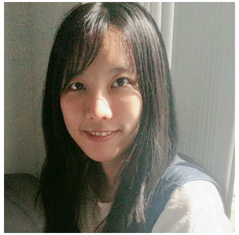
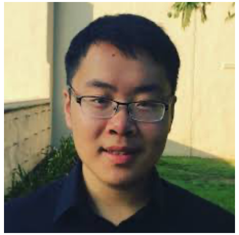
Visual Grounding of Learned Physical Models

ICML 2020

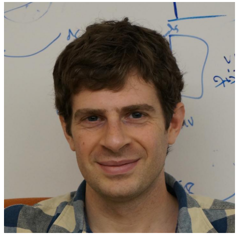
Yunzhu Li



Toru Lin*

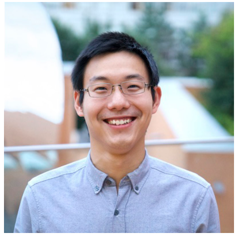


Kexin Yi*



Daniel M. Bear

Daniel L.K. Yamins



Jiajun Wu

Joshua B. Tenenbaum

Antonio Torralba

http://visual-physics-grounding.csail.mit.edu/

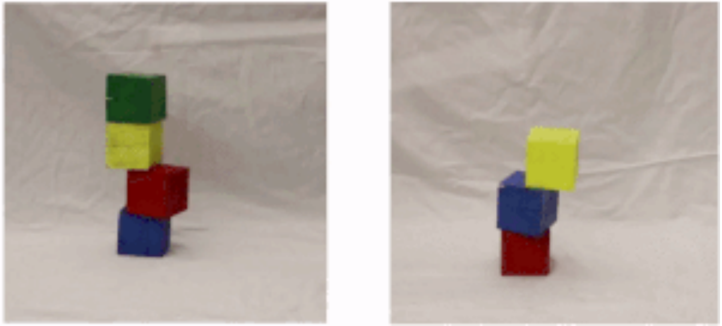
(* indicates equal contribution)

Intuitive Physics

(1) Distinguish between different instances

(2) Recognize objects' physical properties

(3) Predict future movements



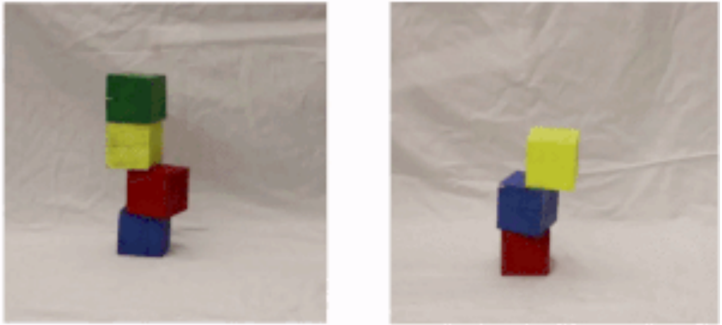
(Wu et al., Learning to See Physics via Visual De-animation)

Intuitive Physics

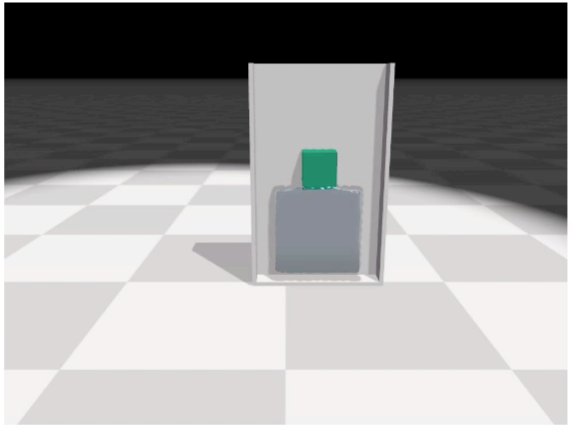
(1) Distinguish between different instances

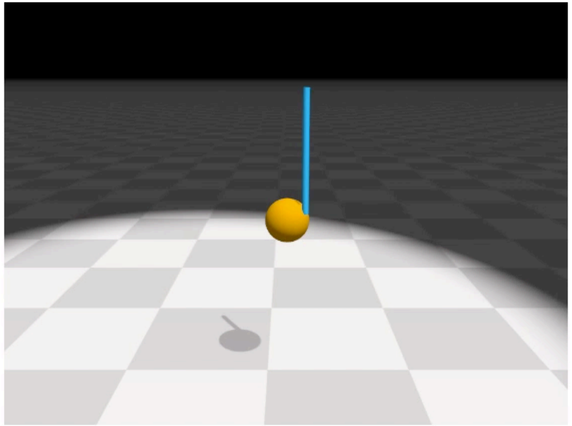
(2) Recognize objects' physical properties

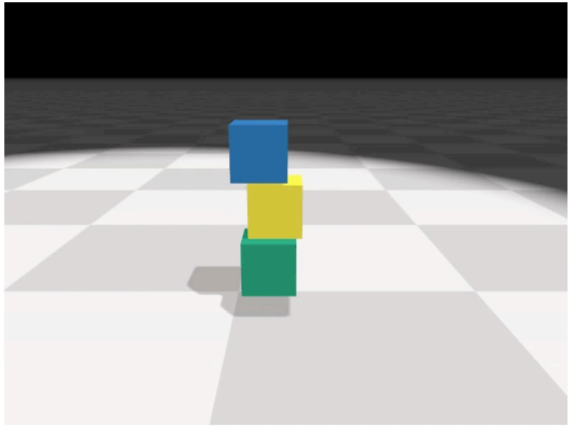
(3) Predict future movements



(Wu et al., Learning to See Physics via Visual De-animation)





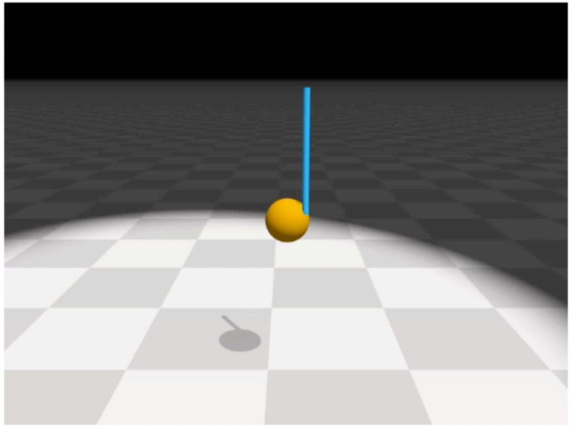


For example

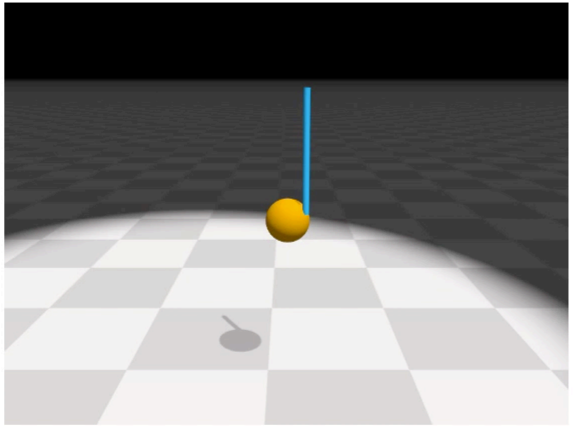
Different physical parameters lead to different motions.

