Visual Object Networks: Image Generation with Disentangled 3D Representations

Jun-Yan Zhu[1] Zhoutong Zhang[1] Chengkai Zhang[1] Jiajun Wu[1] Antonio Torralba[1] Joshua B. Tenenbaum[1] William T. Freeman[1,2]

1 Massachusetts Institute of Technology 2 Google Research

Motivation

Problem: Existing 2D generative models

- ❑ **Do not capture the 3D nature of the world.**
- ❑ **Do not allow 3D-aware image manipulations.**

Given an image of a car:

- What does it look like from a different viewpoint?
- What if we apply its texture to a van?
- Can we mix different 3D designs?

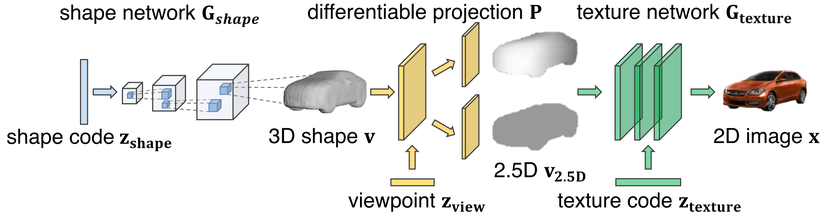
Goal: Joint 3D & 2D generation with disentangled representation.

Samples from WGAN-GP

Our 3D, 2.5D, and 2D output

3D disentanglement

Visual Object Networks (VON)



Training data: Unpaired 3D shapes & Image datasets

Formulation

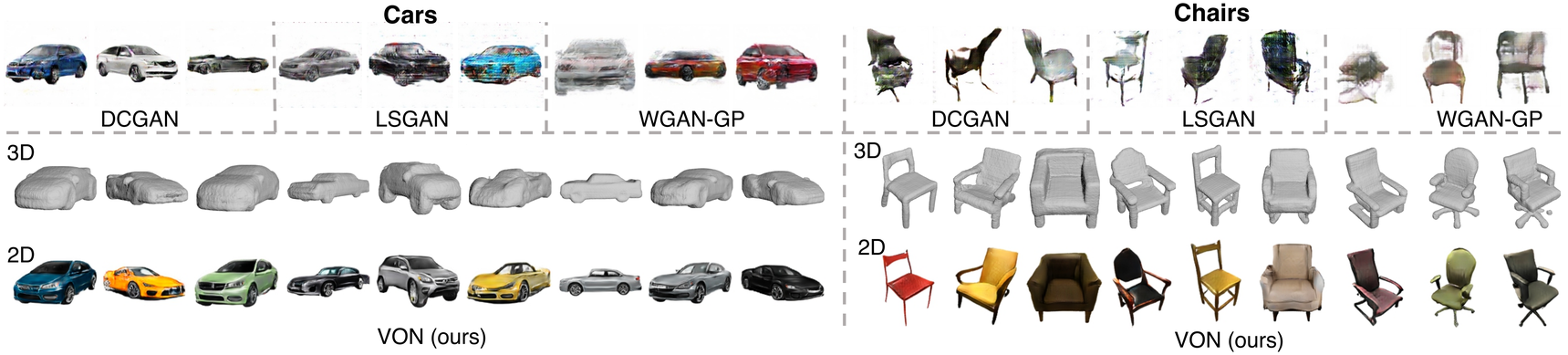
Image Formation: $x = G_{\text{texture}}(\mathcal{P}(G_{\text{shape}}(\mathbf{z}_{\text{shape}}), \mathbf{z}_{\text{view}}), \mathbf{z}_{\text{texture}})$

Total Loss: $\mathcal{L} = \lambda_{\text{shape}}\mathcal{L}_{\text{shape}} + \mathcal{L}_{\text{texture}}$

3D Shape Loss: $\mathcal{L}_{\text{shape}} = \mathbb{E}_{\mathbf{v}}[D_{\text{shape}}(\mathbf{v})] - \mathbb{E}_{\mathbf{z}_{\text{shape}}}[D_{\text{shape}}(G_{\text{shape}}(\mathbf{z}_{\text{shape}}))]$

Texture Loss: $\mathcal{L}_{\text{texture}} = \underbrace{\mathcal{L}^{\text{GAN}}_{\text{image}} + \mathcal{L}^{\text{GAN}}_{\text{2.5D}}}_{\text{Adversarial losses}} + \underbrace{\mathcal{L}^{\text{cyc}}_{\text{image}} + \mathcal{L}^{\text{cyc}}_{\text{2.5D}} + \mathcal{L}^{\text{cyc}}_{\text{texture}}}_{\text{Cycle-consistency losses}} + \underbrace{\mathcal{L}_{\text{KL}}}_{\text{KL loss}}$

Image Generation (2D GANs vs. VON)

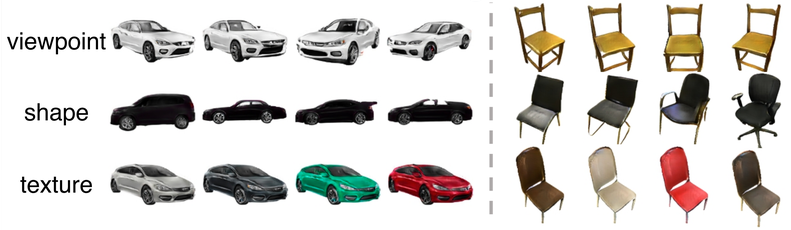


Shape Generation

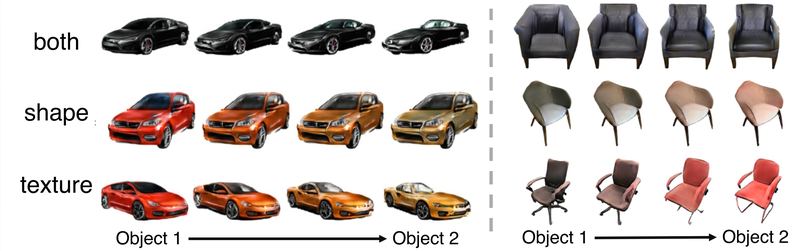
	3D-GAN (voxels)	VON (voxels)
Cars	3.021	**0.021**
Chairs	2.598	**0.082**
	3D-GAN (DF)	VON (DF)
Cars	3.896	**0.002**
Chairs	1.790	**0.006**

3D Fréchet Inception Distance

Applications: Disentangled Generation



Applications: Disentangled Interpolation



Applications: Texture Transfer