Learning to See Physics via Visual De-animation

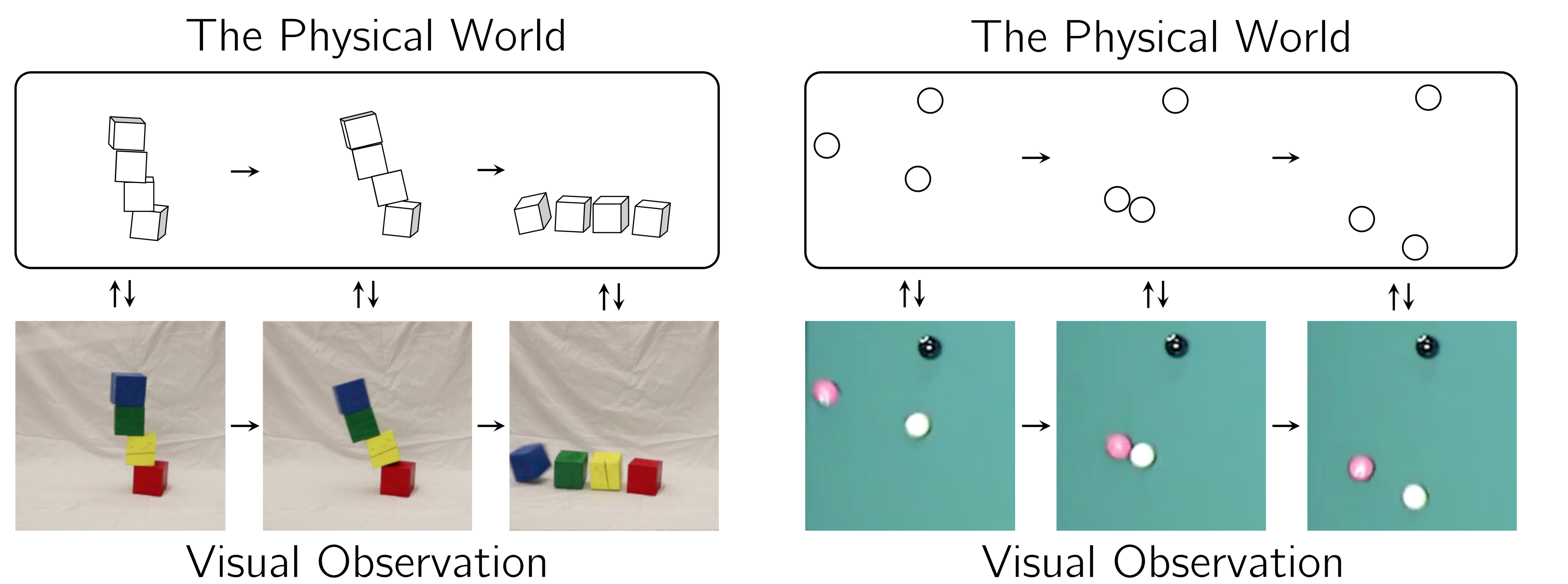
Jiajun Wu[1] Erika Lu[2] Pushmeet Kohli[3] William T. Freeman[1,4] Joshua B. Tenenbaum[1]
1 Massachusetts Institute of Technology 2 University of Oxford 3 DeepMind 4 Google

Learning to See Physics

Goal: **see** an interpretable scene representation, and model its **dynamics**

Motivation

- An object-based, compact, disentangled representation has wide applications.
- Existing models for scene dynamics do not have a perception module.