Estimating physical parameter by comparing mental simulation with observation

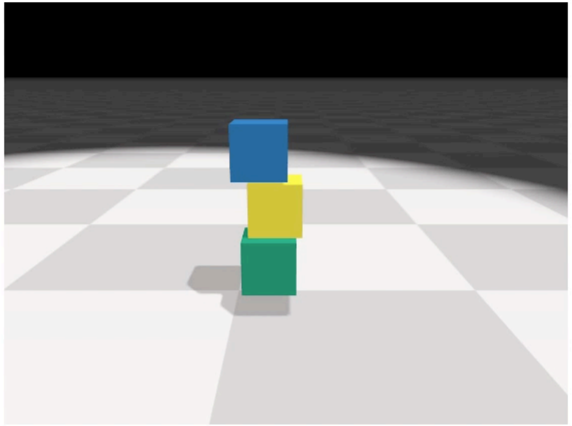
Larger stiffness



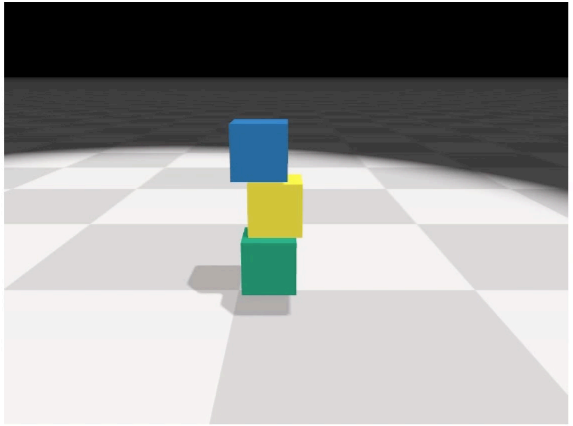
Smaller stiffness

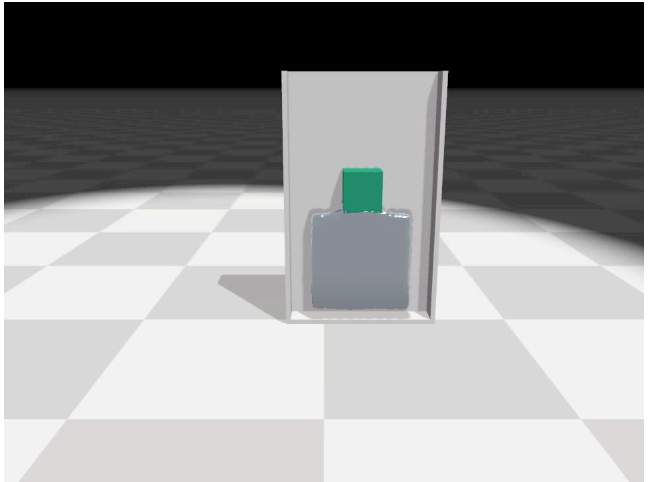


Larger gravity

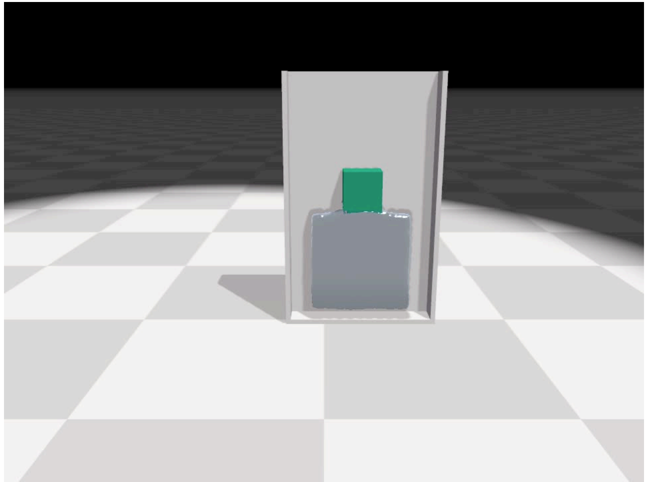


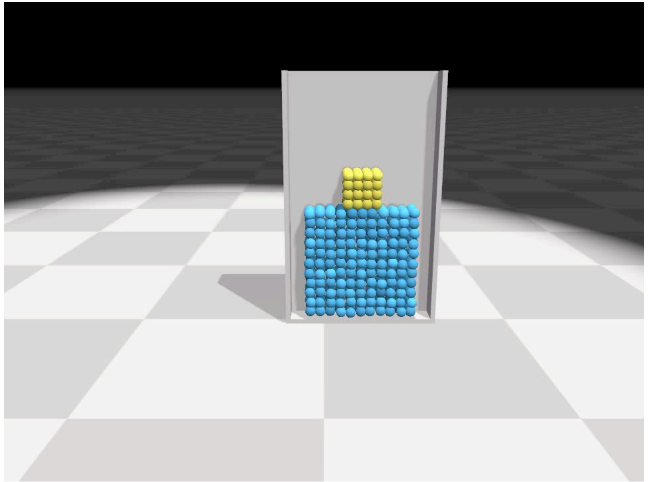
Smaller gravity





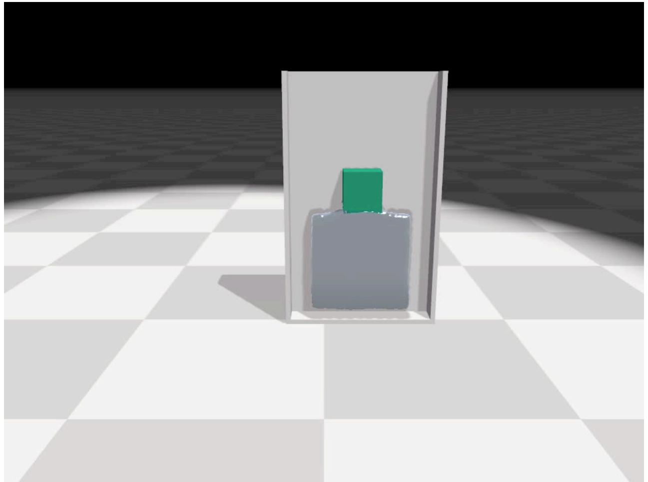
Physical reasoning of deformable objects is challenging.





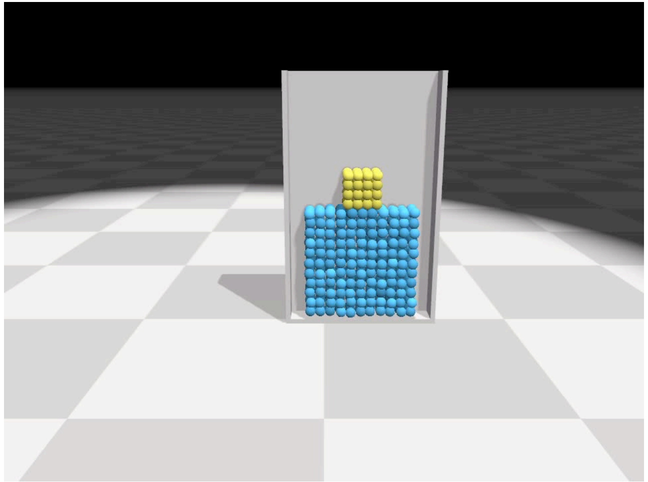
Physical reasoning of deformable objects is challenging.

Particle-based Representation
General & Flexible



Physical reasoning of deformable objects is challenging.

Particle-based Representation
General & Flexible



We propose a model that jointly

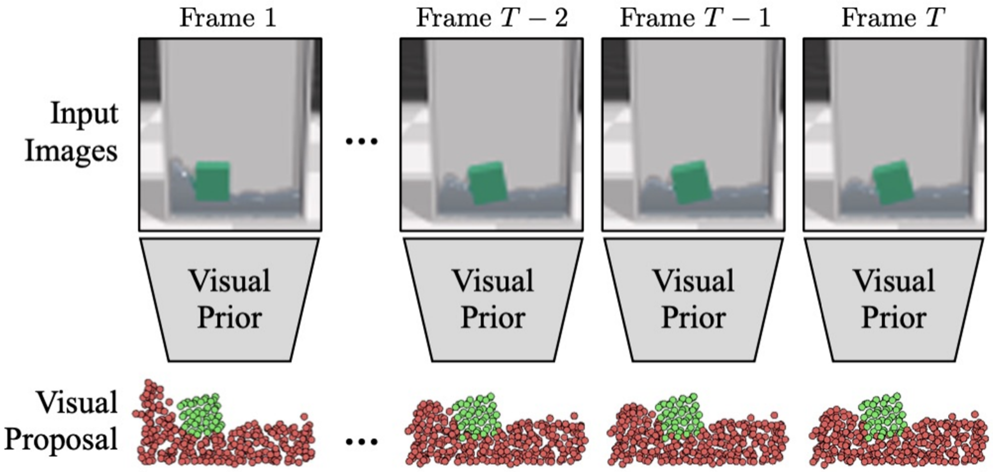
(1) Estimates the physical properties
(2) Refines the particle locations

using

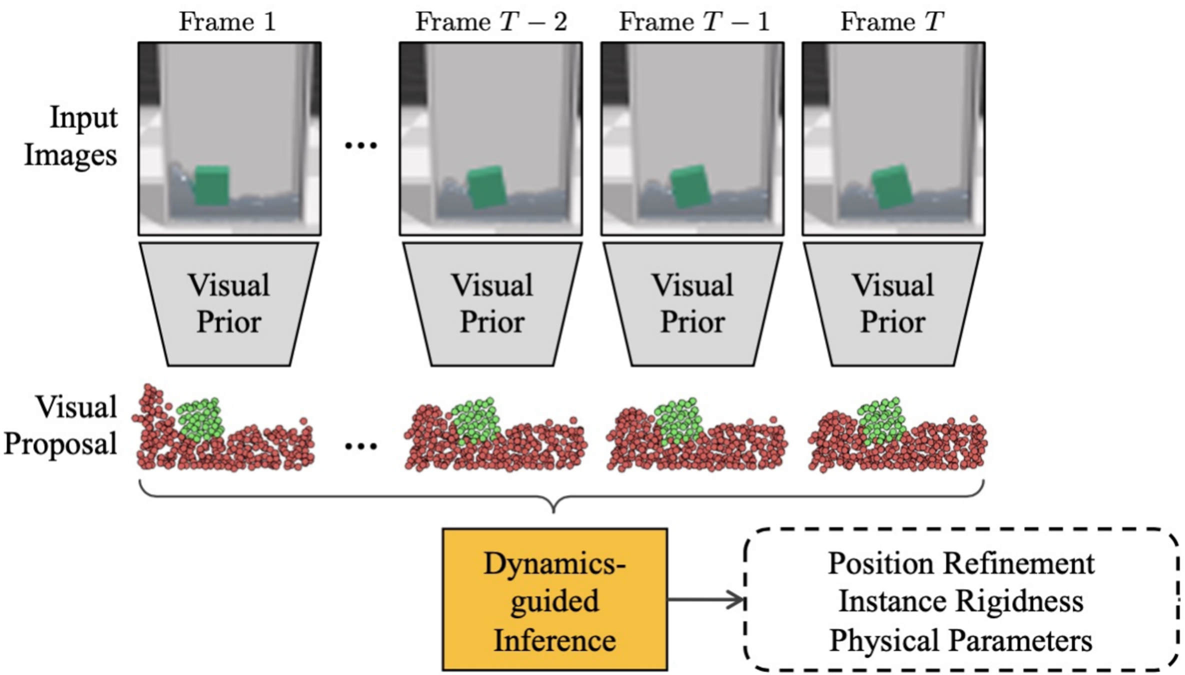
(1) a learned visual prior
(2) a learned dynamics prior

Visually Grounded Physics Learner (VGPL)

Visually Grounded Physics Learner (VGPL)

