A Point Set Generation Network for 3D Object Reconstruction from a Single Image
Hao Su*, Haoqiang Fan*, Leonidas J. Guibas

Motivation & Background
Task: 3D reconstruction from a single image

A deep learning approach

Key challenge: Representation issue of 3D data

Key points:
- Affability to learning
- Flexibility
- Geometric manipulability

Multi-view images
Volumetric occupancy
Depth map

A deep learning approach

Training data generation

Deep network

 Renders

Groundtruth point set

Approach

Basic pipeline

Network

Prediction

Groundtruth point set

Loss: Earth Mover’s Distance
Given two sets of points, measure their

Solves the optimal transportation

\[ d_{EMD}(S_1, S_2) = \min_{\phi \in \mathcal{S}} \sum_{x \in S_1} ||x - \phi(x)||_2 \]

where \( \phi : S_1 \rightarrow S_2 \) is a bijection.

We find EMD is differentiable almost everywhere

We implement a distributed approximation algorithm on CUDA based upon [Bertsekas, 1985], \(|1 + \cdot|\)-approximation

Network design: two-branch architecture

Visualization of two branches

Effect of combining two branches

Comparison w.r.t state-of-the-art

Evaluation

Per-category evaluation

<table>
<thead>
<tr>
<th>category</th>
<th>1 view</th>
<th>3 views</th>
<th>5 views</th>
<th>7 views</th>
</tr>
</thead>
<tbody>
<tr>
<td>plane</td>
<td>0.601</td>
<td>0.313</td>
<td>0.349</td>
<td>0.562</td>
</tr>
<tr>
<td>bench</td>
<td>0.509</td>
<td>0.421</td>
<td>0.502</td>
<td>0.527</td>
</tr>
<tr>
<td>cabinet</td>
<td>0.771</td>
<td>0.716</td>
<td>0.793</td>
<td>0.772</td>
</tr>
<tr>
<td>car</td>
<td>0.831</td>
<td>0.798</td>
<td>0.829</td>
<td>0.836</td>
</tr>
<tr>
<td>chair</td>
<td>0.544</td>
<td>0.466</td>
<td>0.533</td>
<td>0.550</td>
</tr>
<tr>
<td>monitor</td>
<td>0.552</td>
<td>0.468</td>
<td>0.545</td>
<td>0.565</td>
</tr>
<tr>
<td>lamp</td>
<td>0.862</td>
<td>0.781</td>
<td>0.815</td>
<td>0.847</td>
</tr>
<tr>
<td>speaker</td>
<td>0.727</td>
<td>0.662</td>
<td>0.708</td>
<td>0.737</td>
</tr>
<tr>
<td>lamp</td>
<td>0.694</td>
<td>0.544</td>
<td>0.593</td>
<td>0.690</td>
</tr>
<tr>
<td>lamp</td>
<td>0.708</td>
<td>0.633</td>
<td>0.596</td>
<td>0.690</td>
</tr>
<tr>
<td>table</td>
<td>0.655</td>
<td>0.533</td>
<td>0.564</td>
<td>0.580</td>
</tr>
<tr>
<td>monitor</td>
<td>0.749</td>
<td>0.661</td>
<td>0.732</td>
<td>0.754</td>
</tr>
<tr>
<td>waterfill</td>
<td>0.611</td>
<td>0.513</td>
<td>0.596</td>
<td>0.631</td>
</tr>
<tr>
<td>mean</td>
<td>0.680</td>
<td>0.560</td>
<td>0.617</td>
<td>0.631</td>
</tr>
</tbody>
</table>

Analysis

Visualization of two branches

3D shape completion

Effect of combining two branches

Comparison w.r.t state-of-the-art

63% Error reduction!