UniSeg: A Unified Multi-Modal LiDAR Segmentation Network and the OpenPCSeg Codebase

Youquan Liu   Runnan Chen   Xin Li   Lingdong Kong   Yuchen Yang   Zhaoyang Xia   Yeqi Bai

Xinge Zhu   Yuexin Ma   Yikang Li   Yu Qiao   Yuenan Hou
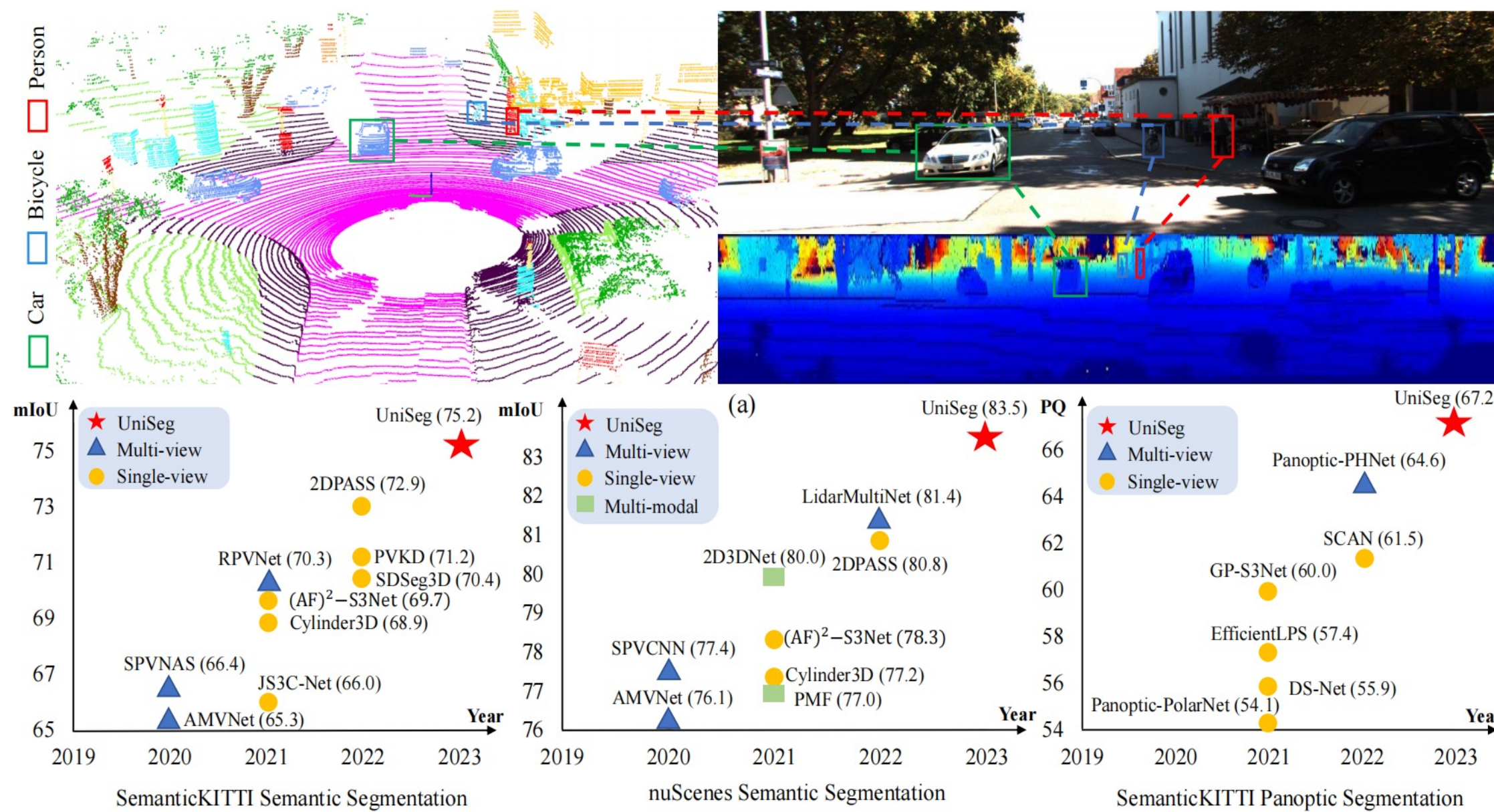
Shanghai Artificial Intelligence Laboratory

ICCV23 PARIS

Motivation & Contribution

TL;DR

➤ We introduce UniSeg, a unified multi-modal fusion network for LiDAR segmentation that leverages the information of RGB images and three representations of LiDAR point clouds for more accurate and robust 3D scene understanding.

