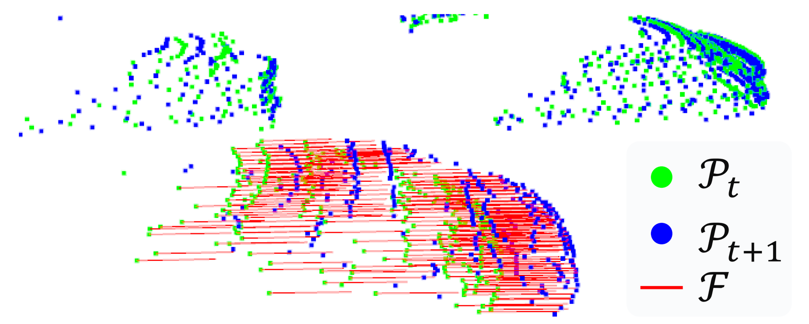
SeFlow: A Self-Supervised Scene Flow Method in Autonomous Driving

Qingwen Zhang[1], Yi Yang[1,2], Peizheng Li[3,4], Olov Andersson[1], Patric Jensfelt[1]

[1] KTH Royal Institute of Technology, Sweden, [2] Scania CV AB, Sweden, [3] University of Tübingen, Germany, [4] Mercedes-Benz AG, Germany

1. Introduction & Motivation

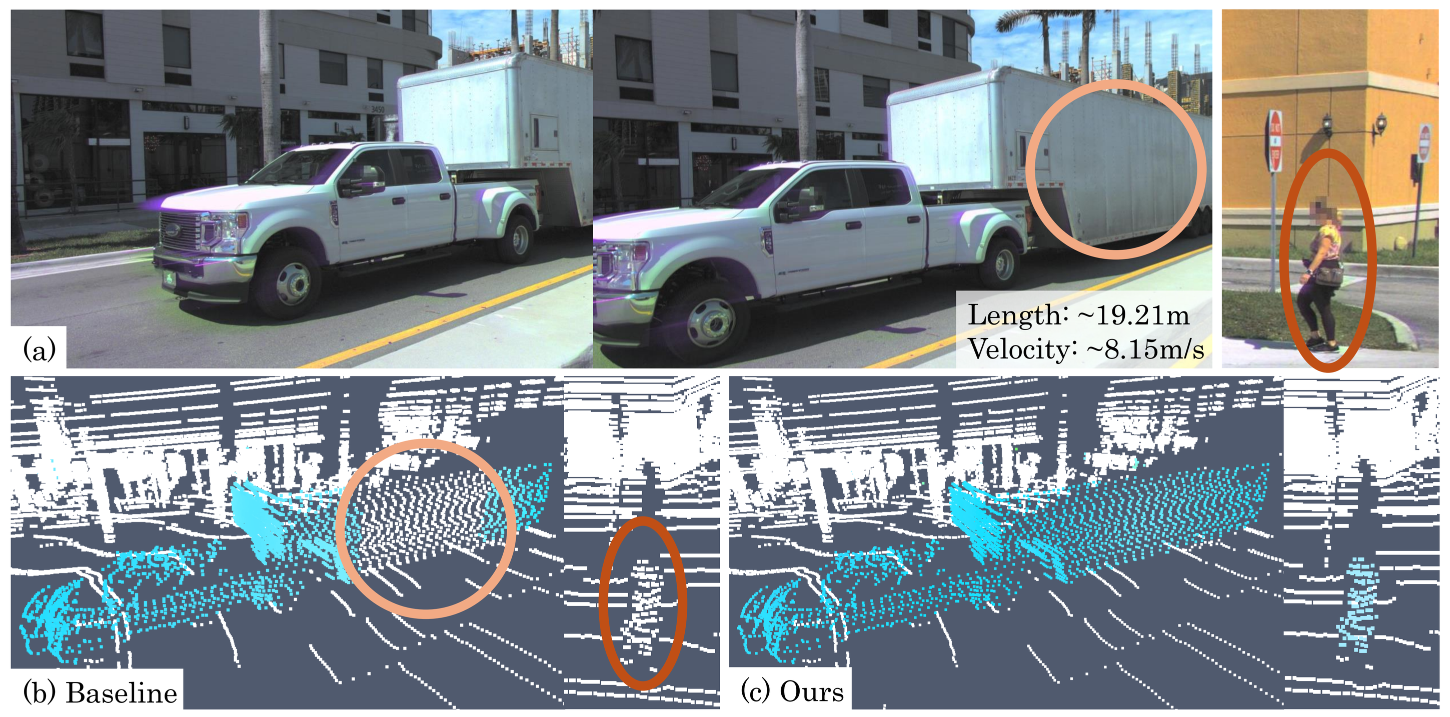
Scene flow estimation determines a scene's 3D motion field, by predicting **the motion of points** in the scene, especially for aiding tasks in autonomous driving.



