

# Supplementary Material for the Paper “Object Detection by 3D Aspectlets and Occlusion Reasoning”

Yu Xiang  
University of Michigan  
[yuxiang@umich.edu](mailto:yuxiang@umich.edu)

Silvio Savarese  
Stanford University  
[ssilvio@stanford.edu](mailto:ssilvio@stanford.edu)

## 1. Voxel Representation

In the 3D objects prior of our Spatial Layout Model (SLM), we compute the intersection and union of two volumes based a voxel representation of our 3D object model. Fig. 1 shows the voxel representations of car, bed, chair, sofa and table in our experiments.

## 2. Learned 3D Aspectlets

We have learned 50, 32, 46, 24 and 25 3D aspectlets for car, bed, chair, sofa and table respectively, which are showed in Fig. 2 - 6. In each 3D model of these figures, the blue atomic aspect parts belong to the 3D aspectlet.

## 3. 3D Aspectlet Detection Examples

In our model, 3D aspectlets provide evidences for partial observation of an object. The likelihood of the location of an 3D aspectlet is modeled with a Conditional Random Field (CRF). By applying the CRF model to an input image, we obtain the detections of the 3D aspectlet. Fig. 7 and Fig. 8 show some anecdotal detection results of the 3D aspectlets of car, bed, chair, sofa and table on the outdoor-scene dataset and the indoor-scene dataset respectively. For each object in the images, we only show one correctly detected 3D aspectlet. We do not show all the correct detections of 3D aspectlets to avoid clutter.

## 4. Anecdotal Results

We present additional anecdotal results on our outdoor-scene dataset and our indoor-scene dataset obtained from SLM in Fig.9 - 12. In each of these figures, “input image” is the image used for testing, “2D detection” shows the detected objects in the input image, “3D spatial layout” shows the spatial layout of the objects and the camera in 3D, and “2D object mask” shows the occlusion order in the image.

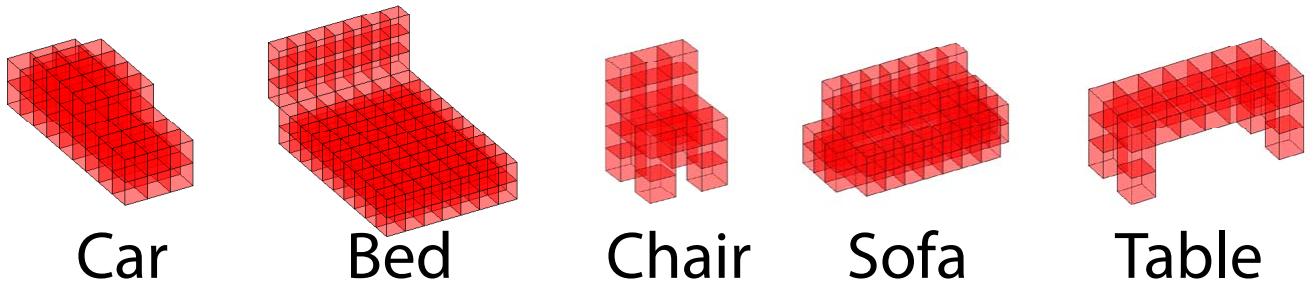


Figure 1. Voxel representations of the five categories in our experiments, which are used in computing the intersection and union of two volumes.

