Self-Supervised Understanding of Dynamic Scenes

Let Data Be the Teacher

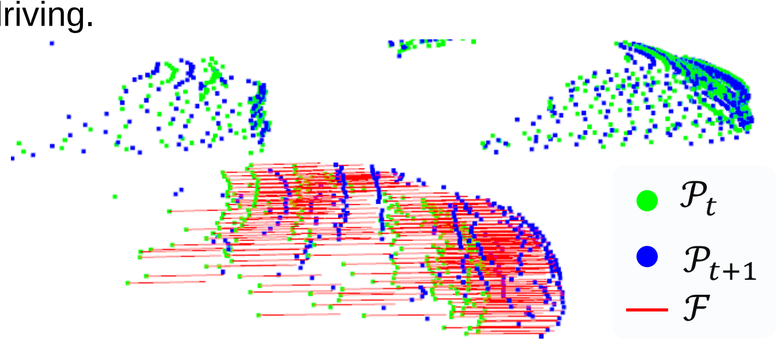
Qingwen Zhang (Supervisor: Patric Jensfelt, Olov Andersson)

KTH Royal Institute of Technology, EECS, RPL

Scan me
Personal Website

1. Introduction & Motivation

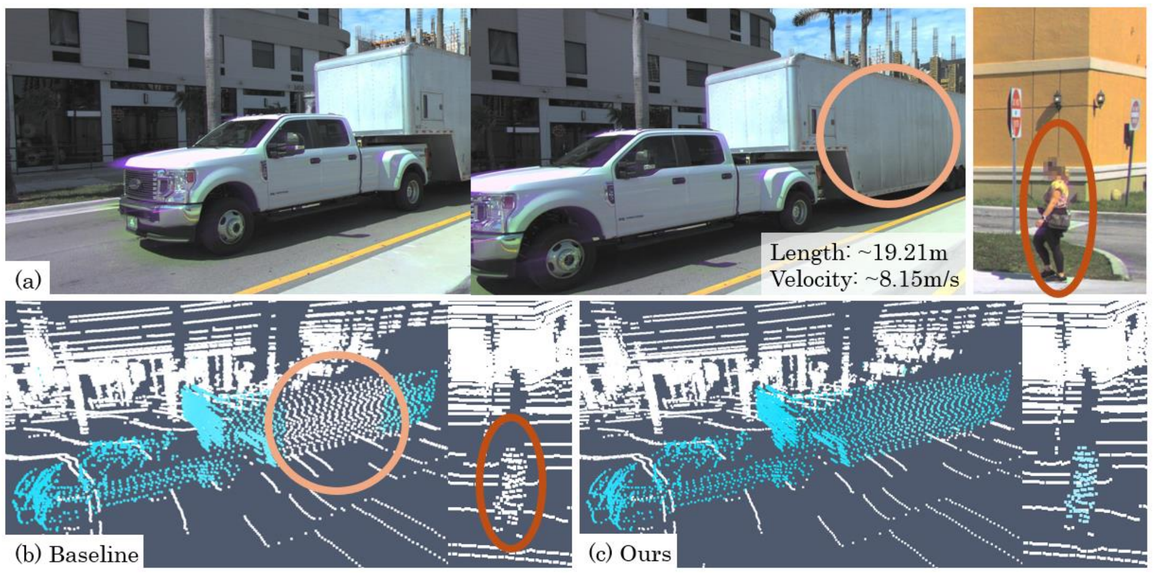
Understanding the surrounding world is key to many emerging robotics applications, currently perhaps most notably in autonomous vehicles. The world is dynamic and pre-computed maps and plans are seldom viable.
Scene flow estimation determines a scene's 3D motion field, by predicting **the motion of points** in the scene, especially for aiding tasks in autonomous driving.



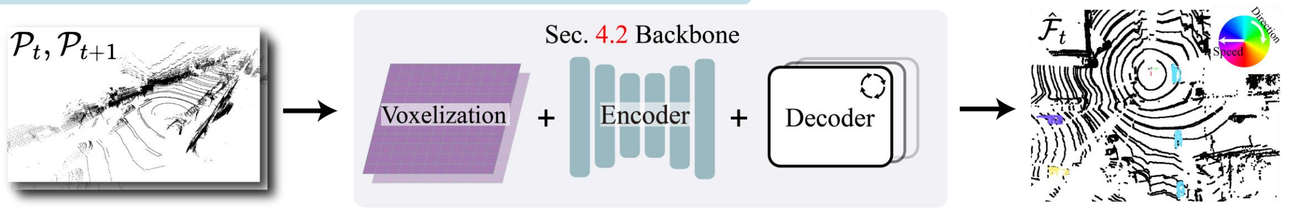
A common paradigm for addressing the scene flow problem is supervised learning by utilizing annotated LiDAR data. However, **expensive labeling inherently** limits the scalability of supervised learning methods**.**

