Camera-Based Road Snow Coverage Estimation

Kai Cordes and Hellward Broszio
VISCODA GmbH
https://roadsc.viscoda.com

Abstract

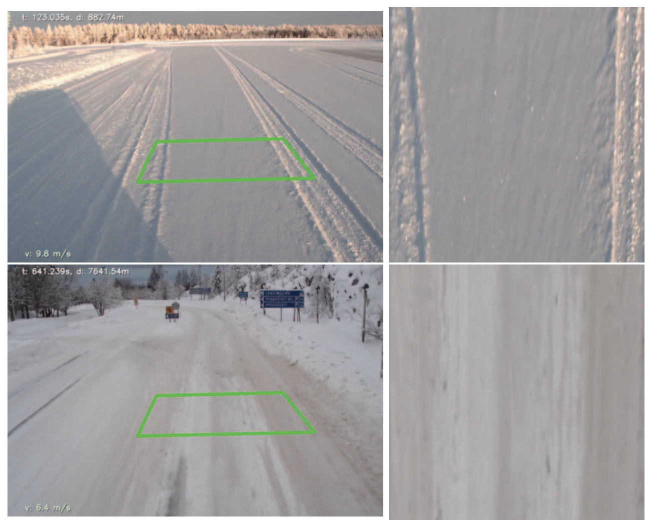
The current road condition is a crucial factor regarding road safety of the ego-vehicle and other road users. Road condition estimation provides essential input data for friction estimation which is used for autonomous and automated driving systems. Camera-based approaches are still far from being practical and other sensors dominate the field of friction estimation. This is due to the limited performance of current approaches and the lack of datasets for the incorporation of learning-based methods.

We propose a novel dataset for a special scenario of road condition, the coverage with snow. It is the first large-scale dataset for camera-based road classification of snow-covered roads with different types of snow coverage. The dataset consists of road patches in bird's eye view perspective and ground truth annotation for the current snow coverage type. It is combinable with RoadSaW [4], a dataset for road surface and wetness estimation, leading to a holistic road condition dataset with 15 categories. The baseline evaluation employs state-of-the-art, real-time capable approaches for classification and uncertainty estimation with RBF (Radial Basis Function) networks. Our experiments demonstrate that the proposed data opens new challenges in the field of camera-based road condition estimation.

1. Introduction

Advanced technologies and various types of sensors in automated vehicles continuously improve the road safety. These sensors include camera, light detection and ranging (LiDAR), radar, inertial measurement unit (IMU), global navigation satellite system (GNSS), and sonar [27, 14, 15, 10]. In terms of usability and cost, cameras are the most popular sensors in autonomous driving vehicles [14]. Key performance indicators of cameras are the large usability range and the high resolution at relatively low cost.

Road conditions effected by weather events such as rain and snow lead to decreased vehicle performance, e.g., decreased friction of the vehicle tires, resulting in fatal crashes and property damage of high amount [11]. For advanced



(a) Camera View (b) Bird's Eye View Patch

Figure 1: Camera view and corresponding bird's eye view ($3.6\,\mathrm{m} \times 3.6\,\mathrm{m}$) of a road region computed from camera calibration. We show *Fresh Fallen Snow* on the test track (top) and *Fully Packed Snow* on a public road (bottom).

driver assistance systems as well as for self-driving vehicles, many tasks such as navigation, self-steering, and automatic breaking can be further improved with the knowledge of the current road condition [22]. Additionally, it is a crucial factor for road safety. Thus, the knowledge of road conditions, e.g., dry, wet, and snowy surfaces, should be considered. In [1], significant effects on the current friction of snow-covered surfaces are derived from the surrounding temperature, snow density, and thickness of the snow. While the temperature is available from in-vehicle sensors, snow density and thickness is a road property and, thus, can only be measured using a dedicated sensor. Three snow states are considered [1], discharging (the mechanical removal of snow), melting (changing to a liquid), and solidifying (changing to a solid). From a vehicles' perspective, a distinction into different snow classes is required for reliable friction estimation.

Table 1: Comparison of related road condition datasets. The proposed dataset *RoadSC*^{3} provides three different snow classes and *Bird's Eye View* (BEV) perspective for the *Regions of Interest* (ROI). *RoadSC*^{15} combines *RoadSaW* with *RoadSC*^{3} leading to 15 categories including three snow densities and three road surface types combined with four wetness levels.

Dataset	Image Source	Calibration	Label Resolution	# Categories: Snow/Other
Khan et al. [11, 24]	*SHRP2 NDS*	✗	full image	2/1
Roychowdhury [21]	*YouTube*	✗	full image	2/4
Busch et al. [18, 3]	compiled	✗	full image / ROI	1/4
RoadSaW [4]	test track	✓	ROI, BEV	0/12
RoadSC^3 **(new)**	test track / public roads	✓	ROI, BEV	**3/0**
RoadSC^{15} **(new)**	test track / public roads	✓	ROI, BEV	**3/12**

2. Related Work

Due to the popularity of machine learning approaches in computer vision tasks, many datasets are available in the automated driving scenario [5, 8, 16]. Some of them consider adverse weather conditions [19, 7]. However, these datasets are focused on object categories and provide 3D object labels of traffic participants only. A few approaches handle snowy conditions as one category out of others [18, 3]. In [21], snowy conditions are divided into snow and slush together with two wetness levels. In [11] two snow conditions, *light snow*, and *heavy snow*, are proposed. However, these datasets only include uncalibrated views. In Tab. 1, existing datasets targeting road surface types including snow categories are listed. The authors of [21] indicate the benefit of patch segmentation for a more detailed view on partially snow-covered road regions observed from a bird's eye view perspective. The *RoadSaW*, dataset [4] includes calibrated bird's eye view perspective. and 12 surface and wetness categories, but no snow.

Our contribution is the construction of a novel dataset *RoadSC* for snow conditions incorporating different categories for the coverage of the road with snow. The data is recorded on eight different days on a test track and on public roads under various snow conditions, including freshly fallen snow, packed snow, and different levels of snow coverage of the road. The scenes are recorded with a calibrated camera setup mounted on a truck. The camera calibration enables the computation of bird's eye view (BEV) road image patches as visualized in Fig. 1. It is the first dataset providing three different snow classes together with high resolution BEV images. Our data is compatible with the *RoadSaW* dataset [4] which includes 12 classes of road surfaces and wetness levels. The combination provides 15 categories: snow densities, wetness levels, and road surface types leading to a holistic road condition estimation dataset. We evaluate the use of deep convolutional neural networks to estimate the road condition for the proposed datasets. Under examination are *RoadSC*^{3} with three snow classes and *RoadSC*^{15}, the combined datasets with 15 categories. As a baseline, a real-time capable approach using the MobileNetV2 [23] architecture and uncertainty estimation with RBF networks [25] is employed. Furthermore, the uncertainty estimation is evaluated using out-of-distribution (OoD) datasets and a detailed investigation of the required hyperparameters. To summarize, our contributions are:

- Unique dataset for road snow coverage estimation, including different snow coverage categories
- Combination with the *RoadSaW* dataset for holistic road condition estimation with 15 categories: snow coverage types, surfaces types, and wetness levels
- Baseline for image-based road condition and uncertainty estimation including hyperparameter study
- The dataset is available at:
`https://roadsc.viscoda.com`

3. RoadSC: Road Snow Coverage Dataset

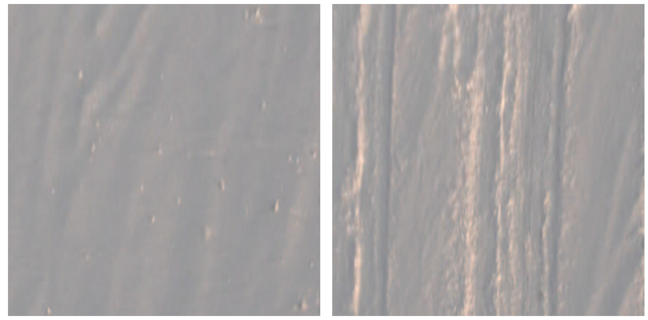
The *Road Snow Coverage* (*RoadSC*) dataset was recorded on a large-scale test track and on public roads in Finland. The image capture was done using two cameras (second camera served as backup) mounted in the cabin of a truck. Eight days of data acquisition was performed resulting in a large-scale dataset designed for road snow classification and uncertainty estimation. Following the approach in [4] camera calibration is used for the extraction of bird's eye view (BEV) road patches of a selectable size at a certain distance to the vehicle. In Fig. 1, examples for the recorded camera views are shown on the left while the corresponding extracted BEV road patches are on the right. As distance to the vehicle *d2v*, $15\,\mathrm{m}$ is selected. The BEV image resolution is 900×900 px to keep the image data compatible with *RoadSaW* [4]. Thus, both datasets can be combined for building a larger dataset for road surface condition estimation including the proposed snow classes together with the surface and wetness classes from *RoadSaW*.

Since snow thickness and density are the main influences on the friction of tires [1], the categories are selected based on the snow densities and the coverage of the snow on the road.

We define three snow coverage categories as

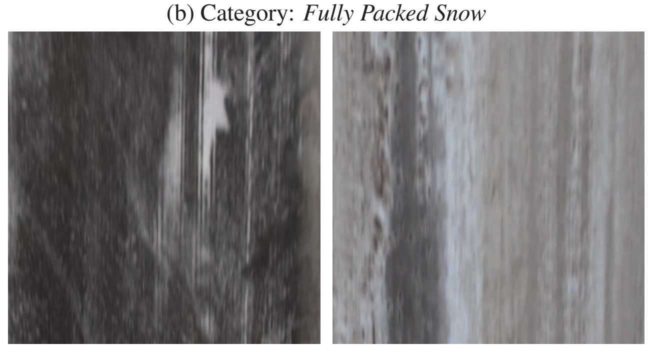
- *Fresh Fallen Snow*: fresh snow surface, no road surface visible, few drive-overs allowed (distinct skid marks), examples in Fig. 2a
- *Fully Packed Snow*: surface fully covered with packed snow, resulting from many drive-overs, no road surface visible, examples in Fig. 2b
- *Partially Covered Snow*: parts of the observed region are snow-covered, road surface is visible, examples in Fig. 2c and Fig. 3

The different categories could be recorded due to the following advantageous conditions: (1) the weather provided fresh-fallen snow; (2) our drivers followed specific maneuvers enabling the recording of many image sequences with the demanded categories (such as driving over as much fresh-fallen snow as possible). The *Partially Covered Snow* category includes a large variety of different coverage levels. Some examples for different levels of coverage are shown in Fig. 3.



(a) Category: *Fresh Fallen Snow*

(b) Category: *Fully Packed Snow*



(c) Category: *Partially Covered Snow*

Figure 2: Example bird's eye view patches of the categories *Fresh Snow*, *Fully Packed Snow*, and *Partially Covered Snow*.

