

Bootstrapping a 4D LiDAR Annotation Tool from Video Foundation Models

Jihun Kim^{1*}, Hyun-Kurl Jang^{1*}, Hyemin Yang^{1*}, Jinnyeon Yang^{1*}, Hyeokjun Kwon^{2*}, Kuk-Jin Yoon¹

¹KAIST, Visual Intelligence Lab ²Chung-Ang University, FoVLab

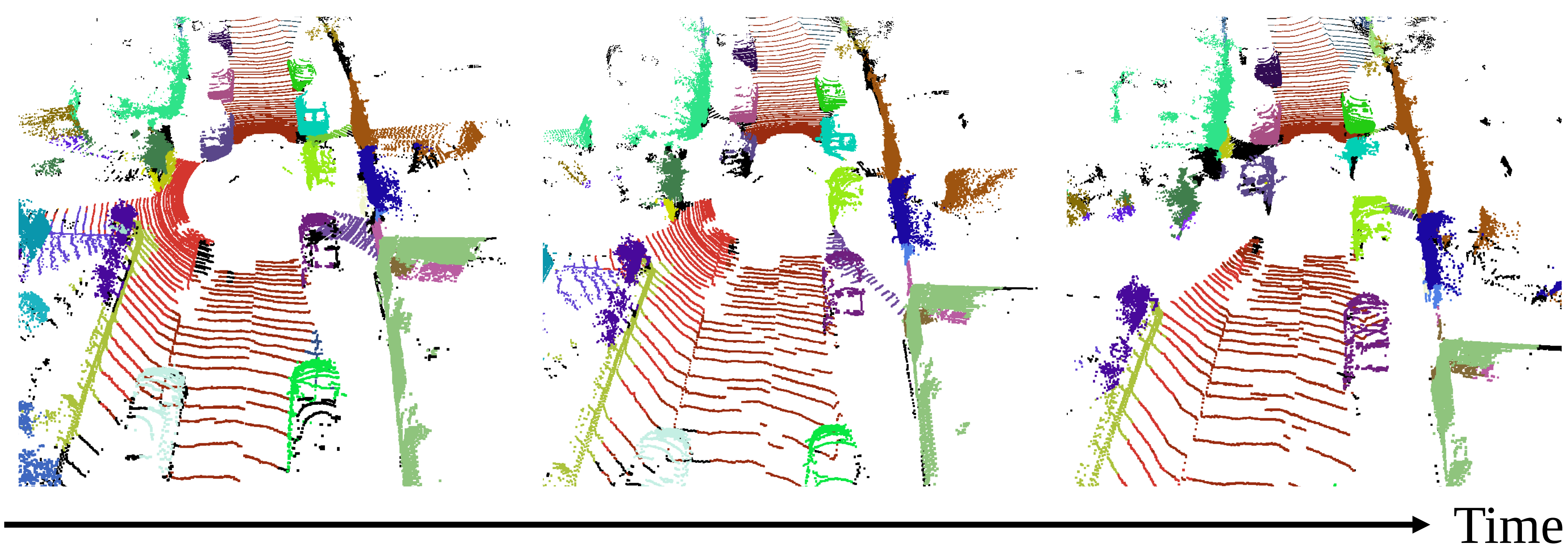
* Equal contribution

Motivation

- Mask Propagation via Segment Anything Model 2

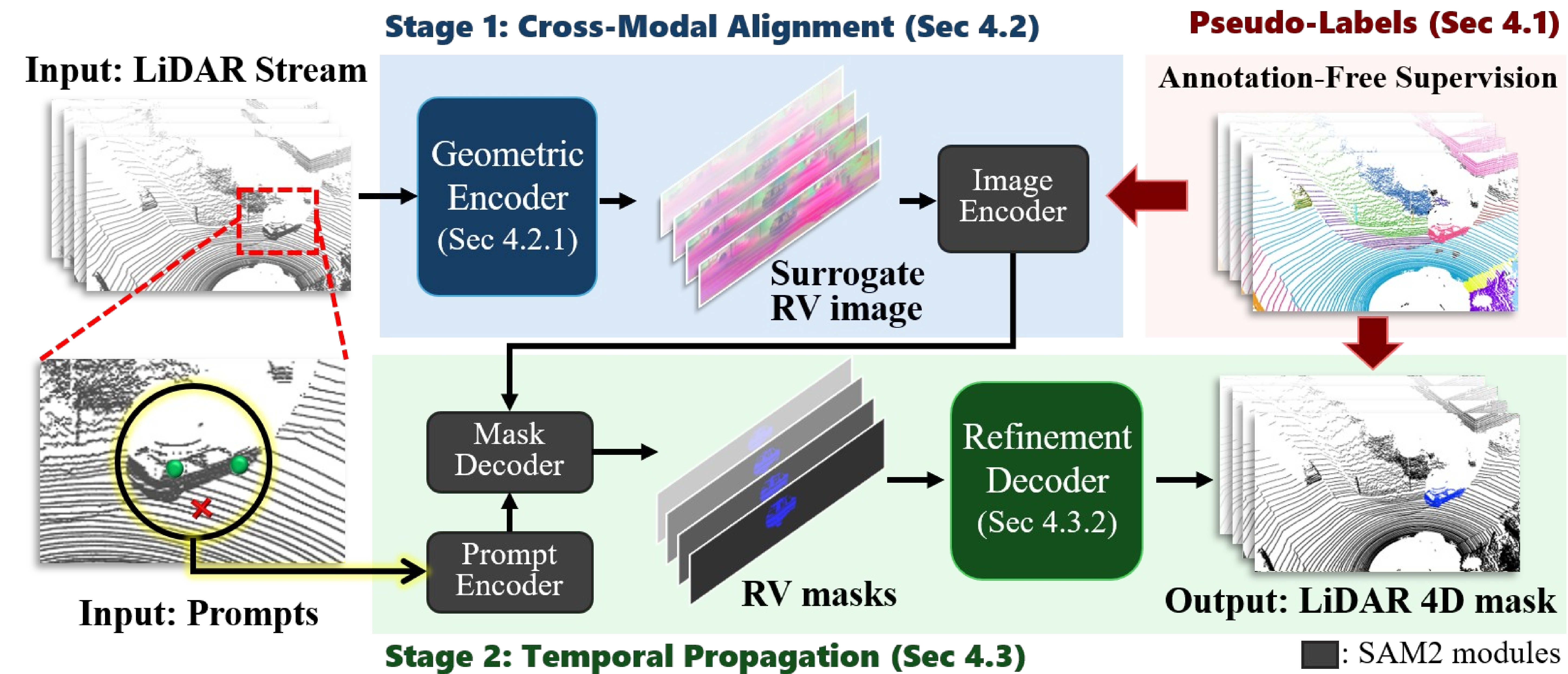


- Temporal Propagation on 4D LiDAR



LiDAR-SAM2

- Stage 1: Learning Cross-Modal Alignment



Experimental Results

- Semantic Segmentation Results

Label type	pts/frame	MinkUNet [6]	PTv2 [36]
Point labels	10	5.2	7.4
SAM2 labels	10	24.3	26.4
Our labels	10	55.2	61.3
Full GT	100k	63.8	70.3

- Panoptic Segmentation Results

Method	LSTQ	S_{assoc}	S_{cls}	IoU_{st}	IoU_{th}
4D-PLS [1] + Our labels	46.9	46.3	47.6	54.3	44.3
4D-PLS [1] + Full GT	62.7	65.1	60.5	65.4	61.3
4D-StOP [13] + Our labels	57.7	64.2	51.7	56.5	51.7
4D-StOP [13] + Full GT	67.0	74.4	60.3	65.3	60.9
Mask4Former [41] + Our labels	64.8	70.7	59.4	56.9	62.8
Mask4Former [41] + Full GT	70.5	74.3	66.9	67.1	66.6

- 4D LiDAR Annotation Results

