Supplement: Just Go with the Flow: Self-Supervised Scene Flow Estimation

Himangi Mittal
Carnegie Mellon University
hmittal@andrew.cmu.edu

Brian Okorn
Carnegie Mellon University
bokorn@andrew.cmu.edu

David Held
Carnegie Mellon University
dheld@andrew.cmu.edu

Self-Supervised Training (nuScenes + KITTI)						
NN Loss	Cycle Loss	Anchor	Flip	EPE (m)↓	ACC (0.05)↑	ACC (0.1)↑
	✓	✓	✓	0.1768	15.90	35.81
✓			✓	0.1102	30.80	73.27
✓	✓		✓	0.1493	22.97	49.58
✓	✓	✓		0.1072	40.03	72.20
✓	✓	✓	✓	**0.1053**	**46.48**	**79.42**
Self-supervised (nuScenes) + Supervised Training (KITTI)						
NN Loss	Cycle Loss	Anchor	Flip	EPE (m)↓	ACC (0.05)↑	ACC (0.1)↑
	✓	✓	✓	0.1572	18.50	52.80
✓			✓	0.1090	34.88	71.32
✓	✓		✓	0.0932	28.18	66.10
✓	✓	✓		0.0926	40.69	74.50
✓	✓	✓	✓	**0.0912**	**47.92**	**79.63**

Table 1: Leave-one-out ablation analysis: We study the effect of removing a single component of self-supervised loss and data augmentation. Top: Models use self-supervised training on nuScenes and KITTI; Bottom: Models use self-supervised training on nuScenes followed by supervised training on KITTI.

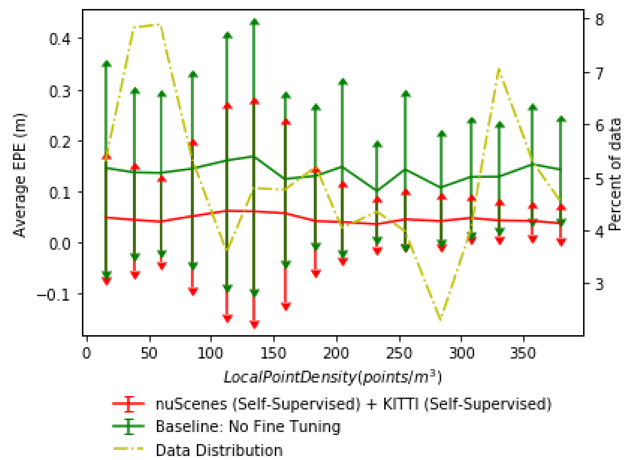

Figure 1: Analysis of average EPE (m) with respect to the local point density (points/m^3). Flow estimates are binned by local point density and a confidence interval of 95% is shown for all results.

1. Hyperparameters

A batch size of 8 is used for all datasets. All models are trained using the Adam optimizer with a learning rate of 1e-4, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 1e-8$.

2. Further Ablation Studies

In this section, we extend our ablation study to further evaluate the importance of each component used in our method. Beginning from our full method, we remove a single component to see the change on the evaluation metrics.

The results can be seen in Table 1. Because "anchoring" is a modification of the cycle loss, when Cycle Consistency Loss is removed, anchoring must be removed too. The results in Table 1 show that each component is important for the performance of our method, and removing any of the components results in a drop in performance. This can especially be seen in the accuracy performance metrics (ACC 0.05 and 0.1), which show large drops when any component is removed.

3. Analysis of Point Density vs End Point Error

We analyze the correlation between local point density and endpoint error. For each point in the point cloud at time t, we compute the number of points within a 0.1m radius neighborhood of that point. Each point is binned based on the density of its local neighborhood. For each bin, the mean end-point error (EPE) is computed for both the baseline, trained only synthetic data, and our method, trained on nuScenes and KITTI using our self-supervised losses. Figure 1 shows no correlation between EPE and neighborhood density for either the baseline or for our method.

4. Error Distribution of End Point Error

We analyze the error distributions of our method and a baseline method. Our method uses self-supervised training on nuScenes followed by supervised training on KITTI. The baseline is trained only on KITTI using supervised learning. By computing the error at every point of every scan in the KITTI test set, we can view the full distribution of end point

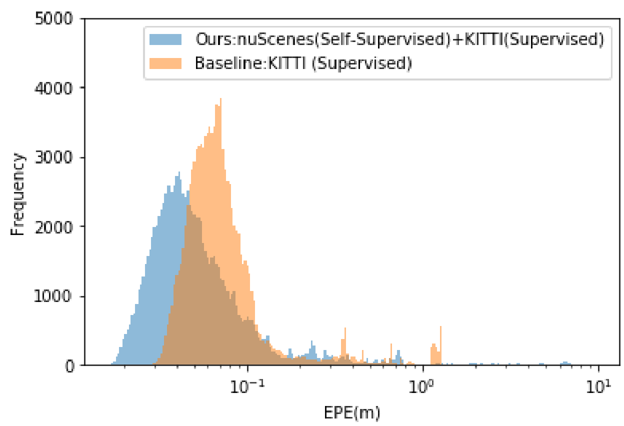

Figure 2: End point error distribution of our method (blue) and the baseline (orange).

errors (EPE), shown in Figure 2. To better show the effects of outliers, we use log binning for the x-axis. Not only is the center of our error distribution lower, with an average EPE of 0.091, but it also shows fewer large outliers than the baseline method.