# ACDC: The Adverse Conditions Dataset with Correspondences for Semantic Driving Scene Understanding

**Conference Paper**

**Author(s):**
Sakaridis, Christos; Dai, Dengxin; Van Gool, Luc

**Publication date:**
2021

**Permanent link:**
https://doi.org/10.3929/ethz-b-000518792

**Originally published in:**
https://doi.org/10.1109/ICCV48922.2021.01059

# ACDC: The Adverse Conditions Dataset with Correspondences for Semantic Driving Scene Understanding

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

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

https://acdc.vision.ee.ethz.ch

## Abstract

*Level 5 autonomy for self-driving cars requires a robust visual perception system that can parse input images under any visual condition. However, existing semantic segmentation datasets are either dominated by images captured under normal conditions or are small in scale. To address this, we introduce ACDC, the Adverse Conditions Dataset with Correspondences for training and testing semantic segmentation methods on adverse visual conditions. ACDC consists of a large set of 4006 images which are equally distributed between four common adverse conditions: fog, nighttime, rain, and snow. Each adverse-condition image comes with a high-quality fine pixel-level semantic annotation, a corresponding image of the same scene taken under normal conditions, and a binary mask that distinguishes between intra-image regions of clear and uncertain semantic content. Thus, ACDC supports both standard semantic segmentation and the newly introduced uncertainty-aware semantic segmentation. A detailed empirical study demonstrates the challenges that the adverse domains of ACDC pose to state-of-the-art supervised and unsupervised approaches and indicates the value of our dataset in steering future progress in the field. Our dataset and benchmark are publicly available.*

## 1. Introduction

Most of the prominent large-scale image-based datasets for driving scene understanding, including Cityscapes [8], Vistas [28] and KITTI [13], are dominated by images captured under normal visual conditions, i.e., at daytime and in clear weather. Yet, vision applications such as autonomous driving impose a strict requirement on perception algorithms to maintain satisfactory performance in adverse domains. Although there are recent efforts to include adverse visual domains in large-scale datasets, such as Oxford RobotCar [27] and BDD100K [55], these efforts focus either on localization/mapping tasks [27,49] or on recognition tasks which *do not involve dense pixel-level outputs*, such as object detection [3,42,55]. For instance, while a notable 40% of the object detection set of BDD100K pertains to nighttime, only 3% of the images in its semantic segmentation set, namely 345 images, are captured at nighttime [40]. In addition, the pixel-level annotation process for adverse-condition images is kept identical in [55] to the normal-condition case, which leads to errors in the ground truth and renders it unreliable [40]. In contrast, seminal previous work [8] has underlined the need for *specialized* techniques and datasets for pixel-level semantic scene understanding in adverse visual conditions, due to the inherent aleatory uncertainty in images captured in such conditions. These render entire image regions indiscernible even for humans.

ACDC constitutes a response to this need for a large-scale driving dataset specialized to adverse conditions, in terms of (i) size, (ii) domain adversity, and (iii) featured tasks. ACDC includes 4006 images with high-quality pixel-level semantic annotations, which are distributed equally among four common adverse conditions in real-world driving environments, namely fog, nighttime, rain, and snow, thus featuring a scale of the same order as Cityscapes. The dataset was deliberately recorded with the respective adverse conditions clearly present. Thus, a large domain shift from the normal clear-weather daytime conditions was achieved. Moreover, for each adverse-condition image, a corresponding normal-condition image of the same scene from approximately the same viewpoint is provided, intended for use by weakly supervised methods.

As to the tasks that our dataset supports, apart from standard semantic segmentation, we add the task of uncertainty-aware semantic segmentation. For the latter we intro-

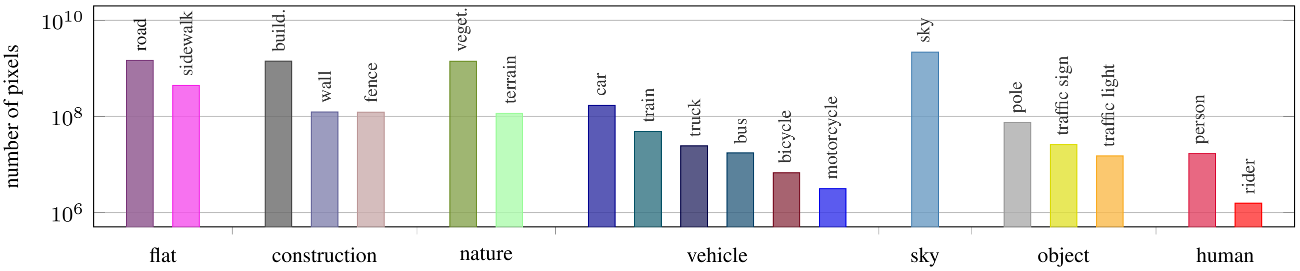

Figure 1. Number of finely annotated pixels per class in ACDC.

duce a specialized annotation protocol and a dedicated performance metric, termed average uncertainty-aware IoU (AUIoU). The key characteristic of uncertainty-aware semantic segmentation is the principled inclusion of image regions with indiscernible semantic content—*invalid* regions—in annotation and evaluation. In particular, the annotation protocol for our adverse-condition images leverages privileged information in the form of the corresponding normal-condition images and the original adverse-condition videos, which enables to *reliably* assign legitimate semantic labels to invalid regions and to include them in the evaluation both for standard and uncertainty-aware semantic segmentation. For the latter task, the separation of labeled pixels into invalid and valid is encoded in a binary mask. While both tasks require a hard semantic prediction, the uncertainty-aware task additionally expects a confidence map prediction. AUIoU is designed to take into account both the semantic and the confidence prediction and to reward predictions with low confidence on invalid pixels and high confidence on valid pixels. The requirement for an additional confidence prediction is relevant for safety-oriented applications, as it can help the downstream decision-making system avoid the fatal consequences of a low-confidence prediction being false, e.g. when a pedestrian is missed.

Apart from being a challenging benchmark for supervised semantic segmentation approaches, ACDC is a well-suited test bed for domain adaptation. A multitude of recent works [7, 15, 22, 23, 26, 41, 43, 44, 46, 48, 51, 53, 59, 60, 62, 65, 66] have focused on unsupervised domain adaptation (UDA) for semantic segmentation, but most of them are validated only on an artificial synthetic-to-real setting, using GTA5 [34] and SYNTHIA [36] as source datasets and Cityscapes [8] as the target dataset. The *normal-to-adverse domain adaptation* scenario for semantic segmentation, which is much more relevant for real-world deployment of autonomous cars due to the difficulty of both acquiring and annotating adverse-condition data, has largely been overlooked. In particular, much fewer works consider normal-to-adverse adaptation in their experiments [10,11,32,37,38,39,40] and whenever they do, they either restrict the target adverse domain to a single condition, e.g. nighttime [10, 39, 40], fog [37, 38], or rain [11], or do not include a quantitative evaluation on the real target domain altogether [32]. We attribute this fragmentation of normal-to-adverse adaptation works to the absence of a general large-scale dataset for semantic segmentation that evenly covers the majority of common adverse conditions and provides reliable ground truth for a sound evaluation in such challenging domains. ACDC answers exactly the need for such a dataset and will serve as a test bed for unsupervised and weakly supervised domain adaptation. Experiments such as Cityscapes→ACDC adaptation are straightforward thanks to the identical label sets of the two datasets, which facilitates validation of new domain adaptation approaches in the normal-to-adverse setting.

