# Physics 101: Learning Physical Object Properties from Unlabeled Videos

Jiajun Wu[1], Joseph J. Lim[2], Hongyi Zhang[1], Joshua B. Tenenbaum[1], William T. Freeman[13]
1 Massachusetts Institute of Technology   2 Stanford University   3 Google Research
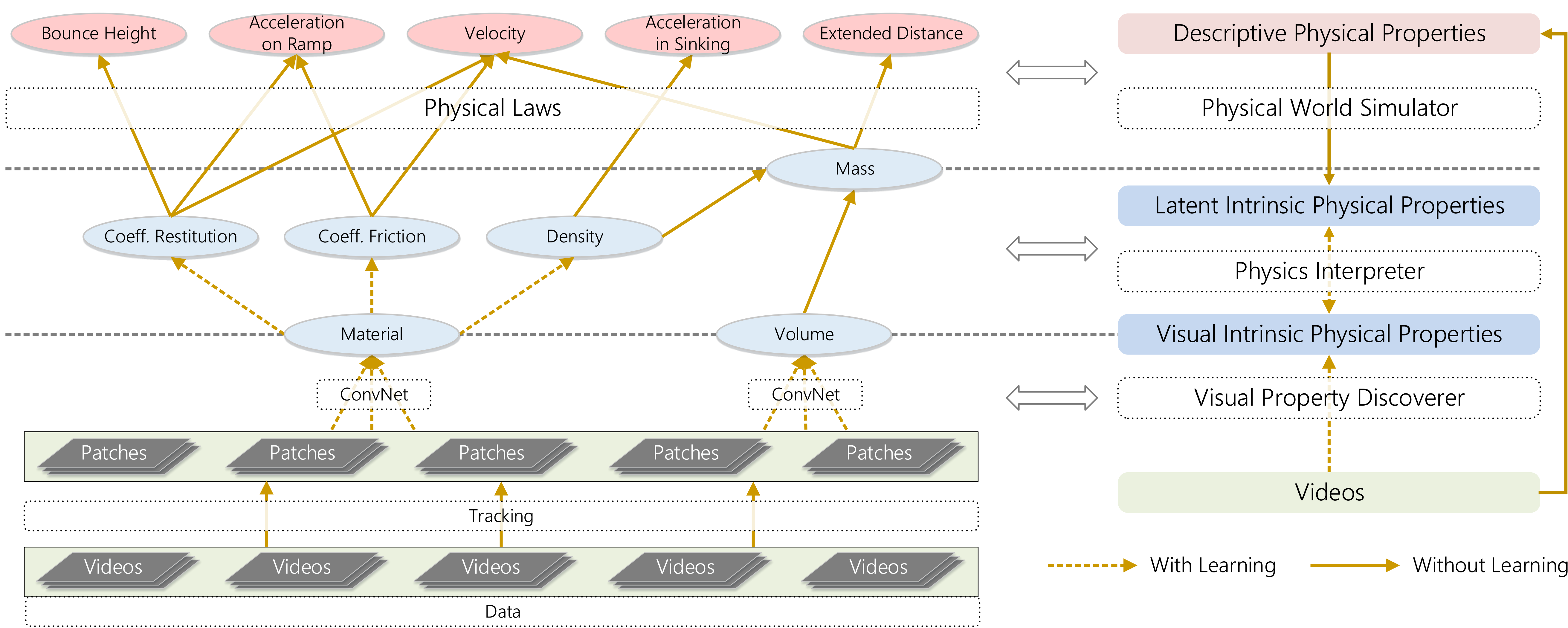
## Overview

**Observations:**
- Human infer rich information from simple visual inputs.
- Infants learn intuition on physics when they are young.
- There is no benchmark for estimating physical properties.

