Hidden Gems: 4D Radar Scene Flow Learning Using Cross-Modal Supervision

Fangqiang Ding, Andras Palffy, Dariu M. Gavrila, Chris Xiaoxuan Lu

WED-AM-106 (Highlight)

THE UNIVERSITY of EDINBURGH

TU Delft

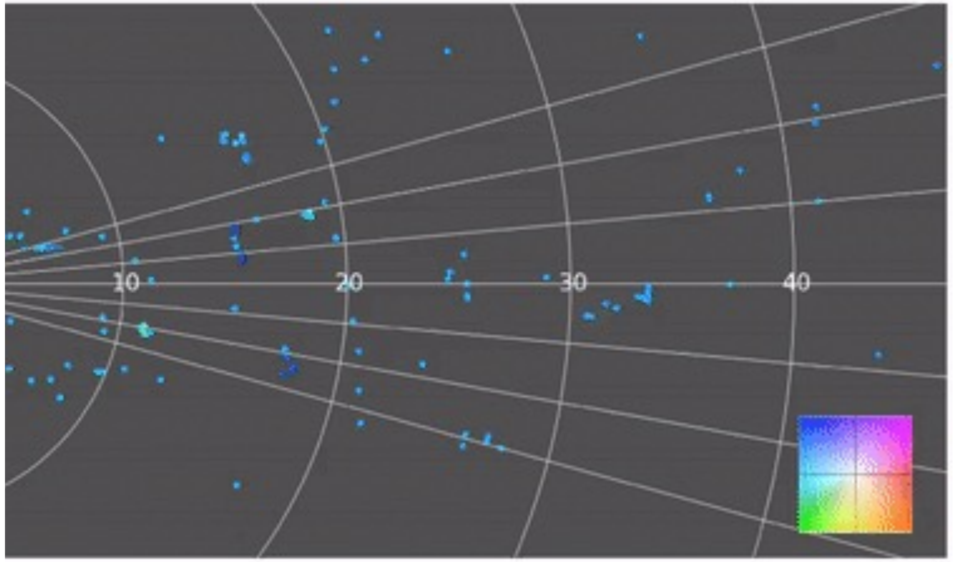
JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Code Available

Problem definition

Input (4D radar point clouds)
Perspective view

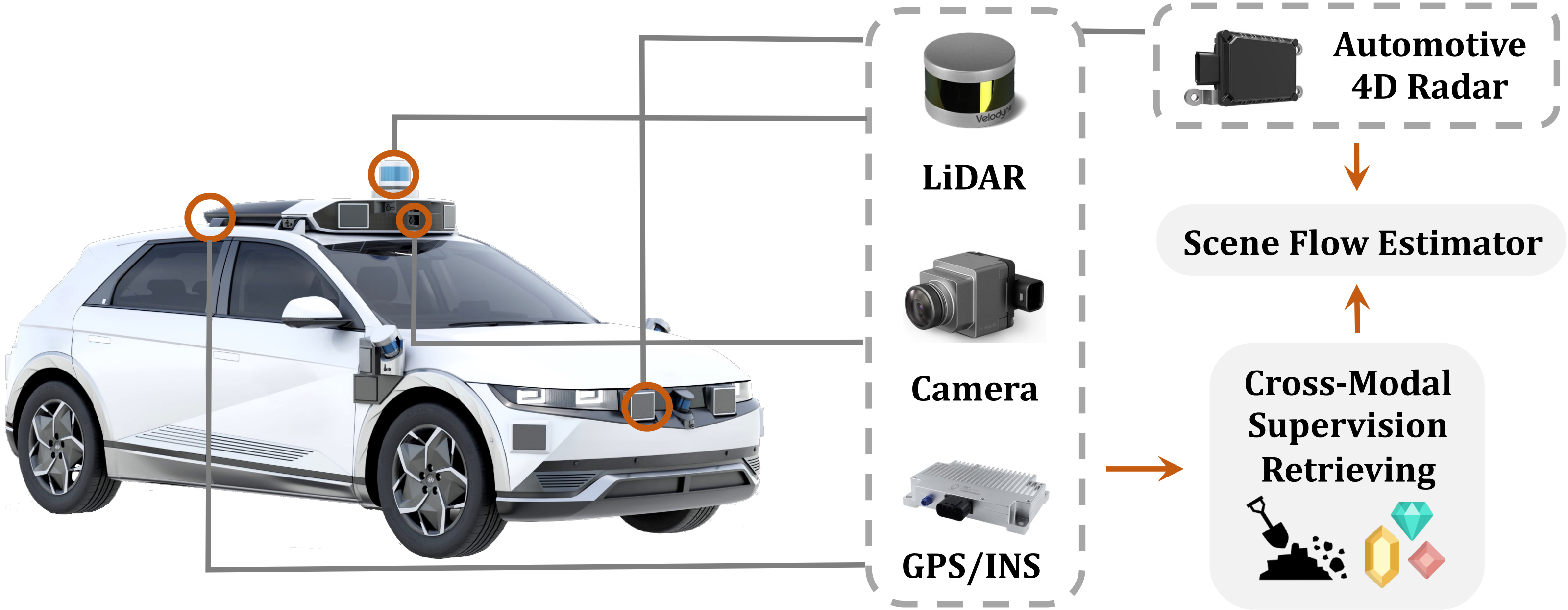
Output (point-level scene flow)
Bird's eye view



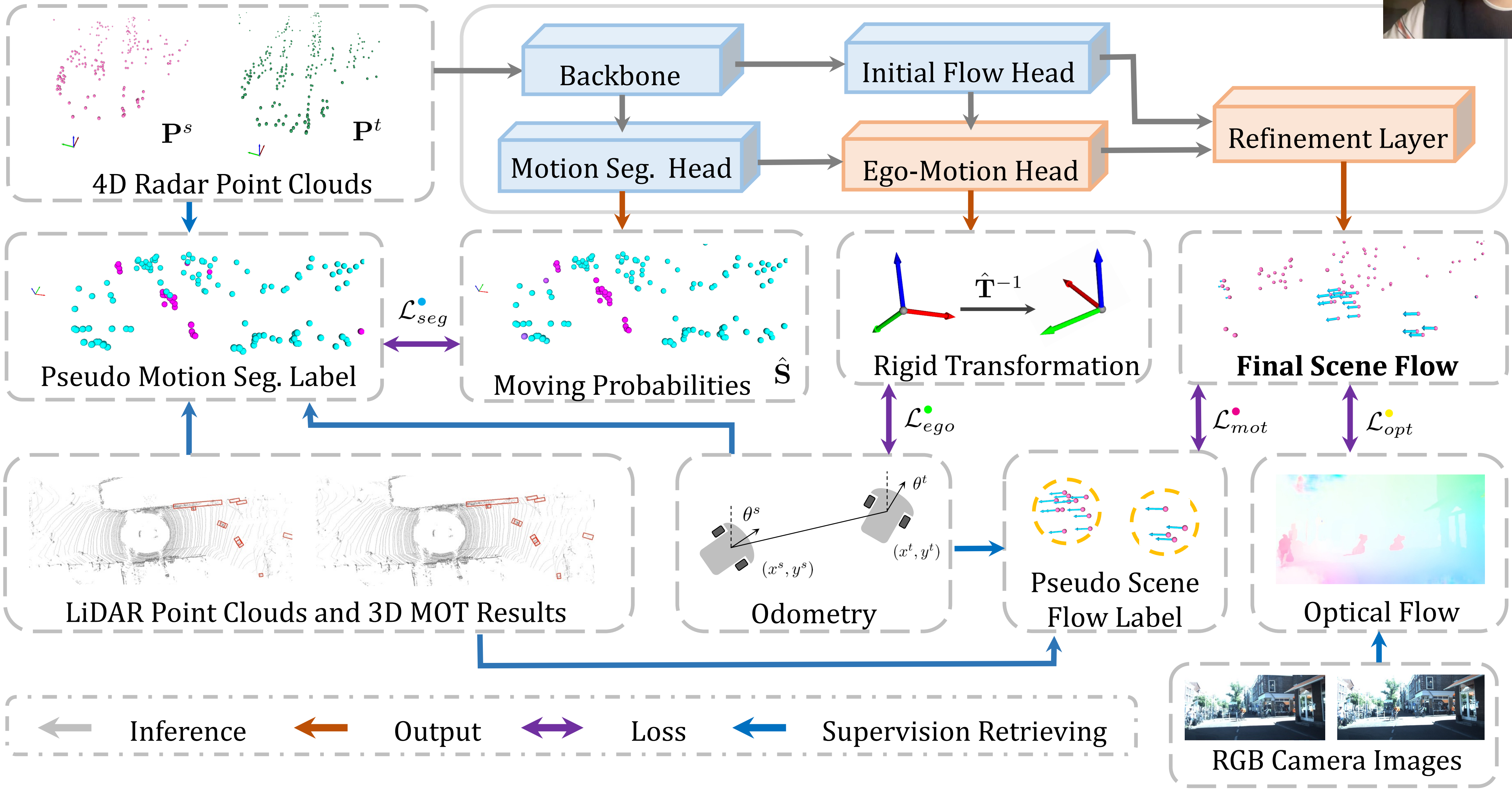
Given consecutive point clouds from 4D radar, we learn to estimate point-level scene flow using cross-modal supervision.

Motivation

- **Fact**: self-driving cars today are equipped with heterogeneous sensors.
- **Insight**: such co-located perception redundancy can be used to provide supervision cues that bootstrap 4D radar scene flow learning.

