ACDC: The Adverse Conditions Dataset with Correspondences for Semantic Driving Scene Understanding –Supplementary Material–

Christos Sakaridis[1], Dengxin Dai[1,2], and Luc Van Gool[1,3]

[1]ETH Zürich, [2]MPI for Informatics, [3]KU Leuven

A. Training Details

We provide the detailed training configurations for the various methods for semantic segmentation that have been used in Sec. 4 of the paper and for the method in [1] for uncertainty-aware semantic segmentation that has been used in Sec. 5 of the paper.

A.1. Normal-to-Adverse Adaptation

For the comparison in Table 2, we use as source-domain model the DeepLabv2 [5] model that is used as the Cityscapes oracle in AdaptSegNet [43], with a performance of 65.1% mIoU on the Cityscapes validation set. For all eight unsupervised domain adaptation (UDA) methods that are compared, we use their default training configurations, including the learning rate schedule and the weights of the various losses. The number of training iterations run for each method as well as the number of self-supervised learning rounds that are used by some of the methods are reported in Table 10. For FDA, SIM and MRNet, we run a first training round without self-training followed by a second training round with self-training, as per default implementation of these methods. For FDA, we train three separate models in each training round, one for each different value of the β parameter from the set $\{0.01,\ 0.05\, 0.09\}$, and use the average prediction of the three models at test time. In all cases, we use the model weights corresponding to the final training iteration for testing.

The same source-domain model is also used for the experiment on adaptation to individual conditions presented in Table 3. Again, we use the default training configurations for all examined methods and across all four conditions. The number of training iterations run for each method to adapt to each condition as well as the number of self-supervised learning rounds that are used by some of the methods are reported in Table 11. For MRNet and fog, the self-supervised training round includes 35k iterations instead of 40k. In addition, for MRNet and rain, the first training round without self-supervised training includes 25k iterations instead of 40k.

Table 10. **Training details for UDA methods in Cityscapes→ACDC adaptation.** "SSL rounds": number of training rounds that include supervision from pseudo-labels; if not relevant for a method, – is reported. "Training iterations": number of SGD iterations for each training round (number of epochs for each training round is alternatively reported).

Method	SSL rounds	Training iterations
AdaptSegNet	–	95k
ADVENT	–	80k
BDL	0	80k
CLAN	–	90k
CRST	3	2 epochs
FDA	1	80k
SIM	1	80k
MRNet	1	50k

Table 11. **Training details for UDA methods in Cityscapes→ACDC adaptation for individual conditions.** "SSL rounds": number of training rounds that include supervision from pseudo-labels; if not relevant for a method, – is reported. "Training iterations": number of SGD iterations for each training round.

Method	SSL rounds	Training iterations
AdaptSegNet	–	40k
ADVENT	–	40k
BDL	0	40k
CLAN	–	40k
FDA	1	40k
SIM	1	40k
MRNet	1	40k

A.2. Supervised Learning on Adverse Conditions

For training the four semantic segmentation methods that are compared in Tables 5 and 6, we have generally used the

Table 12. **Training details for supervised methods on ACDC.**

Method	Base LR	Training epochs
RefineNet	5×10^{-5}	60
DeepLabv2	2.5×10^{-4}	60
DeepLabv3+	10^{-4}	60
HRNet	10^{-4}	60

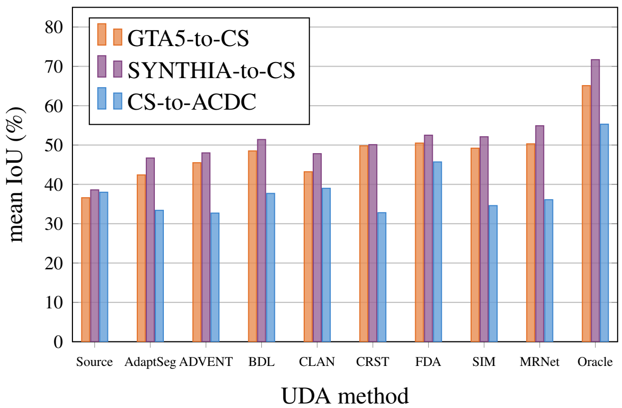


Figure 4. **Comparison of Cityscapes→ACDC to other domain adaptation benchmarks.** We compare the performance of state-of-the-art domain adaptation methods on three benchmarks: GTA5→Cityscapes, SYNTHIA→Cityscapes, and Cityscapes→ACDC. Performance of models trained only on the source domain (Source) and on the target domain with supervision (Oracle) is also presented. "AdaptSeg": AdaptSegNet.

default configuration for each method both in the case of condition experts and uber models. For DeepLabv2 [5], we use the architecture employed in AdaptSegNet [43] in the context of domain adaptation and not the original architecture. We have used the default learning rate schedule for each method, with the base learning rates that are reported in Table 12. We generally use 60 training epochs for all four methods, which yields 96k training iterations for uber models and 24k training iterations for condition experts. Exceptions to this rule are RefineNet and fog where we use 30 epochs, DeepLabv2 and fog where we use 45 epochs, DeepLabv2 and night where we use 240 epochs, and the DeepLabv3+ uber model for which we use 30 epochs. For HRNet, we use the snapshot with the best mIoU performance on the respective validation set of ACDC for predicting on the test set, while for the rest of the methods we use the final training snapshot for the same purpose.

A.3. Uncertainty-Aware Semantic Segmentation

We have used the two-head model designed in [1] and trained it on the entire training set of ACDC for 60 epochs. We use the default learning rate schedule of [1], with a base learning rate of 4×10^{-4}, which is equal to the default. For predicting on the test set, we use the final training snapshot.

B. Comparison to Other Adaptation Benchmarks

In Fig. 4, we present the comparative performance of the eight UDA methods we have used in Table 2 on two additional popular domain adaptation benchmarks, i.e. GTA5→Cityscapes and SYNTHIA→Cityscapes. All examined methods have been configured to optimize performance on the synthetic-to-real UDA setting of these two benchmarks, but this configuration does not generalize well to the real-world normal-to-adverse adaptation setting of Cityscapes→ACDC, as most methods cannot improve significantly upon the source-trained baseline. Furthermore, the comparative performance of the various methods on either of the synthetic-to-real benchmarks does not correlate well with the comparative performance on Cityscapes→ACDC. This indicates that adaptation strategies that work well for the synthetic-to-real setting may not be as beneficial in the real-world normal-to-adverse adaptation setting and that new strategies and algorithms need to be devised for the latter.

C. Detailed Class-level Results

We provide class-level performance for the experiments for which only mean performance over all classes is reported in the paper due to space limitations.

C.1. Normal-to-Adverse Adaptation

In Tables 13–16, we present the class-level IoU performance of the UDA methods that are examined in the setting of adaptation to individual conditions in Table 3 of the paper.

C.2. Evaluation of Pre-trained Models on ACDC

In Tables 17–21, we present the class-level IoU performance of the externally pre-trained models that are evaluated in Table 4 of the paper.

C.3. Supervised Learning on Adverse Conditions

In Tables 22–25, we present the class-level IoU performance of the supervised semantic segmentation methods that are examined in Table 6 of the paper. In particular, we consider the individual conditions of ACDC separately for evaluation, and evaluate on each condition both the respective condition experts that have been trained only on that condition and uber models trained on all conditions.