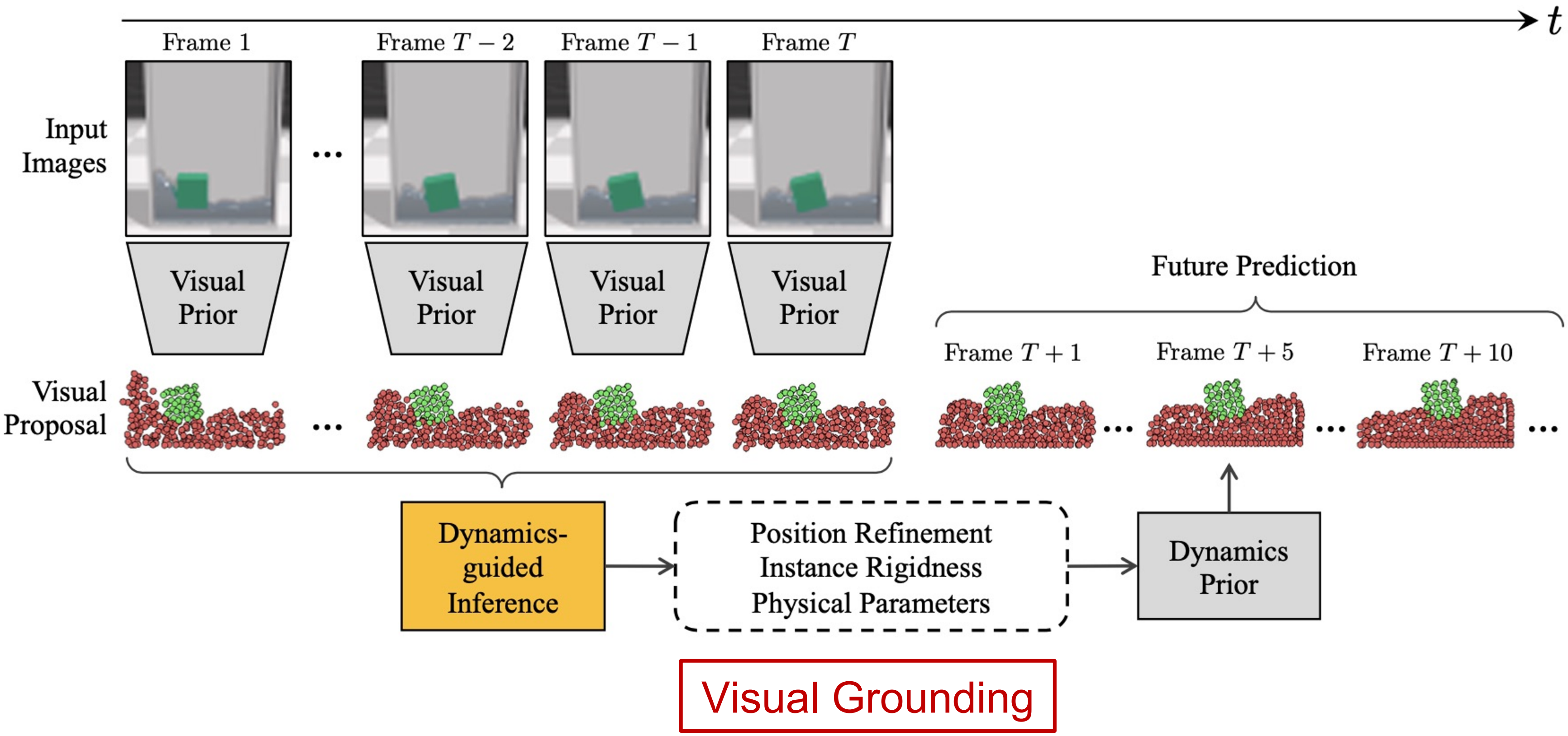


Visually Grounded Physics Learner (VGPL)

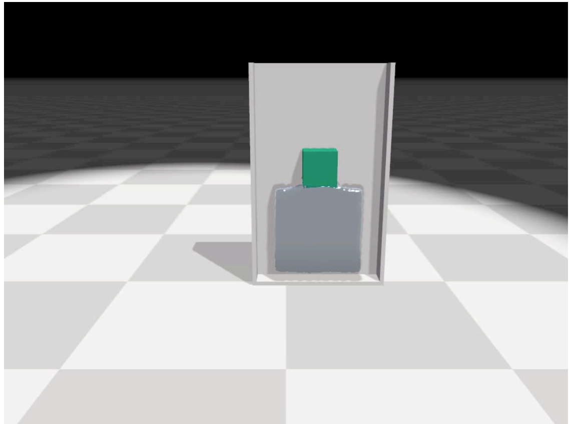
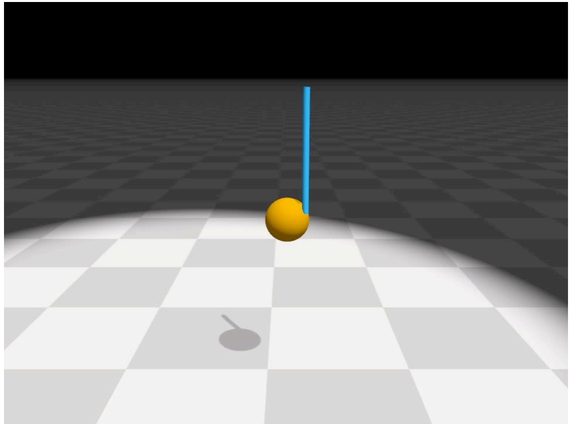
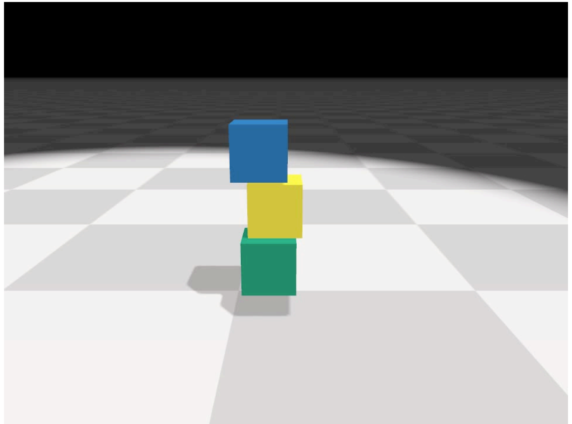


Visual Grounding

Visually Grounded Physics Learner (VGPL)



We evaluate our model in environments involving interactions between rigid objects, elastic materials, and fluids.

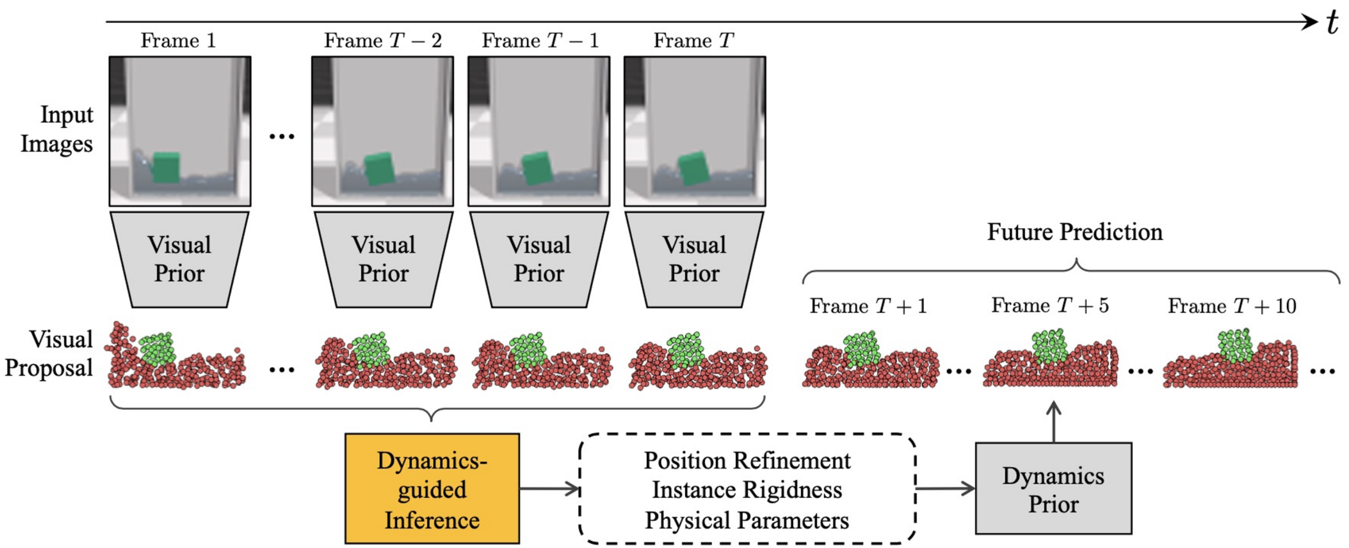




We evaluate our model in environments involving interactions between rigid objects, elastic materials, and fluids.

Within a few observation steps, our model is able to

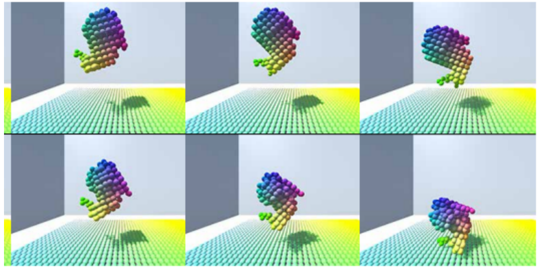
(1) refine the state estimation and reason about the physical properties

(2) make predictions into the future.

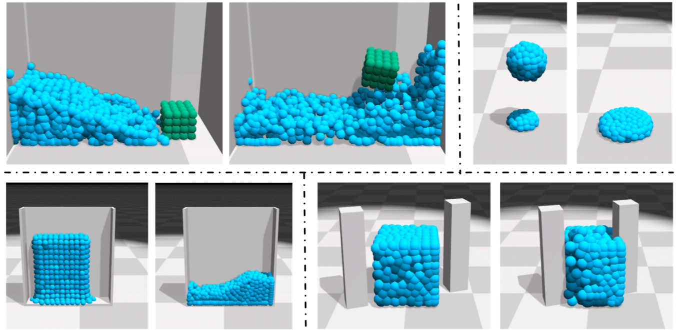


Related Work

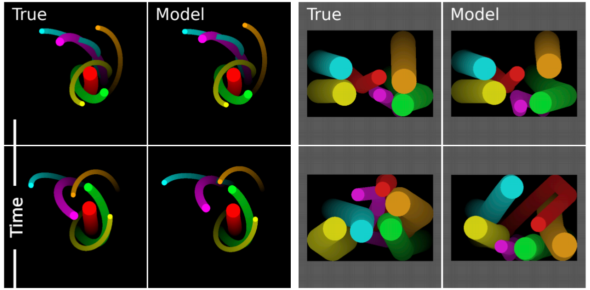
Learning-based particle dynamics



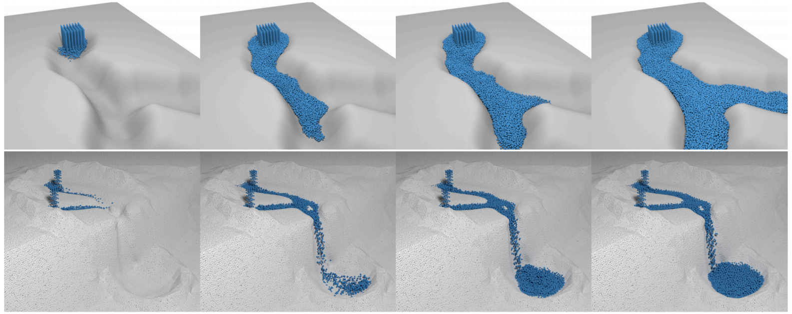
Mrowca, Zhuang, Wang, Haber, Fei-Fei, Tenenbaum, Yamins. NeurIPS'18