3.1. Image Acquisition

For video recording, two redundant FLIR Blackfly cameras, 2048×1536 px resolution, were installed behind the windscreen of the truck at $2.72\,\mathrm{m}$ height. The image capture was done using Nvidia Jetson Nanos. The system worked reliably although the circumstances were challenging with temperatures down to $-20\,°\mathrm{C}/-4\,°\mathrm{F}$ during recording and down to $-40\,°\mathrm{C}/-40\,°\mathrm{F}$ during night. Before the third recording day, fresh snow has been fallen. The drivers used as many fresh snow regions as possible to capture enough video data for the *Fresh Fallen Snow* category.

The road patch format is kept compatible with the dataset *RoadSaW* [4] since a combination of the datasets is intended for a holistic road condition estimation including different types of snow (*RoadSC*) together with road surfaces types and wetness levels (*RoadSaW*). Using camera calibration, the desired road region is selected for the ground plane projection of image content. Following [4], BEV road patches at $15\,\mathrm{m}$ distance are computed with a patch size of $3.6\,\mathrm{m} \times 3.6\,\mathrm{m} = 12.96\,\mathrm{m}^2$ and 900×900 px resolution. Camera view and BEV road patches are visualized in Fig. 1. Patch examples of different categories are shown in Fig. 2 and Fig. 3.

3.2. Dataset Statistics and Balancing

The *RoadSC* dataset employs 238 image sequences divided into the proposed categories *Fresh Fallen Snow*, *Fully Packed Snow*, and *Partially Covered Snow*, cf. Tab. 2. The annotation is done manually using multiple video cuts with respect to their identified category. Only images with clearly identifiable categories are selected for the dataset, there are no misaligned BEVs, i.e., occluded or outside the road. The annotation procedure results in $\approx 186{,}000$ road patches with imbalanced category distribution. The data is divided into training ($\approx 70\%$), validation ($\approx 20\%$), and test ($\approx 10\%$) sets. Images from the same sequence are assigned to the same set ensuring that similar images are not used for both, training and testing. For the balancing of classes, we randomly remove images from the set until an equal number of images per class is achieved leading to 90,759 images in total (30,253 for each of the three categories). The resulting dataset is called *RoadSC*^{3}.

For the combination with *RoadSaW* [4], the number of images in the snow classes is further reduced to match the balanced size of *RoadSaW*^{12} (2138 for each category).

(a) Low Coverage (b) Med. Coverage (c) High Coverage

Figure 3: Examples for different snow coverage levels included in the *Partially Covered Snow* category, cf. Fig. 2c

RoadSaW^{12} includes 12 categories combining surface types (*Asphalt*, *Concrete*, *Cobble*) and surface wetness levels (*Dry*, *Damp*, *Wet*, *Very Wet*). Thus, 15 categories are obtained, each with 2138 road patches. The resulting dataset is called *RoadSC*^{15} containing $15 \cdot 2138 = 32070$ images.

Limitations The dataset is captured on eight days in February of 2022 during daytime (8h00-17h00). The recording is limited to two regions: the test track and public roads near the test track. Fresh-fallen snow could only be recorded on one of the days (02-03) as shown in Tab. 2. The recording on public roads is done on 02-08 only. Here, the category *Fresh Fallen Snow* is not represented.

Table 2: Overview on recording days in Feb. 2022 showing the distribution of the number of sequences by categories.

Capture Day	*Fully Packed Snow*	*Partially Covered Snow*	*Fresh Fallen Snow*
02-01	29	1	✗
02-02	13	✗	✗
02-03	✗	✗	52
02-04	20	✗	✗
02-05	26	18	✗
02-07	9	6	✗
02-08	16	28	✗
02-09	17	3	✗
Σ	130	56	52

4. Evaluation

The targeted use case of automated driving demands small inference time of the selected algorithms. Thus, state-of-the-art, real-time capable methods are chosen for the classification and uncertainty estimation baseline. We use the MobileNetV2 [2, 23] as network architecture since it is widely supported by real-time hardware, e.g., Qualcomm QSC chipsets. The uncertainty estimation is based on RBF (*Radial Basis Function*) networks [25]. The approach estimates uncertainty in a single forward pass. The uncertainty measure is important in cases such as unknown road surfaces or occlusion (i.e., with a car in front of the ego-vehicle), and other visual artefacts such as wipers or raindrops on the windshield.

Our evaluation includes the classification task together with uncertainty estimation on *RoadSC*^{3} with three snow classes and *RoadSC*^{15}, the combination with *RoadSaW* [4] as described in Sec. 3.2 including 15 road condition categories. Results can be directly compared to the evaluations in [4] since a similar setup is used. While the *RoadSC*^{3} evaluation provides a validation of the newly constructed dataset, *RoadSC*^{15} demonstrates a holistic *Road Surface Classification* (RSC) including three snow classes (*Fresh Fallen Snow*, *Fully Packed Snow*, *Partially Covered Snow*) and three road surface types (*Asphalt*, *Concrete*, *Cobble*) with four wetness levels (*Dry*, *Damp*, *Wet*, *Very Wet*).

In Sec. 4.1 the evaluation setup is described. The results are presented in Sec. 4.2. The implementation in a real vehicle in described in Sec. 4.3. A discussion of the results in Sec. 4.4 concludes the evaluation.

