Single-View Reconstruction by Cross-Instance Consistency

**Motivation**

**Goal** → predict 3D from a single image without supervision

**Previous works** → multi-view supervision or strong priors

**Our approach**

1. Autoencoding into explicit factors (shape, texture, pose, bkg)
2. To remove the supervision / hypotheses used in prior works, we leverage the consistency across different instances

---

**Results**

**Comparison on ShapeNet [A]**

**Comparison on CUB-200 [B] & PASCAL Cars [C]**

**Real-world images - CompCars [D] & LSUN [E]**

**Ablation study**

**Qualitative analysis on CompCars [D]**

**Quantitative study on ShapeNet [A]**

---

**UNICORN: UNsupervised Cross-Instance COnsistency for 3D Reconstruction**

1. Progressive conditioning (PC)
2. Swap reconstruction $\mathcal{L}_{\text{swap}}$

---

**Overview**

Unconstrained task → overfitting

Leveraging instance consistency

1. **Progressive conditioning (PC)** = training procedure to gradually specialize to instances by increasing the latent code sizes
2. **Swap reconstruction** = training loss $\mathcal{L}_{\text{swap}}$ enforcing consistency between neighboring instances

Two other technical contributions

A. Custom rendering function to learn from raw photometry
B. Alternate 3D / pose learning to avoid identified failure modes

---

**PyTorch code**