RAD: Realtime and Accurate 3D Object Detection on Embedded Systems Supplementary Material

Hamed H. Aghdam[1], Elnaz J. Heravi[2], Selameab S. Demilew[1], Robert Laganiere[1,2]
[1] University of Ottawa, [2] Sensorcortek
Ottawa, Canada
h.aghdam@uottawa.ca, elena@sensorcortek.ai, sdemi032@uottawa.ca, laganiere@uottawa.ca

1. Error Analysis

1.1. Orientation

Figure 1 illustrates the histogram of θ differences between predicted and ground-truth boxes where the IoU is between $[0.65, 0.7)$. As it turn out, θ differences are mainly distributed around zero and slightly around π and $-\pi$. Samples close to π and $-\pi$ in this graph are the ones whose ground-truth direction are the inverse of predicted boxes. In general, the orientation prediction is accurate since a majority of θ differences is distributed around zero.

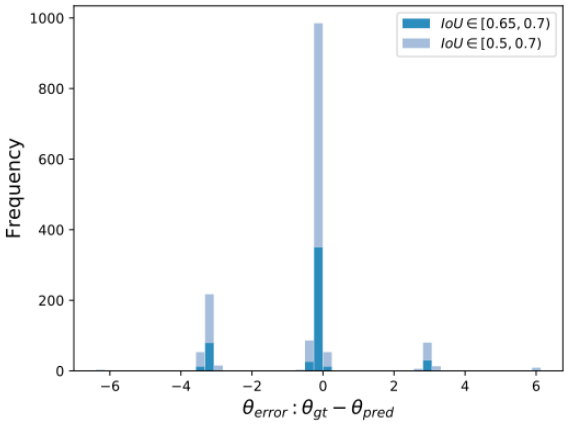

Figure 1. Histogram of yaw differences between predicted and ground-truth boxes where the IoU is between $[0.65, 0.7)$.

1.2. Density

The density of point cloud in the KITTI dataset is a function of distance meaning that distant regions have lower density compared to regions close to the LiDAR sensor. Likewise, partially occluded objects will also form sparse regions in the point cloud. Consequently, the regression head will have to make predictions using only small number of points. We computed the density of points clouds for objects where the IoU between the predicted box and the ground-truth box is between $[0.65, 0.7)$ and calculated the histogram of these objects based on their number of points. We repeated the above procedure for predicted boxes whose IoU with ground-truth boxes are in $[0.5, 0.65)$. Figure 2 shows the histograms. According to this figure, great majority of misalignment are due to sparsity of objects. For instance, there are about 100 samples whose IoUs are in $[0.5, 0.7)$ where there are only 1 to 5 points inside each box (*first bin, light blue bar*). Using LiDAR sensors with higher density at distant regions will potentially increase the accuracy of 3D object detector without affecting its time-to-completion.

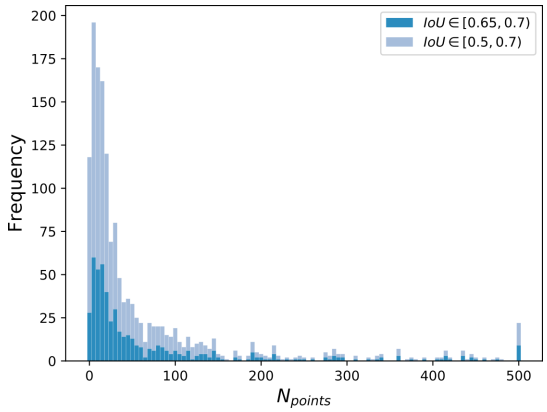

Figure 2. Stacked histogram of point cloud density for boxes whose IoU is in $[0.65, 0.7)$ or $[0.5, 0.7)$.

1.3. Regression Heads

Table 1 presents a more complete version of Table 9 in the paper. Evidently, improving the generalization of regression heads for l and z values will improve the results notably.

Replaced quantity	2D (car)			BEV (car)			3D (car)		
	easy	moderate	hard	easy	moderate	hard	easy	moderate	hard
original	96.24	91.94	89.34	94.86	88.21	85.70	88.70	75.80	72.65
x	96.27	92.25	91.32	95.57	90.76	88.40	90.77	81.26	76.70
y	96.38	94.42	92.06	94.89	88.25	85.74	91.27	82.20	79.62
z	96.16	91.93	89.35	95.40	90.46	88.43	90.10	79.09	76.50
w	96.28	92.03	89.42	95.09	88.40	85.91	89.71	77.97	73.51
h	98.15	91.85	90.83	96.95	90.36	87.83	87.94	76.78	72.45
l	96.24	91.91	89.37	95.04	88.65	87.86	90.08	78.58	74.06
xl	96.30	92.28	91.58	95.77	91.39	90.75	91.81	84.02	81.67
xh	98.34	94.11	91.49	97.75	92.87	90.49	90.22	81.07	78.07
xy	96.39	94.61	92.21	95.54	90.84	88.43	94.34	87.31	84.68
xz	96.18	92.26	91.47	96.13	91.90	91.39	94.13	85.26	82.56
zh	98.23	91.85	90.87	97.66	91.09	90.39	90.86	80.16	77.52
zl	96.26	91.96	89.43	96.21	92.23	91.64	95.05	88.10	85.15
zy	98.28	94.49	92.09	95.41	90.56	88.44	93.74	85.00	84.23
xzh	98.32	94.14	91.56	98.48	94.17	93.45	93.63	86.18	82.14
xzy	98.37	94.69	94.01	96.15	93.44	91.41	95.36	90.67	88.27
xzl	96.33	92.39	91.79	96.31	94.44	93.82	95.77	91.58	89.02

Table 1. Contribution of each regression output to the average precision reduction.