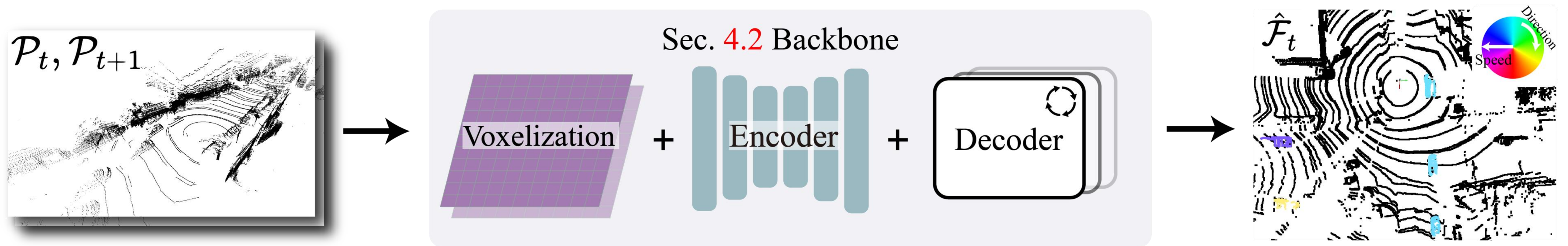
A common paradigm for addressing the scene flow problem is supervised learning by utilizing annotated LiDAR data. However, **expensive labeling inherently** limits the scalability of supervised learning methods**.**

