UniFlow: Zero-Shot LiDAR Scene Flow for Autonomous Driving

Siyi Li[1], Qingwen Zhang[2], Ishan Khatri[3,4], Kyle Vedder[1], Eric Eaton[1], Deva Ramanan[4], and Neehar Peri[4]

[1] University of Pennsylvania
[2] KTH Royal Institute of Technology
[3] StackAV
[4] Carnegie Mellon University

Abstract. LiDAR scene flow is the task of estimating per-point 3D motion between consecutive point clouds. Recent methods achieve centimeter-level accuracy on popular autonomous vehicle (AV) datasets, but are typically only trained and evaluated on a single sensor. In this paper, we aim to learn general motion priors that transfer to diverse and unseen LiDAR sensors. However, prior work in LiDAR semantic segmentation and 3D object detection demonstrate that naively training on multiple datasets yields worse performance than single dataset models. Interestingly, we find that this conventional wisdom does *not* hold for motion estimation, and that state-of-the-art scene flow methods greatly benefit from cross-dataset training *without architectural modification*. We posit that low-level tasks such as motion estimation may be less sensitive to sensor configuration; indeed, our analysis shows that models trained on fast-moving objects (e.g., from highway datasets) perform well on fast-moving objects, even across different datasets. Informed by our analysis, we propose UniFlow, a feedforward model that unifies and trains on multiple large-scale LiDAR scene flow datasets with diverse sensor placements and point cloud densities. Our frustratingly simple solution establishes a new state-of-the-art on Waymo and nuScenes, improving over prior work by 5.1% and 35.2% respectively. Moreover, UniFlow achieves state-of-the-art accuracy on *unseen datasets* like TruckScenes and AEVAScenes, outperforming prior dataset-specific models by 30.1% and 22.5% respectively. See our project page for additional visuals.

Keywords: LiDAR Scene Flow · Autonomous Vehicles · Cross-Domain Generalization

1 Introduction

Recent LiDAR scene flow methods achieve remarkable performance on popular autonomous vehicle (AV) datasets [12, 36, 37, 50] like Argoverse 2, nuScenes, and Waymo [1, 34, 45]. Notably, current state-of-the-art approaches train and evaluate on each dataset independently and do not attempt to generalize to unseen sensor configurations. Intuitively, since each dataset has different sensor placements and point cloud densities, training separate models for each domain

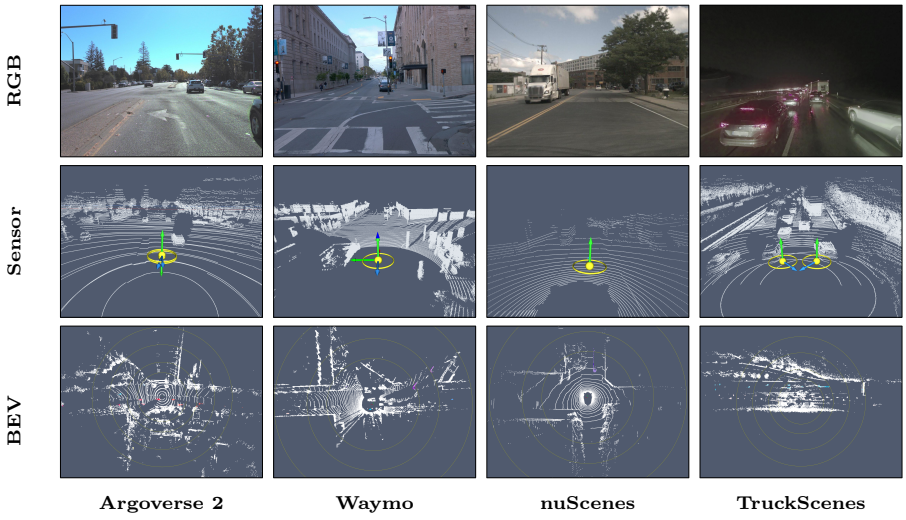

Fig. 1: Dataset Diversity. We visualize the front-center RGB (top), LiDAR sensor positions (middle) and BEV LiDAR point clouds (bottom) for Argoverse 2, Waymo, nuScenes and TruckScenes. Notably, all four datasets use different sensors, and collect data in different environments. Specifically, Argoverse 2, Waymo, and nuScenes collect data in urban city centers with sedans, while TruckScenes primarily collects data on highways with a truck. Due to the diversity of environments and sensor configurations, contemporary LiDAR scene flow methods typically only train and evaluate on each dataset independently. However, we find that multi-dataset training significantly improves both in-domain and out-of-domain generalization. Note that RGB images are shown for visualization purposes only; we address LiDAR-only scene flow for AVs.

seems reasonable (Figure 1). For example, Argoverse 2 (AV2) [45] uses two out-of-phase 32 beam LiDARs, nuScenes [1] uses one 32 beam sensor, and Waymo [34] uses a custom 64 beam sensor. Indeed, it is notoriously challenging to train multi-dataset models for LiDAR-based 3D object detection [32, 42] and semantic segmentation [14, 23, 31] that improve over single dataset models. Notably, prior works that do train on multiple datasets introduce bespoke architectures to explicitly address the problem of cross-domain alignment. Our key contribution is an empirical analysis that shows this does *not* hold for LiDAR-based scene flow; we simply expose standard architectures *without modification* to different sensor data and see a consistent improvement. We posit that "high-level" tasks like LiDAR semantic segmentation or 3D object detection are challenging to train across domains because they rely on human-defined taxonomies [16] that may be inconsistent across datasets [26]. Instead, scene flow models appear to capture properties of the underlying physical 4D world that hold regardless of LiDAR sensors observing it.

Cross-Domain Generalization for Low-Level Vision. We observe similar trends in RGB-based low-level vision tasks, where models learn generalizable geometric representations that transfer well across datasets. For example, FlowNet [5, 10] and RAFT [35] show that optical flow models trained on synthetic

Table 1: Analysis on Cross-Domain Generalization. We train Flow4D [15] on AV2, Waymo, and nuScenes and evaluate each dataset-specific model across datasets. We further break down performance by velocity speed buckets (meters / frame). We find that Flow4D (Waymo) nearly matches Flow4D (AV2)'s overall Dynamic Mean EPE on AV2 (0.199 vs. 0.192, shown in green), and even *outperforms* it on fast-moving objects (0.098 vs. 0.103, shown in gray). This suggests that Flow4D (Waymo) can *already* generalize across datasets. Flow4D (nuScenes) performs considerably worse on out-of-distribution datasets (shown in red) due its limited number of training examples.

Train	Test	FD (cm) ↓	Dyn. Mean ↓	[0, 0.5)	[0.5, 1.0)	[1.0, 2.0)	[2.0, ∞)
AV2	AV2	8.55	0.192	0.020	0.119	0.114	0.103
	Waymo	8.31	0.236	0.016	0.077	0.079	0.350
	nuScenes	20.40	0.463	0.017	0.249	0.334	0.610
Waymo	AV2	8.90	0.199	0.021	0.119	0.125	0.098
	Waymo	4.58	0.215	0.007	0.064	0.051	0.130
	nuScenes	16.04	0.408	0.017	0.217	0.319	0.540
nuScenes	AV2	16.08	0.357	0.039	0.239	0.272	0.395
	Waymo	17.24	0.364	0.022	0.222	0.209	0.598
	nuScenes	7.97	0.230	0.018	0.100	0.109	0.370

datasets (e.g. FlyingChairs, FlyingThings3D, Sintel) generalize surprisingly well to casually captured videos. More recently, zero-shot monocular depth estimators like MoGe [41] and monocular scene flow estimators like ZeroMSF [20] demonstrate that training on diverse datasets significantly improves zero-shot performance. Therefore, we posit that low-level LiDAR-based scene flow may similarly generalize to different sensors if trained at scale.

