A Argoverse 2 2024 Scene Flow Challenge

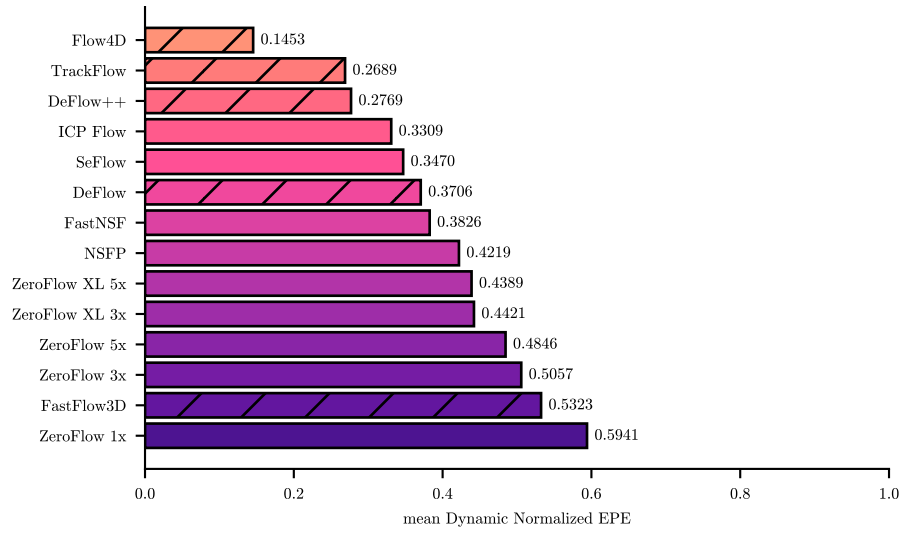


Fig. 9: mean Dynamic Normalized EPE of submissions to *the Argoverse 2 2024 Scene Flow Challenge* on Argoverse 2's *test* split. Supervised methods shown with hatching. Lower is better.

Bucket Normalized EPE was the basis for the *Argoverse 2 2024 Scene Flow Challenge*[4]. The competition featured two tracks: a supervised track, and an unsupervised track, with TrackFlow serving as a baseline in the supervised track. Leaderboards for both tracks are ranked by minimum *mean Dynamic* component of Bucket Normalized EPE.

Notably, Flow4D [14] significantly improved over all prior supervised methods, halving the dynamic error of TrackFlow. Interestingly, Flow4D does not feature any class-aware loss features, instead focusing on architectural improvements over FastFlow3D [12]-based architectures (e.g. ZeroFlow [42], DeFlow [53]). Unsupervised scene flow methods also saw meaningful improvements; ICP-Flow [24] significantly outperformed FastNSF [20], the best performing unsupervised baseline, closely followed by SeFlow [54].

B Bucket Normalized EPE Structure

Table 4 show the structure of the class-speed matrix of Bucket Normalized EPE on Argoverse 2.

[4] Full details about the competition can be found at http://argoverse.org/sceneflow

