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

THE UNIVERSITY of EDINBURGH · TU Delft

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

Contact: fding@ed.ac.uk

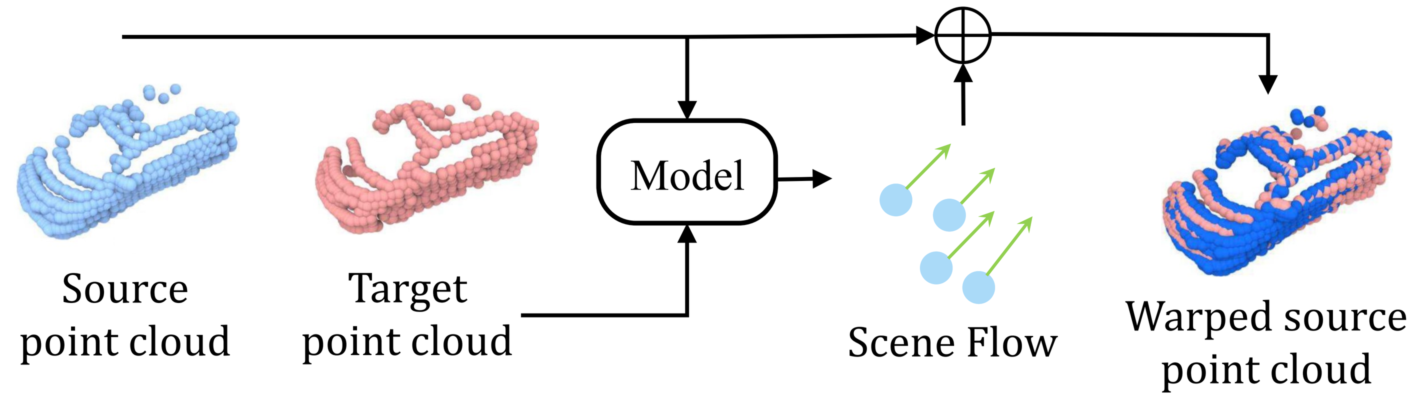
JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Code Available

What is Scene Flow and 4D Radar?

Scene Flow on Point Cloud

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



4D Automotive Radar

- **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

Challenges and Motivation

Challenge: trade-off (annotation efforts and performance)

Strategy	Supervision	Annotation efforts	performance
Self-supervised	None	None	low
Weakly-supervised	GT BG/FG mask	medium	medium
Fully-supervised	GT Scene flow	high	high

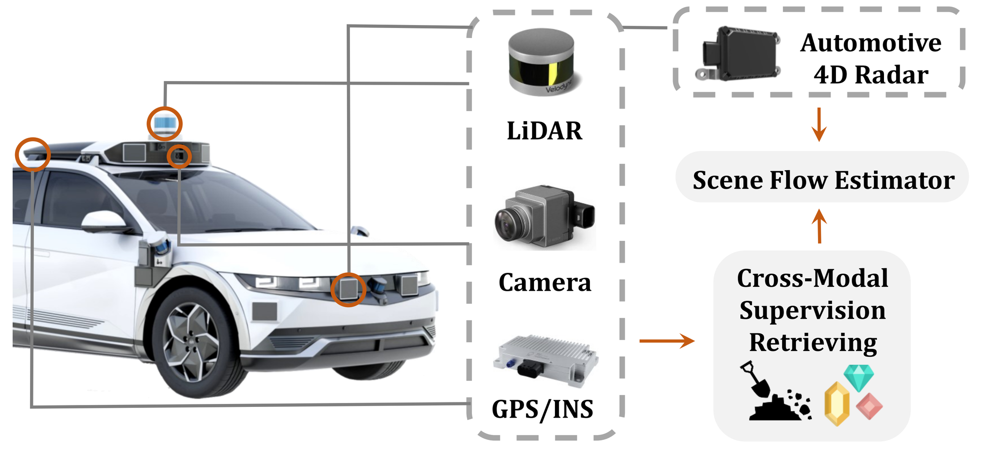
How to overcome such trade-off, i.e. get high performance with low efforts?

