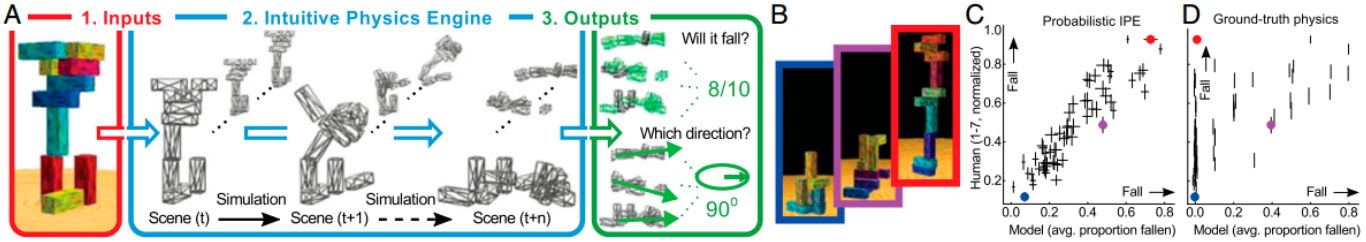




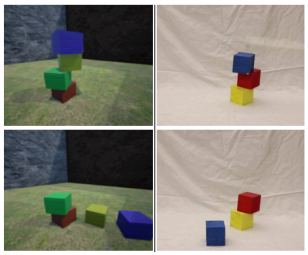


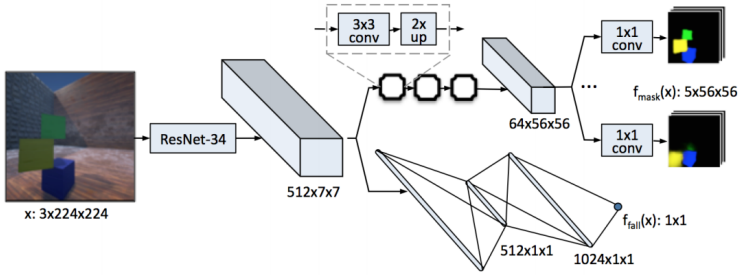
Related Work on Modeling Stability

- [Battaglia'2013]



- [Lerer'2016]





- Ours
  - No simulation at test time (end to end learning)
  - Predicting qualitative outcomes
  - Stability prediction for manipulation

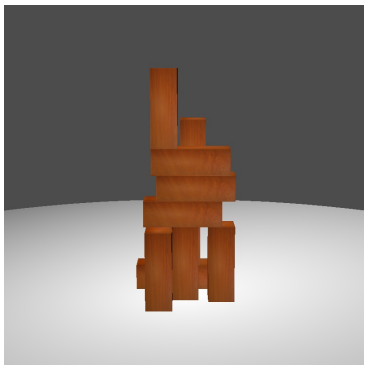
I. Stability Prediction from Visual Appearance

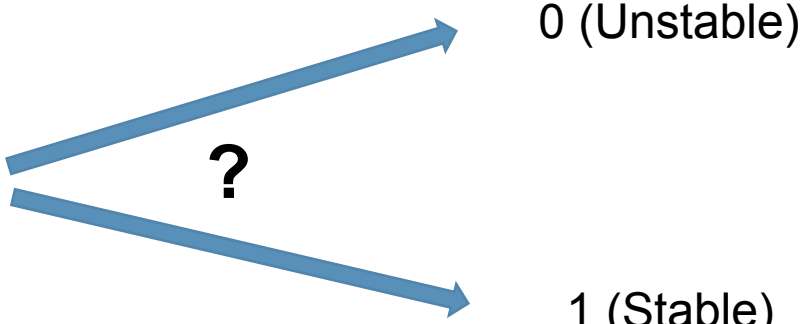
- Model
  - Visual stability prediction
  - Model interpretation
- Data
- Experiment
  - Synthetic data
  - Human subject test
  - Model interpretation