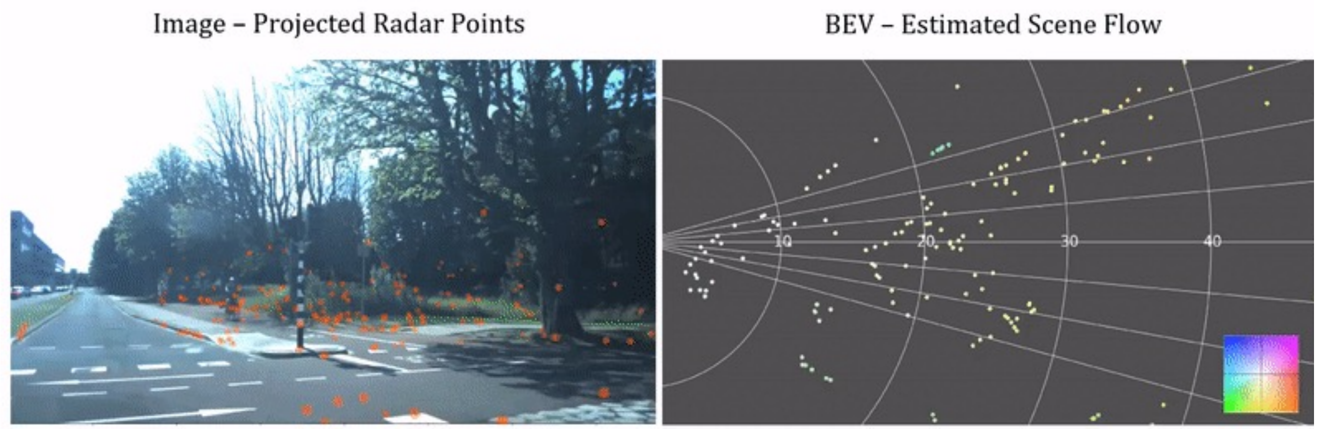


Cross-modal supervised learning pipeline

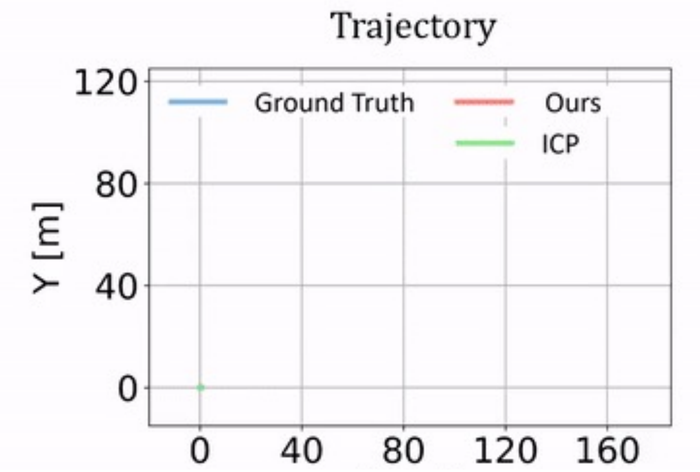


Qualitative results

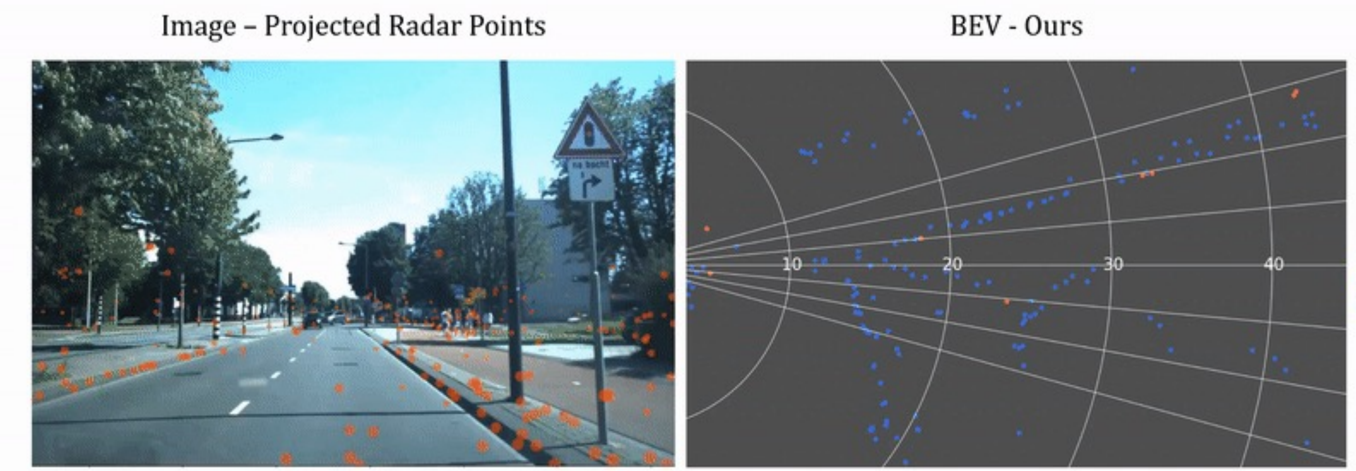
- Scene Flow Estimation

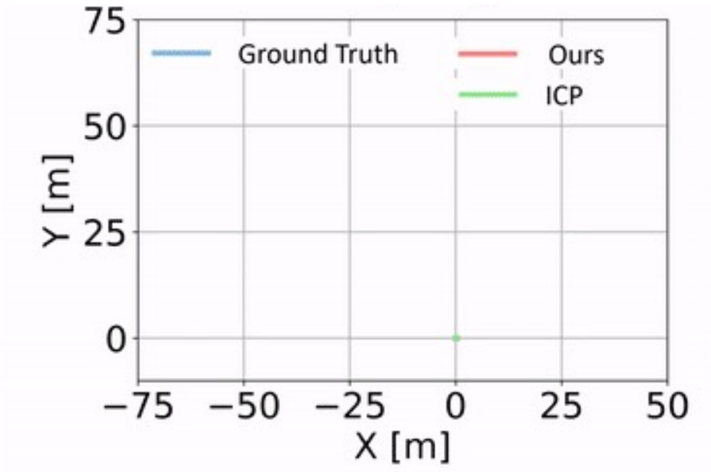


- Ego-motion Estimation



- Motion Segmentation



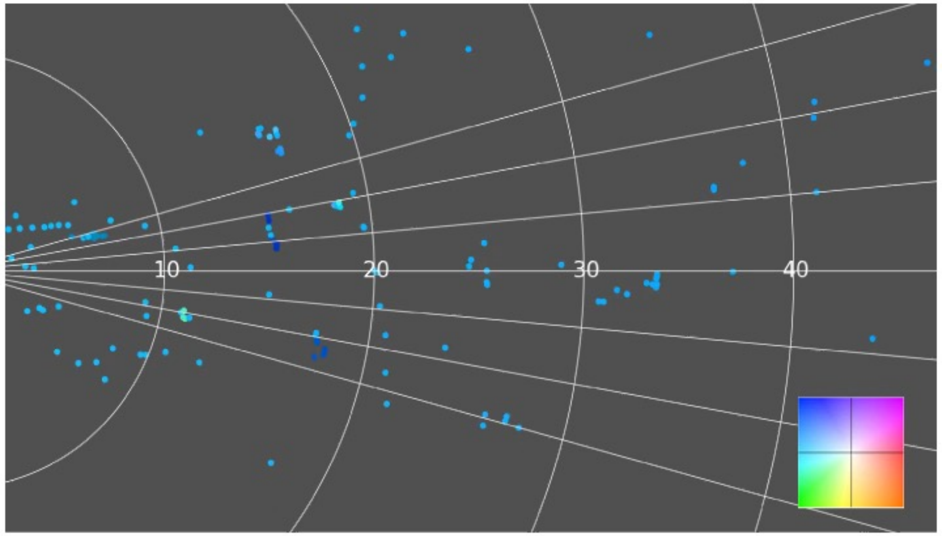


Thanks for watching the quick preview!

Problem definition

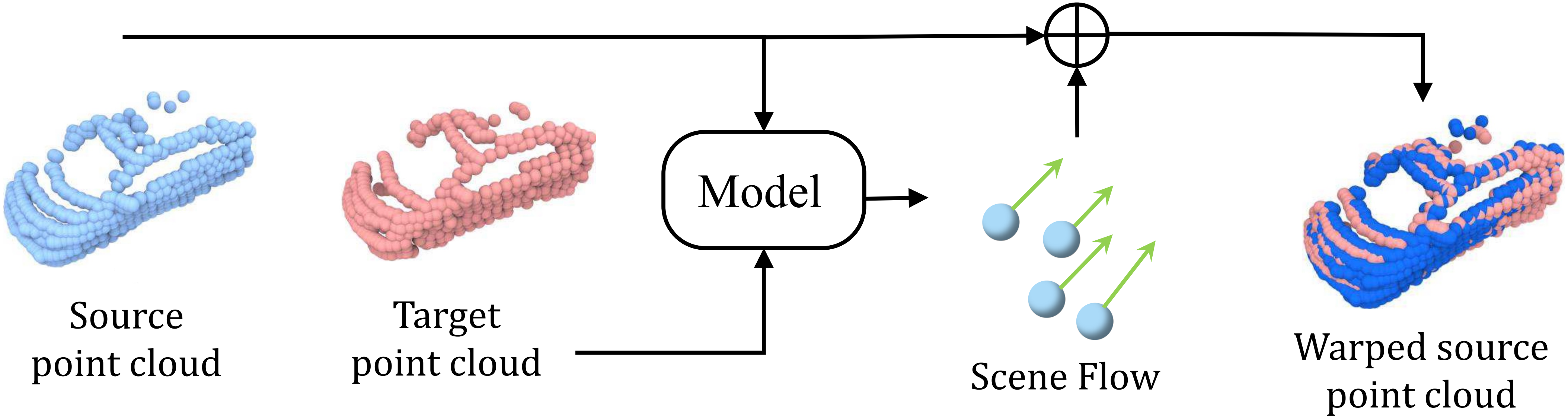
Input (4D radar point clouds)
Perspective view

Output (point-level scene flow)
Bird’s eye view



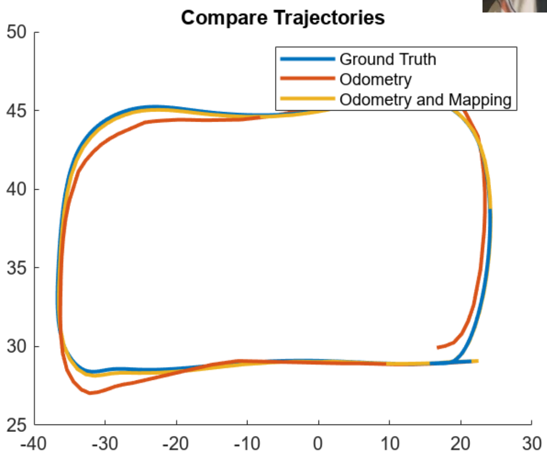
Given consecutive point clouds from 4D radar, we learn to estimate point-level scene flow.