Challenge: radar characteristic (sparsity and noise)

- **Sparse** point cloud: average ~350 points per frame (<1% of LiDAR)
- **Noisy** (ghost) points due to multi-path effects of millimeter-wave
- **Result:** complicate the point-wise scene flow manual annotation; hard to exclusively rely on self-supervision for performance and safety.

Motivation

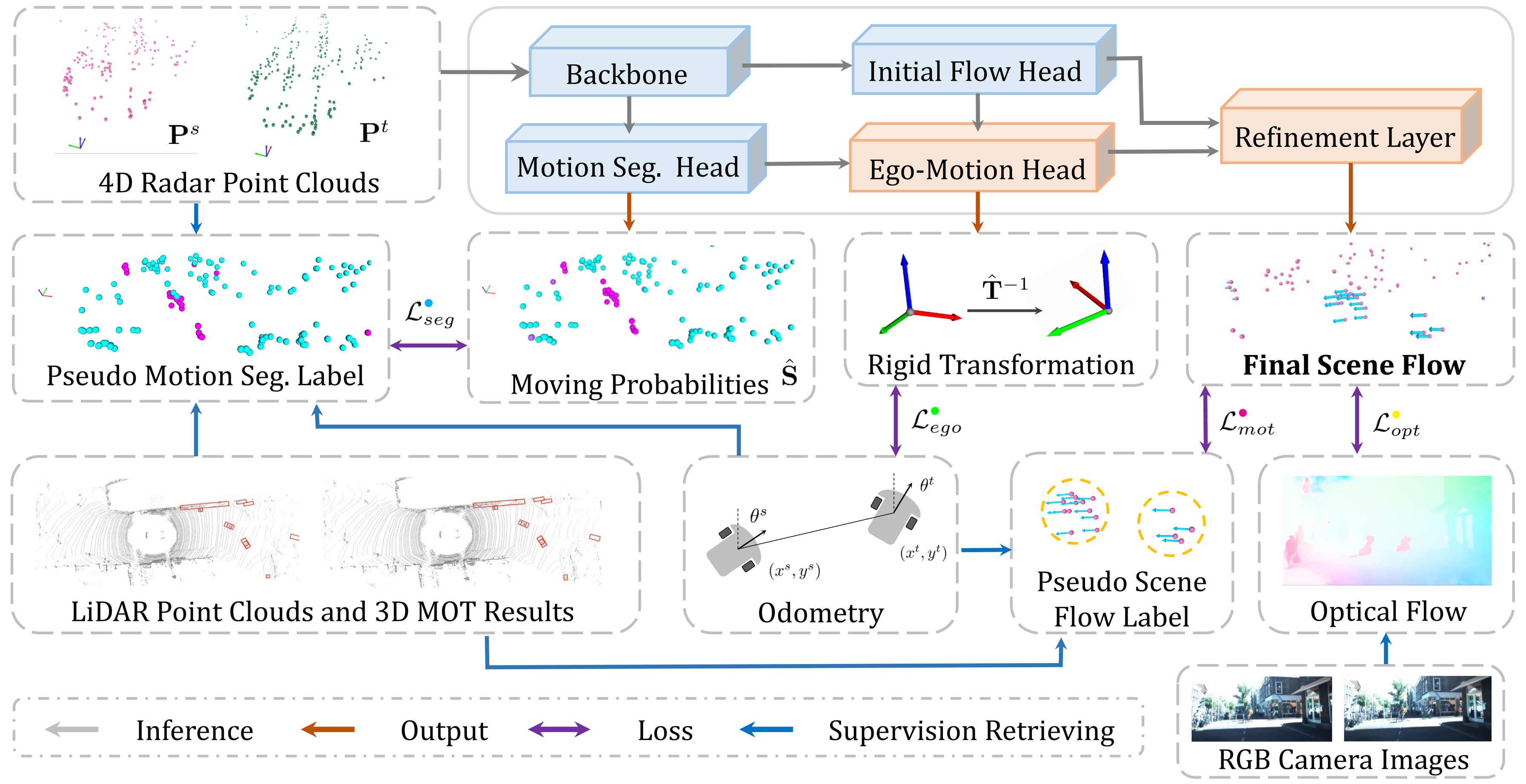
- **Fact**: self-driving cars today are equipped with heterogeneous sensors.
- **Insight**: such co-located perception redundancy can be used to provide supervision cues that bootstrap radar scene flow learning.