Visual Stability Prediction

- Formulation

Image $\mathbf{I}$ → Stability $\mathbf{S}$ $\{0,1\}$



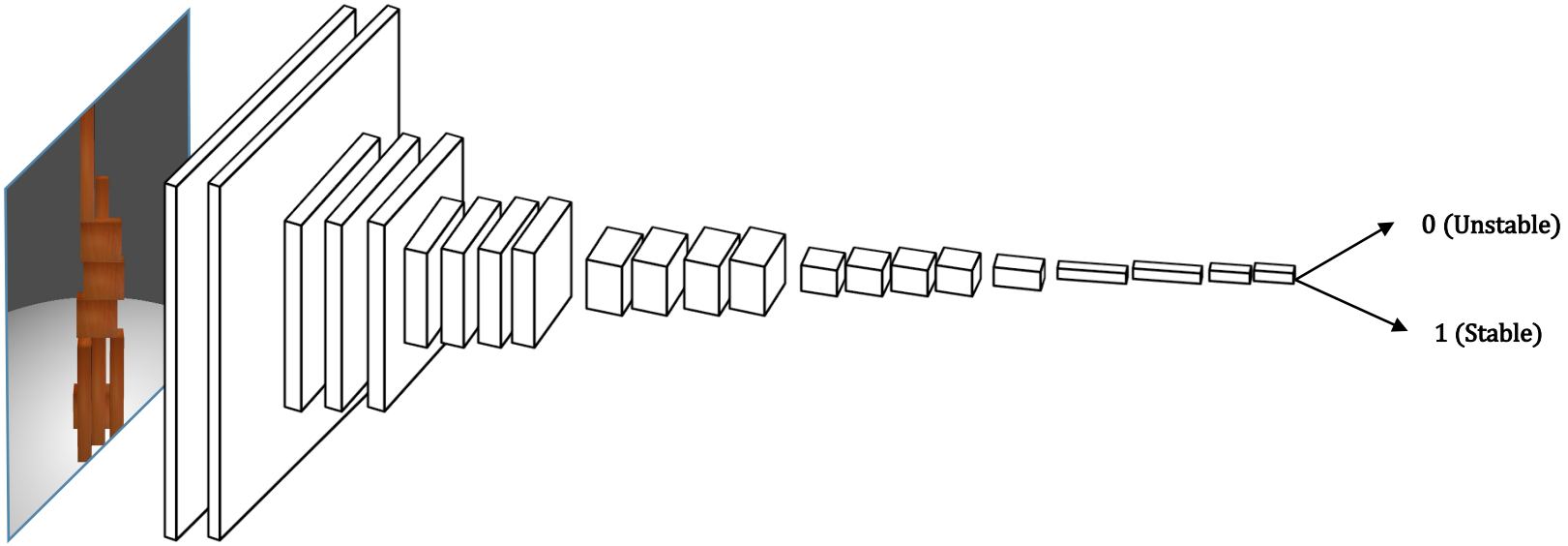


Visual Stability Prediction

- Formulation

Image $\mathbf{I}$ → Stability $\mathbf{S}$ $\{0,1\}$

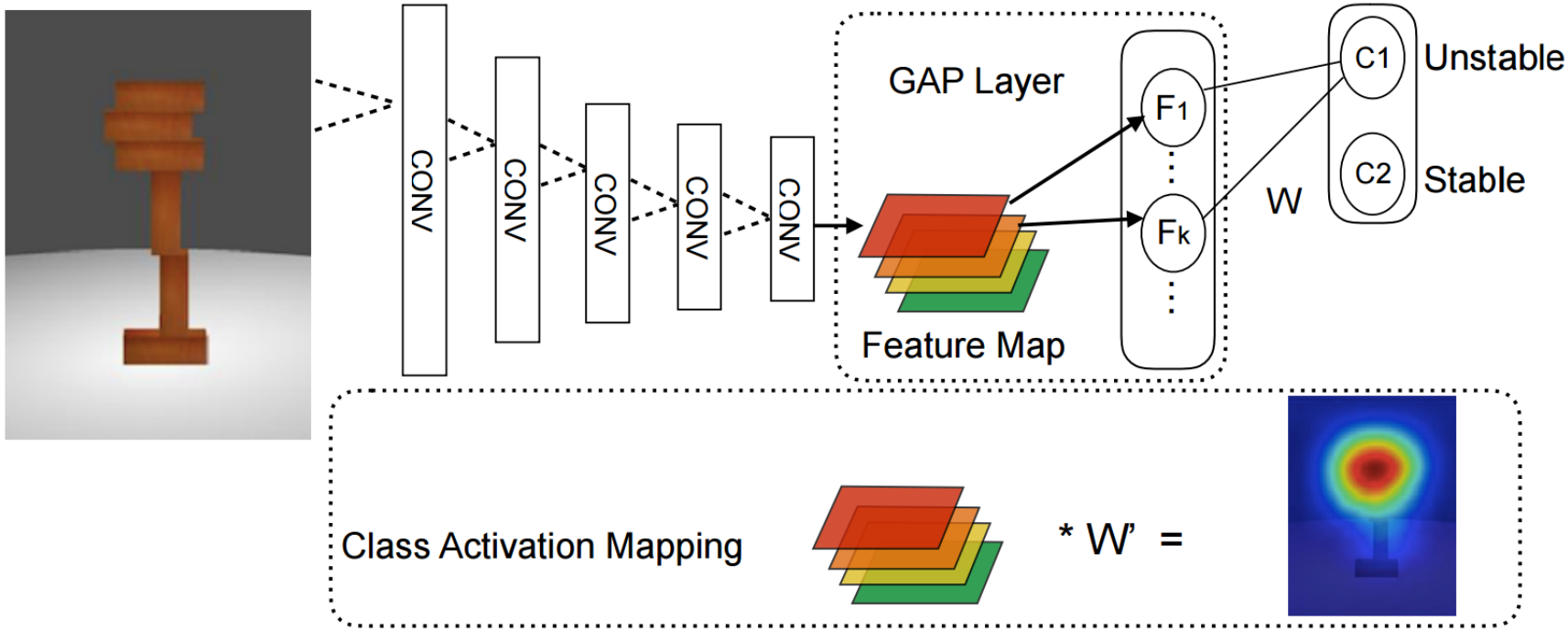
Pilot study: use VGG16



[K. Simonyan et al, "Very deep convolutional networks for large-scale image recognition," ICLR2015]

Model Interpretation

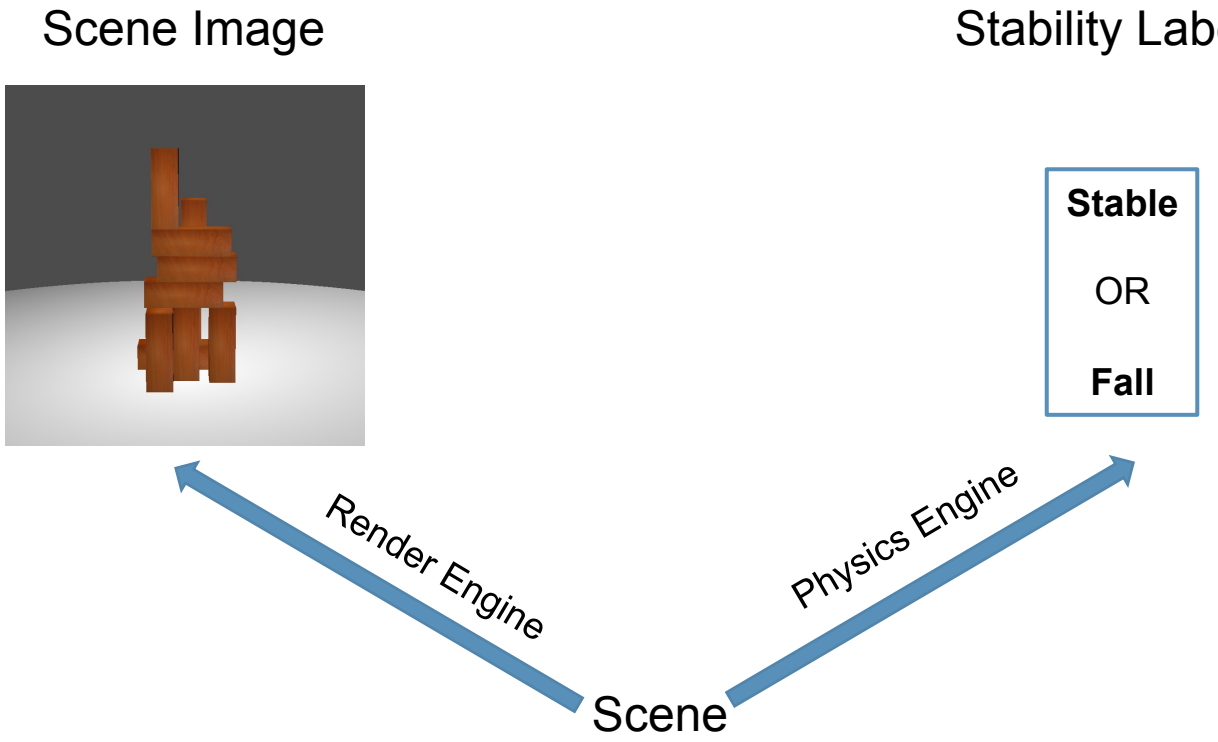
- GAP network for Discriminative region for stability prediction