Point Cloud Scene Flow



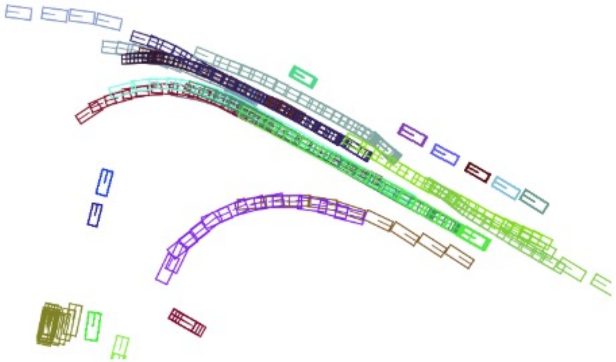
- Represent the 3D inter-frame displacement of each source point
- Induced by the motion of both the ego-vehicle and ambient objects

Downstream tasks

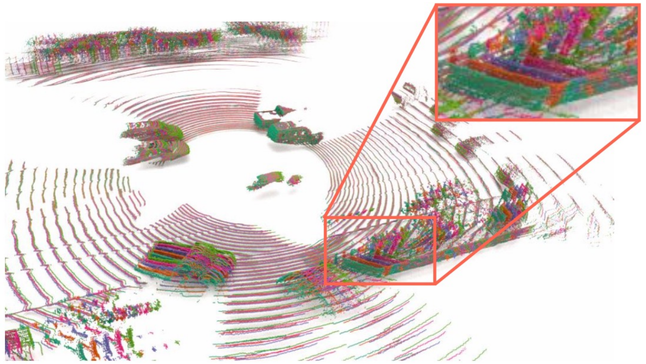


Ego-motion estimation

Motion segmentation



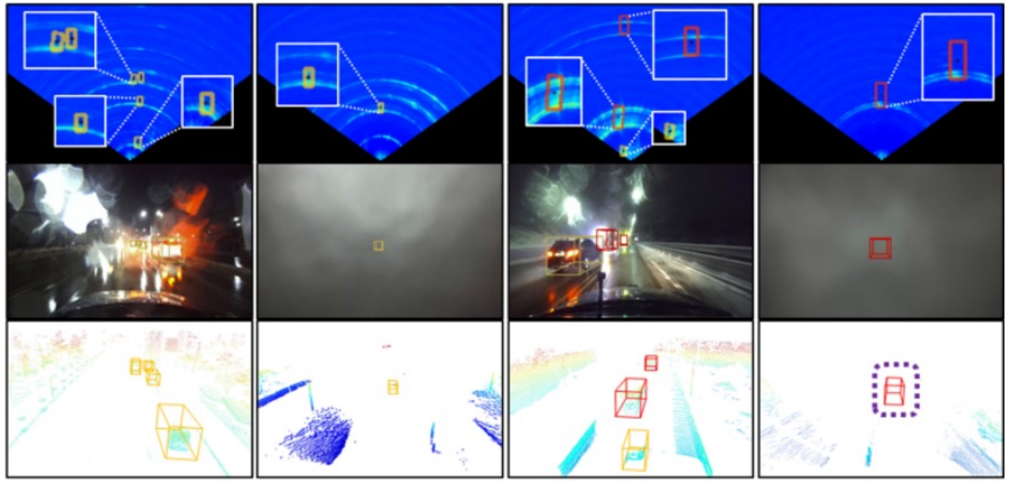
Multi-object tracking



Point cloud accumulation

4D Automotive Radar

- **Emerging** sensor technology in the automotive industry
- **Robust** to adverse weather and poor illumination conditions
- **4D imaging**: 3D position + 1D doppler velocity measurement
- **Radar-on-a-chip**: low-cost (vs. LiDAR), small size and lightweight



(d) rain (e) sleet (f) light snow (g) heavy snow

K-RADAR DATASET

ARBE 4D RADAR

Challenges

- The acquisition of scene flow annotations are costly. In literature, there is a trade-off between annotation efforts and model performance.

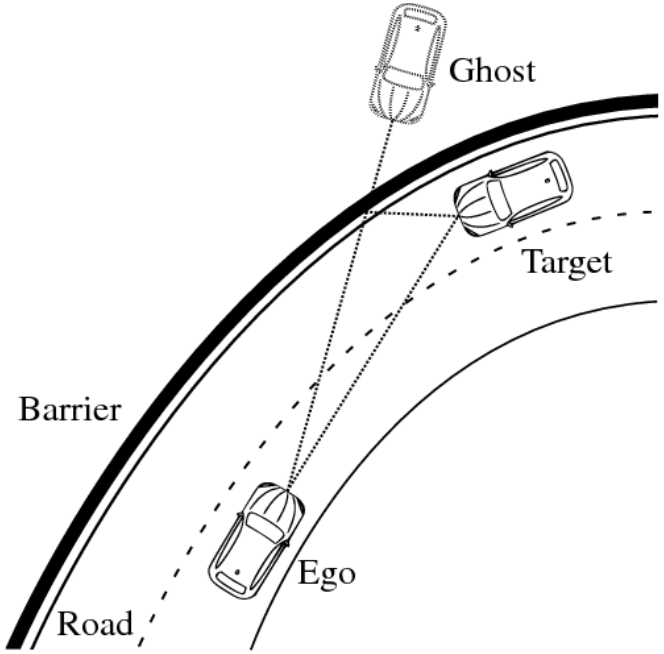
Strategy	Methods	Supervision	Annotation efforts	performance
Self-supervised	JGWTF, SLIM, RaFlow	None	None	low
Weakly-supervised	WsRSF, Dong et al.	GT BG/FG mask	medium	medium
Fully-supervised	FLOT, FlowStep3D	GT Scene flow	high	high

How to overcome such trade-off, i.e. getting a high performance with low or no annotation efforts?

Challenges

- Radar point clouds suffers from sparsity and noise, which further complicate the scene flow annotation and makes self-supervised based methods unfeasible.