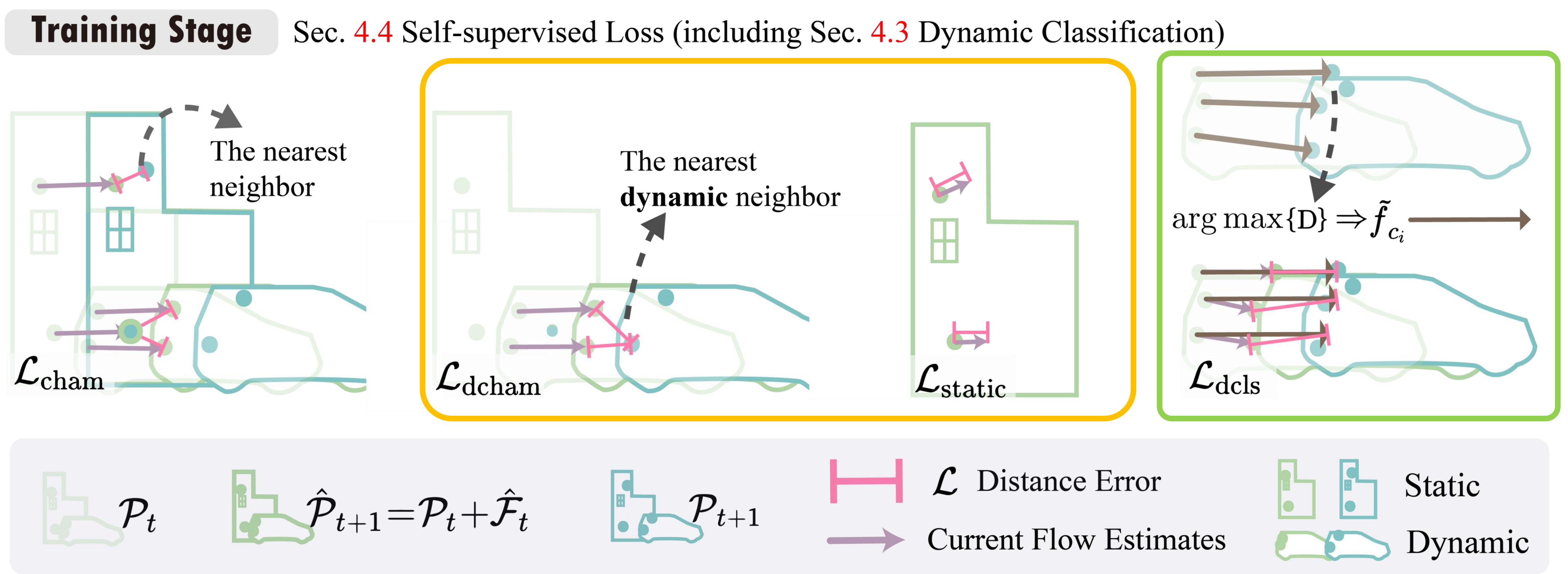
Existing **self-supervised** methods suffer from:

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

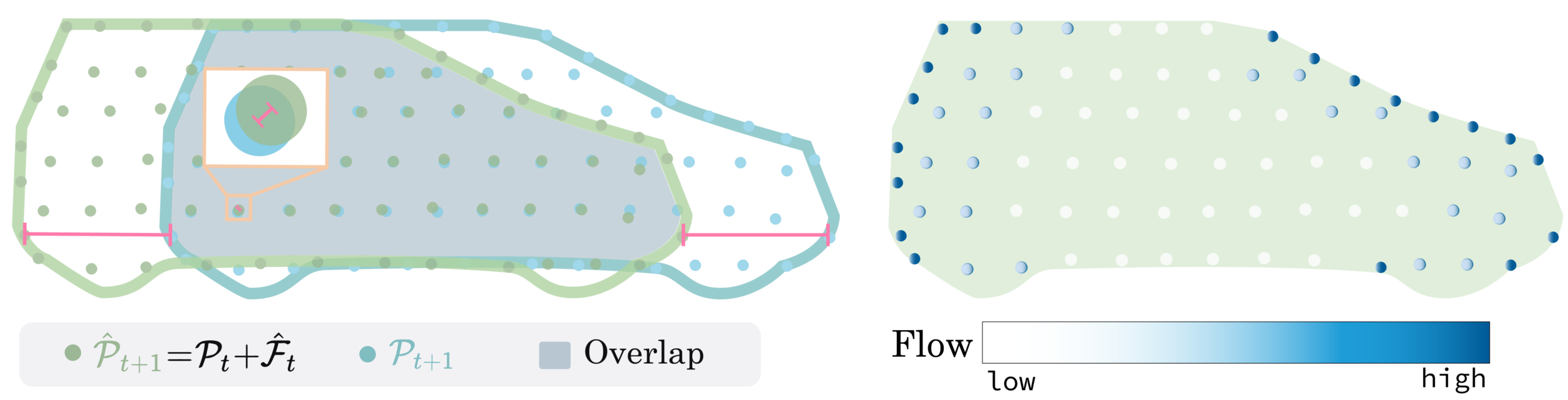


2. Method





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.