[B. Zhou et al, "Learning Deep Features for Discriminative Localization." CVPR, 2016]

Data

- Blocks on the table simulated with physics engine

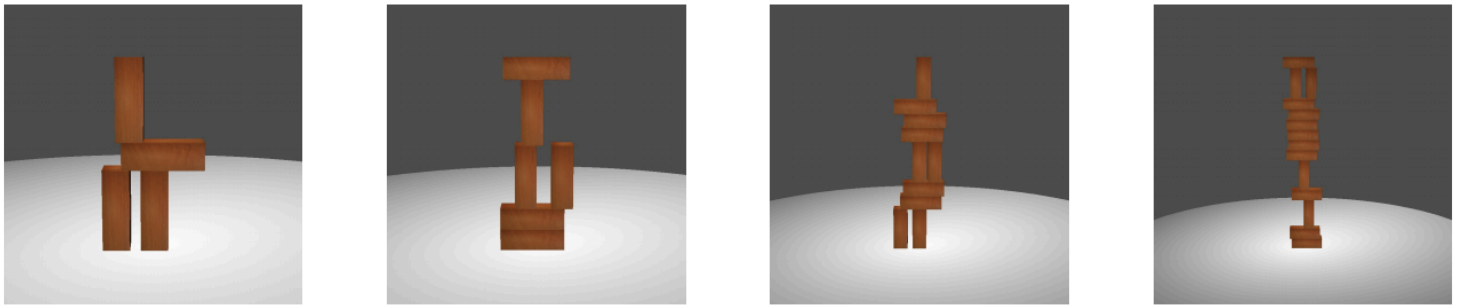


[P. W. Battaglia, et al, “Simulation as an engine of physical scene understanding,” PNAS, 2013]

Data

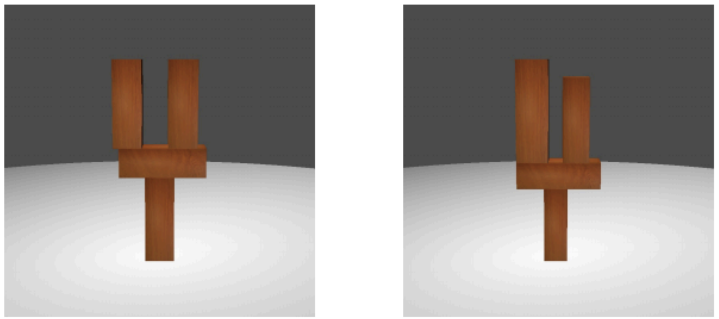
- Scene Parameters: 16 groups, 1K scenes/group

Block Number



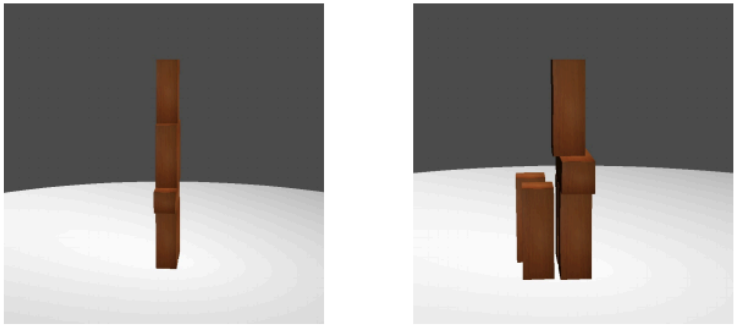
4 Blocks | 6 Blocks | 10 Blocks | 14 Blocks

Block Size



Uniform Size | Non-Uniform Size

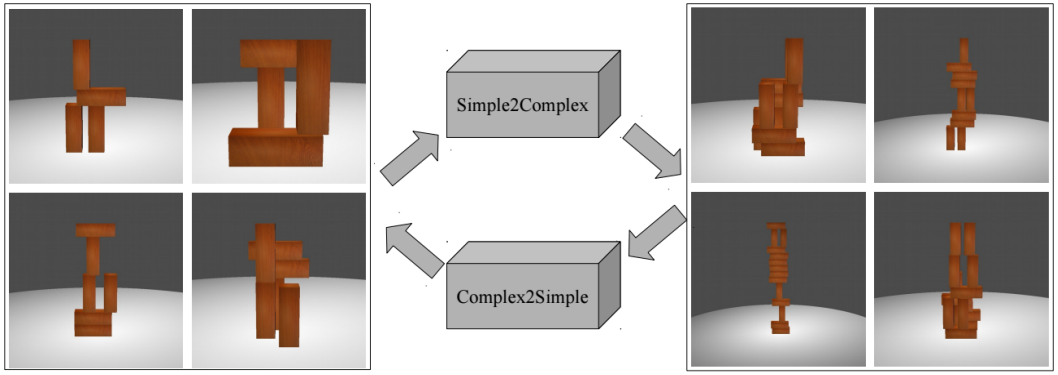
Stacking Depth



2D Stacking | 3D Stacking

Experiment

- Synthetic data -- 3 Groups of experiments
  - **Intra-Group**: Train and test on the scenes with the same scene parameters
  - **Cross-Group**: Train and test on the scenes with different scene parameters
  - **Generalization**: Train a global model and test on the scenes with different scene parameters