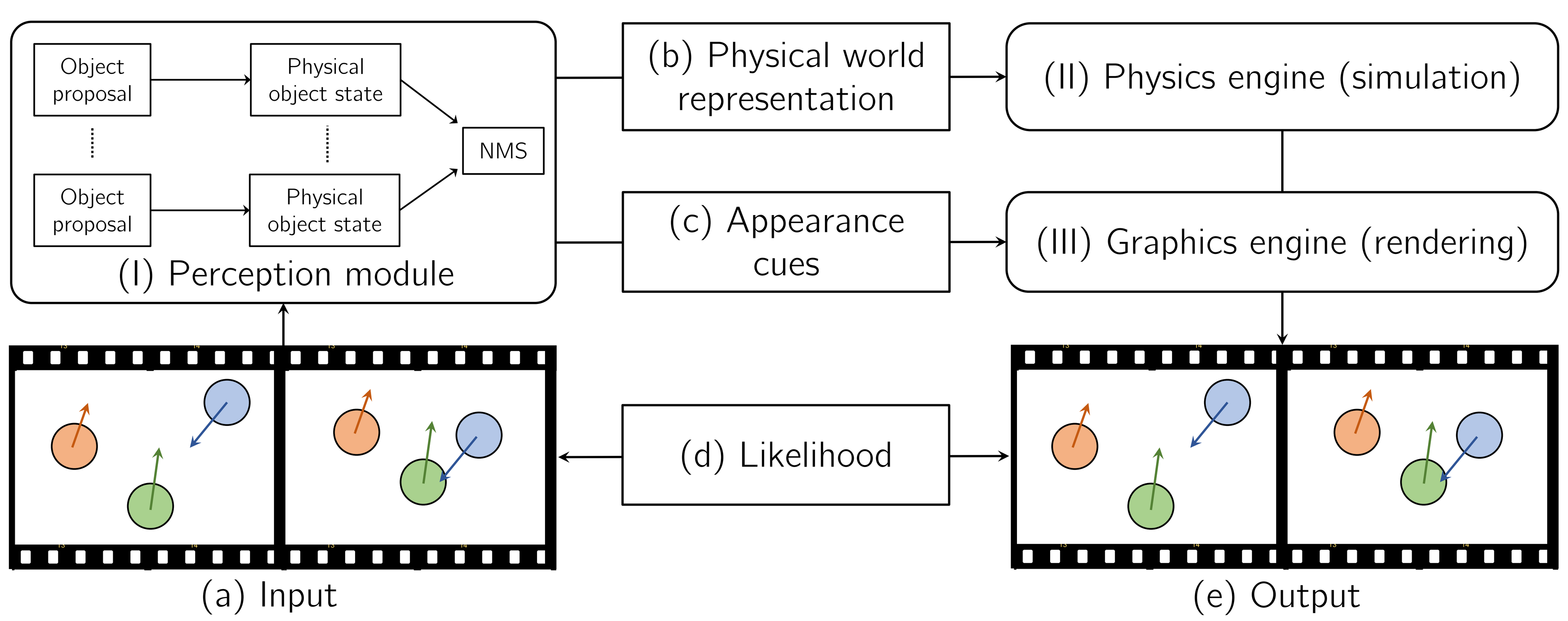


Solution: looping in a forward physics engine and a graphics engine in recognition

Advantages

- Generative, simulation engines bring in symbolic representation naturally.
- The learning paradigm adapts to a variable number of objects in the scene.
- The learned representations have wide applications with simulation engines.

Visual De-animation



- The perception module explains the scene by detecting and describing the object in each segment proposal, in a scene de-rendering style [4].
- The model aims to minimize two losses: the inference loss after the perception module, and the reconstruction loss after the simulation engines.

References

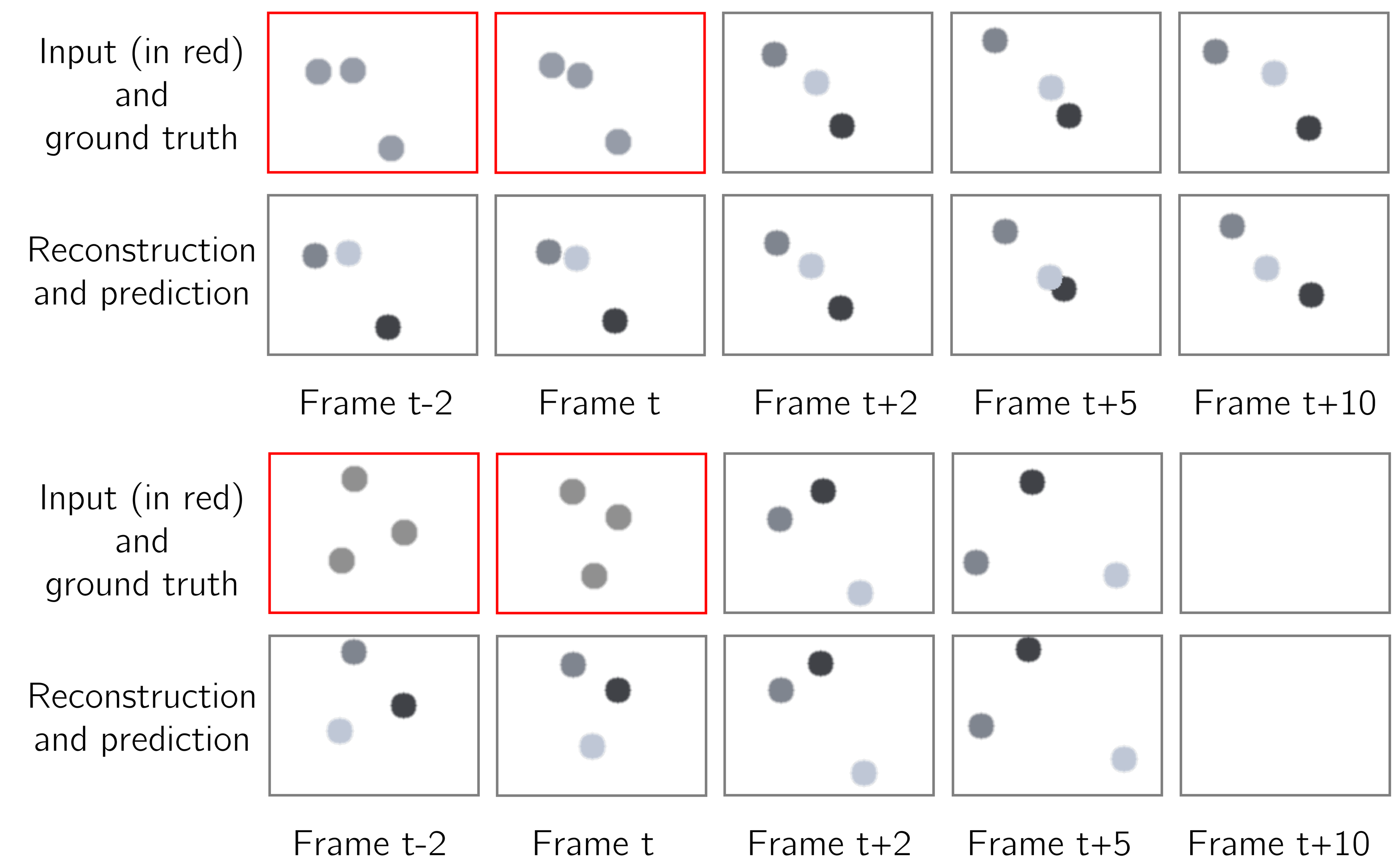
[1] Wu*, Yildirim*, et al. Galileo: Perceiving Physical Object Properties. NIPS 2015
[2] Lerer et al. Learning Physical Intuition of Block Towers by Example. ICML 2016
[3] Chang et al. A Compositional Object-Based Approach to Learning Physical Dynamics. ICLR 2017
[4] Wu et al. Neural Scene De-rendering. CVPR 2017
[5] Watters et al. Visual Interaction Networks. NIPS 2017