4.1. Evaluation Setup

The MobileNetV2 [23] network is pretrained on *ImageNet* [6]. For uncertainty estimation, DUQ (*Deterministic Uncertainty Quantification*) [25] is employed. DUQ is based on RBF (*Radial Basis Function*) networks and proved to provide reasonable results on available datasets (Fashion-MNIST, CIFAR-10, *RoadSaW*). Each class is represented by a centroid and predictions are made by computing a kernel function, which calculates the distance between the feature vector of the model and the centroids. The uncertainty is based on the distance to the closest centroid. To quantify the uncertainty performance, *Out-of-Distribution* (OoD) datasets are evaluated [12, 25]. This data should receive lower confidence scores than the images from the original dataset. The AUROC (*Area Under the Receiver Operator Characteristic*) metric is used to evaluate the OoD performance. All experiments are repeated five times to consider their mean and standard deviation as results. For exemplary visualization of the resulting feature cluster configuration, t-SNE (*t-Distributed Stochastic Neighbor Embedding*) [26] is used.

Training For MobileNetV2, image sizes of 224×224 pixel are used. Standard data augmentation is applied, i.e., random flipping horizontally, scaling, shifting horizontally and vertically, and shearing. The DUQ class centroids [25] are updated after each step using an exponential moving average of data points belonging to that class. The centroid size is set to 1280, corresponding to the size of the feature extractor. As loss function, the categorical cross-entropy, regularized with the two-sided gradient penalty [25] is employed. It is minimized using the

RAdam [13] optimizer. Early stopping is applied to avoid overfitting.

Hyperparameter The authors of DUQ [25] emphasize the importance of two hyperparameters: the length scale σ and the gradient penalty λ. The length scale σ corresponds to the standard deviations of the Gaussians in the Radial Basis Function (RBF) kernel. The λ weights the two-sided gradient penalty in the loss function. The two-sided gradient penalty is integrated to avoid *feature collapse* which is effect of intersecting representation clusters, cf. [25] for details. To determine σ and λ, additional experiments with *Out-of-Distribution* (OoD) data are conducted. For FashionMNIST [25], two different *Out-of-Distribution* (OoD) datasets, MNIST and NotMNIST, are used to derive the choice of λ from the accuracy on MNIST and the AUROC results on FashionMNIST vs. OoD datasets. Here, $\lambda = 0.05$ is determined as the optimal configuration for all datasets. For CIFAR-10, the *In-Distribution* (ID) uncertainty (AUROC) is evaluated to determine a different $\lambda = 0.5$. A similar evaluation in [20] leads to $\lambda = 0$ for CIFAR-10-C and CIFAR-100-C [9]. Setting $\lambda = 0$ clearly optimizes the performance on ID, but decreases the performance on OoD data by switching off the two-sided gradient penalty which is designed for the detection of OoD samples and for the avoidance of feature collapse. In [4], the accuracy on the validation set is used to tune σ; from the ID AUROC, $\lambda = 0.3$ is derived for *RoadSaW*.

To summarize, all procedures [4, 20, 25] lead to the same length scale σ, but to very different values for the gradient weight λ (cf. Tab. 3).

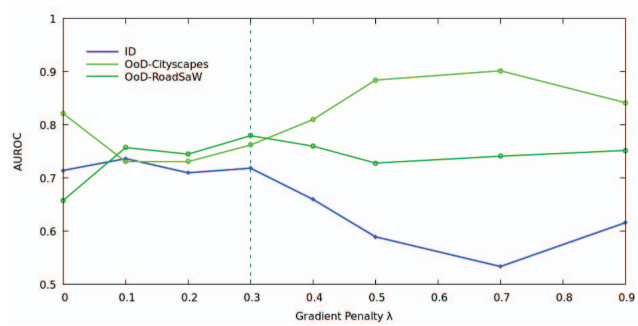
Table 3: Overview on the selection of the hyperparameters (σ, λ) in DUQ [25] for the respective datasets with C categories and N_C number of images per category.

Dataset	C	N_c	σ	λ
FashionMNIST [25]	10	7000	0.1	0.05
CIFAR-10 [25]	10	6000	0.1	0.5
CIFAR-10-C [20]	10	6000	0.1	0.0
CIFAR-100-C [20]	100	600	0.1	0.0
RoadSaW [4]	12	2138	0.1	0.3
RoadSC^3	3	30253	0.1	0.3
RoadSC^{15}	15	2138	0.1	0.3

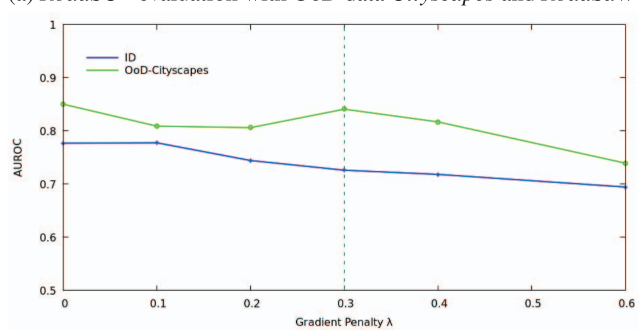
For the evaluation of *RoadSC*^{3}, we follow the more elaborated procedure in [25], using two different *Out-of-Distribution* (OoD) datasets for the AUROC evaluation. Therefore, images from Cityscapes [5] serve as first OoD dataset (as in [4]) while part of the *RoadSaW* dataset (the category *Asphalt* with wetness levels *Damp*, *Wet*, *Very Wet*) is used as the second OoD dataset. The intuition behind the choice of the second OoD dataset is that the targeted application should be able to separate snow surfaces and *Partially Covered Snow*, cf. Fig. 3a, from uncovered asphalt.

