A large-scale CAD dataset of more than 150,000 models in 660 categories.

Motivation: 3D Shape Representation

• Only in Theory
• No State-of-the-arts
• Use 3D Shapes
• Limited to Instance Level Matching

Biederman 1987
Lambert 1988
Biederman et al. 1989
Duda & Hogg 2006
Philbin et al. 2007

Desirable Properties of a Good 3D Shape Representation:

• Data-driven: learn from data rather than predefined shape routines.
• Generic: any simple complex shapes rather than simple shape primitives.
• Compositional: compose simple shapes by assembling simple ones.
• Versatile: applicable to various vision tasks.

3D Deep Learning

• 3D ShapeNets (a Convolutional Deep Belief Network) learns the joint distribution of generic 3D shapes across object category.
• 3D ShapeNets (a Convolutional Deep Belief Network) learns the joint distribution of generic 3D shapes across object category.
• 3D ShapeNets (a Convolutional Deep Belief Network) learns the joint distribution of generic 3D shapes across object category.
• 3D ShapeNets (a Convolutional Deep Belief Network) learns the joint distribution of generic 3D shapes across object category.

Generic:

• Learn from data rather than predefined shape routines.

Versatile:

• Compositional: compose simple shapes by assembling simple ones.

Compositional:

• Versatile: applicable to various vision tasks.

Princeton ModelNet Dataset

A large-scale CAD dataset of more than 150,000 models in 660 categories.

<table>
<thead>
<tr>
<th>Object Categories</th>
<th>Examples of Chairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chair</td>
<td>Bed</td>
</tr>
<tr>
<td>Nightstand</td>
<td>Sofa</td>
</tr>
</tbody>
</table>

3D Feature Extraction

1. Convert 3D ShapeNets to a 3D CNN
2. Feature learning by back-propagation
3. Test for classification and retrieval

3D Shape Generation

1. Inference Process
   - Convert the depth map into volumetric representation.
   - Infer unknown space and label jointly by gibbs sampling.
2. Recognition and Completion
   - Identify free space, observed surface, unknown space.
   - Infer unknown space and label jointly by gibbs sampling.

Recognition and Completion

- Convert the depth map into volumetric representation.
- Infer unknown space and label jointly by gibbs sampling.
- Identify free space, observed surface, unknown space.
- Infer unknown space and label jointly by gibbs sampling.

Deep View Planning

The original entropy of the first view,

\[ H = H(y|x_1, x_2) \]

Given a fixed number of views, the conditional entropy for view \( i \),

\[ H_i = H(y|x_{i-1}, x_i) \]

The reduction of entropy is the mutual information between the new voxel and the label,

\[ I = H(y|x_{i-1}, x_i) - H(y|x_{i-1}) \]

We choose the view with the maximum mutual information

\[ V^* = \arg \max_i I(x_i) \]

Table 2: Accuracy for 2.5D Recognition on NYU dataset.

| View | Precision | Recall | | View | Precision | Recall |
|------|-----------|--------| |------|-----------|--------|
| L1   | 0.85      | 0.80   | | L2   | 0.80      | 0.75   |
| L3   | 0.70      | 0.65   | | L4   | 0.75      | 0.70   |
| L5   | 0.65      | 0.60   | | L6   | 0.70      | 0.65   |

The work is funded by Intel, NVIDIA, and Princeton University.