Not Every Side Is Equal: Localization Uncertainty Estimation for Semi-Supervised 3D Object Detection

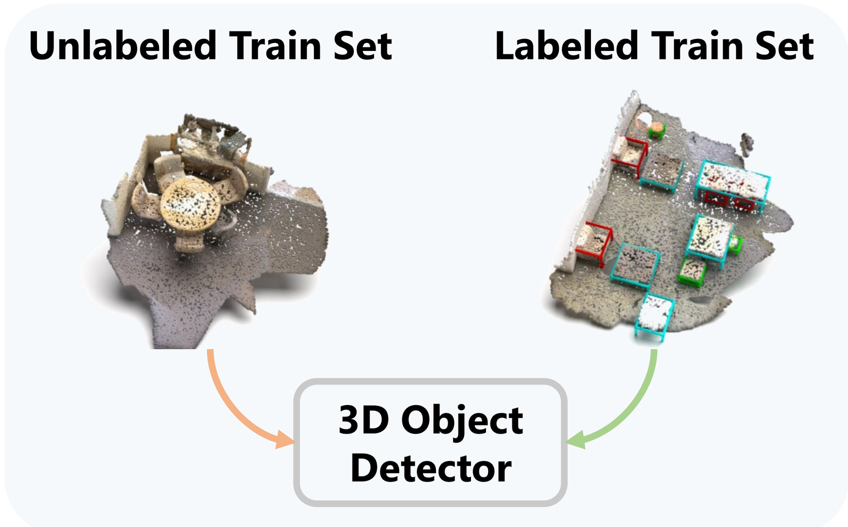
Chuxin Wang[1] Wenfei Yang[1] Tianzhu Zhang[1,2]

[1]University of Science and Technology of China [2]Deep Space Exploration Lab

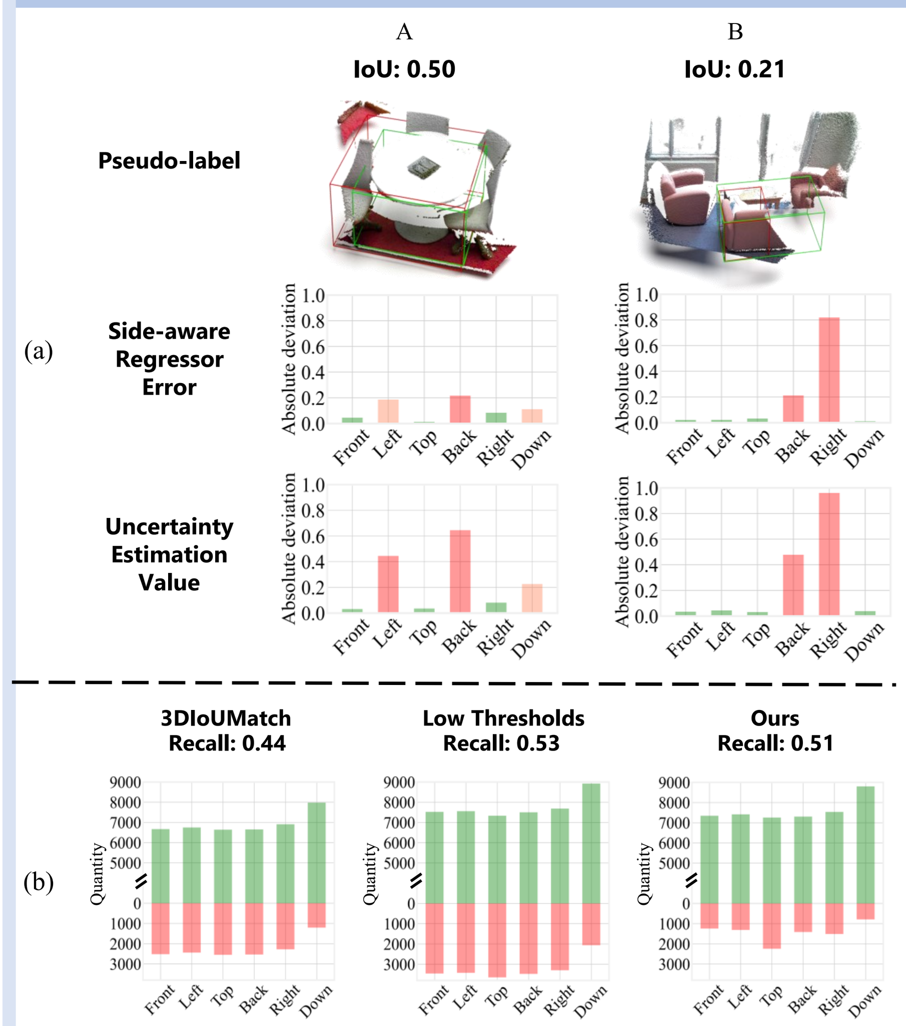
Semi-Supervised 3D Object Detection

3D object detection: Estimate oriented 3D boundary boxes as well as category labels of objects from a point cloud.

Semi-supervised learning: Train a model with a small number of labeled data and a large number of unlabeled data.



Challenges and Contributions



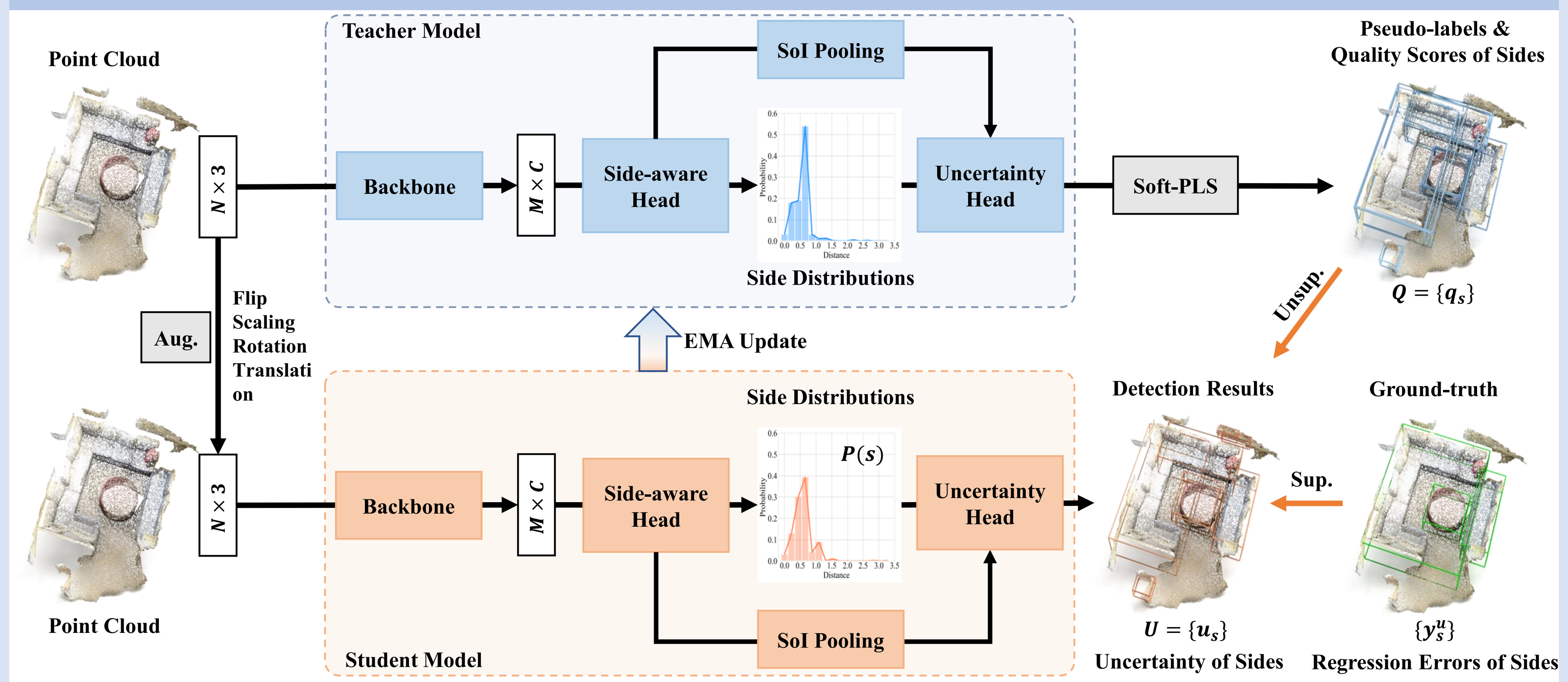
(a) How to estimate the localization quality of each side? *Global scores like IoU are insufficient for pseudo-label selection. Red represents pseudo-labels and green represents ground-truth.*

We decouple the localization task and predict the position of each side as a probability distribution. Then we use a side-aware uncertainty estimation module to evaluate the quality of each side.

(b) How to select pseudo-labels for model training? *The green bars indicate the number of sides with error less than 0.1, while the red bars indicate the number of sides with error greater than 0.1. The recall indicates the proportion of ground-truth that are covered by the pseudo-labels.*

We propose the Soft Pseudo-Label Selection which can suppress the interference of the low-quality sides while fully utilizing the sides with higher quality in the pseudo labels.

Our Framework



Our Approach

(a) Side-aware 3D Object Detection

Side-aware Bounding Box Parameterization: For convenience, we denote the top, down, left, right, front and back sides of a bounding box as:

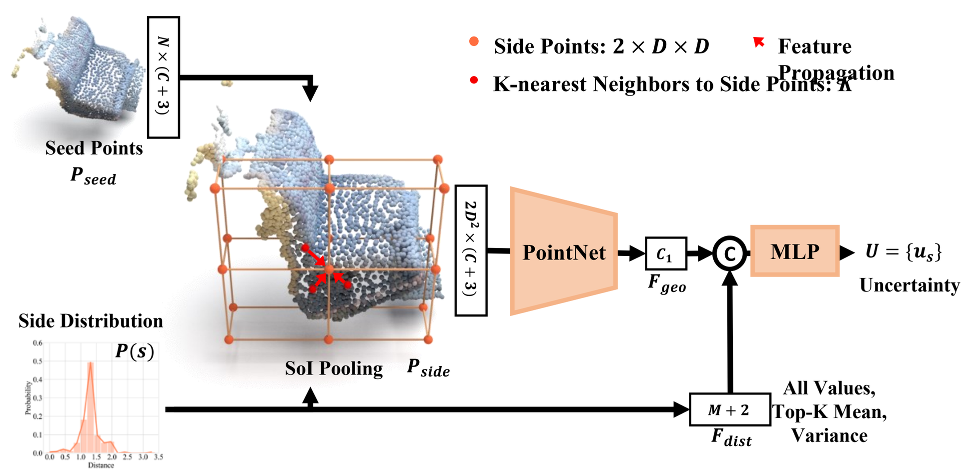
$$s \in B = \{t, d, l, r, f, b\}$$

The model predicts the distribution $P(s)$ defined over the interval $[s_{min}, s_{max}]$. Then we discretize the continuous distribution by dividing the interval into N bins $(s_0, s_1, \cdots, s_{N-1})$. The predicted value $\hat{s}$ of s can be calculated as:

$$\hat{s} = \sum_{i=0}^{N-1} P(s = s_i) s_i$$

Uncertainty Estimation Module (UEM):

We design side of interest (SOI) pooling to obtain the geometric features F_{geo}, then we concatenate the corresponding statistics to obtain the distribution properties F_{dist}. Finally, we fuse the geometric features and distribution properties into a MLP to obtain the uncertainty measure $U = \{u_s \mid s \in B\}$.



Uncertainty Regression Loss:

To guide the training of the UEM, we introduce the uncertainty regression loss. We use the absolute deviation of the predicted side location $\hat{s}$ and ground-truth y_s to compute the uncertainty label $y_s^u = \mathrm{MIN}(\alpha|y_s - \hat{s}|, 1.0)$. The uncertainty regression loss can be computed as the mean square error between y_s^u and u_s.

(b) Soft Pseudo-Label Selection

The Soft Pseudo-Label Selection (soft-PLS) consists of three components: Category specific filter, IoU-guided NMS with low-half strategy, side-aware weight assignment.

Category specific filter: Following FlexMatch, we use category-specific thresholds to filter pseudo-labels.

IoU-guided NMS with low-half strategy: To suppress noise in pseudo-labels caused by duplicated bounding box predictions, we utilize the IoU-guided non-maximal suppression with low-half keeping strategy to eliminate redundant pseudo-labels.

Side-aware weight assignment: We first evaluate the localization quality of each side and compute the quality score $q_s = e^{-\beta u_s}$ of each side. β is a scaling value. When supervising the student model, we use the quality scores $Q = \{q_s | s \in B\}$ of the pseudo-label to weight the loss function as follows:

$$L_{box} = q_B L_{iou}(B) + \sum_{s \in B} (q_s L_{reg}(s))$$

In this way, we reduce the interference of poorly localized sides in model training.

Experiments

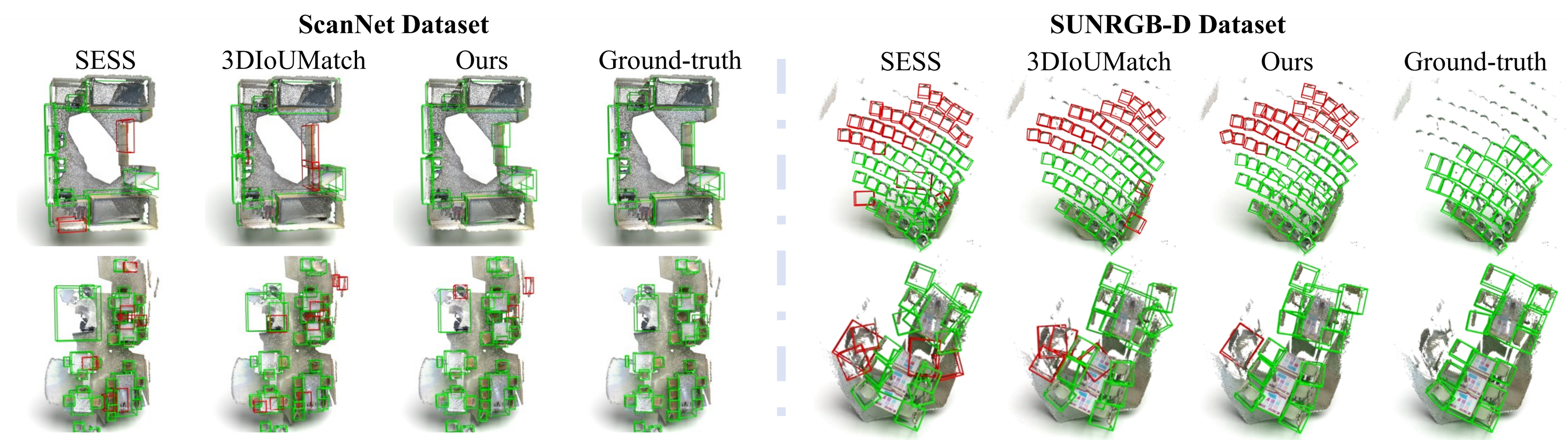
On Indoor Benchmarks

Table 1. **Results on ScanNet Val dataset under different ratios of labeled data.** Results are reported as mean ± standard deviation across 3 runs with random data splits, * represents the re-implemented results on ScanNet 50% labeled data.

Model	5%		10%		20%		50%*		100%	
	mAP_{25}	mAP_{50}	mAP_{25}	mAP_{50}	mAP_{25}	mAP_{50}	mAP_{25}	mAP_{50}	mAP_{25}	mAP_{50}
VoteNet [19]	27.9±0.5	10.8±0.6	36.9±1.6	18.2±1.0	46.9±1.9	27.5±1.2	56.1±1.1	36.5±0.6	57.8	36.0
SESS [37]	32.0±0.7	14.4±0.7	39.5±1.8	19.8±1.3	49.6±1.1	29.0±1.0	57.2±1.2	37.7±0.7	61.3	39.0
3DIoU [30]	40.0±0.9	22.5±0.5	47.2±0.4	28.3±1.5	52.8±1.2	35.2±1.1	59.8±0.7	41.2±0.5	62.9	42.1
Ours	**40.5±1.1**	**23.8±0.8**	**48.8±0.9**	**31.1±1.1**	**54.5±0.8**	**37.3±0.5**	**61.5±1.4**	**43.1±0.8**	**63.8**	**44.1**

Table 2. **Results on SUNRGB-D Val dataset under different ratios of labeled data.** Results are reported as mean ± standard deviation across 3 runs with random data splits, * represents the re-implemented results on SUNRGB-D 50% labeled data.

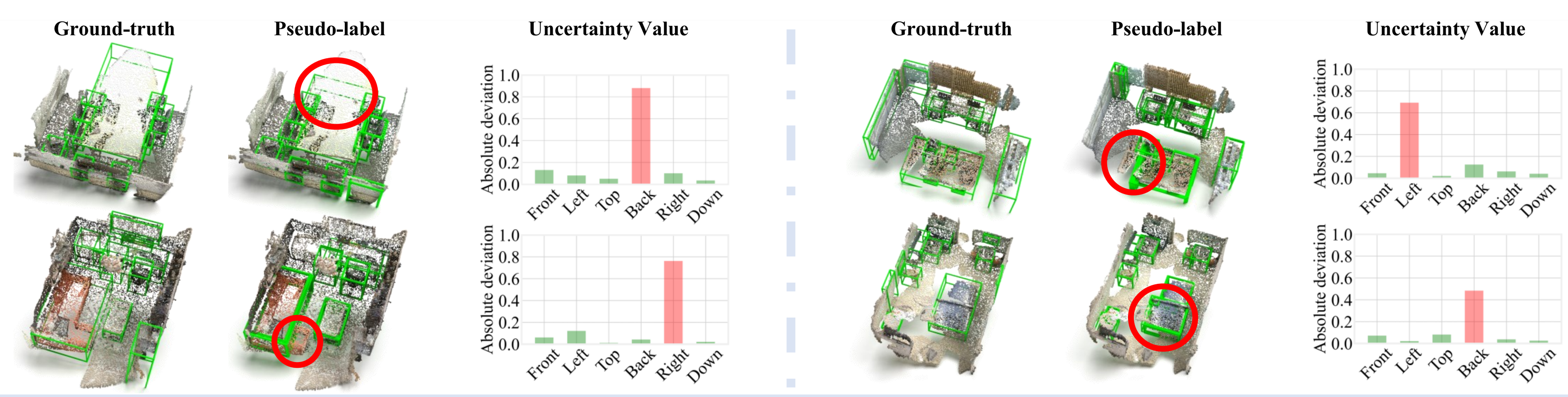
Model	5%		10%		20%		50%*		100%	
	mAP_{25}	mAP_{50}	mAP_{25}	mAP_{50}	mAP_{25}	mAP_{50}	mAP_{25}	mAP_{50}	mAP_{25}	mAP_{50}
VoteNet [19]	29.9±1.5	10.5±0.5	38.9±0.8	17.2±1.3	45.7±0.6	22.5±0.8	55.3±1.1	31.9±0.8	58.0	33.4
SESS [37]	34.2±2.0	13.1±1.0	42.1±1.1	20.9±0.3	47.1±0.7	24.5±1.2	56.2±0.8	33.7±0.7	60.5	38.1
3DIoU [30]	39.0±1.9	21.1±1.7	45.5±1.5	28.8±0.7	49.7±0.4	30.9±0.2	58.3±0.9	35.6±0.4	61.5	41.3
Ours	**41.1±1.2**	**21.8±1.8**	**47.4±0.8**	**29.2±1.2**	**53.4±0.9**	**31.2±1.3**	**60.1±0.4**	**37.8±0.8**	**62.7**	**42.1**



On Outdoor Benchmark

Table 3. **Results on KITTI Val set under different ratios of labeled data.** The results are reported as mean ± standard deviation across 3 runs with random data splits.

Model	mAP(1%)			mAP(10%)			mAP(20%)			mAP(100%)		
	Car	Ped.	Cyc.	Car	Ped.	Cyc.	Car	Ped.	Cyc.	Car	Ped.	Cyc.
PV-RCNN [22]	73.1±0.2	21.4±11.1	28.0±6.0	80.7±1.0	50.0±3.2	60.5±4.7	82.4±0.2	52.4±1.5	65.8±1.3	82.5	58.1	73.5
3DIoU [30]	75.2±1.8	32.9±16.1	31.4±7.8	81.3±0.8	52.6±1.9	62.0±5.8	82.9±0.1	54.5±1.4	67.4±1.7	84.2	60.5	75.2
Ours	**76.3±1.0**	**33.1±13.6**	**33.6±5.2**	**83.1±0.5**	**54.2±1.9**	**65.3±3.6**	**84.1±0.2**	**57.8±1.5**	**70.8±0.5**	**85.3**	**60.9**	**76.3**



Conclusion

- In this paper, we propose a side-aware framework with three specific designs: a probabilistic parameterization method, an uncertainty estimation module, and a soft pseudo-label selection.
- To the best of our knowledge, our approach is the first to consider the quality of local sides for pseudo-label filtering, enabling full exploitation and utilization of valid information in the model prediction results for supervising student models.
- Experiment results indicate that our method outperforms state-of-the-art methods on two indoor datasets and one outdoor dataset.