

Winter adverse driving dataset for autonomy in inclement winter weather

Akhil M. Kurup^{ORCID}* and Jeremy P. Bos^{ORCID}

Michigan Technological University, Electrical and Computer Engineering,
Houghton, Michigan, United States

Abstract. The availability of public datasets with annotated light detection and ranging (LiDAR) point clouds has advanced autonomous driving tasks, such as semantic and panoptic segmentation. However, there is a lack of datasets focused on inclement weather. Snow and rain degrade visibility and introduce noise in LiDAR point clouds. In this article, summarize a 3-year winter weather data collection effort and introduce the winter adverse driving dataset. It is the first multimodal dataset featuring moderate to severe winter weather—weather that would cause an experienced driver to alter their driving behavior. Our dataset features exclusively events with heavy snowfall and occasional white-out conditions. Data are collected using high-resolution LiDAR, visible as well as near infrared (IR) cameras, a long wave IR camera, forward-facing radio detection and ranging, and Global Navigation Satellite Systems/Inertial Measurement Unit units. Our dataset is unique in the range of sensors and the severity of the conditions observed. It is also one of the only data sets to focus on rural and semi-rural environments. Over 36 TB of adverse winter data have been collected over 3 years. We also provide dense point-wise labels to sequential LiDAR scans collected in severe winter weather. We have labeled and will make available around 1000 sequential LiDAR scenes, amounting to over 7 GB or 3.6 billion labeled points. This is the first point-wise semantically labeled dataset to include falling snow. © The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.OE.62.3.031207](https://doi.org/10.1117/1.OE.62.3.031207)]

Keywords: inclement weather; degraded visual environments; light detection and ranging scans; labeled dataset; semantic segmentation.

Paper 20221058SS received Sep. 15, 2022; accepted for publication Nov. 28, 2022; published online Jan. 3, 2023.

1 Introduction

Autonomous vehicles (AV) and robo-taxis have been slowly making their way into our daily lives. Tasks, such as lane-keeping, parking assist, and automated lane changes, are some of the features available in modern production vehicles as part of advanced driver assistance system (ADAS) feature packages. Inclement winter weather, such as heavy rain and snow reduces visibility and free-flowing traffic speeds.¹ Populated North American cities, such as Detroit, Chicago, Minneapolis, and many others, can receive over 1 inch (2.5 cm) of snow per hour, severely affecting the transportation infrastructure. The AVs must be capable of operating in such conditions to ensure universal adaptation. Current technologies, however, lack the capability to operate effectively in adverse winter conditions and one of the leading reasons is the lack of available winter weather datasets. Providing large and varied datasets for training deep learning can be challenging. In the AV space, the problem is sometimes solved by hiring human drivers to drive automobiles in varied traffic scenarios and locations. A popular approach is to use simulation tools in different scenarios.^{2,3} However, it has been recognized that the performance of deep learning approaches is limited by the availability of “corner cases” in the training set. In short, when these algorithms are exposed to scenarios not present in their training data they can fail in sometimes unexpected ways.^{4,5}

*Address all correspondence to Akhil Kurup, amkurup@mtu.edu

The AV perception systems rely on cameras, light detection and ranging (LiDAR), radio DAR (RADAR), and some combination of these sensors to overcome individual shortcomings. Precipitation, such as rain and snow, degrades the performance of perception systems by introducing false detections and reducing visibility.^{6,7} The LiDAR sensors are particularly affected by absorption and scattering effects, exacerbated by their inherent beam divergence and short pulse duration. Snow shows up as a clutter of noise concentrated near the LiDAR⁸ affecting common tasks, such as object detection, tracking, and simultaneous localization and mapping (SLAM). Precipitation is often hard to predict, and severe events are infrequent.

Houghton, Michigan, United States, has severe winter weather between December and February each year. Located on the Keweenaw Peninsula and surrounded by Lake Superior on three sides, lake-effect snow is frequent whenever conditions are favorable. Consequently, the region receives over 200 in (500 cm) of snow on average annually and local records are as much as 360 in (900 cm); more than many snow resorts. Simultaneously, though rural, around 20,000 people live in the area. The region also supports well-developed infrastructure leftover from the region's copper mining days. Blowing snow can result in both intermittent and persistent white-out conditions where visibility is near zero. The frequency of such adverse weather in this area allows reliable collection of large winter driving data featuring extreme snow events. Whereas few events would locally be considered severe winter weather, they would likely pose a challenge for most drivers in large metropolitan areas.

In this paper, we summarize three seasons of data collection efforts and introduce the winter adverse driving dataset (WADS), aptly named after Wadsworth Hall, the largest dorm at Michigan Tech. We have gathered over 36 TB of winter driving data featuring moderate to severe driving conditions.^{8,9} We provide an overview of our autonomy data recorder (ADR) and interchangeable parts for enabling autonomy (IPEA) sensor pod concept. WADS captures active falling snow across different sensors as well as snow accumulated on the sides of the roads from vehicle movement and snow removal management. Our base sensor pod collects data from two side-mounted LiDARs, three forward-facing cameras [(visible, near infrared (NIR), and long wave infrared (LWIR)], Real Time Kinematic-corrected Global Navigation Satellite Systems (GNSS), and an Inertial Measurement Unit (IMU). All are mounted external to the vehicle and connected to a custom-built robot operating system (ROS)-based data recording system. Our sensor pod also includes a mounting point for a high-definition (HD) LiDAR and other sensors. Over the past 3 years, we have tested and evaluated several guest LiDARs. Our data collection hardware also includes an autonomous driving surrogate vehicle. This platform features a single 32 channel LiDAR, a single forward-facing camera behind the windshield, 2 forward-facing RADARs, and a GNSS/IMU unit.

We also make available a semantically labeled portion of WADS, presented here and the first of its kind. Figure 1 shows examples from our labeled dataset collected in urban driving during moderate to severe snow. We believe public access to such data will propel the development of neural networks (NNs) trained to operate in degraded visual environments due to adverse winter weather. Scene understanding in snowy conditions can be used to determine drive-able areas and improve object detection and avoidance, whereas segmentation of active snow can help improve visibility in white-out conditions.

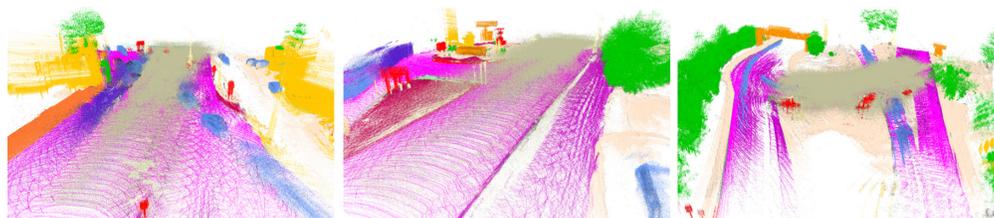


Fig. 1 Our WADS provides dense point-labels to LiDAR scans collected in moderate to severe snow. Semantic labels for active snow (tan color) and accumulated snow (beige color) are unique to our dataset.

Table 1 Publicly available datasets with annotated LiDAR scans. WADS is the first dataset to feature dense point-wise labeled LiDAR scans in severe winter weather.

Dataset	LiDAR	Labeling	Inclement weather?
SemanticKITTI ¹¹	64 channel	Point-wise	No
nuScenes-lidarseg ¹²	32 channel	Point-wise	No
ApolloScape ¹³	64 channel	Point-wise	No
DENSE ¹⁴	32 and 64 channel	Bounding boxes	Yes
CADC ¹⁵	32 channel	Bounding boxes	Yes
WADS (ours)	64 channel	Point-wise	Yes

2 Related Work

Several annotated datasets have been released with LiDAR scans in the recent decade to aid with the development of AV perception tasks, such as segmentation.¹⁰ A complete review of these datasets is outside the scope of this paper. Here, we only discuss the most relevant works addressing inclement weather. Table 1 provides an overview of relevant datasets and our proposed dataset.

Pfeuffer and Dietmayer¹⁶ present an evaluation of various NNs trained for tasks, such as object detection and avoidance. They show that models trained on large datasets, such as KITTI,¹⁷ fail to perform well in adverse weather conditions implying that the availability of data takes precedence over the size of the dataset. The lack of adverse weather data has been addressed in some literature by adding artificial noise such as rain or snow to existing datasets. Sakaridis et al.¹⁸ uses a fog model to add synthetic noise to images and shows an improvement in semantic segmentation using convolutional neural networks. Laser interactions with the environment have been studied by Roy et al.¹⁹ They have modeled the interaction between snow particles and laser pulses to statistically determine the amount of snow per sampled volume based on the characteristics of the laser beam and snow precipitation. Heinzler et al.²⁰ uses a fog and rain model to de-noise point clouds in adverse conditions. They, however, do not present results in snow and extreme weather.

The KITTI¹⁷ and nuScenes¹² datasets provide LiDAR scans annotated with bounding boxes but no data in inclement weather. Correspondingly, SemanticKITTI¹¹ and nuScenes-lidarseg datasets were introduced with point-wise annotations. These include labels for each point in the point cloud, enabling finer details around objects for tasks, such as semantic segmentation and better scene understanding. The ApolloScape¹³ dataset includes LiDAR scans with a semantic mask to extract point-wise annotations. Their current iteration does not include inclement weather, but the authors plan to include fog and snow in later releases. The DENSE¹⁴ dataset includes rain, fog, and snow. Extreme weather is, however, rare thus limiting its usability in training perception systems. Moreover, annotations are limited to bounding boxes. The CADC¹⁵ dataset includes adverse weather data collected in Canada with bounding boxes around vehicles and pedestrians. These annotations are useful for tasks, such as object detection but provide little information for scene understanding. Our dataset provides point-wise annotations for LiDAR scans collected in harsh driving conditions. Unlabeled datasets have been collected by the authors over the last three years and make up the bulk of WADS presented here.^{8,9,21}

3 System Setup

We primarily collect data using two platforms: an IPEA concept system together with our ADR and an AV surrogate. We have introduced our IPEA system, we call the “sensor-pod,” in our



Fig. 2 (a) Our IPEA system (sensor pod)⁸ after year 2 testing campaign⁹ mounted on a Husky A-200 robot with a 64-channel OS1 LiDAR unit at the top, a color camera on the left, an NIR camera on the right, and an LWIR camera in the middle. (b) The sensor pod was mounted on a test vehicle with a 905-nm MEMS LiDAR unit in our year 3 testing campaign.

previous work.^{8,9} It has a common reconfigurable base platform, designed to be easily mounted on any platform to enable autonomous perception and data collection as shown in Fig. 2. The base configuration includes a color camera (towards the left), an LWIR thermal camera (in the center), and an NIR camera (towards the right) of the sensor pod. Over three campaigns, we have tested the performance of several high-resolution LiDARs in inclement weather. The test LiDAR is mounted at the top of the sensor pod and two 16 or 32-channel VLP-32 LiDAR are diagonally mounted on the sides. An Emlid Reach RS GNSS unit is also mounted on the sensor pod.

Our ADR enables near-synchronous data capture from the sensor pod. Fully realized, the ADR consists of a computer platform mounted in a weather-proofed case, capable of being powered from 12 or 24V batteries with quick disconnects for each sensor. In its current form, the ADR consists of a SuperMicro E3 4-core Xeon equipped motherboard mounted to the ADR enclosure. The operating system and system software run on a 256-GB NVMe drive. Data are recorded to a striped RAID 0 array with a 12 TB capacity. The system runs on Ubuntu 18.04 and features several ROS Melodic packages, such as RViz for visualization, robot_description for sensor models, sensor drivers (velodyne, usb_cam, RADAR driver, and other LiDAR drivers) and others. The Supermicro motherboard includes four onboard GigE ports. An additional four ports are available via a PCIe expansion card.

Our primary aim with the IPEA system was to develop the ability to easily change sensor load-out without having to detailed measurements and calibrations. It also provides an easy-to-move solution between platforms. In its current form, it can be easily moved from a car rack to an unmanned ground vehicle (UGV) in under an hour. The xacro-based ROS universal robot description format (URDF) description of the IPEA contains TF transforms between all mounting points and it is straightforward to add or remove sensors with only a cursory understanding of ROS. Throughout our testing campaign, we were able to change add new sensors and change the orientation of others in the field with minimal tooling. Figure 2 shows our IPEA mounted to our UGV (left) and a roof rack (right). Currently, the main obstacle in reducing this time is cable management. Connectorizing our base sensors and standardizing power distribution will also improve switch-over time.

In addition to the sensor pod, we also collect data with an AV surrogate vehicle. A 32-channel VLP-32 LiDAR is mounted on the top of the vehicle with a dedicated GNSS system for positioning. Our AV surrogate platform is pictured in Fig. 3. A single forward-facing camera is mounted inside the vehicle, behind the windshield, to protect it from the elements. In year three, we include two automotive RADARs, operating at 77 GHz. Combining point cloud returns from individual sensors can result in a higher point density as noted in.²² Figure 4 shows that RADAR returns are largely unaffected by snow particles but also highlights the superiority of LiDAR point density in feature recognition.



Fig. 3 Our AV surrogate platform with the forward facing RADAR (a) mounted on the front bumper and a LiDAR (b) mounted on top of the vehicle.

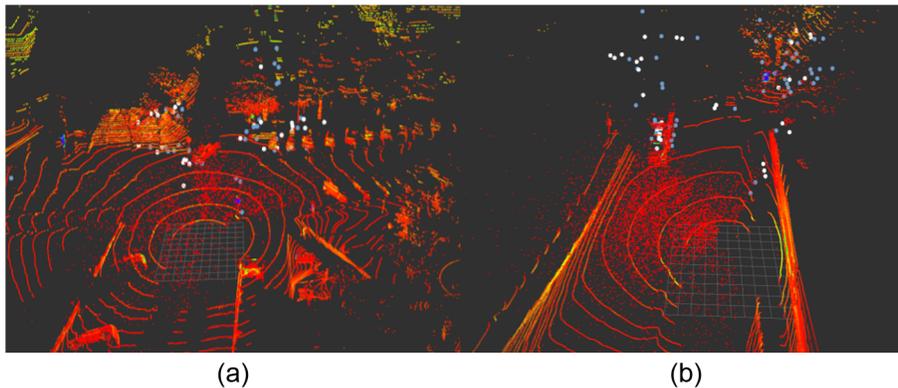


Fig. 4 Example data from 22 February 2022 in snowfall rates of 0.7 in (1.7 cm) per hour. RADAR returns (white and blue points) are visibly unaffected by snow in both urban driving (a) as well as highway driving (b).

4 Inclement Weather Dataset

As mentioned in Sec. 1, winter storms are frequent in the community near Michigan Tech from January through February and enable the reliable collection of winter weather data. Lake effect snow events resulting in 3 to 5 inches (8 to 12 cm) are common, but difficult to predict. Winter storms with snowfall totals of 12 inches (30 cm) are generally more predictable but less common. Blowing winds often accompany snow events, leading to low visibility and poor driving conditions that challenge even seasoned drivers.

Over the past three seasons, we have collected data and tested guest LiDAR sensors over fourteen snow events, resulting in over 36 TB of AV sensor data featuring exclusively adverse driving conditions. In our year 1 campaign, we ended up collecting data for every snow event. In year 2, we focused on high precipitation events (snow rate > 1 in per hour) and tested both 905 nm as well as 1550-nm LiDARs. In year 3, we again focused on high precipitation events and added RADARs. These events have been summarized in Appendix A, Table 2. Details of these events with information, such as weather, sensors tested, and geographical areas can be found in our previous work.^{8,9} Weather conditions reported here are from Michigan Tech's Keweenaw Research Center.²³