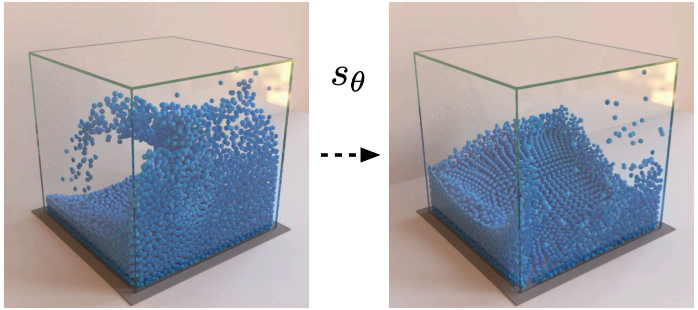
Li, Wu, Tedrake, Tenenbaum, Torralba. ICLR'19



Battaglia, Pascanu, Lai, Rezende, Kavukcuoglu. NeurIPS'16



Ummenhofer, Prantl, Thuerey, Koltun. ICLR'20

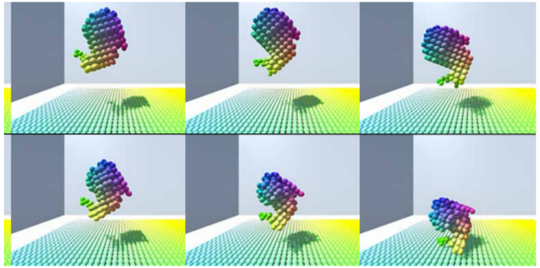


Sanchez-Gonzalez, Godwin, Pfaff, Ying, Leskovec, Battaglia. ICML'20

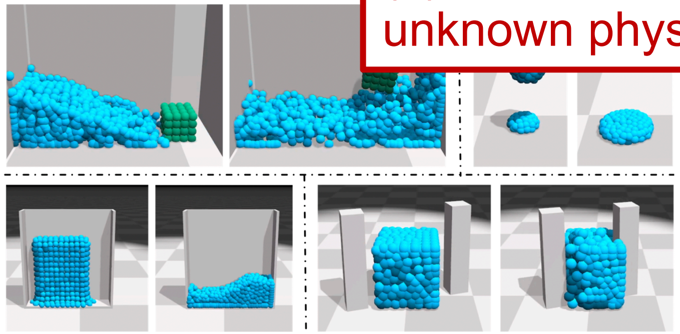
Related Work

Learning-based particle dynamics

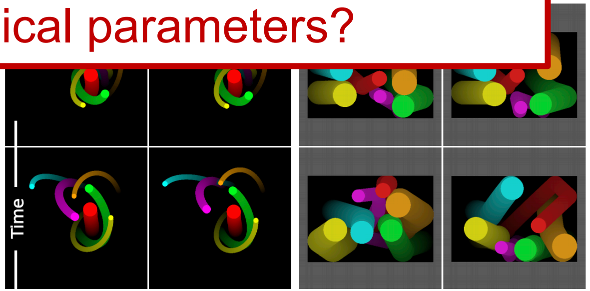
Questions remains:
(1) How well they handle visual inputs?
(2) How to adapt to scenarios of unknown physical parameters?



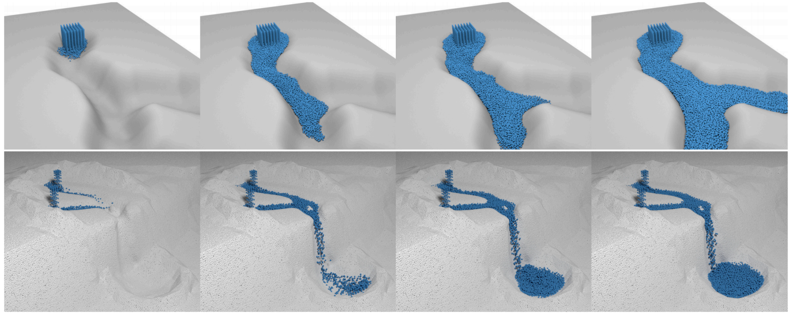
Mrowca, Zhuang, Wang, Haber, Fei-Fei, Tenenbaum, Yamins. NeurIPS'18



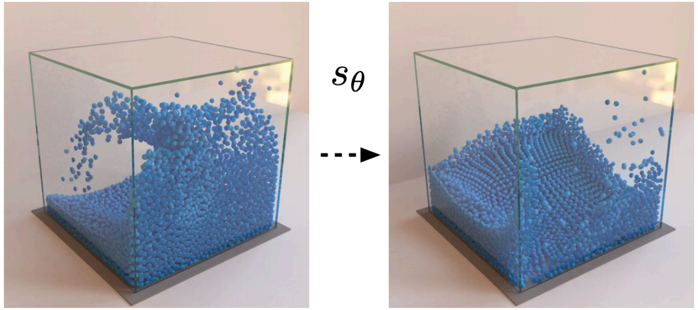
Li, Wu, Tedrake, Tenenbaum, Torralba. ICLR'19



Battaglia, Pascanu, Lai, Rezende, Kavukcuoglu. NeurIPS'16



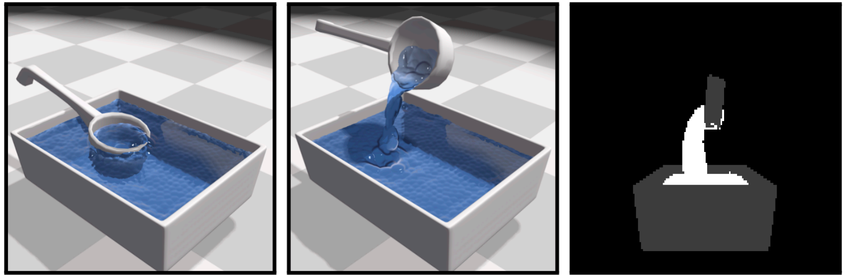
Ummenhofer, Prantl, Thuerey, Koltun. ICLR'20



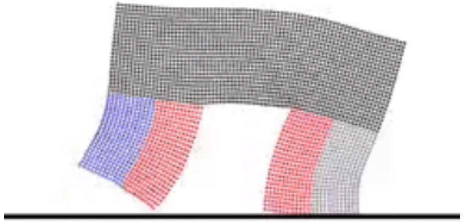
Sanchez-Gonzalez, Godwin, Pfaff, Ying, Leskovec, Battaglia. ICML'20

Related Work

Differentiating through physics-based simulators

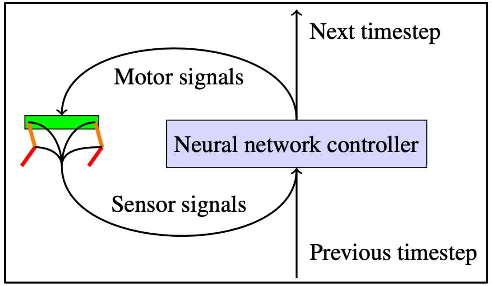


Schenck, Fox. CoRL’18

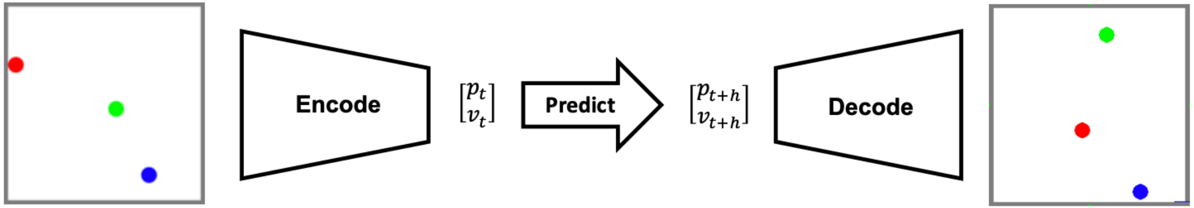


Hu, Liu, Spielberg, Tenenbaum, Freeman, Wu, Rus, Matusik. ICRA’19

Liang, Lin, Koltun. NeurIPS’19



Degrave, Hermans, Dambre, Wyffels. Frontiers in Neurorobotics 2019

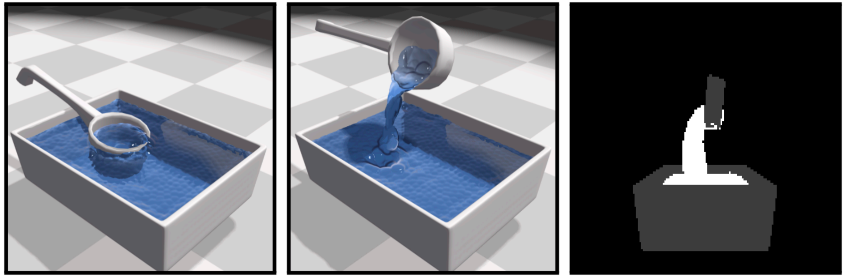


Belbute-Peres, Smith, Allen, Tenenbaum, Kolter. NeurIPS’18

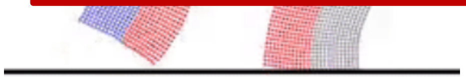
Related Work

Differentiating through physics-based simu

Questions remains:
(1) Make strong assumptions on the structure of the system
(2) Usually time-consuming
(2) Prone to local optimum
(3) Lacking ways to handle visual inputs