We experiment with ACDC in four main directions: evaluation of models pre-trained on normal conditions, supervised learning in adverse conditions, unsupervised and weakly supervised normal-to-adverse domain adaptation, and evaluation of uncertainty-aware semantic segmentation baselines and oracles. Results show that access to ground-truth annotations under adverse conditions is indispensable for achieving high performance, as pre-trained models severely deteriorate under adverse conditions. Moreover, the real-world Cityscapes→ACDC adaptation scenario poses significant challenges to all state-of-the-art UDA methods, which recover at best only a small portion of the performance gain over the source-domain model compared to using full supervision. This underlines the need for UDA methods that perform better when handling adverse target domains and highlights the importance of ACDC in steering future work in this direction. Finally, the uncertainty-aware annotations of ACDC create significant room for improvement over simple confidence prediction baselines and help promote future work on semantic segmentation methods that simultaneously models uncertainty.

## 2. Related Work

**Datasets for driving scene understanding** include real-world and synthetic sets that support geometric and recognition tasks. KITTI [13] and Cityscapes [8] pioneered this area with LiDAR and semantic image annotations, respectively. Subsequent datasets mostly aimed at increasing the scale [17], diversity [28] and number of tasks [55]. As high-quality pixel-level annotations proved hard to acquire [8, 28], another line of work focused on creating synthetic sets at an even larger scale [19, 33, 34, 36] and in which ground truth is automatically generated, as well as translating real datasets to adverse conditions such as fog or rain [14, 37, 38]. Oxford Robotcar [27] was the first real-world large-scale dataset in which adverse visual conditions such as nighttime, rain and snow were significantly represented, but it did not feature semantic annotations. While more recent large-scale sets [2, 30] that cover adverse conditions, such as Waymo Open [42] and nuScenes [3], include bounding boxes, they still lack dense pixel-level semantic annotations, which are vital for real-world autonomous agents [63]. BDD100K [55] is the only exception to this rule, with ca. 13% of its 10000 pixel-level annotations pertaining to adverse conditions but containing severe errors [40], while only a small portion of each of the 1881 adverse-condition images in ADUULM [29] is annotated. On the other hand, several sets with small-scale pixel-level annotations covering adverse conditions [58] were recently presented, focusing on fog [9, 38], nighttime [10, 40], and rain [45]. A notable case is Dark Zurich [40], with 201 fine pixel-level nighttime annotations and a dedicated annotation protocol and evaluation metric that handles regions with ambiguous content. ACDC improves both upon BDD100K, in terms of ground truth quality, and Dark Zurich, in terms of scale and condition diversity, featuring 4006 high-quality fine pixel-level annotations in which fog, night, rain and snow are equally represented.

**Semantic segmentation** has progressed rapidly over the last years, primarily through the design of convolutional neural networks. Based on fully convolutional architectures [25], seminal works introduced atrous convolution [4, 5, 56] and encoder-decoder structures with skip connections [35] to exploit context and improve localization, respectively. Balancing between global and local information was further addressed by parallel branches of different resolutions [24, 31] and global pooling [61]. Other works focused on real-time performance [54], leveraging different modalities such as depth [52], and defining neighborhood-based supervision [20] for segmentation. The current state of the art includes i.a. DeepLabv3+ [6] and ANN [64] with pyramid pooling modules, DANet [12] and CCNet [18] with attention mechanisms, and HRNet [47] and OCR [57] with high-resolution representations. While performance on the popular Cityscapes benchmark is increasingly saturating, we demonstrate that state-of-the-art methods achieve much lower performance on ACDC (see Sec. 4). Thus, ACDC provides a more challenging benchmark for semantic segmentation thanks to the adversity of its domains and is therefore able to foster further progress in the field.

**Adaptation of semantic segmentation** networks to domains where full supervision is not available was launched shortly after the introduction of supervised approaches [16]. A major class of UDA works employs adversarial domain adaptation to implicitly align the source and target domains at the level of pixels and/or features [7, 15, 26, 41, 43, 44, 46, 48, 60]. Other approaches to UDA rely on self-training with pseudo-labels in the target domain [65, 66] or combine self-training with adversarial adaptation [23] or with pixel-level adaptation via explicit transforms from source to target [22, 53]. However, all aforementioned approaches have been evaluated only on the artificial scenario of synthetic-to-real adaptation and overlook *normal-to-adverse adaptation*, which is of higher practical importance for autonomous cars. ACDC constitutes the large-scale target-domain dataset which has been missing so far for such a normal-to-adverse experiment and aims to steer the development of unsupervised and weakly supervised adaptation approaches that can cope with adverse target domains.

## 3. ACDC Dataset

We base the design of ACDC on the same general principles as seminal normal-condition datasets [8] and adapt the collection and annotation process to fit better the adverse condition setting at hand.

### 3.1. Collection

Our data collection is guided by the decision to record the same set of scenes both under adverse and normal conditions. We define the domain of *normal* conditions as the combination of daytime and clear weather, i.e. good visibility and no precipitation or snow cover on the ground. While the focus of ACDC is on adverse conditions, the acquisition of the corresponding normal-condition images is vital both for the subsequent annotation step and to support weakly supervised methods, as the same scene can be much easier to parse in normal conditions, both for humans and machines.

Thus, we recorded several days of video in Switzerland by driving around in a car, primarily in urban areas but also on highways and in rural regions. In order to have a clear domain separation between different adverse conditions, we use the following criterion for the adverse-condition recordings: each recording takes place under only one type of adversity from a set of four items, i.e., fog, nighttime, rain, and snow. For example, our foggy recordings are performed at daytime and without rain or snow. For snow, both snowfall and snow cover on the ground are admissible. Moreover, we keep for further processing only the parts of the adverse-condition recordings that correspond to an intense presence of the respective condition, so as to maximize the domain shift from normal conditions as well as domain adversity.

We record with a 1080p GoPro Hero 5 camera, mounted in front of the windshield at nighttime and in normal conditions and behind the windshield in fog, rain, and snow. The camera records 8-bit RGB frames at a rate of 30 Hz.

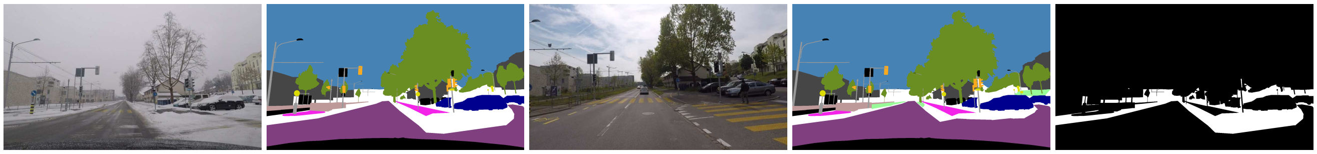
(a) Input image $I$ (b) Stage 1 annotation (draft) (c) Corresponding image $I'$ (d) Stage 2 annotation (GT) (e) Invalid mask $J$

Figure 2. **Illustration of annotation protocol for ACDC.** The color coding of the semantic classes matches Fig. 1. All annotations in (b), (d) and (e) pertain to the input image $I$ in (a). A white color in (b) and (d) denotes unlabeled pixels.

### 3.2. Correspondence Establishment

Our camera also provides GPS readings, which allow us to establish *image-level correspondences* between adverse-condition and normal-condition recordings. In particular, for each adverse-condition recording, we perform a normal-condition recording along exactly the same route. We then use the sequences of GPS measurements of the two recordings to perform a global dynamic-programming-based matching of the adverse GPS sequence to the normal one, where the objective is defined by the Euclidean distances of matched pairs of GPS samples. Our global matching handles routes with loops better than simple nearest neighbors. Each adverse-condition frame is then matched to a normal-condition frame based on the corresponding matched samples of the GPS sequences.

### 3.3. Dataset Splits

ACDC is split into four sets corresponding to the examined conditions. We manually selected 1000 foggy, 1006 nighttime, 1000 rainy and 1000 snowy images from the recordings for dense pixel-level semantic annotation, for a total of 4006 adverse-condition images. The selection process aimed at maximizing the complexity and diversity of captured scenes. Within each recording, any pair of selected images is at least 20 s or 50 m apart (whatever comes first).

The dataset is also split into training, validation, and test sets. We apply a global geographical split across all conditions, so that there is zero overlap between the three sets, even for different conditions. Given the abundance of training data from normal-condition datasets [8, 28, 55] that allow to pre-train semantic segmentation models, we opt for a split with a greater test set size than usual. This aims at providing a highly challenging benchmark for semantic segmentation, both in terms of scale and domain adversity. In particular, we split the set of each adverse condition into 400 training, 100 validation and 500 test images, except the nighttime set with 106 validation images. This results in a total of 1600 training and 406 validation images with public annotations and 2000 test images with annotations withheld for benchmarking purposes, as per standard practice [8].

### 3.4. Annotation

Images captured under adverse conditions contain invalid regions, i.e. regions with indiscernible semantic content, which generally co-exist with valid regions in the same image. We take this into account for creating annotations of ACDC and design a specialized annotation protocol, which leverages privileged information from the corresponding normal-condition images and the original adverse-condition videos and allows (i) the reliable assignment of semantic labels to invalid regions and (ii) the creation of a binary mask that distinguishes valid from invalid regions.

Our annotation protocol consists of two cascaded annotation stages. At stage 1, a semantic labeling draft is manually produced from the adverse-condition image $I$, in which pixels that cannot be unquestionably assigned to a single semantic class are left unlabeled. At stage 2, the corresponding normal-condition image $I'$ and the adverse-condition video from which $I$ was extracted are used to augment and finalize the annotation. In particular, the annotator can assign a legitimate label to pixels that were left unlabeled in stage 1 and correct pixels that were incorrectly labeled in stage 1. Pixels that remain unclear in stage 2 are left unlabeled and are not used for training or evaluation.

The final annotation outputs are twofold: (i) the final semantic annotation $H$ after stage 2, and (ii) a binary invalid mask $J$, where pixels whose label changed from stage 1 to stage 2 are set to 1 (invalid) and pixels with the same semantic label for both stages are set to 0 (valid). $J$ enables the introduction of the new task of uncertainty-aware semantic segmentation, which we detail in Sec. 5.

The 4006 fine-pixel annotations of ACDC were created by a professional team of annotators to ensure high-quality ground truth. Annotators were asked to be conservative in labeling pixels in both stages, so as to minimize errors. Both the initial draft from stage 1 and the final annotation from stage 2 passed through quality control. The total time required for annotating a single image was 3.3 h on average.

The class specifications of ACDC are directly inherited from Cityscapes. In particular, we annotate the 19 evaluation classes of Cityscapes, which include the most common and traffic-related objects in driving scenes. Objects that belong to classes outside this set receive a fall-back label and are not used for training or evaluation. This choice of classes provides full compatibility of ACDC to Cityscapes and other normal-condition datasets for semantic segmentation [28, 55]. Detailed annotation statistics are presented in Fig. 1. An example of our two-stage annotation protocol is shown in Fig. 2 for a snowy image. Note the assignment of a region in the lower right part of the image that is unlabeled

Table 1. **Comparison of ACDC against adverse-condition semantic segmentation datasets.** "Adverse annot.": total annotated adverse-condition images, "Fog"/"Night"/"Rain"/"Snow": annotated foggy/nighttime/rainy/snowy images, "Inv. regions": can invalid regions get legitimate labels?, "Corr. normal": are corresponding normal-condition images available?, "Inv. masks": are invalid masks available?

| Dataset | Adverse annot. | Fog | Night | Rain | Snow | Classes | Reliable GT | Fine GT | Inv. regions | Corr. normal | Inv. masks |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Foggy Driving [38] | 101 | 101 | 0 | 0 | 0 | 19 | ✓ | ✓ | × | × | × |
| Foggy Zurich [9] | 40 | 40 | 0 | 0 | 0 | 19 | ✓ | ✓ | × | × | × |
| Nighttime Driving [10] | 50 | 0 | 50 | 0 | 0 | 19 | ✓ | × | × | × | × |
| Dark Zurich [40] | 201 | 0 | 201 | 0 | 0 | 19 | ✓ | ✓ | ✓ | ✓ | ✓ |
| Raincouver [45] | 326 | 0 | 95 | 326 | 0 | 3 | ✓ | × | × | × | × |
| WildDash [58] | 226 | 10 | 13 | 13 | 26 | 19 | ✓ | ✓ | × | × | × |
| BDD100K [55] | 1346 | 23 | 345 | 213 | 765 | 19 | × | ✓ | × | × | × |
| ACDC | **4006** | **1000** | **1006** | **1000** | **1000** | 19 | ✓ | ✓ | ✓ | ✓ | ✓ |

Table 2. **Comparison of state-of-the-art domain adaptation methods on Cityscapes→ACDC adaptation.** Cityscapes serves as the source domain and the entire ACDC including all four conditions serves as the target domain. The first and second groups of rows present unsupervised and weakly supervised methods, respectively. All unsupervised methods share the same network architecture. The performance of the respective models trained on Cityscapes (Source model) and of the oracle models trained on ACDC with 100 labels (Oracle-100), 200 labels (Oracle-200), and all 1600 labels (Oracle) are 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 [5] | 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 |
| AdaptSegNet [43] | 69.4 | 34.0 | 52.8 | 13.5 | 18.0 | 4.3 | 14.9 | 9.7 | 64.0 | 23.1 | 38.2 | 38.6 | 20.1 | 59.3 | 35.6 | 30.6 | 53.9 | 19.8 | 33.9 | 33.4 |
| ADVENT [46] | 72.9 | 14.3 | 40.5 | 16.6 | 21.2 | 9.3 | 17.4 | 21.2 | 63.8 | 23.8 | 18.3 | 32.6 | 19.5 | 69.5 | 36.2 | 34.5 | 46.2 | 26.9 | 36.1 | 32.7 |
| BDL [23] | 56.0 | 32.5 | 68.1 | 20.1 | 17.4 | 15.8 | 30.2 | 28.7 | 59.9 | 25.3 | 37.7 | 28.7 | 25.5 | 70.2 | 39.6 | 40.5 | 52.7 | 29.2 | 38.4 | 37.7 |
| CLAN [26] | 79.1 | 29.5 | 45.9 | 18.1 | 21.3 | 22.1 | 35.3 | 40.7 | 67.4 | 29.4 | 32.8 | 42.7 | 18.5 | 73.6 | 42.0 | 31.6 | 55.7 | 25.4 | 30.7 | 39.0 |
| CRST [65] | 51.7 | 24.4 | 67.8 | 13.3 | 9.7 | 30.2 | 38.2 | 34.1 | 58.0 | 25.2 | 76.8 | 39.9 | 17.1 | 65.4 | 3.7 | 6.6 | 39.6 | 11.8 | 8.6 | 32.8 |
| FDA [53] | 73.2 | 34.7 | 59.0 | 24.8 | 29.5 | 28.6 | 43.3 | 44.9 | 70.1 | 28.2 | 54.7 | 47.0 | 28.5 | 74.6 | 44.8 | 52.3 | 63.3 | 28.3 | 39.5 | 45.7 |
| SIM [48] | 53.8 | 6.8 | 75.5 | 11.6 | 22.3 | 11.7 | 23.4 | 25.7 | 66.1 | 8.3 | 80.6 | 41.8 | 24.8 | 49.7 | 38.6 | 21.0 | 41.8 | 25.1 | 29.6 | 34.6 |
| MRNet [62] | 72.2 | 8.2 | 36.4 | 13.7 | 18.5 | 20.4 | 38.7 | 45.4 | 70.2 | 35.7 | 5.0 | 47.8 | 19.1 | 73.6 | 42.1 | 36.0 | 47.4 | 17.7 | 37.4 | 36.1 |
| Oracle-100 | 84.4 | 54.8 | 76.4 | 19.3 | 28.9 | 29.5 | 36.5 | 42.6 | 74.2 | 40.3 | 87.7 | 42.5 | 16.5 | 74.9 | 36.5 | 28.6 | 55.9 | 27.3 | 38.6 | 47.1 |
| Oracle-200 | 86.2 | 55.0 | 77.9 | 21.7 | 30.9 | 30.0 | 37.6 | 42.5 | 76.8 | 45.8 | 90.2 | 45.4 | 19.1 | 75.8 | 38.5 | 38.0 | 64.2 | 21.6 | 39.5 | 49.3 |
| Oracle | 88.0 | 62.3 | 80.8 | 37.0 | 35.1 | 33.9 | 49.8 | 49.5 | 80.1 | 50.7 | 92.5 | 51.1 | 26.5 | 79.9 | 49.0 | 41.1 | 72.2 | 26.5 | 44.2 | 55.3 |
| Source model [24] | 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 |
| MGCDA [40] | 73.4 | 28.7 | 69.9 | 19.3 | 26.3 | 36.8 | 53.0 | 53.3 | 75.4 | 32.0 | 84.6 | 51.0 | 26.1 | 77.6 | 43.2 | 45.9 | 53.9 | 32.7 | 41.5 | 48.7 |
| Oracle | 92.5 | 71.2 | 86.2 | 39.0 | 44.0 | 53.2 | 68.8 | 66.0 | 85.1 | 59.3 | 94.9 | 65.2 | 38.5 | 85.8 | 53.8 | 59.7 | 76.2 | 47.5 | 54.5 | 65.3 |

at stage 1 (Fig. 2b) to the *road* label at stage 2 (Fig. 2d), thanks to the clear view from the normal-condition image.

### 3.5. Comparison to Related Datasets

To the best of our knowledge, ACDC constitutes the largest adverse-condition semantic segmentation dataset to date. In Table 1, we compare ACDC to existing datasets that also address semantic segmentation under adverse conditions. Most of these datasets focus on a single condition and are of small scale. WildDash covers a wider variety of adverse conditions but also has a small scale. BDD100K includes 10000 images with semantic segmentation annotations. We inspected these images manually to identify those that pertain to fog, night, rain, and snow. We found that only 1346/10000 images pertain to any of these four conditions. By contrast, ACDC is fully composed of these four common adverse conditions. Notably, it contains one order of magnitude more annotated images than any other competing dataset for each of fog, night and rain. At the same time, our specialized annotation protocol using corresponding normal-condition images ensures *reliable* annotations even for invalid regions, making ACDC a high-quality dataset for training and evaluation for adverse conditions.

## 4. Semantic Segmentation

The first task ACDC supports is standard semantic segmentation. All results in Sec. 4 are reported for the test set of ACDC using the IoU metric. We experiment for our dataset with domain adaptation methods, externally pre-trained models and supervised approaches.

### 4.1. Normal-to-Adverse Adaptation

We present a new benchmark for UDA of semantic segmentation: Cityscapes→ACDC. We select eight representative state-of-the-art UDA methods, train them with their default configurations for adaptation from Cityscapes to the entire ACDC and present the results in Table 2. All eight methods share the same DeepLabv2-based architecture [5]. Whereas these methods have achieved significant performance gains in the popular synthetic-to-real adaptation set-

Table 3. **Comparison of state-of-the-art unsupervised domain adaptation methods on Cityscapes→ACDC adaptation for individual conditions.** We train a separate model on each condition-specific subset of ACDC and evaluate each model on the condition it has been trained for. Performance of the model trained only on the source domain (Source model) and of oracles with access to the target domain labels for each condition (Oracle) is also reported.

| Method | Fog | Night | Rain | Snow |
|---|---|---|---|---|
| Source model | 33.5 | 30.1 | 44.5 | 40.2 |
| AdaptSegNet [43] | 31.8 | 29.7 | 49.0 | 35.3 |
| ADVENT [46] | 32.9 | 31.7 | 44.3 | 32.1 |
| BDL [23] | 37.7 | 33.8 | 49.7 | 36.4 |
| CLAN [26] | 39.0 | 31.6 | 44.0 | 37.7 |
| FDA [53] | 39.5 | 37.1 | 53.3 | 46.9 |
| SIM [48] | 36.6 | 28.0 | 44.5 | 33.3 |
| MRNet [62] | 38.8 | 27.9 | 45.4 | 38.7 |
| Oracle | 52.2 | 45.4 | 57.6 | 56.8 |

ting, we observe that most of them do not improve upon the source-domain baseline in our normal-to-adverse setting. The best-performing UDA method is FDA, which is based on a pixel-level adaptation strategy with an explicit Fourier prior. Even FDA is outperformed by the model that is supervised with only 100 target-domain labels, indicating that there is a lot of room for improvement for UDA methods on this new challenging normal-to-adverse benchmark.

The image-level correspondences of ACDC between adverse and normal conditions act as weak supervision. We experiment with MGCDA, a weakly supervised method that exploits such correspondences. MGCDA outperforms FDA but is still inferior to its fully supervised counterpart.

In addition, we train state-of-the-art UDA methods to adapt from Cityscapes to individual conditions of ACDC in Table 3. The increased uniformity of the target domains in this setting results in larger performance gains overall compared to Table 2. However, night and snow prove particularly challenging for most methods and only FDA brings a performance gain on snow.

### 4.2. Evaluation of Pre-trained Models on ACDC

In Table 4, we use ACDC to evaluate semantic segmentation models which have been pre-trained on external datasets. For models pre-trained on Cityscapes, the performance drop is larger on the nighttime set, implying that the domain shift from the normal-condition domain is larger for this set. Methods that specialize on fog or nighttime generally perform better on that condition compared to models pre-trained on Cityscapes. Moreover, most of these specialized methods also improve the performance on conditions other than the one encountered at training time.

Table 4. **Comparison of externally pre-trained models on ACDC for individual conditions and jointly for all conditions.** The three groups of rows present models pre-trained on normal, foggy, and nighttime conditions respectively. CS: Cityscapes [8], FC: Foggy Cityscapes [38], FC-DBF: Foggy Cityscapes-DBF [37], FZ: Foggy Zurich [37], ND: Nighttime Driving [10], DZ: Dark Zurich [40].

| Method | Trained on | Fog | Night | Rain | Snow | All |
|---|---|---|---|---|---|---|
| RefineNet [24] | CS | 46.4 | 29.0 | 52.6 | 43.3 | 43.7 |
| DeepLabv2 [5] | CS | 33.5 | 30.1 | 44.5 | 40.2 | 38.0 |
| DeepLabv3+ [6] | CS | 45.7 | 25.0 | 50.0 | 42.0 | 41.6 |
| DANet [12] | CS | 34.7 | 19.1 | 41.5 | 33.3 | 33.1 |
| HRNet [47] | CS | 38.4 | 20.6 | 44.8 | 35.1 | 35.3 |
| SFSU [38] | FC | 45.6 | 29.5 | 51.6 | 41.4 | 42.9 |
| CMAda [37] | FC-DBF+FZ | 51.2 | 32.0 | 53.4 | 47.6 | 47.1 |
| DMAda [10] | ND | 50.7 | 32.7 | 54.9 | 48.9 | 47.9 |
| GCMA [39] | CS+DZ | 52.4 | 42.9 | 58.0 | 53.8 | 53.4 |
| MGCDA [40] | CS+DZ | 45.9 | 40.8 | 54.2 | 50.5 | 48.9 |
| DANNet [50] | CS+DZ | – | 47.6 | – | – | – |

### 4.3. Supervised Learning on Adverse Conditions

We use ACDC to train four state-of-the-art supervised semantic segmentation methods and report their performance in Table 5. Qualitative results are shown in Fig. 3 for two supervised methods and one UDA method. We draw the following conclusions: (1) full supervision in adverse conditions is more valuable than designing a better architecture trained solely on normal conditions, as even an earlier method [5] performs better with full supervision than the top-performing externally pre-trained model (cf. Table 4). (2) ACDC is a challenging benchmark for supervised methods due to its hard visual domains; even the very recent HRNet scores only 75.0% mIoU on the test set, which is 5.4% lower than its respective performance of 80.4% on Cityscapes [47]. (3) The rankings of the supervised and the pre-trained models do not correlate well, as can be seen from the results in Tables 5 and 4.

The last point suggests that state-of-the-art networks such as HRNet have enough capacity to overfit to datasets such as Cityscapes, which would explain the low performance of the Cityscapes pre-trained HRNet model on ACDC. We test this hypothesis by training HRNet *jointly on Cityscapes and ACDC*; our expectation is that the jointly trained model will at least match the performance of the individually trained models on each dataset. This is confirmed, as the jointly trained model gets 81.2% mIoU on Cityscapes and 74.8% on ACDC, beating and being on a par with the respective individually trained models. Thus, even if ACDC is not of very large scale, it helps to efficiently regularize segmentation models for normal conditions as well.

Table 6 compares models trained on a single adverse condition, termed condition experts, against models trained on the entire training set, termed uber models. Each condition expert is evaluated on the condition it has been trained on. The uber models generally beat the respective condition experts across different conditions and segmentation networks. This hints that the capacity of these networks

Table 5. **Comparison of state-of-the-art supervised semantic segmentation methods on ACDC.** Training and evaluation are performed using the complete training and test sets, respectively.

| Method | road | sidew. | build. | wall | fence | pole | light | sign | veget. | terrain | sky | person | rider | car | truck | bus | train | motorc. | bicycle | mIoU |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RefineNet [24] | 92.5 | 71.2 | 86.2 | 39.0 | 44.0 | 53.2 | 68.8 | 66.0 | 85.1 | 59.3 | 94.9 | 65.2 | 38.5 | 85.8 | 53.8 | 59.7 | 76.2 | 47.5 | 54.5 | 65.3 |
| DeepLabv2 [5] | 88.0 | 62.3 | 80.8 | 37.0 | 35.1 | 33.9 | 49.8 | 49.5 | 80.1 | 50.7 | 92.5 | 51.1 | 26.5 | 79.9 | 49.0 | 41.1 | 72.2 | 26.5 | 44.2 | 55.3 |
| DeepLabv3+ [6] | 93.4 | 74.8 | 89.2 | 53.0 | 49.0 | 58.7 | 71.1 | 67.4 | 87.8 | 62.7 | 95.9 | 69.7 | 36.0 | 88.1 | 67.7 | 71.8 | 85.1 | 48.0 | 59.8 | 70.0 |
| HRNet [47] | 95.3 | 79.9 | 90.7 | 53.7 | 57.4 | 65.9 | 78.4 | 75.9 | 88.8 | 68.6 | 96.1 | 75.5 | 54.0 | 91.2 | 68.2 | 76.2 | 85.4 | 58.4 | 65.1 | 75.0 |

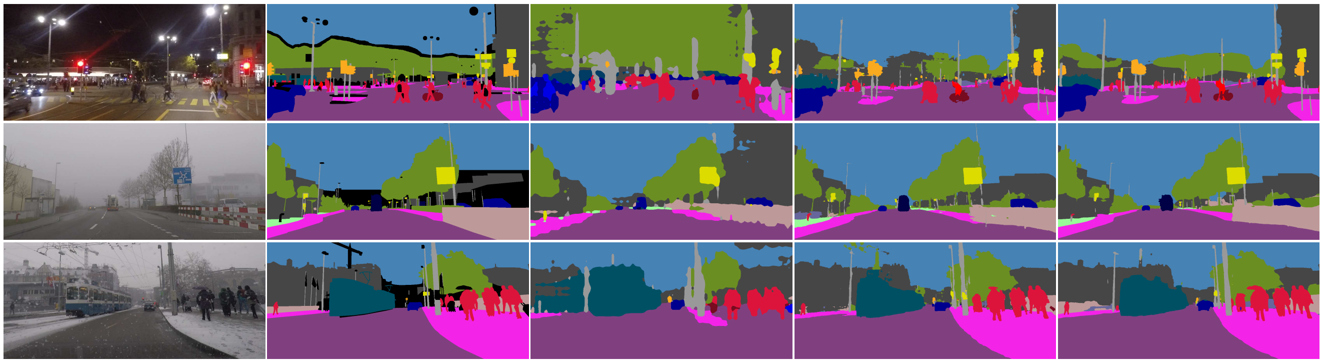

Figure 3. **Qualitative results of selected semantic segmentation methods on ACDC.** From left to right: image, ground-truth annotation, FDA [53], DeepLabv3+ [6], and HRNet [47]. The color coding of the semantic classes matches Fig. 1.

Table 6. **Comparison of condition experts vs. uber models on the different conditions of ACDC.** The first group of rows presents condition-specific expert models trained on a single condition, while the second group presents uber models trained on all conditions. Note that the performance on all conditions is *not* an average of the respective performances on individual conditions.

| Method | Fog | Night | Rain | Snow | All |
|---|---|---|---|---|---|
| RefineNet [24] | 63.6 | 52.2 | 66.4 | 62.5 | 62.8 |
| DeepLabv2 [5] | 52.2 | 45.4 | 57.6 | 56.8 | 54.9 |
| DeepLabv3+ [6] | 68.7 | 59.2 | 73.5 | 70.5 | 69.6 |
| HRNet [47] | 70.8 | 63.2 | 72.7 | 70.2 | 70.9 |
| RefineNet [24] | 65.7 | 55.5 | 68.7 | 65.9 | 65.3 |
| DeepLabv2 [5] | 54.5 | 45.3 | 59.3 | 57.1 | 55.3 |
| DeepLabv3+ [6] | 69.1 | 60.9 | 74.1 | 69.6 | 70.0 |
| HRNet [47] | 74.7 | 65.3 | 77.7 | 76.3 | 75.0 |

is large enough to discover discriminative representations for all conditions simultaneously. We also evaluate ensembles of condition experts against uber models on the complete test set ("All"), where the ensemble uses the expert corresponding to the condition of the input image for prediction. Again, the uber models outperform the ensembles of experts for all examined methods. Moreover, all methods perform worst at nighttime, indicating that the nighttime set of ACDC represents a harder domain than the other sets.

We focus on the widely used DeepLabv3+ network [6] for a detailed study of class-level performance across different conditions and compare the performance of the four condition experts in Table 7. We make the following observations: (1) the lowest performance for *road* and *sidewalk* occurs in snow, which can be attributed to confusion between the two classes due to similar appearance in the presence of snow cover. (2) Classes that usually appear dark or are not well-lit at nighttime, e.g., *building*, *vegetation*, *traffic sign*, and *sky*, are harder to segment at nighttime. (3) Performance on classes with instances of small size, such as *person*, *rider*, and *bicycle*, is lowest on fog, probably due to the combined effect of contrast reduction and low resolution for instances of these classes that are far from the camera.

We also evaluate in Table 8 the four DeepLabv3+ condition experts on conditions that are not encountered at training. Excluding nighttime, the results are close to symmetric with respect to training versus evaluation condition; e.g., training on fog and testing on snow results in a similar performance to training on snow and testing on fog. In contrast, performance of the night expert on other conditions is much higher than performance of other experts at night, implying that representations learned from the nighttime domain can generalize better to the other conditions than vice versa.

## 5. Uncertainty-Aware Semantic Segmentation

Existing works that model uncertainty in semantic segmentation [1, 21] are evaluated only with IoU, which does not assess the predicted confidence. In contrast, for uncertainty-aware semantic segmentation, algorithms are required to output both a hard semantic prediction $\hat{H}$ and a confidence map $C$ with values in the range $[0,\ 1]$. The average UIoU (AUIoU) metric is computed by thresholding $C$ at multiple thresholds across the range $[0,\ 1]$, calculating the UIoU [40] for each threshold and averaging the results. A pixel $p$ with confidence value below the examined threshold is treated as invalid and contributes positively if $J(p) = 1$

Table 7. **Comparison of class-level performance of DeepLabv3+ condition experts on the various conditions of ACDC.** A different model is trained on each individual condition and then evaluated on this condition.

| Condition | road | sidew. | build. | wall | fence | pole | light | sign | veget. | terrain | sky | person | rider | car | truck | bus | train | motorc. | bicycle | mIoU |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fog | 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 |
| Night | 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 |
| Rain | 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 |
| Snow | 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 |

Table 8. **Cross-evaluation of DeepLabv3+ condition experts on the various conditions of ACDC.** Each model is trained on an individual condition and evaluated on each condition separately. Performance of the Cityscapes pre-trained model is also reported.

| Train/Eval | Fog | Night | Rain | Snow |
|---|---|---|---|---|
| Normal | 45.7 | 25.0 | 50.0 | 42.0 |
| Fog | 68.7 | 40.7 | 63.5 | 59.1 |
| Night | 58.5 | 59.2 | 55.6 | 49.6 |
| Rain | 65.2 | 46.0 | 73.5 | 63.5 |
| Snow | 59.2 | 38.0 | 69.3 | 70.5 |

Table 9. **Uncertainty-aware semantic segmentation baseline results using AUIoU.** Supervised methods for standard semantic segmentation are trained and evaluated either separately on each condition or jointly on all conditions for semantic label prediction. Confidence prediction baselines: globally constant and equal to 100% (Constant 100%), max-softmax network outputs (Max-Softmax), ground-truth invalid masks (GT).

| Method | Confidence | Fog | Night | Rain | Snow | All |
|---|---|---|---|---|---|---|
| RefineNet [24] | Constant 100% | 63.6 | 52.2 | 66.4 | 62.5 | 65.3 |
| RefineNet [24] | Max-Softmax | 60.6 | 51.4 | 62.5 | 59.9 | 62.5 |
| RefineNet [24] | GT | 67.9 | 61.1 | 67.9 | 64.0 | 68.8 |
| DeepLabv2 [5] | Constant 100% | 52.2 | 45.4 | 57.6 | 56.8 | 55.3 |
| DeepLabv2 [5] | Max-Softmax | 51.9 | 45.9 | 56.0 | 56.8 | 54.7 |
| DeepLabv2 [5] | GT | 56.7 | 54.7 | 59.1 | 58.4 | 58.9 |
| DeepLabv3+ [6] | Constant 100% | 68.7 | 59.2 | 73.5 | 70.5 | 70.0 |
| DeepLabv3+ [6] | Max-Softmax | 66.4 | 59.1 | 70.6 | 67.9 | 67.8 |
| DeepLabv3+ [6] | GT | 73.1 | 67.1 | 75.0 | 72.0 | 73.3 |

(true invalid) and negatively if $J(p) = 0$ (false invalid).

### 5.1. Baselines and Oracles

We present the results of straightforward baselines for uncertainty-aware segmentation that are based on methods for standard semantic segmentation in Table 9. We first evaluate three state-of-the-art methods using confidence maps that are constant and equal to 1, i.e., not modeling confidence. In this case, AUIoU reduces to IoU. Any sensible method that models confidence should improve upon this baseline. Using the max-softmax scores output by these methods as confidence maps generally yields inferior results to globally constant confidence, as softmax is not a good proxy for confidence. An upper bound for the performance of the examined methods is obtained by using a confidence oracle. More specifically, we use the binary complement of the ground-truth invalid mask $J$ as the confidence prediction. This raises AUIoU performance significantly across all conditions compared to the globally constant confidence baseline. The performance gap between the oracle and the baseline is largest for night, indicating that explicitly modeling uncertainty has the potential to improve performance especially in the nighttime domain.

We have also trained [1] on ACDC, using the GT invalid masks for training its outlier detection part. The learned confidence by [1] leads to lower test set AUIoU (52.0%) than constant confidence (53.0%), indicating that a better modeling of uncertainty is needed in future approaches.

## 6. Conclusion and Outlook

In this paper, we have presented ACDC, a large-scale dataset and benchmark suite for semantic driving scene understanding in adverse conditions. Our dataset covers adverse visual domains that are common in driving scenarios and features high-quality pixel-level annotations which also include visually degraded image regions. Our annotations support both the standard and the new uncertainty-aware semantic segmentation task.

We have evaluated several state-of-the-art approaches on our benchmark, both in the supervised and the unsupervised setting. The conclusions from this evaluation show the importance of ACDC in steering future progress in the field: (i) ACDC provides a challenging target domain for unsupervised domain adaptation approaches in the normal-to-adverse adaptation setting, as most state-of-the-art approaches yield at best marginal performance gains, (ii) ACDC is a hard benchmark for supervised semantic segmentation methods, as the best baseline obtains an IoU of only 75.0%, whereas the same baseline scores 80.4% on Cityscapes, (iii) ACDC can be used jointly with existing normal-condition datasets for training in order to regularize models better and improve their performance both under normal and adverse conditions.

**Acknowledgements.** This work is funded by Toyota Motor Europe via the research project TRACE-Zürich. We thank René Zurbrügg for contributing to the construction of the website, and Anton Obukhov and Yuhang Lu for their advice on running HRNet and DANet.

## References

[1] Petra Bevandić, Ivan Krešo, Marin Oršić, and Siniša Šegvić. Simultaneous semantic segmentation and outlier detection in presence of domain shift. In *German Conference on Pattern Recognition*, 2019. 7, 8

[2] Mario Bijelic, Tobias Gruber, Fahim Mannan, Florian Kraus, Werner Ritter, Klaus Dietmayer, and Felix Heide. Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 3

[3] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuScenes: A multimodal dataset for autonomous driving. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 1, 3

[4] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan Yuille. Semantic image segmentation with deep convolutional nets and fully connected CRFs. In *International Conference on Learning Representations*, May 2015. 3

[5] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848, 2018. 3, 5, 6, 7, 8

[6] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *The European Conference on Computer Vision (ECCV)*, September 2018. 3, 6, 7, 8

[7] Yuhua Chen, Wen Li, and Luc Van Gool. ROAD: Reality oriented adaptation for semantic segmentation of urban scenes. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 2, 3

[8] 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 *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 1, 2, 3, 4, 6

[9] Dengxin Dai, Christos Sakaridis, Simon Hecker, and Luc Van Gool. Curriculum model adaptation with synthetic and real data for semantic foggy scene understanding. *International Journal of Computer Vision*, 128(5):1182–1204, 2020. 3, 5

[10] Dengxin Dai and Luc Van Gool. Dark model adaptation: Semantic image segmentation from daytime to nighttime. In *IEEE International Conference on Intelligent Transportation Systems*, 2018. 2, 3, 5, 6

[11] Shuai Di, Qi Feng, Chun-Guang Li, Mei Zhang, Honggang Zhang, Semir Elezovikj, Chiu C. Tan, and Haibin Ling. Rainy night scene understanding with near scene semantic adaptation. *IEEE Transactions on Intelligent Transportation Systems*, 22(3):1594–1602, 2021. 2

[12] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, and Hanqing Lu. Dual attention network for scene segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 3, 6

[13] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? The KITTI vision benchmark suite. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012. 1, 2

[14] Shirsendu Sukanta Halder, Jean-Francois Lalonde, and Raoul de Charette. Physics-based rendering for improving robustness to rain. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019. 3

[15] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei Efros, and Trevor Darrell. CyCADA: Cycle-consistent adversarial domain adaptation. In *International Conference on Machine Learning*, 2018. 2, 3

[16] Judy Hoffman, Dequan Wang, Fisher Yu, and Trevor Darrell. FCNs in the wild: Pixel-level adversarial and constraint-based adaptation. *arXiv e-prints*, abs/1612.02649, December 2016. 3

[17] Xinyu Huang, Peng Wang, Xinjing Cheng, Dingfu Zhou, Qichuan Geng, and Ruigang Yang. The ApolloScape open dataset for autonomous driving and its application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(10):2702–2719, 2020. 2

[18] Zilong Huang, Xinggang Wang, Lichao Huang, Chang Huang, Yunchao Wei, and Wenyu Liu. CCNet: Criss-cross attention for semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019. 3

[19] Matthew Johnson-Roberson, Charles Barto, Rounak Mehta, Sharath Nittur Sridhar, Karl Rosaen, and Ram Vasudevan. Driving in the matrix: Can virtual worlds replace human-generated annotations for real world tasks? In *IEEE International Conference on Robotics and Automation*, 2017. 3

[20] Tsung-Wei Ke, Jyh-Jing Hwang, Ziwei Liu, and Stella X. Yu. Adaptive affinity fields for semantic segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018. 3

[21] Alex Kendall and Yarin Gal. What uncertainties do we need in Bayesian deep learning for computer vision? In *Advances in Neural Information Processing Systems*, 2017. 7

[22] Myeongjin Kim and Hyeran Byun. Learning texture invariant representation for domain adaptation of semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 2, 3

[23] Yunsheng Li, Lu Yuan, and Nuno Vasconcelos. Bidirectional learning for domain adaptation of semantic segmentation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 2, 3, 5, 6

[24] Guosheng Lin, Anton Milan, Chunhua Shen, and Ian Reid. RefineNet: Multi-path refinement networks with identity mappings for high-resolution semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 3, 5, 6, 7, 8

[25] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015. 3

[26] Yawei Luo, Liang Zheng, Tao Guan, Junqing Yu, and Yi Yang. Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 2, 3, 5, 6

[27] Will Maddern, Geoffrey Pascoe, Chris Linegar, and Paul Newman. 1 year, 1000 km: The Oxford RobotCar dataset. *The International Journal of Robotics Research*, 36(1):3–15, 2017. 1, 3

