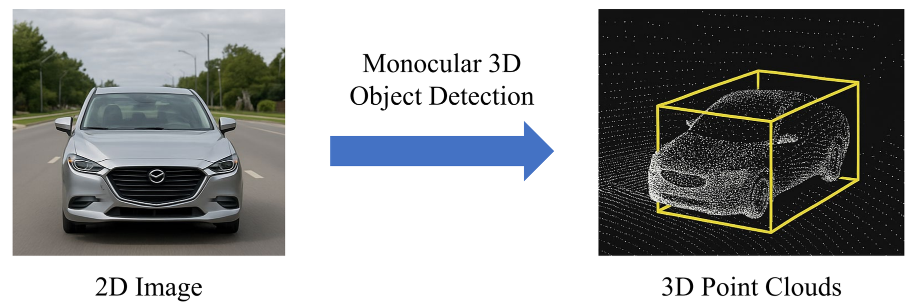
MonoTAKD: Teaching Assistant Knowledge Distillation for Monocular 3D Object Detection

Hou-I Liu[1], Christine Wu[2], Jen-Hao Cheng[2], Wenhao Chai[2], Shian-Yun Wang[3], Gaowen Liu[4], Hugo Latapie[5], Jhih-Ciang Wu[6], Jenq-Neng Hwang[2], Hong-Han Shuai[1], Wen-Huang Cheng[7]

[1]National Yang Ming Chiao Tung University, [2]University of Washington, [3]University of Southern California, [4]Cisco Research, [5]Taijitu AI, [6]National Taiwan Normal University, [7]National Taiwan University

Summary

Goal: Monocular 3D object detection infers object position, size, and orientation from a single RGB image without relying on depth sensors or LiDAR.



Challenges:

- **Depth-Guided:** Training the depth estimator on separate depth datasets may result in suboptimal performance due to domain mismatch.
- **Cross-Modal Distillation:** The effectiveness is limited by the significant feature representation gap between LiDAR-baed teacher and camera-based student.

Contribution:

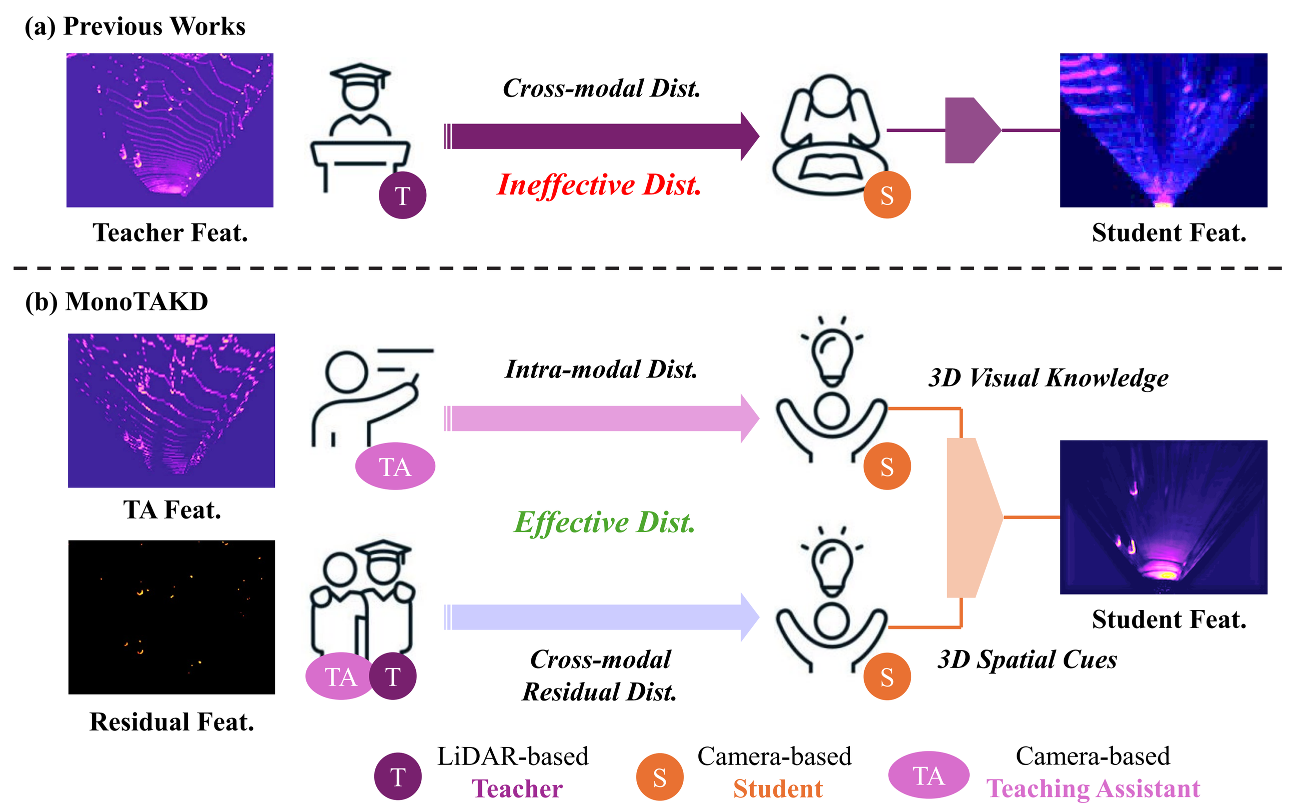
- We propose MonoTAKD, which leverages intra-modal and cross-modal residual distillation to transfer essential 3D knowledge to the student model.
- MonoTAKD achieves state-of-the-art results on KITTI3D and demonstrates strong generalizability on nuScenes and KITTI raw datasets under multi-view and unsupervised settings.