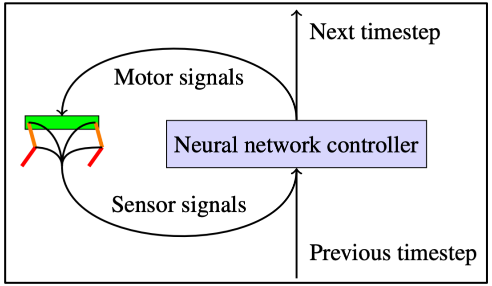


Schenck, Fox. CoRL’18

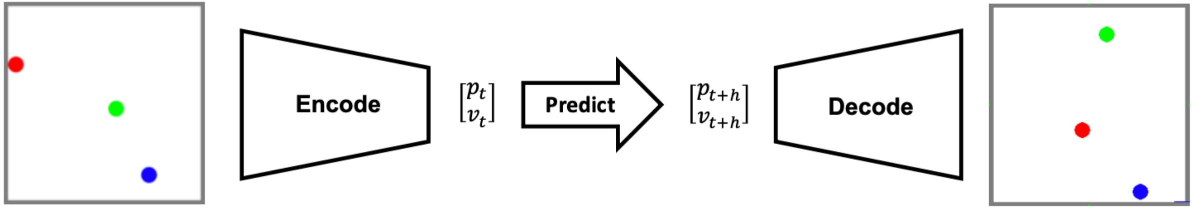


Hu, Liu, Spielberg, Tenenbaum, Freeman, Wu, Rus, Matusik. ICRA’19

Liang, Lin, Koltun. NeurIPS’19



Degrave, Hermans, Dambre, Wyffels. Frontiers in Neurorobotics 2019



Belbute-Peres, Smith, Allen, Tenenbaum, Kolter. NeurIPS’18

Our Work

We proposed **Visually Grounded Physics Learner (VGPL)** to

(1) bridge the perception gap,

(2) enable physical reasoning from visual perception, and

(3) perform dynamics-guided inference to directly predict the optimization results,

which allows quick adaptation to environments with unknown physical properties.

Problem Formulation

Consider a system that contains M objects and N particles.

Problem Formulation

Consider a system that contains M objects and N particles.

$O = \{o^t\}_{t=1}^{T}$: Visual observ.

Problem Formulation

Consider a system that contains M objects and N particles.

Visual prior f_V

$$(\hat{X}', \hat{G}) = f_V(O)$$

$O = \{o^t\}_{t=1}^{T}$: Visual observ.

$\hat{X}'$: Particle position

$\hat{G}$: Instance grouping

Problem Formulation

Consider a system that contains M objects and N particles.

Visual prior f_V Dynamics prior f_D

$$(\hat{X}', \hat{G}) = f_V(O)$$

$$\hat{X}^{T+1} = f_D(\hat{X}, \hat{G}, \quad , \quad)$$

$O = \{o^t\}_{t=1}^{T}$: Visual observ.

$\hat{X}'$: Particle position

$\hat{G}$: Instance grouping

Problem Formulation

Consider a system that contains M objects and N particles.

Visual prior f_V Dynamics prior f_D

$$(\hat{X}', \hat{G}) = f_V(O)$$

$$\hat{X}^{T+1} = f_D(\hat{X}, \hat{G}, \quad , \hat{Q})$$

$O = \{o^t\}_{t=1}^{T}$: Visual observ.

$\hat{X}'$: Particle position

$\hat{G}$: Instance grouping

$\hat{Q}$: Rigidness of each instance

Problem Formulation

Consider a system that contains M objects and N particles.

Visual prior f_V Dynamics prior f_D

$$(\hat{X}', \hat{G}) = f_V(O)$$

$$\hat{X}^{T+1} = f_D(\hat{X}, \hat{G}, \hat{P}, \hat{Q})$$

$O = \{o^t\}_{t=1}^{T}$: Visual observ.

$\hat{X}'$: Particle position

$\hat{G}$: Instance grouping

$\hat{Q}$: Rigidness of each instance

$\hat{P}$: Physical parameters

Problem Formulation

Consider a system that contains M objects and N particles.

Visual prior f_V Dynamics prior f_D Inference module f_I

$$(\hat{X}', \hat{G}) = f_V(O)$$

$$\hat{X}^{T+1} = f_D(\hat{X}, \hat{G}, \hat{P}, \hat{Q})$$

$$(\hat{P}, \hat{Q}, \quad\quad) = f_I(\hat{X}', \hat{G})$$

$O = \{o^t\}_{t=1}^{T}$: Visual observ.

$\hat{X}'$: Particle position

$\hat{G}$: Instance grouping

$\hat{Q}$: Rigidness of each instance

$\hat{P}$: Physical parameters

Problem Formulation

Consider a system that contains M objects and N particles.

Visual prior f_V Dynamics prior f_D Inference module f_I

$$(\hat{X}', \hat{G}) = f_V(O)$$

$$\hat{X}^{T+1} = f_D(\hat{X}, \hat{G}, \hat{P}, \hat{Q})$$

$$(\hat{P}, \hat{Q}, \Delta\hat{X}) = f_I(\hat{X}', \hat{G})$$

$$\hat{X} = \hat{X}' + \Delta\hat{X}$$

$O = \{o^t\}_{t=1}^{T}$: Visual observ.

$\hat{X}'$: Particle position

$\hat{G}$: Instance grouping

$\hat{Q}$: Rigidness of each instance

$\hat{P}$: Physical parameters

$\Delta\hat{X}$: Position refinement

Problem Formulation

Consider a system that contains M objects and N particles.

Visual prior f_V Dynamics prior f_D Inference module f_I

$$(\hat{X}', \hat{G}) = f_V(O)$$

$$\hat{X}^{T+1} = f_D(\hat{X}, \hat{G}, \hat{P}, \hat{Q})$$

$$(\hat{P}, \hat{Q}, \Delta\hat{X}) = f_I(\hat{X}', \hat{G})$$

$$\hat{X} = \hat{X}' + \Delta\hat{X}$$

Objective function

$$(\hat{P}^*, \hat{Q}^*, \Delta\hat{X}^*) = \arg\min_{\hat{P}, \hat{Q}, \Delta\hat{X}} \|\hat{X}^{T+1} - X^{T+1}\|$$

$O = \{o^t\}_{t=1}^T$: Visual observ.

$\hat{X}'$: Particle position

$\hat{G}$: Instance grouping

$\hat{Q}$: Rigidness of each instance

$\hat{P}$: Physical parameters

$\Delta\hat{X}$: Position refinement

Visual Prior f_V

$$(\hat{X}', \hat{G}) = f_V(O)$$

Visual observations : $O = \{o^t\}_{t=1}^{T}$

Visual Prior f_V

$$(\hat{X}', \hat{G}) = f_V(O)$$

Visual observations : $O = \{o^t\}_{t=1}^{T}$

Particle locations : $\hat{X}' = \{(x_i^{t\prime}, y_i^{t\prime}, z_i^{t\prime})\}_{i=1,t=1}^{N,T}$

Instance grouping : $\hat{G} = \{G_i^t\}_{i=1,t=1}^{N,T}$

Visual Prior f_V

$$(\hat{X}', \hat{G}) = f_V(O)$$

Visual observations : $O = \{o^t\}_{t=1}^{T}$

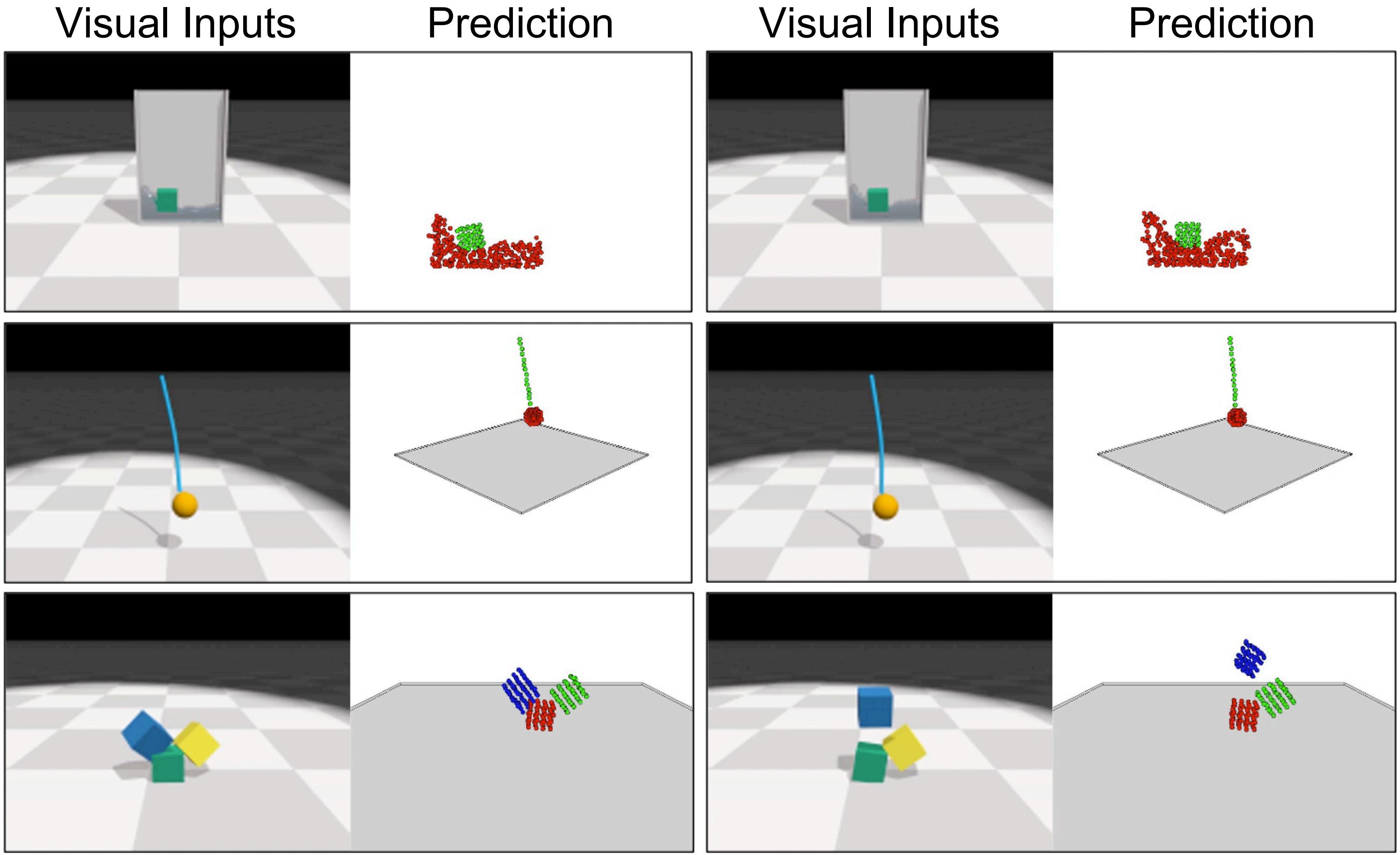
Particle locations : $\hat{X}' = \{(x_i^{t\prime}, y_i^{t\prime}, z_i^{t\prime})\}_{i=1,t=1}^{N,T}$

