Exploring Semi-Supervised Learning for Online Mapping

Supplementary Material

6. Exact numbers

For completeness, we also show the exact performance numbers for the results shown in Figures in Sec. 4. For thresholds and temperatures, the numbers are displayed in Tab. 5 and Tab. 6 for 2.5% of the labels, respectively. For 10%, we refer to Tab. 7 and Tab. 8. In addition to these, the performance for feature similarity loss with different weights is shown in Tab. 9 and Tab. 10 for 2.5% and 10% label utilisation, respectively. The validation performance for teacher fusion over various ranges is provided in Tab. 11 and Tab. 12 for 2.5% and 10% label utilisation. We also experiment with the amount of samples to use in addition to the center sample in Tab. 13. For the test performance across multiple label utilisations, the numbers for both Argoverse 2 (AV2) and nuScenes are shown in Tab. 14. Finally, Tab. 15 provides the performance for city adaptation for both Argoverse 2 and nuScenes when varying the amount of unlabelled data.

Table 5. Validation performance (mIoU) for different thresholds on Argoverse 2 2.5% of the labels utilised.

Thresholds	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9
Thr.	24.1	26.8	25.9	23.6	25.7	23.6	24.6	22.3
Thr.+Hard	24.4	26.6	25.2	25.5	24.2	25.0	23.7	24.6

Table 6. Validation performance (mIoU) for using sharpening with different temperatures on Argoverse 2 2.5% of the labels utilised.

Temperature	0.05	0.1	0.25	0.5	0.75	0.95
No Thr.	27.6	25.3	27.4	27.1	27.0	26.1
Thr.	24.3	26.2	27.4	25.3	25.7	26.2
Thr.+Hard	24.2	25.1	25.3	26.3	26.9	24.4

Table 7. Validation performance (mIoU) for different thresholds on Argoverse 2 10% of the labels utilised.

Thresholds	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9
Thr.	33.6	33.9	33.4	33.3	33.0	32.4	32.1	30.9
Thr.+Hard	34.8	34.7	34.4	33.7	33.3	32.3	31.4	31.4

Table 8. Validation performance (mIoU) for using sharpening with different temperatures on Argoverse 2 10% of the labels utilised.

Temperature	0.05	0.1	0.25	0.5	0.75	0.95
No Thr.	35.0	34.8	34.9	34.7	33.8	34.6
Thr.	34.2	35.3	34.8	34.2	32.9	33.5
Thr.+Hard	34.6	34.8	34.4	34.9	33.9	34.1

Table 9. Validation performance (mIoU) for feature similarity loss with different weights on Argoverse 2 2.5% of the labels utilised.

Weight	0.05	0.1	0.25	0.5	1.0	1.5	2.0	2.5
MSE-pre	24.3	25.4	26.0	22.2	23.3	20.0	23.1	22.5
MSE-post	24.3	22.5	22.5	23.1	24.3	22.5	23.4	24.1
COS-pre	25.3	25.3	24.8	25.0	23.4	24.9	24.6	22.9
COS-post	23.4	25.1	23.4	23.6	25.5	26.1	24.4	25.4

Table 10. Validation performance (mIoU) for feature similarity loss with different weights on Argoverse 2 10% of the labels utilised.

Weight	0.05	0.1	0.25	0.5	1.0	1.5	2.0	2.5
MSE-pre	33.7	32.8	32.3	32.5	32.4	30.0	28.8	24.1
MSE-post	33.6	33.4	32.3	31.9	31.2	29.3	28.0	28.0
COS-pre	34.4	33.8	34.4	34.4	34.1	33.8	34.4	33.2
COS-post	34.0	34.8	35.4	35.1	34.4	34.8	33.4	31.4

Table 11. Validation performance (mIoU) for teacher fusion over various ranges on Argoverse 2 2.5% of the labels utilised.

Ranges	10	15	20	25	30	35	40
Probs.	27.7	27.7	27.2	27.6	28.4	28.7	26.8
Probs.+Thr.	25.5	25.9	27.3	28.2	26.9	27.5	26.7
Feats.	24.1	25.6	24.8	23.6	23.1	25.8	26.2
Feats.+Thr.	26.5	26.6	27.0	26.5	26.1	26.7	25.6

Table 12. Validation performance (mIoU) for teacher fusion over various ranges on Argoverse 2 10% of the labels utilised.

Ranges	10	15	20	25	30	35	40
Probs.	36.2	35.5	36.5	36.1	35.5	35.8	35.9
Probs.+Thr.	35.9	36.2	36.4	35.4	35.7	35.8	36.0
Feats.	34.0	34.5	34.2	34.4	34.3	34.0	34.4
Feats.+Thr.	34.2	34.0	34.8	34.1	34.4	35.4	34.6

Table 13. Multi-step Fusion with varying numbers of additional samples within a 20m range using 10% of labels. The best results are achieved with 4 additional samples.

	Additional samples		
	2	4	6
Probs	34.6	35.0	30.2
Feats	32.4	32.8	26.1

Table 14. Test performance (mIoU) for multiple label utilisations for both Argoverse 2 (AV2) and nuScenes (nuSc).

	+Unlbld Data	Label utilisation 2.5%	5%	10%	25%	50%	100%
AV2	✗	$7.5^{\pm 0.9}$	$9.4^{\pm 0.1}$	$10.6^{\pm 0.2}$	$23.4^{\pm 0.5}$	$33.8^{\pm 0.7}$	$40.2^{\pm 0.7}$
	✓	$\mathbf{28.9}^{\pm 1.2}$	$\mathbf{33.8}^{\pm 0.6}$	$\mathbf{36.7}^{\pm 0.9}$	$\mathbf{37.4}^{\pm 0.8}$	$\mathbf{39.6}^{\pm 0.3}$	–
nuSc	✗	$5.6^{\pm 0.3}$	$7.2^{\pm 2.3}$	$16.6^{\pm 3.0}$	$24.5^{\pm 2.2}$	$27.4^{\pm 2.7}$	$28.6^{\pm 2.4}$
	✓	$\mathbf{15.5}^{\pm 1.2}$	$\mathbf{22.0}^{\pm 1.6}$	$\mathbf{25.1}^{\pm 1.9}$	$\mathbf{27.4}^{\pm 1.8}$	$27.8^{\pm 2.8}$	–

Table 15. Performance (mIoU) for city adaptaion for both Argoverse 2 and nuScenes when increasing amount of unlabelled data.

	Folds	# unlabelled target sequences 0	25	50	100	200	Using target labels
AV2	PIT-MIA → Rest	33.6	34.4	35.5	37.5	37.2	38.9
	MIA-Rest → PIT	36.3	38.7	40.0	40.2	40.7	41.2
	PIT-Rest → Mia	34.5	36.1	37.5	37.8	38.5	40.2
nuSc	Boston → Singapore	11.1	12.6	15.3	17.0	17.7	27.0