Supplementary Material for CaKDP

This document supplements our main submission “CaKDP: Category-aware Knowledge Distillation and Pruning Framework for Lightweight 3D Object Detection”. We first report the experiment details of our proposed framework in Section 1. In Section 2, we illustrate more technical details. Besides, We provide the statistical analysis of Figure 2 in Section 3. In Section 4, we demonstrate more results of CaKDP. Moreover, in Section 5, we conduct ablations to illustrate the influence of factor of KD loss (α). In Section 6 and Section 7, we demonstrate that the comparisons focus exclusively on distillation and pruning, respectively. Furthermore, in Section 8, we discuss the pipeline of our CaKDP framework. Finally, in Section 9, we further visualize the predictions before and after IOU-aware refinement.

1. Experiment Details

1.1. Dataset

KITTI Dataset: KITTI 3D object benchmark contains 7481 training samples and 7518 test samples. The training samples are further divided into a training set with 3712 samples and a validation set with 3769 samples. The dataset has three categories (Car, pedestrian and cyclist). Also, the dataset has three difficulty levels (easy, moderate, and hard) based on the object size, occlusion, and truncation levels.

To detect the objects of KITTI, we voxelize the input point cloud into a grid of resolutions $[0.05, 0.05, 0.1]$ meters in ranges $[0, 70.4]$, $[-40, 40]$, and $[-3, 1]$ meters along the X, Y, and Z axes, respectively. The maximum number of points in each voxel is set to 5. To demonstrate the detection ability of the detectors on KITTI dataset, we evaluate the results by moderate average precision @R40 (40 recall positions) for each category and also calculate the moderate mean average precision @R40 (moderate mAP@R40). Following the default settings in OpenPCDet [10], for cars we require an 3D bounding box overlap of 0.70, while for pedestrians and cyclists we require a 3D bounding box overlap of 0.50.

Waymo Open Dataset (WOD): WOD is a large-scale public autonomous driving dataset, which contains 1150 sequences in total, with 798 for training, and 202 for validation. It is collected by one long-range LiDAR sensor at 75 meters and four near-range sensors. We detect the categories of vehicle, pedestrian and cyclist in the WOD.

For WOD, point clouds are clipped into $[-75.2, 75.2]$ meters for X- or Y-axis, and $[-2, 4]$ meters for Z-axis. Voxel size is $[0.1, 0.1, 0.15]$ meters by default. The maximum number of points in each voxel is set to 5. Following [10], we evaluate the results by LEVEL 1/LEVEL 2 AP (L1 AP/L2 AP) and LEVEL 1/LEVEL 2 APH (L1 APH/L2 APH) of each category and also calculate LEVEL 2 mAP (L2 mAP) and LEVEL 2 mAPH (L2 mAPH).

For both KITTI dataset and WOD, following previous works on pruning and KD [4, 6, 14], we leverage the number of parameters and FLOPs evaluate the efficiency of the detectors.

1.2. Implementation Details

Implementation Details of KITTI Dataset: For all the experiments on KITTI dataset, the factor of IOU loss (β) is set to 1.0. In training phase, same as the default configurations in OpenPCDet [10] and SpareseKD [12], the batch size, number of epochs, weight decay, momentum are set to 4, 80, 0.01 and 0.9 respectively. The initial learning rate is set to 0.003, and it is multiplied by 0.1 at the 35-th and 45-th epochs. In inference phase, the threshold of IOU-aware refinement module is set to 0.1 (i.e., $\delta = 0.1$).

Additionally, when SECOND is student model, to select the representative samples from student detectors, we set the threshold of NMS to 0.7, and samples with the confidence (category prediction) less than 0.25 are ignored in the NMS pipeline. For the combinations of “SECOND & Voxel-RCNN” and “SECOND & PV-RCNN”, when retaining ratios are equal to 1.0, 0.75 and 0.5, the factor of KD loss (α) is set to 1.0; and when retaining ratio is equal to 0.30, the factor of KD loss is set to 0.1. Moreover, for the combination of “SECOND & PartA2”, when retaining ratios are equal to 1.0 and 0.75, the factor of KD loss is set to 1.0; and when retaining ratios are equal to 0.50 and 0.30, the factor of KD loss is set to 0.1.

Furthermore, when training the higher performance CenterPoint, we set the threshold and confidence (category prediction) in NMS pipeline to 0.7 and 0.1, respectively. For the combination of “CenterPoint & Voxel-RCNN”, when retaining ratios are equal to 1.0, 0.75 and 0.5, the factor of KD loss is set to 2.0; and when retaining ratio is equal to 0.35, the factor of KD loss is set to 1.0. Besides, for “Cen-

Type	Model	Along y-axis		Along x-axis				
		above red	between red and blue	(0, 0.2]	(0.2, 0.4]	(0.4, 0.6]	(0.6, 0.8]	(0.8, 1.0]
One-stage	SECOND [11]	44.10%	34.70%	1.19%	1.16%	11.51%	40.81%	40.67%
	CenterPoint [13]	15.46%	64.61%	0.82%	3.59%	12.20%	39.82%	41.18%
Two-stage	Voxel-RCNN [2]	79.10%	16.09%	0.76%	2.60%	10.25%	39.12%	44.44%
	PV-RCNN [7]	77.65%	16.64%	3.70%	0.66%	10.73%	37.87%	45.18%
	PartA2 [8]	81.36%	13.70%	0.92%	2.95%	10.80%	38.02%	44.48%

Table 1. Proportions in Figure 2. For the statistics along y-axis and x-axis, we exclude the predictions with IOU of 0, and the predictions with Cate-Pred of 0, respectively.

terPoint & PV-RCNN" and "CenterPoint & PartA2", when retaining ratios are equal to 1.0, 0.75 and 0.5; the factor of KD loss is set to 1.0, and when retaining ratio is equal to 0.35, the factor of KD loss is set to 0.1.