C.4. Uncertainty-aware Semantic Segmentation

In Tables 26–30, we present the class-level average uncertainty-aware IoU (AUIoU) performance of the baselines and oracles that are examined in Table 9 of the paper. More specifically, Table 26 considers methods trained

Table 13. **Comparison of state-of-the-art unsupervised domain adaptation methods on Cityscapes→ACDC adaptation for fog.** Performance of the model trained only on the source domain (Source model) and of the oracle with access to the target domain labels (Oracle) is also reported.

Method	road	sidew.	build.	wall	fence	pole	light	sign	veget.	terrain	sky	person	rider	car	truck	bus	train	motorc.	bicycle	mIoU
Source model	66.4	31.2	26.8	22.9	18.6	8.2	32.3	10.7	70.7	39.0	31.3	17.6	41.1	65.0	30.0	34.3	18.3	42.3	29.0	33.5
AdaptSegNet	35.4	45.9	35.4	25.6	17.5	9.0	32.5	23.1	70.5	47.4	11.6	22.3	28.2	44.4	43.9	35.0	46.0	15.6	15.0	31.8
ADVENT	44.2	38.9	26.4	20.7	20.1	7.9	34.4	23.6	70.7	35.6	8.3	17.3	43.5	60.0	48.6	46.8	40.5	19.9	17.6	32.9
BDL	36.9	37.8	47.0	28.2	21.6	13.7	37.2	34.5	67.2	49.4	27.6	29.1	51.3	58.5	49.4	51.8	30.3	21.4	22.5	37.7
CLAN	48.8	41.3	29.6	27.2	21.0	16.1	41.1	39.6	67.7	50.2	15.4	36.2	30.8	72.2	52.2	54.4	47.2	27.1	22.6	39.0
FDA	68.8	37.3	27.1	27.6	19.8	21.6	37.5	43.3	74.9	43.7	33.1	35.0	21.5	65.7	44.6	45.3	47.1	41.5	15.8	39.5
SIM	76.7	43.1	23.5	23.6	17.9	10.9	32.1	15.3	70.4	50.5	21.4	34.8	44.3	58.4	50.5	55.2	34.7	23.0	8.8	36.6
MRNet	78.6	26.1	19.6	29.0	13.5	12.0	41.9	49.0	78.2	59.0	6.6	39.8	26.1	72.5	44.8	37.9	59.6	19.1	24.1	38.8
Oracle	89.9	65.6	81.2	39.1	25.9	28.1	45.9	47.7	83.0	67.4	96.7	35.2	38.4	73.5	46.1	29.8	37.9	28.4	31.6	52.2

Table 14. **Comparison of state-of-the-art unsupervised domain adaptation methods on Cityscapes→ACDC adaptation for nighttime.** Performance of the model trained only on the source domain (Source model) and of the oracle with access to the target domain labels (Oracle) is also reported.

Method	road	sidew.	build.	wall	fence	pole	light	sign	veget.	terrain	sky	person	rider	car	truck	bus	train	motorc.	bicycle	mIoU
Source model	77.0	22.9	56.3	13.5	9.2	23.8	22.9	25.6	41.4	16.1	2.9	44.1	17.5	64.1	11.9	34.5	42.4	22.6	22.7	30.1
AdaptSegNet	84.9	39.9	66.8	17.2	17.7	13.4	17.6	16.4	39.6	16.1	5.7	42.8	21.4	44.8	11.9	13.0	39.1	27.5	28.4	29.7
ADVENT	86.5	45.3	60.8	23.2	12.5	15.4	18.0	19.4	41.2	18.3	2.7	43.8	21.3	61.6	12.6	19.1	43.0	30.2	27.6	31.7
BDL	87.1	49.6	68.8	20.2	17.5	16.7	19.9	24.1	39.1	23.7	0.2	42.0	20.4	63.7	18.0	27.0	45.6	27.8	31.3	33.8
CLAN	82.3	28.8	65.9	15.1	9.3	22.1	16.1	26.5	39.2	23.4	0.4	45.9	25.4	63.6	9.5	24.2	39.8	31.5	31.1	31.6
FDA	82.7	39.4	57.0	14.7	7.6	26.1	37.8	30.5	53.2	14.0	15.3	48.0	28.8	62.6	26.6	47.5	51.5	27.0	35.0	37.1
SIM	87.0	48.4	42.1	6.3	8.3	15.8	8.4	17.6	21.7	22.8	0.1	39.3	22.1	60.3	8.7	18.2	42.3	30.1	32.9	28.0
MRNet	83.6	36.3	65.6	8.1	8.2	21.5	30.0	23.7	39.4	24.2	0.0	44.1	26.0	64.9	0.8	3.6	7.6	10.3	31.8	27.9
Oracle	90.5	63.7	78.0	30.0	29.6	32.9	37.0	41.2	61.9	25.2	75.3	47.9	23.4	69.5	2.7	15.4	60.3	39.7	37.9	45.4

Table 15. **Comparison of state-of-the-art unsupervised domain adaptation methods on Cityscapes→ACDC adaptation for rain.** Performance of the model trained only on the source domain (Source model) and of the oracle with access to the target domain labels (Oracle) is also reported.

Method	road	sidew.	build.	wall	fence	pole	light	sign	veget.	terrain	sky	person	rider	car	truck	bus	train	motorc.	bicycle	mIoU
Source model	71.2	26.7	73.8	20.8	27.1	29.9	39.3	44.4	87.3	25.2	82.0	42.0	14.3	76.2	36.3	26.6	49.8	30.3	42.2	44.5
AdaptSegNet	81.2	43.2	83.3	27.3	31.4	23.0	41.4	40.5	87.2	35.0	93.1	40.2	15.5	73.9	45.7	34.9	57.0	27.1	49.1	49.0
ADVENT	77.0	31.0	52.5	35.0	34.2	23.4	42.1	41.0	85.3	34.2	26.7	41.3	14.1	75.6	47.3	40.4	64.3	29.6	46.2	44.3
BDL	79.1	39.0	82.8	30.0	34.5	28.1	40.1	47.3	87.0	28.7	91.8	40.6	17.8	74.6	46.3	36.7	60.4	33.2	46.3	49.7
CLAN	77.5	40.0	46.8	24.9	30.3	28.1	37.7	48.3	83.8	37.0	6.6	45.7	17.4	79.7	43.7	42.9	63.7	35.0	46.1	44.0
FDA	76.6	45.0	82.9	37.0	35.6	34.8	49.8	52.0	88.7	37.8	88.8	43.6	17.4	76.8	46.5	53.6	64.8	34.5	45.5	53.3
SIM	76.6	29.6	85.7	20.4	28.7	21.3	37.4	34.2	87.3	34.8	94.0	29.4	16.6	73.2	46.1	22.3	46.2	21.8	39.3	44.5
MRNet	70.5	9.9	46.5	35.6	36.1	36.5	56.4	56.2	90.2	41.3	4.3	53.0	23.5	81.6	39.3	26.7	57.8	43.6	54.5	45.4
Oracle	87.3	63.9	89.0	50.3	40.6	38.4	52.2	53.4	89.2	42.2	96.7	51.5	13.0	81.9	47.9	47.2	72.2	29.1	48.8	57.6

jointly on all conditions of ACDC and also evaluated jointly on all conditions, while Tables 27–30 present methods trained and evaluated on individual conditions. The results corresponding to the baseline that uses constant confidence equal to 1 are omitted, as they are identical by definition to IoU results and are thus already included in Table 5 of the paper and Tables 22–25.

