# Supplementary material
# Deep MANTA: A Coarse-to-fine Many-Task Network for joint 2D and 3D vehicle analysis from monocular image

Florian Chabot[1] , Mohamed Chaouch[1] , Jaonary Rabarisoa[1] , Céline Teulière[2] , Thierry Chateau[2]
[1] CEA-LIST Vision and Content Engineering Laboratory, [2] Pascal Institute, Blaise Pascal University
[1]{florian.chabot, mohamed.chaouch, jaonary.rabarisoa}@cea.fr
[2]{celine.teuliere, thierry.chateau}@univ-bpclermont.fr

In this supplementary material, we provide some qualitative results for the Deep MANTA approach: results on the Kitti test set (Figure 2) and a video demonstration (link). Figure 1 is a screenshot from the video demonstration. Note that no temporal information are used to generate it.

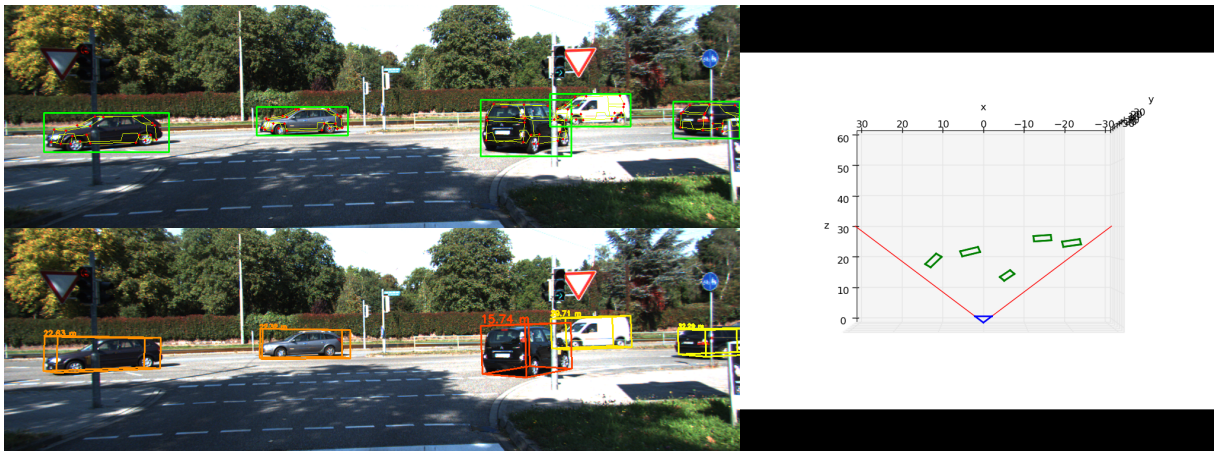

Figure 1. Screenshot from the demonstration video. *Top left*: 2D detection, part localization and part visibility. Red dots: visible parts, green dots: occluded parts. Self-occluded parts are not represented for better readability. *Bottom left*: projected 3D bounding boxes and the distance to the camera. *Right*: top view of the scene (the camera is represented in blue).

Figure 2. Image results on Kitti test set. For each image, 2D detections, parts localization and part visiblity are represented. Red dots: visible parts, green dots: occluded parts, blue dots: self-occluded parts. Under each image, we show 3D parts and 3D bounding boxes. The camera is represented in blue.