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
- viewpoint
- shape
- texture
- 3D disentanglement

Our 3D, 2.5D, and 2D output

Training data: Unpaired 3D shapes & Image datasets

Visual Object Networks (VON)

Visual Object Networks (VON) VON (ours) VON (ours) VON (ours)

Formulation

**Image Formation:**
\[ x = G_{\text{texture}}(P(G_{\text{shape}}(z_{\text{shape}}, z_{\text{view}}), z_{\text{texture}})) \]

**Total Loss:**
\[ \mathcal{L} = \lambda_{\text{shape}} \mathcal{L}_{\text{shape}} + \mathcal{L}_{\text{texture}} \]

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

**Texture Loss:**
\[ \mathcal{L}_{\text{texture}} = \mathcal{L}_{\text{GAN}}^\text{image} + \mathcal{L}_{2.5D}^\text{image} + \mathcal{L}_{2.5D}^\text{texture} + \mathcal{L}_{\text{cyc}}^\text{2.5D} + \mathcal{L}_{\text{cyc}}^\text{image} + \mathcal{L}_{\text{KL}} \]

Adversarial losses
Cycle-consistency losses

Applications: Disentangled Generation

Applications: Disentangled Interpolation

Applications: Texture Transfer