Table 16. **Comparison of state-of-the-art unsupervised domain adaptation methods on Cityscapes→ACDC adaptation for snow.** Performance of the model trained only on the source domain (Source model) and of the oracle with access to the target domain labels (Oracle) is also reported.

Method	road	sidew.	build.	wall	fence	pole	light	sign	veget.	terrain	sky	person	rider	car	truck	bus	train	motorc.	bicycle	mIoU
Source model	68.5	26.6	52.7	18.8	26.9	22.2	35.7	40.7	76.5	3.6	49.9	50.4	27.1	73.7	27.6	39.1	60.9	21.1	42.5	40.2
AdaptSegNet	51.3	32.5	47.3	21.5	31.5	13.2	37.8	23.2	76.0	2.6	4.5	49.9	23.1	68.7	38.3	31.8	51.5	21.7	45.0	35.3
ADVENT	50.8	24.8	46.2	15.5	26.0	15.5	27.9	23.0	70.0	2.1	9.5	44.2	25.3	68.5	22.9	24.9	50.1	23.9	38.9	32.1
BDL	42.3	36.4	60.2	15.7	30.4	15.1	41.4	30.4	71.3	1.7	11.2	46.8	27.8	57.7	38.6	34.1	59.2	28.1	43.7	36.4
CLAN	71.8	26.0	37.3	12.5	27.0	21.1	32.0	41.1	78.5	1.9	0.9	50.9	23.9	82.4	43.2	39.5	61.6	25.2	39.4	37.7
FDA	74.6	30.9	56.1	20.5	34.8	28.7	53.9	47.8	80.5	1.1	55.9	53.1	37.9	79.7	40.5	51.9	67.4	34.3	41.8	46.9
SIM	72.1	26.7	39.4	13.3	29.5	15.3	26.4	17.9	76.4	4.8	5.1	45.9	32.0	76.2	29.8	26.6	48.3	23.2	24.2	33.3
MRNet	67.7	3.5	36.8	8.3	24.8	18.0	52.6	55.4	82.4	0.5	0.1	62.2	30.2	79.2	32.1	59.3	58.4	29.1	35.8	38.7
Oracle	89.1	61.7	82.7	26.4	40.9	35.5	56.5	54.1	85.2	39.0	95.1	55.0	25.7	84.3	38.6	53.8	77.6	29.0	49.5	56.8

Table 17. **Comparison of externally pre-trained models on the complete test set of ACDC including all conditions.** The three groups of rows present models pre-trained on normal, foggy, and nighttime conditions respectively. CS: Cityscapes, FC: Foggy Cityscapes, FC-DBF: Foggy Cityscapes-DBF, FZ: Foggy Zurich, ND: Nighttime Driving, DZ: Dark Zurich.

Method	Trained on	road	sidew.	build.	wall	fence	pole	light	sign	veget.	terrain	sky	person	rider	car	truck	bus	train	motorc.	bicycle	mIoU
RefineNet	CS	66.3	28.9	67.6	19.2	25.9	36.7	50.0	47.5	69.4	28.8	83.0	42.1	17.7	72.6	30.9	31.6	48.9	26.1	36.7	43.7
DeepLabv2	CS	71.9	26.2	51.1	18.8	22.5	19.7	33.0	27.7	67.9	28.6	44.2	43.1	22.1	71.2	29.8	33.3	48.4	26.2	35.8	38.0
DeepLabv3+	CS	75.1	32.8	65.9	17.5	20.2	32.2	46.7	45.2	70.5	33.5	80.9	23.9	14.7	71.5	40.1	20.3	51.2	20.2	28.8	41.6
DANet	CS	58.0	6.0	57.3	6.8	22.3	27.7	41.3	42.1	66.4	19.9	69.2	32.2	10.2	46.5	22.4	19.1	43.1	13.2	25.5	33.1
HRNet	CS	55.6	10.9	55.4	7.7	15.9	21.7	37.8	42.5	67.4	13.3	59.0	38.7	14.0	68.3	23.8	48.0	48.3	17.9	23.6	35.3
SFSU	FC	72.9	28.8	68.3	19.6	23.9	37.3	49.3	47.0	60.4	33.4	72.3	43.1	14.8	72.7	31.7	31.2	47.0	25.4	35.5	42.9
CMAda	FC-DBF+FZ	79.9	32.5	69.5	14.7	24.7	41.1	53.6	51.3	67.4	34.8	83.8	49.0	19.9	77.0	34.1	38.5	51.1	29.6	42.7	47.1
DMAda	ND	75.3	35.5	67.4	19.2	27.1	40.0	53.7	50.9	74.6	30.9	84.9	48.8	23.1	76.6	39.7	37.4	52.5	29.1	42.1	47.9
GCMA	CS+DZ	79.7	48.7	71.5	21.6	29.9	42.5	56.7	57.7	75.8	39.5	87.2	57.4	29.7	80.6	44.9	46.2	62.0	37.2	46.5	53.4
MGCDA	CS+DZ	76.0	49.4	72.0	11.3	21.7	39.5	52.0	54.9	73.7	24.7	88.6	54.1	27.2	78.2	30.9	41.9	58.2	31.1	44.4	48.9

Table 18. **Comparison of externally pre-trained models on ACDC for fog.** The three groups of rows present models pre-trained on normal, foggy, and nighttime conditions respectively. CS: Cityscapes, FC: Foggy Cityscapes, FC-DBF: Foggy Cityscapes-DBF, FZ: Foggy Zurich, ND: Nighttime Driving, DZ: Dark Zurich.

Method	Trained on	road	sidew.	build.	wall	fence	pole	light	sign	veget.	terrain	sky	person	rider	car	truck	bus	train	motorc.	bicycle	mIoU
RefineNet	CS	64.4	40.0	69.6	24.2	19.7	36.5	52.7	55.2	71.1	35.4	93.9	27.4	19.2	72.7	42.0	42.1	69.3	30.3	15.8	46.4
DeepLabv2	CS	66.4	31.2	26.8	22.9	18.6	8.2	32.3	10.7	70.7	39.0	31.3	17.6	41.1	65.0	30.0	34.3	18.3	42.3	29.0	33.5
DeepLabv3+	CS	82.3	57.6	61.5	18.1	16.4	33.3	49.6	54.5	76.0	44.1	90.0	9.6	28.7	69.0	35.1	34.5	28.9	41.7	37.5	45.7
DANet	CS	52.1	14.5	49.7	5.5	16.9	30.0	47.9	51.5	72.2	23.3	80.1	24.2	3.0	44.7	32.4	27.5	65.1	10.8	7.7	34.7
HRNet	CS	57.3	19.3	49.1	12.8	17.8	27.3	44.0	54.7	72.8	15.5	81.7	28.3	3.9	66.6	28.4	52.0	72.7	7.2	18.1	38.4
SFSU	FC	72.3	37.9	74.4	28.9	19.3	37.5	49.4	54.6	58.0	43.7	77.9	28.6	5.3	73.6	42.4	44.0	72.7	31.4	14.9	45.6
CMAda	FC-DBF+FZ	81.7	43.5	72.8	25.6	19.5	39.8	51.0	58.9	80.5	51.3	95.3	36.9	12.7	76.5	45.2	51.2	77.1	33.2	19.9	51.2
DMAda	ND	75.5	44.7	72.6	26.4	20.8	38.3	52.9	57.8	75.9	38.6	96.3	35.5	26.8	75.8	47.7	50.7	73.9	35.8	17.3	50.7
GCMA	CS+DZ	80.8	53.5	70.1	29.2	20.7	38.4	53.0	60.9	70.2	46.5	95.4	44.2	38.0	76.6	52.4	49.7	56.8	41.0	17.6	52.4
MGCDA	CS+DZ	71.7	47.3	65.7	18.2	15.3	34.4	48.6	59.9	64.9	24.7	95.4	44.8	23.8	73.3	36.1	45.4	63.9	23.9	15.4	45.9

D. Additional Details on ACDC Dataset

We provide additional details on the construction and the characteristics of ACDC. We have implemented a website and evaluation server for the ACDC benchmark and have made it publicly available. An indicative screenshot from the submission page of the website is provided in Fig. 6.

D.1. Collection

Our recordings were performed in Switzerland. Therefore, the geographic distribution of ACDC is similar to Cityscapes, which was also recorded in central Europe.

Table 19. **Comparison of externally pre-trained models on ACDC for nighttime.** The three groups of rows present models pre-trained on normal, foggy, and nighttime conditions respectively. CS: Cityscapes, FC: Foggy Cityscapes, FC-DBF: Foggy Cityscapes-DBF, FZ: Foggy Zurich, ND: Nighttime Driving, DZ: Dark Zurich.

Method	Trained on	road	sidew.	build.	wall	fence	pole	light	sign	veget.	terrain	sky	person	rider	car	truck	bus	train	motorc.	bicycle	mIoU
RefineNet	CS	66.5	24.0	50.3	16.9	11.6	26.4	34.2	25.5	44.2	21.6	0.1	40.8	24.8	57.4	6.8	37.3	20.5	23.9	19.1	29.0
DeepLabv2	CS	77.0	22.9	56.3	13.5	9.2	23.8	22.9	25.6	41.4	16.1	2.9	44.1	17.5	64.1	11.9	34.5	42.4	22.6	22.7	30.1
DeepLabv3+	CS	73.0	20.8	50.4	22.2	5.4	22.6	31.8	23.0	42.9	16.1	6.6	19.2	11.7	48.9	0.9	13.9	42.4	10.5	13.7	25.0
DANet	CS	67.1	4.5	46.7	5.5	5.1	13.1	29.3	19.6	36.6	15.6	0.1	29.3	12.4	29.1	4.5	12.3	9.0	10.3	13.3	19.1
HRNet	CS	50.0	10.1	59.9	0.7	6.0	14.2	25.6	22.3	19.1	3.4	0.1	37.6	7.9	49.4	6.9	45.9	13.9	7.8	11.3	20.6
SFSU	FC	76.9	26.2	50.4	18.1	9.6	27.4	33.3	25.3	41.0	21.5	0.0	41.5	25.3	58.7	7.3	40.7	17.9	22.0	17.9	29.5
CMAda	FC-DBF+FZ	82.6	25.4	53.9	10.1	11.2	30.5	36.7	30.0	38.7	16.5	0.1	46.0	26.2	65.8	13.9	50.9	20.4	24.8	23.8	32.0
DMAda	ND	74.7	29.5	49.4	17.1	12.6	31.0	38.2	30.0	48.0	22.8	0.2	47.0	25.4	63.8	12.8	46.1	23.1	24.7	24.6	32.7
GCMA	CS+DZ	78.6	45.9	58.5	17.7	18.6	37.5	43.6	43.5	58.7	39.2	22.4	57.9	29.9	72.1	21.5	56.2	41.8	35.7	35.4	42.9
MGCDA	CS+DZ	74.5	52.5	69.4	7.7	10.8	38.4	40.2	43.3	61.5	36.3	37.6	55.3	25.6	71.2	10.9	46.4	32.6	27.3	33.8	40.8
DANNet	CS+DZ	90.7	61.1	75.5	35.9	28.8	26.6	31.4	30.6	70.8	39.4	78.7	49.9	28.8	65.9	24.7	44.1	61.1	25.9	34.5	47.6

Table 20. **Comparison of externally pre-trained models on ACDC for rain.** The three groups of rows present models pre-trained on normal, foggy, and nighttime conditions respectively. CS: Cityscapes, FC: Foggy Cityscapes, FC-DBF: Foggy Cityscapes-DBF, FZ: Foggy Zurich, ND: Nighttime Driving, DZ: Dark Zurich.

Method	Trained on	road	sidew.	build.	wall	fence	pole	light	sign	veget.	terrain	sky	person	rider	car	truck	bus	train	motorc.	bicycle	mIoU
RefineNet	CS	73.9	29.9	82.9	26.3	37.2	46.3	61.8	57.9	89.4	42.5	96.6	44.2	13.2	80.5	40.7	22.9	66.8	32.0	53.5	52.6
DeepLabv2	CS	71.2	26.7	73.8	20.8	27.1	29.9	39.3	44.4	87.3	25.2	82.0	42.0	14.3	76.2	36.3	26.6	49.8	30.3	42.2	44.5
DeepLabv3+	CS	74.4	29.8	82.3	18.1	28.8	41.7	54.3	55.6	88.7	32.8	97.2	36.7	8.5	84.7	51.7	34.0	61.5	29.7	40.0	50.0
DANet	CS	59.9	2.4	75.9	12.9	31.5	37.7	49.5	53.3	85.5	35.5	91.1	35.4	8.4	53.5	26.0	16.4	57.8	17.9	38.9	41.5
HRNet	CS	65.0	6.7	70.3	16.1	20.2	29.5	48.5	54.7	87.5	36.1	80.1	40.6	8.6	78.2	34.1	44.6	67.3	29.4	34.6	44.8
SFSU	FC	74.6	29.9	81.4	24.1	33.8	46.2	59.9	56.7	86.8	40.8	93.4	46.4	15.1	80.5	40.5	18.6	65.7	33.6	52.5	51.6
CMAda	FC-DBF+FZ	78.1	34.8	80.7	18.9	33.3	50.0	63.1	62.2	87.4	38.8	96.6	51.1	16.9	83.3	37.9	21.9	68.7	36.5	55.1	53.4
DMAda	ND	78.3	37.7	82.5	24.2	36.8	49.0	64.5	61.5	90.6	42.8	97.3	49.6	18.2	83.4	45.1	21.6	70.2	35.2	54.8	54.9
GCMA	CS+DZ	81.1	48.0	84.8	25.0	37.3	49.8	66.5	66.2	92.1	43.5	97.6	54.5	20.4	85.5	47.3	34.6	71.3	40.3	56.7	58.0
MGCDA	CS+DZ	80.5	46.5	79.9	16.0	28.8	44.9	60.0	61.5	90.3	44.8	97.1	51.1	23.1	82.3	33.4	30.2	69.1	36.5	53.8	54.2

Table 21. **Comparison of externally pre-trained models on ACDC for snow.** The three groups of rows present models pre-trained on normal, foggy, and nighttime conditions respectively. CS: Cityscapes, FC: Foggy Cityscapes, FC-DBF: Foggy Cityscapes-DBF, FZ: Foggy Zurich, ND: Nighttime Driving, DZ: Dark Zurich.