Implementation Details of WOD and WOD-mini: For all the experiments on WOD-mini and WOD, the factor of IOU loss (β) is also set to 1.0. Besides, we set the threshold of NMS to 0.7, and samples with the confidence (category prediction) less than 0.1 are ignored in the NMS pipeline. In training phase, we also keep the common configurations used in OpenPCDet [10] and SpareseKD [12], the batch size, number of epochs, initial learning rate, weight decay, momentum are set to 4, 30, 0.003, 0.01 and 0.9 respectively. In inference phase, the threshold of IOU-aware refinement module is set to 0.05 (i.e., $\delta = 0.05$).

Additionally, the factors of KD loss (α) are set to 0.5 and 1.0 for distillation in the combinations of "CenterPoint & PV-RCNN++" and "CenterPoint & Voxel-RCNN", respectively.

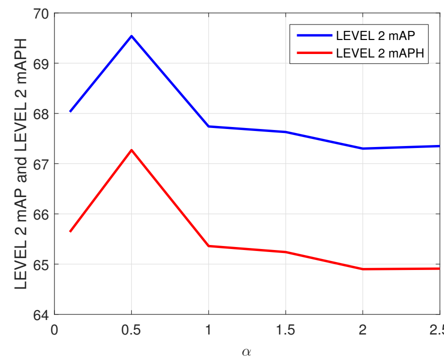


Figure 1. Results of CaKDP with different α.

2. More Technical Details

2.1. Selection Process for Representative Samples

In Section 3.2 of the main paper, we leverage non-maximum suppression (NMS) to select representative samples (RSs) from the student detector, and then the selected RSs are used to match the teacher's knowledge at corresponding location for distillation. We here demonstrate the detailed selection process for RSs:

(1) Firstly, all anchors are sorted from large to small based on their category predictions (confidence scores), and the top-N samples with higher confidence scores are selected to form the current processing bank. Here, the processing bank represents the set of samples that currently needs to be processed.

(2) Afterwards, we select the prediction with the highest confidence score in the current processing bank, and move it from the processing bank to the candidate bank, where the candidate bank represents the set of RSs.

(3) After that, we calculate the intersection over union (IOU) between the current sample with the highest confidence and other anchors in the processing bank. Subsequently, we remove samples from the processing bank whose IOU exceeds the IOU threshold.

(4) We repeat step (2) and step (3) until there are no anchors left in the processing bank. The final candidate bank comprises the set of RSs.

2.2. Details of IOU-aware Refinement Module

In Section 3.4, we propose the modified IOU-aware refinement module to remove the redundant false positive (FP) samples. We here demonstrate the architecture for predicting IOU and the ground truth label of IOU:

Architecture of IOU head: We set the IOU head to be the same as the classification head. In anchor-based detectors (e.g., SECOND), it entails a single-layer 1x1 convolution. In center-based detectors (e.g., CenterPoint), it involves a stack of one-layer 3x3 convolution (including, batch norm and ReLU) followed by one-layer 1x1 convolution.

Label of IOU: For the anchor-based detectors, the label is the IOU between the anchor and the ground truth bounding box, which ranges from 0 to 1. For center-based detectors, the label takes values of either 0 or 1, where the IOU label corresponding to the object center is set to 1, and the IOU label corresponding to other position is set to 0.

3. Statistically Analyze of Figure 2

In Figure 2 of the main paper, we illustrate the gap between heterogeneous detectors by demonstrating the distribution of predictions of different detectors. In this subsection, Table 1 statistically analyzes the proportion of predictions in different intervals. This table numerically explains that the

