

Motivation & Contributions

- Humans are good at *perceiving* and *simulating* the world with 3D structure in mind
- Previous deep generative models are often limited to a single object, hard to interpret, and missing the 3D structure

3D-SDN (ours) vs. 2D Method

Original image

3D-SDN (ours)

pix2pixHD



• **3D-SDNs** learn and incorporate

- (1) Scene semantic labels
- (2) Texture encodings for objects and the background
- (3) 3D geometry and pose for objects

Scene Manipulation via 3D-SDN



3D-Aware Scene Manipulation via Inverse Graphics

Shunyu Yao^{*1}, Tzu-Ming Harry Hsu^{*2}, Jun-Yan Zhu², Jiajun Wu², Antonio Torralba², William T. Freeman²³, and Joshua B. Tenenbaum² ¹Tsinghua University, ²MIT CSAIL, ³Google Research