Method	2D (car)			BEV (car)			3D (car)		
	easy	moderate	hard	easy	moderate	hard	easy	moderate	hard
sequential augmentation	96.24	91.94	89.34	94.86	88.21	85.70	88.70	75.80	72.65
random augmentation	96.27	91.78	89.17	95.17	88.02	85.51	90.19	75.88	72.79

Table 2. The effect of augmentation on the performance.

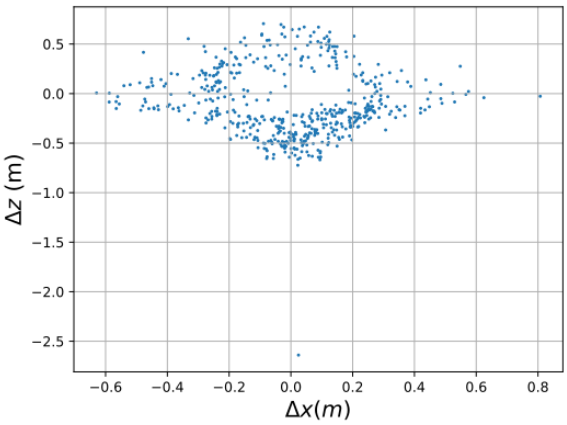

Figure 3. Correlation prediction errors of x and z.

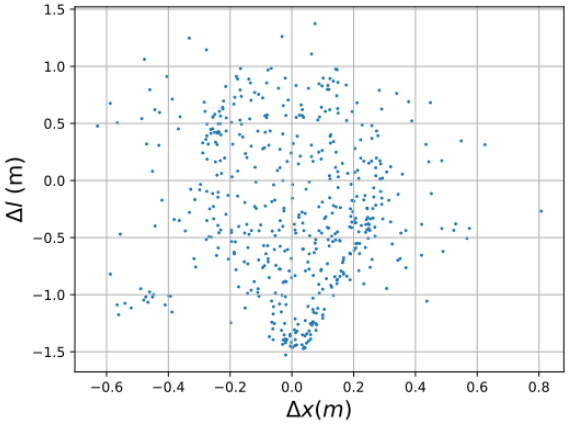

Figure 4. Correlation prediction errors of x and l.

1.4. Correlation

Figure 3 shows the correlation between prediction errors of x and z. One interesting observation is that as the magnitude of the prediction error for x increases, the magnitude of the prediction error for z reduces. In other words, there is a linear correlation between the magnitude (*i.e.* absolute value) of these two quantities.

Similarly, we computed the correlation between prediction error of x and l as well as z and l. According to the results in Figure 4 and Figure 5, there is a weak correlation between prediction errors of z and l and there is not a correlation between prediction errors of x and l.

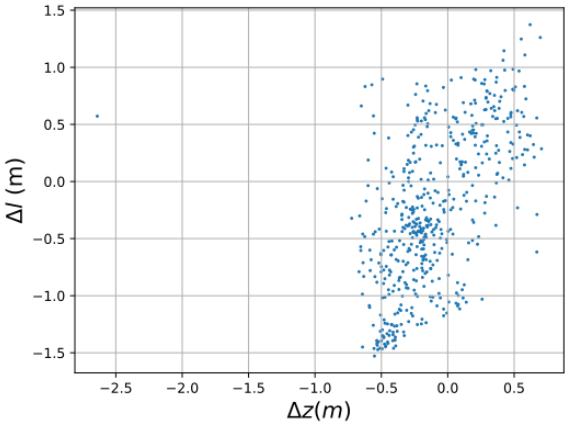

Figure 5. Correlation prediction errors of z and l.

2. Augmentation

Previously, we showed how different augmentation methods will increase the accuracy of the network. Augmentation methods can be applied in two different modes. Whereas the first mode applies augmentation technique sequentially with a predefined order, the second mode applies augmentation methods randomly.

Specifically, the sequential mode starts with applying random sampling followed by per-box rotation/translation, random flipping, global rotation, scaling and global translation. The random mode starts with applying random sampling followed by picking 3 of remaining methods randomly with a random order and applying them on the data. Table 2 compares these two methods.

The results show the accuracy of the network on all metrics are very similar except for the easy class in 3D metrics where random augmentation produces more accurate results. In addition, the random mode is computationally faster to execute and it will potentially increase the diversity of the dataset more than the sequential mode.