The comparative study for the gradient penalty λ on *RoadSC*^{3} is shown in Fig. 4a. In contrast to the results in [25], there is not an optimal λ for all metrics. As expected, the *In-Distribution* (ID) performance is higher for smaller λ-values and decreases for larger λ when the influence of the two-sided gradient penalty increases. For larger λ, the OoD AUROC improves. The *RoadSC*^{15} results show a similar, but less distinct trend, cf. Fig. 4b. Here, for large λ, i. e. $\lambda > 0.6$, the probability of a diverging optimization during training increases significantly.

Based on our evaluations, we select $\lambda = 0.3$ on both, *RoadSC*^{3} and *RoadSC*^{15}, for the results reported in the following Sec. 4.2.



(a) *RoadSC*^{3} evaluation with OoD data *Cityscapes* and *RoadSaW*.

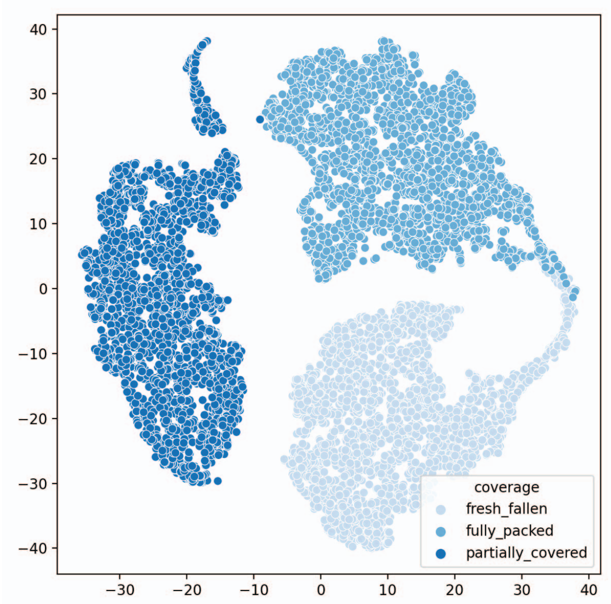


(b) *RoadSC*^{15} evaluation with OoD data from *Cityscapes*.

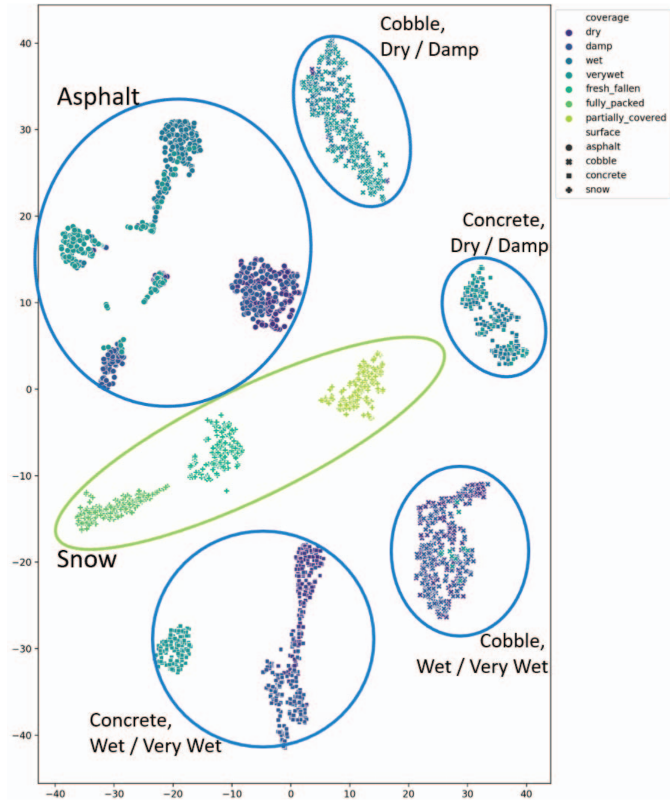
Figure 4: AUROC results on ID (*In-Distribution*) and OoD (*Out-of-Distribution*) datasets with respect to the gradient penalty λ. In our experimental results, we use $\lambda = 0.3$.

4.2. Experimental Results

For the evaluation of the classification and uncertainty estimation performance, we show F1-scores for the *In-Distribution* (ID) test set and the AUROC measure for ID and *Out-of-Distribution* (OoD) data. We compare the results of the proposed data sets *RoadSC*^{3} and *RoadSC*^{15} with the results of *RoadSaW*^{6} and *RoadSaW*^{12} taken from [4]. For the OoD analysis, images from the *Cityscapes* dataset are used (as in [4]). The F1-score comparison is



(a) The *RoadSC*^{3} embeddings with three snow coverage types *fresh-fallen*, *fully-packed*, and *partially-covered*, visualized with three shades of blue. The embeddings have three distinct clusters with a transition from *fresh-fallen* to *fully-packed*.
.



(b) The *RoadSC*^{15} embeddings with distinct surface clusters *Snow*, *Asphalt*, *Cobble*, and *Concrete*. The cluster representation indicates feature collapse between *Wet* vs. *Very Wet* and *Dry* vs. *Damp* wetness classes for *Cobble*, and *Concrete*.

Figure 5: Feature embedding visualizations of the RBF network DUQ on *RoadSC* using t-SNE

Table 4: Comparison of F1-Scores for *RoadSaW* [4] (*RoadSaW*^{6} and *RoadSaW*^{12}) and *RoadSC* (*RoadSC*^{3} and *RoadSC*^{15}. The superscript shows the number of categories.

