

BEV-ConvFusion: A Synergistic 2D Fusion Framework for Real-Time Autonomous Perception



2025 CUA이 중앙대학교 인공지능 학회 하계 컨퍼런스
Proceeding of 2025 Chung-Ang University Artificial Intelligence Summer Conference

Abstract

- 자율주행은 정확도와 실시간 효율성이 모두 요구된다. 기존 LiDAR-카메라 융합은 3D 백본에 의존하여 계산량이 크다.
- 본 연구는 2D 기반 BEV-ConvFusion을 제안하여 계산 효율성을 높이고, 핵심 모듈인 SynCAM을 통해, 이미지와 BEV 특징을 상호 보완적으로 정제한다.

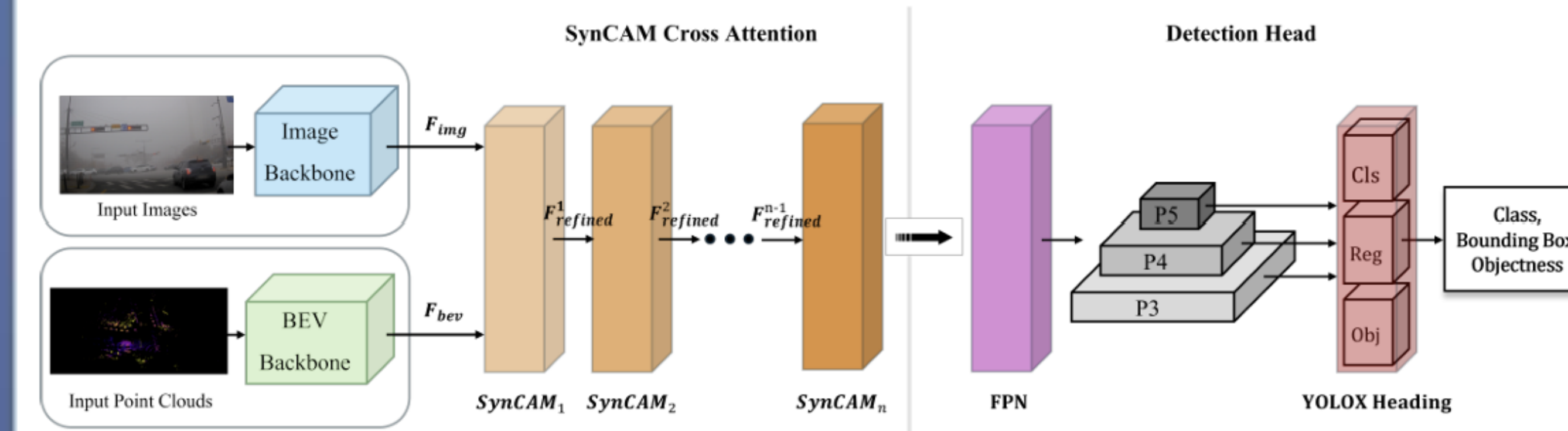
Introduction

- 카메라는 풍부한 의미 정보를 가지고 있지만, 조명 및 날씨에 취약하다.
- 반대로, LiDAR는 강건한 기하학 정보를 지니고 있지만, 의미가 부족하다.
- 현재, 카메라-LiDAR의 융합은 3D 백본을 사용하며, 실시간 적용이 어렵다는 단점이 있다.
- 해결책으로, 본 프로젝트는 2D BEV 표현과 SynCAM 기반으로 효율적으로 카메라와 LiDAR를 통합한다.