In laying out our testing, we generally observed the weather forecast for the week and planned to collect data on days or evenings where substantial snowfall was expected to occur. Routes varied but commonly included the loop from the Michigan Tech Advanced Power System Research Center (APSRC) to Houghton County Memorial Airport (airport code CMX) and back to the APSRC along Airpark Blvd. Another common route starts at the APSRC and goes to US-41 via Airpark Blvd. US-41 then takes us to Michigan Tech's campus. Testing around campus involved driving from a parking location around campus on US-41, Cliff Dr., and Phoenix Dr. The latter brings us down to the Portage canal and features a large hill to the south severely reducing the number of visible GNSS satellites. From starting points on campus,

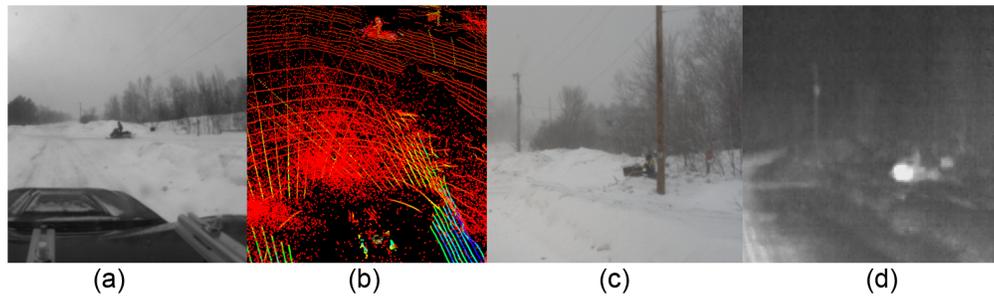


Fig. 5 Example data from 12 February 2020 showing a snowmobile crossing the Boston Location road: (a) NIR, (b) LiDAR, (c) visible, and (d) LWIR imagery taken from our sensor pod.

in Houghton, or at the APSRC we commonly drove to Calumet, Michigan, United States. At 1214 ft (370 m), compared with Houghton's 643 ft (196 m) Calumet often receives significantly more snowfall. Routes running from Houghton to Eagle River Michigan take US-41 to M26 into Eagle River and back. This corridor in Keweenaw County often features some of the worst winter weather. Other routes were selected at random based on weather radar and perceived or predicted chances of precipitation.

As far as we are aware, this is the first AV dataset containing coplanar LWIR, visible, and NIR imagery. Our unique dataset features items that stand out and are not likely seen on roadways in areas that do not have persistent snow on the ground over the winter. An interesting example is the presence of snowmobiles adjacent to or even on roadways as shown in Fig. 5. Not uncommon throughout the rest of the United States would be the presence of deer on or adjacent to the road. However, detecting deer behind snowbanks or at tree lines without some sort of LWIR camera is likely difficult if not impossible (see Fig. 14). In the year 2 data, we observe "blooming" effects around objects with high reflectivity (traffic signs), in the presence of ice on the sensor surface (shown in Fig. 6). Rapid ice buildup has often resulted in short segments of data collection followed by manual clean up of the sensors. High snowfall rates are another unique feature of our data. Most of the data collections from Year 3 feature snowfall rates in excess of one inch (2.5 cm) per hour (Fig. 7).

Lane lines are generally not visible during the winter months in Houghton, Michigan, United States. In fact, the concepts of lanes on roads that are frequently snow-covered is ambiguous and may depend on local tradition. On infrequently traveled roads drivers may center themselves on the roadway moving to their right only when another vehicle approaches. On snow-covered three or four lane roads lanes are often defined by the path taken by the vehicle ahead of you or wherever

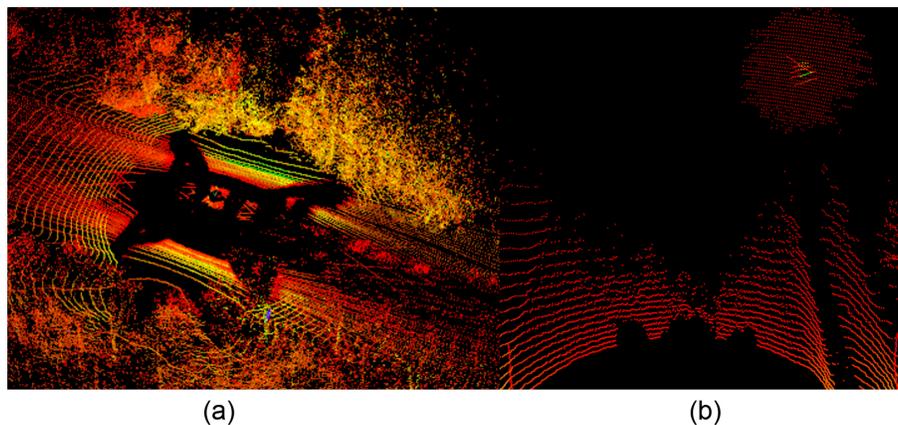


Fig. 6 (a) Example LiDAR image from the OS1-64 and two diagonally mounted VLP-32s; a vortex trail is present behind the vehicle. (b) LiDAR image showing degraded performance when ice has accumulated on the front of the Ouster as well as blooming at the upper right of the image around a stop sign. Note that the crossed lines from the Velodynes are coincident with the bloom.



Fig. 7 Example images collected during snowfall rates around 1 in (2.5 cm) per hour. The visible imagery (a) and NIR (b) show visible snowflakes, whereas the LWIR imagery (c) shows no effect of snowflakes.

tracks are located. Similarly, pedestrian behavior also changes in the winter. Especially on side streets you are likely to see people walking in the roadway because sidewalks are not present or are snow covered. All of these behaviors are present in various portions of WADS. In Appendix D, we break down each of the data files found in the WADS year 3 set as well as the type of unique winter features found therein. Example images of each type are also included.

