

# 3D Object Detection in Point Clouds

Yen-Yu Chang   Tzu-Sheng Kuo   Shao-Yuan Ho

yenyu, tskuo, yuan1995@stanford.edu

## Highlights

- **Reproduction** of VoteNet [1], the state-of-the-art neural network for 3D object detection in point clouds.
- **Error Analysis** on our reproduced results and report quantitative and qualitative findings.
- **Improvement Strategies** for better performance:
  - **Hyperparameter Tuning**: we tune the weights for the loss and learning rates
  - **Input Features**: we include RGB features as input
  - **Backbone Module**: we augment the backbone [2] with VoxelEncoder [3]

## Introduction

- 3D object detection aims at detecting and classifying objects in 3D scenes.
- The task usually takes RGB-D data such as point clouds as input and performs object detection.

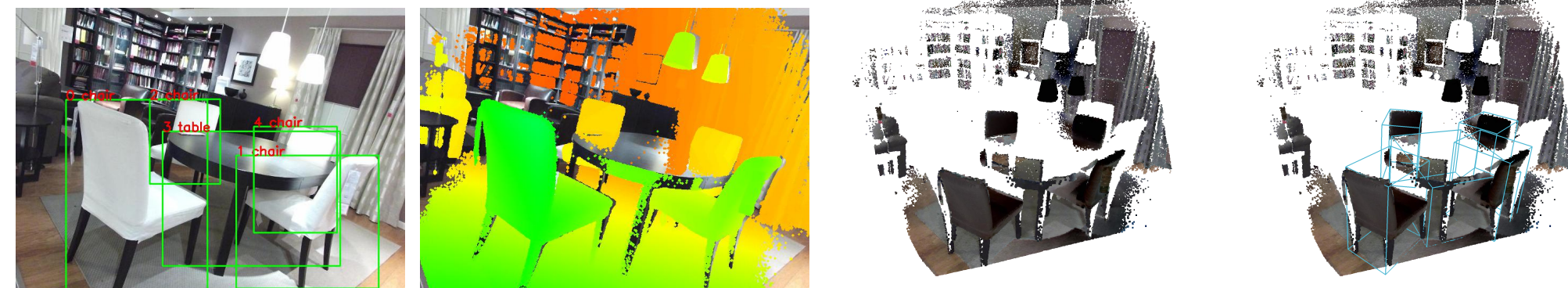


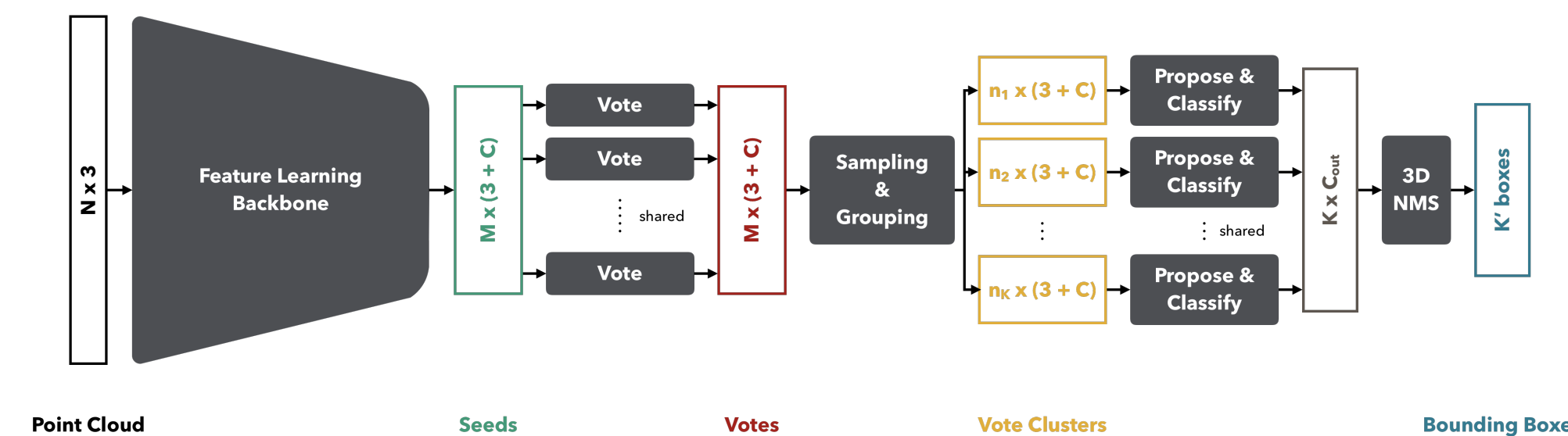
Figure 1: Example SUN RGB-D dataset [4] and 3D detection output.

## Related Work

- **Non Deep Learning Methods** compare RGB-D data with 3D shapes in databases using, for example, sliding windows.
- **2D Object Detectors** such as R-CNN are extended into 3D versions and applied to 3D voxels.
- **Projection-based methods** such as MV3D project 3D information onto multiple 2D planes and apply 2D object detectors.

These approaches sacrifice the valuable geometric information and sparsity in 3D point clouds.

## VoteNet



$$L = L_{vote-reg} + aL_{obj-cls} + bL_{box} + cL_{sem-cls},$$

Figure 2: The architecture and the loss function of VoteNet.

