Adaptive O-CNN: A Patch-based Deep Representation of 3D Shapes

Peng-Shuai Wang\textsuperscript{1,2} Chun-Yu Sun\textsuperscript{1,2} Yang Liu\textsuperscript{2} Xin Tong\textsuperscript{2}

\textsuperscript{1} Tsinghua University \quad \textsuperscript{2} Microsoft Research Asia
3D Learning for Shape Analysis and Synthesis
Our Goal

- A good 3D representation for shape learning
  - Compact: low memory and computational cost
  - Informative: good shape generation quality
Full Voxel based Representation

- Related work: [Wu et al. 2015]; [Choy et al. 2016]; [Wu et al. 2016] ...
- ✔ Intuitive extension of images
- ✗ Low resolution

3D ShapeNet [Wu et al. 2015]

3D R²N² [Choy et al. 2016]
Sparse Voxel based Representation

- Related work: [Wang et al. 2017]; [Tatarchenko et al. 2017]; [Riegler et al. 2017]...

- Support high resolution

- Low surface quality

O-CNN [Wang et al. 2017]  
OctGen [Tatarchenko et al. 2017]
Point based Representation

- Related work: [Qi et al. 2017]; [Su et al. 2017]; [Yin et al. 2018] ...
- ▶️ Flexible to use, effective for point cloud input
- ✗️ Generate scatter points, hard to extract surface

Images:
- PointNet [Qi et al. 2017]
- PSG [Su et al. 2017]
Mesh based Representation

- Related work: [Groueix et al. 2018]; [Kato et al. 2018]; [Sinha et al. 2017]...
- ✔ Better visual quality compared with PSG [Su et al. 2017]
- ✗ Irregular and distorted mesh elements, restricted topology

[Kato et al. 2018]  
AtlasNet [Groueix et al. 2018]
Key Observation

- Subdivide the octree considering the geometry variation

Octree: subdivide if non-empty

Adaptive octree: subdivide if non-empty && plane fitting error < $\delta$
Patch-Based Adaptive Octree

Planar patches at 4-level

Planar patches at 5-level

Planar patches at 6-level
Patch-Based Adaptive Octree

Planar patches
\[ \mathbf{n} \ast (\mathbf{x} - \mathbf{c}) + d = 0 \]

Planar patches at 4-level

Planar patches at 5-level

Planar patches at 6-level
Technical Challenges: Adaptive O-CNN

- Encoder network
  - How to deal with multi-resolution inputs

Adaptive Octree

Encoder

Feature Vector
Technical Challenges: Adaptive O-CNN

- Decoder network
  - How to generate the adaptive Octree

![Diagram of Feature Vector to Decoder]

12
Adaptive O-CNN Encoder

- Reference: O-CNN encoder
Adaptive O-CNN Encoder

Element max
Adaptive O-CNN Decoder

Prediction module

Empty

Surface-poorly-approximated
Adaptive O-CNN Decoder

Prediction module

Empty
Surface-poorly-approximated
Surface-well-approximated
Adaptive O-CNN Decoder

Prediction module

Empty

Surface-poorly-approximated

Surface-well-approximated
Adaptive O-CNN Decoder

Prediction module

Empty

Surface-poorly-approximated

Surface-well-approximated
Adaptive O-CNN Decoder
Adaptive O-CNN Decoder: Loss Function

- Octree node status: \( L_{structure} = \sum_l H_l \)
- Patch parameters: \( L_{patch} = \sum_l \frac{1}{N_l} \sum_i \| n_i - n_i^g \|^2 + | d_i - d_i^g | \)
Efficiency of Adaptive O-CNN

- Adaptive Octree: much less voxels compared with Octree
  - Titan X GPU; Batch size 32

Full Voxel $N^3$

Octree about $2.5 \times N^2$

Adaptive Octree about $0.8 \times N^2$
Time Efficiency

Average time for each forward and backward iteration

Time (msec)

Voxel CNN

O-CNN

Adaptive O-CNN

32^3  64^3  128^3  256^3
Memory Efficiency

![Memory Efficiency Graph](image)

- **GPU Memory**
- **Voxel CNN**
- **O-CNN**
- **Adaptive O-CNN**

Memory (GB) vs. Voxel Count

Memory Efficiency table:

<table>
<thead>
<tr>
<th>Voxel Count</th>
<th>Memory (GB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>32^3</td>
<td>0</td>
</tr>
<tr>
<td>64^3</td>
<td>1</td>
</tr>
<tr>
<td>128^3</td>
<td>2</td>
</tr>
<tr>
<td>256^3</td>
<td>3</td>
</tr>
</tbody>
</table>
## Results – Shape Classification

- **Dataset**: ModelNet40
- **Comparable testing accuracy**

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>PointNet [Qi et al. 2017]</td>
<td>89.2%</td>
<td>PointNet++ [Qi et al. 2017]</td>
<td>91.9%</td>
</tr>
<tr>
<td>VRN Ensemble [Brock et al. 2016]</td>
<td>95.5%</td>
<td>SubVolSup [Qi et al. 2016]</td>
<td>89.2%</td>
</tr>
<tr>
<td>OctNet [Riegler et al. 2017a]</td>
<td>86.5%</td>
<td>O-CNN [Wang et al. 2017]</td>
<td>90.6%</td>
</tr>
<tr>
<td>Kd-Network [Klokov and Lempitsky 2017]</td>
<td>91.8%</td>
<td>Adaptive O-CNN</td>
<td>90.5%</td>
</tr>
</tbody>
</table>
Results – 3D Autoencoder

• Dataset: ShapeNet55
  • 39,715 3D models from 13 categories

• Network

![Diagram of Encoder and Decoder]

![Bar chart showing Chamfer Distance]

Chamfer Distance

- PSG
- AtlasNet
- O-CNN
- AO-CNN
Visual Comparison

Visual Comparison

Ablation Study: Patch Primitive

- Patch primitive enables sub-voxel precision
Ablation Study: Adaptive Patches

- Adaptive octree produces less holes on the output

O-CNN(Patch)  Adaptive O-CNN
Results – Shape Completion

(a) Incomplete shape  (b) Ground-truth  (c) O-CNN(patch)  (d) Our results
Results – Shape from a single image

• Dataset: ShapeNet55 and its renderings
  • 39,715 3D models from 13 categories

• Network

![Diagram of encoder and decoder]

Chamfer Distance

- PSG
- AtlasNet
- AO-CNN
Visual Comparison

(a) Input image  (b) Ground-truth  (c) PSG  (d) AtlasNet  (e) Our results
Limitation and Future Work

• The output is not seamless mesh
  • Post-processing, mesh repair

• Currently only planar patch is used
  • Extension: general primitive such as quadratic surface patches
Conclusion

• Adaptive O-CNN
  • Patch-Guided adaptive octree
  • High memory and computational efficiency
  • High shape generation quality

• Code and data online
  • https://github.com/Microsoft/O-CNN