Snowbanks create their problems as they change as often as daily in the winter months. Localization using HD LiDAR maps would be difficult without adding a heuristic or including them as a ground plane component. To that point, snow on the roads whether piled, smooth or tracked is likely to create issues with ground plane identification and subtraction. We anticipate these situations may trouble ADAS and AV systems that rely on machine learning in particular.

5 Labeled LiDAR Dataset

As mentioned above, we have collected over 36 TB of winter driving data over the past three winters. We selected data collected on February 12, 2020, to label around 1000 scans with more to be added as and when they are labeled and verified. The temperature on this day fell from a high of 28 F (−2 C) at 9 am to 7 F (−14 C) at 6 pm. Data collection on this day started around 1 pm from the Keweenaw Research Center. Low visibility due to blowing snow coupled with heavy winds (up to 25 mph; 40 kph) made for challenging driving conditions. Scans from our dataset have been split into sequences of approximately 100 scans each. Every scan has associated pose information which is used to aggregate scans to further the development of algorithms using spatial information. Multiple suburban scenes have been captured, including two-lane highways, residential areas, and parking as well as moving vehicles. Figure 1 shows a few labeled scenes collected during moderate snow. Points have been labeled into one of 22 classes including active snow and accumulated snow, which are exclusive to our dataset.

5.1 Labeling

Bounding boxes provide vector annotations and often include undesired background objects which can be detrimental for AV perception tasks, such as semantic segmentation. We have opted for point-wise labels as they are more precise and enable fine details in the environment to be highlighted, such as individual snowflakes. Manual labeling of point clouds is a tedious process, exacerbated by having to work around suspended snow particles. To maintain compatibility with existing systems and ensure the adoption of inclement weather data into existing frameworks, we use the popular KITTI format.¹⁷ We leverage the point-cloud labeling tool introduced by Behley et al.¹¹

To speed up the process, annotators superimpose several scans using pose information, available with our dataset. Figure 8 (a) shows a single labeled scan and (b) shows several scans superimposed using pose information. On average, annotators need approximately 6 hrs per sequence of scans to label and resolve occlusions. Labeled scans are assessed by a second annotator to correct any errors and ensure data quality.



Fig. 8 (a) A single labeled scan showing active falling snow and accumulated snow. (b) Several scans have been labeled and superimposed using pose information enabling spatial scene understanding.

Each scan is stored as a floating-point binary (.bin) format in the velodyne directory while corresponding labels are stored as label files in the labels directory. Both of these can be easily read using most programming languages. The poses.txt file holds pose information for every scan. This provides spatial information to users. Note that the use of “velodyne” in this file does not imply the pointclouds were captured by a Velodyne LiDAR.

5.2 Statistics

In our labeled dataset, every point in a LiDAR scan has been labeled into one of 22 classes as shown in Fig. 9. Here, classes are grouped into categories for easy viewing. Around 1000 LiDAR scans have been completely labeled amounting to over 7 GB or 3.6 billion points in all. The majority of labeled points lie in urban driving scenarios with roads, buildings, and various types of vehicles representing most of our labeled data. A good proportion of vegetation and other terrain exists as well making our dataset valuable for training NNs. The number of labeled points varies per class leading to an unbalanced dataset which is common for datasets collected outdoors. For example, because this is a rural adverse weather dataset, we expect fewer vehicles to be outdoors which is why we see fewer labeled points representing different vehicles.

In addition to these classes, we introduce two new classes to represent snow that is noticeably not found in other datasets: “active-snow” captures falling snow particles and associated clutter noise in a LiDAR return whereas “accumulated-snow” captures snow that builds upon the sides

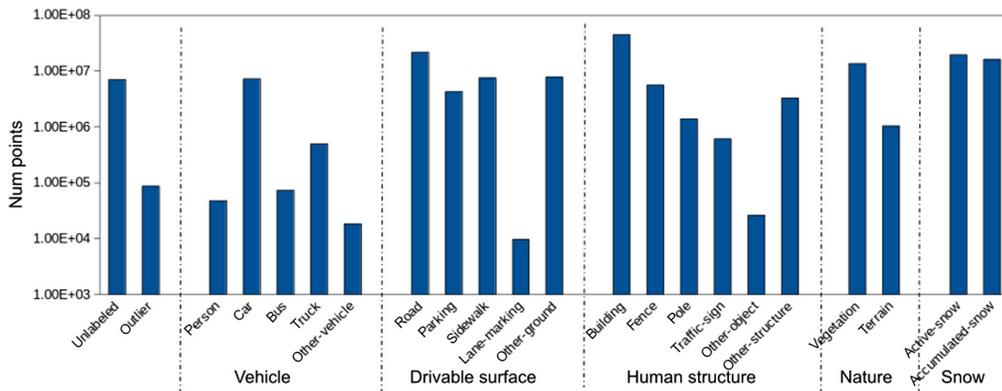


Fig. 9 Distribution of classes in the WADS dataset. Scenes from suburban driving including vehicles, roads, and man-made structures are included. Two novel classes: active-snow and accumulated-snow are introduced to improve AV perception in adverse winter weather.

