SPARC: Sparse Render-and-Compare for CAD model alignment in a Single RGB Image
Florian Langer, Gwangbin Bae, Ignas Budvytis, Roberto Cipolla
Department of Engineering, University of Cambridge

Problem
From a single RGB image predict aligned 3D shapes that represent the given scene.

RGB Image → Aligned 3D Shapes → Overlays

Method

From an RGB image, predict the normals, depths, and segmented regions.

2D Image Information
RGB and Predicted Normals → Predicted Depth

3D CAD Model Information
CAD model, Initial pose: Q, T, S → Sparse Render

Extra Information
Sparse Render → Reprojected NOCs

Updated Pose
Iteration 1 → Iteration 2 → Iteration 3

Discussion
- Sparse Render-and-Compare produces more accurate alignments compared to NOCs.
- Sparse Inputs can be rendered very fast and processed efficiently due to cross attention.
- Predicting pose update steps reduces the number of iterations needed to just 3 leading to a total runtime of 110 ms (compared to traditional render and compare requiring 100s of update steps with run times of ca. 30s).
- SPARC-Net does not seem able to make full use of all available information (particularly precise normal estimates) and does not improve when using more than 3 refinement steps.

Quantitative Results - ScanNet

<table>
<thead>
<tr>
<th>Method</th>
<th>Instance</th>
<th>Val Loss Sparse</th>
<th>Train Loss Sparse</th>
<th>Val Loss Whole Image</th>
<th>Train Loss Whole Image</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROCA</td>
<td>120</td>
<td>0.30</td>
<td>0.28</td>
<td>0.32</td>
<td>0.26</td>
</tr>
<tr>
<td>Ours</td>
<td>25.0</td>
<td>0.45</td>
<td>0.43</td>
<td>0.47</td>
<td>0.41</td>
</tr>
<tr>
<td>GT</td>
<td></td>
<td>0.30</td>
<td>0.28</td>
<td>0.32</td>
<td>0.26</td>
</tr>
</tbody>
</table>

Limitations:
- Depth information is required large number of iterations (12): 300, (3) 1000
- More iterations are needed to perform full render (Total time: 2.4 min, [3] 344)
- Requires very good initialization

References