Learning to See Physics via Visual De-animation

**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
- 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.

**Advantages**
- 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

### Study 1: Billiard Tables

**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)

### 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)

### References