Method	Trained on	road	sidew.	build.	wall	fence	pole	light	sign	veget.	terrain	sky	person	rider	car	truck	bus	train	motorc.	bicycle	mIoU
RefineNet	CS	61.0	25.5	73.7	11.7	31.1	37.2	53.1	57.7	71.3	0.9	92.7	44.1	14.7	77.0	30.3	26.9	57.2	18.4	38.5	43.3
DeepLabv2	CS	68.5	26.6	52.7	18.8	26.9	22.2	35.7	40.7	76.5	3.6	49.9	50.4	27.1	73.7	27.6	39.1	60.9	21.1	42.5	40.2
DeepLabv3+	CS	73.9	32.6	71.3	11.1	25.6	31.4	50.6	54.4	77.8	4.1	87.0	25.1	14.6	82.7	39.5	17.2	55.2	12.0	31.2	42.0
DANet	CS	47.6	5.4	57.5	2.9	29.1	29.3	41.4	51.2	71.1	0.5	64.8	32.7	11.7	56.5	14.5	27.9	53.7	8.1	25.9	33.3
HRNet	CS	59.6	9.3	43.9	4.0	17.8	17.6	35.6	47.0	77.0	0.0	32.5	39.4	39.2	74.2	13.4	54.0	61.1	15.9	26.1	35.1
SFSU	FC	64.5	24.0	72.6	10.9	28.8	37.8	54.9	58.1	62.4	0.8	78.4	44.2	9.5	76.0	29.5	25.6	55.2	16.7	37.3	41.4
CMAda	FC-DBF+FZ	74.6	31.6	73.6	9.4	30.3	43.1	61.9	61.7	75.7	0.7	93.5	53.1	19.1	79.6	29.7	31.6	61.9	22.9	50.3	47.6
DMAda	ND	73.6	34.4	74.9	12.3	33.4	41.1	58.4	60.1	79.9	0.6	95.4	53.1	23.0	80.4	40.3	34.5	62.9	22.7	48.6	48.9
GCMA	CS+DZ	79.7	49.5	75.3	17.5	37.9	43.2	59.0	61.9	78.8	2.2	95.5	62.5	33.6	83.2	42.5	43.4	72.1	32.2	51.1	53.7
MGCDA	CS+DZ	80.1	49.5	70.2	6.1	27.8	39.6	55.4	58.0	76.0	0.3	95.5	57.5	35.7	81.0	28.6	48.9	70.3	27.8	50.5	50.5

This eliminates geographic location from the set of factors that introduce a domain shift between Cityscapes and ACDC and allows to study in isolation the effect of visual conditions at time of capture on the performance of semantic segmentation methods, both in the supervised setting and the unsupervised domain adaptation setting.

D.2. Correspondence Establishment

We present in Algorithm 1 the dynamic programming algorithm that we use for matching the GPS sequences of adverse-condition recordings and normal-condition recordings of ACDC. The algorithm takes into account the sequen-

Table 22. **Comparison of state-of-the-art supervised methods on ACDC for fog.** The first group of rows presents condition-specific expert models trained only on fog, while the second group presents uber models trained on all conditions.

Method	road	sidew.	build.	wall	fence	pole	light	sign	veget.	terrain	sky	person	rider	car	truck	bus	train	motorc.	bicycle	mIoU
RefineNet	93.2	75.5	86.1	44.1	37.6	46.0	64.2	64.8	85.5	70.8	97.9	46.1	34.8	79.3	59.4	64.8	82.4	36.6	38.8	63.6
DeepLabv2	89.9	65.6	81.2	39.1	25.9	28.1	45.9	47.7	83.0	67.4	96.7	35.2	38.4	73.5	46.1	29.8	37.9	28.4	31.6	52.2
DeepLabv3+	93.8	77.4	88.8	51.0	43.3	54.2	68.2	71.7	87.7	74.6	98.2	53.5	32.1	83.8	69.3	84.4	85.3	47.2	40.1	68.7
HRNet	94.6	79.6	89.9	53.6	44.9	59.4	74.3	76.1	88.9	77.6	98.3	61.5	53.3	86.0	66.6	80.0	88.5	41.1	30.2	70.8
RefineNet	93.5	75.6	87.2	42.3	39.2	49.8	68.5	67.2	85.6	70.1	97.9	52.6	48.2	81.0	62.6	62.0	69.1	57.7	37.4	65.7
DeepLabv2	90.9	67.2	81.6	38.7	29.5	29.7	51.2	50.7	81.4	61.9	96.0	34.8	40.5	74.1	53.4	53.1	59.9	8.3	32.5	54.5
DeepLabv3+	93.6	77.6	89.2	54.0	44.8	55.8	67.6	72.0	88.0	73.5	98.2	49.5	24.4	83.9	72.2	84.2	89.2	52.8	42.4	69.1
HRNet	94.9	81.0	90.5	58.9	53.7	61.9	79.0	78.7	89.3	78.7	98.3	63.2	54.6	87.2	72.3	87.8	90.6	58.7	38.9	74.7

Table 23. **Comparison of state-of-the-art supervised methods on ACDC for nighttime.** The first group of rows presents condition-specific expert models trained only on nighttime, while the second group presents uber models trained on all conditions.

Method	road	sidew.	build.	wall	fence	pole	light	sign	veget.	terrain	sky	person	rider	car	truck	bus	train	motorc.	bicycle	mIoU
RefineNet	93.4	70.3	78.6	34.3	34.1	46.9	52.2	54.2	66.3	18.7	78.1	60.3	35.5	76.2	4.7	47.8	59.4	36.0	45.3	52.2
DeepLabv2	90.5	63.7	78.0	30.0	29.6	32.9	37.0	41.2	61.9	25.2	75.3	47.9	23.4	69.5	2.7	15.4	60.3	39.7	37.9	45.4
DeepLabv3+	94.7	75.9	85.0	48.4	38.6	52.2	55.8	54.4	76.1	30.3	84.2	67.4	41.1	85.0	8.3	62.3	80.6	35.6	49.8	59.2
HRNet	95.5	78.8	86.5	49.2	44.1	58.0	64.5	63.2	75.6	41.0	83.9	71.7	48.8	84.6	15.5	76.9	81.2	25.9	55.9	63.2
RefineNet	93.5	70.9	80.3	32.0	32.0	46.0	53.9	54.1	69.2	31.9	78.0	61.0	35.4	80.2	11.6	60.0	69.4	48.9	46.8	55.5
DeepLabv2	86.6	57.8	71.7	30.3	23.6	31.8	37.4	38.9	60.0	26.8	72.8	47.6	25.1	71.1	16.9	27.8	65.1	30.6	38.5	45.3
DeepLabv3+	94.7	75.3	84.9	46.9	37.8	53.8	57.3	52.1	75.7	41.2	82.9	66.6	40.2	83.6	24.7	67.9	80.8	41.7	49.4	60.9
HRNet	95.7	79.0	86.2	46.8	43.5	59.2	64.9	64.5	75.3	40.3	82.7	72.1	52.6	86.9	18.8	78.8	83.6	52.5	57.3	65.3

Table 24. **Comparison of state-of-the-art supervised methods on ACDC for rain.** The first group of rows presents condition-specific expert models trained only on rain, while the second group presents uber models trained on all conditions.