**Goal:** building a benchmark and a model to interpret physical scenes

**Contributions:**
- A benchmark for learning physical object properties from videos
- A novel, generally applicable model of physical scenes
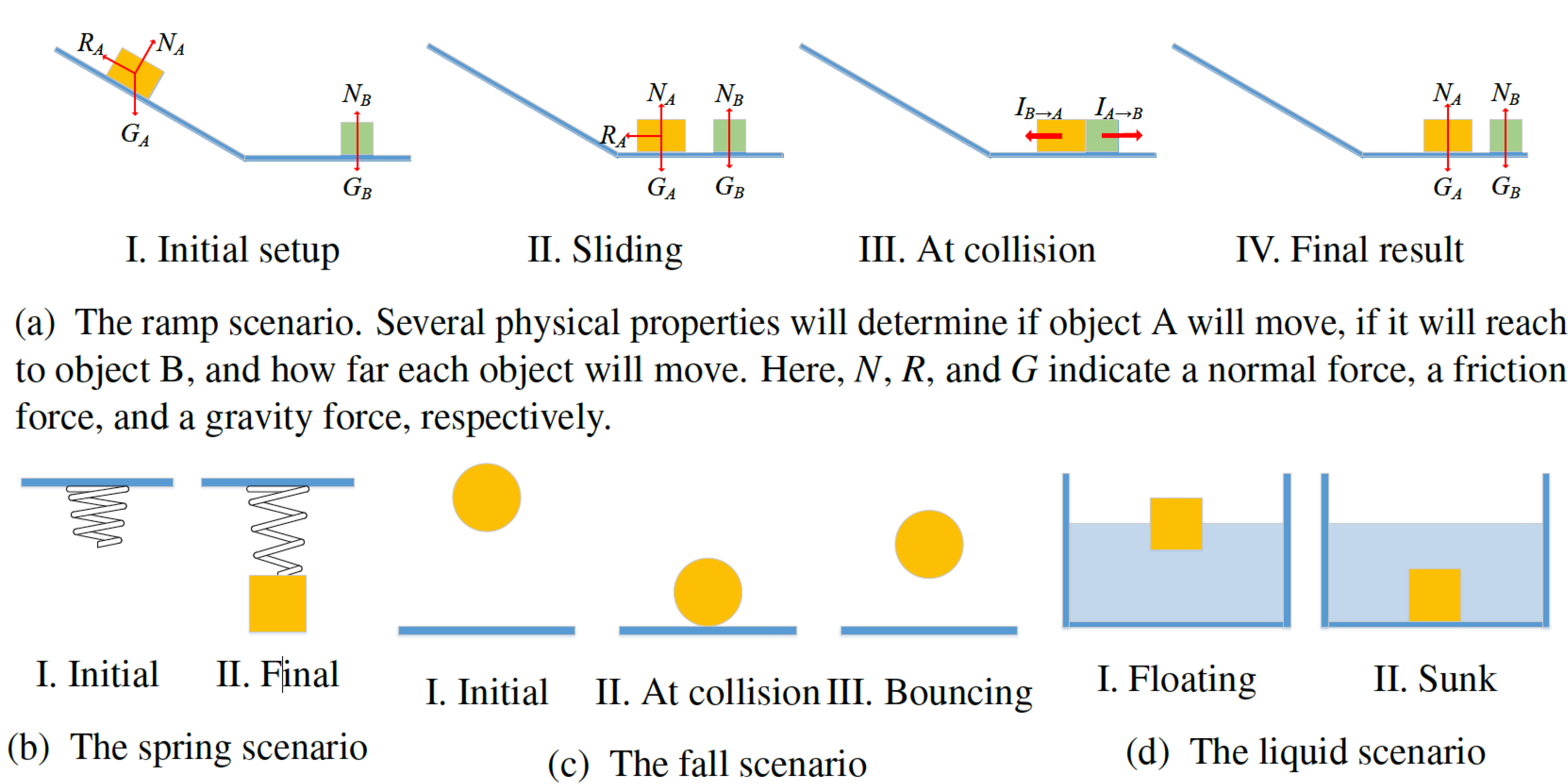- A neural net realization with good empirical performance