To better understand this problem, we first train dataset-specific models for AV2 [45], Waymo [34] and nuScenes [1] and evaluate zero-shot cross-domain performance. Surprisingly, Flow4D [15] trained on Waymo achieves similar Dynamic Mean EPE on AV2 as Flow4D trained on AV2. Importantly, Waymo's 64 beam LiDAR is considerably more dense than AV2's two out-of-phase 32-beam LiDARs (Figure 1), indicating that point density does not significantly impact LiDAR scene flow generalization. Further, Flow4D (Waymo) achieves lower EPE (i.e. better performance) than Flow4D (AV2) on fast moving objects in AV2 since there are more fast movers in Waymo than in AV2 (Table 1). This suggests that the accuracy of motion prediction is highly correlated with object velocity (Figure 2); in order to accurately predict fast motion, one should train on fast motion. This suggests that the key to LiDAR scene flow generalization is training on a diverse combination of datasets with a wide range of object velocities.

Towards Zero-Shot LiDAR Scene Flow. We argue that scaling up LiDAR-based scene flow methods will be a key enabler for 3D motion understanding and dynamic reconstruction in diverse environments. To this end, we retrain state-of-the-art methods *without modification* on supervised data from AV2, Waymo, and nuScenes and demonstrate significant improvements for both in-distribution and out-of-distribution generalization. We denote models trained with multiple datasets as UniFlow. *To the best of our knowledge, UniFlow is the first to demonstrate the benefits of cross-domain training for LiDAR scene flow.* Further, UniFlow models demonstrate remarkable zero-shot accuracy on TruckScenes [6]

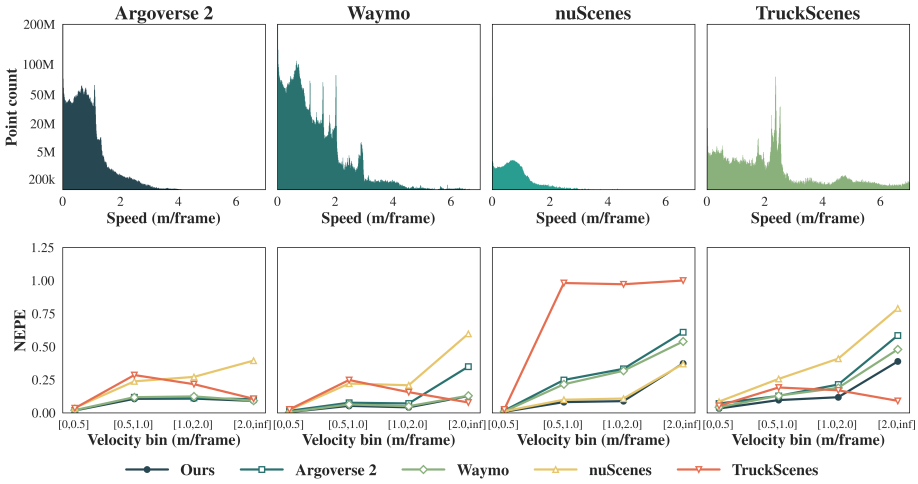

Fig. 2: Cross-Dataset Generalization Correlates with Velocity Distribution. We plot the velocity distributions for the AV2, Waymo, nuScenes, and TruckScenes train sets (top) and the Dynamic Mean EPE per velocity bin of Flow4D trained on AV2, Waymo, nuScenes, TruckScenes, and UniFlow (bottom). Notably, Flow4D trained on TruckScenes outperforms Flow4D trained on any other dataset for fast moving objects (2.0, ∞) across all datasets because it has the largest number of fast moving objects.

(an autonomous truck dataset collected at highway speeds) and AEVAScenes (a small-scale dataset with a novel FMCW LiDAR sensor), outperforming prior dataset-specific models by 30.1% (Table 6) and 22.5% (Table 7), respectively. Lastly, we find that cross-domain training on near-range (i.e $\leq$ 35 meters from the ego-vehicle) points improves scene flow generalization to unseen far-range points (Table 8). We hypothesize that sparse LiDAR points near the ego-vehicle mimic the distribution of dense LiDAR points far away from the vehicle [29].

Why Does Multi-Dataset Training Work? Our experiments suggest that multi-dataset training improves LiDAR scene flow performance because scene flow is inherently class-agnostic. This allows us to avoid complications arising from dataset-specific label definitions (Table 9). For example, nuScenes defines `bicycle` as excluding the rider, while Waymo includes the rider [26,30]. Such label ambiguity poses a significant challenge for semantic tasks, potentially explaining why naively combining datasets for LiDAR-based 3D object detection and semantic segmentation yields worse performance than single-dataset models. Furthermore, we find that with careful augmentations, we can simulate diverse sensor configurations that improve cross-dataset generalization (Figure 4). For example, dropping out LiDAR beams from a Waymo point cloud makes the data look more like a nuScenes point cloud, while augmenting point cloud height makes a sensor mounted on a sedan look like it was mounted on a truck.

Contributions. We present three major contributions. First, we highlight that dataset-specific scene flow models *already* show signs of cross-domain generalization, motivating our investigation into multi-dataset training. Next, we

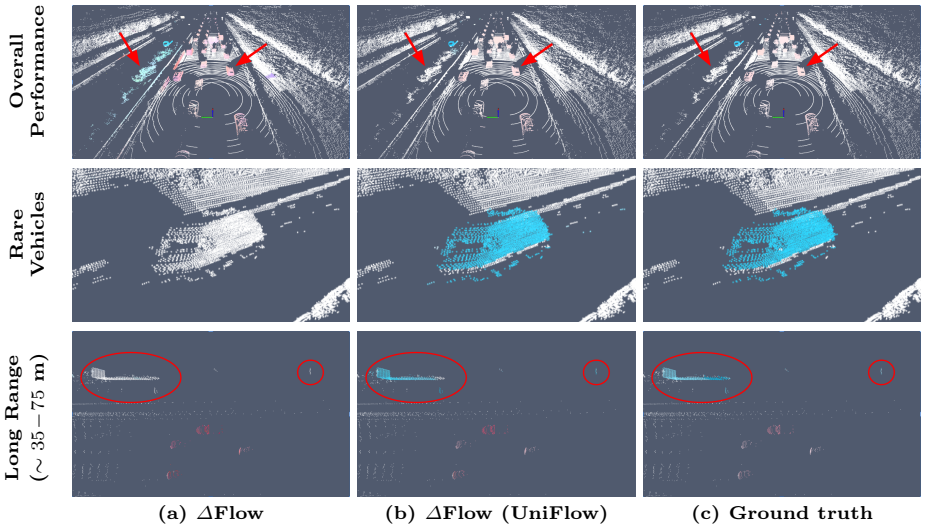

Fig. 3: Zero-Shot Generalization on TruckScenes. We qualitatively compare the dataset-specific ΔFlow model and our ΔFlow (UniFlow) model above. ΔFlow (UniFlow) produces more accurate vehicle motion estimates in general, and avoids falsely predicting motion for rain artifacts (on the top left) as seen in ΔFlow (top row). Next, ΔFlow (UniFlow) correctly estimates the motion of the van since it has been trained on more examples of rare classes (middle row). Lastly, ΔFlow (UniFlow) generalizes significantly better at long-range, accurately estimating the truck's motion at $\sim$35 m and the car's motion at $\sim$70 m, despite only being trained on LiDAR points up to 35m (bottom row).

demonstrate that re-training existing scene flow methods on diverse, large-scale data yields state-of-the-art performance on nuScenes and Waymo, improving over our baselines by 35.2% and 5.1% respectively. Finally, we show that UniFlow generalizes well to unseen sensors, outperforming dataset-specific models by 30.1% on TruckScenes, and by 22.5% on AEVAScenes.