Method	road	sidew.	build.	wall	fence	pole	light	sign	veget.	terrain	sky	person	rider	car	truck	bus	train	motorc.	bicycle	mIoU
RefineNet	89.2	69.8	91.7	52.2	51.3	57.9	71.0	69.9	93.6	50.5	98.4	65.8	25.1	88.1	49.4	55.4	74.8	47.0	60.2	66.4
DeepLabv2	87.3	63.9	89.0	50.3	40.6	38.4	52.2	53.4	89.2	42.2	96.7	51.5	13.0	81.9	47.9	47.2	72.2	29.1	48.8	57.6
DeepLabv3+	92.8	77.4	93.9	67.3	58.1	64.1	74.4	75.9	94.2	50.8	98.6	70.8	33.4	90.4	67.7	79.2	86.8	54.6	66.1	73.5
HRNet	94.8	81.8	94.9	69.6	63.7	69.5	79.6	80.7	94.8	51.2	98.7	73.5	27.0	93.1	75.4	40.9	61.4	59.6	70.8	72.7
RefineNet	91.5	73.5	91.1	51.0	51.6	58.3	72.5	73.7	92.9	51.2	97.9	65.5	29.5	89.2	59.8	68.2	80.3	48.0	59.5	68.7
DeepLabv2	87.4	64.8	88.1	48.2	40.4	38.4	52.0	56.9	89.3	40.2	96.5	52.3	17.4	83.9	55.5	63.0	75.8	28.9	47.2	59.3
DeepLabv3+	92.7	76.5	93.5	64.8	58.0	63.8	75.8	77.3	94.1	50.0	98.0	70.5	33.1	91.2	75.9	85.1	86.2	55.8	65.0	74.1
HRNet	95.6	83.1	94.2	60.1	66.3	71.2	82.3	82.4	94.6	55.1	98.6	75.2	39.7	93.4	73.8	86.2	85.9	66.4	71.3	77.7

Table 25. **Comparison of state-of-the-art supervised methods on ACDC for snow.** The first group of rows presents condition-specific expert models trained only on snow, while the second group presents uber models trained on all conditions.

Method	road	sidew.	build.	wall	fence	pole	light	sign	veget.	terrain	sky	person	rider	car	truck	bus	train	motorc.	bicycle	mIoU
RefineNet	90.1	65.7	86.4	31.2	48.1	58.0	76.7	70.3	89.7	45.7	97.3	70.8	15.4	87.1	35.0	43.1	79.1	38.7	59.9	62.5
DeepLabv2	89.1	61.7	82.7	26.4	40.9	35.5	56.5	54.1	85.2	39.0	95.1	55.0	25.7	84.3	38.6	53.8	77.6	29.0	49.5	56.8
DeepLabv3+	91.9	70.9	90.1	48.9	52.0	62.2	79.2	74.5	92.0	47.0	97.6	78.2	35.9	90.4	61.7	64.3	89.2	43.9	69.4	70.5
HRNet	93.6	75.2	89.0	42.0	55.6	67.7	83.3	78.9	93.0	48.9	97.8	78.1	16.4	92.6	54.8	61.6	87.0	50.0	68.9	70.2
RefineNet	90.2	65.7	86.5	33.7	50.6	57.8	78.0	71.5	89.2	44.5	97.0	73.8	46.0	88.4	50.0	48.0	79.9	40.6	60.3	65.9
DeepLabv2	88.7	62.5	82.5	35.3	41.7	35.0	59.0	52.8	84.4	36.0	95.2	58.1	29.8	84.8	48.9	30.9	77.9	32.9	48.4	57.1
DeepLabv3+	91.4	69.6	88.8	48.8	53.9	60.6	79.5	72.9	90.5	44.7	97.4	77.4	37.2	90.0	64.3	55.0	87.8	41.7	70.0	69.6
HRNet	94.4	77.3	91.5	53.1	63.6	70.2	85.1	81.4	92.1	57.7	97.7	83.3	69.6	93.6	71.8	54.5	86.3	52.7	73.1	76.3

tial nature of the GPS measurements from the two recordings in computing the correspondence function A. In particular, we enforce $k < i \Rightarrow A(k) \leq A(i)$. That is, for a given sample i of the adverse-condition sequence P, its matched sample $A(i)$ of the normal-condition sequence R is restricted to not precede in time any sample of R that has

Table 26. **Uncertainty-aware semantic segmentation baseline results on the complete test set of ACDC including all conditions.** Supervised methods for standard semantic segmentation are trained and evaluated jointly on all conditions for semantic label prediction. Confidence prediction baselines: max-softmax network outputs (Max-Softmax) and ground-truth invalid masks (GT).

Method	Confidence	road	sidew.	build.	wall	fence	pole	light	sign	veget.	terrain	sky	person	rider	car	truck	bus	train	motorc.	bicycle	mAUIoU
RefineNet	Max-Softmax	91.3	67.6	84.4	34.3	42.1	49.9	64.7	64.2	85.8	54.6	95.3	59.6	34.4	84.6	51.9	60.6	70.6	43.3	48.9	62.5
RefineNet	GT	92.9	73.1	89.1	43.1	50.7	57.0	72.9	70.7	90.1	63.4	97.7	67.6	43.1	87.3	57.3	61.4	77.1	54.1	58.3	68.8
DeepLabv2	Max-Softmax	87.1	60.4	79.7	36.1	35.7	32.6	47.3	48.7	80.2	49.2	92.2	49.0	24.7	79.0	51.1	43.3	72.3	26.3	45.1	54.7
DeepLabv2	GT	88.5	64.4	84.2	40.9	41.8	37.8	54.0	54.2	86.4	54.9	96.0	53.6	30.3	81.8	52.5	42.7	73.6	33.3	47.6	58.9
DeepLabv3+	Max-Softmax	92.1	71.3	88.2	49.0	47.3	54.9	68.7	65.6	88.0	60.7	96.0	65.0	33.9	87.5	66.7	72.6	81.3	43.8	55.0	67.8
DeepLabv3+	GT	93.8	76.5	91.4	56.6	55.4	62.3	75.0	72.3	91.8	66.5	98.0	72.0	41.0	89.5	71.1	74.0	86.5	55.4	63.7	73.3

Table 27. **Uncertainty-aware semantic segmentation baseline results on ACDC for fog.** Supervised methods for standard semantic segmentation are trained and evaluated on fog for semantic label prediction. Confidence prediction baselines: max-softmax network outputs (Max-Softmax) and ground-truth invalid masks (GT).

Method	Confidence	road	sidew.	build.	wall	fence	pole	light	sign	veget.	terrain	sky	person	rider	car	truck	bus	train	motorc.	bicycle	mAUIoU
RefineNet	Max-Softmax	92.6	71.9	82.9	40.7	35.8	42.7	62.1	62.6	84.1	64.1	97.5	45.0	26.8	77.1	57.8	59.9	79.8	35.2	33.4	60.6
RefineNet	GT	93.4	76.5	87.6	48.7	45.5	49.4	68.2	68.9	87.3	73.0	98.1	55.6	40.3	80.9	61.3	65.4	83.7	53.6	51.7	67.9
DeepLabv2	Max-Softmax	89.7	63.0	79.2	39.4	25.9	25.0	41.4	46.6	82.5	66.7	95.6	36.4	35.6	72.7	49.5	29.6	44.5	29.2	33.3	51.9
DeepLabv2	GT	90.2	66.7	82.8	44.2	35.3	31.5	49.5	52.2	84.8	69.4	96.9	44.2	44.5	76.0	48.3	30.1	39.0	48.0	42.7	56.7
DeepLabv3+	Max-Softmax	92.9	74.8	87.2	51.3	41.7	49.9	65.6	69.8	87.1	72.3	97.6	51.9	27.1	82.8	67.4	79.1	84.1	42.6	36.4	66.4
DeepLabv3+	GT	93.9	78.3	90.0	55.5	52.0	57.9	72.3	75.9	89.2	76.6	98.4	63.2	38.5	85.0	71.7	85.1	86.7	66.0	53.3	73.1

Table 28. **Uncertainty-aware semantic segmentation baseline results on ACDC for nighttime.** Supervised methods for standard semantic segmentation are trained and evaluated on nighttime for semantic label prediction. Confidence prediction baselines: max-softmax network outputs (Max-Softmax) and ground-truth invalid masks (GT).