	Model	Retaining Ratio	KD Loss	Car			Pedestrian			Cyclist			Para.	FLOPs	Moderate mAP@R40
				Easy	Mod.	Hard	Easy	Mod.	Hard	Easy	Mod.	Hard			
Stu	SECOND [11]	1.00	✘	89.59	81.33	78.50	58.07	52.95	48.51	83.44	65.89	62.44	5.3	80.7	66.72
	CenterPoint [13]	1.00	✘	86.61	78.50	76.44	56.26	51.85	47.22	84.77	67.78	64.06	5.8	96.5	66.04
Tea	Voxel-RCNN [2]	1.00	✘	92.81	84.97	82.47	63.56	57.74	52.86	91.73	73.73	69.27	11.0	81.6	72.15
Stu	SECOND [11]	1.00	✔	92.64	83.27	80.61	68.41	60.88	55.36	94.29	75.13	69.17	5.3	81.0	73.09
		0.75	✔	91.85	83.09	80.27	69.71	62.91	57.31	91.67	73.48	68.65	3.3	54.2	73.16
		0.50	✔	91.79	82.72	80.01	68.02	60.44	54.78	90.58	72.15	67.60	1.5	30.2	71.77
		0.30	✔	89.23	79.18	74.97	60.23	53.78	49.15	82.79	66.61	62.48	0.6	17.7	66.52
	CenterPoint [13]	1.00	✔	90.05	82.85	80.22	67.71	60.84	55.29	92.21	73.73	69.20	5.8	97.9	72.48
		0.75	✔	90.22	82.72	80.18	65.36	59.59	53.15	90.15	73.51	69.03	3.5	67.3	71.94
		0.50	✔	89.62	80.40	77.64	67.72	60.65	55.21	88.53	72.18	67.72	1.8	39.5	71.07
		0.35	✔	85.93	73.98	68.90	66.18	58.65	52.22	84.05	66.73	62.73	1.1	31.3	66.45
Tea	PV-RCNN [7]	1.00	✘	91.44	84.25	82.06	65.69	57.67	52.40	90.20	72.33	67.76	13.1	93.1	71.42
Stu	SECOND [11]	1.00	✔	90.06	83.01	80.34	65.37	58.61	53.47	92.60	73.82	69.30	5.3	81.0	71.81
		0.75	✔	91.66	82.69	80.14	67.93	59.92	54.59	90.84	71.99	67.50	3.3	54.2	71.53
		0.50	✔	90.35	82.40	79.69	68.67	60.12	53.73	90.29	71.17	66.75	1.5	30.2	71.23
		0.30	✔	89.39	79.46	74.93	58.79	52.96	48.47	82.21	65.86	61.68	0.6	17.7	66.10
	CenterPoint [13]	1.00	✔	90.45	83.40	80.77	64.71	57.56	52.34	89.40	70.89	67.90	5.8	97.9	70.62
		0.75	✔	90.35	83.32	80.60	65.16	58.10	52.86	91.37	72.82	68.71	3.5	67.3	71.41
		0.50	✔	89.80	80.53	77.75	65.45	59.11	52.97	88.55	70.25	65.82	1.8	39.5	69.96
		0.35	✔	85.84	74.06	69.22	65.80	58.12	52.52	84.61	67.08	62.69	1.1	31.3	66.42
Tea	PartA2 [8]	1.00	✘	91.62	82.22	79.92	66.39	60.42	55.27	90.56	72.65	68.15	63.6	93.3	71.77
Stu	SECOND [11]	1.00	✔	90.05	82.66	80.29	67.25	60.08	54.86	91.54	73.51	68.81	5.3	81.0	72.08
		0.75	✔	90.31	82.85	80.20	68.02	60.51	54.40	92.03	73.86	69.13	3.3	54.2	72.41
		0.50	✔	90.13	82.71	79.81	64.53	57.62	51.55	90.48	72.10	67.63	1.5	30.2	70.81
		0.30	✔	88.74	77.18	74.41	58.82	53.38	48.92	82.23	66.5	62.09	0.6	17.7	65.69
	CenterPoint [13]	1.00	✔	90.23	82.99	80.39	64.90	58.12	52.70	90.42	72.07	67.41	5.8	97.9	71.06
		0.75	✔	90.22	83.05	80.47	63.00	58.12	52.01	92.28	73.29	68.77	3.5	67.3	71.49
		0.50	✔	89.87	80.14	77.50	66.02	59.28	53.87	89.73	72.24	67.74	1.8	39.5	70.55
		0.35	✔	85.55	74.85	70.35	63.71	57.73	53.31	83.26	65.15	61.44	1.1	31.3	65.91

Table 2. Results of CaKDP on KITTI Dataset. 'Tea' and 'Stu' represent teacher and student models, respectively. 'Mod.' represents moderate, and 'Para.' represents parameter.

	Model	Retaining Ratio	KD Loss	Vehicle		Pedestrian		Cyclist		Para.	FLOPs	L2 mAP/mAPH
				L1 AP/APH	L2 AP/APH	L1 AP/APH	L2 AP/APH	L1 AP/APH	L2 AP/APH			
Stu	CenterPoint [13]	1.00	✘	72.65 / 72.12	64.56 / 64.08	74.16 / 67.95	62.23 / 60.53	70.75 / 69.56	68.16 / 67.02	7.8	114.8	66.32 / 63.88
Tea	PV-RCNN++ [9]	1.00	✘	77.50 / 77.02	69.08 / 68.64	79.56 / 73.38	71.04 / 65.32	72.75 / 71.66	70.09 / 69.04	16.1	123.5	70.07 / 67.67
Stu	CenterPoint [13]	1.00	✔	74.72 / 74.24	66.32 / 65.88	77.97 / 72.25	69.43 / 64.14	73.25 / 72.15	70.56 / 69.50	7.8	116.2	68.77 / 66.51
		0.70	✔	74.31 / 73.81	65.98 / 65.52	77.83 / 72.02	69.38 / 64.00	73.02 / 71.89	70.37 / 69.28	4.7	79.1	68.58 / 66.27
		0.50	✔	73.28 / 72.77	64.94 / 64.47	76.90 / 70.93	68.41 / 62.90	72.87 / 71.71	70.21 / 69.09	2.8	55.6	67.85 / 65.48
		0.35	✔	70.87 / 70.32	62.49 / 61.99	74.69 / 68.39	66.02 / 60.28	70.42 / 69.19	67.82 / 66.64	1.8	39.0	65.44 / 62.97
Tea	Voxel-RCNN [2]	1.00	✔	76.88 / 76.44	68.54 / 68.13	79.31 / 73.59	70.74 / 65.42	72.49 / 71.45	69.83 / 68.83	18.7	117.6	69.70 / 67.46
Stu	CenterPoint [13]	1.00	✔	74.91 / 74.42	66.55 / 66.11	77.96 / 72.29	69.49 / 64.23	72.96 / 71.83	70.28 / 69.20	7.8	116.2	68.78 / 66.51
		0.70	✔	74.44 / 73.94	66.11 / 65.66	77.76 / 71.94	69.27 / 63.88	73.29 / 72.15	70.62 / 69.52	4.7	79.1	68.67 / 66.36
		0.50	✔	72.94 / 72.44	64.56 / 64.11	76.92 / 70.96	68.34 / 62.85	72.78 / 71.62	70.12 / 69.00	2.8	55.6	67.67 / 65.32
		0.35	✔	70.06 / 69.53	61.77 / 61.30	74.77 / 68.51	66.07 / 60.36	70.75 / 69.53	68.17 / 66.99	1.8	39.0	65.33 / 62.88

Table 3. Results on WOD-mini. 'L1' and 'L2' represent LEVEL 1 and LEVEL 2, respectively.

main discrepancy between heterogeneous detectors lies in the category predictions.

4. Results of CaKDP

In Table 1, Table 3 and Table 5 of the main submitted manuscript, we only report some of the key metric values (moderate AP @R40, moderate mAP @R40, L2 mAP and L2 mAPH) due to the page limitation policy. Hence, in this section, we demonstrate all the results of our proposed CaKDP in Table 2, Table 3 and Table 4.