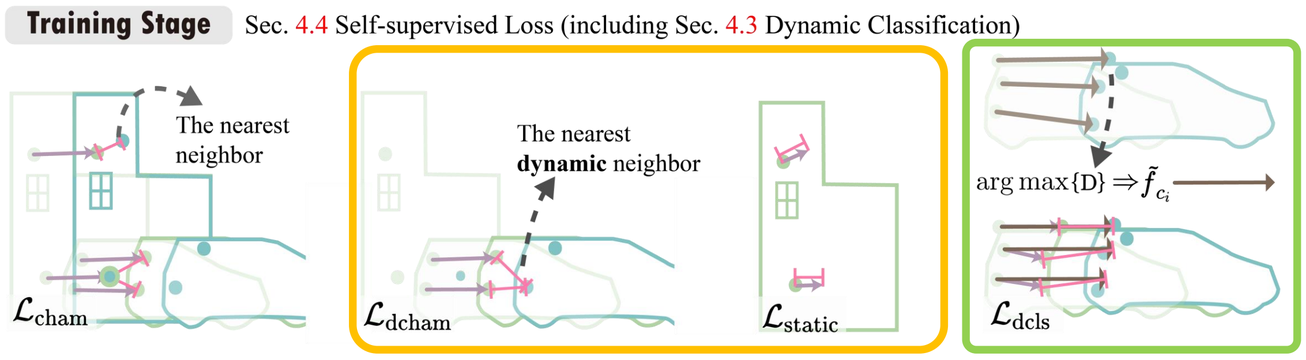
Existing **self-supervised** methods suffer from:

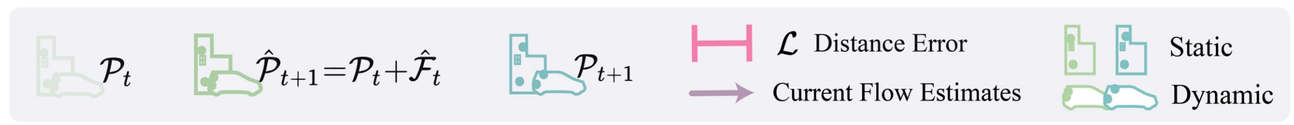
Point data imbalance | Disregard Object-level motion | Running efficiency



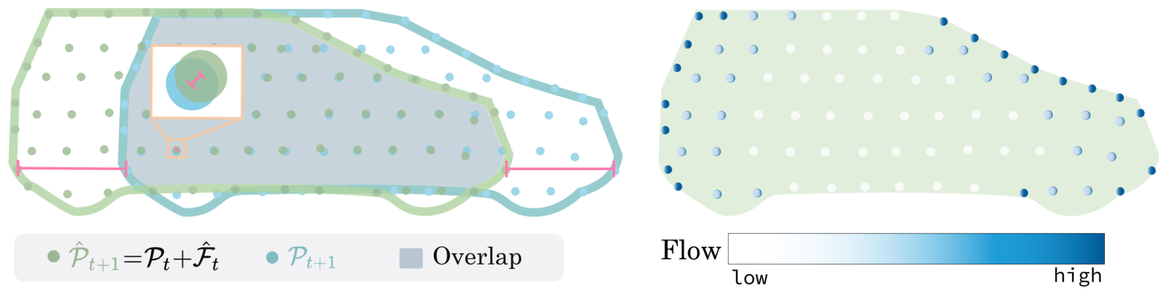
2. Method







***SeFlow* Architecture**. Top: With two consecutive point clouds as inputs, our model predicts the estimated flows of all points. Bottom: Conceptual visualization of the Chamfer loss and the three proposed training losses.



(a) Chamfer distance (b) Point flow magnitude

Classify Dynamic and Static using DUFOMap, a ray-casting map based dynamic awareness method. *Core idea* for our loss:

1. Constraint on dynamic nearest neighbor
2. Make sure flow of static points are zero.
3. For points from the same cluster, flow must be consistent.

$$\mathcal{L}_{\text{dcls}} = \frac{1}{|\mathcal{P}_{t,d}|} \sum_{c_i \in \mathcal{C}_{t,d}} \left(\sum_{p_j \in \mathcal{P}_{c_i}} ||\hat{f}_{p_j} - \tilde{f}_{c_i}||_2^2 \right)$$

$$\kappa_i = \arg\max_k \{\mathrm{D}(p_k, \mathcal{P}_{t+1,d}) | p_k \in \mathcal{P}_{c_i}\}$$

$$\tilde{f}_{c_i} = p'_{\kappa_i} - p_{\kappa_i}$$

3. Experiment

Table 1 and Table 2: ***SeFlow*** achieve **state-of-art (SOTA, 1st rank)** performance in the self-supervised scene flow task.