Study 1: Billiard Tables

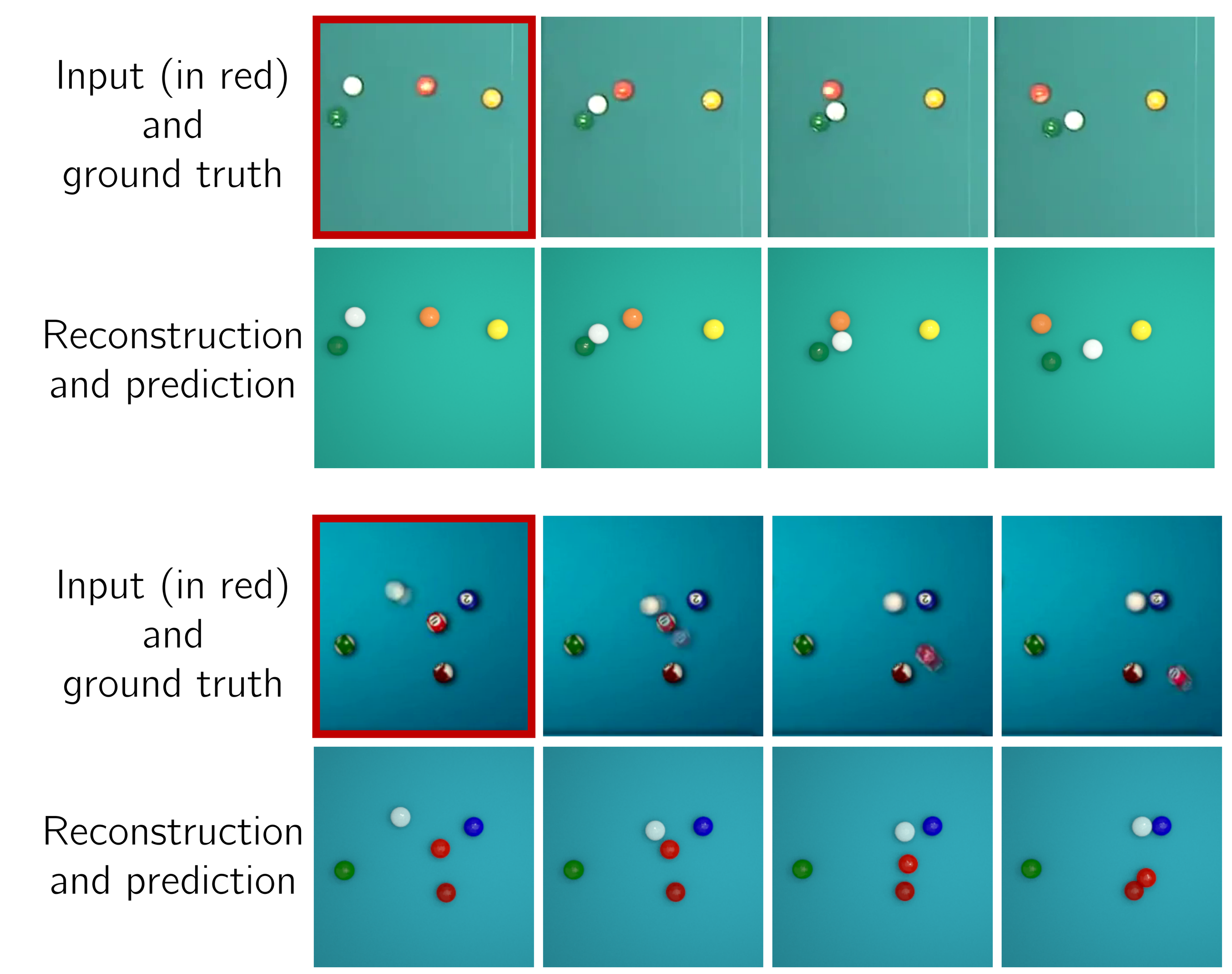
Setup

- 2D physics simulation with a neural, differentiable physics engine [3]
- Pre-training on data synthesized by the graphics and physics engines
- End-to-end fine-tuning with the reconstruction loss using back-propagation, as simulation engines are differentiable