➤ This work ranks 1st on two LiDAR segmentation challenges of SemanticKITTI and nuScenes, strongly demonstrating the efficacy of the proposed multi-modal approaches.

The OpenPCSeg Codebase

➤ There only exists a few open-source projects for the LiDAR segmentation task and their released implementations are often difficult to reproduce.

➤ To facilitate relevant research, this work construct the largest and most comprehensive LiDAR segmentation codebase dubbed OpenPCSeg that tailored to ensure reproducible implementations of competitive LiDAR segmentation models.

➤ The OpenPCSeg project is built based on the well-known OpenPCDet project and has reproduced more than ten outdoor LiDAR segmentation algorithms from different LiDAR representations, such as voxel and multi-modal fusion.

Table 1: Comparisons between existing codebase.

Codebase	Task	Task Difficulty	#Method
MMDetection3D	Indoor Seg	Relatively Easy	3
OpenPCSeg	Outdoor Seg	Difficult	14

## Methodology

Unified Multi-Modal Network

➤ This work presents the Learnable cross-Modal Association (LMA) module that automatically fuses voxel and range-view features with image features, fully utilizes the rich semantic information of images and are robust to calibration errors.

➤ The enhanced voxel and range-view features are transformed to the point space, where three views of point cloud features are further fused adaptively by the designed Learnable cross-View Association (LVA) module.

LMA: Learnable Cross-Modal Association
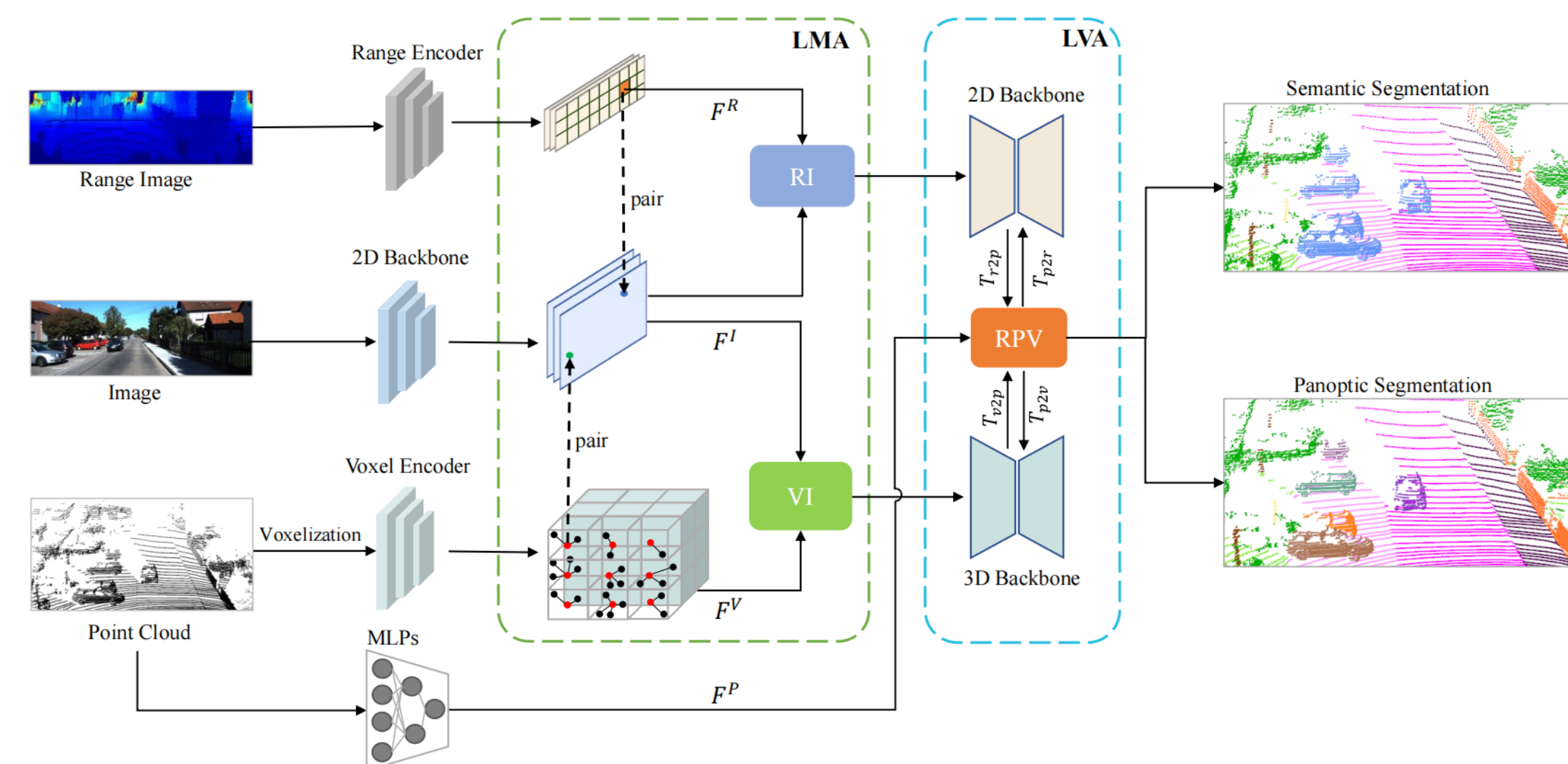
MHCA: Multi-Head Cross-Attention
$\copyright$: Concatenation