5. Influence of Factor of KD Loss

In this section, we conduct experiments on the combination of "CenterPoint & PV-RCNN++" (WOD) to illustrate the influence of factor of KD loss (α) in Eq. (7) of main submitted manuscript. We set the retaining ratio to 0.5, and other configurations keep unchanged. The results of CaKDP with different KD loss factors (α) are shown in Fig. 1. When α = 0.5, the L2 mAP and L2 mAPH reach their peak values,

	Model	Retaining Ratio	KD Loss	Vehicle		Pedestrian		Cyclist		Para.	FLOPs	L2 mAP/mAPH
				L1 AP/APH	L2 AP/APH	L1 AP/APH	L2 AP/APH	L1 AP/APH	L2 AP/APH			
Stu	CenterPoint [13]	1.00	✘	74.21 / 73.67	66.25 / 65.76	76.26 / 70.16	68.50 / 62.86	72.09 / 70.96	69.47 / 68.38	7.8	114.8	68.07 / 65.66
Tea	PV-RCNN++ [9]	1.00	✘	78.64 / 78.20	70.32 / 69.91	81.33 / 75.83	73.04 / 67.87	73.76 / 72.69	71.06 / 70.03	16.1	123.5	71.47 / 69.27
Stu	CenterPoint [13]	1.00	✔	76.17 / 75.71	67.76 / 67.34	79.07 / 73.73	70.53 / 65.55	73.60 / 72.50	70.92 / 69.86	7.8	116.2	69.74 / 67.59
		0.70	✔	76.03 / 75.55	67.72 / 67.27	79.20 / 73.74	70.80 / 65.69	72.84 / 71.75	70.16 / 69.11	4.7	79.1	69.56 / 67.36
		0.50	✔	75.35 / 74.84	67.00 / 66.53	78.78 / 73.17	70.42 / 65.19	73.88 / 72.72	71.21 / 70.09	2.8	55.6	69.54 / 67.27
		0.35	✔	73.21 / 72.70	64.87 / 64.41	77.27 / 71.28	68.76 / 63.23	73.08 / 71.91	70.39 / 69.26	1.8	39.0	68.01 / 65.63

Table 4. Results on full Waymo Open Dataset. 'L1' and 'L2' represent LEVEL 1 and LEVEL 2, respectively.

"SECOND & Voxel-RCNN" (Para./ FLOPs: 5.3M/80.7G)	Method	-	Vanilla KD [3]	GID [1]	PD [15]	SparseKD [12]	CaKD
	mAP	66.72	68.62	68.63	67.20	67.14	**72.83**
"CenterPoint & PV-RCNN++" (Para./ FLOPs: 7.8M/114.8G)	Method	-	Vinalla KD [3]	GID [1]	PD [15]	SparseKD [12]	CaKD
	mAPH	63.88	64.81	64.86	64.43	65.15	**65.97**

Table 5. Exclusive comparison on distillation.

which are 67.67% and 65.32%, respectively.

6. Exclusive Comparison on Distillation

In this subsection, we compare our CaKD with other KD methods. Table 5 demonstrates the results of "SECOND & Voxel-RCNN" on the KITTI dataset, and those of "CenterPoint & PV-RCNN++" on the WOD-mini dataset. As shown, our CaKD achieves higher accuracy student detectors on both datasets. Particularly, our method significantly improves the performance of student models on KITTI dataset, where CaKD provides SECOND with mAP of 72.83%, while Vanilla KD, GID, PD and SparseKD achieve 68.62%, 68.63%, 67.20%, 67.14%, respectively. Therefore, our CaKD has ability to obtain the student detector with higher performance.

7. Exclusive Comparison on Pruning

In this subsection, we compare our CaPr with L1 pruning method [5], which leverages the L1 norm to evaluate the importance of each filter. We set different pruning ratios to compress SECOND on KITTI dataset while keeping the default training configurations unchanged for retraining. As shown in Table 6, our CaPr achieves higher mAP while reducing more parameters and FLOPs. Therefore, CaPr demonstrates its capability to produce lightweight student detectors with appropriate architecture and parameters.

8. Pipeline of CaKDP Framework

CaKD and CaPr are two important modules in the training phase of our proposed CaKDP framework. In this section, we empirical study the influence of the order of these two modules. We list four different pipelines as:

- #**Mode 1: (1) KD:** We first conduct CaKD to get the complete one-stage student detector; **(2) Pruning:** After that, we leverage CaPr to prune the distilled student detector; **(3) Fine-tuning without KD loss:** Then, we conduct fine-tuning (without CaKD loss) to restore the accuracy of the pruned detector. **(4) KD:** Finally, we further leverage CaKD to train the detector after fine-tuning, and get the compact student detector with higher performance.
- #**Mode 2: (1) KD:** We first conduct CaKD to get the complete one-stage student detector; **(2) Pruning:** After that, we leverage CaPr to prune the distilled student detector; **(3) KD:** Finally, we retrain the pruned detector by final loss (containing CaKD loss) to restore the accuracy of the pruned detector.
- #**Mode 3: (1) Pruning:** We first prune the pretrained one-stage student detector; **(2) Fine-tuning without KD loss:** After that, we conduct fine-tuning (without CaKD loss) to restore the accuracy of the pruned detector. **(3) KD:** Finally, CaKD is leveraged to improve the performance of compact student detector after fine-tuning.
- #**Mode 4 (Ours): (1) Pruning:** We first prune the pretrained one-stage student detector; **(2) KD:** After that, we retrain the pruned student detector by final loss (containing CaKD loss) to restore the detection ability of the pruned detector.