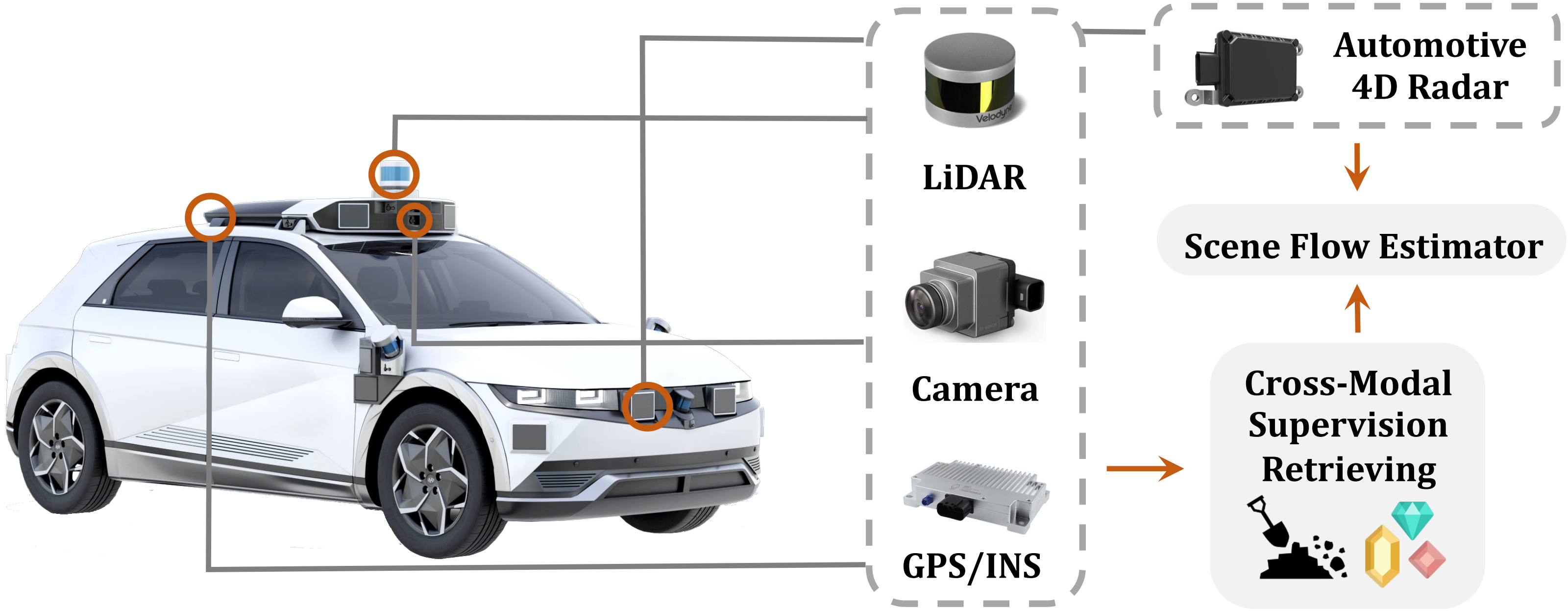
LiDAR vs. RADAR



MULTI-PATH EFFECT

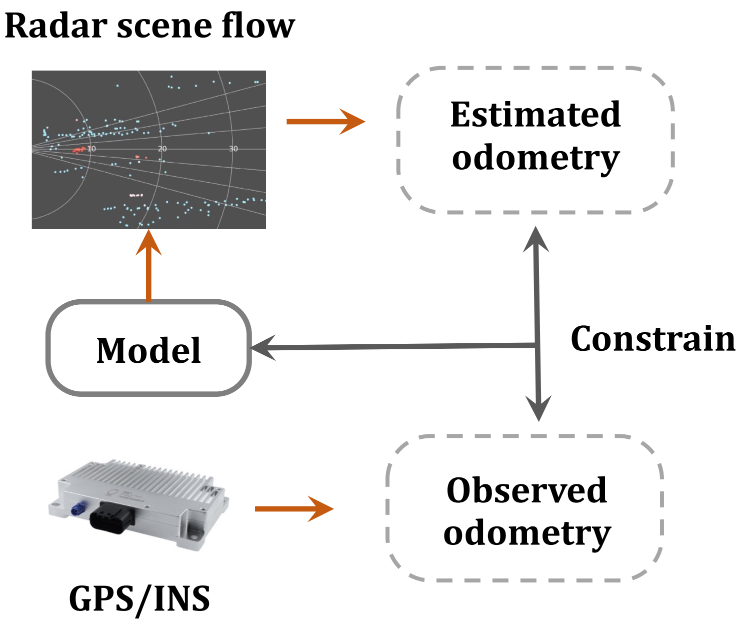
Motivation

- **Fact**: self-driving cars today are equipped with heterogeneous sensors.
- **Insight**: such co-located perception redundancy can be used to provide supervision cues that bootstrap 4D radar scene flow learning.

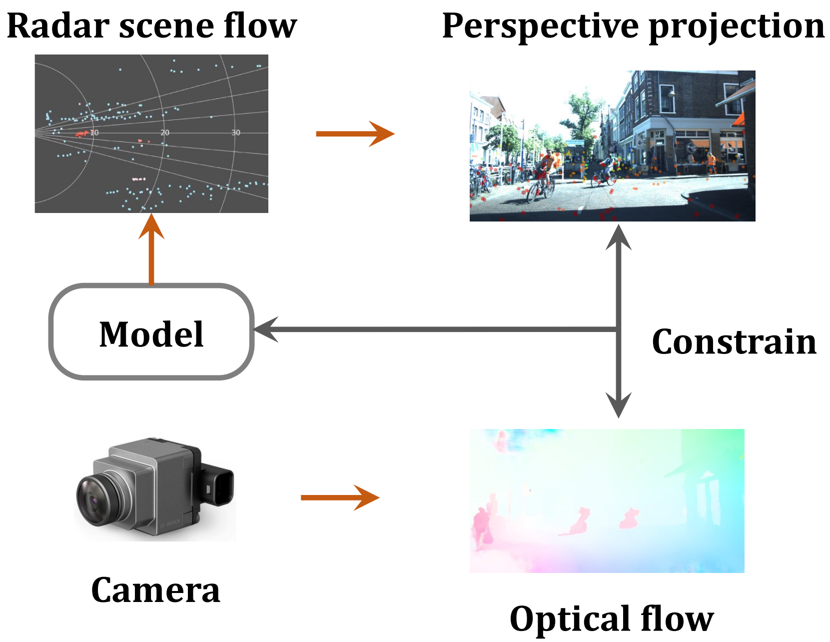


Motivation

- Example: odometry consistency

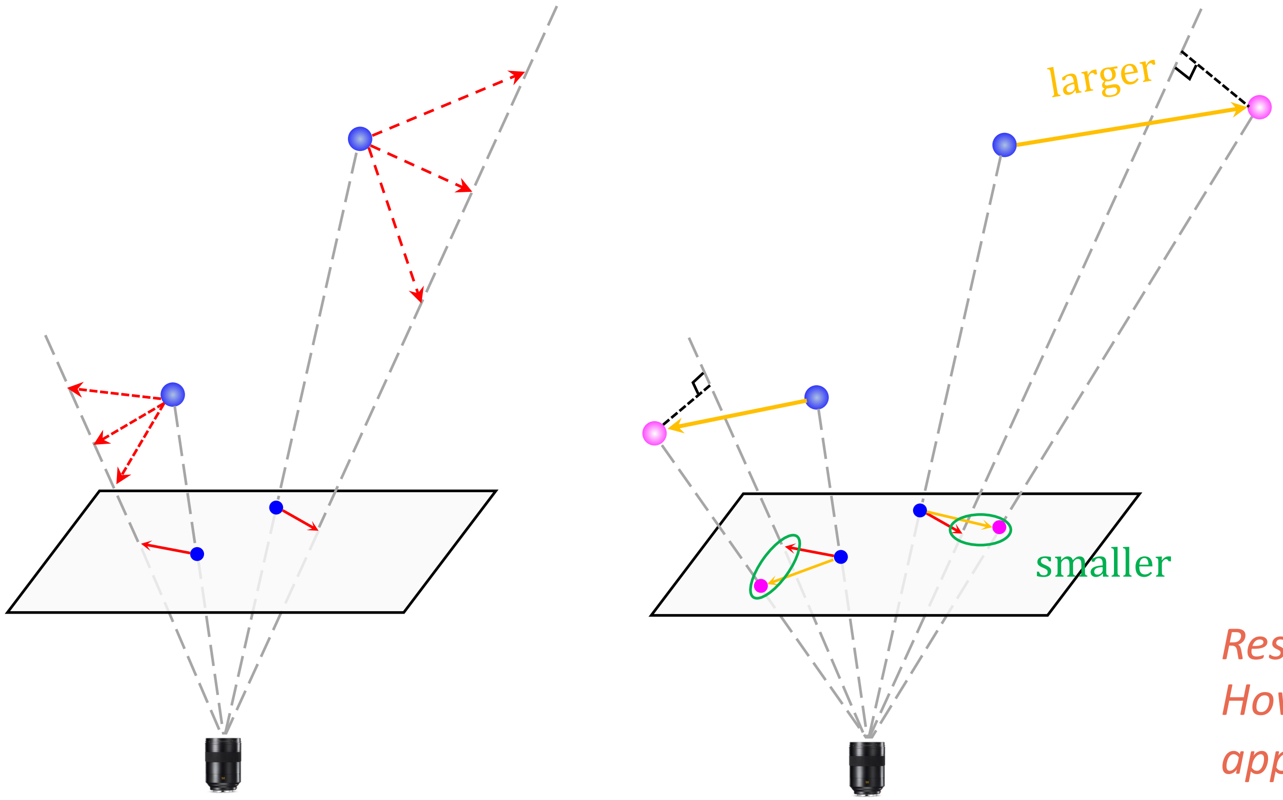


- Example: perspective consistency



Motivation

- Retrieving accurate supervision signals from co-located sensors and effectively use them are non-trivial. For example:



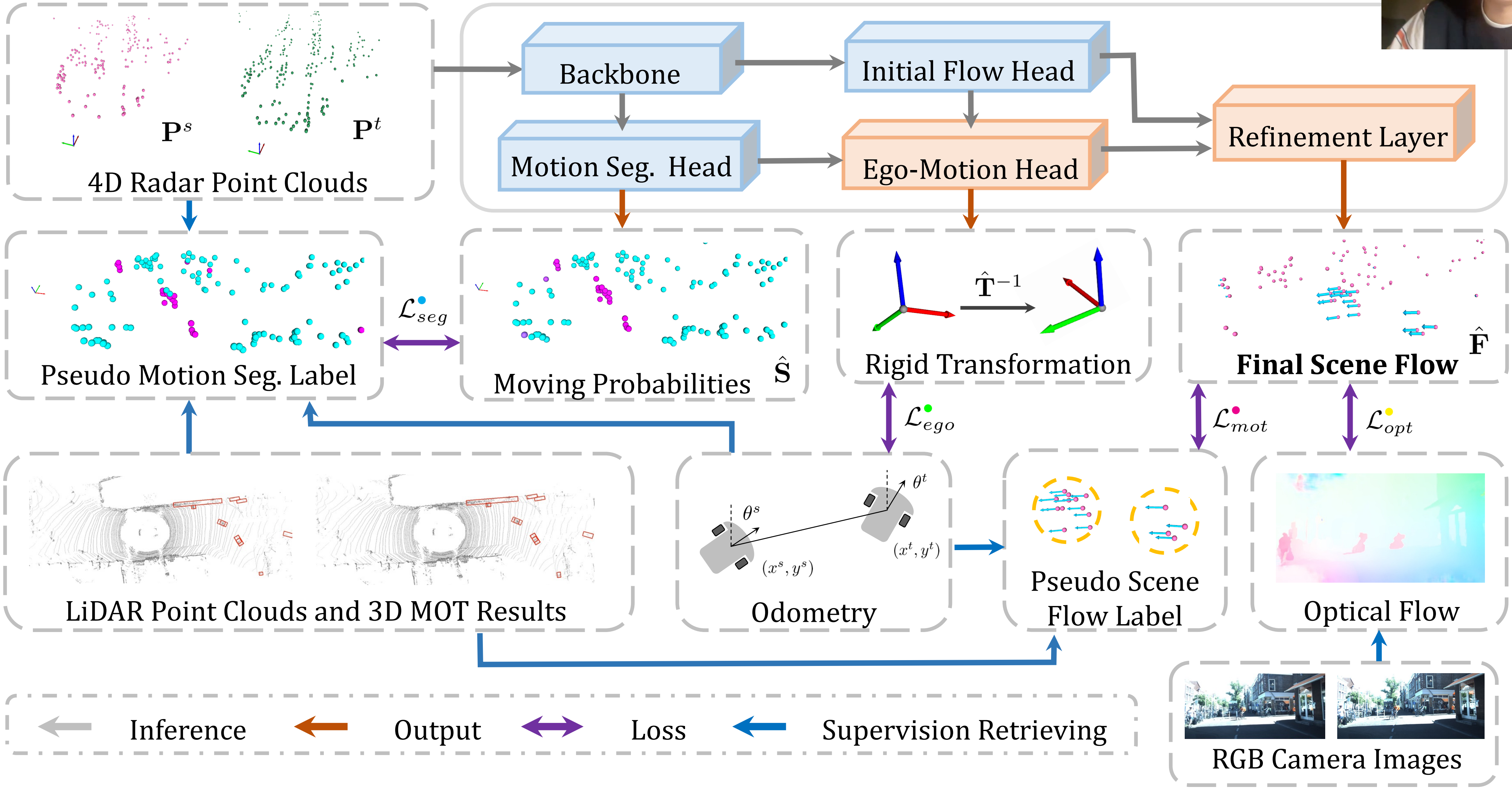
Depth-unaware perspective projection potentially incurs weaker constraints to the scene flow of far points.