LVA: Learnable Cross-View Association
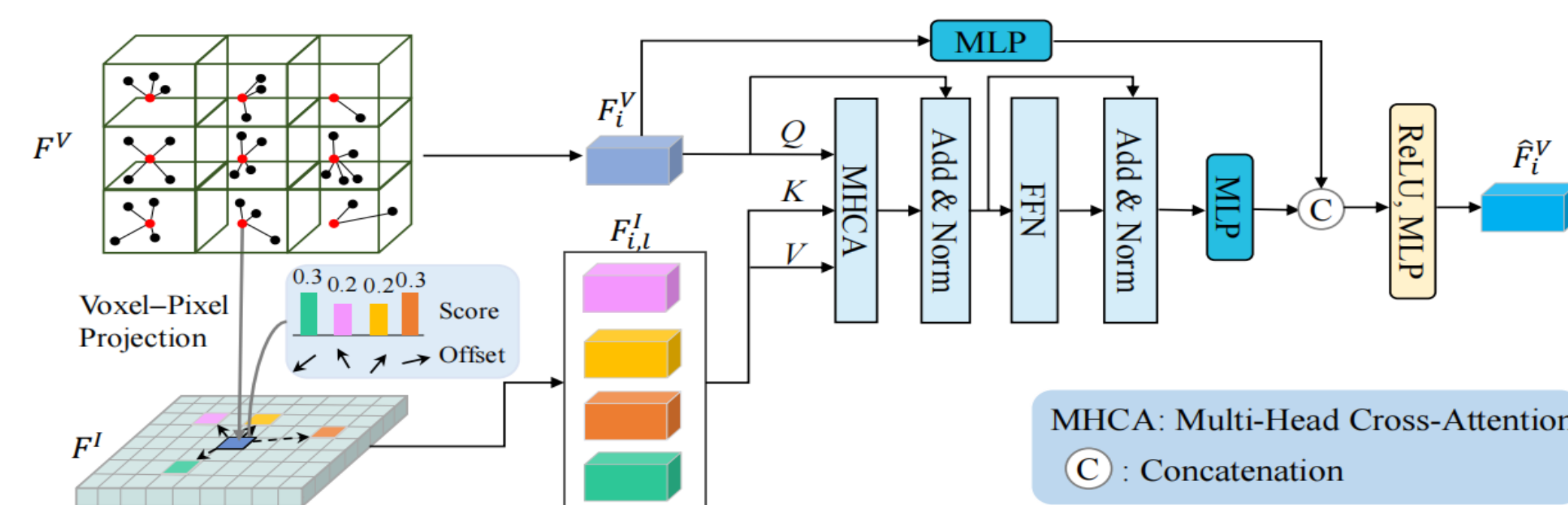
✕ : Multiplication
✚ : Addition
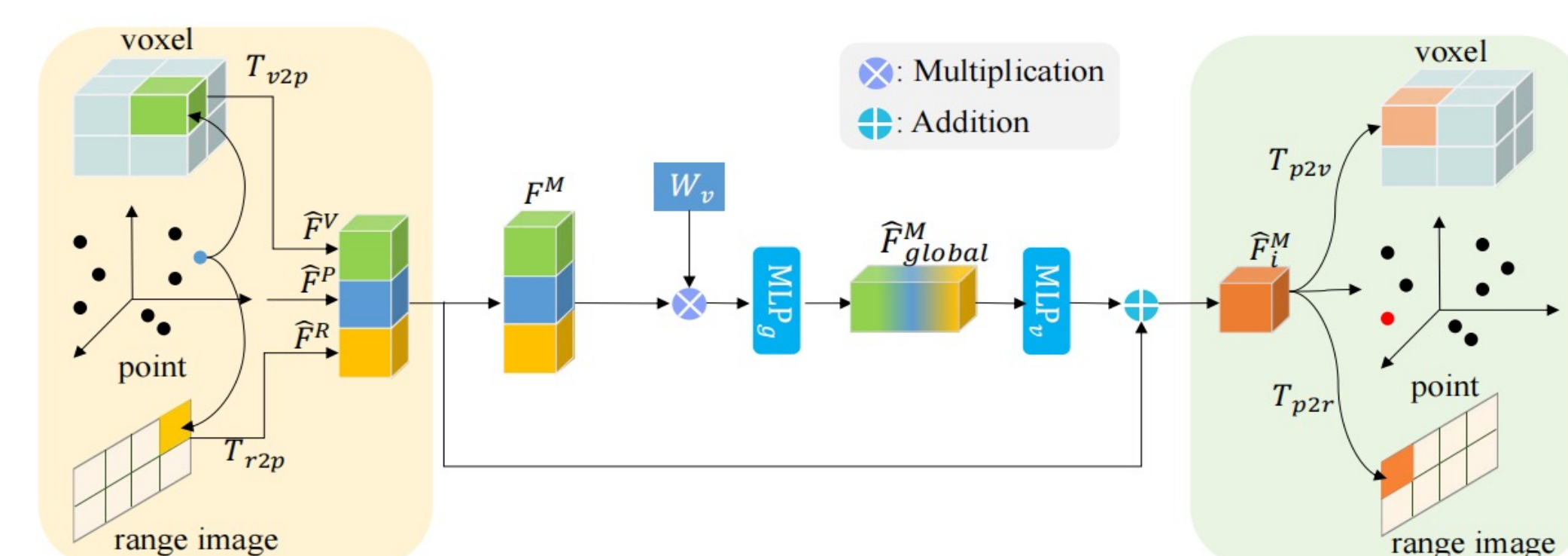
Experiments & Analysis

Accurate LiDAR Segmentation

➤ Influence of Different Modalities and Views.

Voxel	Point	Range image	RGB Image	mIoU
✓				68.4
	✓			13.7
		✓		55.8
✓			✓	69.7
		✓	✓	58.1
✓	✓			68.5
✓		✓		69.7
✓	✓	✓	✓	71.3

➤ Effect of the Learnable cross-Modal Association (LMA) module.

Method	mIoU	Δ
Early Fusion Add (Baseline)	70.1	+0.0
Early Fusion Concat	69.4	−0.7
PointPainting*	70.4	+0.3
PointAugmenting*	70.5	+0.4
LMA (Ours)	71.3	+1.2

➤ Effect of the Learnable cross-View Association (LVA) module.

Method	mIoU	Δ
Add (Baseline)	70.4	+0.0
Concat	70.5	+0.1
Self-Attention	70.4	+0.0
LVA	71.3	+0.9

➤ Qualitative results of the UniSeg framework and the baseline.

(a) Error map by baseline   (b) Error map by UniSeg   (c) Ground Truth

Paper

Code