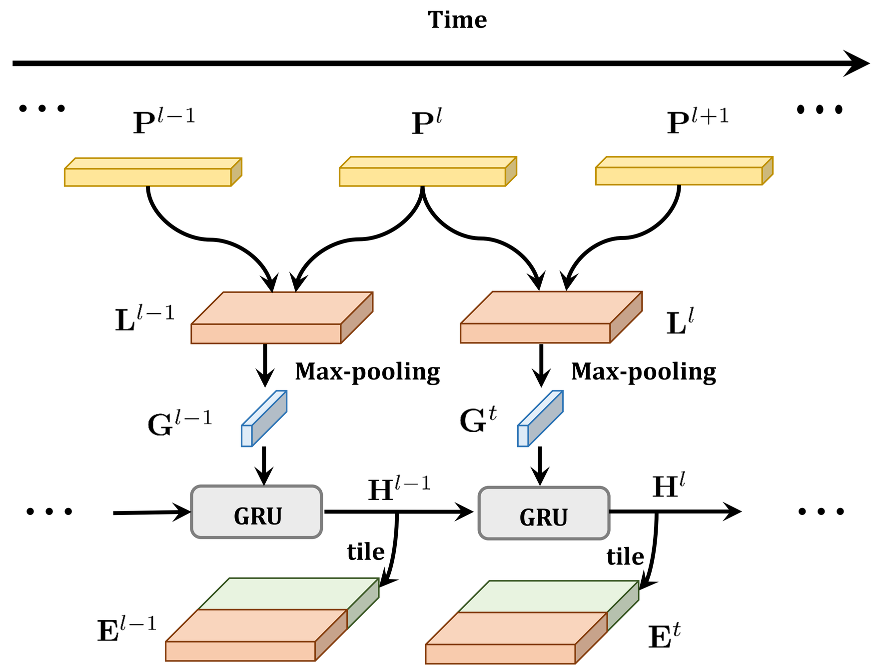
How to retrieve useful cross-modal supervision cues and apply them to bootstrap radar scene flow learning?

Proposed Method

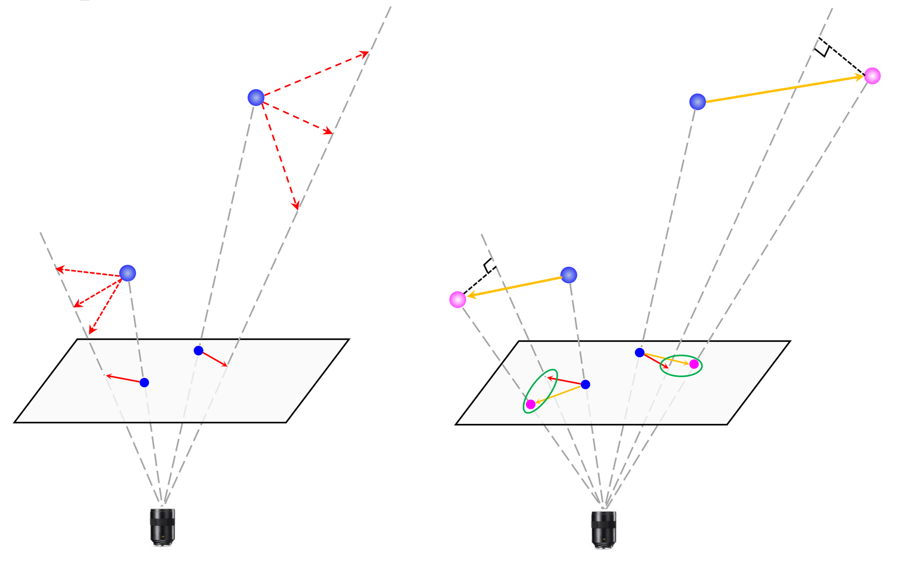
- **Cross-modal supervised learning pipeline:**
  a) end-to-end two stages (blue/orange block colors for stage 1/2) model.
  b) multi-task problem: scene flow, motion segmentation, ego-motion transform.
  c) no need for pretraining any models, no need for annotating any labels manually.



