Cycle-Consistent Generative Rendering for 2D-3D Modality Translation

3DV 2020

Tristan Aumentado-Armstrong, Alex Levinshtein, Stavros Tsogkas, Konstantinos G. Derpanis, and Allan D. Jepson
Samsung AI Centre Toronto

SAMSUNG AI CENTRE

Motivation

2D-3D paired data is not easy to obtain.
But **unpaired** data is more accessible.

Inspired by CycleGAN, can we learn a **modality translation** model between 2D to 3D?

Shape-to-Image Translation

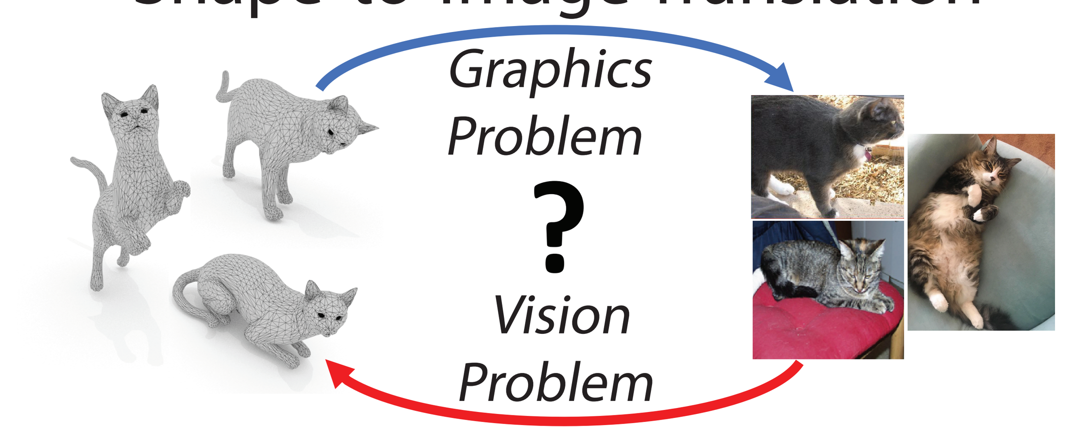
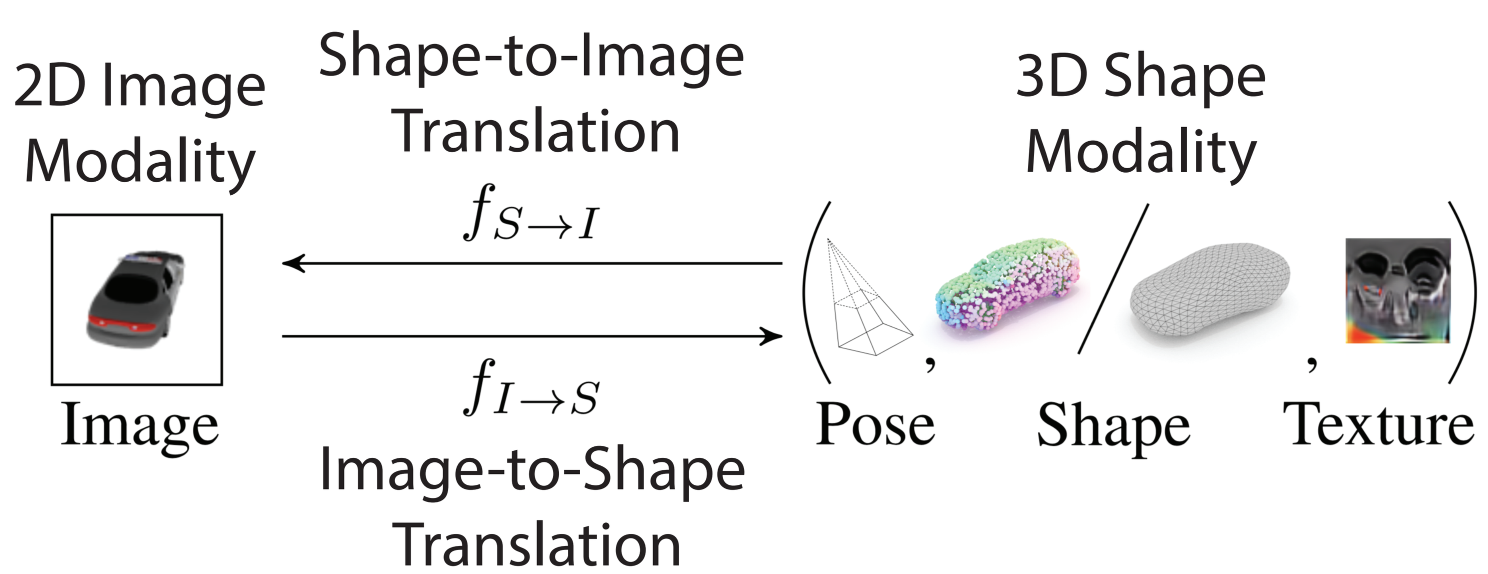