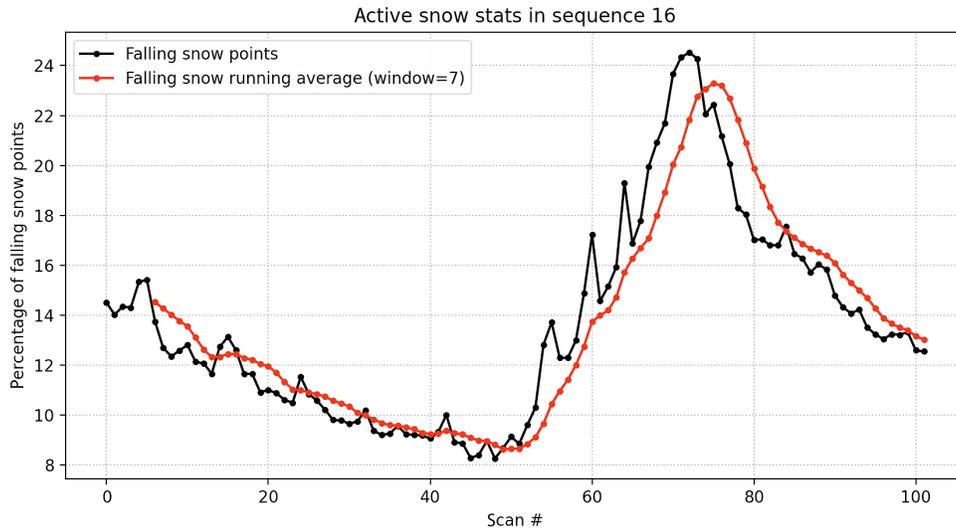


Fig. 10 Percentage of falling snow in sequence 16 ranges from 10% to over 20% of the total points.

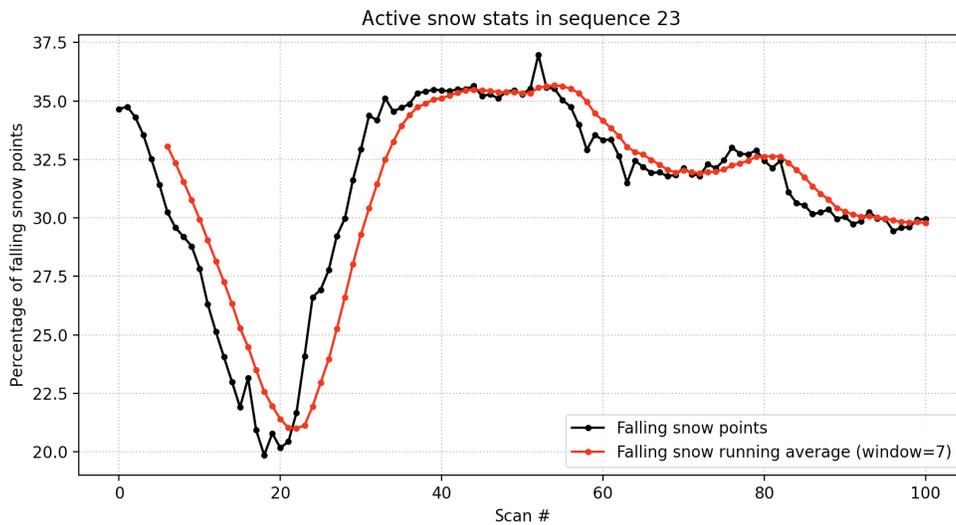


Fig. 11 Percentage of falling snow in sequence 23 ranges from 20% to over 32% of the total points.

of drive-able surfaces due to vehicle traffic and snow removal. Accumulated snow often changes, sometimes throughout the day, and may confuse feature-based algorithms. Overall, active snow makes up 10%, whereas, accumulated snow accounts for 21% of our labeled dataset. As seen in Figs. 10 and 11, individual rate of falling snow can vary from scan to scan even within sequences. Access to such data will be useful for AV tasks, such as object detection, localization and mapping, and semantic and panoptic segmentation, in adverse weather.

6 Conclusion and Future Work

Adverse weather conditions negatively affect perception systems used in AV. In particular, LiDAR point clouds suffer from false detections (both positive and negative) introduced by falling rain and snow. Until now, a lack of datasets focused on inclement winter weather has limited the development of AVs to good clear weather conditions. In this work, we have summarized

a 3-year campaign of winter data collection in adverse driving conditions in Michigan's Keweenaw Peninsula. Our WADS is composed of over 36 TB of multimodal data and is the first to feature severe snow and white-out conditions. Our data also feature exclusive events, such as snowmobiles and wildlife, which are absent from other datasets and may negatively impact ADAS functions. We also introduced dense point-wise labels for our dataset to further AV tasks, such as object detection, localization and mapping, and semantic and panoptic segmentation, in adverse weather. We propose two class labels, falling snow, and accumulated snow to represent conditions that are notably absent from other open-source datasets.

Going forward, we would like to provide annotated images and possibly RADAR data to enable sensor fusion in winter weather. We have also touched upon processing the AV data, however, in future works, we hope to compare the performance of common AV tasks, such as fusion, detection and classification, and SLAM.

7 Appendix A: Winter Data Collection Events

In this section we provide a full description of the individual data collection events that make up WADS. In Table 2 we attempt to capture not only the date and times of the collections. We also provide a subjective description of the test conditions that would be familiar to those oriented to the local climatology.

Table 2 Summary of winter data collection events across three seasons. Precise details of specific events, sensors used, and interesting observations can be found in the individual works.^{8,9}

Events	Conditions	Summary
18 January, 2020	12 in (30 cm) snow; 27°F	First test with IPEA and ADR
23 January, 2020	3 in (7.6 cm) snow; 33°F	Lake effect snow leading to eventual ice accumulation on sensors
12 February, 2020	3.2 in (8 cm) snow; 15°F	Low visibility due to blowing snow with strong winds as high as 25 mph (40 kph)
17 February, 2020	2.2 in (5.7 cm) snow; 33°F	Lake effect snow
27 February, 2020	6.5 in (16 cm) snow; 14°F	Lake effect snow
05 March, 2020	3.25 in (8 cm) snow; 31°F	Wet snow with occasional rain (seen as LiDAR blooming)
16 January, 2021	1 in (2.5 cm) snow; 32°F	Initial integration test for the second season
04 February, 2021	12 in (30 cm) snow; 32°F	High precipitation arctic weather system with wet and heavy snow
06 February, 2021	5 in (13 cm) snow; 32°F	Dry and fine blowing snow with winds up beyond 30 mph (13 m/s)
28 February, 2021	10 in (25 cm) snow; 20°F	Wet and dry lake effect snow, turning into dry fluffy snow towards the end
05 January, 2022	8 in (20 cm) snow; 29°F	Low visibility due to wet and heavy blowing snow with wind gusts as high as 45 mph (72 kph)
09 January, 2022	4 in (10 cm) snow; 26°F	Lake effect snow with wind gusts as high as 25 mph (40 kph)
21 February, 2022	10.5 in (26 cm) snow; 14°F	Small blowing snow particles leading to poor visibility
22 February, 2022	16 in (40 cm) snow; 14°F	Low visibility due to blowing lake effect snow

8 Appendix B: Examples from the WADS Dataset Years 1 and 2

Included here are example images of the data collected in years one and two of the WADS effort. These include unfamiliar arrangements of persons and devices as well as snow-moving equipment on roadways (Fig. 12). Figures 13 and 14 highlight the usefulness of LWIR camera and a high mounted lidar in detecting occluded obstacles during nighttime conditions. This portion of

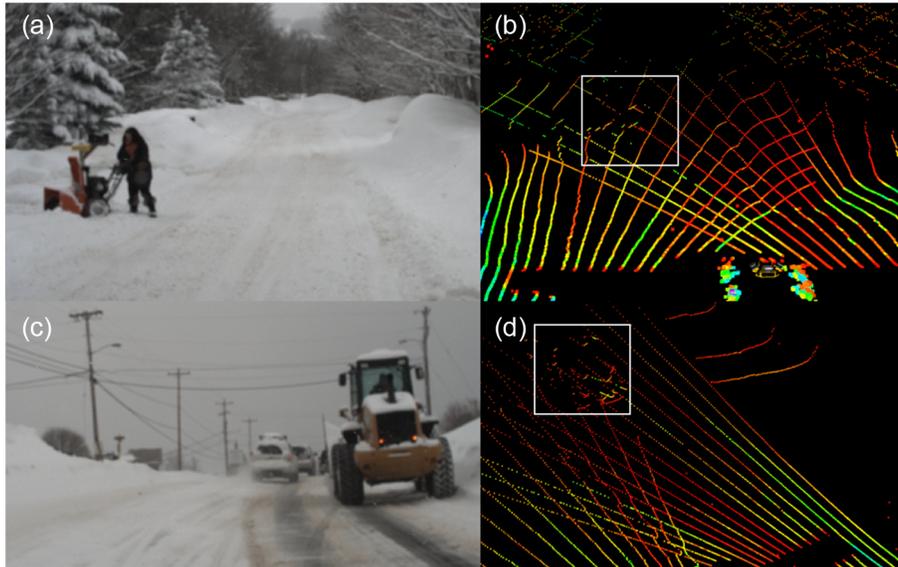


Fig. 12 (a) A Portage township resident clears their driveway using a snowblower in the aftermath of the storm. (b) LiDAR point cloud with the person and snowblower highlighted. (c) A front end loader on Sharon Ave. in Houghton following the storm. (d) LiDAR point cloud with front end loader highlighted.

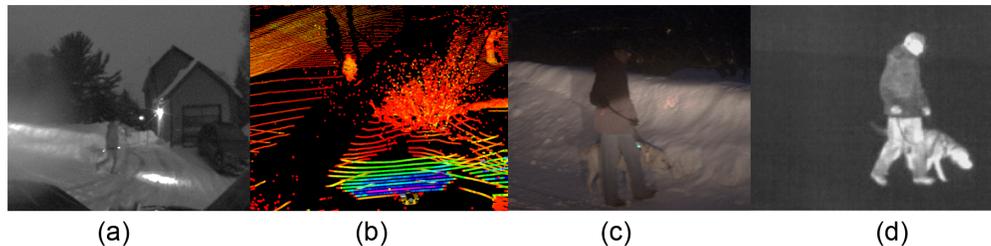


Fig. 13 (a) NIR, (b) LiDAR, (c) visible, and (d) LWIR imagery of a pedestrian walking a dog.