[28] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bulò, and Peter Kontschieder. The Mapillary Vistas dataset for semantic understanding of street scenes. In *The IEEE International Conference on Computer Vision (ICCV)*, 2017. 1, 2, 3, 4

[29] Andreas Pfeuffer, Markus Schön, Ditzel Carsten, and Klaus Dietmayer. The ADUULM-Dataset - a semantic segmentation dataset for sensor fusion. In *Proceedings of the British Machine Vision Conference (BMVC)*, 2020. 3

[30] Matthew Pitropov, Danson Evan Garcia, Jason Rebello, Michael Smart, Carlos Wang, Krzysztof Czarnecki, and Steven Waslander. Canadian adverse driving conditions dataset. *The International Journal of Robotics Research*, 2020. 3

[31] Tobias Pohlen, Alexander Hermans, Markus Mathias, and Bastian Leibe. Full-resolution residual networks for semantic segmentation in street scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 3

[32] Horia Porav, Tom Bruls, and Paul Newman. Don't worry about the weather: Unsupervised condition-dependent domain adaptation. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 33–40, 2019. 2

[33] Stephan R. Richter, Zeeshan Hayder, and Vladlen Koltun. Playing for benchmarks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, October 2017. 3

[34] Stephan R. Richter, Vibhav Vineet, Stefan Roth, and Vladlen Koltun. Playing for data: Ground truth from computer games. In *European Conference on Computer Vision*. Springer, 2016. 2, 3

[35] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, 2015. 3

[36] German Ros, Laura Sellart, Joanna Materzynska, David Vazquez, and Antonio M. Lopez. The SYNTHIA dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 2, 3

[37] Christos Sakaridis, Dengxin Dai, Simon Hecker, and Luc Van Gool. Model adaptation with synthetic and real data for semantic dense foggy scene understanding. In *The European Conference on Computer Vision (ECCV)*, 2018. 2, 3, 6

[38] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, 126(9):973–992, 2018. 2, 3, 5, 6

[39] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Guided curriculum model adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation. In *The IEEE International Conference on Computer Vision (ICCV)*, 2019. 2, 6

[40] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Map-guided curriculum domain adaptation and uncertainty-aware evaluation for semantic nighttime image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 1, 2, 3, 5, 6, 7

[41] Swami Sankaranarayanan, Yogesh Balaji, Arpit Jain, Ser Nam Lim, and Rama Chellappa. Learning from synthetic data: Addressing domain shift for semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 2, 3

[42] 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 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 1, 3

[43] Yi-Hsuan Tsai, Wei-Chih Hung, Samuel Schulter, Kihyuk Sohn, Ming-Hsuan Yang, and Mammohan Chandraker. Learning to adapt structured output space for semantic segmentation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 2, 3, 5, 6

[44] Yi-Hsuan Tsai, Kihyuk Sohn, Samuel Schulter, and Manmohan Chandraker. Domain adaptation for structured output via discriminative patch representations. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019. 2, 3

[45] Frederick Tung, Jianhui Chen, Lili Meng, and James J. Little. The Raincouver scene parsing benchmark for self-driving in adverse weather and at night. *IEEE Robotics and Automation Letters*, 2(4):2188–2193, 2017. 3, 5

[46] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Perez. ADVENT: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 2, 3, 5, 6

[47] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, Wenyu Liu, and Bin Xiao. Deep high-resolution representation learning for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 3, 6, 7

[48] Zhonghao Wang, Mo Yu, Yunchao Wei, Rogerio Feris, Jinjun Xiong, Wen-mei Hwu, Thomas S. Huang, and Honghui Shi. Differential treatment for stuff and things: A simple unsupervised domain adaptation method for semantic segmen-

tation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 2, 3, 5, 6

[49] Patrick Wenzel, Rui Wang, Nan Yang, Qing Cheng, Qadeer Khan, Lukas von Stumberg, Niclas Zeller, and Daniel Cremers. 4Seasons: A cross-season dataset for multi-weather SLAM in autonomous driving. In *Proceedings of the German Conference on Pattern Recognition (GCPR)*, 2020. 1

[50] Xinyi Wu, Zhenyao Wu, Hao Guo, Lili Ju, and Song Wang. DANNet: A one-stage domain adaption network for unsupervised nighttime semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021. 6

[51] Zuxuan Wu, Xintong Han, Yen-Liang Lin, Mustafa Gokhan Uzunbas, Tom Goldstein, Ser Nam Lim, and Larry S. Davis. DCAN: Dual channel-wise alignment networks for unsupervised scene adaptation. In *The European Conference on Computer Vision (ECCV)*, 2018. 2

[52] Dan Xu, Wanli Ouyang, Xiaogang Wang, and Nicu Sebe. PAD-Net: Multi-tasks guided prediction-and-distillation network for simultaneous depth estimation and scene parsing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 3

[53] Yanchao Yang and Stefano Soatto. FDA: Fourier domain adaptation for semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 2, 3, 5, 6, 7

[54] Changqian Yu, Jingbo Wang, Chao Peng, Changxin Gao, Gang Yu, and Nong Sang. BiSeNet: Bilateral segmentation network for real-time semantic segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018. 3

[55] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. BDD100K: A diverse driving dataset for heterogeneous multitask learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 1, 2, 3, 4, 5

[56] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. In *International Conference on Learning Representations*, 2016. 3

[57] Yuhui Yuan, Xilin Chen, and Jingdong Wang. Object-contextual representations for semantic segmentation. In *The European Conference on Computer Vision (ECCV)*, pages 173–190, 2020. 3

[58] Oliver Zendel, Katrin Honauer, Markus Murschitz, Daniel Steininger, and Gustavo Fernandez Dominguez. WildDash - creating hazard-aware benchmarks. In *The European Conference on Computer Vision (ECCV)*, 2018. 3, 5

[59] Yang Zhang, Philip David, and Boqing Gong. Curriculum domain adaptation for semantic segmentation of urban scenes. In *The IEEE International Conference on Computer Vision (ICCV)*, 2017. 2

[60] Yiheng Zhang, Zhaofan Qiu, Ting Yao, Dong Liu, and Tao Mei. Fully convolutional adaptation networks for semantic segmentation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 2, 3

[61] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 3

[62] Zhedong Zheng and Yi Yang. Rectifying pseudo label learning via uncertainty estimation for domain adaptive semantic segmentation. *International Journal of Computer Vision*, 2021. 2, 5, 6

[63] Brady Zhou, Philipp Krähenbühl, and Vladlen Koltun. Does computer vision matter for action? *Science Robotics*, 4(30), 2019. 3

[64] Zhen Zhu, Mengde Xu, Song Bai, Tengteng Huang, and Xiang Bai. Asymmetric non-local neural networks for semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019. 3

[65] Yang Zou, Zhiding Yu, Xiaofeng Liu, B.V.K. Vijaya Kumar, and Jinsong Wang. Confidence regularized self-training. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019. 2, 3, 5

[66] Yang Zou, Zhiding Yu, B.V.K. Vijaya Kumar, and Jinsong Wang. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *The European Conference on Computer Vision (ECCV)*, 2018. 2, 3