Aim

- 실시간성과 정확성을 모두 갖춘 자율주행 인식 모델 개발
- LiDAR-카메라 상호 보완적 특징을 효과적으로 융합

Methods

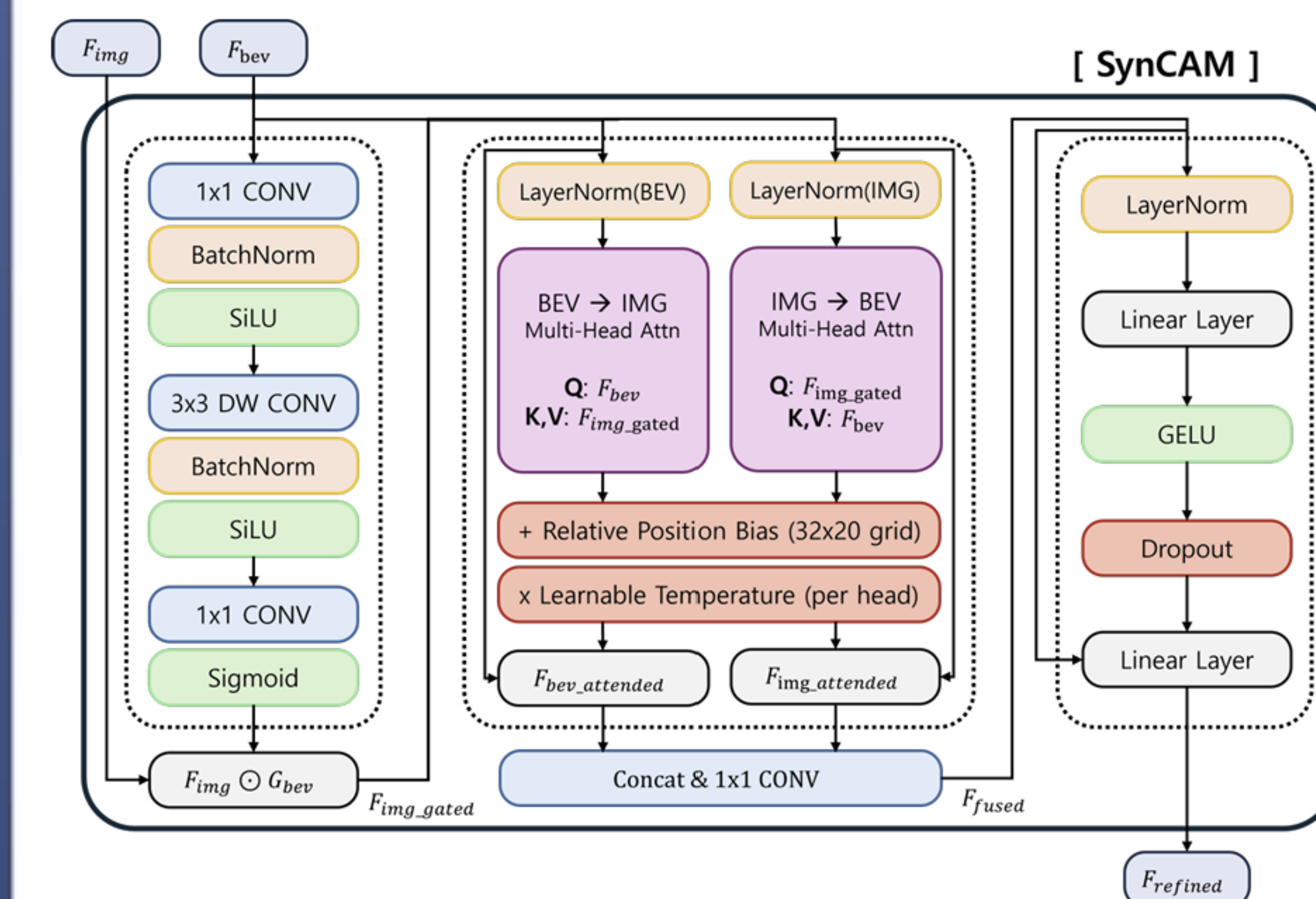


BEV 변환

LiDAR Point Cloud를 Height, Density, Intensity 3채널 BEV 맵으로 변환한다.

Dual Stream Encoder

카메라는 ConvNext-Tiny를 사용하며, BEV는 ResNet-18을 사용하였다.



SynCAM (3 Stages)

- Spatial Gating
- Bidirectional Cross-Attention
- Feature Refinement

Detection Head: YOLOX 기반

Results

TABLE I
PERFORMANCE OF CAMERA-ONLY 2D OBJECT DETECTION MODELS.

Backbone	mAP@[.5:.95]	mAP@.50
EfficientNet [12]	50.3	69.1
ConvNeXt-Tiny [10]	54.3	73.3
YOLOv8-m [15]	33.4	51.6
Swin-Tiny [9]	52.0	71.0

TABLE II
PERFORMANCE COMPARISON WITH 2D AND 3D MODELS.

Modality	Method	mAP@.50
2D-Only	ConvNeXt-Tiny [10]	73.3
3D-Only	PV-RCNN [16]	72.3
	Voxel R-CNN [17]	71.6
	SECOND [18]	68.6
	PointRCNN [19]	66.4
Fusion	BEV-ConvFusion	75.3

TABLE III
ABLATION STUDY ON THE NUMBER OF STACKED SYNCAM BLOCKS.

# SynCAM Blocks	mAP@.50
1 Block	73.6
2 Blocks	74.7
3 Blocks	75.3
4 Blocks	73.9

Conclusion

- 2D 기반의 BEV-ConvFusion은 실시간 적용 가능성과 높은 정확도를 보였다.
- SynCAM은 이미지-LiDAR 상호 보완적 정제를 한다.
- 기존의 단일 모달 및 3D 백본 대비 효율적, 강건한 융합 성능을 달성했다.