We take "SECOND & Voxel-RCNN" on KITTI dataset as example to illustrate the influence of different pipelines. The retaining ratio is set to 0.5, and other configurations remain unchanged. In Table 5, similar results are obtained by four different pipelines. However, #**Mode 4 (ours)** has the simplest and fastest training phase, while #**Mode 1** to #**Mode 3** need more epochs to train the compact student detector. Compared with #**Mode 4**, #**Mode 1** contains a fine-tuning step (without CaKD loss) and an additional KD step (by training with final loss); #**Mode 2** has an additional KD step (by training with final loss); and #**Mode 3** provides a fine-tuning step (without CaKD loss). Therefore, in all the experiments of our main submission, we utilize

Method	Retaining Ratio	Car			Pedestrian			Cyclist			Para.	FLOPs	Moderate mAP@R40
		Easy	Mod.	Hard	Easy	Mod.	Hard	Easy	Mod.	Hard			
SECOND [11]	1.00	89.59	81.33	78.50	58.07	52.95	48.51	83.44	65.89	62.44	5.3	80.7	66.72
L1 [5]	0.7	89.12	80.80	77.78	54.40	48.98	44.71	82.45	65.73	62.26	2.6	39.5	65.17
CaPr	0.5	90.28	81.23	78.05	55.54	50.24	45.73	83.95	67.23	63.89	1.5	30.1	**66.24**
L1 [5]	0.4	87.63	76.33	73.43	51.05	44.89	41.10	72.75	57.47	54.07	0.9	13.2	59.56
CaPr	0.3	89.02	78.23	75.29	49.97	46.65	43.41	79.05	63.56	60.13	0.6	17.6	**62.81**

Table 6. Exclusive comparison on pruning.

Model	Mode	Car			Pedestrian			Cyclist			Para.	FLOPs	Moderate mAP@R40
		Easy	Mod.	Hard	Easy	Mod.	Hard	Easy	Mod.	Hard			
SECOND-×0.5	#Mode 1	90.42	83.05	80.29	67.51	60.35	54.60	90.34	72.61	67.95	1.5	29.5	72.00
	#Mode 2	90.56	82.74	79.96	67.53	60.18	54.30	91.46	72.58	67.89	1.5	29.5	71.83
	#Mode 3	90.48	82.89	80.17	67.91	60.88	54.04	91.22	72.49	67.63	1.5	30.2	72.09
	#Mode 4 (ours)	91.79	82.72	80.01	68.02	60.44	54.78	90.58	72.15	67.60	1.5	30.2	71.77

Table 7. Comparison of Different Pipelines of CaKDP Framework.

#**Mode 4** as the pipeline of our proposed CaKDP framework.

9. Visualization

In this section, we demonstrate more examples to visualize the influence of modified IOU-aware refinement module. As shown in Fig. 2, predictions without IOU-aware refinement contain more false positive (FP) samples. Therefore, our modified IOU-aware refinement module in CaKDP framework has ability to remove redundant FP samples, and further helps the framework to improve the accuracy of the student detectors.

References

[1] Xing Dai, Zeren Jiang, Zhao Wu, Yiping Bao, Zhicheng Wang, Si Liu, and Erjin Zhou. General instance distillation for object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7842–7851, 2021. 4

[2] Jiajun Deng, Shaoshuai Shi, Peiwei Li, Wengang Zhou, Yanyong Zhang, and Houqiang Li. Voxel r-cnn: Towards high performance voxel-based 3d object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1201–1209, 2021. 2, 3

[3] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015. 4

[4] Zejiang Hou, Minghai Qin, Fei Sun, Xiaolong Ma, Kun Yuan, Yi Xu, Yen-Kuang Chen, Rong Jin, Yuan Xie, and Sun-Yuan Kung. Chex: Channel exploration for cnn model compression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12287–12298, 2022. 1

[5] Hao Li, Asim Kadav, Igor Durdanovic, Hanan Samet, and Hans Peter Graf. Pruning filters for efficient convnets. *arXiv preprint arXiv:1608.08710*, 2016. 4, 5

[6] Mingbao Lin, Rongrong Ji, Yan Wang, Yichen Zhang, Baochang Zhang, Yonghong Tian, and Ling Shao. Hrank: Filter pruning using high-rank feature map. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1529–1538, 2020. 1

[7] Shaoshuai Shi, Chaoxu Guo, Li Jiang, Zhe Wang, Jianping Shi, Xiaogang Wang, and Hongsheng Li. Pv-rcnn: Point-voxel feature set abstraction for 3d object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10529–10538, 2020. 2, 3

[8] Shaoshuai Shi, Zhe Wang, Jianping Shi, Xiaogang Wang, and Hongsheng Li. From points to parts: 3d object detection from point cloud with part-aware and part-aggregation network. *IEEE transactions on pattern analysis and machine intelligence*, 43(8):2647–2664, 2020. 2, 3

[9] Shaoshuai Shi, Li Jiang, Jiajun Deng, Zhe Wang, Chaoxu Guo, Jianping Shi, Xiaogang Wang, and Hongsheng Li. Pv-rcnn++: Point-voxel feature set abstraction with local vector representation for 3d object detection. *International Journal of Computer Vision*, 131(2):531–551, 2023. 3, 4

[10] OpenPCDet Development Team. Openpcdet: An open-source toolbox for 3d object detection from point clouds. https://github.com/open-mmlab/OpenPCDet, 2020. 1, 2

