Supplementary Material FAST3D: Flow-Aware Self-Training for 3D Object Detectors

Christian Fruhwirth-Reisinger[1,2]
christian.reisinger@icg.tugraz.at
Michael Opitz[1,2]
micopitz@amazon.de
Horst Possegger[2]
possegger@icg.tugraz.at
Horst Bischof[1,2]
bischof@icg.tugraz.at

[1] Christian Doppler Laboratory for Embedded Machine Learning
[2] Institute of Computer Graphics and Vision
Graz University of Technology

In the following, we present additional evaluations of our flow-aware self-training approach. In particular, we demonstrate the improved pseudo-label quality (Sec. 1) and present a detailed ablation study (Sec. 2).

1 Pseudo-Label Quality

Table 1 demonstrates the effectiveness of our pseudo-label refinement for unsupervised domain adaptation of PointRCNN from KITTI [1, 2] to Waymo Open Dataset (WOD) [2]. We show the detection quality of the initial model Φ^{src}, the re-trained model after the first re-training cycle Φ_1^{tar}, both followed by their corresponding refined pseudo-labels. This means that the target model Φ_1^{tar} has been re-trained with the refined pseudo-labels of Φ^{src}. Re-training Φ_1^{tar} with its refined pseudo-labels results in the final target model Φ_2^{tar} as reported in our main manuscript. Since the training results saturate and do not change significantly anymore, we omit these from the table for the sake of readability.

We report the pseudo-label quality in terms of true positives (TP) and false positives (FP). We omit false negatives (FN), as these can be easily calculated by subtracting TPs from the total number of instances. We report these scores for three different intersection over union (IoU) thresholds. On the one hand, this shows that we can generate a lot of high-quality labels (IoU $\geq 0.7/0.5$). On the other hand, lower IoU thresholds show the ability to collect additional new labels albeit with less overlap.

We observe that the initial model Φ^{src} has low recall for high quality labels, *e.g.* 53984 out of 303204 *vehicle* instances (*i.e.* only 17.8%), and produces a high number of FPs. With FAST3D, we can significantly improve TPs for the same class **by 159.1%** (*i.e.* 139868 pseudo-labels) while decreasing the FPs **by 36.5%**, **already in the first self-training cycle**. Although the TP improvement for *cyclists* and *pedestrians* is smaller (64.0% and 23%, respectively), it is still remarkable, also in terms of FP reduction (7.4% and 67.0%). Addi-

	Class	# True Positive (TP) ↑			# False Positive (FP) ↓		
		IoU 0.7 / 0.5	**0.5 / 0.25**	**0.25 / 0.15**	**0.7 / 0.5**	**0.5 / 0.25**	**0.25 / 0.15**
				1^{st} Self-training cycle			
Φ^{src}	Vehicle	53984	153444	172505	137984	38524	19467
	Pedestrian	22133	30796	31380	14772	6177	5818
	Cyclist	883	1154	1175	1473	1202	1181
FAST3D	Vehicle	139868 (+159.1%)	210476 (+37.2%)	219858 (+27.5%)	87567 (-36.5%)	16959 (-56.0%)	7577 (-61.1%)
	Pedestrian	36289 (+ 64.0%)	47663 (+54.8%)	49431 (+57.5%)	13685 (- 7.4%)	2614 (-57.7%)	1614 (-72.3%)
	Cyclist	1086 (+ 23.0%)	1472 (+27.6%)	1529 (+30.1%)	486 (-67.0%)	100 (-91.7%)	43 (-96.4%)
				2^{nd} Self-training cycle			
Φ_1^{tar}	Vehicle	149874	197187	203850	100115	52802	46141
	Pedestrian	37130	41124	41725	19731	15942	15926
	Cyclist	1687	2130	2188	3499	3056	2997
FAST3D	Vehicle	163556 (+ 9.1%)	225668 (+14.4%)	236592 (+16.1%)	81536 (-18.6%)	19424 (-63.2%)	8500 (-81.6%)
	Pedestrian	48029 (+29.4%)	60668 (+47.5%)	62857 (+50.6%)	16379 (-17.0%)	4127 (-74.1%)	3206 (-79.9%)
	Cyclist	1749 (+ 3.7%)	2306 (+ 8.3%)	2403 (+ 9.8%)	1083 (-69.0%)	526 (-82.8%)	429 (-85.7%)

Table 1: Pseudo-label quality for the PointRCNN detector. We use 200 WOD sequences (∼ 25%) from the official training split for pseudo-label collection and re-training. Total number of instances: *vehicle* (303204), *pedestrian* (172377), *cyclist* (4203). We list three different IoU thresholds (denoted IoU *vehicle* / IoU *pedestrian* and *cyclist*). Results are shown for two self-training cycles: First, using the initial model Φ^{src} and second, using Φ_1^{tar}, *i.e.* re-trained with the refined pseudo-labels of the first cycle. Improvements (in parentheses) are relative to the respective previous detection step.

tionally, we observe that the number of FPs decreases with relaxing the IoU threshold. This implies that overall, our approach finds a high number of TP bounding boxes.

After the first self-training cycle, the target model Φ_1^{tar} achieves a much better detection performance compared to the initial model Φ^{src}, *e.g.* **177.6**% increase *w.r.t.* TPs for the *vehicle* class (IoU ≥ 0.7). Since training with pseudo-labels introduces incorrect labels, the performance *w.r.t.* FPs degrades slightly. However, applying FAST3D again has a positive effect on FPs and FNs.

