Utilizing color data in VoteNet for 3D Object Detection in Point Clouds

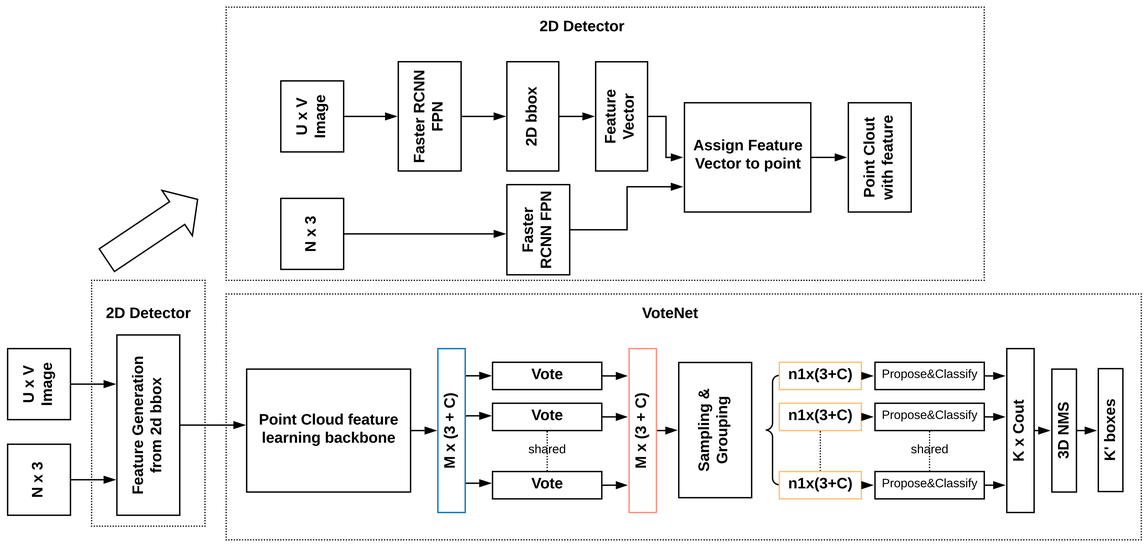
{ Xue Dong }

Stanford University

Introduction

RGB-D object detection has become a popular research topic due to the demand from application such as automated car and LIDAR data processing. Recent progress of deep learning techniques has greatly improved model precision and processing time. In this project, we study one of the state-of-the-art 3D object detection framework – VoteNet, and try to improve it using RGB information.

Method



- The baseline VoteNet framework is shown at the bottom of the diagram. It only uses (x, y, z) location of point cloud.
- We try to use 2D object detection as a hint to 3D object detection.
- We use 2D faster RCNN FPN to detect 2D bounding boxes $[xmin, ymin, xmax, ymax]$ and project this bounding box back to 3D space to get a frustum. If a point from point cloud $\{p_i\}_{i=1}^{N}$ is inside the frustum, then we assign the class score $b_{i,j}$ to it. We combine the score feature with it's 3D coordinate using

$$g_i = p_i, b_{i,j} \tag{1}$$

- We use SUN RGBD dataset for training and evaluation. The dataset contains 10335 scenes.

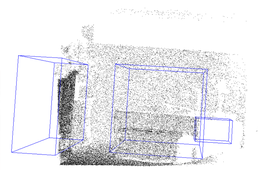
Precision and Recall

	bathtub	bed	bookshelf	chair	desk	dresser	nightstand	sofa	mAP
VoteNet	0.74	0.83	0.31	0.74	0.25	0.25	0.62	0.65	0.58
VoteNet use_color	0.69	0.83	0.33	0.74	0.25	0.27	0.59	0.65	0.57
Frustum PointNet	0.43	0.81	0.33	0.64	0.24	0.32	0.58	0.61	0.54
VoteNet+2D GT	0.84	0.95	0.71	0.82	0.63	0.82	0.87	0.89	0.82
VoteNet+2D	0.34	0.74	0.28	0.62	0.15	0.21	0.41	0.46	0.43

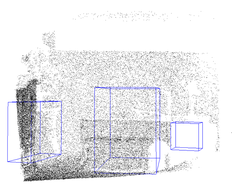
	bathtub	bed	bookshelf	chair	desk	dresser	nightstand	sofa	AR
VoteNet	0.90	0.95	0.67	0.86	0.78	0.80	0.88	0.90	0.86
VoteNet use_color	0.84	0.94	0.67	0.85	0.79	0.76	0.88	0.90	0.84
VoteNet+2D GT	0.86	0.96	0.82	0.88	0.81	0.89	0.94	0.93	0.89
VoteNet+2D	0.48	0.83	0.58	0.74	0.62	0.57	0.64	0.71	0.68

Object Detection Result

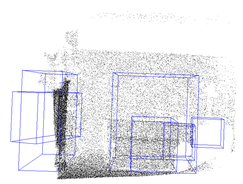
2D Ground Truth



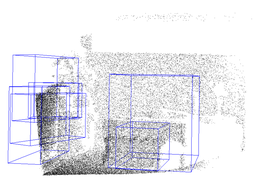
3D Ground Truth



VoteNet



VoteNet + 2D GT



VoteNet + 2D

References

[1] Charles R Qi, Or Litany, Kaiming He, and Leonidas J Guibas. Deep hough voting for 3d object detection in point clouds. 2019.

[2] Charles R. Qi, Wei Liu, Chenxia Wu, Hao Su, and Leonidas J. Guibas. Frustum pointnets for 3d object detection from rgb-d data. Jun 2018.

Conclusion and Future Research

Before we use a real 2D detector, we tried a proof of concept method using ground truth 2D bbox. Results are shown as VoteNet + 2D GT in the tables. Precision rate is improved by a lot.
Then we use a faster RCNN FPN to generate 2D bounding box on evaluation set. Results are shown as VoteNet + 2D. The precision and recall rate have dropped a lot.
From the point cloud image we can see that even in VoteNet + 2D GT there are a lot of false positive results. This could be a sign of over fitting.
To improve this, we can split training set into two parts, one for 2D detector training and one for 3D detector training. This will be our future work.

https://youtu.be/CBaVkr4Cv1o