Method	Confidence	road	sidew.	build.	wall	fence	pole	light	sign	veget.	terrain	sky	person	rider	car	truck	bus	train	motorc.	bicycle	mAUIoU
RefineNet	Max-Softmax	92.3	66.4	78.6	31.8	37.2	46.2	48.3	53.3	73.5	16.9	83.6	54.9	34.6	77.4	8.5	43.1	53.6	35.2	41.6	51.4
RefineNet	GT	93.6	72.4	88.2	42.0	53.0	55.5	61.6	61.7	89.0	31.3	97.1	63.3	41.9	80.0	18.2	50.3	60.8	49.5	51.9	61.1
DeepLabv2	Max-Softmax	90.2	62.2	78.6	29.9	32.9	33.7	36.5	40.3	65.6	25.2	77.9	45.2	23.2	70.2	5.0	14.6	62.1	40.3	38.8	45.9
DeepLabv2	GT	90.8	65.8	87.2	37.8	45.3	43.3	48.1	49.6	87.8	37.5	97.0	51.1	29.8	74.3	17.3	17.3	63.0	51.8	43.8	54.7
DeepLabv3+	Max-Softmax	93.8	73.3	85.2	47.0	43.4	51.3	53.7	54.3	80.7	28.7	87.9	62.1	40.9	84.8	10.4	65.2	78.8	34.7	47.2	59.1
DeepLabv3+	GT	94.9	77.5	91.5	54.7	53.4	60.2	64.8	62.5	92.7	41.3	98.5	70.2	49.3	88.3	22.4	65.5	82.4	50.5	55.0	67.1

Table 29. **Uncertainty-aware semantic segmentation baseline results on ACDC for rain.** Supervised methods for standard semantic segmentation are trained and evaluated on rain for semantic label prediction. Confidence prediction baselines: max-softmax network outputs (Max-Softmax) and ground-truth invalid masks (GT).

Method	Confidence	road	sidew.	build.	wall	fence	pole	light	sign	veget.	terrain	sky	person	rider	car	truck	bus	train	motorc.	bicycle	mAUIoU
RefineNet	Max-Softmax	86.0	67.8	89.9	44.9	45.7	53.2	65.1	67.3	92.1	48.4	97.8	58.6	23.6	86.6	44.1	53.1	65.6	40.3	56.6	62.5
RefineNet	GT	89.5	70.8	92.1	54.1	53.2	59.9	72.6	72.3	93.9	52.1	98.4	67.4	26.6	88.7	52.4	56.4	75.5	51.4	62.9	67.9
DeepLabv2	Max-Softmax	85.9	62.3	87.2	48.3	38.9	35.8	48.6	51.5	87.3	41.8	95.9	47.2	13.5	80.8	46.2	50.2	69.3	23.9	50.0	56.0
DeepLabv2	GT	87.8	65.1	89.4	52.1	42.5	40.2	53.7	56.1	89.6	43.6	96.8	53.4	13.8	82.7	50.2	48.1	72.9	33.3	51.4	59.1
DeepLabv3+	Max-Softmax	91.2	75.3	92.8	62.2	53.7	60.0	71.3	72.2	93.2	50.0	98.0	65.4	30.8	90.0	63.5	77.0	83.1	48.0	63.9	70.6
DeepLabv3+	GT	93.2	78.4	94.2	68.8	60.0	66.0	75.8	78.2	94.5	52.5	98.6	72.4	35.0	91.0	70.4	80.4	87.4	58.8	69.0	75.0

been matched to a sample k of P that precedes i. This constraint is based on the fact that the routes of the two recordings are driven in the same direction and thus in the same order. Consequently, for routes that contain loops, our formulation prevents the matching of samples that are nearest neighbors but correspond to *different* passes from the same location and are thus potentially associated with different driving directions and 3D rotations of the camera.

D.3. Annotation

In Fig. 5, we show the percentage of the pixels of each semantic class in ACDC that are marked as invalid in the ground-truth invalid mask J. For the majority of the classes, a notable percentage of more than 5% of the pixels are la-

Table 30. **Uncertainty-aware semantic segmentation baseline results on ACDC for snow.** Supervised methods for standard semantic segmentation are trained and evaluated on snow for semantic label prediction. Confidence prediction baselines: max-softmax network outputs (Max-Softmax) and ground-truth invalid masks (GT).

Method	Confidence	road	sidew.	build.	wall	fence	pole	light	sign	veget.	terrain	sky	person	rider	car	truck	bus	train	motorc.	bicycle	mAUIoU
RefineNet	Max-Softmax	89.1	59.9	83.8	25.8	43.8	53.1	72.6	69.2	88.6	43.5	96.8	65.9	11.7	85.8	39.5	48.4	74.1	36.9	48.8	59.9
RefineNet	GT	91.3	69.1	86.8	32.4	49.9	59.0	78.2	72.8	90.0	52.5	97.3	71.8	16.1	87.6	37.6	44.7	79.5	39.8	60.1	64.0
DeepLabv2	Max-Softmax	89.1	61.7	82.7	26.4	40.9	35.5	56.5	54.1	85.2	39.0	95.1	55.0	25.7	84.3	38.6	53.8	77.6	29.0	49.5	56.8
DeepLabv2	GT	90.3	65.1	83.1	27.6	42.7	36.5	57.9	56.7	85.5	46.3	95.1	56.4	26.4	85.0	41.1	55.0	78.2	30.2	49.8	58.4
DeepLabv3+	Max-Softmax	90.6	67.0	88.8	45.1	48.9	57.8	76.6	72.9	90.8	45.7	97.0	74.8	28.4	89.2	63.3	67.8	87.8	36.7	61.1	67.9
DeepLabv3+	GT	92.9	74.0	90.4	50.3	53.9	63.4	80.5	77.4	92.2	53.6	97.6	79.2	36.6	90.9	64.4	65.9	90.0	45.2	69.8	72.0

Algorithm 1 Dynamic programming algorithm for GPS sequence matching

Input: Adverse-condition GPS sequence $P = (\mathbf{p}_1, \ldots, \mathbf{p}_n)$, normal-condition GPS sequence $R = (\mathbf{r}_1, \ldots, \mathbf{r}_m)$
Output: Correspondence function $A : \{1, \ldots, n\} \to \{1, \ldots, m\}$