Instance grouping : $\hat{G} = \{G_i^t\}_{i=1,t=1}^{N,T}$

Objective function

$$\mathcal{L}_\mathcal{V} = \frac{1}{TN} \sum_{t=1,i=1}^{T,N} \left[\|\hat{X}_i^{t\prime} - X_i^t\|^2 + H(\hat{G}_i^t, G_i^t) \right]$$

Results of the Visual Prior f_V



Dynamics Prior f_D

$$\hat{X}^{T+1} = f_D(\hat{X}, \hat{G}, \quad , \quad).$$

$\hat{X}$: Particle position

$\hat{G}$: Instance grouping

Dynamics Prior f_D

$$\hat{X}^{T+1} = f_D(\hat{X}, \hat{G}, \hat{P}, \hat{Q}).$$

$\hat{X}$: Particle position

$\hat{G}$: Instance grouping

$\hat{Q}$: Rigidness of each instance

$\hat{P}$: Physical parameters

Dynamics Prior f_D

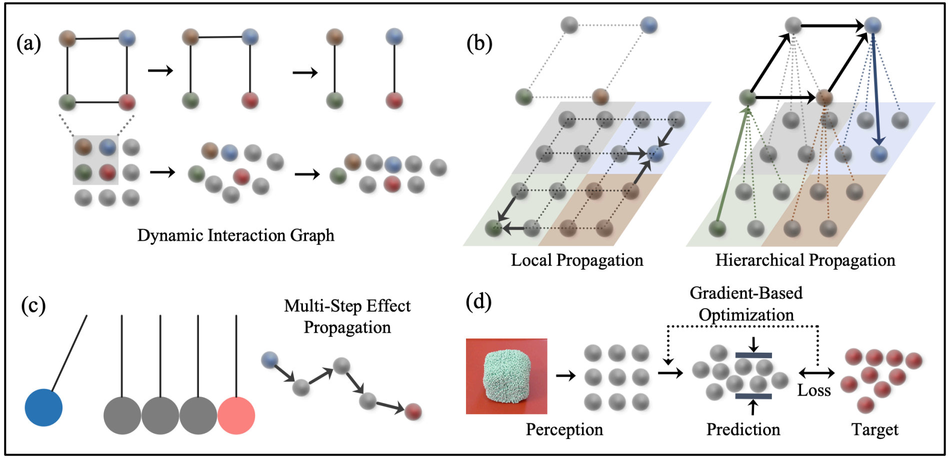
$$\hat{X}^{T+1} = f_D(\hat{X}, \hat{G}, \hat{P}, \hat{Q}).$$

$\hat{X}$: Particle position

$\hat{G}$: Instance grouping

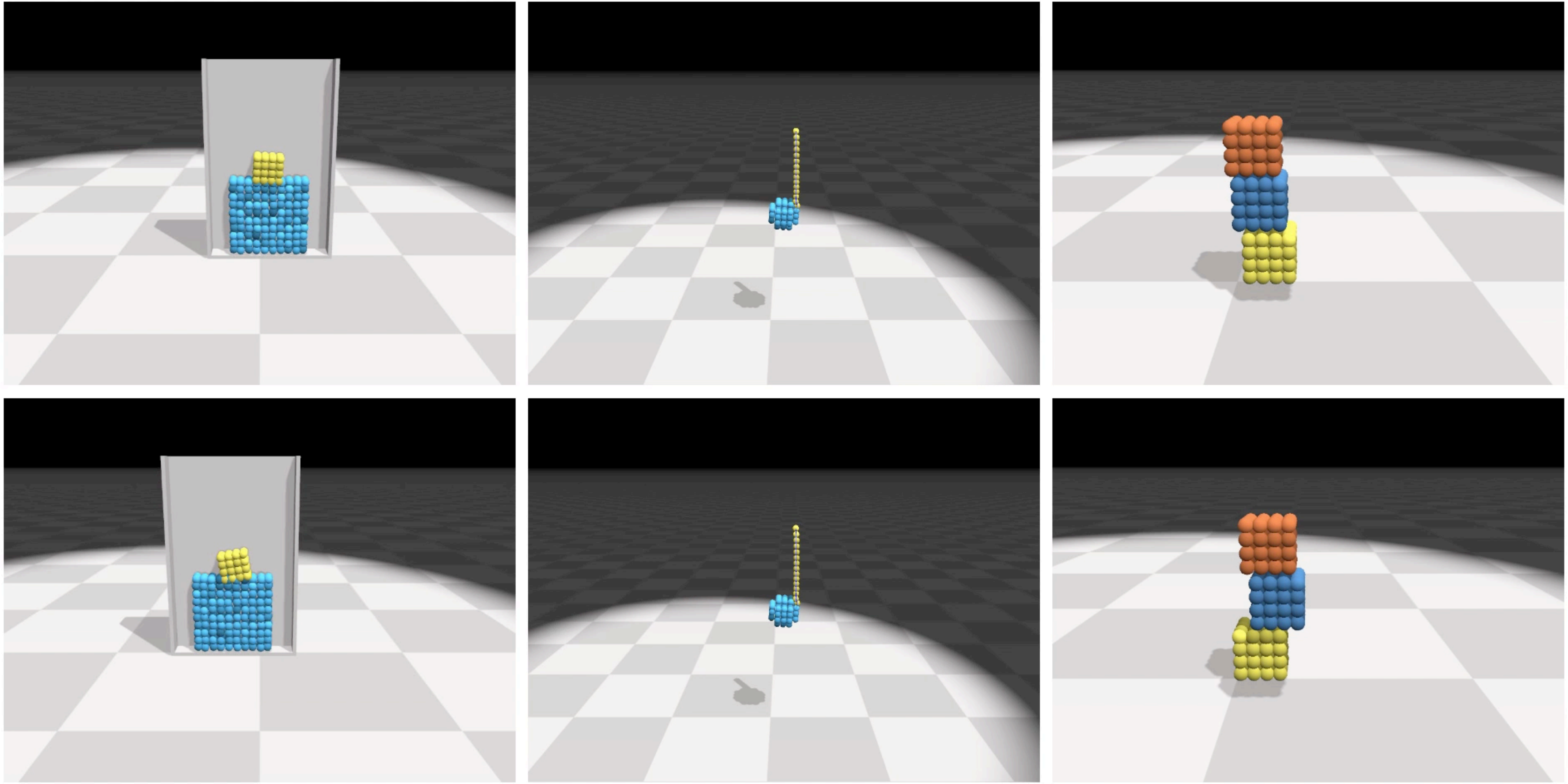
$\hat{Q}$: Rigidness of each instance

$\hat{P}$: Physical parameters



Li, Wu, Tedrake, Tenenbaum, Torralba, "Learning Particle Dynamics for Manipulating Rigid Bodies, Deformable Objects, and Fluids," ICLR'19

Results of the Dynamics Prior f_D



Dynamics-Guided Inference

Dynamics-Guided Inference

$\hat{Q}$: Rigidness of each instance

$\hat{P}$: Physical parameters

Dynamics-Guided Inference

$\hat{Q}$: Rigidness of each instance

$\hat{P}$: Physical parameters

$\hat{X}'$: Particle position

$\hat{G}$: Instance grouping

$$(\hat{P}, \hat{Q}, \qquad) = f_I(\hat{X}', \hat{G})$$

Dynamics-Guided Inference

$\hat{Q}$: Rigidness of each instance

$\hat{P}$: Physical parameters

$\hat{X}'$: Particle position

$\hat{G}$: Instance grouping

$\Delta\hat{X}$: Position refinement

$$(\hat{P}, \hat{Q}, \Delta\hat{X}) = f_I(\hat{X}', \hat{G})$$

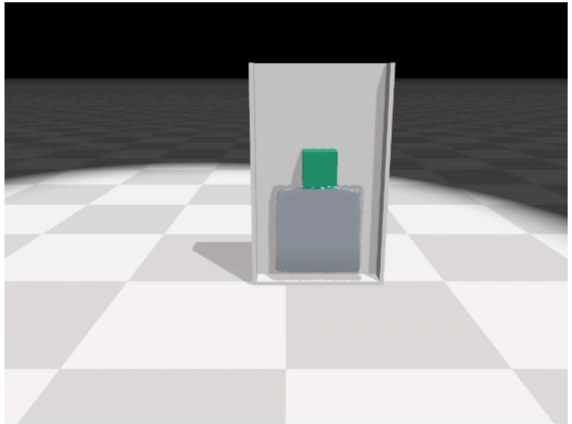
Results

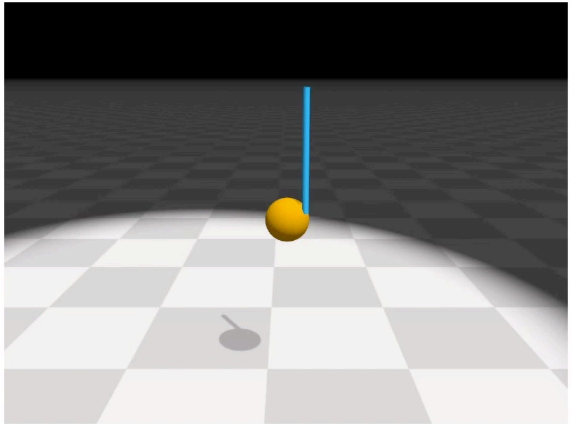
We will mainly investigate how accurate the following estimations are and whether they help with future prediction:

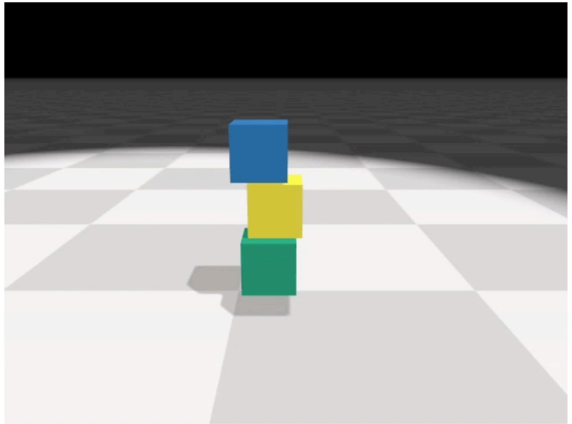
(1) $\hat{Q}$: Rigidness estimation

(2) $\hat{P}$: Parameter estimation

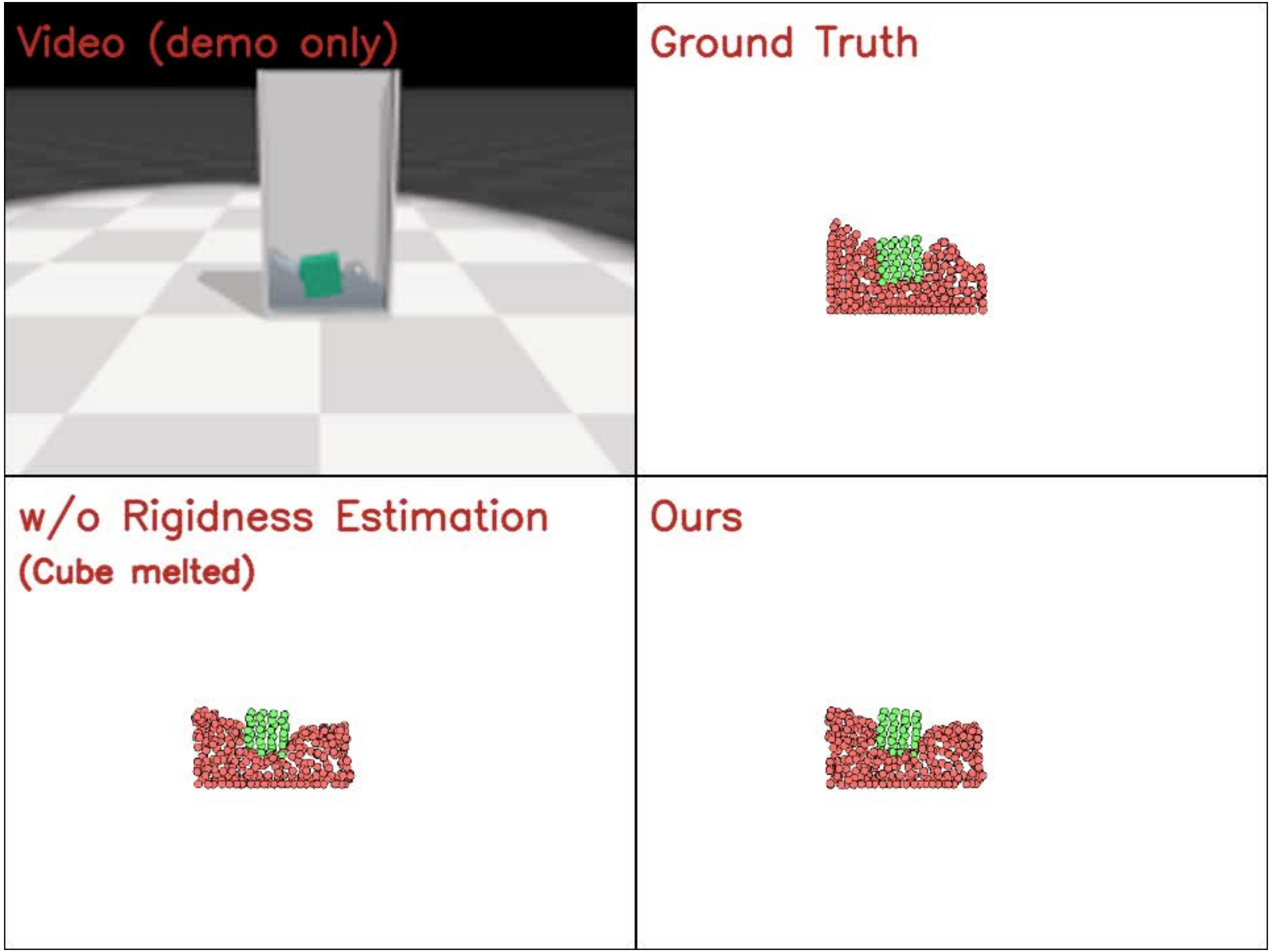
(3) $\Delta\hat{X}$: Position refinement



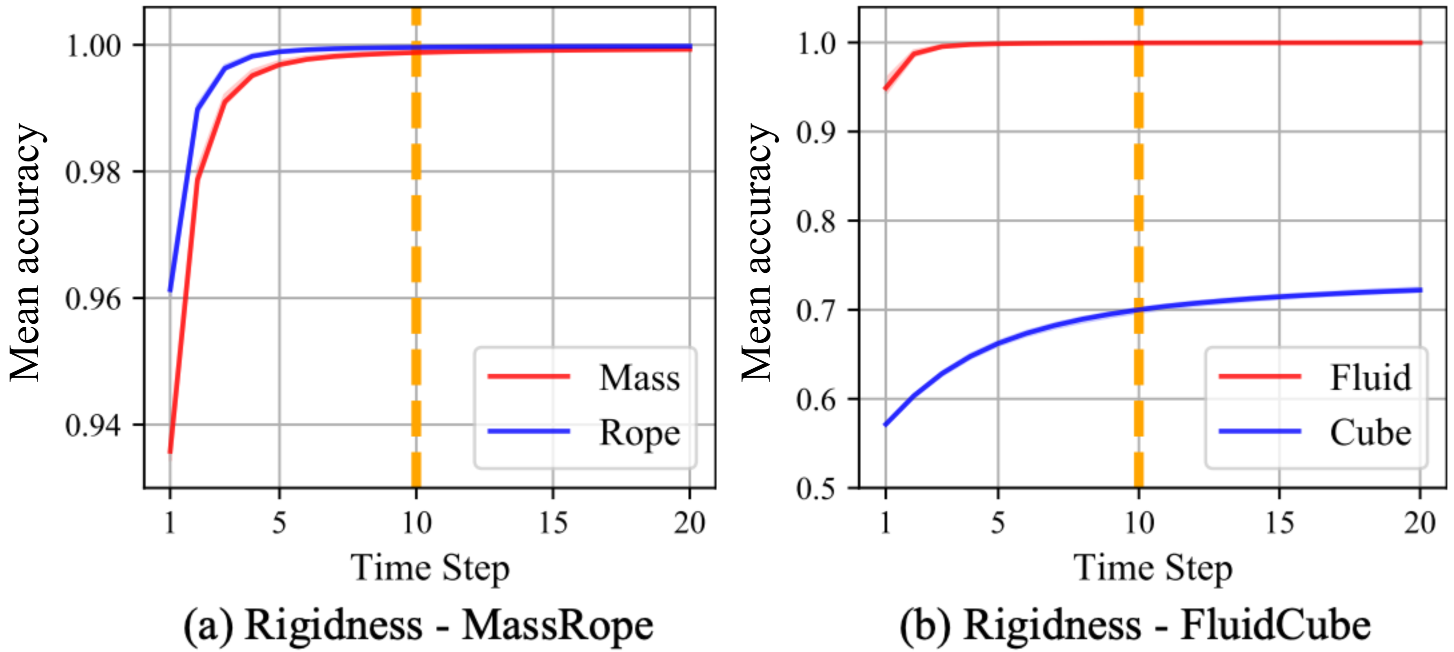




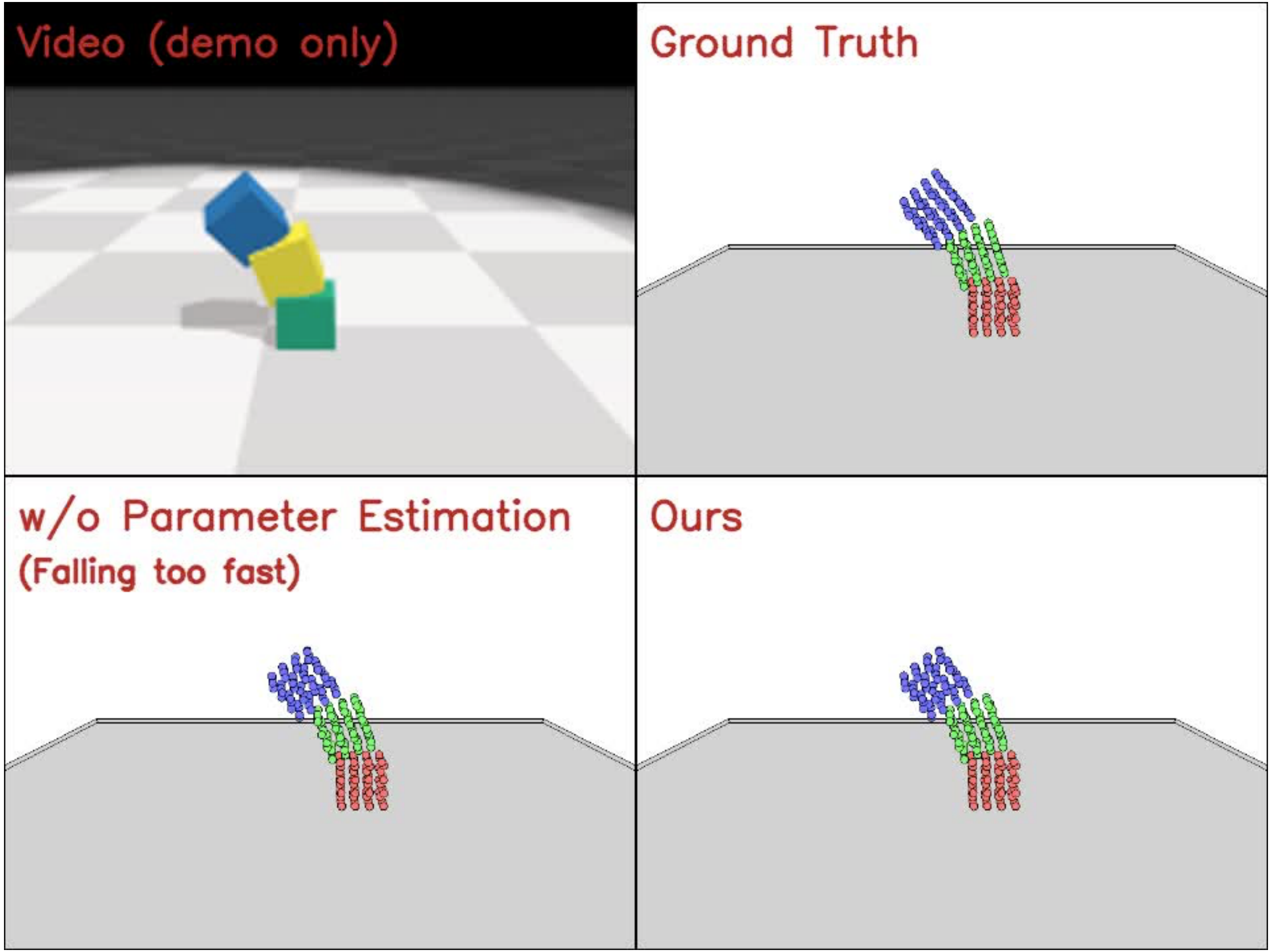
Qualitative results on **Rigidness Estimation**