For the sake of completeness, we also show the results in terms of AP_{BEV}/AP_{3D} for the *vehicle* class and two IoU thresholds (0.7, 0.5) in Table 2. We again see that the last refinement step shows the best results. Note that the AP metrics consider the confidence scores which we currently don't use in our self-training approach. Instead, we handle all refined pseudo-labels as if they have a confidence score of 1.0.

	IoU 0.7				IoU 0.5			
	Overall 0m - 75m	**Near** 0m - 30m	**Medium** 30m - 50m	**Far** 50m - 75m	**Overall** 0m - 75m	**Near** 0m - 30m	**Medium** 30m - 50m	**Far** 50m - 75m
1^{st} Self-training cycle								
Φ^{src}	29.5 / 8.8	45.7 / 14.8	31.9 / 8.6	11.3 / 2.1	58.5 / 52.4	78.0 / 73.0	66.2 / 59.4	35.2 / 26.0
FAST3D	62.3 / 34.4	74.1 / 46.5	63.1 / 34.0	43.0 / 19.4	73.5 / 71.3	87.5 / 86.5	77.2 / 74.9	55.6 / 49.6
2^{nd} Self-training cycle								
Φ_1^{tar}	65.8 / **45.0**	82.9 / **62.7**	70.4 / **49.2**	42.0 / **24.2**	72.2 / 69.0	88.3 / 87.8	76.9 / 73.8	50.8 / 47.3
FAST3D	**73.6** / 39.8	**86.4** / 51.1	**73.3** / 43.3	**52.6** / 23.9	**83.0 / 80.6**	**92.9 / 92.5**	**82.4 / 81.1**	**70.6 / 61.8**

Table 2: Pseudo-label quality for the PointRCNN detector in terms of AP_{BEV}/AP_{3D} for the *vehicle* class and two different IoU thresholds (0.7, 0.5) on KITTI→WOD.

2 Ablation Study

We conduct a detailed ablation study to show the contribution of each step of our approach. Table 3 lists the results for the *vehicle* class on the same KITTI→WOD scenario as our previous experiments. In particular, we evaluate the following configurations:

a) denotes the performance of the initial PointRCNN model Φ^{src}, trained only on the source (*i.e.* KITTI) data.

b) adds test-time augmentation, where we only need 2 additional scales in contrast to [3].

c) additionally considers only highly confident detections as pseudo-labels (standard self-training technique to reduce FP).

d) adds our flow-aware pseudo-label propagation step to increase the recall of high quality labels.

e) includes bounding box correction, where we update box sizes and additionally positions for static objects in order to enhance IoU overlaps.

f) additionally applies track filtering to remove unreliable pseudo-labels.

g) uses *flow consistency* to recover incorrectly filtered pseudo-labels.

h) finally adds backward completion to mitigate late track initialisation due to potentially missed detections.

As shown by the results, all steps contribute to the effectiveness of our flow-aware self-training approach, allowing us to surpass the current state-of-the-art in unsupervised domain adaptation.

	Configuration	# True Positive (TP) ↑		# False Positive (FP) ↓		Precision ↑		Recall ↑		AP_{3D}
		0.7	**0.25**	**0.7**	**0.25**	**0.7**	**0.25**	**0.7**	**0.25**	**0.7**
a)	Φ^{src}	53984	172505	137984	19467	28.1	89.9	17.8	56.9	8.79
b)	Φ^{src} + TTA	109727	199079	117927	28578	48.2	87.5	36.2	65.7	25.73
c)	Φ^{src} + TTA + T	106152	178934	77239	**4457**	57.9	97.6	35.1	59.0	25.73
d)	Φ^{src} + TTA + T + P	118063	209221	116174	25016	50.4	89.3	38.9	69.0	29.29
e)	Φ^{src} + TTA + T + P + C	132066	209235	101977	24808	56.4	89.4	43.6	69.0	35.18
f)	Φ^{src} + TTA + T + P + C + F	129133	196056	**71685**	4762	**64.3**	**97.6**	42.6	64.7	35.19
g)	Φ^{src} + TTA + T + P + C + F + R	129212	196453	72118	4877	64.2	**97.6**	42.6	64.8	**35.20**
h)	Φ^{src} + TTA + T + P + C + F + R + B	**139868**	**219858**	87567	7577	61.5	96.7	**46.1**	**72.5**	34.40

Table 3: Ablation study with PointRCNN for the *vehicle* class (in total 303204 instances) at the first self-training cycle of KITTI→WOD. We denote the different contributions as: Test-time augmentation (TTA), high confidence thresholding (T), label propagation (P), bounding box correction (C), track filtering (F), recovery (R) and backward tracking (B).

References

[1] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In *Proc. CVPR*, 2012.

[2] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets Robotics: The KITTI Dataset. *IJRR*, 2013.

[3] Cristiano Saltori, Stéphane Lathuiliére, Nicu Sebe, Elisa Ricci, and Fabio Galasso. SF-UDA3D: Source-Free Unsupervised Domain Adaptation for LiDAR-Based 3D Object Detection. In *Proc. 3DV*, 2020.

[4] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, Vijay Vasudevan, Wei Han, Jiquan Ngiam, Hang Zhao, Aleksei Timofeev, Scott Ettinger, Maxim Krivokon, Amy Gao, Aditya Joshi, Yu Zhang, Jonathon Shlens, Zhifeng Chen, and Dragomir Anguelov. Scalability in Perception for Autonomous Driving: Waymo Open Dataset. In *Proc. CVPR*, 2020.