Experiments on Synthetic Videos Generated by Physics Engines



Experiments on Real Videos from YouTube



Study 2: Block Towers

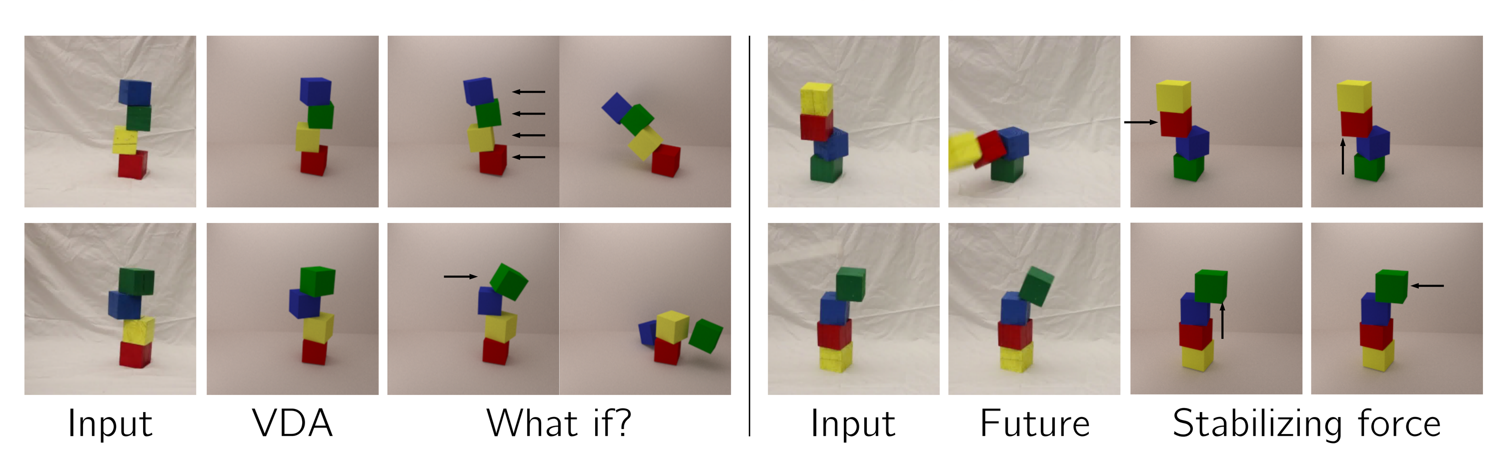
Setup

- Rigid body simulation with a non-differentiable physics engine
- Pre-training on data synthesized by the generative models
- End-to-end fine-tuning with the reconstruction loss and the loss on stability prediction (with REINFORCE)

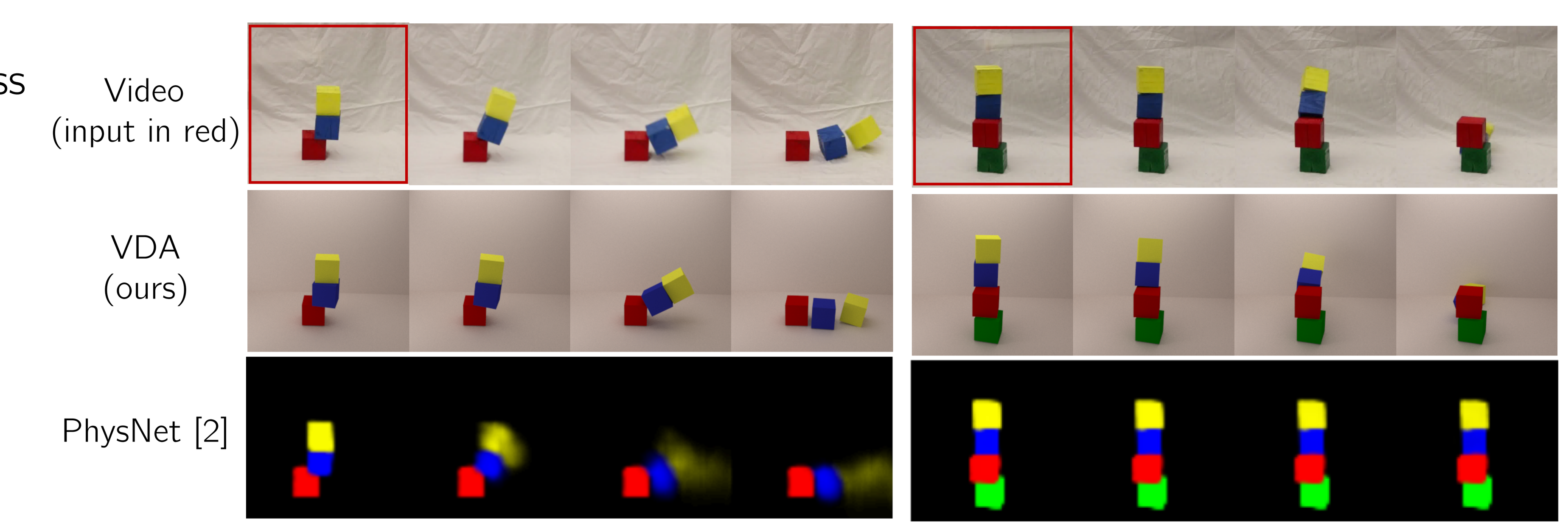
Accuracies on Stability Prediction

Methods	2 Blocks	3 Blocks	4 Blocks	Mean
VDA (ours)	75	76	73	75
PhysNet [2]	66	66	73	68
GoogleNet	70	70	70	70
Chance	50	50	50	50

What if ... ? How to ... ?



Results on the Block Tower Dataset [2]



More Qualitative Results