Image-to-Shape Translation

Model Overview

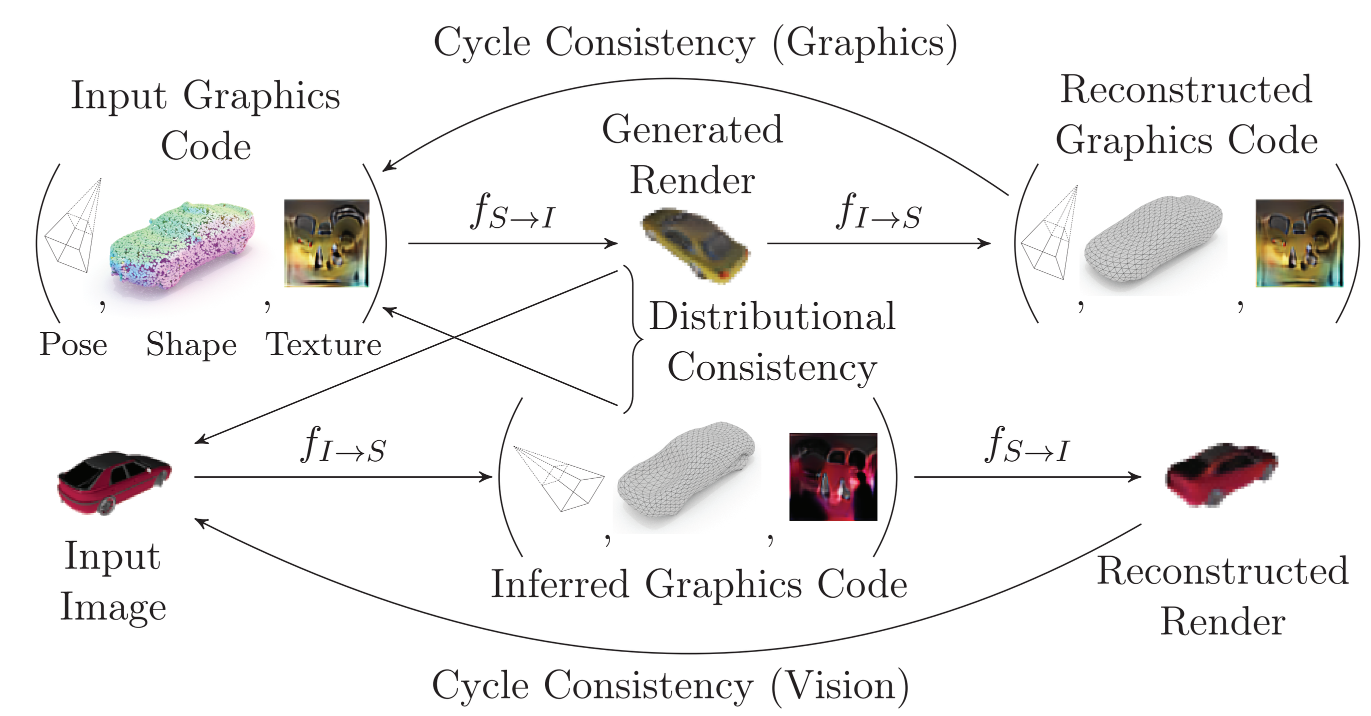
Goal: Learn a bidirectional mapping between the 2D and 3D modalities of objects from unpaired data.