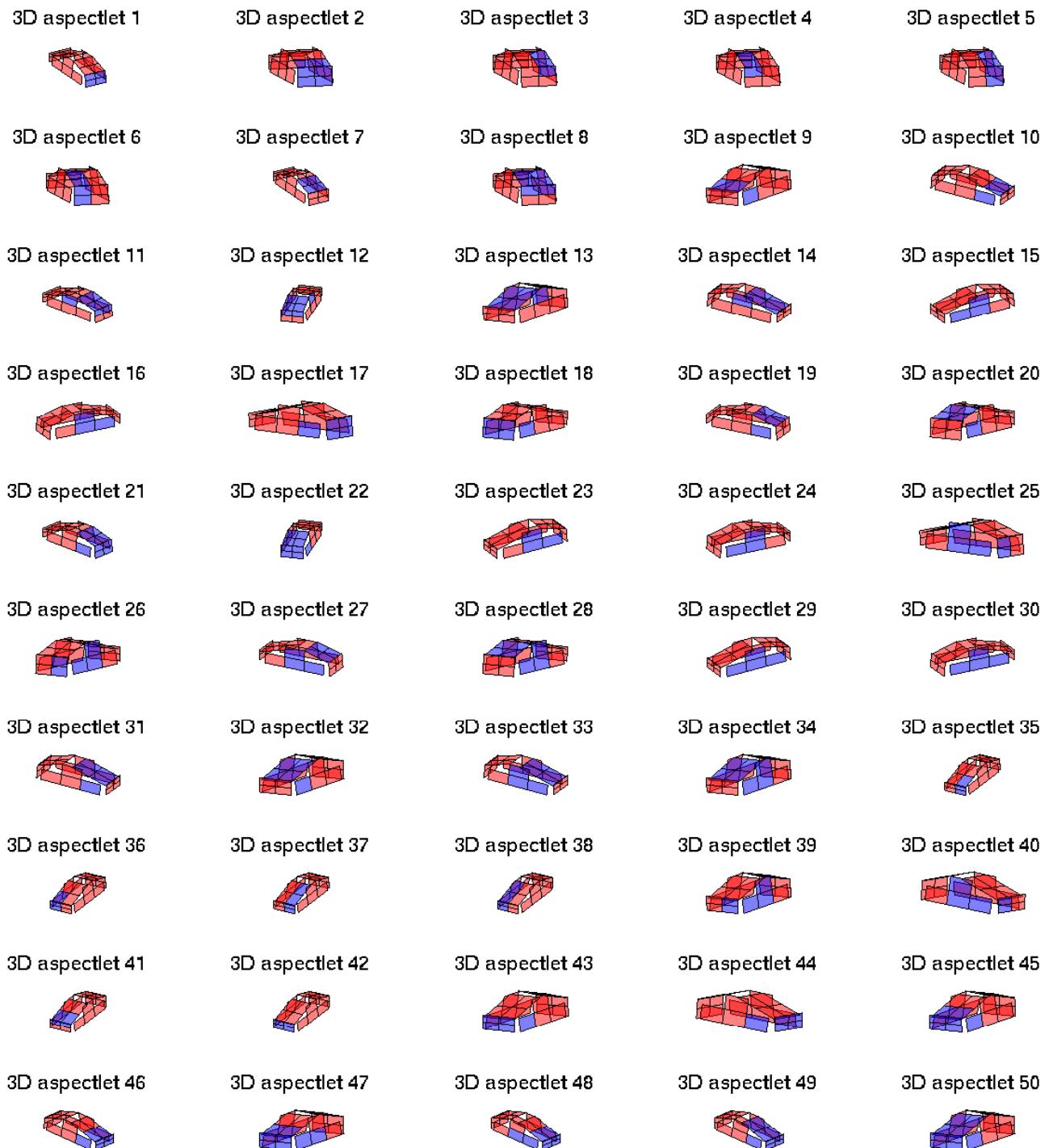


Figure 2. 50 3D aspectlets of car learned in our experiments. In each 3D model, the blue atomic aspect parts belong to the 3D aspectlet.

