# SIDE: Center-based Stereo 3D Detector with Structure-aware Instance Depth Estimation
## *Supplementary Material*

Xidong Peng[1], Xinge Zhu[2], Tai Wang[2], and Yuexin Ma[1]

[1]ShanghaiTech University
[2]The Chinese University of Hong Kong
{linmo1533, zhuxinge123, taiwang.me}@gmail.com, mayuexin@shanghaitech.edu.cn

## A. Qualitative Results

We show the qualitative results of some scenarios in the kITTI dataset in the Fig.1. We show the corresponding stereo box, 3D box and bird's eye view on the left and right images. Our method can detect objects accurately in most scene and the detected three-dimensional boxes are well aligned on both the front-view images and LiDAR point cloud. It also can be seen that our method also perform well for the detection of the distant or largely occluded objects because of the structure-aware instance depth estimation.

## B. Perspectives about Cars

The keypoints of perspective observed in different perspectives of the same car are different. As shown in Fig.2, when the camera captures the same car from different perspectives, only one corner is visible in the bounding box of the object. Therefore, not only the distance of the keypoint relative to the border but also the type of keypoint needs to be predicted.

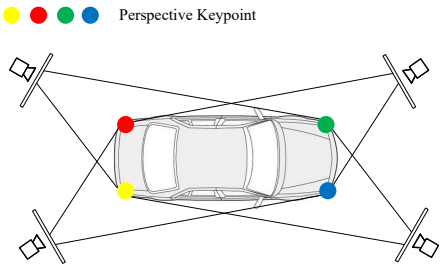


Figure 2: The same car in different perspectives

## C. Object Orientation and Viewpoint

Although we want to obtain the prediction of objects' steering angle $\theta$, we regress the $\alpha$ angle of each object instead of directly regressing the steering angle $\theta$. As shown in Fig.3, $\theta$ represents the angle between the car's direction and the x-axis, and $\alpha$ represents the angle between the car's direction and the x-axis when the car rotates around the y-axis to the front of the camera. The two angles above can be connected by $\beta$ in Eq.1 and $\beta$ can be expressed as the tangent about x and z of the car's center.

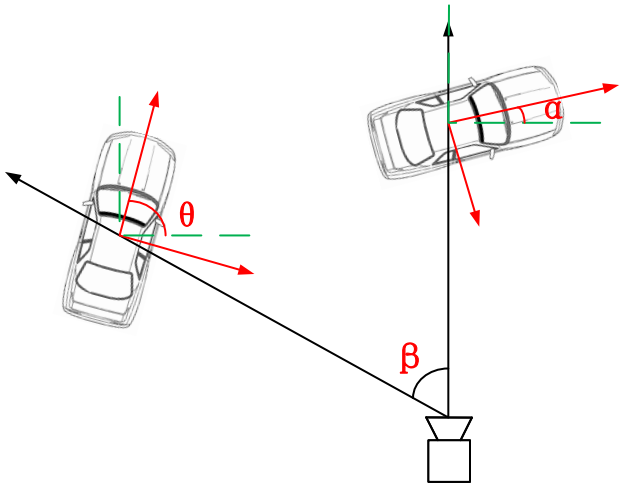


Figure 3: Relation between object orientation $\theta$ and the viewpoint angle $\alpha$.

$$\theta + \frac{\pi}{2} - \beta = \alpha + \frac{\pi}{2} \Rightarrow \theta = \alpha + \arctan(\frac{x}{z}) \quad (1)$$

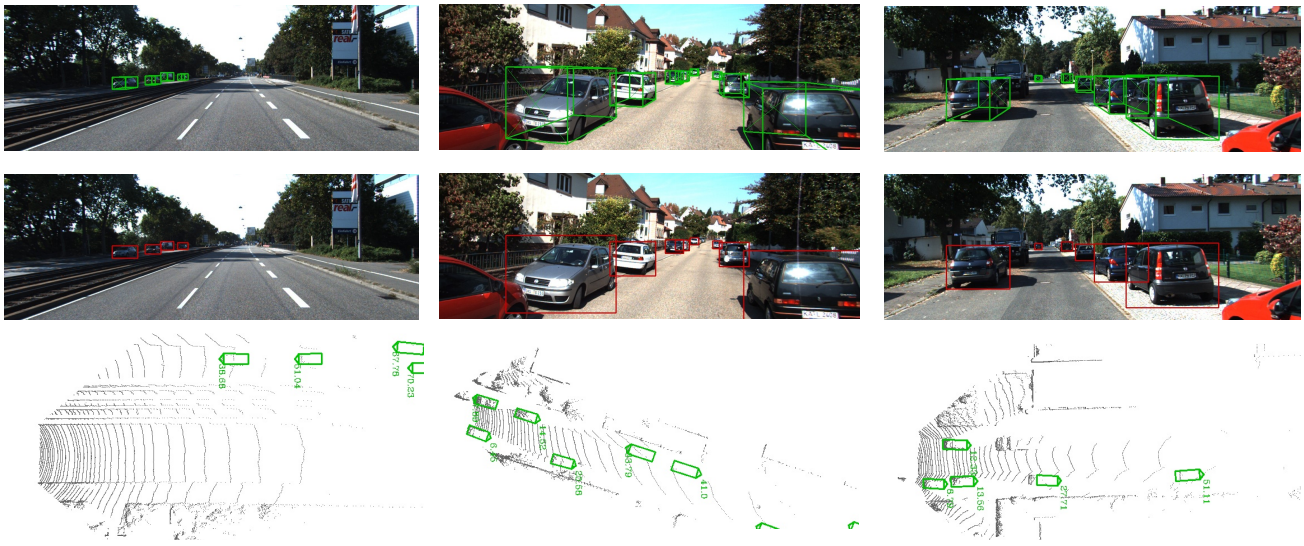

Figure 1: Quantitative results of multiple scenarios in the KITTI data set. From top to bottom: 3D detection on left image, 2D detection on left image, and bird's eye view image. Our method can perform well for the detection of the distant or largely occluded objects.