
Haotian Xue\(^1\), Antonio Torralba\(^2\), Joshua B. Tenenbaum\(^2\), Daniel LK Yamins\(^3\), Yunzhu Li\(^3\), Hsiao-Yu Tung\(^2\)

**Background: Intuitive Physics**

- Humans have the ability to gain strong intuition about the physical world around them, allowing us to predict the movement of complex dynamics in 3D space without knowing the underlying dynamics.

- Learning from visual inputs of old scenes, humans can generalize the acquired 3D-aware intuition to new scenes.

**Motivation: Learning 3D-IntPhys from Video**

- We want a framework to enable machine to learn such kind of 3D-aware intuitive physics from solely visual inputs.

- We want to impose strong inductive bias, making it possible to learn reasonable intuitive physics from visual inputs with a strong generalization ability to unseen settings.

**Our Approach: 3D-IntPhys**

**Key Idea:** Learn explicit 3D-based representation to model dynamics from visual inputs, which can be well-generalized to unseen scenes.

**Input Video from 6 Camera Viewpoints**

![3D-Aware Intuitive Physics](image)

**Dynamics Prediction in Particle Space**

- Our method is composed of a conditional NeRF-style visual frontend and a 3D point-based dynamics prediction backend, imposing strong structural inductive bias.

- We first train conditional NeRF to reconstruct explicit 3D representation, then we learn explicit 3D dynamics with Chamfer Loss and Merge Loss.

- We generate multi-view dataset for interpolate and extrapolate settings.

**Motivation: Learning 3D-IntPhys from Video**

- We want a framework to enable machine to learn such kind of 3D-aware intuitive physics from solely visual inputs.

- We want to impose strong inductive bias, making it possible to learn reasonable intuitive physics from visual inputs with a strong generalization ability to unseen settings.

**Results**

**Visual-head Reconstruction:** 3D-IntPhys has a more generalized visual head.

**Long-term Rollout Prediction:** The error of 3D-IntPhys vs Baseline.

More video about the dataset and rollout prediction can be found by scanning the QR code:

More video results can be found here.