Quantitative results on **Rigidness Estimation**



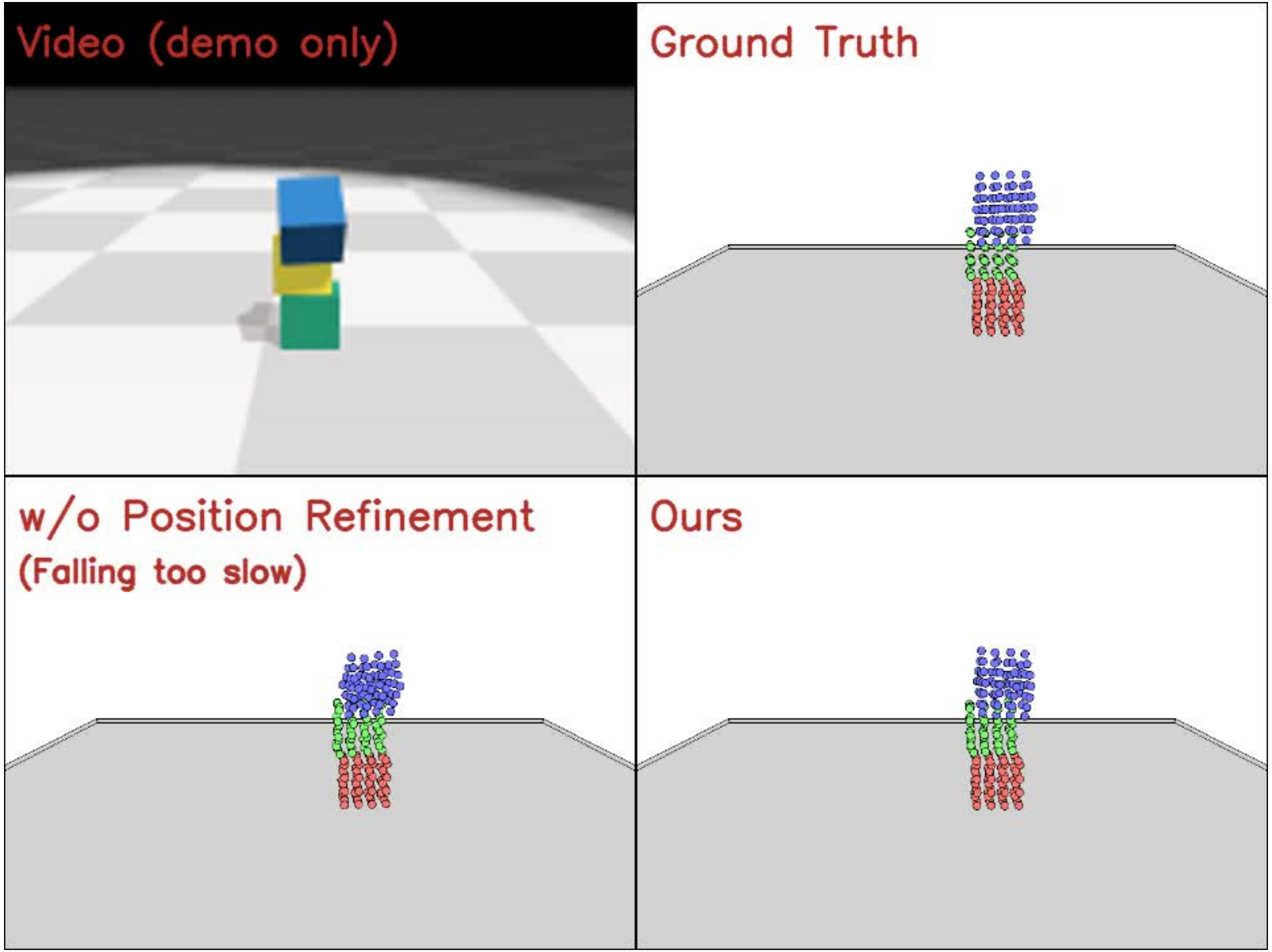
Qualitative results on **Parameter Estimation**



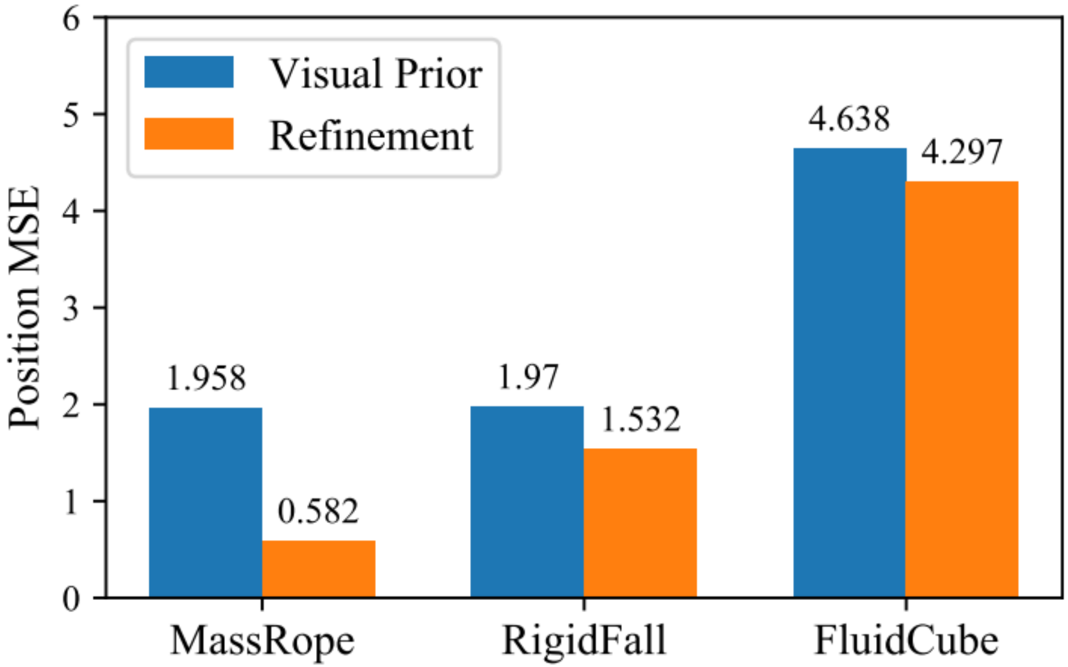
Quantitative results on **Parameter Estimation**

Methods	MassRope	RigidFall	FluidCube
DensePhysNet	24.5% (15.1)	25.7% (15.4)	28.6% (15.0)
Ours w/o Rigidness	3.4% (2.2)	7.4% (4.1)	22.2% (14.7)
VGPL (ours)	**2.9% (1.3)**	**3.7% (2.7)**	**17.5% (13.6)**

Qualitative results on **Position Refinement**



Quantitative results on **Position Refinement**



Quantitative results on **Future Prediction**

Methods	FluidCube				RigidFall				MassRope		
	$T+1$	$T+5$	$T+10$	$T+20$	$T+1$	$T+5$	$T+10$	$T+20$	$T+1$	$T+5$	$T+10$
w/o Rigidness	**3.864**	5.100	7.631	13.62	2.283	10.68	43.93	198.1	0.898	4.849	16.40
w/o Refinement	4.530	6.349	8.584	10.50	2.640	6.720	16.71	57.10	2.298	3.628	7.493
w/o Param. Est.	3.894	5.363	7.557	10.19	**2.110**	6.229	16.04	51.91	0.845	4.612	24.48
VGPL (ours)	3.887	**5.038**	**6.531**	**7.998**	2.112	**6.190**	**15.73**	**50.78**	**0.807**	**2.724**	**7.338**

In summary

We proposed **Visually Grounded Physics Learner (VGPL)** to

(1) simultaneously reason about physics and make future predictions
based on visual and dynamics priors.

In summary

We proposed **Visually Grounded Physics Learner (VGPL)** to

(1) simultaneously reason about physics and make future predictions based on visual and dynamics priors.

(2) We employ a particle-based representation to handle rigid bodies, deformable objects, and fluids.

In summary

We proposed **Visually Grounded Physics Learner (VGPL)** to

(1) simultaneously reason about physics and make future predictions based on visual and dynamics priors.

(2) We employ a particle-based representation to handle rigid bodies, deformable objects, and fluids.

(3) Experiments show that our model can infer the physical properties within a few observations, which allows the model to quickly adapt to unseen scenarios and make accurate predictions into the future.

Thank you for watching!