2 Related Works

Scene Flow Estimation is the task of describing the 3D motion field between temporally successive point clouds [12, 38, 50], and is an important primitive for 3D motion understanding and dynamic scene reconstruction [40]. Early approaches [17,43,44,51] learned point-wise features to estimate per-point flow but struggled to scale to large outdoor environments [36, 48]. More recent work [9, 13,15,25,36] jointly estimates flow for all points in a scene. Contemporary methods can be broadly classified into feedforward models and optimization-based methods. Feedforward models directly learn a mapping between point cloud pairs and flow fields, but require large-scale human annotations [11, 15]. For real world datasets (typically from the autonomous vehicle domain), these human annotations are provided in the form of 3D bounding boxes and tracks for every object in the scene. In contrast, optimization-based methods do not require

labeled data, and instead optimize a learned representation per-scene [18, 19]. This per-scene optimization is prohibitively expensive, making it difficult to scale to large datasets. For example, EulerFlow [37] achieves state-of-the-art unsupervised accuracy on many popular AV datasets, but takes over 24 hours to optimize each scene. In summary, feedforward models achieve efficient inference but demonstrate limited generalization beyond their training data, while optimization-based methods produce high quality flow for diverse scenes, but are too slow for real-time applications. Our work aims to reconcile this trade-off by training a single feedforward model that achieves robust generalization across diverse datasets and sensors.

Cross-Domain Generalization for LiDAR is a long-standing challenge in LiDAR-based perception. Prior works in 3D object detection [7, 27] and semantic segmentation [2, 14, 46] show that naively training models on multiple LiDAR datasets often performs worse than models trained on a single dataset. This performance drop can be largely attributed to diverse sensor hardware (e.g., number of beams, scan pattern, point cloud density), environmental conditions (e.g., weather, geography), and sensor placement across datasets. Often, such methods also design bespoke architectures to address cross-domain generalization. In addition, several methods have introduced data augmentation strategies [33], such as random point dropping [39] and object scaling [3], while others [24, 28] have proposed unsupervised domain adaptation techniques that align feature distributions across datasets. However, such data augmentation strategies primarily address the *geometric domain gap* between datasets. For motion-centric tasks such as scene flow, we argue that understanding the distribution of object velocities between datasets represents a critical yet overlooked axis for addressing cross-domain generalization. To this end, our work presents the first systematic study of this *velocity domain gap*.

3 UniFlow: Towards Zero-Shot LiDAR Scene Flow for Autonomous Vehicles

UniFlow aims to learn general motion priors that transfer to diverse and unseen LiDAR sensors by training a single scene flow model across diverse autonomous driving datasets. In this section, we describe the process of unifying five popular AV datasets and our training strategy.

Dataset Unification. We unify five widely used AV datasets to facilitate multi-dataset training. Importantly, we standardize LiDAR sensor frame rates and annotation frequencies to ensure that the time interval between LiDAR sweeps is consistent across datasets. This step is crucial because state-of-the-art scene flow methods parameterize motion as the displacement between two consecutive LiDAR frames. Specifically, we down-sample high-frequency datasets like nuScenes (which provides LiDAR sweeps at 20 Hz) to 10 Hz to match AV2 and Waymo. For datasets with sparse scene flow annotations (e.g., nuScenes annotates tracks at 2 Hz), we train solely on the annotated frames and their real adjacent scans. This ensures that we do not introduce any label noise or

Table 2: Dataset Statistics. We summarize the five datasets used for training and evaluation below. We include both the (# Total Frames / # Annotated Frames) for nuScenes and TruckScenes.

Dataset	Split	# Scenes	# Frames	LiDAR Setup	Ego-Vehicle	Scenario
Argoverse 2	Train	700	110,071	2 x 32-beam	Car	Urban / City
	Val	150	23,547			
Waymo	Train	798	155,000	1 x 64-beam	Car	Urban / Suburban
	Val	202	40,000			
nuScenes	Train	700	137,575 / 27,392	1 x 32-beam	Car	Urban / City
	Val	150	29,126 / 5,798			
TruckScenes	Train	524	101,902 / 20,380	2 x 64-beam	Truck	Highway
	Val	75	14,625 / 2,925			
AEVAScenes	Val	65	6,500	6× FMCW	Car	Urban / Highway

motion artifacts due to interpolation. Lastly, we apply LineFit [8] for ground point removal for all datasets.

Unified Multi-Dataset Training. We retrain existing state-of-the-art scene flow architectures (SSF [13], Flow4D [15], and ΔFlow [50]) on the unified dataset mixture of nuScenes, AV2, and Waymo. Importantly, we do not re-weight dataset frequencies or apply dataset-specific augmentations. Instead, we employ a minimal yet effective augmentation set including height randomization to account for sensor placement ($z \pm 0.5$ - 2m, $p = 0.8$) and LiDAR ray dropping to simulate varying LiDAR point densities (alternating odd/even rays, $p = 0.35$). Ray dropping is not performed on nuScenes due to its inherent sparsity. Further, we do not discard any scans and use dataset provided ego-motion for motion compensation. We do not temporally align sensor measurements between datasets beyond frame rate. This unified training paradigm encourages models to learn domain-agnostic geometric priors rather than memorizing dataset-specific motion patterns, resulting in significant zero-shot generalization improvements. See the supplementary material for more details.

4 Experiments

In this section, we evaluate UniFlow on Argoverse 2, Waymo, nuScenes, TruckScenes, and AEVAScenes. We ablate the impact of dataset characteristics on model performance, and identify scaling trends. We encourage the reader to see the supplement for additional analysis and visuals.

Datasets. We train and evaluate UniFlow on five large-scale autonomous driving datasets to rigorously evaluate cross-domain generalization to diverse sensor configurations, ego-vehicle platforms, and driving scenarios (Table 2). AV2 and nuScenes collect data in dense urban environments with 32-beam LiDARs. Waymo uses a custom LiDAR sensor to collect denser point clouds from urban and suburban environments with more varied driving dynamics. We evaluate zero-shot generalization using TruckScenes [6] since it collects data in different environments (high-speed highway scenes vs. urban and suburban scenes),

Table 3: Argoverse 2 Test Set. We compare UniFlow with recent methods on the Argoverse 2 test set. Notably, Flow4D-XL (UniFlow) outperforms Flow4D by 9.0%, while SSF (UniFlow) outperforms SSF by 13.8% Dynamic Mean EPE. Notably, ΔFlow (UniFlow) achieves slightly worse mean Dynamic EPE than ΔFlow. This suggests that performance on AV2 may be saturating, and further improvements may be overfitting to label noise.

Methods	Three-way EPE (cm) ↓				Dynamic Bucket-Normalized ↓				
	Mean	FD	FS	BS	Mean	Car	Other	Pedestr.	VRU
Unsupervised									
NSFP	6.06	11.58	3.16	3.44	0.422	0.251	0.331	0.722	0.383
FastNSF	11.18	16.34	8.14	9.07	0.383	0.296	0.413	0.500	0.322
SeFlow	4.86	12.14	1.84	0.60	0.309	0.214	0.291	0.464	0.265
ICP Flow	6.50	13.69	3.32	2.50	0.331	0.195	0.331	0.435	0.363
EulerFlow	4.23	4.98	2.45	5.25	0.130	0.093	0.141	0.195	**0.093**
Supervised									
TrackFlow	4.73	10.30	3.65	0.24	0.269	0.182	0.305	0.358	0.230
DeFlow	3.43	7.32	2.51	0.46	0.276	0.113	0.228	0.496	0.266
SSF	2.73	5.72	1.76	0.72	0.181	0.099	0.162	0.292	0.169
Flow4D	2.24	4.94	1.31	0.47	0.145	0.087	0.150	0.216	0.127
ΔFlow	2.11	**4.33**	1.37	0.64	**0.113**	0.077	**0.129**	**0.149**	0.096
SSF (UniFlow, Ours)	2.23	4.89	1.31	0.51	0.156	0.102	0.147	0.241	0.134
Flow4D-XL (UniFlow, Ours)	2.07	4.50	1.31	**0.40**	0.132	**0.072**	0.131	0.218	0.107
ΔFlow (UniFlow, Ours)	**2.07**	4.41	**1.26**	0.54	0.118	0.073	0.136	0.164	0.098

Table 4: Waymo Validation Set. We compare UniFlow with recent methods on the Waymo val set. Notably, supervised methods achieve lower EPE than unsupervised methods, with ΔFlow (UniFlow) beating SeFlow by 62.0% on Foreground Dynamic.