## Reproduction Results

Table 1: Evaluation metric is AP with 3D IoU threshold = 0.25.

model	bathtub	bed	bookshelf	chair	desk	dresser	nightstand	sofa	table	toilet	mAP
paper	74.4	83.0	28.8	75.3	22.0	29.8	62.2	64.0	47.3	90.1	57.7
ours 1	77.8	82.7	28.1	74.2	24.2	24.4	61.7	62.2	49.0	89.4	57.4
ours 2	77.1	83.1	28.4	74.7	25.1	25.2	61.7	61.6	49.4	87.7	57.4
ours 3	76.8	83.7	27.9	74.3	24.3	26.4	60.9	63.4	48.6	88.6	57.5

## Example Visualization



Figure 3: Reproduced qualitative results.

## Error Analysis

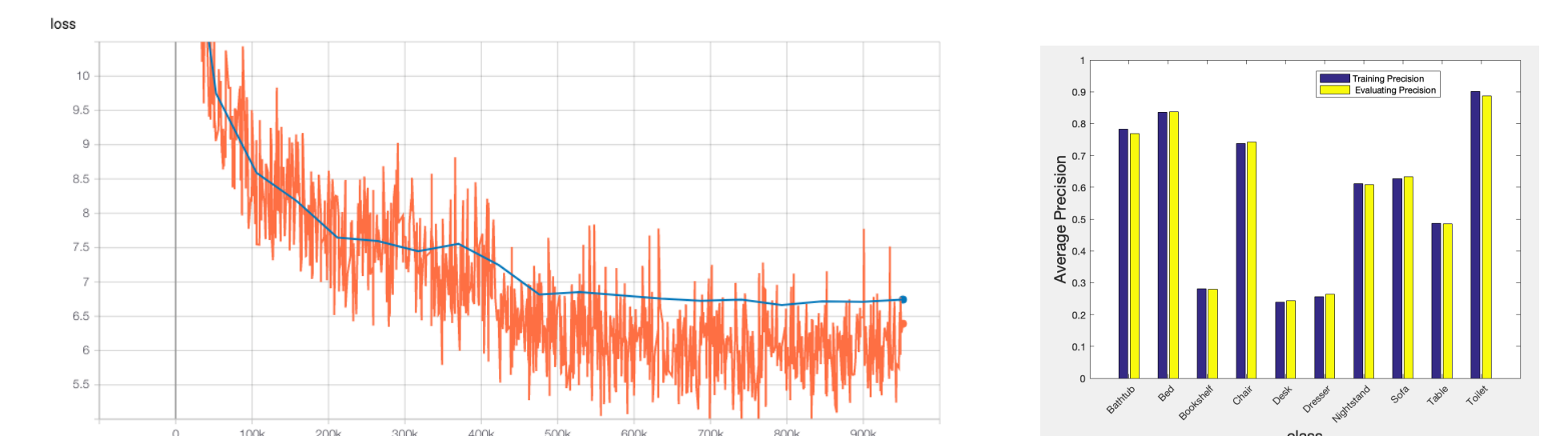


Figure 4: Visualization of training loss and AP across categories.

## Improvement Strategies

- **Tune hyperparameters** such as learning rate and weights of loss.
- **Include RGB feature** as input and keep its along with XYZ.
- **Modify backbone** by adding Voxel Feature Encoding layers [3].

## Improvement Results

Table 2: Results for improvements.

model	bathtub	bed	bookshelf	chair	desk	dresser	nightstand	sofa	table	toilet	mAP
paper	74.4	83.0	28.8	75.3	22.0	29.8	62.2	64.0	47.3	90.1	57.7
$\alpha_{100,130,160}$	76.4	82.5	30.2	74.5	23.9	27.2	59.2	62.9	48.3	88.8	57.4
$\alpha_{80,110,140}$	77.3	83.8	28.5	74.5	23.3	26.1	63.1	62.3	49.5	90.6	57.9
$a = 0.8$	79.5	83.3	28.5	74.3	23.7	26.9	60.0	63.3	49.4	89.0	57.8
$b = 1.3$	74.6	82.0	26.3	74.7	23.3	26.5	58.4	60.9	49.8	88.3	56.5
$c = 0.5$	71.6	83.3	26.1	74.4	22.4	26.7	61.1	58.9	46.5	89.3	56.0
RGB	72.4	84.3	27.0	73.4	24.6	24.2	58.3	62.7	49.7	87.7	56.4
VFE	4.75	23.2	2.95	36.9	2.00	0.68	79.1	14.2	11.6	47.8	14.6

## Selected References

- [1] C. R. Qi, O. Litany, K. He, and L. J. Guibas, "Deep hough voting for 3d object detection in point clouds," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019.
- [2] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," in *Advances in Neural Information Processing Systems 30* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), pp. 5099--5108, Curran Associates, Inc., 2017.
- [3] Y. Zhou and O. Tuzel, "Voxelnet: End-to-end learning for point cloud based 3d object detection," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4490--4499, June 2018.
- [4] S. Song, S. P. Lichtenberg, and J. Xiao, "Sun rgb-d: A rgb-d scene understanding benchmark suite," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 567--576, June 2015.
- [5] Y. Zhou and O. Tuzel, "Voxelnet: End-to-end learning for point cloud based 3d object detection," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4490--4499, June 2018.