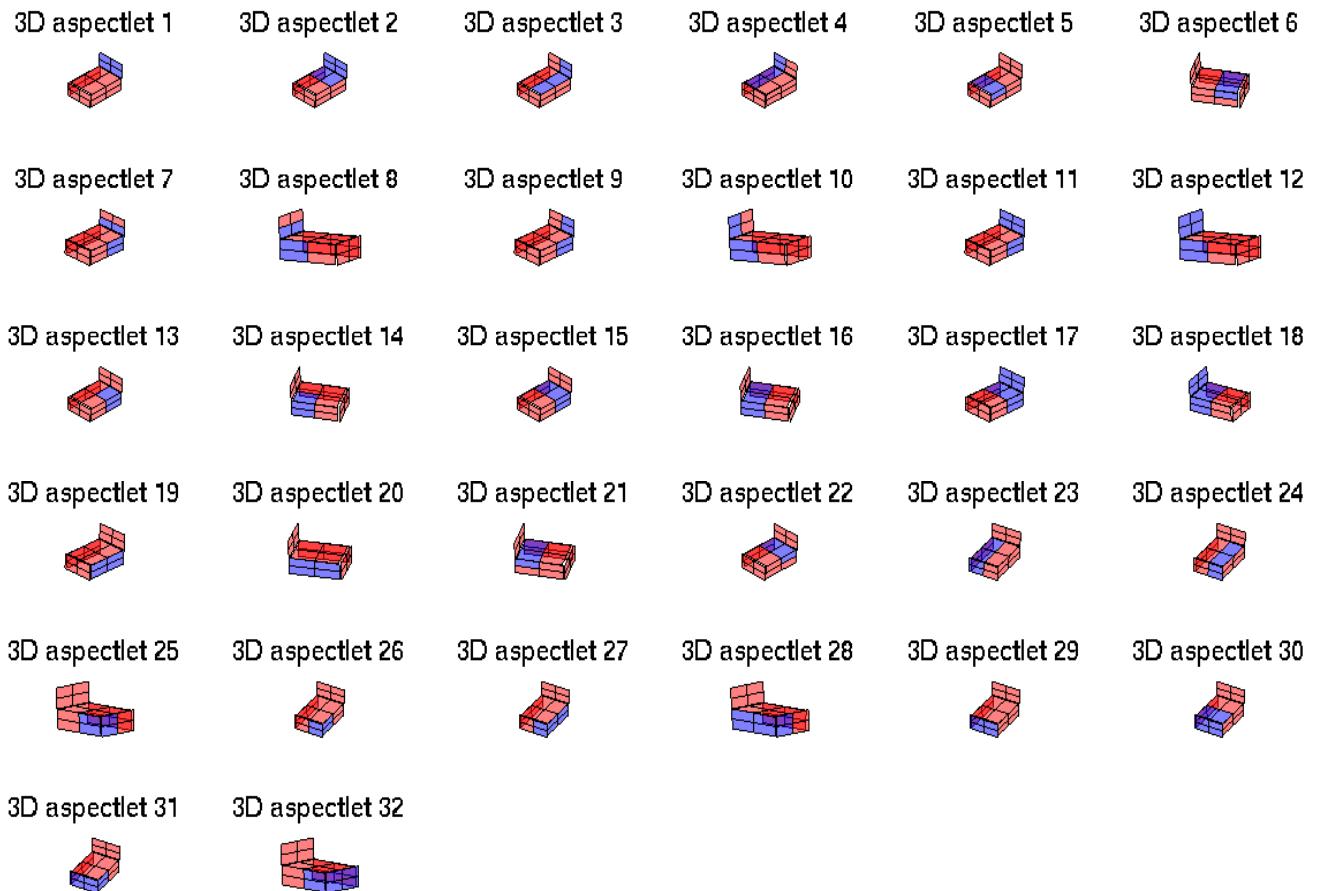


Figure 3. 32 3D aspectlets of bed learned in our experiments. In each 3D model, the blue atomic aspect parts belong to the 3D aspectlet.

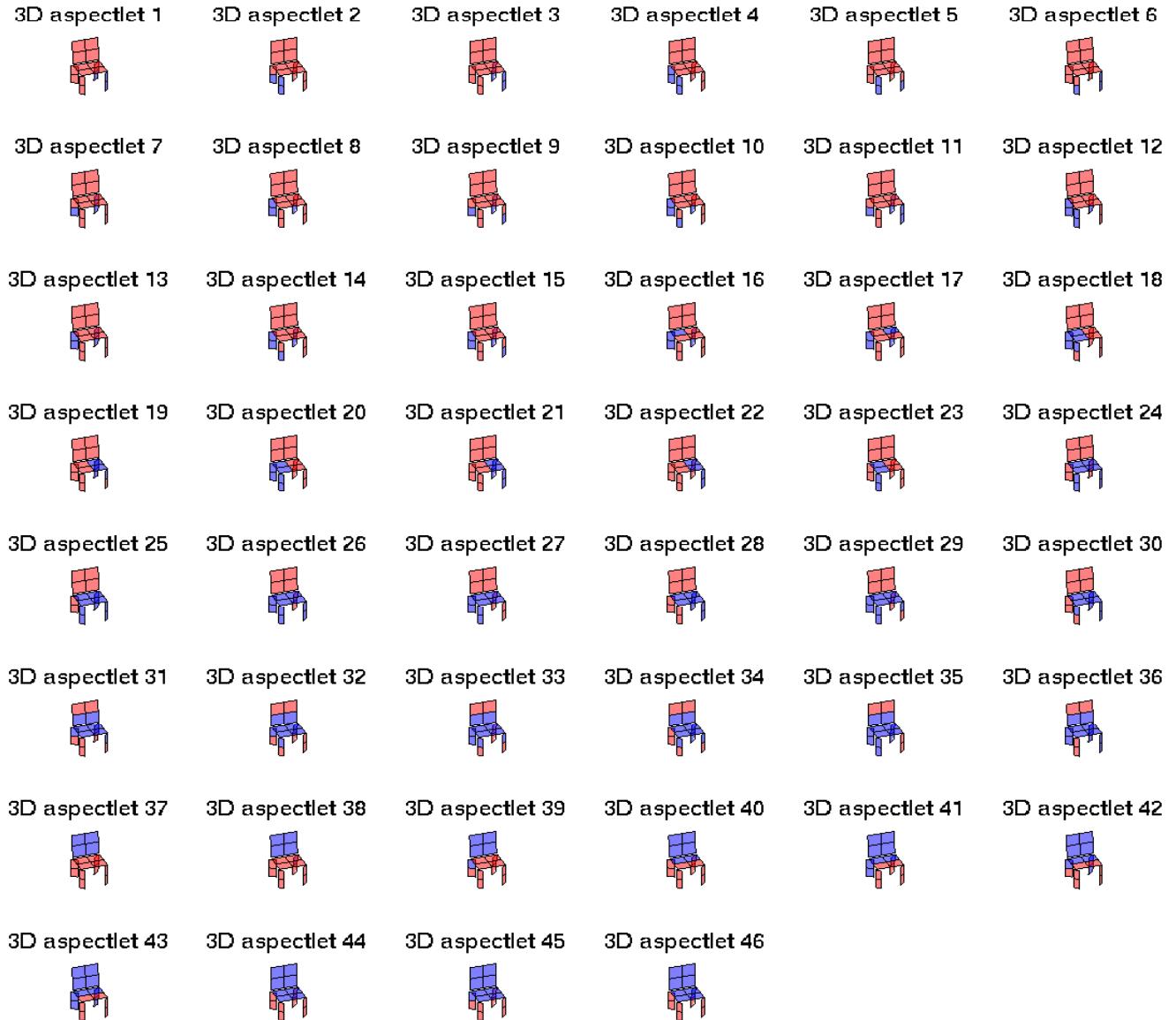


Figure 4. 46 3D aspectlets of chair learned in our experiments. In each 3D model, the blue atomic aspect parts belong to the 3D aspectlet.