Methods	Three-way EPE (cm) ↓				Dynamic Bucket-Normalized ↓			
	Mean	FD	FS	BS	Mean	Car	Pedestr.	VRU
Unsupervised								
NSFP	10.05	17.12	10.81	2.21	0.574	0.315	0.823	0.584
ZeroFlow	8.64	22.40	1.57	1.96	0.770	0.444	0.982	0.884
SeFlow	8.50	20.81	2.14	2.57	0.328	0.305	0.485	0.185
SeFlow++	3.63	9.30	0.87	0.71	0.232	0.201	0.521	0.247
Supervised								
DeFlow	4.46	9.80	2.59	0.98	-	-	-	-
SSF	2.19	5.62	0.74	0.23	0.264	0.097	0.469	0.225
Flow4D	1.82	4.58	0.61	0.28	0.215	0.074	0.423	0.146
ΔFlow	1.40	3.27	0.58	0.35	0.198	0.073	0.407	0.114
SSF (UniFlow, Ours)	1.85	4.74	**0.58**	**0.22**	0.234	0.075	0.434	0.195
Flow4D-XL (UniFlow, Ours)	1.60	3.83	0.68	0.30	0.191	**0.064**	**0.400**	0.110
ΔFlow (UniFlow, Ours)	**1.38**	**3.07**	0.61	0.46	**0.188**	0.065	0.403	**0.096**

with different sensors (32-beam LiDARs vs. 64-beam LiDARs), using a different vehicle type (truck vs. sedan). We also evaluate zero-shot generalization using AEVAScenes since it uses a novel FMCW LiDAR. These five datasets allow us to systematically analyze the impact of point density (32-beam vs. 64-beam), ego-vehicle perspective (car vs. truck), and velocity (urban vs. highway) on cross-domain generalization.

Metrics. We use *Dynamic Bucket-Normalized EPE* [12] as our primary metric to evaluate cross-domain generalization, particularly across diverse velocity distributions. This protocol is specifically designed to be both class-aware and speed-normalized. By normalizing motion estimation error by an object's speed, Dynamic Bucket-Normalized EPE measures the fraction of motion *not* described, enabling fair comparison between slow-moving pedestrians and fast-moving ve-

Table 5: nuScenes Validation Set. We compare UniFlow with recent methods on the nuScenes val set. Interestingly, SSF (UniFlow) outperforms Flow4D-XL (UniFlow) by 26.53% Dynamic Mean EPE, suggesting that SSF's architecture more effectively learns to estimate flow from sparse point clouds.

Methods	Three-way EPE (cm) ↓				Dynamic Bucket-Normalized ↓				
	Mean	FD	FS	BS	Mean	Car	Other	Pedestr.	VRU
Unsupervised									
NSFP	10.79	20.26	4.88	7.23	0.602	0.463	0.456	0.829	0.662
FastNSF	12.16	18.20	6.11	12.18	0.560	0.436	0.523	0.737	0.543
SeFlow	8.19	16.15	3.97	4.45	0.554	0.396	0.635	0.726	0.419
Supervised									
DeFlow	3.98	6.99	3.45	1.50	0.314	0.163	0.286	0.533	0.275
SSF	3.00	6.55	2.04	0.41	0.220	0.142	0.197	0.398	0.144
Flow4D	3.46	7.97	1.85	0.55	0.230	0.160	0.241	0.345	0.176
ΔFlow	2.33	4.83	1.37	0.79	0.216	0.138	0.219	0.327	0.181
SSF (UniFlow, Ours)	1.97	4.33	1.38	**0.20**	0.144	**0.081**	**0.131**	0.267	0.097
Flow4D-XL (UniFlow, Ours)	3.01	7.28	1.47	0.28	0.196	0.137	0.219	0.272	0.157
ΔFlow (UniFlow, Ours)	**1.63**	**3.19**	**1.29**	0.42	**0.140**	0.098	0.137	**0.232**	**0.091**

Table 6: TruckScenes Validation Set. We compare UniFlow with recent methods on the TruckScenes val set. According to three-way EPE, Flow4D achieves the best performance due to low Foreground Dynamic error. However, Dynamic Bucket Normalized EPE highlights that Flow4D-XL (UniFlow) outperforms it across all moving object categories, reinforcing that three-way EPE is a biased metric [12].

Methods	Three-way EPE (cm) ↓				Dynamic Bucket-Normalized ↓				
	Mean	FD	FS	BS	Mean	Car	Other	Pedestr.	VRU
Unsupervised									
NSFP	45.63	120.45	2.07	14.38	0.658	0.303	0.350	1.221	0.758
FastNSF	30.72	59.44	3.35	29.38	0.588	0.218	0.376	1.124	0.635
SeFlow	37.41	103.97	1.56	6.68	0.681	0.494	0.752	0.886	0.591
ICP Flow	58.82	169.91	1.43	5.12	0.472	0.302	0.614	0.596	0.376
Supervised									
DeFlow	7.30	16.47	1.67	3.77	0.570	0.180	0.410	0.970	0.730
SSF	6.20	15.17	**1.01**	2.44	0.453	0.215	0.308	0.875	0.414
Flow4D	16.14	44.87	1.71	1.85	0.456	0.176	0.351	0.885	0.413
ΔFlow	**7.28**	**16.26**	1.36	4.52	0.402	0.196	0.400	0.690	0.323
Zero-shot									
SSF (UniFlow, Ours)	35.23	103.72	1.69	**0.27**	0.435	0.149	0.455	0.669	0.466
Flow4D-XL (UniFlow, Ours)	23.59	68.41	1.78	0.57	**0.281**	**0.088**	**0.277**	0.530	0.230
ΔFlow (UniFlow, Ours)	16.04	45.76	1.02	1.32	0.283	0.101	0.293	**0.513**	**0.226**

hicles. We also report three-way EPE [4] for completeness, but note that it is a biased estimator of performance.

Baselines. We compare UniFlow against both self-supervised and fully-supervised methods: NSFP [18], FastNSF [19], ICPFlow [21], SeFlow [49], SeFlow++ [47], EulerFlow [37], VoteFlow [22], TrackFlow [12], DeFlow [48], SSF [13], Flow4D [15] and ΔFlow [50]. We design a new model called Flow4D-XL, which simply adds more parameters to Flow4D to address model underfitting and improve its scaling behavior (see the supplementary material). To ensure fairness, Argoverse 2 test-set results are obtained directly from the public leaderboard. We use Waymo, nuScenes and TruckScenes baselines as implemented in Open-

Table 7: AEVAScenes Validation Set. We compare UniFlow with recent methods on the AEVAScenes val set. Notably, due to the limited number of logs in the dataset, all methods are evaluated zero-shot. UniFlow models consistently outperform single-dataset models.