Research Question:
How to retrieve useful cross-modal supervision cues and apply them to bootstrap 4D radar scene flow learning?

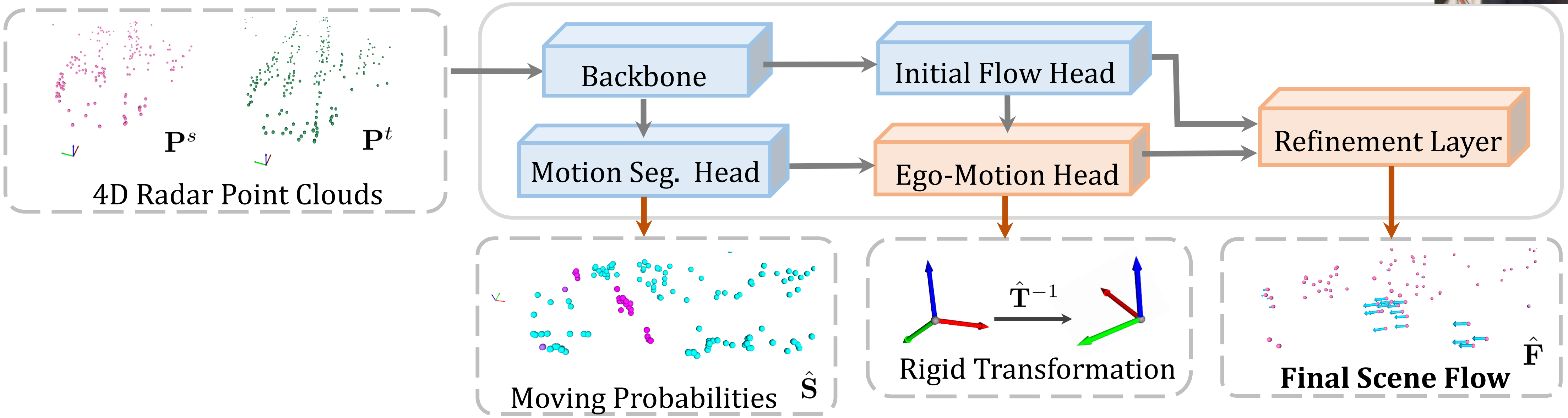
Contribution

- **The first** 4D radar scene flow learning using cross-modal supervision from co-located heterogeneous sensors on an autonomous vehicle.
- **A pipeline** that consists of a multi-task model architecture and loss functions to using multiple cross-modal constraints for model training.
- **State-of-the-art** performance of the proposed CMFlow method was demonstrated on a public dataset and show its effectiveness in downstream tasks as well.

Cross-modal supervised learning pipeline



Model architecture

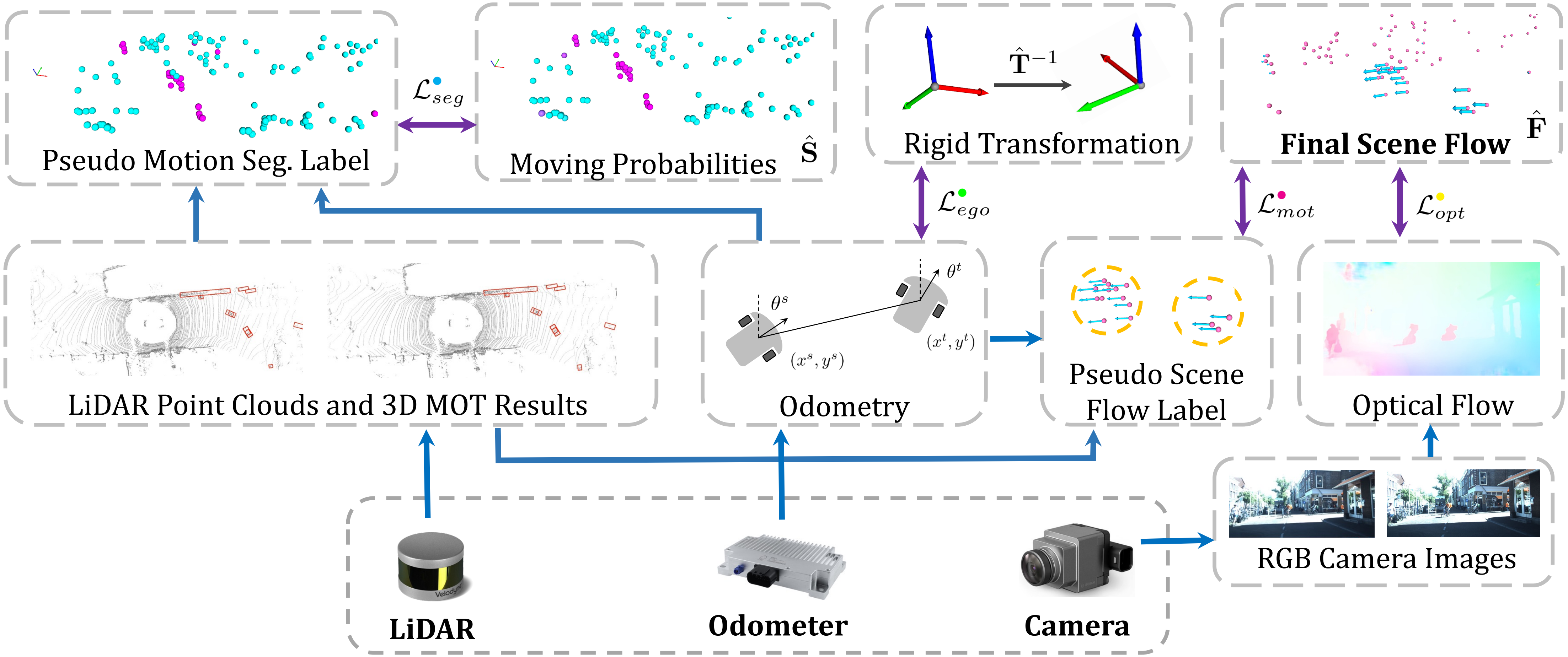


Takeaway:

- Two-stage fashion: blue/orange block colors for stage 1/2
- Multi-task model: scene flow, motion segmentation, ego-motion estimation
- The flow vectors of static points are only caused by the radar's ego-motion, we can regularize them with the more reliable rigid transformation

Cross-modal supervision

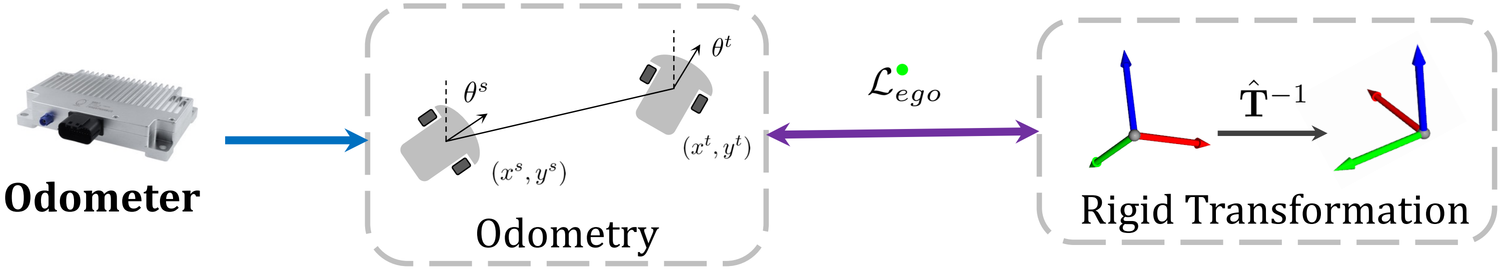
- Overall loss: $\mathcal{L} = \mathcal{L}_{ego} + \mathcal{L}_{seg} + \mathcal{L}_{mot} + \lambda_{opt}\mathcal{L}_{opt}$



Cross-modal supervision

Ego-motion loss:

$$\mathcal{L}_{ego} = \frac{1}{N}\sum_{i=1}^{N}\left\|(\hat{\mathbf{T}}-\mathbf{T})[\mathbf{c}_i^s\ 1]^\top\right\|_2$$