Method	Run Time per frame [ms]	Argoverse 2 3-way	Argoverse 2 FD	Waymo 3-way	Waymo FD
FastFlow3D† [11]	34 ± 5	0.0782	0.2072	0.0782	0.1954
DeFlow† [38]	48 ± 4	0.0534	0.1340	0.0446	0.0980
FastNSF [15]	507 ± 312	0.1657	0.3540	0.1579	0.3012
NSFP [14]	32,060 ± 10,112	0.0685	0.1503	0.1005	0.1712
ZeroFlow† [29]	34 ± 5	0.0814	0.2109	0.0921	0.2162
SeFlow (Ours)†	48 ± 4	**0.0628**	0.1525	**0.0598**	0.1506

Fig. 4: Training datasets size - Less but better

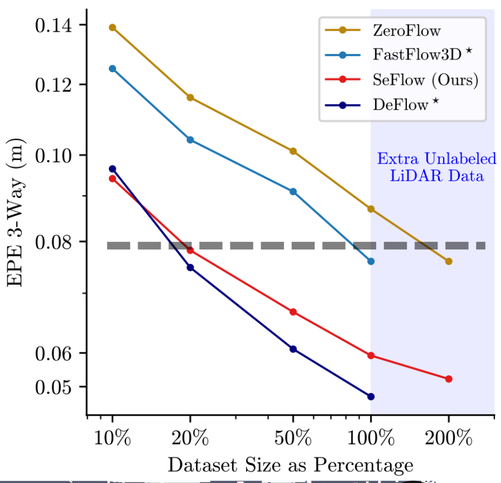
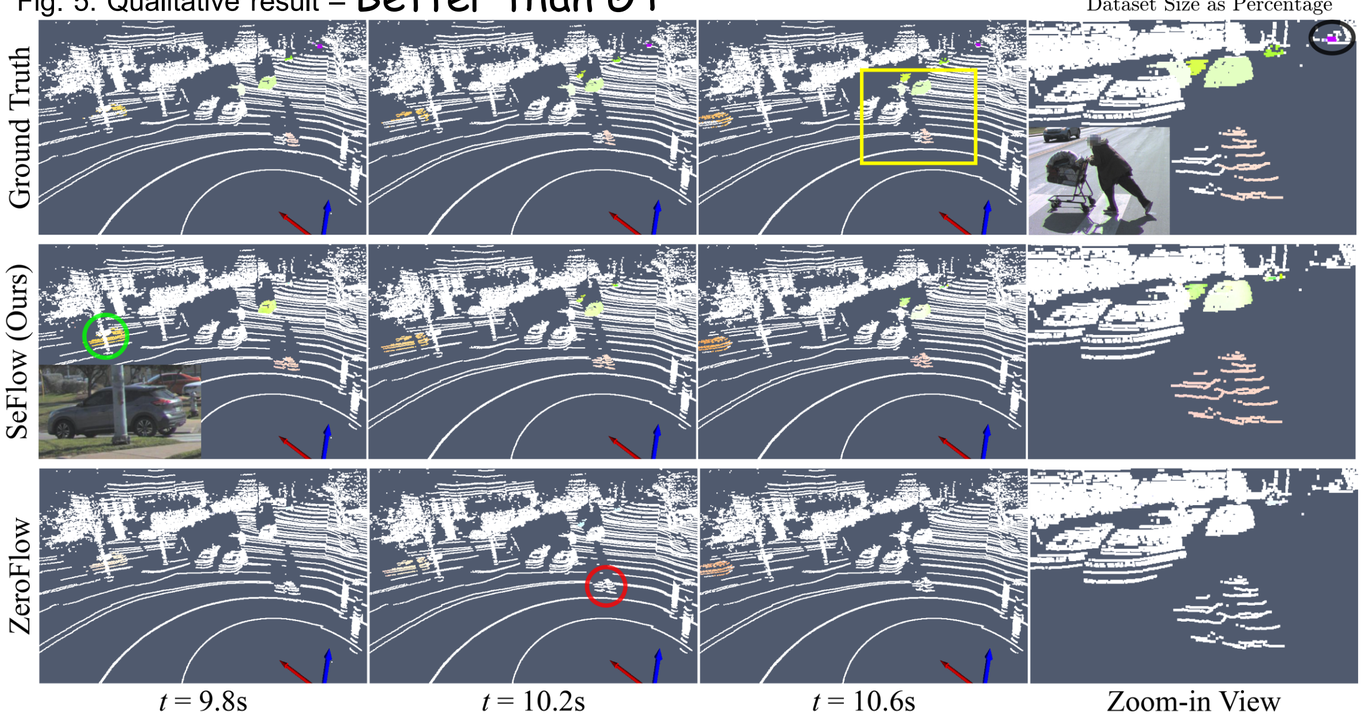
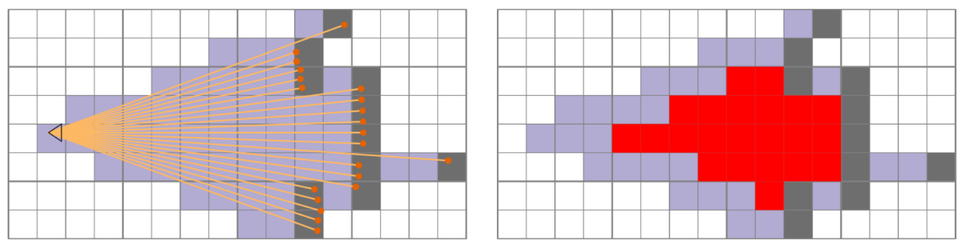


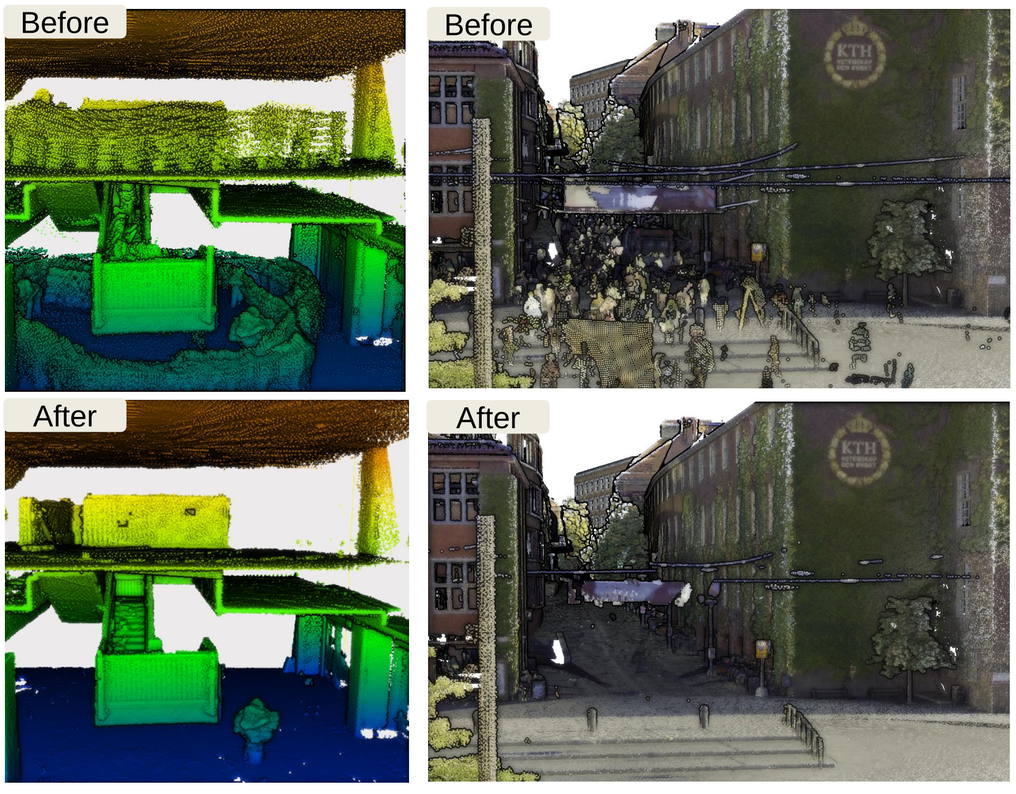
Fig. 5: Qualitative result – Better than GT



The key insight of DUFOMap is that if **a region has been observed as empty** at one time, **points observed** inside this region at another time have to be **dynamic**.



(a) (b)



4. Conclusion

- We propose ***DUFOMap***, a method for detecting dynamics by finding parts of space that has been observed as free taking into account sensor noise and localization errors. It achieves SOTA performance across different scenarios and sensors.
- We propose ***SeFlow***, a novel method that integrates a dynamic classification method in formulating efficient self-supervision objectives. It construct loss functions to learn dynamic flow estimation in imbalanced data and ensure consistent object-level flow.
- Future work: autonomous driving downstream task. (Under review, stay tuned :)

Paper Lists (abbreviation)

A. ECCV'24 ***SeFlow***
B. IEEE RA-L ***DUFOMap***
C. ICRA'24 DeFlow
D. IEEE RA-L BeautyMap
E. ITSC'23 DynamicMap Benchmark

WASP WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM