Cycle-Consistent Generative Rendering for 2D-3D Modality Translation
Tristan Aumentado-Armstrong, Alex Levinshtein, Stavros Tsogkas, Konstantinos G. Derpanis, and Allan D. Jepson
Samsung AI Centre Toronto

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.

Unpaired Data

2D Image Modality
Shape-to-Image Translation
3D Shape Modality
Image-to-Shape Translation

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

Shape-to-image translation learns generative modelling
Learned representation disentangles pose, shape, and texture

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

Results

Cross-modality object fidelity (cycle consistency)

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

\[ \mathcal{L}(\Phi) = D\left[S \mid \tilde{S}(I_f)\right] + D\left[I \mid \tilde{I}(S)\right] + \text{Regularization} \]
\[ C_{I \rightarrow \tilde{I}} \left[I, \tilde{I}(S)\right] + C_{\tilde{S} \rightarrow I}(\tilde{S}, \tilde{I}(\tilde{S})) + \mathcal{L}_R \]

Shape-to-image translation learns 3D reconstruction
Image-to-shape translation learns generative modelling

Change pose and shape
Change latent texture