### Advantages:

- Better optimization and learning
- Leverage large-scale RGBD data for constraints
- 3D Scene Understanding

### Geometry-driven Parts (gDPM)

Parts based on consistent underlying 3D geometry

### What is the right way to select parts?

**Unsupervised Parts**

- Heuristic initialization, e.g., gradient magnitudes.

**Supervised Parts**

- Key-point/part annotation, e.g., anatomical.

### Our Approach

**Effectiveness of DPMs + Richness of Geometric Representation**

**Geometry-driven Dictionary of 3D Elements**

Desired properties:

- Representative: frequent among objects
- Spatially consistent w.r.t the object

**Learning gDPM**

Enforcing geometric constraints using large-scale RGBD data.

Input: $x = \{I, F^l, l\}, y \in \{-1, 1\}$


Minimize: $L_D(\beta) = \frac{1}{2}\|\beta\|^2 + C \sum_{i=1}^{N} \max(0, 1 - y_i f_\beta(x_i))$

$$f_\beta(x) = \begin{cases} S_{\text{Appearance}}(I, z, \beta_i) & \text{Standard Appearance Score of root and part filters.} \\
S_{\text{Appearance}}(I, z, \beta_i) + \sum_{i=1}^{n} S_G(e', \omega(l^2, l'_i)), & \text{Geometry constraints on part placements.} \\
\omega(l) = \text{Raw surface normal at } S_C(e') = \text{Similarity function between surface normal maps} \\
\end{cases}$$

**Check Geometric Consistency and Merge.**

Dictionary ($D$) of Geometrically-consistent 3D elements ($100$s clusters, $3$ Aspect Ratio)

**From 3D Parts to Object Hypothesis**

Greedyly select N parts based on frequency and geometric consistancy to make an object hypothesis: $p = [p^1, \ldots, p^N]$, where $p^k = (e, l^2)$, $(N \sim [6, 12])$

**Quantitative Results**

<table>
<thead>
<tr>
<th>Type</th>
<th>Bed</th>
<th>Chair</th>
<th>M+TV</th>
<th>Sofa</th>
<th>Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPM (No Parts)</td>
<td>20.94</td>
<td>10.69</td>
<td>6.38</td>
<td>5.51</td>
<td>2.73</td>
</tr>
<tr>
<td>DPM</td>
<td>22.39</td>
<td><strong>14.44</strong></td>
<td>8.10</td>
<td>7.16</td>
<td>3.53</td>
</tr>
<tr>
<td>DPM (Our Parts, No Latent)</td>
<td>26.59</td>
<td>5.71</td>
<td>2.35</td>
<td>6.82</td>
<td>3.41</td>
</tr>
<tr>
<td>DPM (Our Parts)</td>
<td>29.15</td>
<td>11.43</td>
<td>4.17</td>
<td>8.30</td>
<td>1.76</td>
</tr>
<tr>
<td>gDPM</td>
<td><strong>33.39</strong></td>
<td>13.72</td>
<td><strong>9.28</strong></td>
<td><strong>11.04</strong></td>
<td><strong>4.05</strong></td>
</tr>
</tbody>
</table>

### Experimental Results

**Qualitative Results**

- **Input Image**
- **DPM Detection**
- **gDPM Detection**
- **Predicted Geometry**

**Test Set: NYU v2 RGB Images**

**Figure 3. A few examples of resulting elements in dictionary after refinement procedure.**

**Figure 2. A few elements from the dictionary after the initialization step.**

**Table 1. AP performance on the task of object detection.**

**Building Part-based Object Detectors via 3D Geometry**

Abhinav Shrivastava and Abhinav Gupta