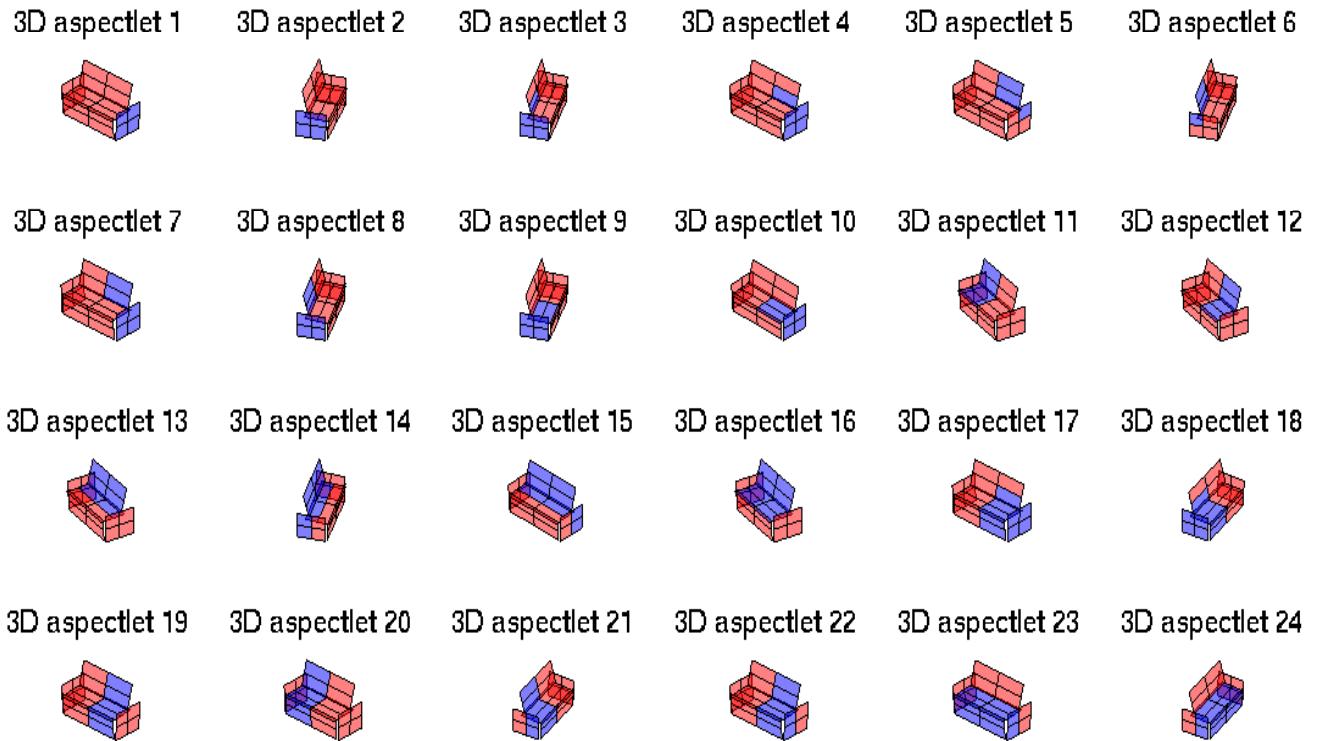


Figure 5. 24 3D aspectlets of sofa learned in our experiments. In each 3D model, the blue atomic aspect parts belong to the 3D aspectlet.