- **Temporal update module** embedded in backbone: propagate previous latent global information to the current frame



- Optical flow loss $\mathcal{L}_{opt}$: use **point-to-ray** distance instead of flow divergence in the pixel scale. See motivation below:



Due to **perspective projection**: a larger point-to-ray distance leads to a low flow error on image for a farther point.

Scene Flow Evaluation

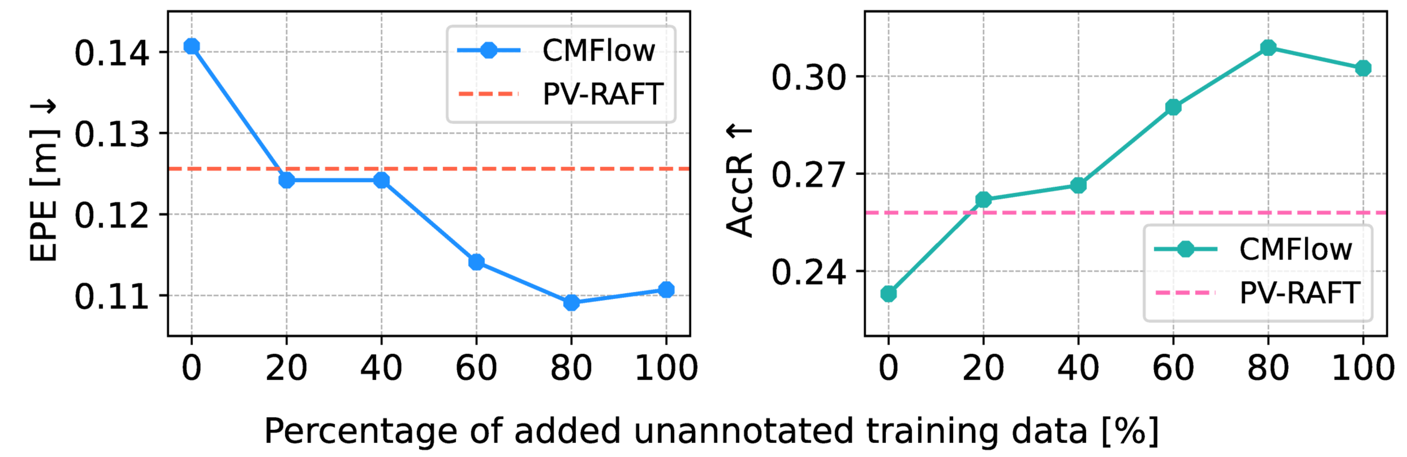
- **State-of-the-art comparison:** compare **CMFlow** (T denotes temporal update) with baselines that also demand no annotation efforts on the *View-of-Delft* dataset.

Method	EPE [m]	AccR	Method	EPE [m]	AccR
Graph Prior	0.445	0.104	SLIM	0.323	0.170
JGWTF	0.375	0.103	RaFlow	0.226	0.390
PointPWC	0.422	0.113	**CMFlow**	**0.141**	**0.499**
FlowStep3D	0.292	0.161	**CMFlow (T)**	**0.130**	**0.539**

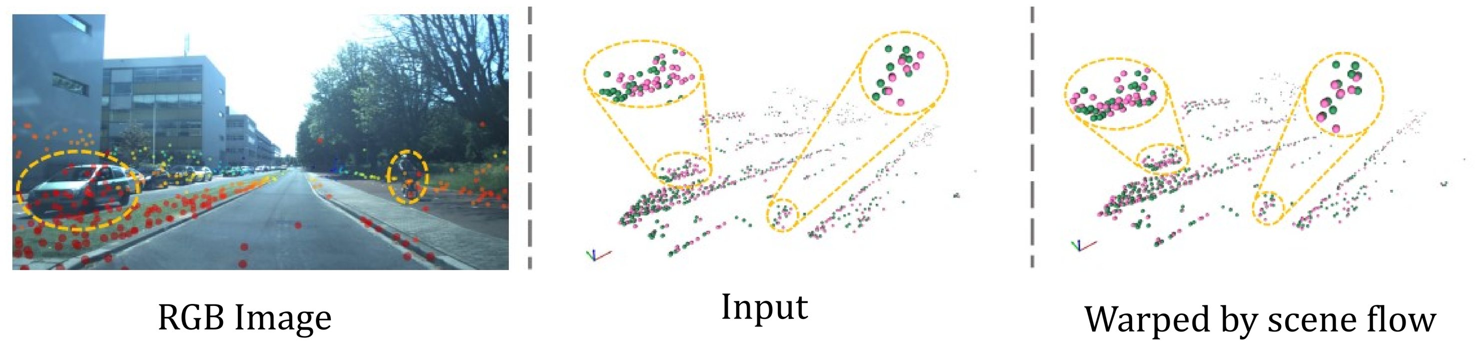
- **Breakdown results:** combine cross-modal supervision signals from different modalities. Abbrev.: odometer (O), LiDAR (L), camera (C).

	O	L	C	EPE [m]	AccS	AccR
(a)				0.228	0.184	0.392
(b)	√			0.161	0.203	0.442
(c)	√	√		0.145	0.228	0.482
(d)	√		√	0.159	0.216	0.458
(e)	√	√	√	**0.141**	**0.233**	**0.499**

- **Impact of the amount of unannotated training data.** PV-RAFT is fully-supervised trained with limited annotated samples.

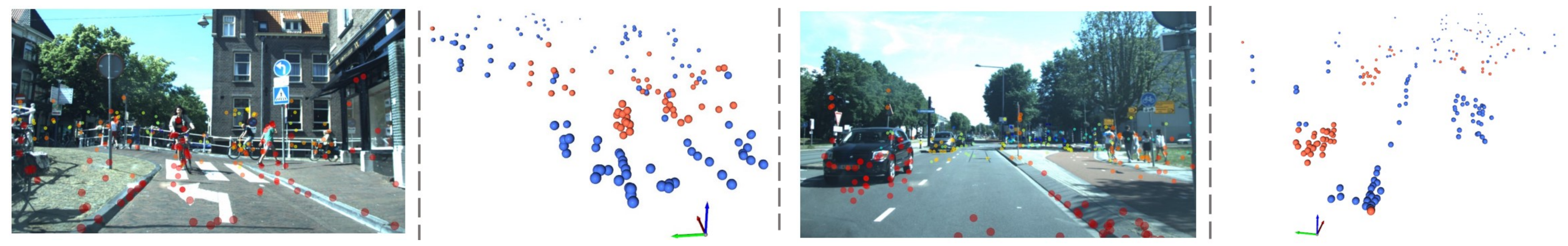


- **Qualitative results:** pink/green for source/target point cloud



Sub-Task Evaluation

- **Motion segmentation:** orange/blue color for moving/static points



- **Ego-motion estimation:** We accumulate our inter-frame ego-motion transform estimations and plot the long-term trajectories.