Method	Three-way EPE (cm) ↓				Dynamic Bucket-Norm. ↓				
	Mean	FD	FS	BS	Mean	Car	Other	Pedestr.	VRU
SSF (AV2)	32.13	92.92	2.09	1.39	0.822	0.822	0.851	0.858	0.859
SSF (Waymo)	20.59	59.38	1.86	0.53	0.811	0.513	0.957	0.877	0.899
SSF (nuScenes)	31.04	88.68	3.08	1.36	0.772	0.689	0.865	0.775	0.759
SSF (UniFlow, Ours)	10.80	29.76	2.05	0.60	0.544	0.239	0.686	0.612	0.639
Flow4D (AV2)	16.13	43.20	3.19	2.01	0.419	0.317	0.434	0.491	0.433
Flow4D (Waymo)	8.14	18.89	4.03	1.50	0.408	0.152	0.560	0.491	0.560
Flow4D (nuScenes)	25.61	70.06	3.86	2.93	0.535	0.512	0.678	0.459	0.489
Flow4D-XL (UniFlow, Ours)	**6.14**	13.16	3.89	1.37	0.338	0.114	**0.312**	0.480	0.448
ΔFlow (AV2)	15.63	43.37	2.73	0.79	0.444	0.302	0.560	0.460	0.453
ΔFlow (Waymo)	7.01	16.27	3.59	1.19	0.391	0.123	0.648	0.426	0.426
ΔFlow (nuScenes)	21.80	49.91	4.12	1.56	0.457	0.396	0.596	0.435	0.402
ΔFlow (UniFlow, Ours)	6.65	13.17	5.16	1.62	**0.303**	**0.112**	0.359	**0.398**	**0.344**

SceneFlow[5] with standard training hyperparameters. We include additional implementation details in the supplement.

Comparison to State-of-the-Art Methods. We demonstrate the effectiveness of the UniFlow framework by retraining SSF, Flow4D, and ΔFlow on multiple datasets and compare their performance against recent approaches on popular scene flow benchmarks (Tables 3, 4, and 5). On the Argoverse 2 test set, Flow4D-XL (UniFlow) improves over Flow4D by 9.0% and SSF (UniFlow) improves over SSF by 13.8% in Dynamic Mean EPE. Moreover, Flow4D-XL (UniFlow) achieves near-parity with EulerFlow at a fraction of the runtime. We see similar improvements between Flow4D and SSF and their UniFlow counterparts on the Waymo (Table 4) and nuScenes (Table 5) validation sets. Interestingly, SSF (UniFlow) outperforms Flow4D-XL (UniFlow) on nuScenes, suggesting that the SSF architecture is naturally better suited for processing sparse point clouds.

TruckScenes Zero-Shot Performance. We demonstrate UniFlow's strong zero-shot generalization capabilities with extensive evaluations on TruckScenes (Table 6). All baseline methods (except UniFlow) were trained on TruckScenes. Despite never being explicitly trained on high speed highway scenes, Flow4D-XL (UniFlow) significantly outperforms the dataset-specific Flow4D on Dynamic Mean EPE. Notably, Flow4D-XL (UniFlow) was only trained on urban driving scenes collected from a sedan, indicating that our approach can successfully generalize across significant domain gaps. Table 6 also highlight the importance of the speed normalized and bucketed EPE metric [12], as three-way EPE results don't adequately measure Flow4D-XL's (UniFlow) significant improvements on pedestrian and VRU performance. We ablate this further in the supplementary material.

[5] https://github.com/KTH-RPL/OpenSceneFlow

Table 8: Ablation on Long-Range Generalization. All $^{\text{LR}}$ results are obtained by evaluating models at extended voxel ranges while keeping the voxel size fixed and reusing the original trained weights. $^{\text{ZS}}$ denotes zero-shot evaluation, where the model has never seen the target domain during training. We report Dynamic Mean EPE for each distance interval from the ego-vehicle. ΔFlow (UniFlow) demonstrates strong zero-shot generalization to extended spatial ranges for both in- and out-of-domain datasets.

Dataset	Method	Mean	0–35	35–50	50–75	75–100	100–inf
AV2	ΔFlow	0.526	0.117	0.492	0.745	0.708	0.568
	ΔFlow$^{\text{LR}}$	0.259	0.117	0.205	0.240	0.325	0.409
	ΔFlow$^{\text{LR}}$ (UniFlow, Ours)	**0.226**$_{\downarrow 12.7\%}$	0.114$_{\downarrow 2.6\%}$	0.198$_{\downarrow 3.4\%}$	0.219$_{\downarrow 8.8\%}$	0.260$_{\downarrow 20.0\%}$	0.339$_{\downarrow 17.1\%}$
Waymo	ΔFlow	0.379	0.055	0.455	0.628	-	-
	ΔFlow$^{\text{LR}}$	0.115	0.068	0.112	0.166	-	-
	ΔFlow$^{\text{LR}}$ (UniFlow, Ours)	**0.105**$_{\downarrow 9.1\%}$	0.063$_{\downarrow 7.4\%}$	0.104$_{\downarrow 7.1\%}$	0.149$_{\downarrow 10.2\%}$	-	-
nuScenes	ΔFlow	0.389	0.172	0.380	0.633	0.606	-
	ΔFlow$^{\text{LR}}$	0.267	0.181	0.242	0.318	0.447	-
	ΔFlow$^{\text{LR}}$ (UniFlow, Ours)	**0.248**$_{\downarrow 7.1\%}$	0.166$_{\downarrow 7.9\%}$	0.209$_{\downarrow 13.6\%}$	0.301$_{\downarrow 5.3\%}$	0.405$_{\downarrow 9.4\%}$	-
TruckScenes	ΔFlow	1.679	1.211	1.514	1.901	1.835	1.936
	ΔFlow$^{\text{LR}}$	1.479	1.184	1.311	1.487	1.582	1.831
	ΔFlow$^{\text{LR,ZS}}$ (UniFlow, Ours)	**0.880**$_{\downarrow 40.6\%}$	0.786$_{\downarrow 33.6\%}$	0.826$_{\downarrow 37.0\%}$	0.694$_{\downarrow 53.3\%}$	0.779$_{\downarrow 50.8\%}$	1.317$_{\downarrow 28.1\%}$

Table 9: Ablation on Semantic Cross-Domain Generalization. We validate our claim that semantics are less generalizable than motion by retraining Flow4D-XL with an additional per-point semantic classification head. Notably, training Flow4D-XL on the unified dataset improves zero-shot scene flow estimation on TruckScenes, but hinders mIoU on out-of-distribution datasets.

Dataset		**Scene Flow↓**		**Semantic Classification ↑**					
Train	**Test**	**Three-Way (cm)**	**Dyn. Norm.**	**mIoU**	**Accuracy**	**CAR**	**OTHER**	**PED**	**VRU**
AV2	AV2	3.46	0.204	75.3%	93.9%	90.9%	86.2%	88.2%	36.0%
Waymo	Waymo	1.74	0.202	79.1%	98.9%	99.0%	-	87.1%	51.3%
nuScenes	nuScenes	2.99	0.214	70.5%	89.8%	82.8%	83.3%	80.5%	35.4%
TruckScenes	TruckScenes	15.56	0.462	64.6%	95.3%	91.0%	92.9%	32.6%	42.1%
Unified	AV2	3.26	0.178	78.8%	94.2%	91.4%	85.8%	90.0%	47.9%
Unified	Waymo	1.57	0.190	64.0%	99.4%	99.4%	0.0%	93.1%	63.5%
Unified	nuScenes	3.19	0.215	79.9%	93.4%	88.0%	89.1%	82.3%	60.3%
Unified	TruckScenes	24.69	0.288	40.5%	66.1%	53.5%	46.4%	20.1%	41.9%

AEVAScenes Zero-Shot Performance. To further evaluate zero-shot generalization to unseen LiDARs, we benchmark our models on the AEVAScenes dataset [6], which features high-resolution FMCW LiDARs that differs substantially from rotating LiDAR sensors used in existing benchmarks. Due to the limited scale of AEVAScenes, all methods are evaluated zero-shot. Notably, UniFlow variants consistently outperform their respective best single-dataset counterparts on AEVAScenes, with SSF (UniFlow) improving Dynamic Mean EPE by 33.8%, Flow4D-XL (UniFlow) by 17.2%, and ΔFlow (UniFlow) by 22.5%. We include visuals of zero-shot performance in the supplementary material.

Analysis of Long-Range Generalization. To evaluate whether unified training produces range-robust representations, we evaluate how models trained on short-range LiDAR data transfer to longer-range settings *without retraining*.