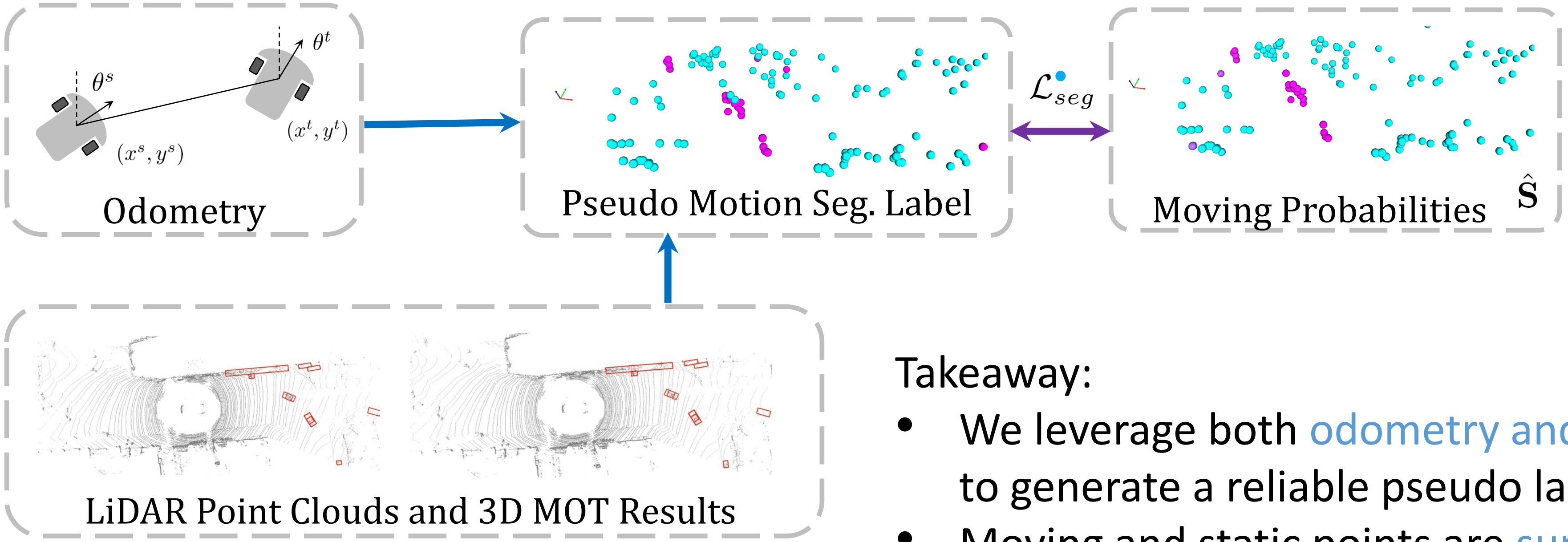
Takeaway:

- The odometry can be used to explicitly supervise the rigid transformation and implicitly constrain the initial and final scene flow output

Cross-modal supervision

Motion segmentation loss:

$$\mathcal{L}_{seg} = \frac{1}{2}\left(\frac{\sum_{i=1}^{N}(1-s_i)\log(1-\hat{s}_i)}{\sum_{i=1}^{N}(1-s_i)} + \frac{\sum_{i=1}^{N} s_i \log(\hat{s}_i)}{\sum_{i=1}^{N} s_i}\right).$$



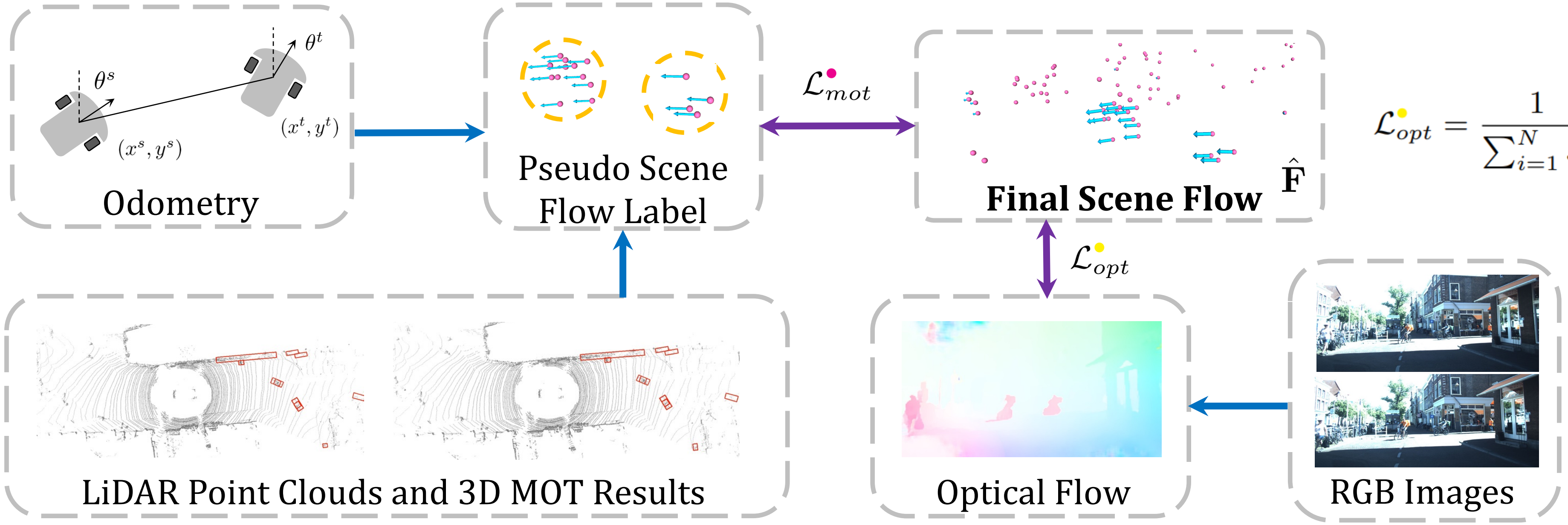
Takeaway:

- We leverage both odometry and LiDAR 3D MOT results to generate a reliable pseudo label.
- Moving and static points are supervised separately to balance their impact.

Cross-modal supervision

Scene flow loss:

$$\mathcal{L}_{mot} = \frac{1}{\sum_{i=1}^{N} s_i^l} \sum_{i=1}^{N} \left\| s_i^l (\hat{\mathbf{f}}_i - \mathbf{f}_i^{fg}) \right\|_2$$



$$\mathcal{L}_{opt} = \frac{1}{\sum_{i=1}^{N} s_i} \sum_{i=1}^{N} s_i \mathcal{D}(\mathbf{c}_i^s + \hat{\mathbf{f}}_i, \mathbf{m}_i + \mathbf{w}_i, \theta)$$

Takeaway:

- We supervise foreground points scene flow with LiDAR 3D MOT Results
- In the optical loss, we take the point-to-ray distance as the training objective, which is more insensitive to points at different ranges.

Main results

Method	Sup.	EPE [m]↓	AccS↑	AccR↑	RNE [m]↓	MRNE [m]↓	SRNE [m]↓
ICP [4]	None	0.344	0.019	0.106	0.138	0.148	0.137
Graph Prior* [33]	None	0.445	0.070	0.104	0.179	0.186	0.176
JGWTF* [31]	Self	0.375	0.022	0.103	0.150	0.139	0.151
PointPWC [52]	Self	0.422	0.026	0.113	0.169	0.154	0.170
FlowStep3D [21]	Self	0.292	0.034	0.161	0.117	0.130	0.115
SLIM* [2]	Self	0.323	0.050	0.170	0.130	0.151	0.126
RaFlow [9]	Self	0.226	0.190	0.390	0.090	0.114	0.087
CMFlow	Cross	0.141	**0.233**	0.499	0.057	0.073	0.054
CMFlow (T)	Cross	**0.130**	0.228	**0.539**	**0.052**	**0.072**	**0.049**

Takeaway:

- The state-of-the-art performance compared with baselines that also demand no annotation efforts
- The performance is further improved when applying the temporal information (i.e., T)

Breakdown results

	O	L	C	EPE [m]↓	AccS↑	AccR↑	RNE [m]↓
(a)				0.228	0.184	0.392	0.091
(b)	✓			0.161	0.203	0.442	0.065
(c)	✓	✓		0.145	0.228	0.482	0.058
(d)	✓		✓	0.159	0.216	0.458	0.064
(e)	✓	✓	✓	**0.141**	**0.233**	**0.499**	**0.057**

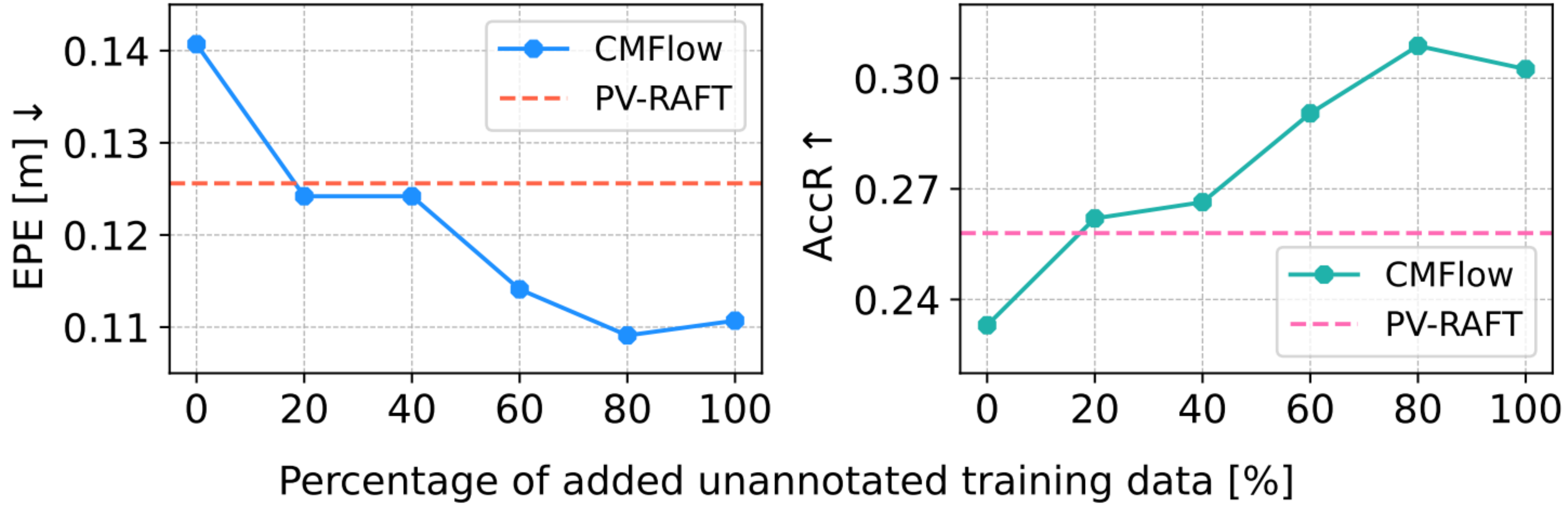
	L (seg)	L (flow)	EPE [m]↓	AccS↑	AccR↑	RNE [m]↓
(a)			0.159	0.216	0.458	0.064
(b)	✓		0.156	0.221	0.467	0.063
(c)		✓	0.152	0.217	0.477	0.061
(d)	✓	✓	**0.141**	**0.223**	**0.499**	**0.057**

Illustration of the causes of noisy supervision

Takeaway:

- All modalities contribute to our method, and the odometer leads to the biggest performance gain.
- Due to their noisy labels, the gains brought by camera and LiDAR are smaller than that of odometer.