Dataset	F1-Score
RoadSaW^6	$90.81\% \pm 0.64$
RoadSaW^{12}	$61.60\% \pm 0.75$
RoadSC^3	$97.23\% \pm 0.01$
RoadSC^{15}	$70.92\% \pm 0.03$

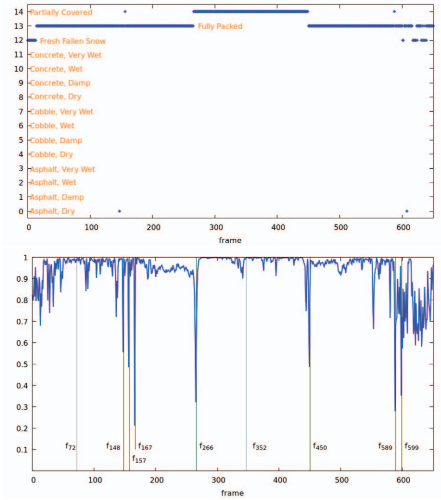
Table 5: Uncertainty estimation results for the RBF network DUQ. We measure the AUROC score on *In-Distribution* (ID) and *Out-of-Distribution* (OoD) datasets.

Dataset	ID	OoD
RoadSaW^6	$81.87\% \pm 5.06$	$98.59\% \pm 0.84$
RoadSaW^{12}	$74.86\% \pm 3.00$	$96.17\% \pm 3.29$
RoadSC^3	$71.87\% \pm 0.13$	$76.21\% \pm 0.21$
RoadSC^{15}	$76.50\% \pm 0.03$	$82.76\% \pm 0.11$

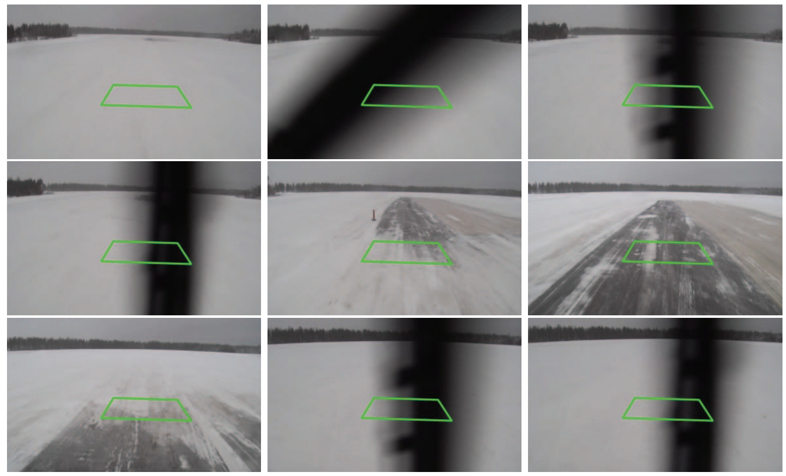
shown in Tab. 4 while the AUROC results are shown in Tab. 5.

The F1-scores of the new datasets *RoadSC*^{3} and *RoadSC*^{15} are in alignment with the results of the *RoadSaW* datasets. With decreasing number of categories, the F1-scores tend to increase. However, the *RoadSC*^{15} result shows larger accuracy than those of *RoadSaW*^{12} and provides a smaller standard deviation. While the ID AUROC results (cf. Tab. 5) are on par with *RoadSaW*, significantly lower scores with OoD AUROC are achieved for the datasets *RoadSC*^{3} and *RoadSC*^{15}. Better OoD AUROC of up to 90% could be achieved with a different selection of hyperparameters, but this comes with a performance loss on ID data, cf. Fig. 4a.

In general, small OoD AUROC results indicate feature collapse since there is potential for decreased discrimination in the feature representation space [25]. This is likely to occur in the new datasets due to the smooth transition between *Fresh Fallen Snow* and *Fully Packed Snow*, cf. Fig. 5a. For *RoadSC*^{15}, the effect of feature collapse is



(a) Classification (top) for *RoadSC*^{15} with 15 categories, including 14: *Partially Covered*, 13: *Fully Packed Snow*, and 12: *Fresh Fallen Snow* with the respective confidences (bottom).



(b) Nine selected frames $(f_{72}, f_{148}, f_{157}, f_{167}, f_{266}, f_{347}, f_{450}, f_{589}, f_{599})$ from the evaluation in (a) with the respective confidences $(0.99, 0.55, 0.48, 0.21, 0.32, 0.96, 0.48, 0.28, 0.35)$

Figure 6: Evaluation of an additional example sequence, not included in the dataset. The adversarial examples with wipers occluding the region of interest receive low confidence scores $(f_{148}, f_{157}, f_{167}, f_{266}, f_{589}, f_{599})$ because of their absence in the dataset. The frames f_{266}, f_{450} receive low confidence scores because their feature representations are in-between two classes. The estimations for f_{72} and f_{347} have high confidence.

visualized in Fig. 5b. The feature representation clusters of (*Concrete, Dry*) vs. (*Concrete, Damp*), (*Cobble, Dry*) vs. (*Cobble, Damp*), and (*Cobble, Wet*) vs. (*Cobble, Very Wet*) intersect. Thus, misclassifications between these categories are to be expected. The confusion between *Dry* vs. *Damp* and *Wet* vs. *Very Wet* may not be too harmful for the application, but nevertheless, leads to low accuracy scores of the classifier and therefore to low OoD accuracy.