[6] https://scenes.aeva.com/

Table 10: Ablation on TruckScenes. We ablate the impact of multi-dataset training, data augmentation, and model scaling on zero-shot TruckScenes performance. All the model architectures we test benefit from multi-dataset training, while data augmentation continues to provide notable improvements even with 3× larger training sets, suggesting that scaling up UniFlow may yield further improvements.

Method	FD (cm) ↓	Dynamic EPE ↓	[0, 0.5)	[0.5, 1.0)	[1.0, 2.0)	[2.0, ∞)
Flow4D (Waymo)	116.01	0.336	0.071	0.129	0.215	0.586
+ Unified Dataset	93.76 ↓22.25	0.310 ↓0.026	0.040 ↓0.031	0.116 ↓0.013	0.170 ↓0.045	0.594 ↑0.008
+ Augmentation	68.32 ↓25.44	0.301 ↓0.009	0.047 ↑0.007	0.111 ↓0.005	0.141 ↓0.029	0.407 ↓0.187
+ XL Backbone	68.41 ↑0.09	0.281 ↓0.020	0.033 ↓0.014	0.097 ↓0.014	0.119 ↓0.022	0.389 ↓0.018
SSF (Waymo)	170.65	0.737	0.049	0.555	0.740	0.971
+ Unified Dataset	133.04 ↓37.61	0.577 ↓0.160	0.042 ↓0.007	0.257 ↓0.298	0.514 ↓0.226	0.827 ↓0.144
+ Augmentation	103.72 ↓29.32	0.435 ↓0.142	0.036 ↓0.006	0.149 ↓0.108	0.285 ↓0.229	0.637 ↓0.190
ΔFlow (Waymo)	117.83	0.357	0.067	0.114	0.302	0.714
+ Unified Dataset	84.16 ↓33.67	0.331 ↓0.026	0.036 ↓0.031	0.119 ↑0.005	0.268 ↓0.034	0.659 ↓0.055
+ Augmentation	16.04 ↓68.40	0.283 ↓0.048	0.039 ↑0.003	0.094 ↓0.025	0.127 ↓0.141	0.429 ↓0.230

Table 11: Ablation on Frame-Rate. We evaluate Flow4D-XL (UniFlow), SSF (UniFlow) and ΔFlow's (UniFlow) generalization to different frame rates as a stress test outside standard operating conditions. Note that slower frame-rates effectively mimic a slower ego-vehicle, while faster-frame rates mimic a faster ego-vehicle. Although we train all models on 10 Hz annotations, multi-dataset training yields better results than single-dataset models, particularly at slower frame rates. Interestingly, faster frame rates do not negatively impact performance nearly as much as slower frame-rates.

Method	Three-way EPE (cm)				Dynamic Bucket-Normalized				
	Mean ↓	FD ↓	FS ↓	BS ↓	Dyn. Mean ↓	Car ↓	Other ↓	Pedestr. ↓	VRU ↓
2 Hz									
Flow4D	30.79	88.10	3.26	1.02	0.495	0.364	0.494	0.613	0.509
SSF	31.16	89.82	3.07	0.59	0.520	0.363	0.498	0.684	0.535
ΔFlow	26.41	74.43	3.12	1.66	0.457	0.370	0.471	0.545	0.440
Flow4D-XL (UniFlow, Ours)	27.52	79.54	**2.36**	0.65	0.444	0.297	**0.415**	0.587	**0.476**
SSF (UniFlow, Ours)	28.67	83.11	2.50	**0.39**	0.503	**0.296**	0.432	0.671	0.611
ΔFlow (UniFlow, Ours)	**23.87**	**68.26**	2.46	0.88	**0.398**	0.306	0.420	**0.494**	0.494
5 Hz									
Flow4D	9.58	25.64	2.45	0.65	0.323	0.215	0.312	0.432	0.333
SSF	8.74	23.02	2.74	0.47	0.311	0.179	0.286	0.489	0.291
ΔFlow	6.43	16.54	1.98	0.76	0.264	0.173	0.271	0.369	0.242
Flow4D-XL (UniFlow, Ours)	8.02	21.78	1.92	0.37	0.290	0.190	0.265	0.405	0.302
SSF (UniFlow, Ours)	6.75	17.96	2.06	**0.22**	0.285	**0.135**	**0.218**	0.454	0.333
ΔFlow (UniFlow, Ours)	**5.26**	**13.44**	**1.81**	0.58	**0.232**	0.150	0.220	**0.343**	**0.214**
10 Hz (Standard Frame Rate)									
Flow4D	3.46	7.97	1.85	0.55	0.230	0.160	0.241	0.241	0.176
SSF	3.00	6.55	2.04	0.41	0.220	0.142	0.197	0.398	0.144
ΔFlow	2.05	4.19	1.36	0.59	0.173	0.132	0.184	0.261	0.117
Flow4D-XL (UniFlow, Ours)	3.01	7.28	1.47	0.28	0.196	0.137	0.219	0.272	0.157
SSF (UniFlow, Ours)	1.97	4.33	1.38	**0.20**	0.144	**0.081**	**0.131**	0.267	0.097
ΔFlow (UniFlow, Ours)	**1.63**	**3.18**	**1.29**	0.42	**0.140**	0.098	0.137	**0.232**	**0.091**
20 Hz									
Flow4D	2.55	5.79	1.35	0.50	0.272	0.196	0.315	0.377	0.200
SSF	2.94	6.72	1.73	0.38	0.316	0.177	0.303	0.533	0.250
ΔFlow	1.60	3.28	0.48	0.48	0.220	0.167	0.231	0.327	0.156
Flow4D-XL (UniFlow, Ours)	1.95	4.65	1.00	0.21	0.230	0.171	0.269	0.315	0.163
SSF (UniFlow, Ours)	2.13	5.16	1.11	**0.12**	0.246	**0.127**	0.275	0.395	0.189
ΔFlow (UniFlow, Ours)	**1.30**	**2.60**	**0.96**	0.34	**0.196**	0.153	**0.215**	**0.277**	**0.138**

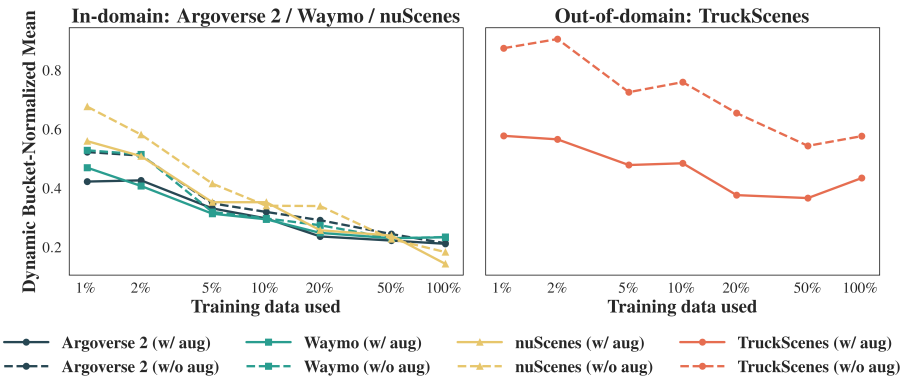

Fig. 4: Scaling Laws. We evaluate both the in-distribution (on AV2, nuScenes, and Waymo) and out-of-distribution (on TruckScenes) performance of Flow4D-XL (UniFlow) with different amounts of training data. Unsurprisingly, increasing data reduces Dynamic Mean EPE. However, we find that data augmentation is significantly more important for out-of-distribution performance, and has minimal impact on in-distribution performance. Lower is better.

