Hierarchical Surface Prediction

Christian Häne, Shubham Tulsiani, Jitendra Malik
Problem Statement

• Predict full geometry

Input image → Output geometry
Problem Statement

• Predict full geometry

Input image  →  Optional Surface Color Prediction
Related Work

• Coarse resolution voxel grid

  ![Diagram](image1)

  Girdhar et al. 2016

• We predict high resolution
  – Hierarchical Surface Prediction, Häne et al., 3DV 2017

• Concurrent work
  – OctNetFusion, Riegler et al., 3DV 2017
  – Octree Generating Networks, Tatarchenko et al., ICCV 2017
Approach

• Earlier works coarse resolution
  – Dense 3D occupancy grid

• Surfaces are 2D

• Hierarchical prediction
  – Fine resolution only around surface
  – Octree
Prediction in Depth-First Manner

- Image encoding
- Decode first level
- Determine octants with boundary

3 Labels (free space / boundary / occupied space)
Prediction in Depth-First Manner

- Cropping
- Upsampling
- Boundary octants
Prediction in Depth-First Manner

- Cropping
- Upsampling
- Boundary octants
Prediction in Depth-First Manner

- Cropping
- Upsampling
- Boundary octants
Prediction in Depth-First Manner

• Cropping
• Upsampling to final resolution
Prediction in Depth-First Manner

- Back to Closest Level with Unevaluated Blocks
Prediction in Depth-First Manner

- Cropping
- Upsampling
- Boundary Octants
Prediction in Depth-First Manner

- Cropping
- Upsampling to Final Resolution
Prediction in Depth-First Manner

- Back to Closest Level with Unevaluated Blocks
Prediction in Depth-First Manner

- Cropping
- Upsampling to final resolution
• Predicting all blocks builds the octree
• Supervision at each level
• Only voxels around surface predicted
• Evaluation of whole tree (too) slow for training
• Subsampling of the blocks
• Test time quality saturates
• We use 30%
Computational Benefit

- Number of evaluated voxels at each resolution
Baselines

- Dense Low Resolution (LR) Prediction at $32^3$
- Two ground truths
  - Hard (H), Max-Pooling of high resolution
  - Soft (S), Average-Pooling of high resolution
- Baselines upsampled to $256^3$ for evaluation
Quantitative Evaluation, Shapenet

• Mean over 13 Classes
  – Intersection over Union (IoU)
  – Chamfer Distance (CD)

<table>
<thead>
<tr>
<th>Method</th>
<th>RGB Input</th>
<th></th>
<th>Depth Input</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IoU</td>
<td>CD</td>
<td>IoU</td>
<td>CD</td>
</tr>
<tr>
<td>LR H</td>
<td>0.438</td>
<td>0.0246</td>
<td>0.414</td>
<td>0.0293</td>
</tr>
<tr>
<td>LR S</td>
<td>0.443</td>
<td>0.0287</td>
<td>0.417</td>
<td>0.0339</td>
</tr>
<tr>
<td>HSP (ours)</td>
<td>0.484</td>
<td>0.0211</td>
<td>0.526</td>
<td>0.0184</td>
</tr>
</tbody>
</table>
Qualitative Evaluation, Shapenet
Qualitative Evaluation, Shapenet
Color Results

• Output of our system at different resolutions
Conclusion

• High resolution voxel prediction
• Surface color

• Future work
  – Multi-view / depth map fusion
  – Scenes
  – Geometric loss function