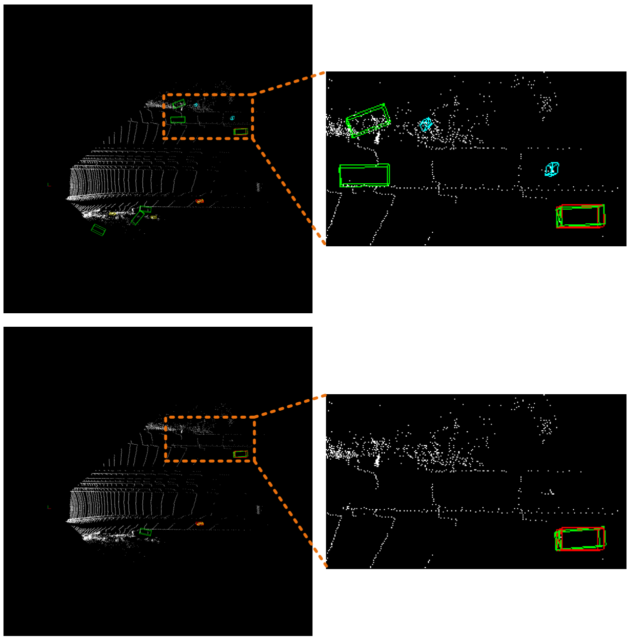
[11] Yan Yan, Yuxing Mao, and Bo Li. Second: Sparsely embedded convolutional detection. *Sensors*, 18(10):3337, 2018. 2, 3, 5

[12] Jihan Yang, Shaoshuai Shi, Runyu Ding, Zhe Wang, and Xiaojuan Qi. Towards efficient 3d object detection with knowledge distillation. *Advances in Neural Information Processing Systems*, 35:21300–21313, 2022. 1, 2, 4

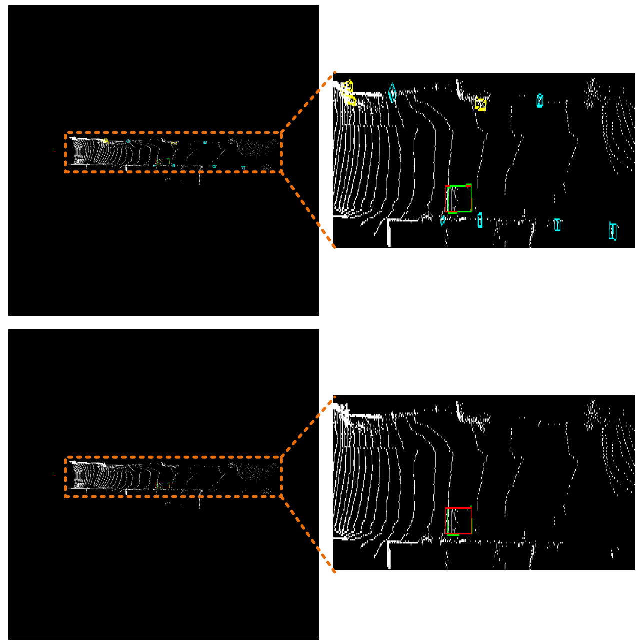
[13] Tianwei Yin, Xingyi Zhou, and Philipp Krahenbuhl. Center-based 3d object detection and tracking. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11784–11793, 2021. 2, 3, 4

[14] Haonan Zhang, Longjun Liu, Hengyi Zhou, Wenxuan Hou, Hongbin Sun, and Nanning Zheng. Akecp: Adaptive knowledge extraction from feature maps for fast and efficient channel pruning. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 648–657, 2021. 1

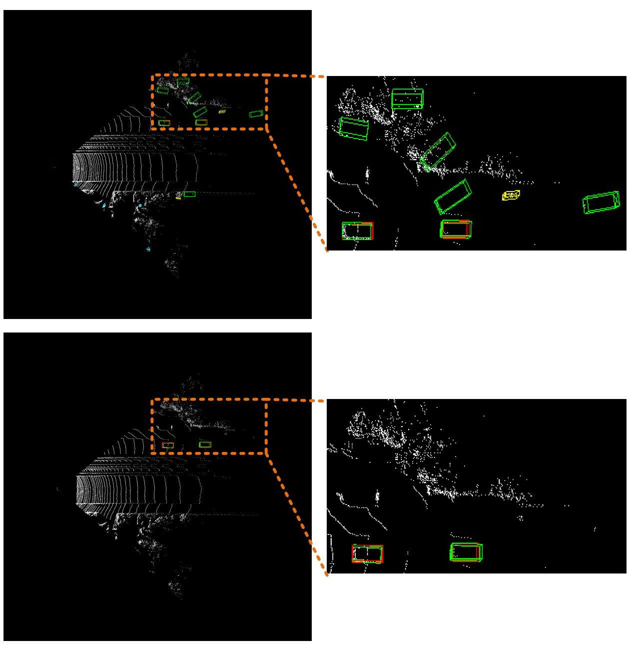
[15] Linfeng Zhang, Runpei Dong, Hung-Shuo Tai, and Kaisheng Ma. Pointdistiller: Structured knowledge distillation towards efficient and compact 3d detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 21791–21801, 2023. 4



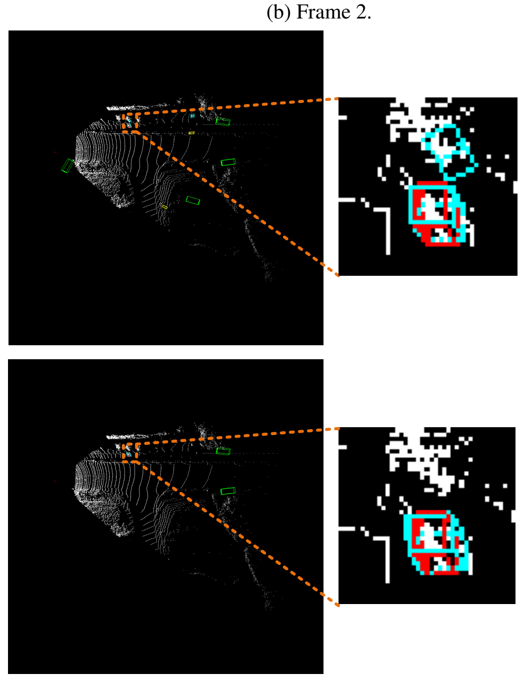
(a) Frame 1.