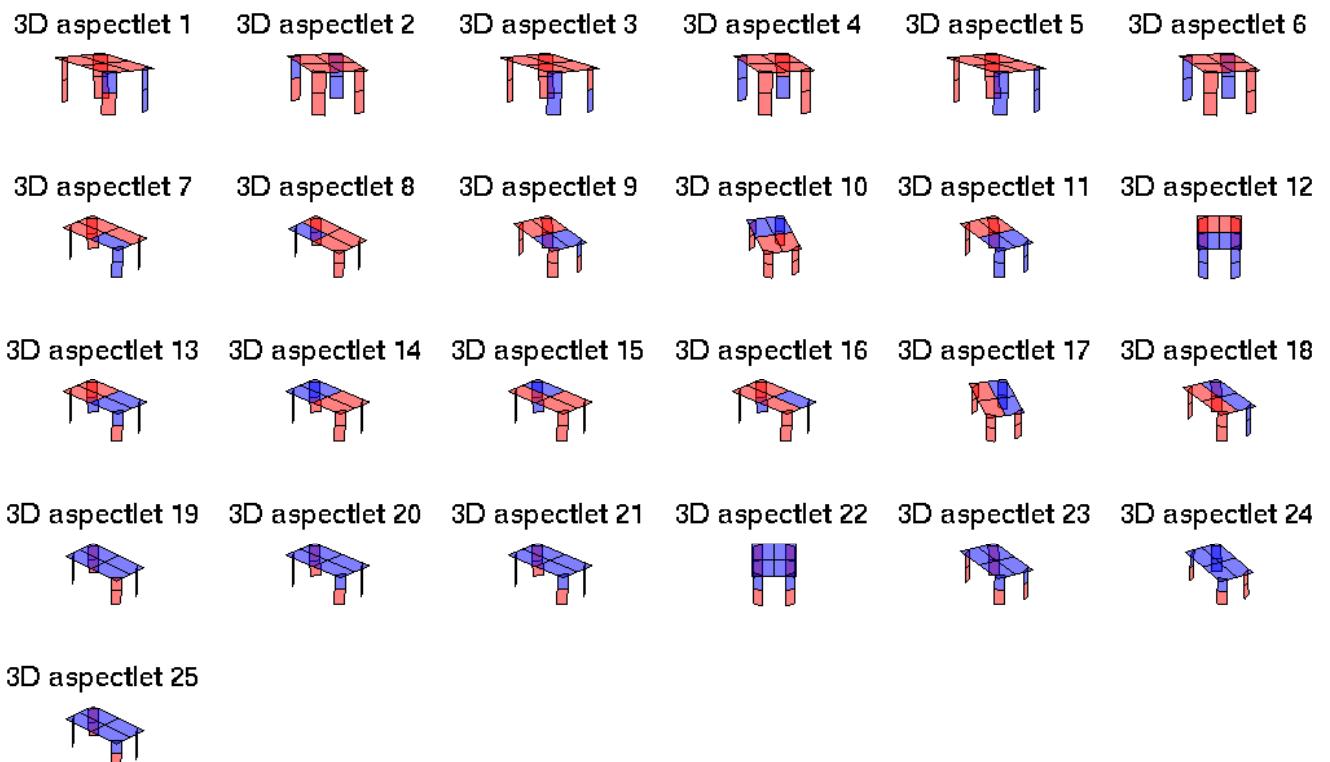


Figure 6. 25 3D aspectlets of table learned in our experiments. In each 3D model, the blue atomic aspect parts belong to the 3D aspectlet.

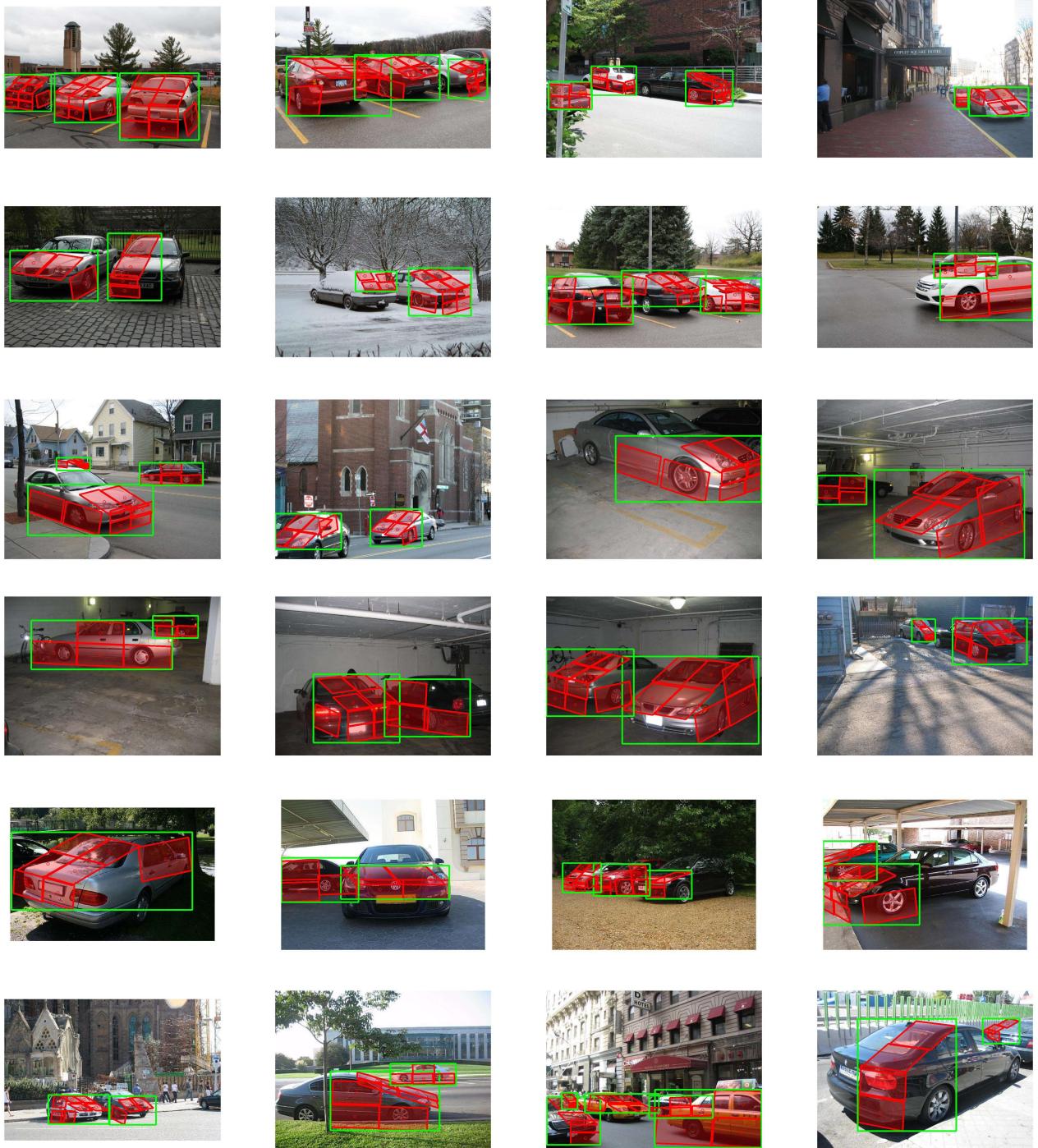
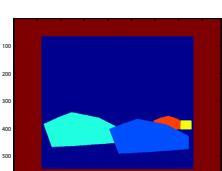
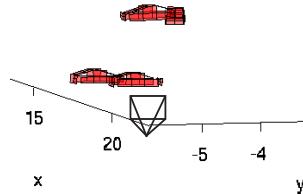
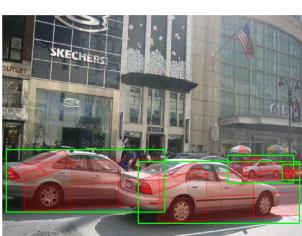
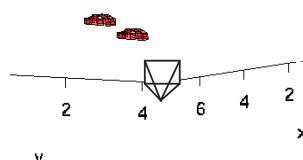