## Scene Modeling



## Physics 101: A Benchmark
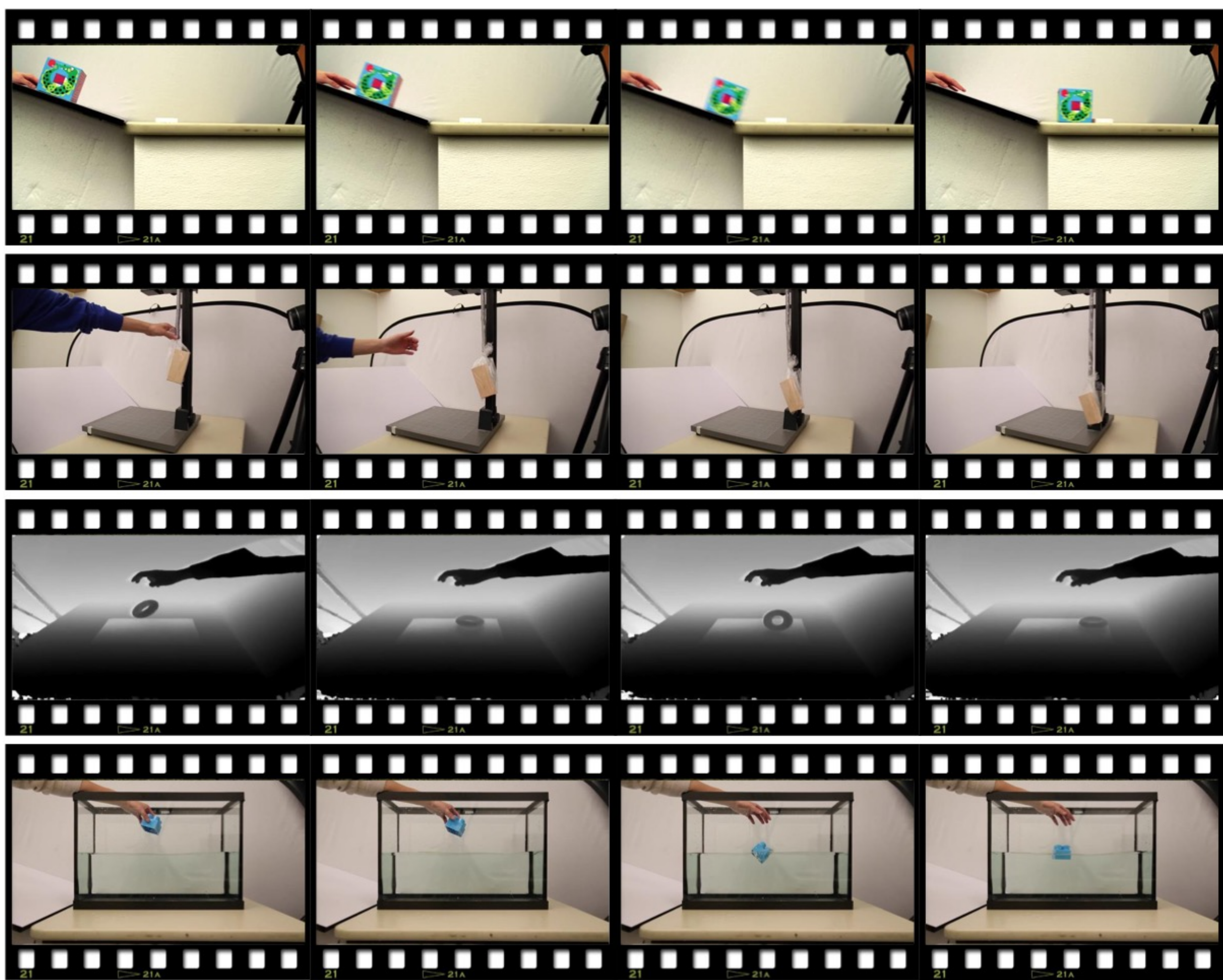


### 101 Objects from 15 Categories



|  I. Initial setup | II. Sliding | III. At collision | IV. Final result |

(a) The ramp scenario. Several physical properties will determine if object A will move, if it will reach to object B, and how far each object will move. Here, $N$, $R$, and $G$ indicate a normal force, a friction force, and a gravity force, respectively.



| I. Initial | II. Final | I. Initial | II. At collision | III. Bouncing | I. Floating | II. Sunk |

(b) The spring scenario    (c) The fall scenario    (d) The liquid scenario

### 4 Scenarios

## Experiments

### Inferring Intrinsic Physical Properties

| Methods | Mass | | | Density | | | Volume | | |
|---|---|---|---|---|---|---|---|---|---|
|  | Frame | Trial | Object | Frame | Trial | Object | Frame | Trial | Object |
| RGB (oracle) | 0.79 | 0.72 | 0.67 | 0.83 | 0.74 | 0.65 | 0.77 | 0.67 | 0.61 |
| Depth (oracle) | 0.67 | 0.51 | 0.33 | 0.69 | 0.61 | 0.49 | 0.40 | 0.41 | 0.33 |
| RGB (spring) | 0.40 | 0.35 | 0.20 | N/A | N/A | N/A | N/A | N/A | N/A |
| RGB (liquid) | N/A | N/A | N/A | 0.33 | 0.27 | 0.30 | N/A | N/A | N/A |
| RGB (joint) | 0.52 | 0.32 | 0.26 | 0.33 | 0.34 | 0.33 | 0.40 | 0.37 | 0.30 |
| Depth (joint) | 0.33 | 0.27 | 0.25 | 0.29 | 0.23 | 0.17 | 0.30 | 0.20 | 0.22 |
| Uniform | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

### Inferring Descriptive Physical Properties

| Tasks | Methods | Frame | Trial | Object |
|---|---|---|---|---|
| Moving Distance | Ours (joint) | 0.65 | 0.42 | 0.33 |
| Moving Distance | CNN | 0.63 | 0.39 | 0.21 |
| Moving Distance | Uniform | 0 | 0 | 0 |
| Bounce Height | Ours (joint) | 0.35 | 0.31 | 0.23 |
| Bounce Height | Ours (transfer) | 0.32 | 0.27 | 0.16 |
| Bounce Height | CNN | 0.36 | 0.25 | 0.15 |
| Bounce Height | Uniform | 0 | 0 | 0 |
| Float | Ours (joint) | 0.94 | 0.87 | 0.84 |
| Float | CNN | 0.96 | 0.86 | 0.81 |
| Float | Uniform | 0.70 | 0.70 | 0.70 |



### 17,000+ Videos from 4 Viewpoints