Experiment on Synthetic Data

- Recognition rate w.r.t. number of blocks

Num.of Blks	Uni.		NonUni.
	$2D$	$3D$	$2D$
$4B$	93.0	99.2	93.2
$6B$	88.8	91.6	88.0
$10B$	76.4	68.4	69.8
$14B$	71.2	57.0	74.8

Intra-Group

Num.of Blks	Uni.		NonUni.	
	$2D$	$3D$	$2D$	$3D$
4B	93.2	99.0	95.4	99.8
6B	89.0	94.8	87.8	93.0
10B	83.4	76.0	77.2	74.8
14B	82.4	67.2	78.4	66.2

Generalization

Experiment on Synthetic Data

- Recognition rate w.r.t. stacking depth

Num.of Blks	Uni.		NonUni.
	2D	3D	2D
4B	93.0	99.2	93.2
6B	88.8	91.6	88.0
10B	76.4	68.4	69.8
14B	71.2	57.0	74.8

Intra-Group

Num.of Blks	Uni.		NonUni.	
	2D	3D	2D	3D
4B	93.2	99.0	95.4	99.8
6B	89.0	94.8	87.8	93.0
10B	83.4	76.0	77.2	74.8
14B	82.4	67.2	78.4	66.2

Generalization

Experiment on Synthetic Data

- Recognition rate w.r.t. block size

Num.of Blks	Uni.		NonUni.
	$2D$	$3D$	$2D$
$4B$	93.0	99.2	93.2
$6B$	88.8	91.6	88.0
$10B$	76.4	68.4	69.8
$14B$	71.2	57.0	74.8

Intra-Group

Num.of Blks	Uni.		NonUni.	
	$2D$	$3D$	$2D$	$3D$
4B	93.2	99.0	95.4	99.8
6B	89.0	94.8	87.8	93.0
10B	83.4	76.0	77.2	74.8
14B	82.4	67.2	78.4	66.2

Generalization

Experiment on Synthetic Data

- Generalization test

Num.of Blks	Uni.		NonUni.
	$2D$	$3D$	$2D$
$4B$	93.0	99.2	93.2
$6B$	88.8	91.6	88.0
$10B$	76.4	68.4	69.8
$14B$	71.2	57.0	74.8

Intra-Group

Num.of Blks	Uni.		NonUni.	
	$2D$	$3D$	$2D$	$3D$
4B	93.2	99.0	95.4	99.8
6B	89.0	94.8	87.8	93.0
10B	83.4	76.0	77.2	74.8
14B	82.4	67.2	78.4	66.2

Generalization

Setting	Simple → Complex	Complex → Simple
Accuracy (%)	69.9	86.9

Cross-Group

Human Subject Study

- Settings
  - Sample a subset of test data to 8 subjects
  - Each given 96 images across all 16 scenes groups
  - For each scene, a rating from 1-5 is required
  - Compare the human's performance vs our model

Definitely Unstable	Probably Unstable	Cannot Tell	Probably Stable	Definitely Stable
1	2	3	4	5