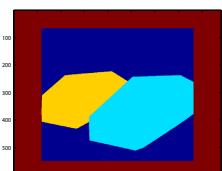
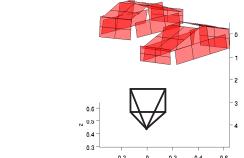
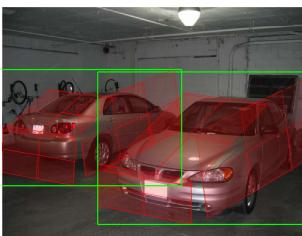
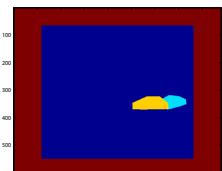
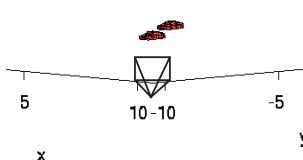
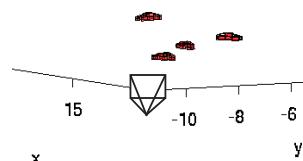
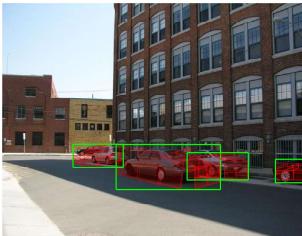
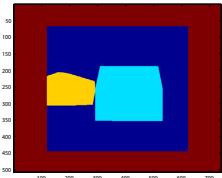
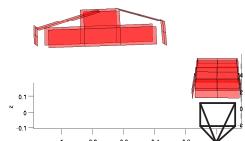
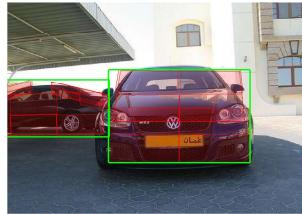


