自動運転に向けた3D 物体検出手法の知識蒸留の検討

大橋 有[1], 池畑 諭[2,3], 石川 康太[2], 佐藤 育郎[1,2]

[1]東京科学大学, [2]デンソーITラボラトリ, [3]国立情報学研究所

Recognition, Control and Learning Algorithm Lab.

導入

背景

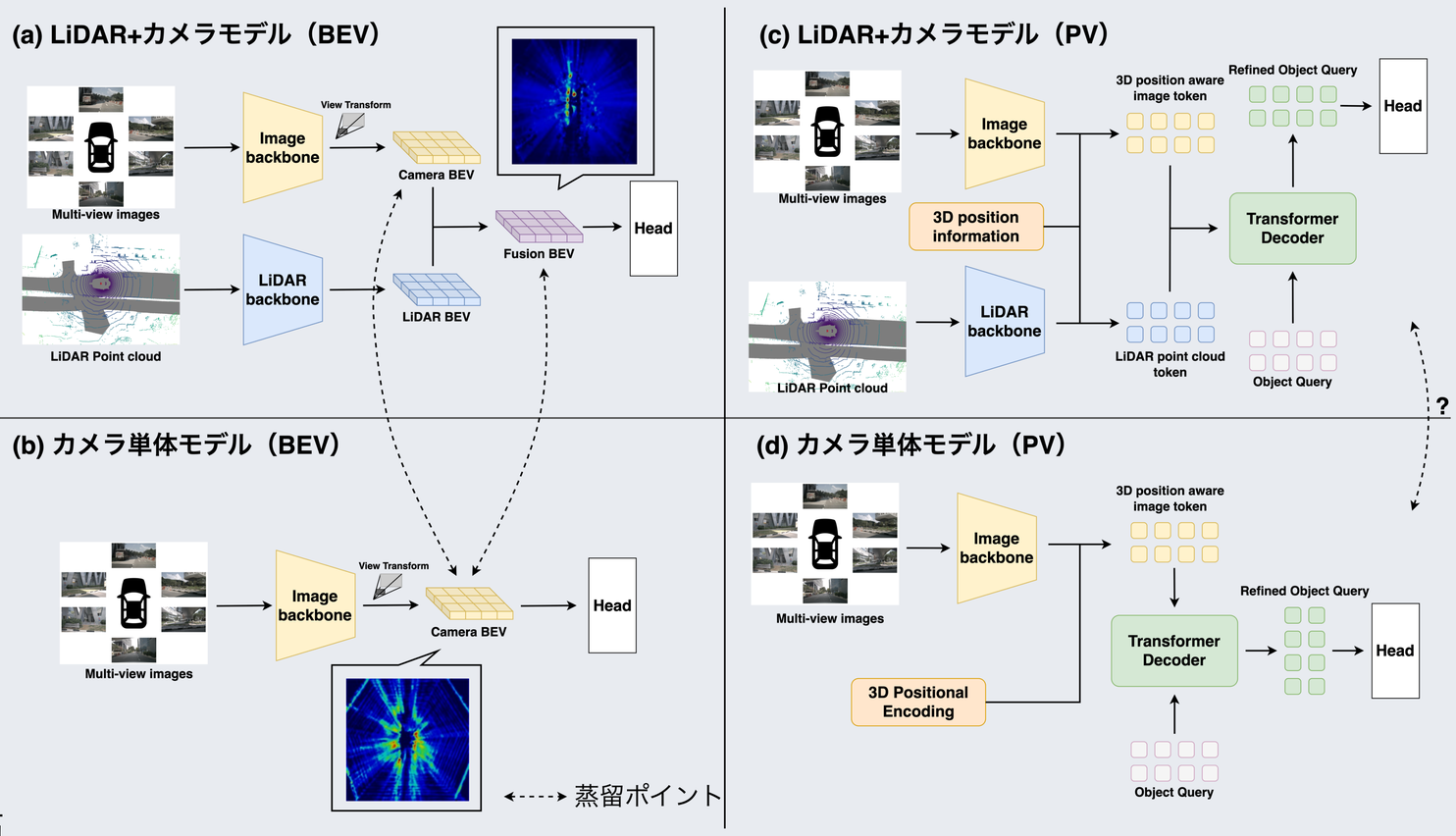
- カメラのみの3D物体検出は低コストな反面，位置推定精度が課題.
- LiDAR+カメラモデルからカメラ単体モデルへのクロスモーダル知識蒸留が有望視されている.