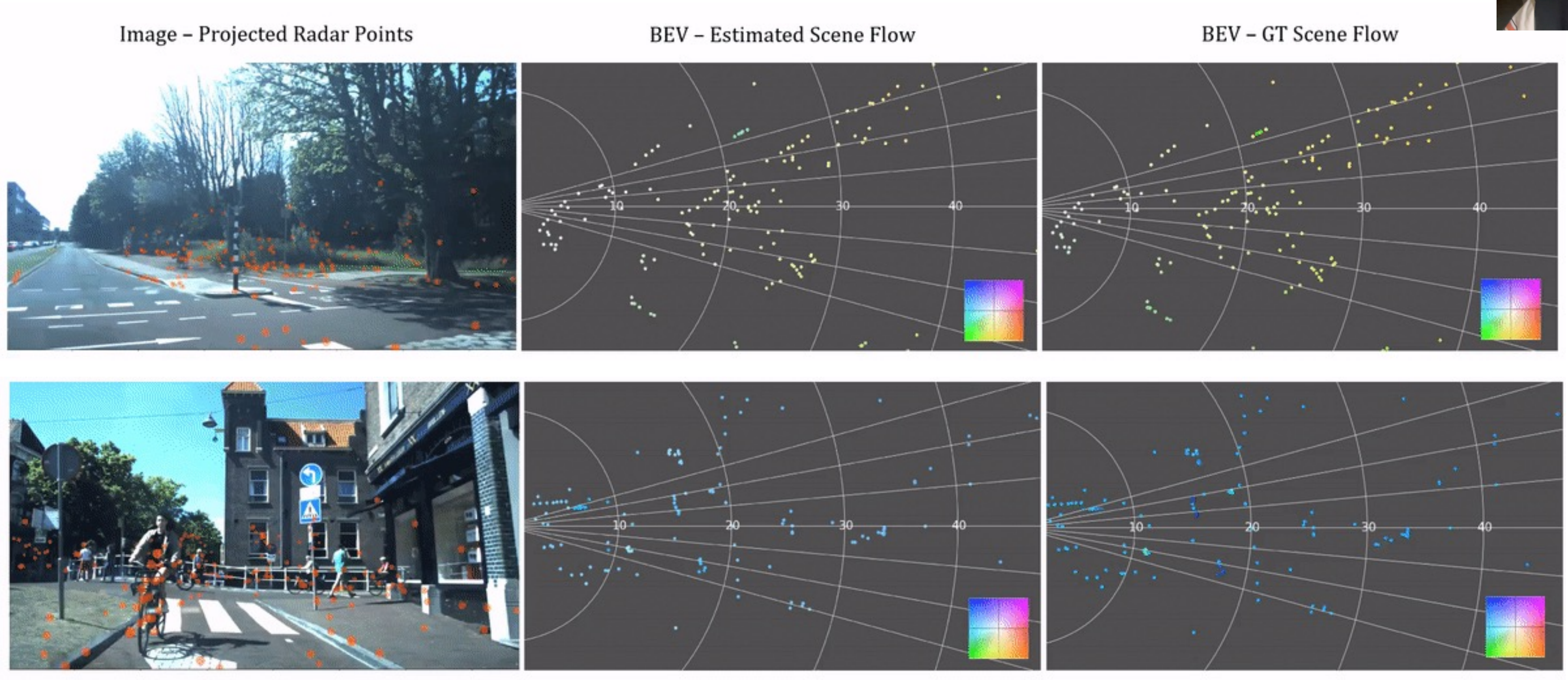
Impact of the amount of unannotated data



Takeaway:

- The performance of CMFlow improves by a large margin by using extra unannotated training data.
- After adding only 20% extra samples, CMFlow can already outperform PV-RAFT trained with less annotated samples.

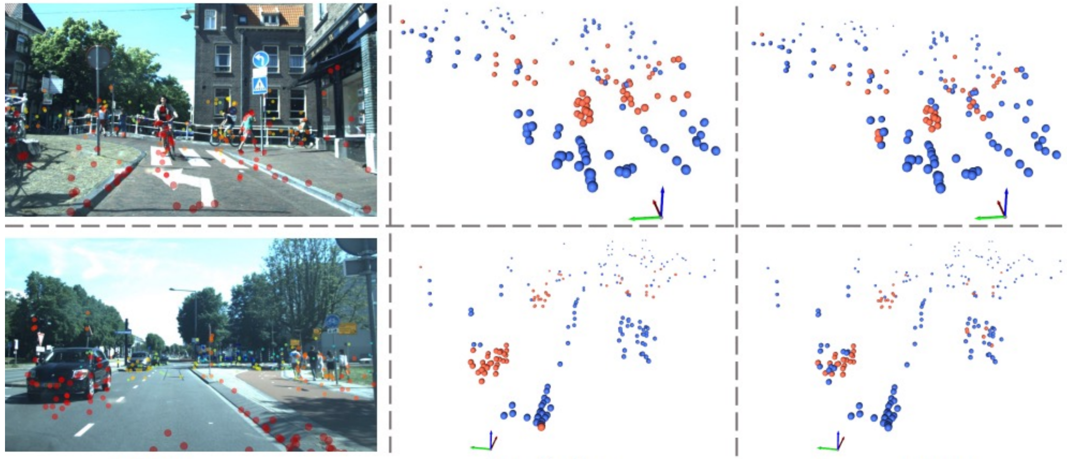
Scene flow demo



Color of points in the BEV image represents the magnitude and direction of scene flow vectors.

Subtask – motion segmentation evaluation

	Label $\mathbf{S}^v$	Label $\mathbf{S}^l$	A.D.	mIoU (%)	Gain (%)
(a)				46.9	-
(b)	✓			52.8	+5.9
(c)	✓	✓		54.1	+1.3
(d)	✓	✓	✓	57.1	+3.0

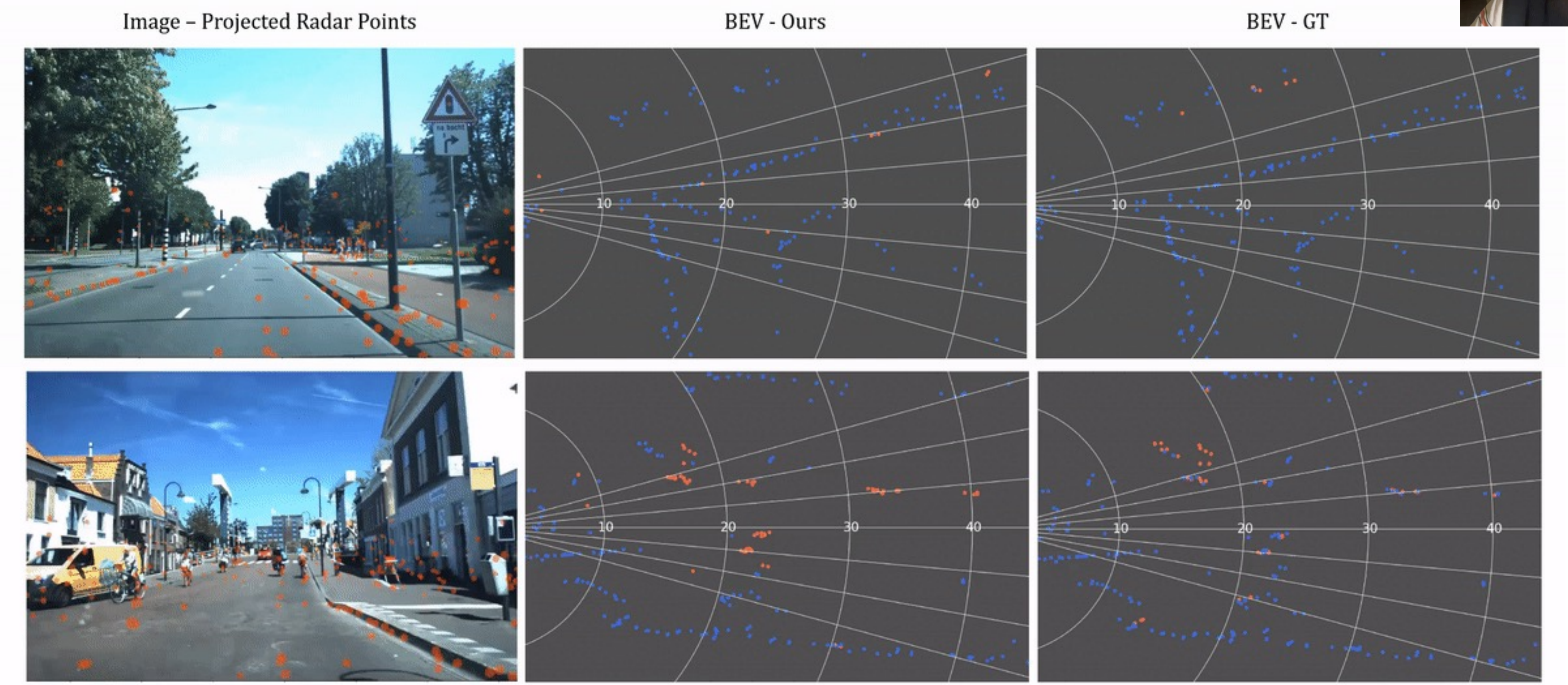


RGB Image Prediction Ground Truth

Takeaway:

- Two ingredients of the pseudo motion segmentation label contributes to our performance improvement on motion segmentation.