Figure 7. Anecdotal detection results of the 3D aspectlets of car on the outdoor-scene dataset. For each object in the images, we only show one correctly detected 3D aspectlet. We do not show all the correct detections of 3D aspectlets to avoid clutter.



Figure 8. Anecdotal detection results of the 3D aspectlets of bed, chair, sofa and table on the indoor-scene dataset. For each object in the images, we only show one correctly detected 3D aspectlet. We do not show all the correct detections of 3D aspectlets to avoid clutter.



input image

2D detection

3D spatial layout

2D object mask

Figure 9. Anecdotal detection results on the outdoor-scene dataset. The 2D detections show the detected objects in the images. The 3D plots show the spatial layout of objects and camera in 3D. The 2D object masks show the occlusion order in the images.

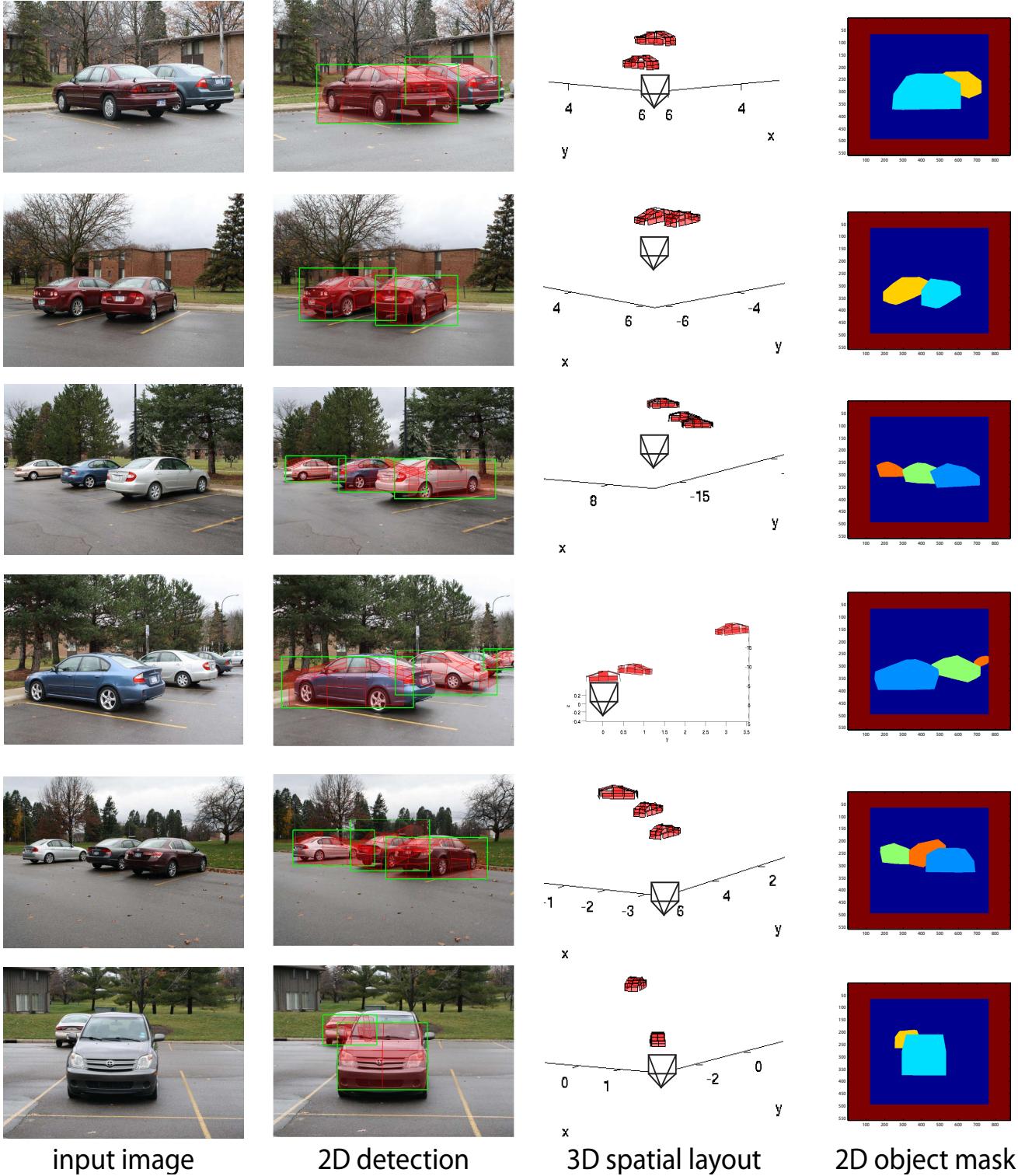
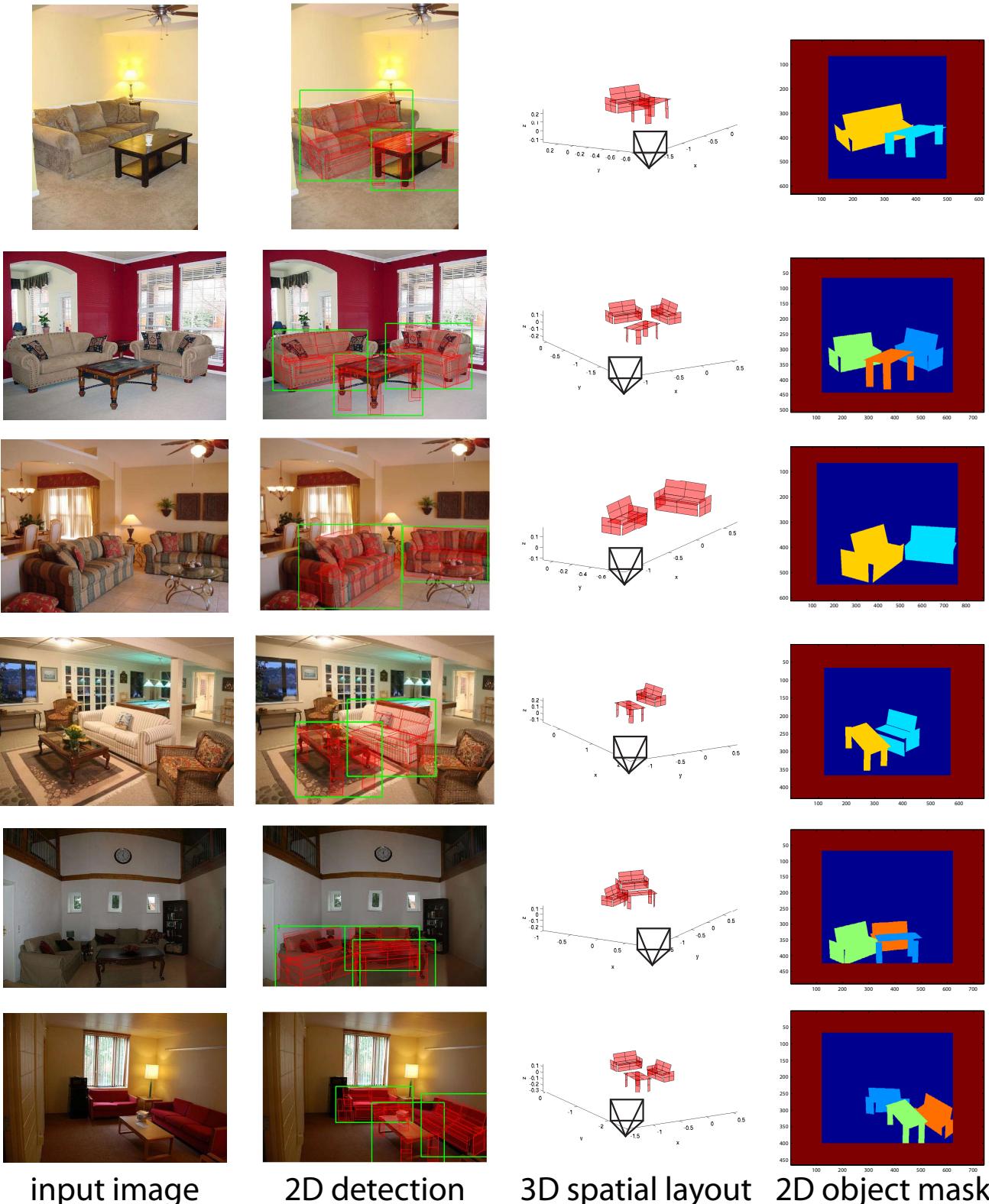


Figure 10. Anecdotal detection results on the outdoor-scene dataset. The 2D detections show the detected objects in the images. The 3D plots show the spatial layout of objects and camera in 3D. The 2D object masks show the occlusion order in the images.