4.3. In-Vehicle Implementation

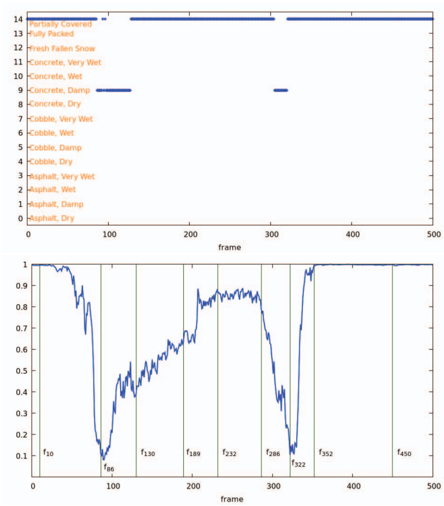
The approach is implemented on a Jetson Nano TX2 with Nvidia-Jetpack v.4.6, torchvision v.0.9, libtorch v.1.8, and Linux Ubuntu 18.04. For the image capture, FLIR Blackfly cameras (3.2 MP) are used, same as for dataset generation. The classification and uncertainty estimation results are transferred to the in-vehicle system using a USB-CAN interface. The inference time for the real-time relevant batch size of 1 with MobileNetV2 provides 28.5 fps.

The overall system running time including image capture, preprocessing (BEV, scaling), inference, and message passing via the CAN-interface is 15 Hz leading to a practical, real-time capable solution for road condition classification and uncertainty estimation.

4.4. Discussion

The evaluations show reasonable results in most cases, but also misclassifications between visually similar categories. For *RoadSC*^{3}, a transition between *Fresh Fallen Snow* and *Fully Packed Snow* is visible (Fig. 5a). For *RoadSC*^{15}, feature representation clusters intersect, e.g., *Dry* vs. *Damp* for the *Cobble* and *Concrete* surface types (Fig. 5b). The collapsing categories contain images of visual appearance with a seamless transition between them. We infer, that the reason for these failure cases is due to the coarse discretization of the category labels. A higher resolution for the labels *Wetness* and *Snow Density* is desirable. The two-sided penalty is not able to prevent the feature collapse, even with carefully selected hyperparameters (cf. Sec. 4.1). The authors of [20] indicate that deterministic uncertainty methods suffer from suboptimal calibration and that DUQ has issues regarding its convergence for complex classification problems, shown on CIFAR-100-C [9]. Additionally, recent works show that feature space fitting is possible without finetuning on OoD data [17]. We also observed convergence issues on the proposed dataset, especially for large gradient penalty weights λ. Thus, the new dataset provides new challenges and requires improved models for classification and uncertainty estimation. There is substantial potential for accuracy improvements.

Nevertheless, the RBF network determines large uncertainties for *Out-of-Distribution* data. In Fig. 6 and Fig. 7, result examples are visualized. Fig. 6 demonstrates the confidence estimation when wipers occlude the scene. As expected, all images with wiper occlusion receive low con-



(a) Classification (top) for *RoadSC*[15] with 15 categories, including 14:*Partially Covered*, 13:*Fully Packed Snow*, and 12:*Fresh Fallen Snow* with the respective confidences (bottom).

(b) Nine selected frames $(f_{10}, f_{86}, f_{130}, f_{189}, f_{232}, f_{286}, f_{322}, f_{352}, f_{450})$ from the evaluation in (a) with the respective confidences $(0.99, 0.16, 0.38, 0.63, 0.88, 0.83, 0.19, 0.99, 0.99)$

Figure 7: Evaluation of an additional example sequence, not included in the dataset. The occlusions of the region of interest with a car in 7 of the 9 images shall lead to low confidence scores. This is the case in the frames $(f_{86}, f_{130}, f_{189})$, and f_{322}, but not for $f_{232}, f_{286})$. Here, the visual appearance if close to *Partially Covered Snow* with a homogeneous region in the center and snow to the left and right borders.

fidence scores. Two images show low confidence because their feature representation is located between two feature clusters in the feature space. The first effect (OoD) is due to large *epistemic* uncertainty, the latter is due to *aleatoric* uncertainty [25]. A distinction of *aleatoric* and *epistemic* uncertainty would be helpful to distinguish between these situations [17].

In Fig. 7, the preceding vehicle occludes the evaluated *Region Of Interest* (ROI). Again, low confidence scores are expected. Indeed, most of the images receive low confidence scores $(f_{86}, f_{130}, f_{189}, f_{322})$, sometimes together with devious classification *Concrete, Damp*. For the frames (f_{232}, f_{286}) a high confidence is estimated which is due to the similarity of the ROI to *Partially Covered Snow* although the car occludes the road. The classification is comprehensible since snow is present at the left/right borders with a rather homogeneous region in-between. It is questionable if we can consider this classification *correct* since main parts of the considered region are from the car roof. Nevertheless, these examples demonstrate the usability of the uncertainty estimation for unknown situations.

5. Conclusions

We propose the novel *RoadSC* (*Road Surface Coverage*) dataset for camera-based road surface condition estimation. The dataset focuses on snow coverage of the road with three different coverage categories. It includes $90{,}759$ bird's eye view patches recorded on a test track as well as on public roads. It is combinable with the *RoadSaW* (*Road Surface and Wetness*) dataset leading to a comprehensive dataset including three snow classes and three road surface types with four wetness levels. Thus, a holistic computer vision road condition dataset with 15 categories and high-quality bird's eye views is provided for development and evaluation.

The baseline evaluation shows that current real-time machine learning based approaches are not accurate enough for an optimal solution, even with carefully selected hyperparameters. For visually similar categories, feature collapse occurs leading to suboptimal performance on *Out-of-Distribution* data. Nevertheless, the significance of the approach is demonstrated with natural application-relevant scenarios. Thus, the dataset provides the possibility to develop and benchmark new models for *Road Condition Estimation*, an important application for automated driving.