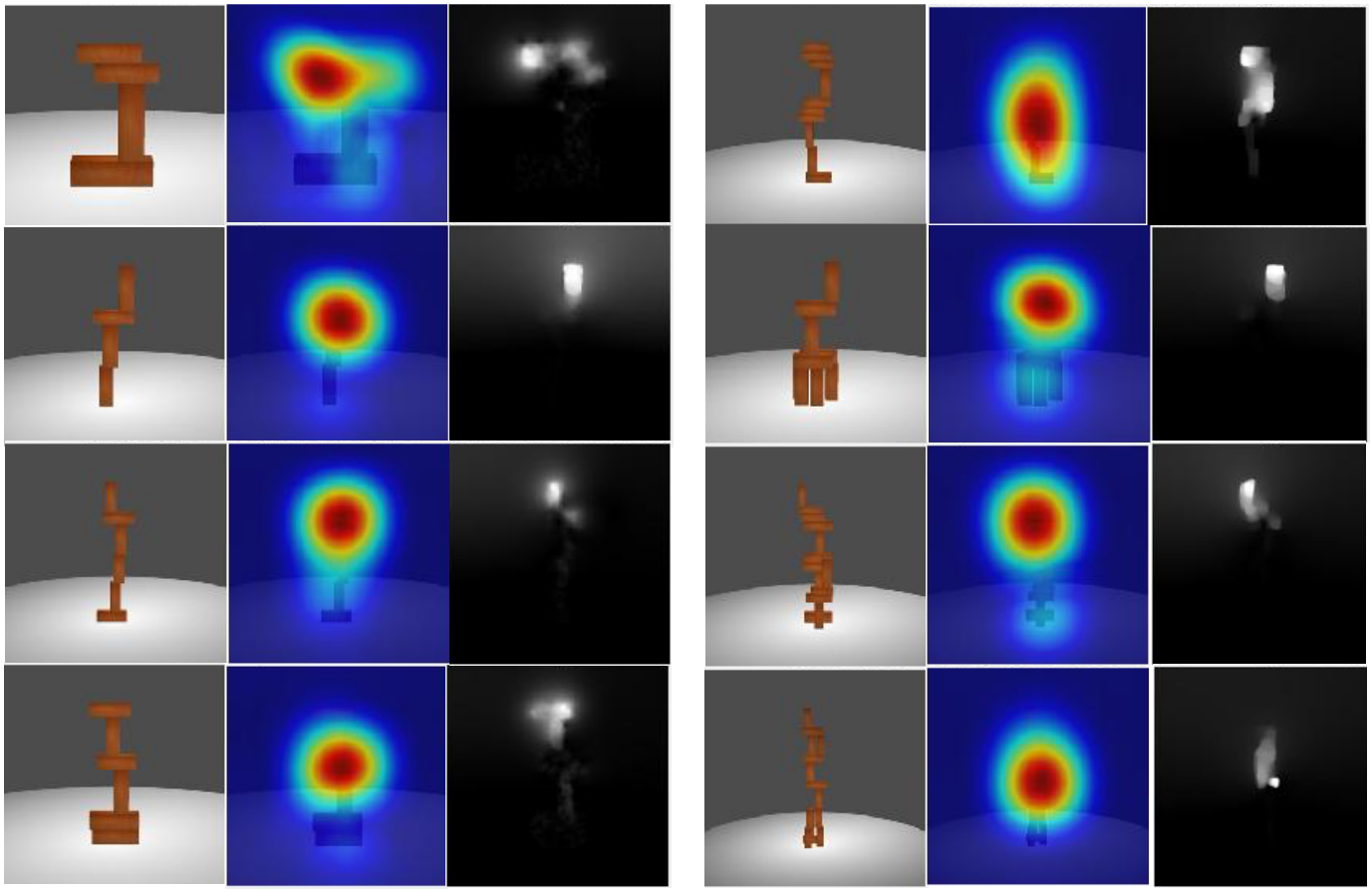
Human Subject Study

- Settings
  - Sample a subset of test data to 8 subjects
  - Each given 96 images across all 16 scenes groups
  - For each scene, a rating from 1-5 is required
  - Compare the human's performance vs our model

Num.of Blks	Uni.		NonUni.	
	$2D$	$3D$	$2D$	$3D$
4B	79.1/**91.7**	93.8/**100.0**	72.9/**93.8**	92.7/**100.0**
6B	78.1/**91.7**	83.3/**93.8**	71.9/**87.5**	89.6/**93.8**
10B	67.7/**87.5**	72.9/72.9	66.7/**72.9**	71.9/68.8
14B	71.9/**79.2**	68.8/66.7	**71.9**/**81.3**	59.3/**60.4**

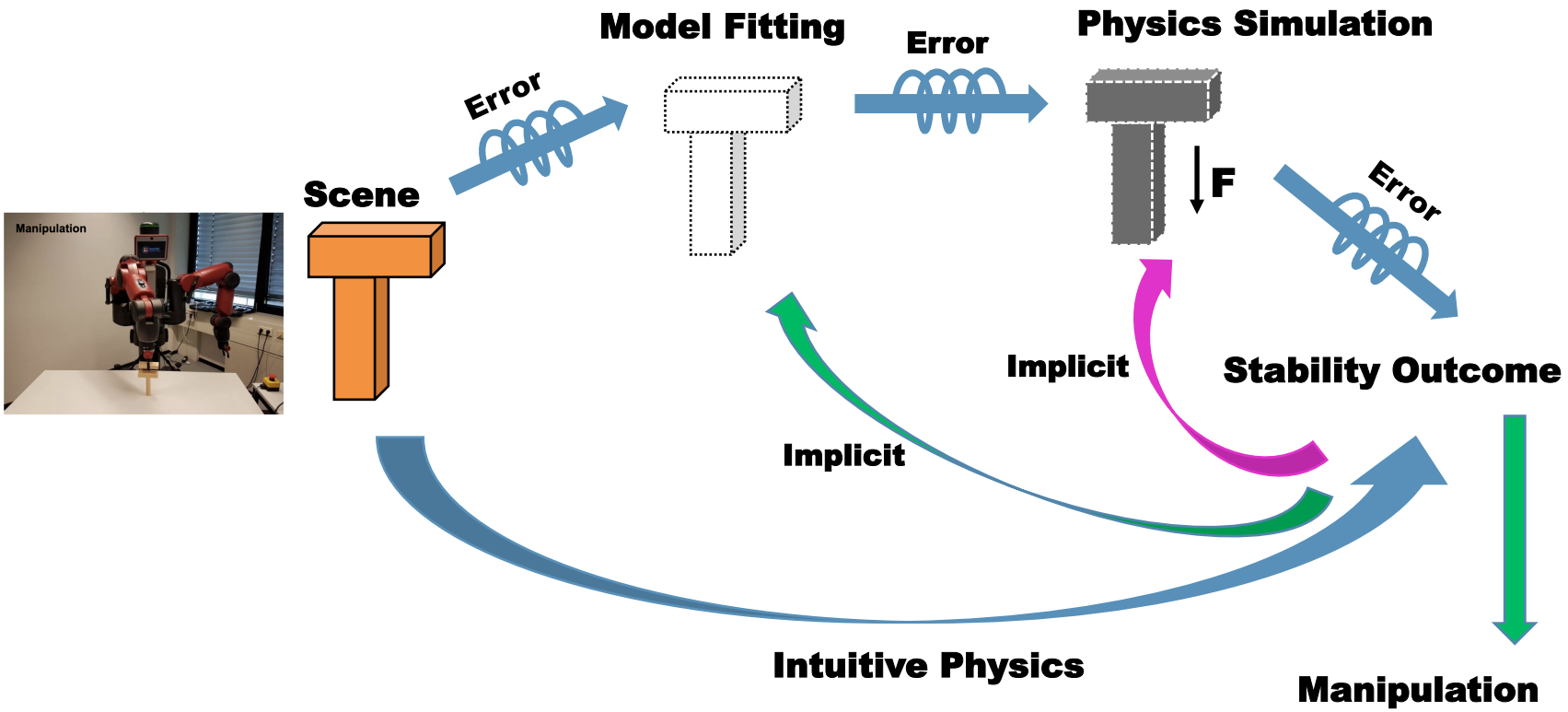
For entry a/b, human result a, image-based prediction b