Our 3D shape representation is a **graphics code**, consisting of a rigid pose, a deformable mesh, and a 3D texture.

Methods

Training cycles and losses



Generated renders (3D to 2D) ≈ real images
Inferred renders (2D to 3D) ≈ real 3D models

$$\mathcal{L}(\Phi) = \mathfrak{D}\left[S \,||\, \widetilde{S}(I)\right] + \mathfrak{D}\left[I \,||\, \widetilde{I}(S)\right] + \mathcal{C}_{I\to\widetilde{S}\to\widehat{I}}\left[I, \widehat{I}(\widetilde{S})\right] + \mathcal{C}_{S\to\widetilde{I}\to\widehat{S}}\left[S, \widehat{S}(\widetilde{I})\right] + \overbrace{\mathcal{L}_R}^{\text{Regularization}}$$

Cross-modality object fidelity (cycle consistency)

Shape-to-image translation

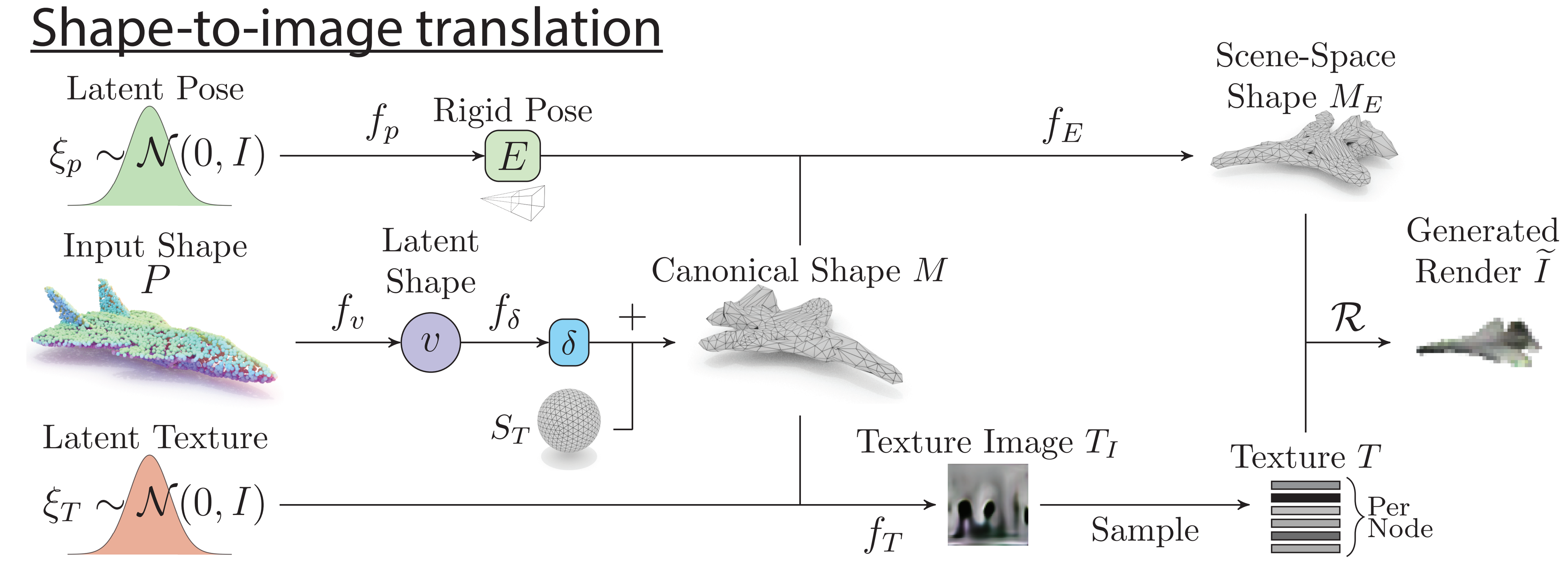
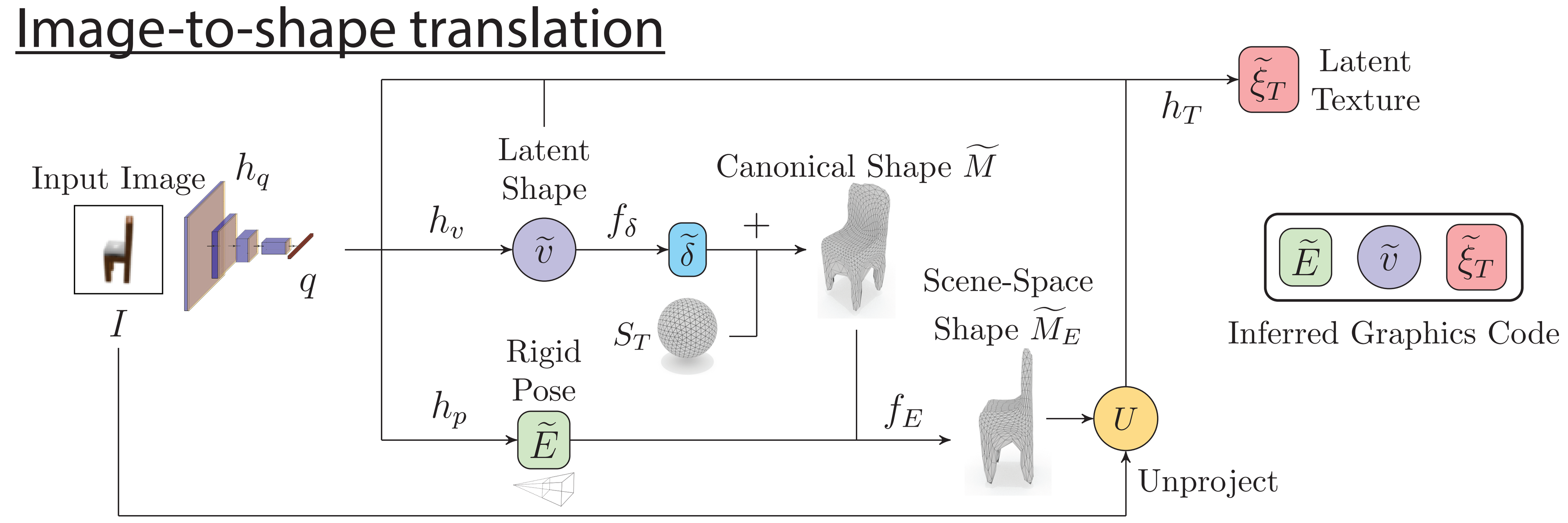
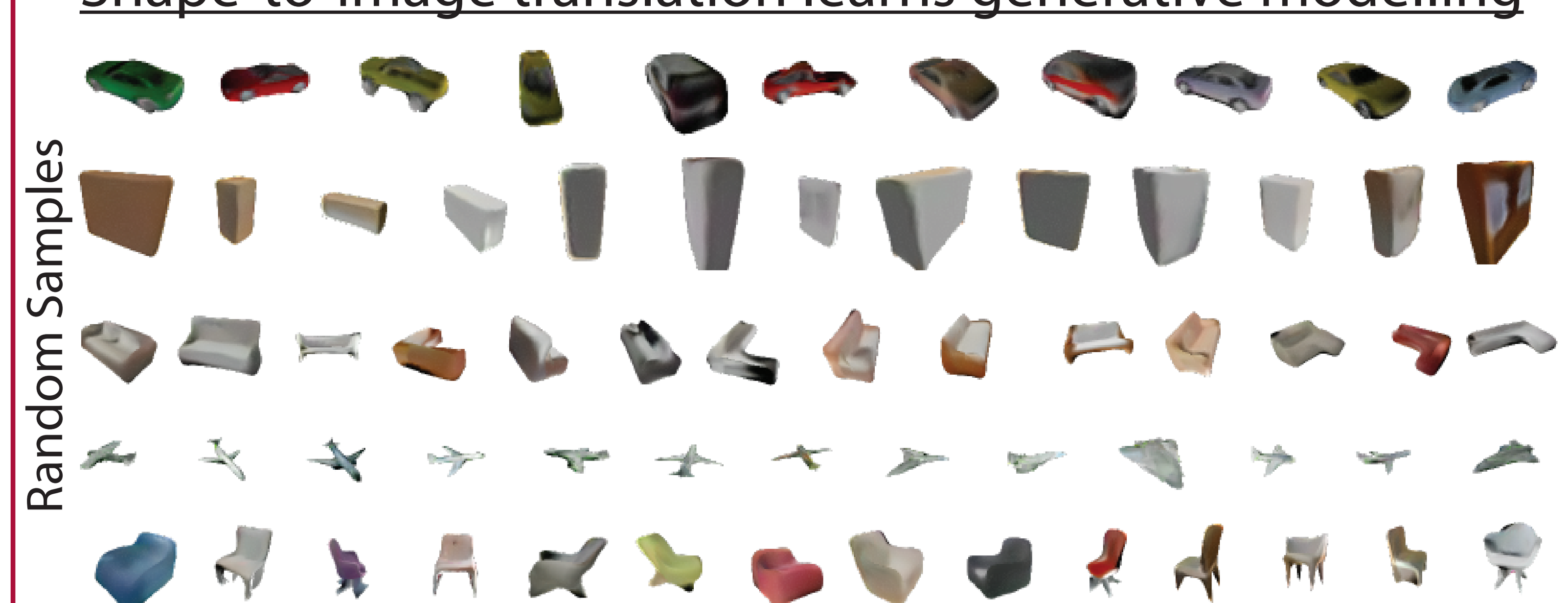