Official Code:
`https://github.com/hoiliu-0801/MonoTAKD`

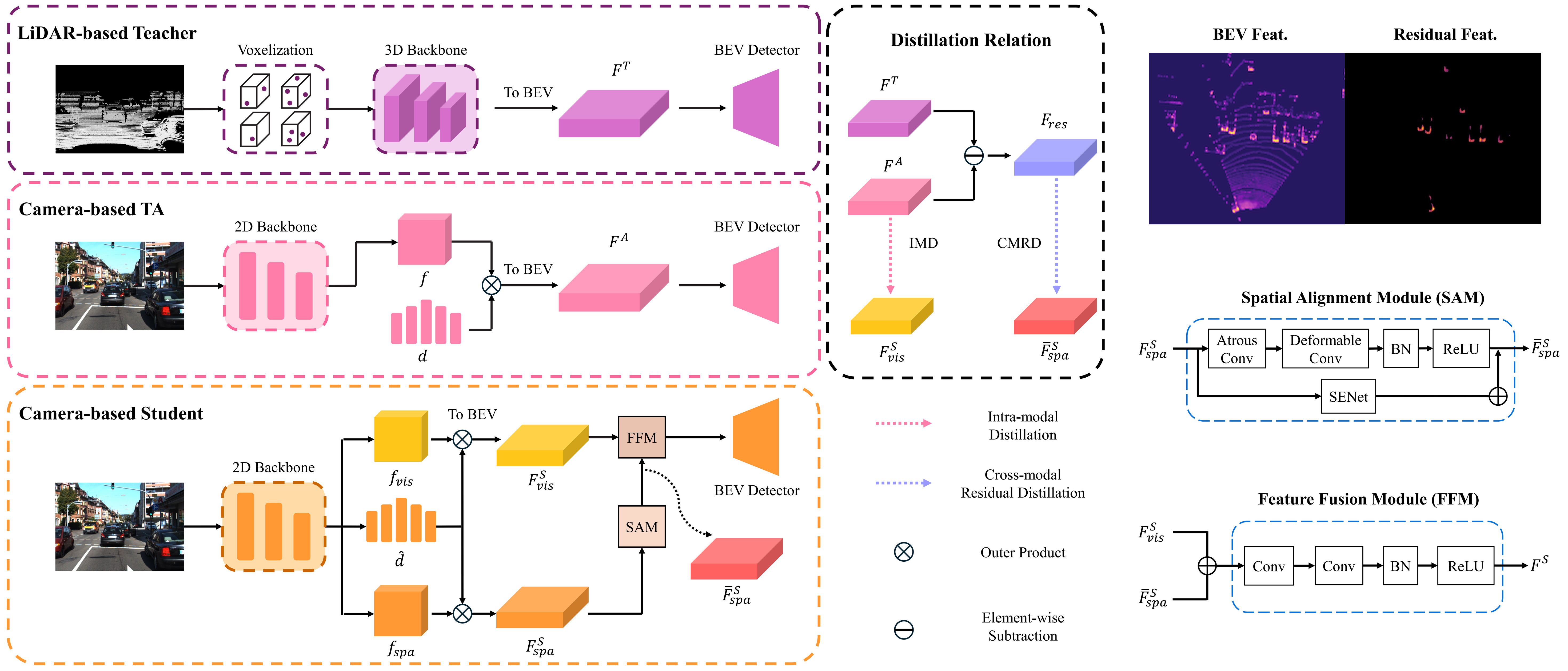
Motivation

We propose a camera-based Teaching Assistant (TA) to transfer knowledge from a LiDAR teacher to a camera student via two feature distillation losses.