標準的な特徴表現の方式

- Bird's-Eye View上の特徴表現を利用するもの (‘BEV’)
- BEVを介さないPerspective View上の特徴表現を利用するもの (‘PV’)

検討方針

1. BEV方式とPV方式の有効性の見極め
2. PV方式を用いた知識蒸留手法の開発・評価



既存研究調査：BEV方式 - PV方式

調査範囲

- 主に2021~2025年にCV系トップ会議で発表された論文のうちnuScenesデータセットでの評価結果を調査した (右表1, 2).

調査結果

- C単体では最新のPV方式とBEV方式の手法はほぼ同性能.

  それぞれのSOTAであるRayDNとGeoBEVはどちらも0.52mAP，0.61NDS程度
- C+Lでは，PV方式がBEV方式を上回る.

  PV方式のSOTAであるSparseLIFは，BEV方式のSOTAであるBEVFusionに比べて+0.027mAP，+0.032NDS
- BEV蒸留を用いた生徒モデルは最新のPV方式の手法の性能には届いていない.

  DistillBEV(S)はRayDNと比較して−0.068mAP，−0.057NDS
- PV方式はBEV方式に比べて速い推論速度になっている（右表2）.

  StreamPETRはBEVDetに対して約1.6倍の推論速度

表1. nuScenesデータセットにおける3D物体検出評価結果

Method		Reference	Modaliry	Backbone		validation set	
				Camera	LiDAR	mAP ↑	NDS ↑
BEV	BEVDet	arXiv2021	C	ResNet101	-	0.302	0.381
	BEVFormer	ECCV2022	C	ResNet101	-	0.416	0.517
	BEVDepth	AAAI2023	C	ResNet101	-	0.412	0.535
	GeoBEV	AAAI2025	C	ResNet101	-	0.526	0.615
	BEVFusion	ICRA2023	C+L	Swin-T	VoxelNet	0.685	0.714
PV	DETR3D	CoRL2022	C	ResNet101	-	0.349	0.434
	PETR	ECCV2022	C	ResNet101	-	0.370	0.442
	PETRv2	ICCV2023	C	ResNet101	-	0.421	0.524
	StreamPETR	ICCV2023	C	ResNet101	-	0.504	0.592
	RayDN	ECCV2024	C	ResNet101	-	0.518	0.604
	CMT	ICCV2023	C+L	V2-99	VoxelNet	0.703	0.729
	SparseFusion	ICCV2023	C+L	Swin-T	VoxelNet	0.710	0.731
	SparseLIF	ECCV2024	C+L	V2-99	VoxelNet	0.712	0.746
BEV 蒸留	UniDistill(S)	CVPR2023	C	ResNet50	-	0.265	0.378
	DistillBEV(S)	ICCV2023	C	ResNet101	-	0.450	0.547
	SimDistill(S)	AAAI2024	C	Swin-T	-	0.404	0.453

※CはCameraを表し，LはLiDARを表している．また，BEV蒸留の欄の (S)は生徒モデルを表す.

表2. BEV方式とPV方式のFPS比較

Method	Frames	FPS↑
BEVDet	1	16.7
BEVDepth	2	15.7
PETRv2	2	18.9
StreamPETR	8	27.1

結論

BEV方式は異なるモダリティ同士の統合が容易だが計算負荷が大きく，より高速かつ高精度なPV方式の方が有力な選択肢である.

手法の開発方針

- 既存研究調査の結果からPV方式による手法は高速かつ高精度であることが確認できた.

  → PV方式を基盤とした知識蒸留を検討する.
- LiDAR教師の幾何的情報を，BEV表現を介さずにCameraベース生徒へ伝達するクロスモーダル蒸留を設計する.

今後の計画

生徒モデルの構造の選定

教師の豊富な幾何情報を効率的に取り込めるPV方式のモデルを選定する.

蒸留損失の適用箇所の検討

蒸留を適用する層を探索的に選定する.

本研究は東京科学大学「DENSO IT LAB 認識・制御・学習アルゴリズム共同研究講座」における研究プロジェクト「機械学習の実応用に向けた品質基盤開発」の成果である