Towards Unsupervised Online Domain Adaptation for Semantic Segmentation: Supplementary Materials

Yevhen Kuznietsov[1] Marc Proesmans[1] Luc Van Gool[1,2]
[1]KU Leuven/ESAT-PSI [2]ETH Zurich/CVL
{yevhen.kuznietsov, marc.proesmans, luc.vangool}@esat.kuleuven.be

Contents

References

[1] Hassan Alhaija, Siva Mustikovela, Lars Mescheder, Andreas Geiger, and Carsten Rother. Augmented reality meets computer vision: Efficient data generation for urban driving scenes. *International Journal of Computer Vision (IJCV)*, 2018.

[2] Yohann Cabon, Naila Murray, and Martin Humenberger. Virtual kitti 2, 2020.

[3] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016.

[4] Adrien Gaidon, Qiao Wang, Yohann Cabon, and Eleonora Vig. Virtual worlds as proxy for multi-object tracking analysis. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 4340–4349, 2016.

[5] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *International Journal of Robotics Research (IJRR)*, 2013.

[6] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

1. Virtual KITTI v1 vs v2

While Virtual KITTI v1 [4] and v2 [4, 2] are very similar, we noticed that the labeling policy of v1 for such classes as "pole", "light" and "sign" is inconsistent with KITTI [1] labeling policy for all 5 videos (see Fig. 1). We observed a similar inconsistency for only one (the shortest) video of Virtual KITTI v2. In particular, if a pole is attached to a traffic sign or to a traffic light in Virtual KITTI v1, it is labeled as "sign" or "light" respectively, which is different for KITTI and most videos of Virtual KITTI v2. The labeling policy for other classes seems to be identical in Virtual KITTI v1 and v2.

Thus, for our resource-expensive ablation experiments, we selected Virtual KITTI v2 as a source dataset with more target-consistent labeling.

Figure 1. Examples of GT labeling for 5 videos from KITTI (left), Virtual KITTI v1 (middle) and Virtual KITTI v2 (right).

2. KITTI Frame Mapping: Semantic Segmentation to Raw and Odometry

Semantic segmentation ID	KITTI subset	Sequence name	Frame ID
2	Raw [5]	09-26-0005	10
3	Raw	09-26-0005	59
7	Raw	09-26-0009	354
8	Raw	09-26-0009	364
9	Raw	09-26-0009	374
10	Raw	09-26-0009	384
11	Raw	09-26-0009	394
12	Raw	09-26-0009	414
13	Raw	09-26-0011	111
14	Raw	09-26-0011	127
15	Raw	09-26-0011	147
16	Raw	09-26-0011	157
17	Raw	09-26-0011	167
18	Raw	09-26-0013	10
19	Raw	09-26-0013	20
20	Raw	09-26-0013	40
21	Raw	09-26-0013	70
22	Raw	09-26-0014	10
23	Raw	09-26-0014	20
24	Raw	09-26-0014	30
25	Raw	09-26-0014	50
26	Raw	09-26-0014	60
27	Raw	09-26-0014	129
28	Raw	09-26-0014	141
29	Raw	09-26-0014	152
30	Raw	09-26-0014	172
31	Raw	09-26-0014	192
32	Raw	09-26-0014	213
33	Raw	09-26-0014	240
34	Raw	09-26-0015	187
35	Raw	09-26-0015	197
36	Raw	09-26-0015	209
37	Raw	09-26-0015	219
38	Raw	09-26-0015	229
39	Raw	09-26-0015	239
40	Raw	09-26-0015	264
41	Raw	09-26-0015	273
42	Raw	09-26-0015	286
43	Raw	09-26-0017	10
44	Raw	09-26-0017	30
45	Raw	09-26-0017	40
46	Raw	09-26-0017	50
47	Raw	09-26-0018	46
48	Raw	09-26-0018	66
49	Raw	09-26-0018	76
50	Raw	09-26-0018	86
51	Raw	09-26-0018	96
52	Raw	09-26-0018	106

Semantic segmentation ID	KITTI subset	Sequence name	Frame ID
53	Raw	09-26-0018	133
54	Raw	09-26-0019	30
55	Raw	09-26-0019	87
56	Raw	09-26-0019	97
57	Raw	09-26-0022	634
58	Raw	09-26-0022	644
59	Raw	09-26-0022	654
60	Raw	09-26-0027	53
61	Raw	09-26-0027	103
62	Raw	09-26-0028	71
63	Raw	09-26-0028	118
64	Raw	09-26-0028	228
65	Raw	09-26-0028	269
66	Raw	09-26-0028	284
67	Raw	09-26-0028	303
68	Raw	09-26-0028	313
69	Raw	09-26-0028	378
70	Raw	09-26-0029	16
71	Raw	09-26-0029	123
72	Raw	09-26-0032	95
73	Raw	09-26-0032	114
74	Raw	09-26-0032	125
75	Raw	09-26-0032	207
76	Raw	09-26-0032	218
77	Raw	09-26-0032	330
78	Raw	09-26-0032	340
79	Raw	09-26-0032	350
80	Raw	09-26-0032	360
81	Raw	09-26-0032	378
83	Raw	09-26-0036	54
84	Raw	09-26-0036	402
85	Raw	09-26-0046	52
86	Raw	09-26-0046	62
88	Raw	09-26-0051	23
89	Raw	09-26-0051	218
90	Raw	09-26-0051	230
91	Raw	09-26-0051	282
92	Raw	09-26-0051	292
93	Raw	09-26-0051	302
94	Raw	09-26-0051	312
95	Raw	09-26-0051	322
96	Raw	09-26-0051	342
97	Raw	09-26-0051	356
98	Raw	09-26-0051	379
105	Raw	09-26-0056	10
106	Raw	09-26-0056	82
107	Raw	09-26-0056	122
108	Raw	09-26-0056	132
109	Raw	09-26-0056	191
110	Raw	09-26-0056	201
111	Raw	09-26-0056	282
112	Raw	09-26-0057	125

Semantic segmentation ID	KITTI subset	Sequence name	Frame ID
113	Raw	09-26-0057	140
114	Raw	09-26-0057	176
115	Raw	09-26-0057	299
116	Raw	09-26-0057	319
117	Raw	09-26-0057	339
118	Raw	09-26-0059	26
119	Raw	09-26-0059	46
120	Raw	09-26-0059	137
121	Raw	09-26-0059	150
122	Raw	09-26-0059	260
123	Raw	09-26-0059	280
124	Raw	09-26-0059	290
125	Raw	09-26-0059	300
126	Raw	09-26-0059	310
127	Raw	09-26-0059	320
128	Raw	09-26-0070	69
129	Raw	09-26-0070	224
130	Raw	09-26-0084	84
131	Raw	09-26-0084	179
132	Raw	09-26-0084	238
141	Raw	09-26-0096	20
142	Raw	09-26-0096	278
143	Raw	09-26-0096	381
144	Raw	09-26-0101	109
145	Raw	09-26-0101	175
146	Raw	09-26-0101	447
147	Raw	09-26-0101	457
148	Raw	09-26-0101	809
149	Raw	09-26-0104	15
150	Raw	09-26-0104	35
155	Raw	09-28-0002	343
157	Raw	09-29-0004	36
158	Raw	09-29-0004	79
159	Raw	09-29-0004	94
160	Raw	09-29-0004	105
161	Raw	09-29-0004	162
162	Raw	09-29-0004	258
163	Raw	09-29-0004	285
164	Raw	09-29-0004	308
168	Raw	09-29-0071	59
169	Raw	09-29-0071	943
180	Odometry [6]	17	14
181	Odometry	17	45
182	Odometry	17	55
183	Odometry	17	94
184	Odometry	17	229
185	Odometry	17	297
199	Raw	10-03-47	556

3. Weather Conditions: Detailed Results

	Model	Terrain	Sky	Veg.	Tree	Build.	Road	T.sign	T.light	Pole	Truck	Car	Van	Mean
Fog	NA	88.4	51.9	52.9	66.9	28.2	94.9	74.0	23.5	46.1	11.2	91.3	63.8	57.8
	OA	**93.7**	**68.2**	**74.0**	**79.5**	**73.1**	**97.3**	**78.5**	**32.3**	**49.8**	**47.9**	**91.5**	**70.2**	**71.4**
Rain	NA	95.4	70.5	89.5	78.9	47.4	97.9	83.7	29.9	41.2	**56.4**	**94.5**	**80.8**	72.2
	OA	**96.2**	**95.6**	**94.7**	**89.4**	**88.0**	**98.4**	**89.2**	**46.8**	**62.3**	53.1	93.0	79.4	**82.2**

Table 1. Semantic segmentation results for online UDA for changing weather conditions in Virtual KITTI v2. Left column specifies target domain weather. NA stands for non-adapted model trained under normal weather conditions, OA – for online adaptation from normal conditions to the specified target domain.