Image-to-shape translation



Results

Shape-to-image translation learns generative modelling



Learned representation disentangles pose, shape, and texture

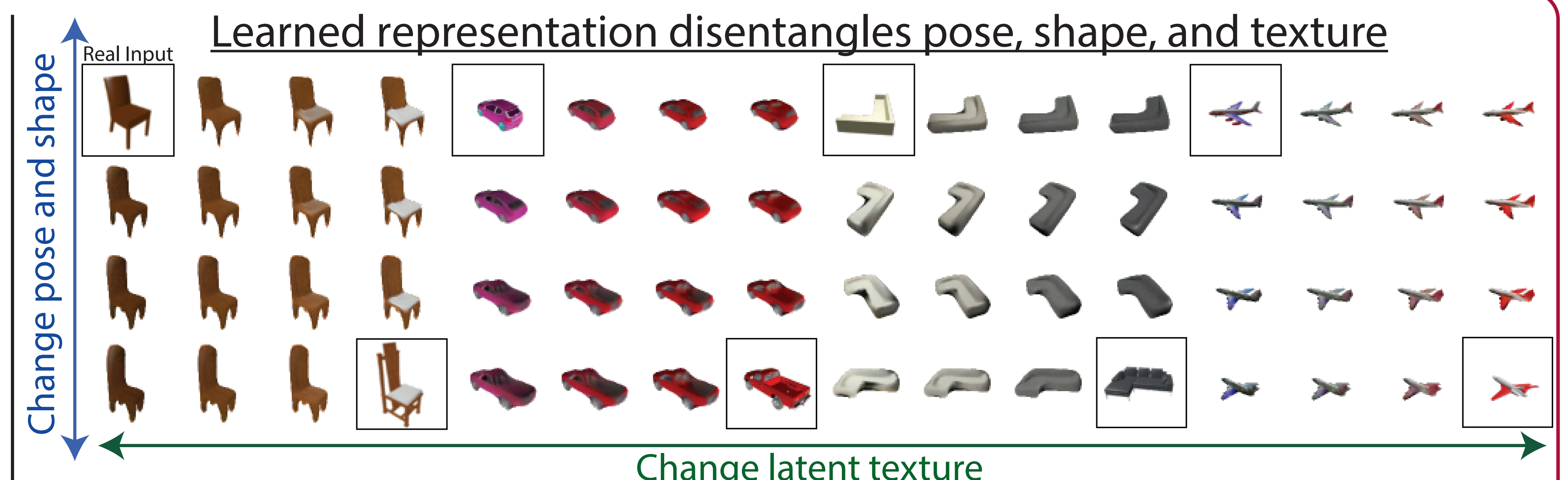
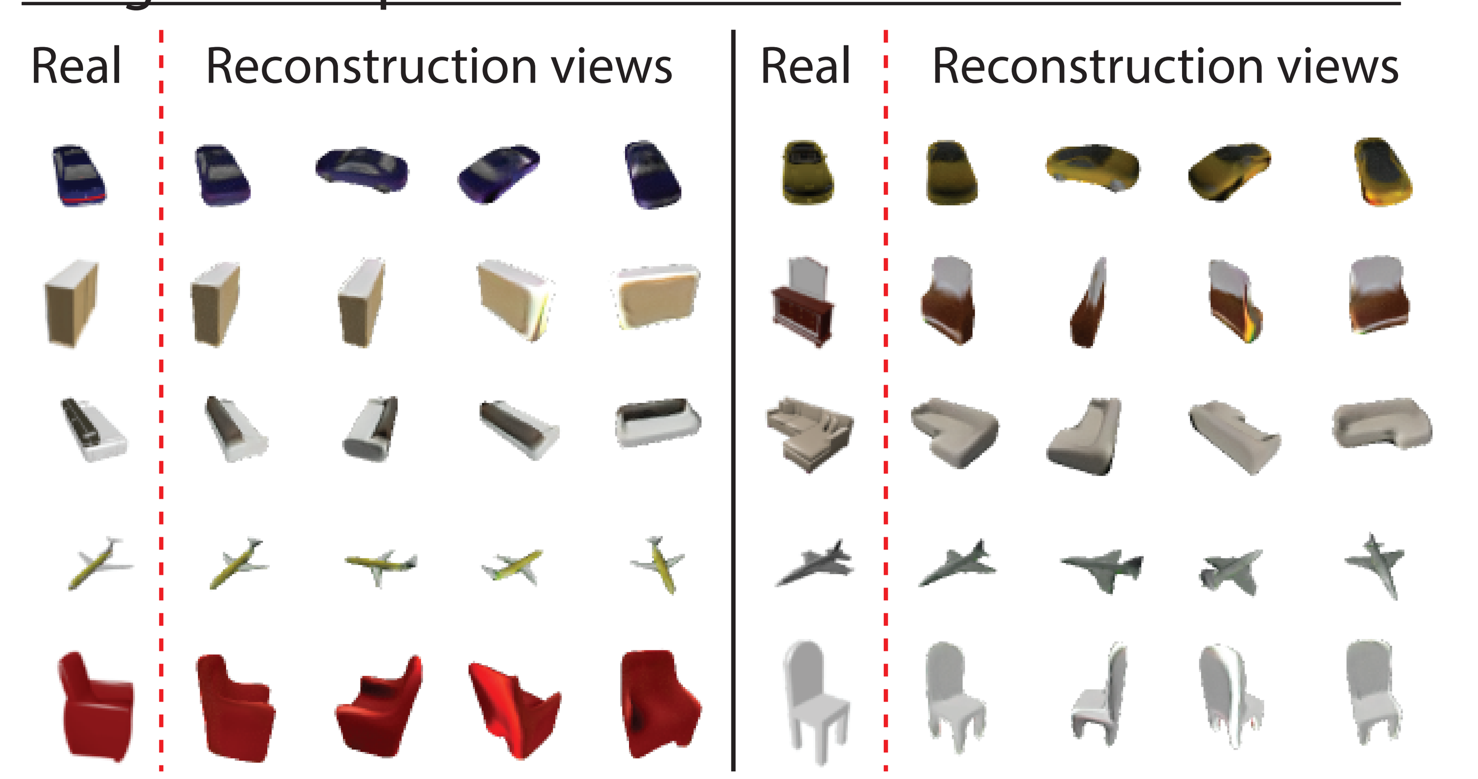


Image-to-shape translation learns 3D reconstruction



Unsupervised image correspondence

Markers: same colour = same template node

Limitations

- Noisy/low-res textures & geometry
- No explicit lighting model

- Topological constraints on template

- Occluded parts under-constrained