Acknowledgements This work has been performed in the framework of the *InFusion* project supported by the *Bundesministerium für Verkehr und digitale Infrastruktur*. The authors would like to acknowledge the contributions of their colleagues from *InFusion*.

References

[1] T. Abe, A. Kanda, K. Kasamura, Y. Nakashima, and Y. Nakanishi. Investigating tire friction on snow-covered runways using electrical conductivity. *Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, 233(7):1036–1045, 2019. 1, 2

[2] S. Bianco, R. Cadene, L. Celona, and P. Napoletano. Benchmark analysis of representative deep neural network architectures. *IEEE Access*, 6:64270–64277, 2018. 4

[3] A. Busch, D. Fink, M.-H. Laves, Z. Ziaukas, M. Wielitzka, and T. Ortmaier. Classification of road surface and weather-related condition using deep convolutional neural networks. In *International Symposium on Dynamics of Vehicles on Roads and Tracks*, pages 1042–1051. Springer, 2019. 2

[4] K. Cordes, C. Reinders, P. Hindricks, J. Lammers, B. Rosenhahn, and H. Broszio. Roadsaw: A large-scale dataset for camera-based road surface and wetness estimation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 4439–4448, 2022. 1, 2, 3, 4, 5, 6

[5] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The cityscapes dataset for semantic urban scene understanding. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 2, 5

[6] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009. 4

[7] C. A. Diaz-Ruiz et al. Ithaca365: Dataset and driving perception under repeated and challenging weather conditions. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21383–21392, 2022. 2

[8] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the KITTI vision benchmark suite. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361. IEEE, 2012. 2

[9] D. Hendrycks and T. Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. *International Conference on Learning Representations*, 2019. 5, 7

[10] P. Jonsson, J. Casselgren, and B. Thörnberg. Road surface status classification using spectral analysis of nir camera images. *IEEE Sensors Journal*, 15(3):1641–1656, 2014. 1

[11] M. N. Khan and M. M. Ahmed. Snow detection using in-vehicle video camera with texture-based image features utilizing k-nearest neighbor, support vector machine, and random forest. *Transportation research record*, 2673(8):221–232, 2019. 1, 2

[12] B. Lakshminarayanan, A. Pritzel, and C. Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30, pages 6402–6413, 2017. 4

[13] L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao, and J. Han. On the variance of the adaptive learning rate and beyond. In *International Conference on Learning Representations*, 2020. 5

[14] L. Liu, S. Lu, R. Zhong, B. Wu, Y. Yao, Q. Zhang, and W. Shi. Computing systems for autonomous driving: State of the art and challenges. *IEEE Internet of Things Journal*, 8(8):6469–6486, 2020. 1

[15] Y. Ma, M. Wang, Q. Feng, Z. He, and M. Tian. Current non-contact road surface condition detection schemes and technical challenges. *Sensors*, 22(24):9583, 2022. 1

[16] W. Maddern, G. Pascoe, C. Linegar, and P. Newman. 1 Year, 1000km: The Oxford RobotCar Dataset. *The International Journal of Robotics Research*, 36(1):3–15, 2017. 2

[17] J. Mukhoti, A. Kirsch, J. van Amersfoort, P. H. Torr, and Y. Gal. Deep deterministic uncertainty: A new simple baseline. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24384–24394, 2023. 7, 8

[18] M. Nolte, N. Kister, and M. Maurer. Assessment of deep convolutional neural networks for road surface classification. In *IEEE International Conference on Intelligent Transportation Systems*, pages 381–386, 2018. 2

[19] M. Pitropov, D. E. Garcia, J. Rebello, M. Smart, C. Wang, K. Czarnecki, and S. Waslander. Canadian adverse driving conditions dataset. *The International Journal of Robotics Research*, 40(4-5):681–690, 2021. 2

[20] J. Postels, M. Segu, T. Sun, L. D. Sieber, L. Van Gool, F. Yu, and F. Tombari. On the practicality of deterministic epistemic uncertainty. In *International Conference on Machine Learning*, volume 162, pages 17870–17909, 2022. 5, 7

[21] S. Roychowdhury, M. Zhao, A. Wallin, N. Ohlsson, and M. Jonasson. Machine learning models for road surface and friction estimation using front-camera images. In *IEEE International Joint Conference on Neural Networks*, pages 1–8, 2018. 2

[22] E. Šabanovič, V. Žuraulis, O. Prentkovskis, and V. Skrickij. Identification of road-surface type using deep neural networks for friction coefficient estimation. *Sensors*, 20(3):612, 2020. 1

[23] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen. MobileNetV2: Inverted residuals and linear bottlenecks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4510–4520, 2018. 2, 4

[24] O. Smadi, N. Hawkins, Z. Hans, B. A. Bektaş, S. Knickerbocker, I. Nlenanya, R. Souleyrette, and S. Hallmark. Naturalistic driving study: development of the roadway information database. 2015. 2

[25] J. Van Amersfoort, L. Smith, Y. W. Teh, and Y. Gal. Uncertainty estimation using a single deep deterministic neural network. In *International Conference on Machine Learning*, pages 9690–9700, 2020. 2, 4, 5, 6, 8

[26] L. Van der Maaten and G. Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008. 4

[27] S. Zang, M. Ding, D. Smith, P. Tyler, T. Rakotoarivelo, and M. A. Kaafar. The impact of adverse weather conditions on autonomous vehicles: How rain, snow, fog, and hail affect the performance of a self-driving car. *IEEE vehicular technology magazine*, 14(2):103–111, 2019. 1