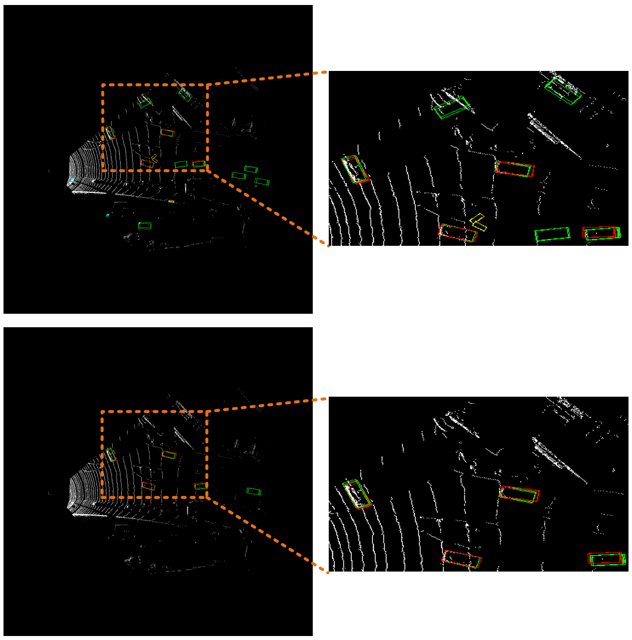
(b) Frame 2.



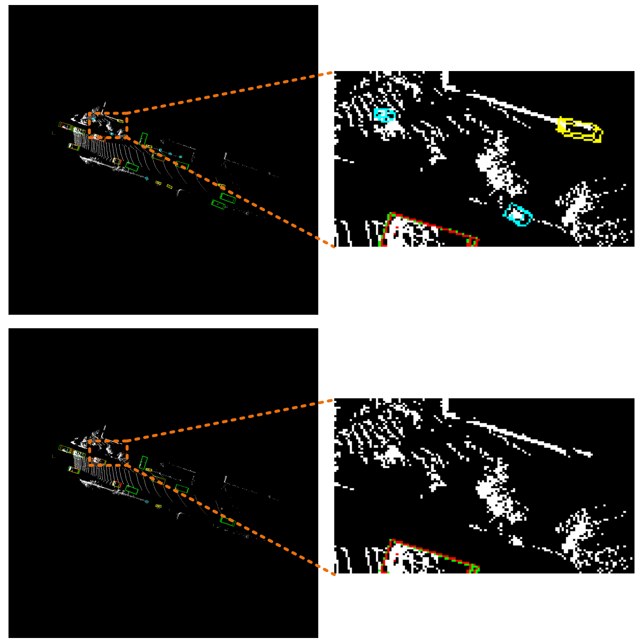
(c) Frame 3.



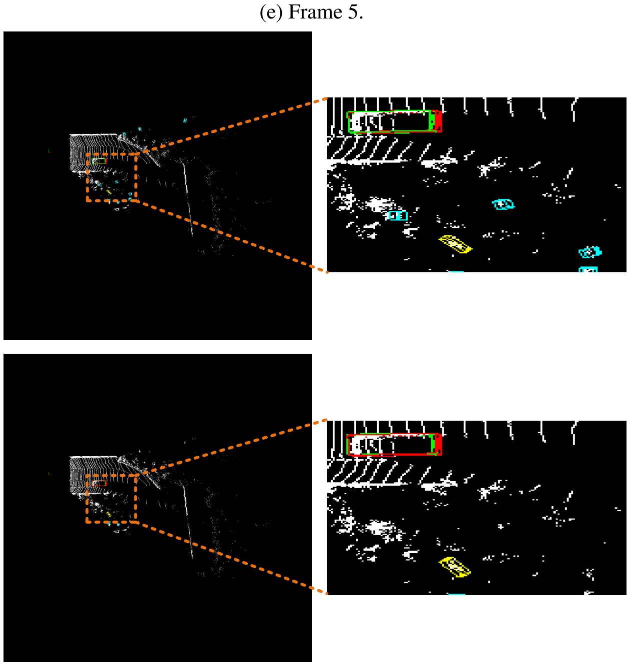
(d) Frame 4.



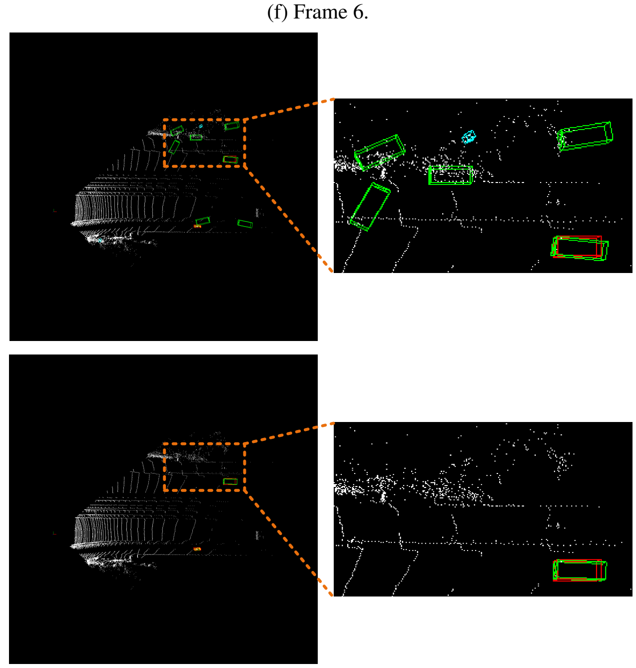
(e) Frame 5.



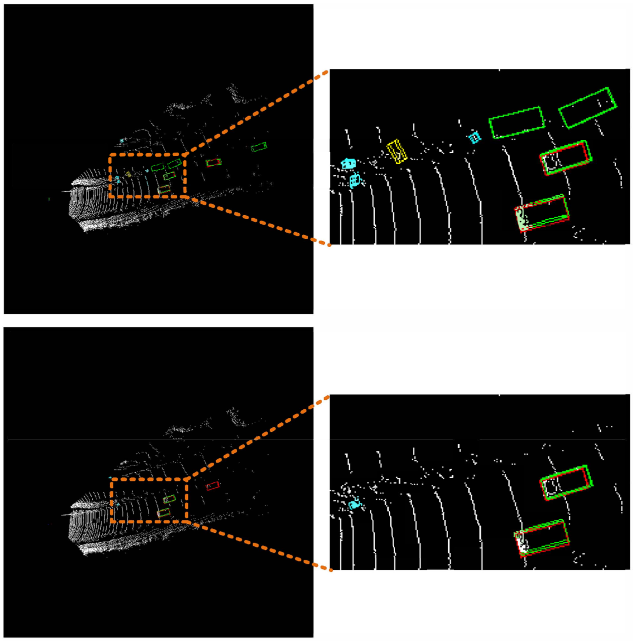
(f) Frame 6.



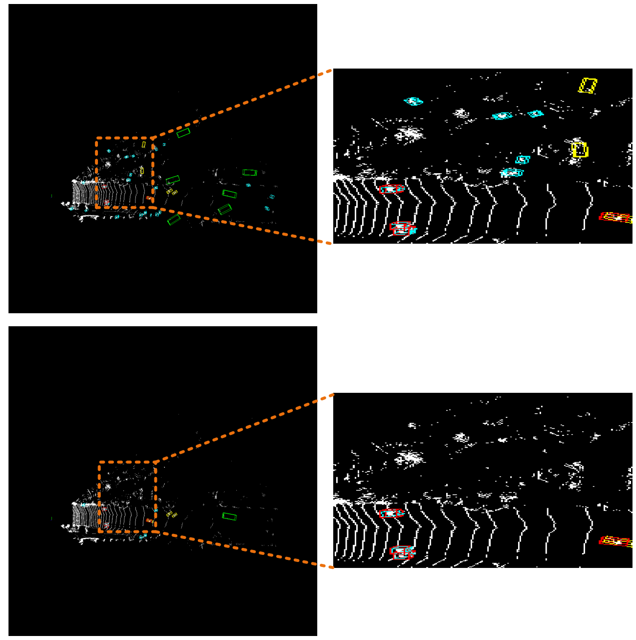
(g) Frame 7.



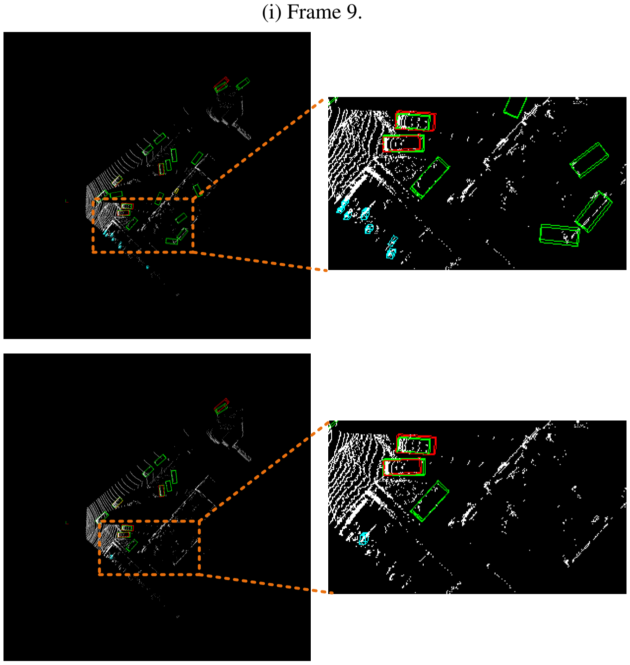
(h) Frame 8.



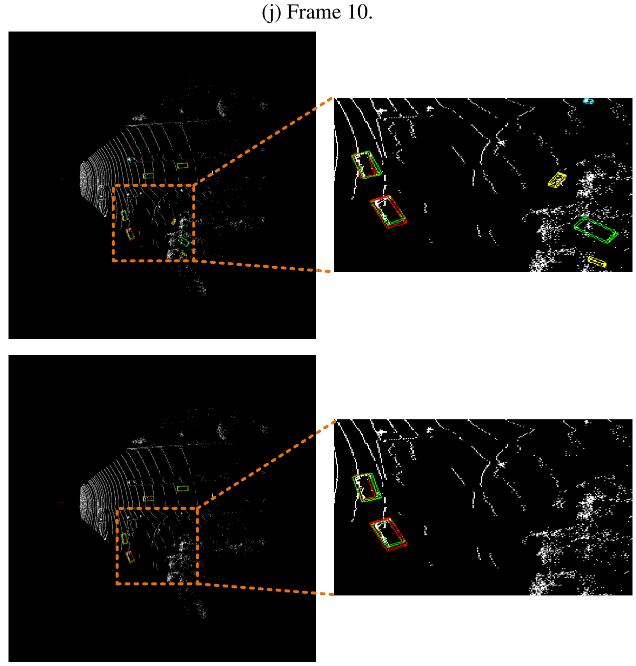
(i) Frame 9.



(j) Frame 10.



(k) Frame 11.



(l) Frame 12.

Figure 2. Visualization of the predictions before and after IOU-aware refinement. In each subfigure, the upper one of the two images represents the predicted results without refinement, while the lower one represents predicted results with our proposed IOU-aware refinement. The Green, yellow and blue boxes represent the predicted bounding boxes of Car, Pedestrian and Cyclist, respectively. The red boxes represent the ground truth bounding boxes.