- ▷ Compute pairwise Euclidean distances of GPS samples
- $d_{ij} \leftarrow \|\mathbf{p}_i - \mathbf{r}_j\|,\ 1 \leq i \leq n,\ 1 \leq j \leq m$
- ▷ Compute cost matrix C $(n \times m)$
- $C_{1j} \leftarrow d_{1j},\ 1 \leq j \leq m$
- $C_{ij} \leftarrow \min_{k \leq j}\{C_{i-1,k}\} + d_{ij},\ 2 \leq i \leq n,\ 1 \leq j \leq m$
- ▷ Compute backtracking indices matrix α
- $\alpha_{ij} \leftarrow \arg\min_{k \leq j}\{C_{i-1,k}\},\ 2 \leq i \leq n,\ 1 \leq j \leq m$
- ▷ Backtracking
- $A(n) \leftarrow \arg\min_{j}\{C_{nj}\}$
- $A(i) \leftarrow \alpha_{i+1,A(i+1)},\ 1 \leq i \leq n-1$

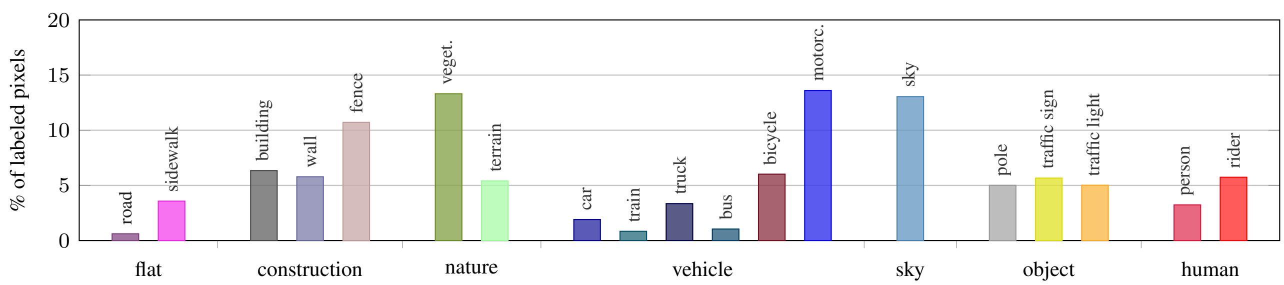


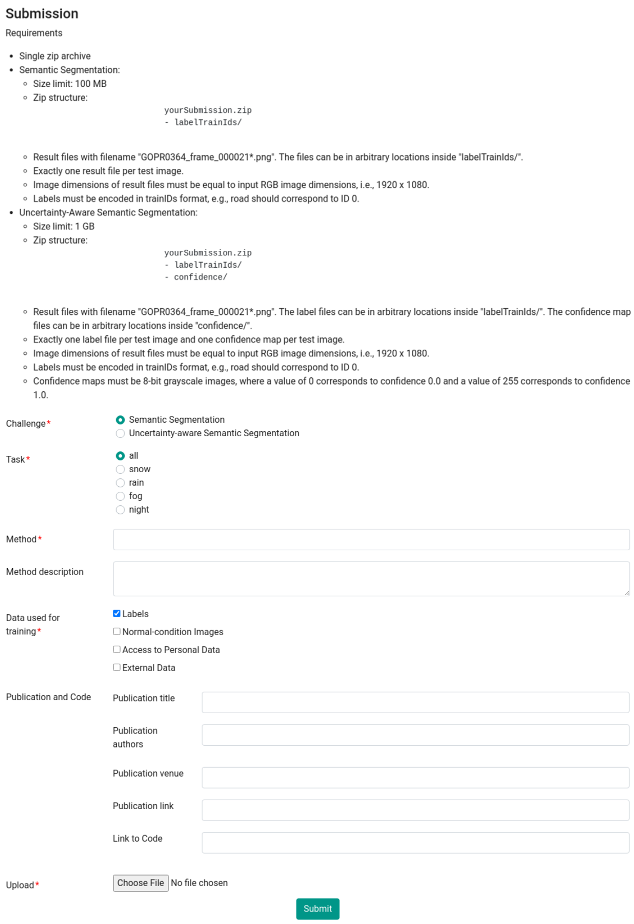
Figure 5. Per-class percentages of labeled pixels that are marked as invalid in ACDC.

beled as invalid, which demonstrates the ability of our specialized annotation protocol with privileged information to assign a legitimate semantic label even to invalid regions with ambiguous semantic content.

The total number of annotated pixels in ACDC is presented in Table 31. Note that labeled pixels that are marked as valid in the ground-truth invalid masks J constitute ca. 85% of the total pixels. From the remaining 15% of pixels that did not receive a legitimate semantic label in stage 1 of the annotation because of their ambiguity, it was possible to label *half* of them (7.5%) with a legitimate semantic label in stage 2 of the annotation, by making use of the additional privileged information in the form of corresponding normal-condition images and original adverse-condition videos. Note that for stage 2 of the annotation, we explicitly set the time budget (excluding quality control) to 20 minutes and asked the annotators to prioritize labeling of (i) traffic participants and (ii) distant and/or unclear objects that were

Table 31. **Overall annotation statistics for ACDC.** We report the total number of pixels assigned to a legitimate semantic label (Labeled) and of pixels not assigned to any semantic label (Unlabeled) as well as the respective percentages.

	Number of pixels	Percentage of pixels (%)
Labeled	7.682×10^9	92.47
-out of which Valid	7.055×10^9	84.93
-out of which Invalid	0.627×10^9	7.54
Unlabeled	0.625×10^9	7.53
Total	8.307×10^9	100.00



Submission

Requirements

- Single zip archive
- Semantic Segmentation:
  - Size limit: 100 MB
  - Zip structure:

```
yourSubmission.zip
- labelTrainIds/
```

  - Result files with filename "GOPR0364_frame_000021*.png". The files can be in arbitrary locations inside "labelTrainIds/".
  - Exactly one result file per test image.
  - Image dimensions of result files must be equal to input RGB image dimensions, i.e., 1920 x 1080.
  - Labels must be encoded in trainIDs format, e.g., road should correspond to ID 0.
- Uncertainty-Aware Semantic Segmentation:
  - Size limit: 1 GB
  - Zip structure:

```
yourSubmission.zip
- labelTrainIds/
- confidence/
```

  - Result files with filename "GOPR0364_frame_000021*.png". The label files can be in arbitrary locations inside "labelTrainIds/". The confidence map files can be in arbitrary locations inside "confidence/".
  - Exactly one label file per test image and one confidence map per test image.
  - Image dimensions of result files must be equal to input RGB image dimensions, i.e., 1920 x 1080.
  - Labels must be encoded in trainIDs format, e.g., road should correspond to ID 0.
  - Confidence maps must be 8-bit grayscale images, where a value of 0 corresponds to confidence 0.0 and a value of 255 corresponds to confidence 1.0.

Challenge*
Semantic Segmentation
Uncertainty-aware Semantic Segmentation

Task*
all
snow
rain
fog
night

Method*

Method description

Data used for training*
Labels
Normal-condition Images
Access to Personal Data
External Data

Publication and Code
Publication title
Publication authors
Publication venue
Publication link
Link to Code

Upload*
Choose File No file chosen

Submit

Figure 6. **The submission page of our benchmark website.** Our evaluation server supports the two tasks and five condition configurations of ACDC, accepting submissions both for individual conditions and for all conditions. Best viewed on a screen.

affected the most by the adverse conditions at the time of capture.