ApolloCar3D: A Large 3D Car Instance Understanding Benchmark for Autonomous Driving - Supplementary Materials

Xibin Song[1,2], Peng Wang[1,2], Dingfu Zhou[1,2], Rui Zhu[3], Chenye Guan[1,2],
Yuchao Dai[4], Hao Su[3], Hongdong Li[5,6] and Ruigang Yang[1,2]

[1]Baidu Research [2]National Engineering Laboratory of Deep Learning Technology and Application, China [3]University of California, San Diego [4] Northwestern Polytechnical University, Xi'an, China [5] Australian National University, Australia [6]Australian Centre for Robotic Vision, Australia

{songxibin,wangpeng54,zhoudingfu,guanchenye,yangruigang}@baidu.com,
{rzhu,haosu}@eng.ucsd.edu, daiyuchao@gmail.com and hongdong.li@anu.edu.au

Content.

In this supplementary materials, we provide

1. Detailed annotation definition of each keypoint.

2. Implementation details and performance of our trained keypoint detector.

3. Details of the function $B(K_c)$ and neighborhood function $N(c, M, \kappa)$ in Eq. (3) of the main paper.

4. Additional visualizations and analyses of our proposed ApolloCar3D dataset.

1. Keypoints annotation definition

Here we list the definition of the 66 semantic keypoints, as shown in Fig. 1,

- 0: Top left corner of left front car light
- 1: Bottom left corner of left front car light
- 2: Top right corner of left front car light

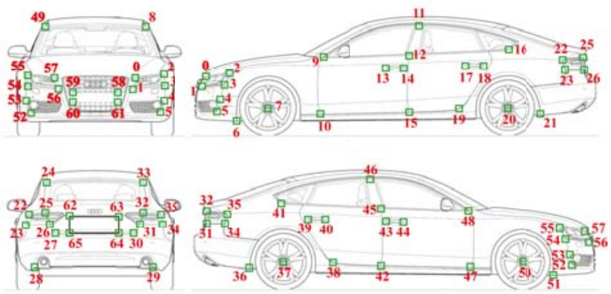

Figure 1: 3D keypoints definition for car models. 66 keypoints are defined for each model.

- 3: Bottom right corner of left front car light
- 4: Top right corner of left front fog light
- 5: Bottom right corner of left front fog light
- 6: Front section of left front wheel
- 7: Center of left front wheel
- 8: Top right corner of front glass
- 9: Top left corner of left front door
- 10: Bottom left corner of left front door
- 11: Top right corner of left front door
- 12: Middle corner of left front door
- 13: Front corner of car handle of left front door
- 14: Rear corner of car handle of left front door
- 15: Bottom right corner of left front door
- 16: Top right corner of left rear door
- 17: Front corner of car handle of left rear door
- 18: Rear corner of car handle of left rear door
- 19: Bottom right corner of left rear door
- 20: Center of left rear wheel
- 21: Rear section of left rear wheel
- 22: Top left corner of left rear car light

- 23: Bottom left corner of left rear car light
- 24: Top left corner of rear glass
- 25: Top right corner of left rear car light
- 26: Bottom right corner of left rear car light
- 27: Bottom left corner of trunk
- 28: Left corner of rear bumper
- 29: Right corner of rear bumper
- 30: Bottom right corner of trunk
- 31: Bottom left corner of right rear car light
- 32: Top left corner of right rear car light
- 33: Top right corner of rear glass
- 34: Bottom right corner of right rear car light
- 35: Top right corner of right rear car light
- 36: Rear section of right rear wheel
- 37: Center of right rear wheel
- 38: Bottom left corner of right rear car door
- 39: Rear corner of car handle of right rear car door
- 40: Front corner of car handle of right rear car door
- 41: Top left corner of right rear car door
- 42: Bottom left corner of right front car door
- 43: Rear corner of car handle of right front car door
- 44: Front corner of car handle of right front car door
- 45: Middle corner of right front car door
- 46: Top left corner of right front car door
- 47: Bottom right corner of right front car door
- 48: Top right corner of right front car door
- 49: Top left corner of front glass
- 50: Center of right front wheel
- 51: Front section of right front wheel
- 52: Bottom left corner of right fog light
- 53: Top left corner of right fog light
- 54: Bottom left corner of right front car light
- 55: Top left corner of right front car light

Method	mean pixel error	detection rate
CPM [1]	$4.39(px)$	75.41%
Human label	$2.67(px)$	92.40%

Table 1: Keypoints accuracy.

- 56: Bottom right corner of right front car light
- 57: Top left corner of right front car light
- 58: Top right corner of front license plate
- 59: Top left corner of front license plate
- 60: Bottom left corner of front license plate
- 61: Bottom right corner of front license plate
- 62: Top left corner of rear license plate
- 63: Top right corner of rear license plate
- 64: Bottom right corner of rear license plate
- 65: Bottom left corner of rear license plate

2. Keypoints Accuracy

Tab. 1 shows the accuracy of 2d keypoints. For each predicted keypoint, if its distance to ground truth keypoint is less than $10(pixel)$, we regard it as positive, otherwise, it is regarded as negative. We first crop out each car using its ground truth mask, then use CPM [1] to train the 2d keypoints detector. The detection rate is $75.41\ \%$(rate of number of positive keypoints and all ground truth), and the mean pixel error is $4.39\ px$. We also show the accuracy of human labeled keypoints. The detection rate of human labeled 2d keypoints is 92.40%, and the mean pixel error of detected 2d keypoints is $2.67(pixel)$. As discussed in the paper, the mis-labelling of human is primarily because humans cannot accurately memorize the semantic meaning of all the 66 keypoints. However, it is still much better than a trained CPM keypoint detector because the robustness of human with respect to appearance and occlusion changes.

3. Car pose estimation

To judge whether a car needs to use contextual constrains, we define the condition $B(K_c)$ in Eq. (3) for a car instance as the number of annotated keypoints is greater than 6, and the labelled keypoints are lying on more than two predefined car surfaces (detailed in tab. 2).

Otherwise, we additionally use $N(c, M, \kappa)$, which is a κ nearest neighbor function, to find spatial close car instances and regularize the solved poses. Specifically, the metric for retrieve neighborhood is the distance between mean coordinates of labelled keypoints. Here we set $\kappa = 2$.

(a) (b) (c) (d)

Figure 2: Visualization results of different approaches, in which (a) the input image, (b) and (c) are the results with direct regression method and key points-based method with context constraint. (d) gives the ground truth results.

4. Additional qualitative results

We show additional results in Fig. 2, the key point based approach provides more accurate 3D estimation than the direct approach due to the use of geometric constraints, and inter-car relationship.

In particular, for the direct approach, most errors occur in depth prediction. As explained in the paper, the method predicts the global 3D property of depth purely based on object appearance in 2D, and it faces the difficulty of down-sampled image size and none-clear appearance etc. On the other hand, keypoint-based approach leverages the cue of absolute car model size and re-projection geometry, which well constraints the 3D position of the car. However, keypoint-based approach failed when no key points are detected within a certain mask, especially for cars of unusual appearance.

References

[1] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh. Convolutional pose machines. In *Proceedings of the IEEE Con-*

Surface name	Keypoints label
Front surface	0, 1, 2, 3, 4, 5, 6, 8, 49, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61
Left surface	7, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21
Rear surface	24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 62, 63, 64, 65
Right surface	36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 50

Table 2: We divided a car into four visible surfaces, and manually define the correspondence between keypoints and surfaces.

ference on Computer Vision and Pattern Recognition, pages 4724–4732, 2016.