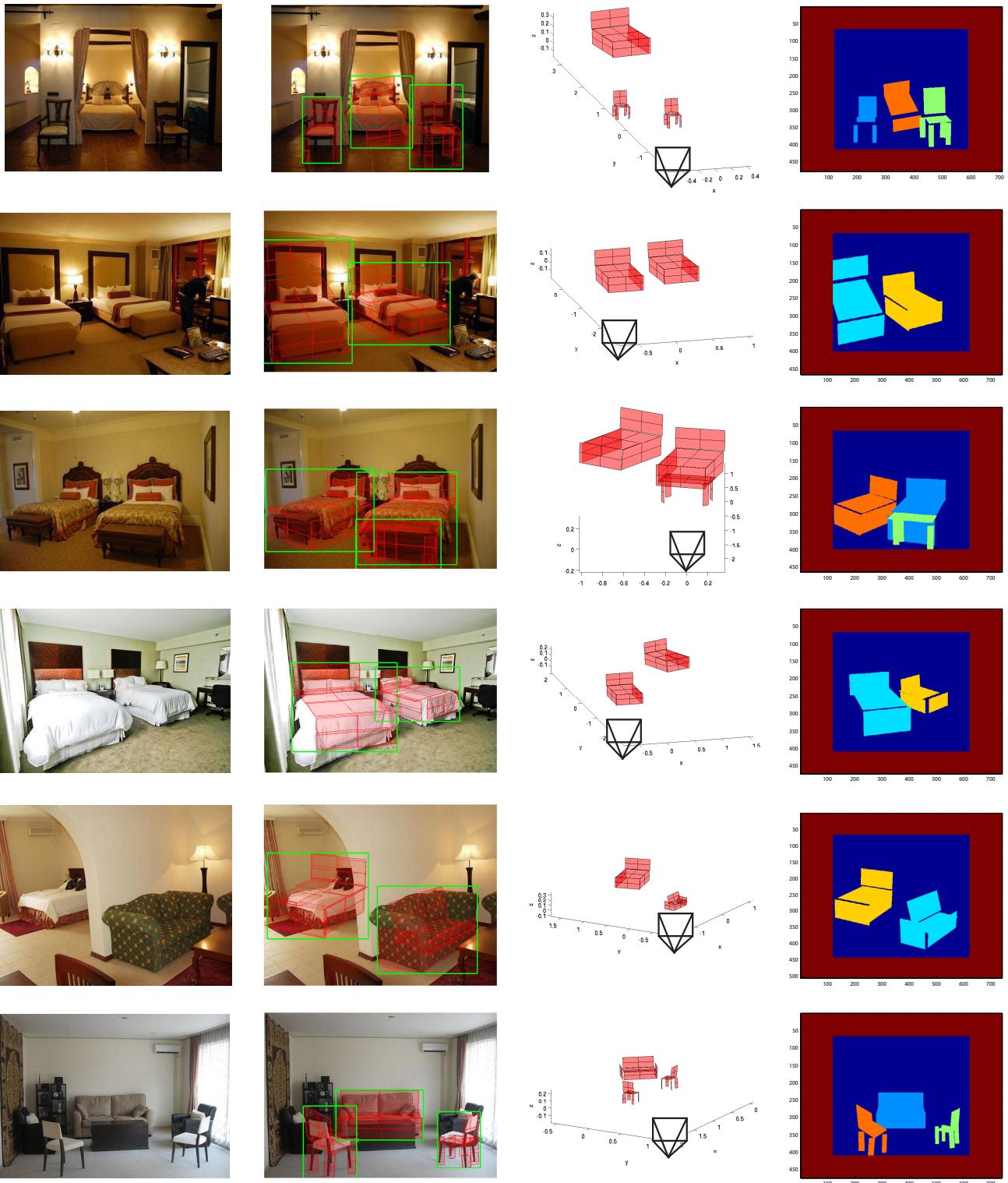
**input image**

**2D detection**

**3D spatial layout**

**2D object mask**

Figure 11. Anecdotal detection results on the indoor-scene dataset. The 2D detections show the detected objects in the images. The 3D plots show the spatial layout of objects and camera in 3D. The 2D object masks show the occlusion order in the images.



**input image**

**2D detection**

**3D spatial layout**

**2D object mask**

Figure 12. Anecdotal detection results on the indoor-scene dataset. The 2D detections show the detected objects in the images. The 3D plots show the spatial layout of objects and camera in 3D. The 2D object masks show the occlusion order in the images.