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