- **Intra-modal Distillation:** Transfers 3D visual knowledge from TA to the student within the same modality, minimizing the feature gap.
- **Cross-modal Residual Distillation:** Provides 3D spatial cues to the student via residual features computed from the difference between TA and teacher.



Architecture



Experiments

Result on KITTI3D:

Method	Extra Data	AP_{3D} Easy	AP_{3D} Mod.	AP_{3D} Hard	AP_{BEV} Easy	AP_{BEV} Mod.	AP_{BEV} Hard
DDMP-3D	Pre-trained Depth	19.71	12.78	9.80	28.08	17.89	13.44
DD3D		23.22	16.34	14.20	30.98	22.56	20.03
Kinematic3D	Temporal	19.07	12.72	9.17	26.69	17.52	13.10
Dfm		22.94	16.82	14.65	31.71	22.89	19.97
CaDDN	LiDAR Auxiliary	19.17	13.41	11.46	27.94	18.91	17.19
MonoDTR		21.99	15.39	12.73	28.59	20.38	17.14
MonoNerd		22.75	17.13	15.63	31.13	23.46	20.97
MonoPGC		24.68	17.17	14.14	32.50	23.14	20.30
OccupancyM3D		25.55	17.02	14.79	35.38	24.18	21.37
Cube R-CNN	None	23.59	15.01	12.56	31.70	21.20	18.43
MonoUNI		24.75	16.73	13.49	33.28	23.05	19.39
MonoATT		24.72	17.37	15.00	36.87	24.42	21.88
MonoDETR		25.00	16.47	13.58	33.60	22.11	18.60
CMKD		25.09	16.99	15.30	33.69	23.10	20.67
MonoCD		25.53	16.59	14.53	33.41	22.81	19.57
MonoTAKD		**27.91**	**19.43**	**16.51**	**38.75**	**27.76**	**24.14**

Result on nuScenes:

Method	Modality	Backbone	NDS↑	mAP↑
BEVFormer	C	R50	0.423	0.352
+DistillBEV	L→C	R50	0.476	0.367
+STXD	L→C	R50	0.481	0.374
+TAKD (Ours)	L→C	R50	**0.490**	**0.392**
BEVFormer	C	R101	0.445	0.374
+DistillBEV	L→C	R101	0.545	0.446
+STXD	L→C	R101	0.543	0.440
+TAKD (Ours)	L→C	R101	**0.558**	**0.451**
BEVDepth	C	R50	0.440	0.317
+DistillBEV	L→C	R50	0.510	0.403
+LabelDistill	L→C	R50	0.528	0.419
+TAKD (Ours)	L→C	R50	**0.537**	**0.430**
BEVDepth	C	R101	0.535	0.412
+DistillBEV	L→C	R101	0.547	0.450
+LabelDistill	L→C	R101	0.553	0.451
+TAKD (Ours)	L→C	R101	**0.564**	**0.466**

Ablation on Feature Distillation Losses:

Settings	Loss: IMD	Loss: CMD	Loss: CMRD	AP_{3D} Easy	AP_{3D} Mod.	AP_{3D} Hard
1				24.35	16.16	13.44
2	✓			31.11	20.24	16.91
3		✓		28.22	18.29	15.10
4			✓	29.68	19.57	16.22
5	✓	✓		30.73	19.88	16.43
6	✓		✓	**34.36**	**22.61**	**19.88**

Ablation on SAM:

Components: ASPP	DConv.	SENet	FFM	AP_{3D} Easy	AP_{3D} Mod.	AP_{3D} Hard
				31.36	20.80	16.53
✓				32.28	21.45	17.18
	✓			32.46	21.51	17.99
✓	✓			33.19	21.81	18.23
✓	✓	✓		33.26	21.47	18.61
✓	✓	✓	✓	**34.36**	**22.61**	**19.88**

Efficiency analysis:

Model	Params. (M)	FLOPs (G)	AP_{3D} Easy	AP_{3D} Mod.	AP_{3D} Hard
MonoNeRD	83.0	356.57	20.64	15.44	13.99
DD3Dv2	80.3	163.00	26.23	21.21	18.83
MonoDETR	47.4	57.51	28.84	20.61	16.38
CMKD	**45.1**	**41.32**	23.53	16.33	14.44
TAKD-Lite	**45.1**	**41.32**	31.36	20.80	16.53
TAKD	47.8	44.90	**34.36**	**22.61**	**19.88**

Qualitative Visualization: