Planes vs. Chairs: Category-guided 3D shape learning without any 3D cues
Zixuan Huang1, Stefan Stojanov1, Anh Thai1, Varun Jampani2, James M. Rehg1
1Georgia Institute of Technology, 2Google Research

Motivation

Large-scale annotated data empowered the great success of learning-based method in 2D computer vision tasks. However, 3D reconstruction from single images is still quite challenging. One key reason is the lack of annotated data at scale. How can we learn 3D shape reconstruction in a more scalable way?

Problem Framing

We propose learning under Multi-Category Single-View (MCSV) setting:
- No 3D cues such as viewpoints or multiple views for supervision;
- Learn a single unified model that works for all the categories.

MCSV learning
- is more scalable
- enables data pooling to learn category-agnostic features

But this learning setting makes the shape-viewpoint entanglement problem even harder to solve:

Can we better constrain the shape learning?

A 3D-unsupervised model that learn shapes of multiple object categories at once.
Project Page: https://zixuanh.com/multiclass3D

Categorical Metric Learning

We regularize the shape learning via category labels.

Architecture

E: image encoder  f_s: implicit SDF MLP  f_T: implicit RGB MLP
H: hypernetwork  V: viewpoint estimator  R: learnable renderer

Results

Quantitative ablation and SOTA comparison on ShapeNet-13:

<table>
<thead>
<tr>
<th>Methods</th>
<th>F-Score@1.0°F-Score@5.0°F-Score@10.0°</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>w/o category</td>
<td>0.1589</td>
<td>0.6261</td>
</tr>
<tr>
<td>w/o (L_{\text{metric}})</td>
<td>0.1875</td>
<td>0.6864</td>
</tr>
<tr>
<td>w/o (L_{\text{sim}})</td>
<td>0.1837</td>
<td>0.6741</td>
</tr>
<tr>
<td>w/o (L_{\text{ours}})</td>
<td>0.1846</td>
<td>0.6437</td>
</tr>
<tr>
<td>Ours</td>
<td>0.2005</td>
<td>0.7168</td>
</tr>
</tbody>
</table>

SDF-SRN: 0.1606 | 0.5441 | 0.7584 | 0.682 |

Limitation:
- training instability due to the adversarial regularization, particularly on real-world images with many categories

Qualitative ablation and SOTA comparison on ShapeNet-13:

Results on ShapeNet-55:

Results on Pascal3D+: