Towards Unsupervised Learning of Generative Models for 3D Controllable Image Synthesis

Yiyi Liao*, Katja Schwarz*, Lars Mescheder, Andreas Geiger
Motivation

3D Controllable Image Synthesis

- 3D controllability is essential in many applications, e.g., gaming, simulation, virtual reality and data augmentation
- 3D controllable properties: 3D pose, shape, appearance of multiple objects and camera viewpoint
Motivation

Classical Rendering Pipeline

- **3D Controllable**
- Expensive and inefficient to design 3D models
Motivation

2D Generative Models

- Efficient, learned from only 2D images
- Geometry and appearance not disentangled → no 3D control
Motivation

Ours

- 3D Controllable
- Efficient, learned from only 2D images
- Unsupervised, disentangled 3D representation learning
Motivation

Idea: Learning the image synthesis pipeline jointly in 3D and 2D space
Method Overview

3D Generator

Differentiable Rendering

2D Generator

\( g_{3D} : z \rightarrow I \)

\( g_{2D} : z \rightarrow I \)

\( \alpha_{bg} \)

\( \alpha_i \)

\( X_{bg} \)

\( X_i \)

\( A_i \)

\( D_i \)

\( X'_{bg} \)

\( X'_i \)

\( A'_i \)

\( D'_i \)

\( \mathcal{L}_{com} \)

\( \mathcal{L}_{geo} \)

\( \mathcal{L}_{adv} \)

Real Image

Real Background Image

\( I \)

\( I_{bg} \)
Experiments

Datasets

Car w/o BG  Car with BG  Indoor  Fruit
Experiments

Car Dataset

- Layout2Im and our 2D baseline are only controllable for 2D translation
- Layout2Im fails to disentangle object identity and pose
- Our method is controllable for 3D translation and rotation with coherent object identity
Experiments

Indoor Dataset

- Layout2Im and our 2D baseline are only controllable for 2D translation
- Layout2Im fails to disentangle object identity and pose
- Our method is controllable for 3D translation and rotation with coherent object identity
Experiments

Fruit Dataset

- We collect 800 images with 5 fruits and 5 backgrounds
- Our method is able to synthesize plausible images from real data
Failure Cases

A single primitive generates multiple objects occasionally.

Identity might flip wrt. large viewpoint change.

Stronger inductive biases are required to tackle these problems.
Thank you!