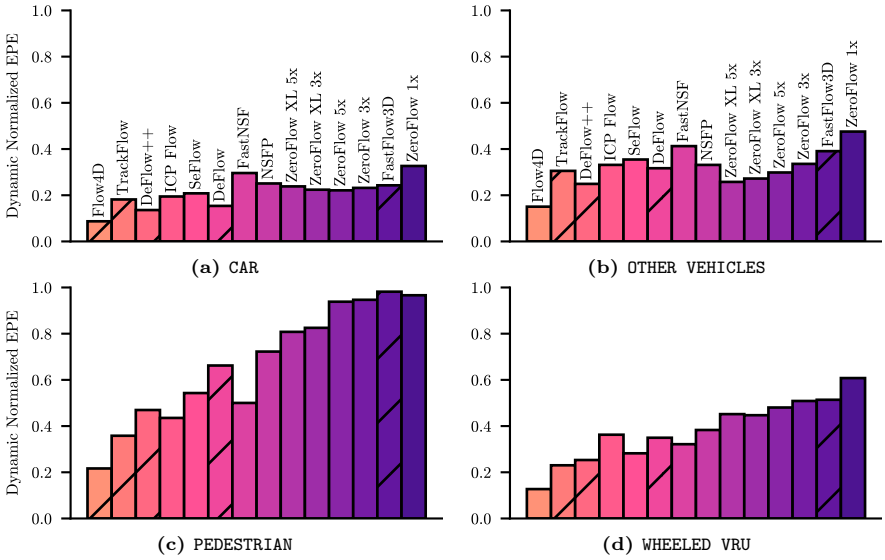


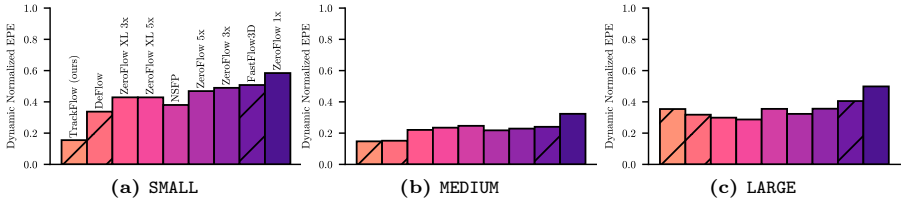
Fig. 10: Per meta-class Dynamic Normalized EPE of submissions to *the Argoverse 2 2024 Scene Flow Challenge* on Argoverse 2's *test* split. Supervised methods shown with hatching. Lower is better. Method color and position is consistent between plots.

Class	Speed Columns				
	0-0.4m/s	0.4-0.8m/s	0.8-1.2m/s	...	20-∞m/s
BACKGROUND	-	-	-	-	-
CAR	-	-	-	-	-
OTHER VEHICLES	-	-	-	-	-
PEDESTRIAN	-	-	-	-	-
WHEELED VRU	-	-	-	-	-

Table 4: Example of Bucket Normalized EPE's class-speed matrix.

C Bucket Normalized EPE Without Semantics

In Section 3 we present Bucket Normalized EPE with the object distribution broken down by semantic classes. While this makes sense when semantics are available, this is not a fundamental requirement for Bucket Normalized EPE. To demonstrate this, we break down Argoverse 2's bounding boxes by *size* instead of semantics. We group the ground truth boxes into one of three volume based clusters: SMALL: $< 9.5m^3$, MEDIUM: $\geq 9.5m^3 \wedge < 40m^3$, or LARGE: $\geq 40m^3$. As shows in Fig. 11, this distribution breakdown still highlights the poor performance of prior art on small objects.



(a) `SMALL` **(b)** `MEDIUM` **(c)** `LARGE`

Fig. 11: Bucket Normalized EPE using ground truth size based clustering.

D FAQ

D.1 TrackFlow is *just* a tracking method

Yes, TrackFlow is a tracking method applied to the scene flow problem. The state-of-the-art performance of TrackFlow suggests that Scene Flow via Tracking is a fruitful area of exploration for future work on supervised scene flow.

D.2 TrackFlow uses bounding boxes and thus can only estimate rigid flow — what does this paper have to say about non-rigid scene flow?

It's true that TrackFlow operates on the level of bounding boxes, but as we discuss in Section 2.1, public real-world datasets derive motion annotations from bounding box tracks. If non-rigid labels were available, one could train a detector to also regress keypoints (or use an off-the-shelf pretrained method [51]) and track across those keypoints.

D.3 TrackFlow uses bounding boxes from a detector — does this mean it cannot detect open-set objects?

TrackFlow uses a class-aware object detector as its bounding box proposer. However, the Scene Flow via Tracking framework does not require class annotations – nothing prevents the use of a class agnostic open world bounding box proposer, either trained like FasterRCNN's RPN [8,39], Object Localization Network [13], or via geometric priors [11].

D.4 Our metric is "just" Threeway EPE extended to multiple classes and multiple speed buckets with normalization, and our method "just" combines a detector and tracker. Where is the novelty in this idea?

The ideas presented in this paper are simple and post-hoc obvious, but serve to highlight catastrophic failures currently overlooked in existing approaches.