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

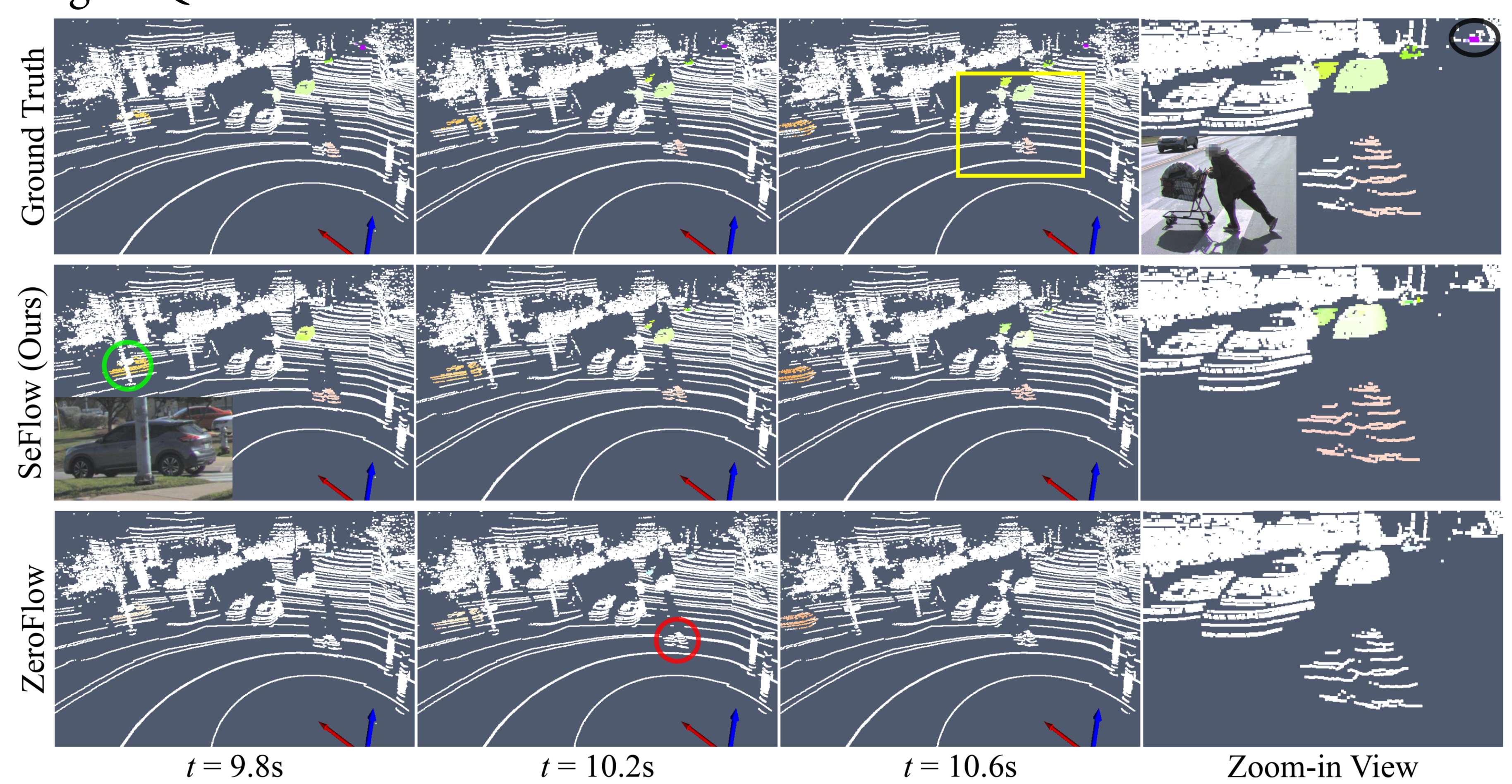
$$\kappa_i = \arg\max_k\{\mathrm{D}(p_k,\mathcal{P}_{t+1,d}) \mid p_k\in\mathcal{P}_{c_i}\}$$

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

Table 3: Ablation study of loss term

$\mathcal{L}_{cham}$	$\mathcal{L}_{dcham}$	$\mathcal{L}_{static}$	$\mathcal{L}_{dcls}$	EPE ↓ 3-way	FD	FS	BS
✓				0.0962	0.203	0.052	0.033
✓	✓			0.0916	0.181	0.059	0.035
✓	✓	✓		0.0779	0.220	**0.012**	**0.002**
✓	✓	✓	✓	**0.0643**	**0.160**	0.029	0.004

Fig. 5: Qualitative result – Better than GT



3. Experiments

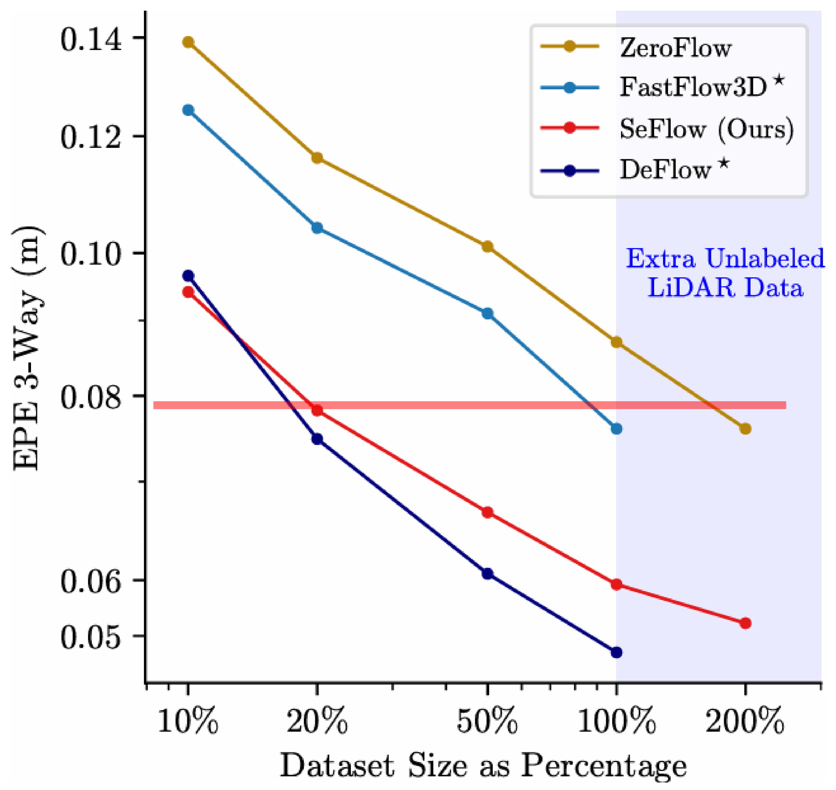
Table 1 and Table 2: **SeFlow** achieve **state-of-art** (**SOTA, 1st rank**) performance in the self-supervised scene flow task on Argoverse 2 and Waymo.

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

Table 4 and Fig. 4: Training datasets size - Less but better

EPE = End point error; FD = Foreground dynamic; BS = Background

Dataset	EPE ↓ 3-way	FD	FS	BS
10%	0.094	0.234	0.040	0.006
20%	0.078	0.197	0.032	0.004
50%	0.066	0.167	0.028	0.004
100%	**0.059**	0.147	0.026	0.004
ZF 100%	0.088	0.231	0.022	0.011
ZF 200%	0.076	0.198	0.018	0.011



4. Conclusion

- We propose **SeFlow**, a novel method that integrates a dynamic classification method in formulating efficient self-supervision objectives.
- We construct loss functions to learn dynamic flow estimation in imbalanced data and ensure consistent object-level flow, mitigating the effects of correspondence errors
- Our **SeFlow** ranks **1st** in Argoverse 2 self-supervised scene flow online leaderboard (Date: Sep 2, 2024).
- Future work: autonomous driving downstream task. (Under review, stay tuned :)

Scan Me