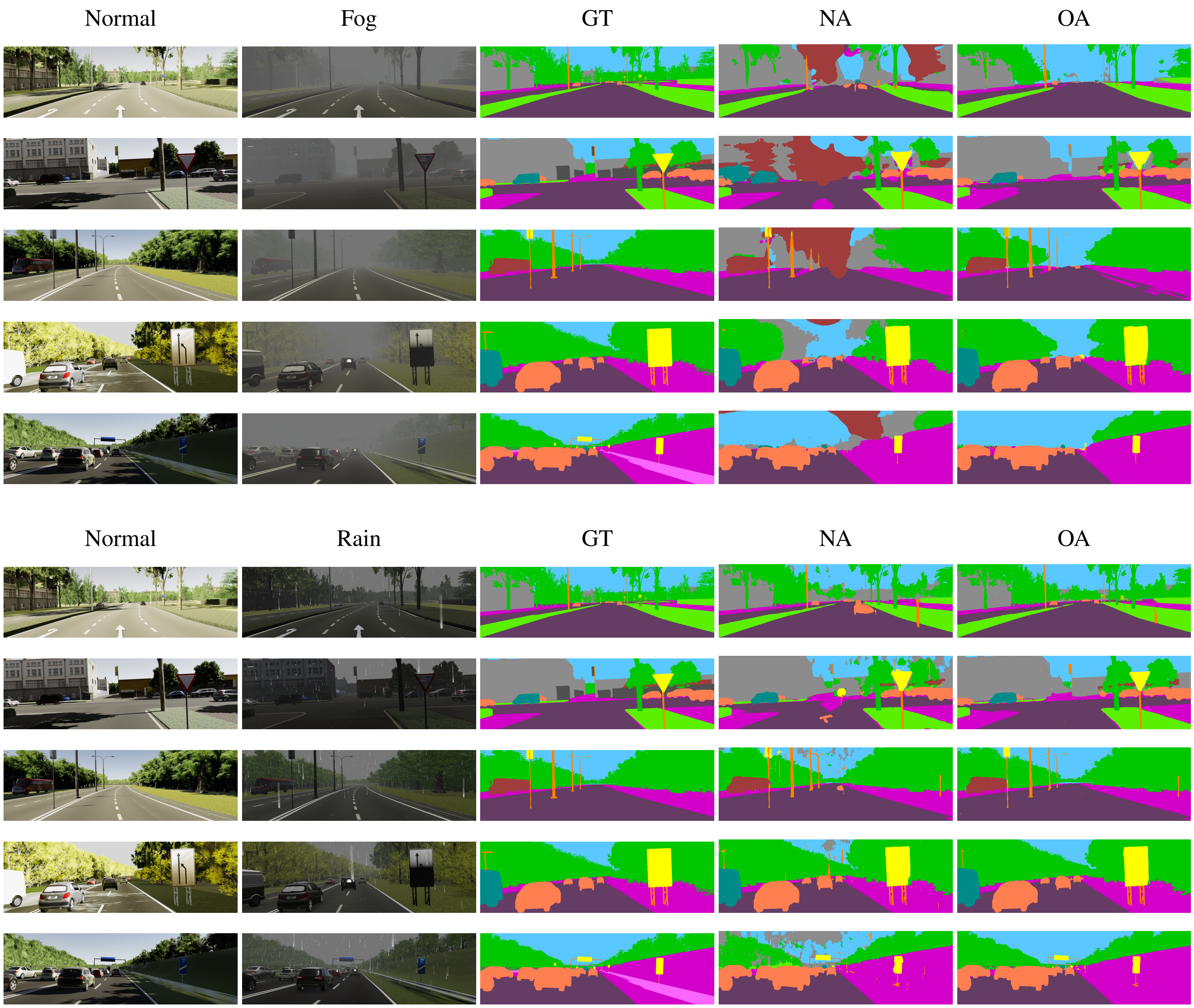

Figure 2. Qualitative results for various weather conditions.

4. Cityscapes: Qualitative Results

Input	GT	Source	Online

Figure 3. Qualitative results for Cityscapes [3] Frankfurt video.