Specifically, we extend the voxelization range during inference while keeping the voxel size identical to the original trained model. This change alters only the spatial coverage of the dynamic voxelization module, requiring a simple weight transfer step (i.e., model weights trained for short-range scene flow are loaded into a model with identical voxel size but a larger voxel grid for long-range processing) to adapt to the new grid dimensions while preserving all learned features. As shown in Table 8, $\Delta\text{Flow}^{\text{LR,ZS}}$ (UniFlow) consistently outperforms $\Delta\text{Flow}^{\text{LR}}$ across all distance intervals on AV2, reducing Dynamic Mean EPE by 12.7% on average and up to 20.0% in the 75–100 m range. This indicates that unified multi-dataset training produces representations that remain robust even when the perception range is significantly expanded. On TruckScenes, $\Delta\text{Flow}^{\text{LR,ZS}}$ (UniFlow) achieves a remarkable 40.6% reduction in overall error compared to $\Delta\text{Flow}^{\text{LR}}$, with improvements exceeding 50% in the 50–100 m range, despite never being trained on this dataset. These results highlight that the representations learned through unified multi-dataset training naturally scale to larger spatial extents.

Analysis of Semantic Cross-Domain Generalization. We evaluate the benefits of unified training on both geometric and semantic understanding by extending Flow4D-XL with a multi-task semantic head. This new Flow4D-XL variant simultaneously predicts scene flow and per-point semantics using the same backbone features. Since all datasets follow different taxonomies, we map their annotations to four coarse categories (e.g. `car`, `other`, `pedestrian`, and `VRU`) and jointly train the model with an additional class-balanced cross-entropy loss. As shown in Table 9, adding the semantic head has almost no effect on scene flow accuracy, indicating that the auxiliary task does not interfere with motion estimation. Consistent with our original findings, unified multi-dataset training continues to improve both in-domain and zero-shot performance for scene flow.

However, training the semantic prediction head on the unified dataset yields mixed results. Models trained on unified datasets show small in-domain gains, but perform substantially worse than dataset-specific model when evaluated on unseen domains. This supports our hypothesis that dataset unification primarily benefits low-level geometric representations, while high-level semantics remain sensitive to dataset-specific biases.

Analysis of Training Recipe. We ablate the impact of multi-dataset training, data augmentation, and model scaling on zero-shot TruckScenes performance in Table 10. All baselines are trained on Waymo only to ensure fair zero-shot evaluation. All model architectures show clear gains from multi-dataset training. Notably, despite training on a dataset three times larger, data augmentation continues to show substantial improvements, indicating that further scaling may yield additional performance gains.

Analysis on Scaling Laws. We evaluate Flow4D-XL (UniFlow) with varying amounts of training data on both in-distribution benchmarks (AV2, nuScenes, and Waymo) and an out-of-distribution benchmark (TruckScenes) in Fig 4. As expected, larger training sets reduce Dynamic Mean EPE. However, data augmentation proves far more critical for out-of-distribution performance, while its effect on in-distribution performance remains minimal.

Analysis on Frame-Rate. We evaluate the generalization of Flow4D-XL (UniFlow), SSF (UniFlow) and ΔFlow (UniFlow) across different frame rates in Table 11. These settings serve as a stress test of temporal generalization outside standard operating conditions: slower frame rates approximate a slower-moving ego-vehicle, while faster frame rates approximate a faster one. Although all models are trained using 10 Hz annotations, multi-dataset training consistently improves performance over single-dataset models, especially at slower frame rates. Interestingly, higher frame rates have a smaller negative effect on performance, likely because models only need to estimate smaller displacement vectors for each timestep.

5 Conclusion

In this paper, we introduce UniFlow, a frustratingly simple approach for scalable LiDAR scene flow that re-trains off-the-shelf models with diverse data from multiple datasets. Although prior work in LiDAR-based 3D object detection and segmentation suggest that naively training on multiple LiDAR datasets yields worse results than single-dataset trained models, we posit that "low-level" tasks such as motion estimation may be less sensitive to sensor configuration. Notably, our model establishes a new state-of-the-art on Waymo and nuScenes, improving over prior work by 5.1% and 35.2% respectively. Although UniFlowis primarily trained on urban city driving scenarios and standard spinning LiDARs, we demonstrate that it generalizes surprisingly well to highway truck driving and novel FMCW sensors *without any dataset-specific fine-tuning*. Our extensive analysis shows that such improvements are model-agnostic, suggesting that future scene flow methods should adopt this multi-dataset training strategy.

References

1. Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., Beijbom, O.: nuscenes: A multimodal dataset for autonomous driving. In: CVPR (2020) 1, 2, 3
2. Caunes, A., Chateau, T., Fremont, V.: seg_3d_by_pc2d: Multi-view projection for domain generalization and adaptation in 3d semantic segmentation. arXiv preprint arXiv:2505.15545 (2025) 6
3. Cen, J., Yun, P., Zhang, S., Cai, J., Luan, D., Tang, M., Liu, M., Yu Wang, M.: Open-world semantic segmentation for lidar point clouds. In: European Conference on Computer Vision. pp. 318–334. Springer (2022) 6
4. Chodosh, N., Ramanan, D., Lucey, S.: Re-evaluating lidar scene flow. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 6005–6015 (January 2024) 9
5. Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., Van Der Smagt, P., Cremers, D., Brox, T.: Flownet: Learning optical flow with convolutional networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2758–2766 (2015) 2
6. Fent, F., Kuttenreich, F., Ruch, F., Rizwin, F., Juergens, S., Lechermann, L., Nissler, C., Perl, A., Voll, U., Yan, M., et al.: Man truckscenes: A multimodal dataset for autonomous trucking in diverse conditions. Advances in Neural Information Processing Systems **37**, 62062–62082 (2024) 3, 7
7. Hegde, D., Lohit, S., Peng, K.C., Jones, M., Patel, V.: Multimodal 3d object detection on unseen domains. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 2499–2509 (2025) 6
8. Himmelsbach, M., Hundelshausen, F.V., Wuensche, H.J.: Fast segmentation of 3d point clouds for ground vehicles. In: Intelligent Vehicles Symposium (IV), 2010 IEEE. pp. 560–565. IEEE (2010) 7
9. Hoffmann, D.T., Raza, S.H., Jiang, H., Tananaev, D., Klingenhoefer, S., Meinke, M.: Floxels: Fast unsupervised voxel based scene flow estimation. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 22328–22337 (2025) 5
10. Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., Brox, T.: Flownet 2.0: Evolution of optical flow estimation with deep networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2462–2470 (2017) 2
11. Jund, P., Sweeney, C., Abdo, N., Chen, Z., Shlens, J.: Scalable scene flow from point clouds in the real world. IEEE Robotics and Automation Letters **7**(2), 1589–1596 (2021) 5
12. Khatri, I., Vedder, K., Peri, N., Ramanan, D., Hays, J.: I can't believe it's not scene flow! In: European Conference on Computer Vision. pp. 242–257. Springer (2024) 1, 5, 8, 9, 10
13. Khoche, A., Zhang, Q., Sanchez, L.P., Asefaw, A., Mansouri, S.S., Jensfelt, P.: SSF: Sparse long-range scene flow for autonomous driving. In: 2025 IEEE International Conference on Robotics and Automation (ICRA). pp. 6394–6400 (2025). https://doi.org/10.1109/ICRA55743.2025.11128770 5, 7, 9
14. Kim, J., Woo, J., Kim, J., Im, S.: Rethinking lidar domain generalization: Single source as multiple density domains. In: European Conference on Computer Vision. pp. 310–327. Springer (2024) 2, 6

