A Coarse-to-Fine Model for 3D Pose Estimation and Sub-category Recognition
–Supplementary Material–

Roozbeh Mottaghi¹, Yu Xiang²,³, and Silvio Savarese³
¹Allen Institute for AI, ²University of Michigan-Ann Arbor, ³Stanford University

We provide per-category results in Tables 1 and 2. These tables correspond to Tables 1 and 2 in the paper, respectively.

### Table 1

<table>
<thead>
<tr>
<th>Bounding Box</th>
<th>All</th>
<th>Sub-category &amp; Viewpoint</th>
<th>Sub-category</th>
<th>Viewpoint (8 views)</th>
</tr>
</thead>
<tbody>
<tr>
<td>aeroplane</td>
<td>69.1</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>boat</td>
<td>29.3</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>car</td>
<td>53.8</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

**RCNN [2]**

| aeroplane    | 40.5| 28.6        |
| boat         | 6.5 | 3.2         |
| car          | 41.6| 16.6        |

**DPM-VOC+VP [3]**

| aeroplane    | 39.8| 23.8        |
| boat         | 5.8 | 1.0         |
| car          | 37.3| 23.8        |

**V-DPM [1]**

| aeroplane    | 41.8| 24.1        |
| boat         | 6.7 | 4.7         |
| car          | 34.9| 23.0        |

| aeroplane    | 40.0| 25.8        |
| boat         | 5.4 | 4.7         |
| car          | 32.0| 19.5        |

**Flat model**

| aeroplane    | 69.2| 40.3        |
| boat         | 29.5| 21.9        |
| car          | 56.1| 23.2        |

| Full model   | 69.2| 40.3        |
| Separate     | ✓   | 27.0        |

**Separate**

| aeroplane    | 60.2| 47.6        |
| boat         | 29.5| 24.8        |
| car          | 56.1| 20.6        |

Table 1. Results of variation of DPM [1], DPM-VOC+VP [3] and RCNN [2] on PASCAL3D+ [4] for all three or a subset of tasks. The result of DPM-VOC+VP [3] was adopted from [4]. The first column (‘Bounding Box’) is equivalent to the standard detection AP of PASCAL VOC. The meaning of ✓ is that the method is not capable of doing that task.

### Table 2

<table>
<thead>
<tr>
<th>Bounding Box</th>
<th>All</th>
<th>Sub-category &amp; Viewpoint</th>
<th>Sub-category</th>
<th>Viewpoint (8 views)</th>
</tr>
</thead>
<tbody>
<tr>
<td>aeroplane</td>
<td>69.1</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>boat</td>
<td>26.3</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>car</td>
<td>53.0</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

**1-layer hierarchy (ours)**

| aeroplane    | 40.5|         |
| boat         | 9.8 |         |
| car          | 36.4|         |

**2-layer hierarchy (ours)**

| aeroplane    | 28.6| 41.3    |
| boat         | 21.3| 9.6     |
| car          | 16.1| 37.7    |

**3-layer hierarchy (ours)**

| aeroplane    | 28.7| 40.8    |
| boat         | 8.8 | 10.6    |
| car          | 13.4| 18.2    |

| Full model   | 69.2| 40.3    |
| Separate     | ✓   | 27.0    |

**Flat model**

| aeroplane    | 69.2| 38.3    |
| boat         | 29.5| 10.6    |
| car          | 56.1| 29.0    |

**Separate**

| aeroplane    | 60.2| 39.2    |
| boat         | 29.5| 10.1    |
| car          | 56.1| 36.7    |

Table 2. Results of variations our hierarchical model, a flat model that uses the same set of features as those of the 3-layer hierarchy, and also separate classifiers on PASCAL3D+[4]. ³ We consider the same confidence values as the 3-layer model. So the bounding box detection results are identical.
Sub-category information:

We considered the following sub-categories in our experiments:

- **Aeroplane**: Airline, Fighter, Propeller, Shuttle
- **Boat**: Cabin, Cruise, Rowing, Sailing
- **Car**: Hatchback, Mini, Minivan, Race, Sedan, SUV, Truck, Wagon

Confusion matrices for sub-category recognition:

![Confusion matrices](image)

Figure 1. Confusion matrices for sub-category recognition. The recall rate is 81.7, 66.0, and 71.1% for aeroplane, boat, and car, respectively.

References