5. Post-Adaptation

As mentioned in the paper, we describe another interesting setup for online UDA – post-adaptation. In particular, given a model $\mathcal{M}_a$ already offline-adapted from source domain to target domain train set, one can further adapt it to target domain test set videos using the online protocol described in the paper (except that the replay buffer is initialized using target domain train set).

Table 2 demonstrates the results of post-adaptation (Virtual KITTI v2 $\rightarrow_{offline}$ KITTI train $\rightarrow_{online}$ KITTI test) compared to offline baseline presented in the paper (Virtual KITTI v2 $\rightarrow_{offline}$ KITTI train). While 0.4% mean IoU improvement is not much, one should take into account that this result was obtained after approximately 200 backpropagation iterations on average, while offline baseline was already trained for 40 epochs on target domain train set. Moreover, the domain shift between train and test set of KITTI is small. As shown in Table 2, post-adaptation never caused mean IoU drop in 10 runs, with different random seeds used for experience sampling (the random seeds were also randomly selected).

	Model	Road	Building	Pole	T.light	T.sign	Vegetation	Terrain	Sky	Car	Truck	Mean
Avg.	Baseline	**89.7**	**66.0**	**40.1**	34.1	30.4	**83.1**	62.1	90.6	*83.4*	9.9	59.0
	Baseline+PA	89.5	65.4	38.7	**34.9**	**35.0**	82.9	**62.4**	**90.7**	*83.4*	**10.7**	**59.4**
Baseline+PA	Seed 0	90.4	65.6	38.9	34.6	35.2	82.7	63.7	90.7	83.0	11.6	59.6
	Seed 1	89.1	66.0	38.7	34.9	34.1	82.7	61.2	90.8	83.4	11.1	59.2
	Seed 2	89.4	65.7	38.8	35.8	35.7	83.1	62.8	90.7	83.5	11.4	59.7
	Seed 3	90.2	64.4	38.5	35.1	34.0	82.7	63.3	90.7	83.5	9.7	59.2
	Seed 4	88.9	64.0	38.9	35.8	36.2	82.8	61.6	90.7	83.2	9.5	59.2
	Seed 5	89.7	65.5	38.9	35.4	33.8	83.3	64.0	91.0	83.3	9.2	59.4
	Seed 6	89.4	66.1	38.5	35.6	35.4	82.8	61.9	90.8	83.4	11.2	59.5
	Seed 7	89.5	65.5	38.8	35.7	35.7	82.6	62.2	90.7	83.6	11.9	59.6
	Seed 8	89.0	65.5	38.3	33.7	35.6	82.8	61.1	90.7	83.4	10.4	59.1
	Seed 9	89.5	66.2	39.2	32.9	34.6	82.9	62.4	90.6	83.6	10.9	59.3
	Min	88.9	64.0	38.3	32.9	33.8	82.6	61.1	90.6	83.0	9.2	59.1
	Max	90.4	66.2	39.2	35.8	36.2	83.3	64.0	91.0	83.6	11.9	59.7

Table 2. Semantic segmentation results for post-adaptation (PA) on KITTI online semantics subset.

6. Detailed Pipeline for Experience Replay

Require: Target domain video v, adaptation batch size b, random seed s, maximum replay buffer size M

$\mathcal{E} \leftarrow \emptyset$ ▷ Initialize replay buffer, always keep samples sorted according to their timestamps

$r \leftarrow 1$ ▷ Experience buffering rate

for $t \in [2..\textbf{end}(v)]$ **do**

  if $t\%r == 0$ **then** ▷ Add samples to buffer every r frames to keep them uniformly distributed over time

    if size($\mathcal{E}$) $< M$ **then**

      $\mathcal{E} \leftarrow$ **insert_sample**$(I^v_{t\text{-}2}, I^v_{t\text{-}1}, I^v_t)$

    else ▷ Handle buffer overflow here

      $\mathcal{E} \leftarrow$ **remove_every_2nd_sample** $(\mathcal{E})$ ▷ Halving the size of the buffer

      $\mathcal{E} \leftarrow$ **insert_sample**$(I^v_{t\text{-}2}, I^v_{t\text{-}1}, I^v_t)$

      $r \leftarrow r \cdot 2$

  $\mathcal{B} \leftarrow (I^v_{t\text{-}2}, I^v_{t\text{-}1}, I^v_t) \oplus$ **sample**$(\mathcal{E}, b\text{ - }1, s)$ ▷ Triplet batch used for optimization

  Do adaptation using $\mathcal{B}$