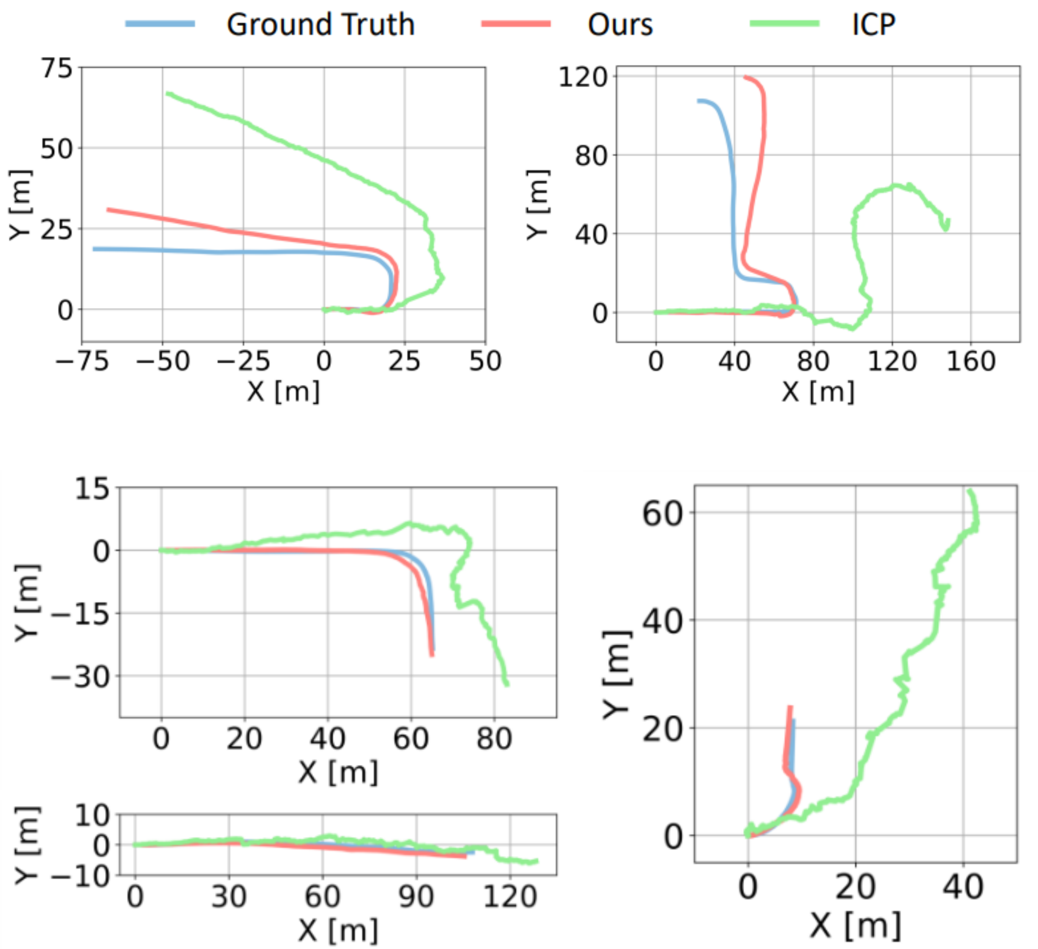
Motion segmentation demo



In the BEV images, blue/orange denotes static and moving points respectively.

Subtask – ego-motion estimation

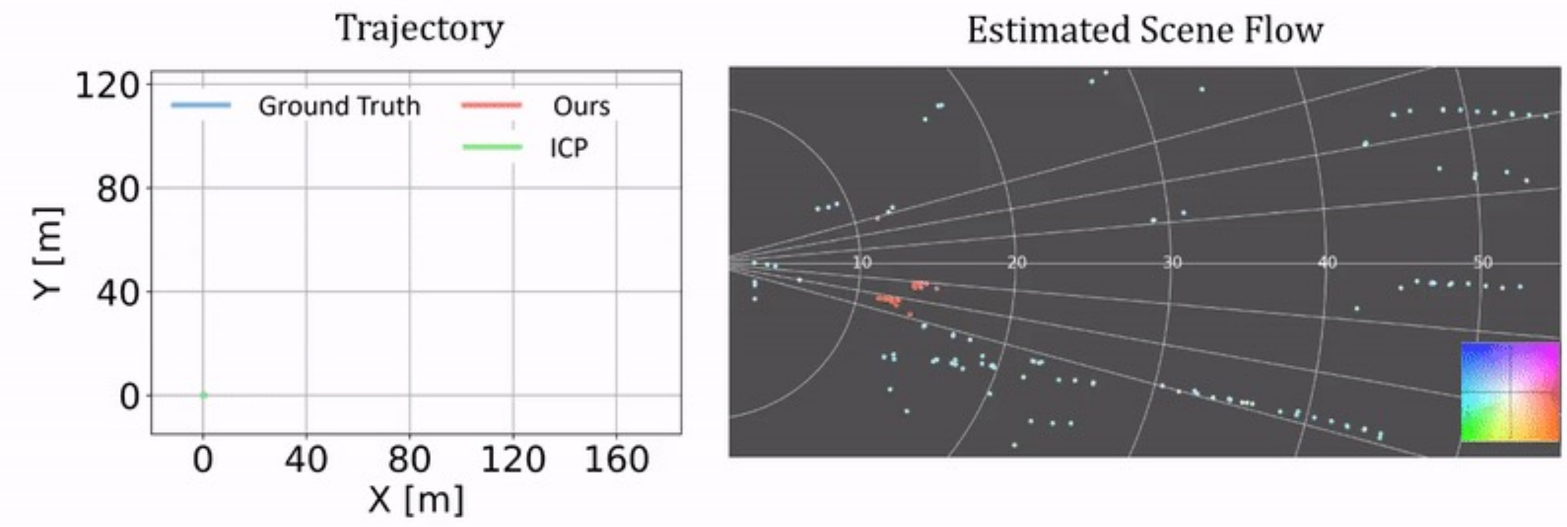
	O	L + C	A.D.	T	RTE [m]	RAE [°]
(a)					0.090	0.336
(b)	✓				0.086	0.183
(c)	✓	✓			0.085	0.145
(d)	✓	✓	✓		0.071	**0.089**
(e)	✓	✓	✓	✓	**0.066**	0.090

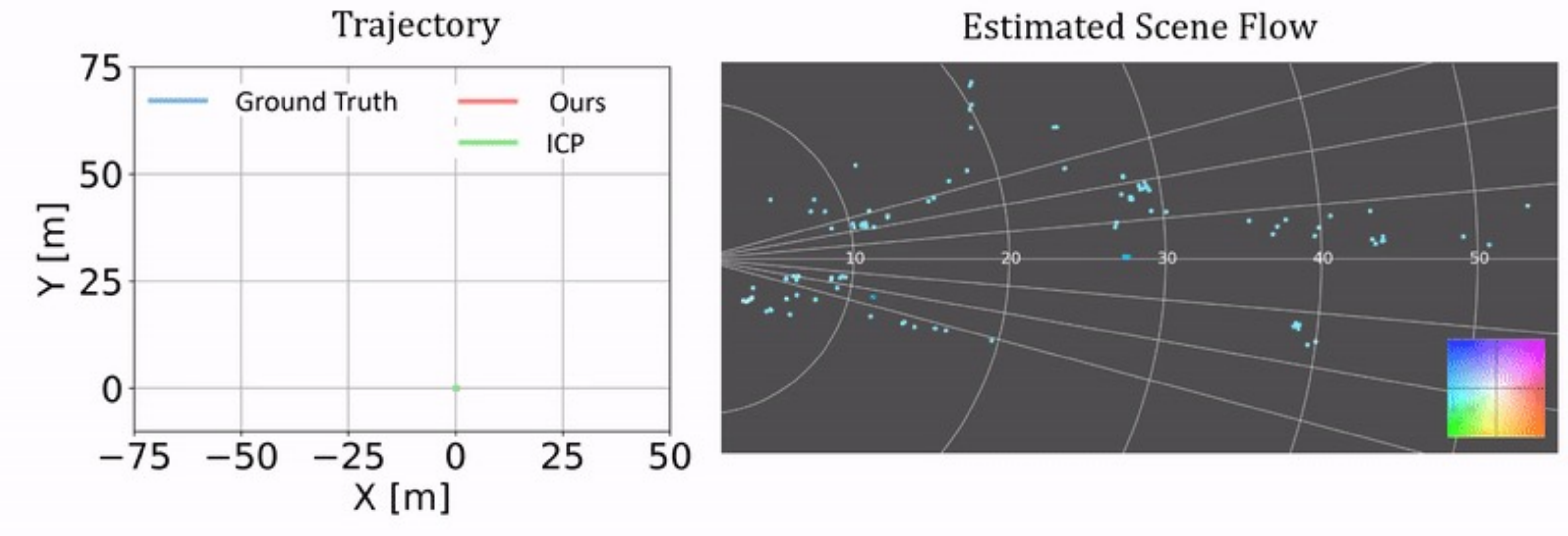


Takeaway:

- Both odometer and LiDAR/camera contribute to our ego-motion estimation results
- By accumulating inter-frame ego-motion, our method can support the long-term odometry.

Ego-motion demo





Thanks for watching the presentation!

Code

Paper

Demo

Page

Hidden Gems: 4D Radar Scene Flow Learning Using Cross-Modal Supervision

Fangqiang Ding, Andras Palffy, Dariu M. Gavrila, Chris Xiaoxuan Lu

WED-AM-106 (Highlight)

THE UNIVERSITY of EDINBURGH

TU Delft

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Code Available