15. Kim, J., Woo, J., Shin, U., Oh, J., Im, S.: Flow4D: Leveraging 4d voxel network for lidar scene flow estimation. IEEE Robotics and Automation Letters pp. 1–8 (2025). https://doi.org/10.1109/LRA.2025.3542327 3, 5, 7, 9
16. Lambert, J., Liu, Z., Sener, O., Hays, J., Koltun, V.: Mseg: A composite dataset for multi-domain semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2879–2888 (2020) 2
17. Lang, I., Aiger, D., Cole, F., Avidan, S., Rubinstein, M.: Scoop: Self-supervised correspondence and optimization-based scene flow. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5281–5290 (2023) 5
18. Li, X., Kaesemodel Pontes, J., Lucey, S.: Neural scene flow prior. Advances in Neural Information Processing Systems **34**, 7838–7851 (2021) 6, 9
19. Li, X., Zheng, J., Ferroni, F., Pontes, J.K., Lucey, S.: Fast neural scene flow. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9878–9890 (2023) 6, 9
20. Liang, Y., Badki, A., Su, H., Tompkin, J., Gallo, O.: Zero-shot monocular scene flow estimation in the wild. In: Proceedings of the Computer Vision and Pattern Recognition Conference (2025) 3
21. Lin, Y., Caesar, H.: Icp-flow: Lidar scene flow estimation with icp. In: CVPR (2024) 9
22. Lin, Y., Wang, S., Nan, L., Kooij, J., Caesar, H.: Voteflow: Enforcing local rigidity in self-supervised scene flow. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 17155–17164 (2025) 9
23. Liu, Y., Kong, L., Wu, X., Chen, R., Li, X., Pan, L., Liu, Z., Ma, Y.: Multi-space alignments towards universal lidar segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14648–14661 (2024) 2
24. Liu, Y., Kong, L., Wu, X., Chen, R., Li, X., Pan, L., Liu, Z., Ma, Y.: Multi-space alignments towards universal lidar segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 14648–14661 (June 2024) 6
25. Luo, J., Cheng, J., Tang, X., Zhang, Q., Xue, B., Fan, R.: Mambaflow: A novel and flow-guided state space model for scene flow estimation. arXiv preprint arXiv:2502.16907 (2025) 5
26. Madan, A., Peri, N., Kong, S., Ramanan, D.: Revisiting few-shot object detection with vision-language models. arXiv preprint arXiv:2312.14494 (2023) 2, 4
27. Malić, D., Fruhwirth-Reisinger, C., Schulter, S., Possegger, H.: Gblobs: Explicit local structure via gaussian blobs for improved cross-domain lidar-based 3d object detection. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 27357–27367 (2025) 6
28. Michele, B., Boulch, A., Puy, G., Vu, T.H., Marlet, R., Courty, N.: Saluda: Surface-based automotive lidar unsupervised domain adaptation. In: 2024 International Conference on 3D Vision (3DV). pp. 421–431. IEEE (2024) 6
29. Peri, N., Li, M., Wilson, B., Wang, Y.X., Hays, J., Ramanan, D.: An empirical analysis of range for 3d object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4074–4083 (2023) 4
30. Robicheaux, P., Popov, M., Madan, A., Robinson, I., Nelson, J., Ramanan, D., Peri, N.: Roboflow100-vl: A multi-domain object detection benchmark for vision-language models. arXiv preprint arXiv:2505.20612 (2025) 4
31. Saltori, C., Osep, A., Ricci, E., Leal-Taixé, L.: Walking your lidog: A journey through multiple domains for lidar semantic segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 196–206 (2023) 2

32. Soum-Fontez, L., Deschaud, J.E., Goulette, F.: Mdt3d: Multi-dataset training for lidar 3d object detection generalization. In: 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 5765–5772. IEEE (2023) 2
33. Sun, J., Qing, C., Xu, X., Kong, L., Liu, Y., Li, L., Zhu, C., Zhang, J., Xiao, Z., Chen, R., et al.: An empirical study of training state-of-the-art lidar segmentation models. arXiv preprint arXiv:2405.14870 (2024) 6
34. Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., Guo, J., Zhou, Y., Chai, Y., Caine, B., et al.: Scalability in perception for autonomous driving: Waymo open dataset. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2446–2454 (2020) 1, 2, 3
35. Teed, Z., Deng, J.: Raft: Recurrent all-pairs field transforms for optical flow. In: European conference on computer vision. pp. 402–419. Springer (2020) 2
36. Vedder, K., Peri, N., Chodosh, N., Khatri, I., Eaton, E., Jayaraman, D., Ramanan, Y.L.D., Hays, J.: ZeroFlow: Fast Zero Label Scene Flow via Distillation. International Conference on Learning Representations (ICLR) (2024) 1, 5
37. Vedder, K., Peri, N., Khatri, I., Li, S., Eaton, E., Kocamaz, M.K., Wang, Y., Yu, Z., Ramanan, D., Pehserl, J.: Neural eulerian scene flow fields. In: The Thirteenth International Conference on Learning Representations (2025), https://openreview.net/forum?id=0CieWy9ONY 1, 6, 9
38. Vedula, S., Rander, P., Collins, R., Kanade, T.: Three-dimensional scene flow. IEEE transactions on pattern analysis and machine intelligence **27**(3), 475–480 (2005) 5
39. Wang, C., Ma, C., Zhu, M., Yang, X.: Pointaugmenting: Cross-modal augmentation for 3d object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11794–11803 (2021) 6
40. Wang, J., Che, H., Chen, Y., Yang, Z., Goli, L., Manivasagam, S., Urtasun, R.: Flux4d: Flow-based unsupervised 4d reconstruction. In: The Thirty-ninth Annual Conference on Neural Information Processing Systems (2025), https://openreview.net/forum?id=FeUGQ6AiKR 5
41. Wang, R., Xu, S., Dong, Y., Deng, Y., Xiang, J., Lv, Z., Sun, G., Tong, X., Yang, J.: Moge-2: Accurate monocular geometry with metric scale and sharp details. arXiv preprint arXiv:2507.02546 (2025) 3
42. Wang, Y., Chen, X., You, Y., Li, L.E., Hariharan, B., Campbell, M., Weinberger, K.Q., Chao, W.L.: Train in germany, test in the usa: Making 3d object detectors generalize. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11713–11723 (2020) 2
43. Wang, Z., Wei, Y., Rao, Y., Zhou, J., Lu, J.: 3d point-voxel correlation fields for scene flow estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence (2023) 5
44. Wei, Y., Wang, Z., Rao, Y., Lu, J., Zhou, J.: PV-RAFT: Point-Voxel Correlation Fields for Scene Flow Estimation of Point Clouds. In: CVPR (2021) 5
45. Wilson, B., Qi, W., Agarwal, T., Lambert, J., Singh, J., et al.: Argoverse 2: Next generation datasets for self-driving perception and forecasting. In: Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks (NeurIPS Datasets and Benchmarks 2021) (2021) 1, 2, 3
46. Xu, X., Kong, L., Shuai, H., Zhang, W., Pan, L., Chen, K., Liu, Z., Liu, Q.: Superflow++: Enhanced spatiotemporal consistency for cross-modal data pretraining. arXiv preprint arXiv:2503.19912 (2025) 6

47. Zhang, Q., Khoche, A., Yang, Y., Ling, L., Sina, S.M., Andersson, O., Jensfelt, P.: HiMo: High-speed objects motion compensation in point cloud. arXiv preprint arXiv:2503.00803 (2025) 9
48. Zhang, Q., Yang, Y., Fang, H., Geng, R., Jensfelt, P.: DeFlow: Decoder of scene flow network in autonomous driving. In: 2024 IEEE International Conference on Robotics and Automation (ICRA). pp. 2105–2111 (2024). https://doi.org/10.1109/ICRA57147.2024.10610278 5, 9
49. Zhang, Q., Yang, Y., Li, P., Andersson, O., Jensfelt, P.: SeFlow: A self-supervised scene flow method in autonomous driving. In: European Conference on Computer Vision (ECCV). pp. 353–369. Springer (2024). https://doi.org/10.1007/978-3-031-73232-4_20 9
50. Zhang, Q., Zhu, X., Zhang, Y., Cai, Y., Andersson, O., Jensfelt, P.: DeltaFlow: An efficient multi-frame scene flow estimation method. arXiv preprint arXiv:2508.17054 (2025) 1, 5, 7, 9
51. Zhang, Y., Edstedt, J., Wandt, B., Forssén, P.E., Magnusson, M., Felsberg, M.: Gmsf: Global matching scene flow. Advances in Neural Information Processing Systems **36** (2024) 5