Model Interpretation



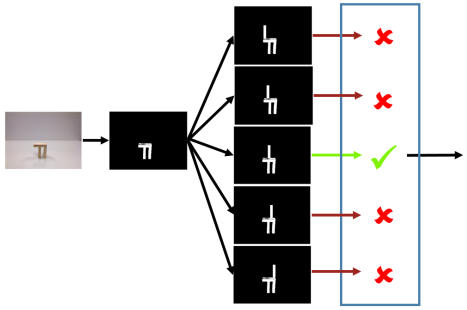
Scene GAP Flow Scene GAP Flow

Acquiring Intuitive Physics

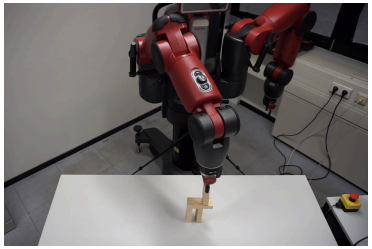


II. Manipulation Guided by Stability Prediction

- Approach: Integrate stability prediction into manipulation

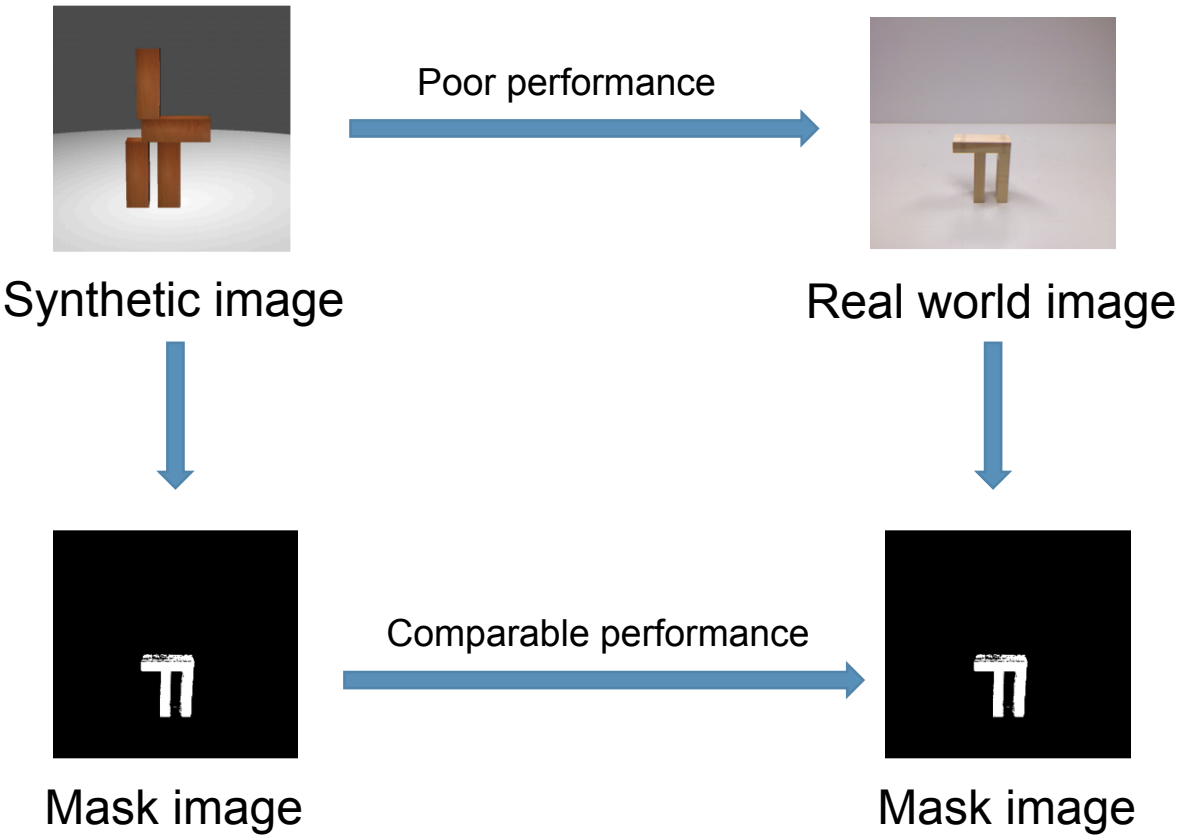


- Experiment: Robot stack block

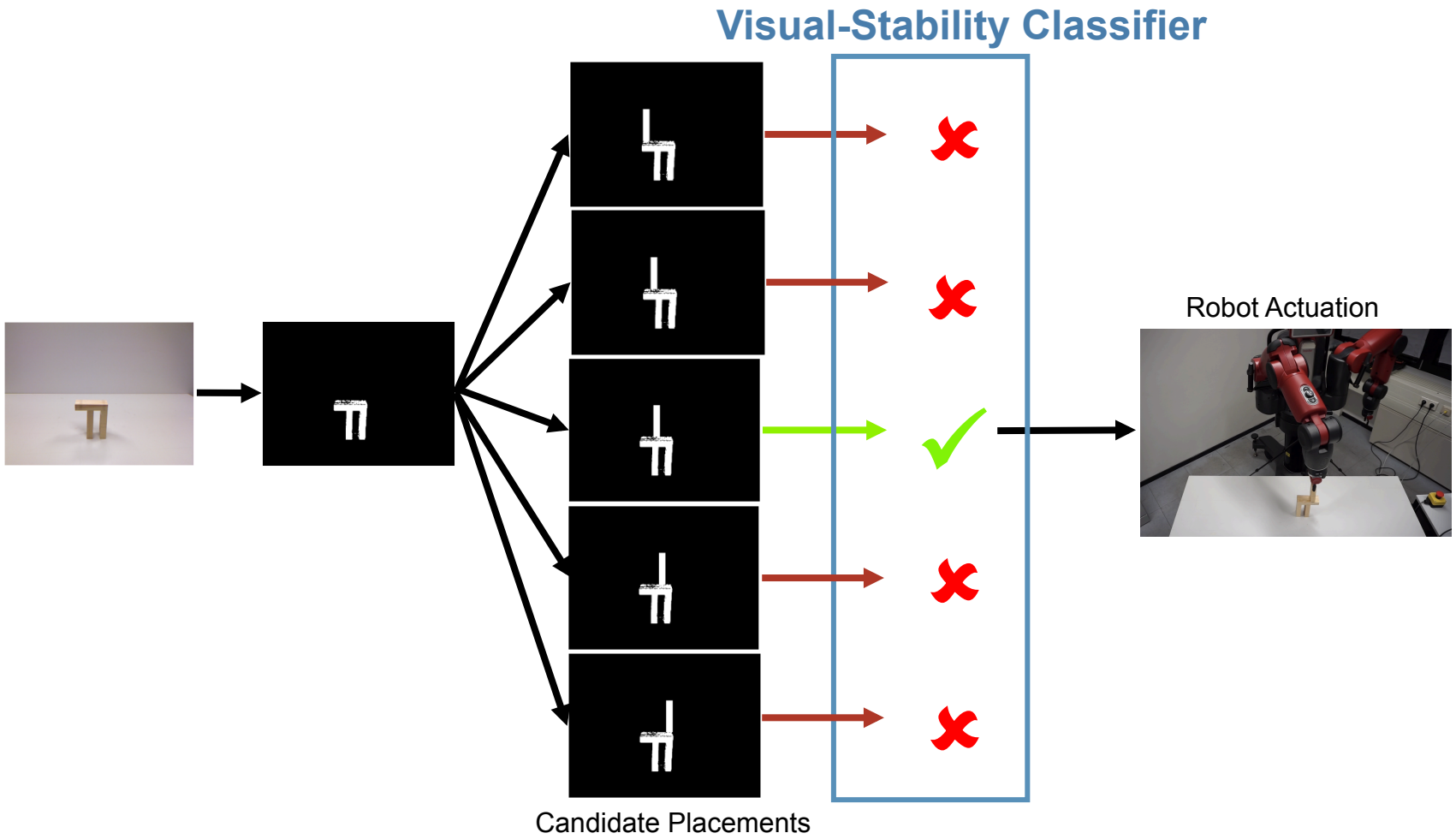


Approach

- Difference in appearance between real world data and synthetic data

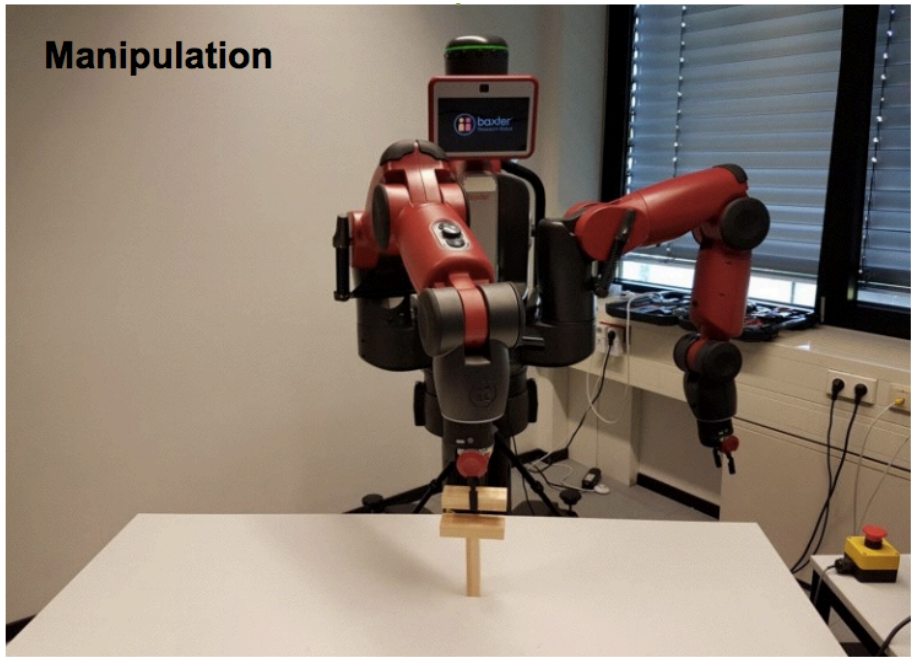


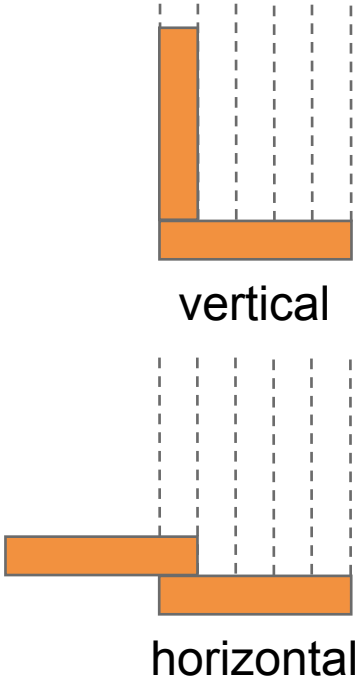
Approach: Integrating Visual Stability Prediction into Manipulation



Experiment

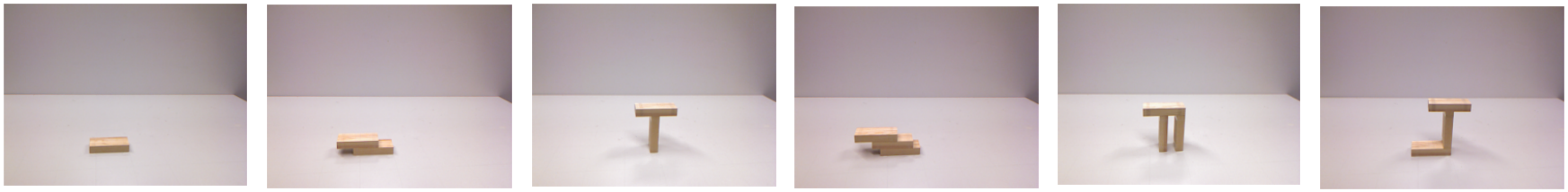
- Set up
  - Robot places a block on a give structure without breaking its stability





Experiment

- Prediction Rate:
  - Accuracy for stability prediction, 78.6% over all the scenes
- Manipulation Rate:
  - *# robot predict stable ∩ successfully put a block/# stable configurations*



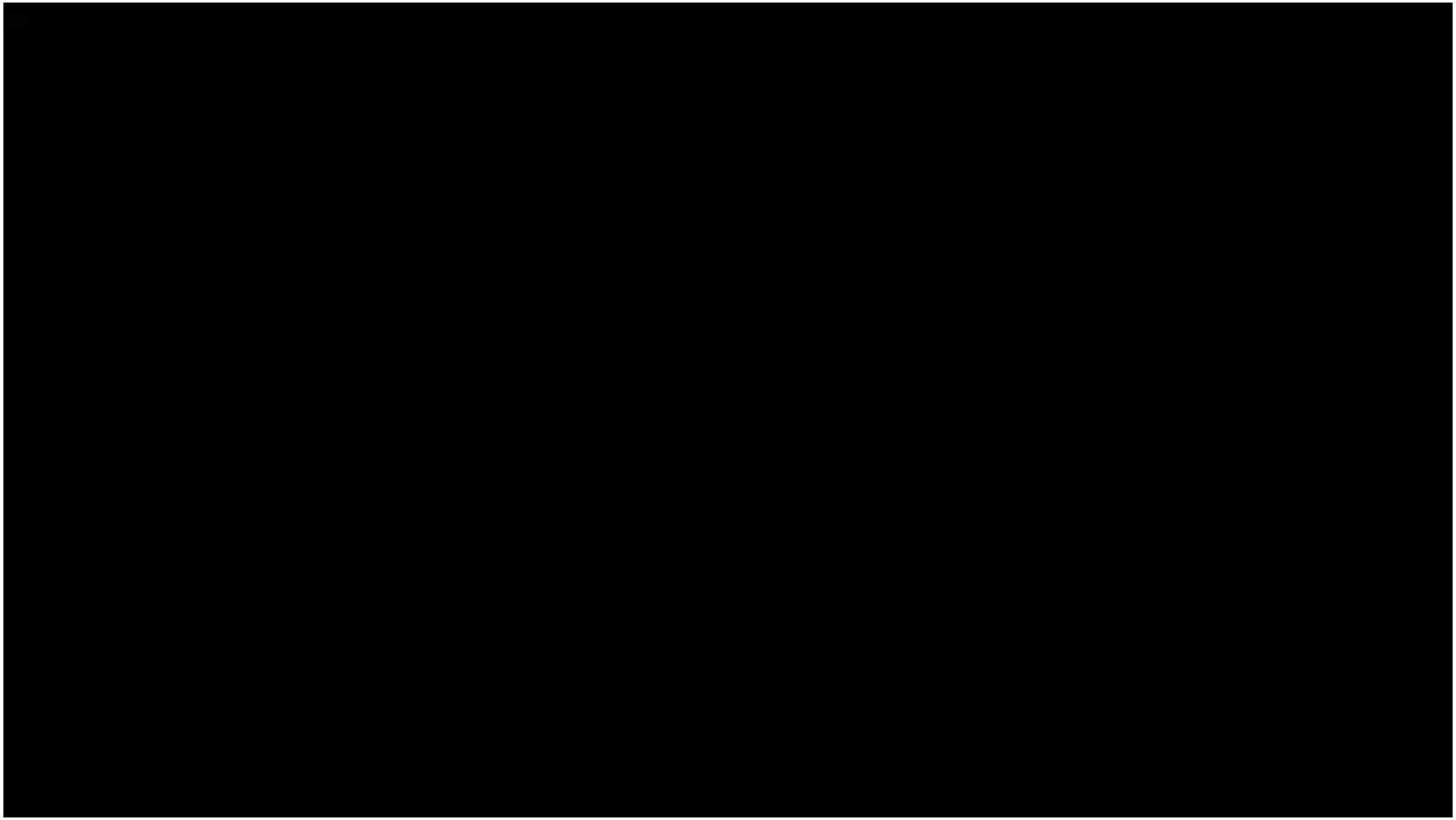
Horizontal Placement

Pred.	66.7	66.7	88.9	77.8	100.0	66.7
Mani.	80.0 (4/5)	66.7 (2/3)	66.7 (2/3)	66.7 (2/3)	100.0 (3/3)	0.0 (0/3)

Vertical Placement

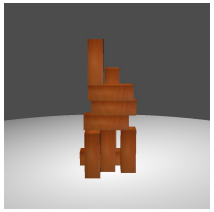
Pred.	100.0	60.0	100.0	80.0	40.0	60.0
Mani.	100.0 (5/5)	100.0 (3/3)	100.0 (1/1)	66.7 (2/3)	25.0 (1/4)	0.0 (0/1)

Result



Conclusion

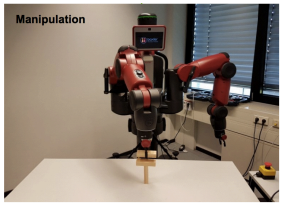
- End-to-end learning for intuitive physics
- Integrate visual stability prediction into manipulation

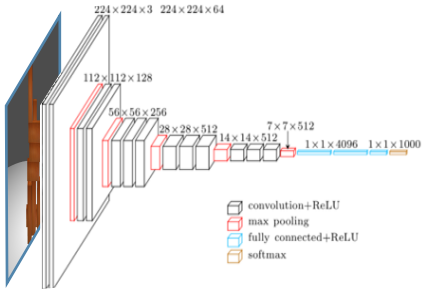


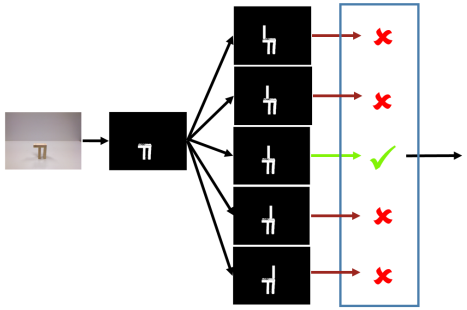
Intuitive Physics

Visual Stability Prediction

Manipulation







- **Questions?**