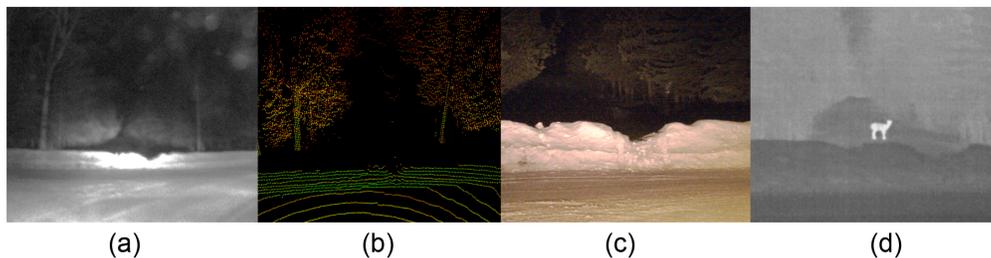


Fig. 14 (a) NIR, (b) LiDAR, (c) visible, and (d) LWIR imagery of a whitetail deer behind a snowbank.

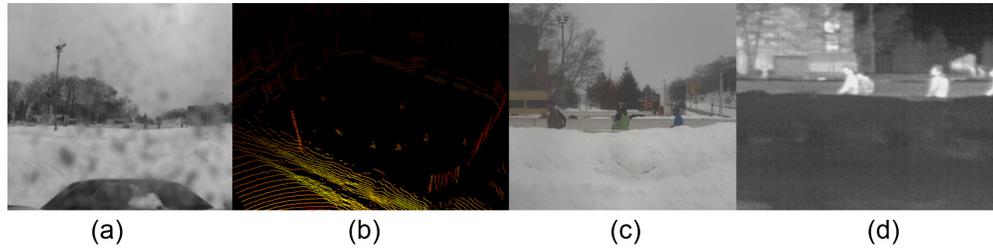


Fig. 15 (a) NIR, (b) LiDAR, (c) visible, and (d) LWIR imagery of Michigan Tech students playing broomball. Ice accumulation is visible on the NIR image but also present on the other sensors.

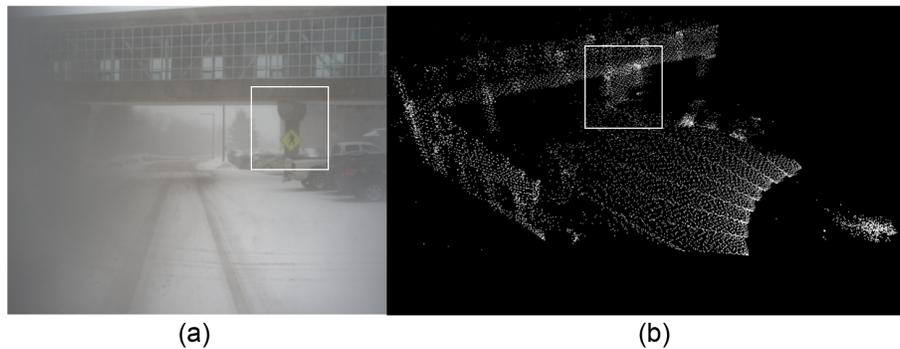


Fig. 16 Example of the blooming present when water droplets were present on a LiDAR optical window. (a) The reflective pedestrian crossing sign is highlighted. It is matched to the highlighted area in the LiDAR point cloud on the right.

the data set also includes novel arrangement of persons (Fig. 15) and blooming from accumulated water-ice on a lidar optical window (Fig. 16).

9 Appendix C: Example Labeled Scans from the WADS Dataset

Here we feature some examples of the labelled pointclouds available in WADS as well as highlighting some unique features in the dataset. These features include a water-crossing lift bridge (Fig. 17), complex intersections (Figs. 18 and 19) as well as locally intense traffic and multi-story buildings (Figs. 18–20).

Labeled scenes from our WADS dataset are shown here. Moving objects span across point clouds and show up as streaks. Tan-colored active snow is detected close to the sensor and

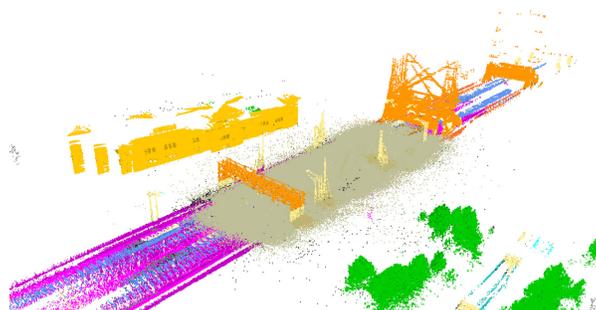


Fig. 17 A labeled scan showing a portion of the Portage Lake Lift Bridge that connects the Keweenaw to the Upper Peninsula of Michigan.

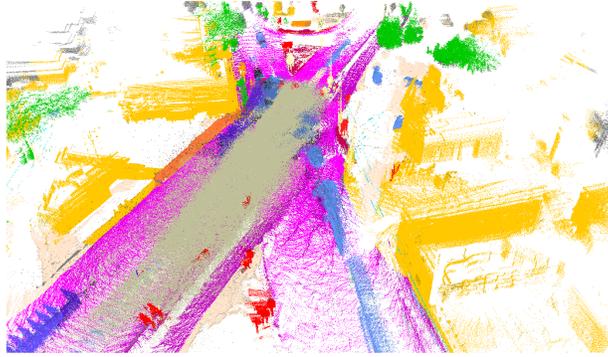


Fig. 18 A labeled scan showing a three-way intersection in Hancock, Michigan, United States, during a busy hour.

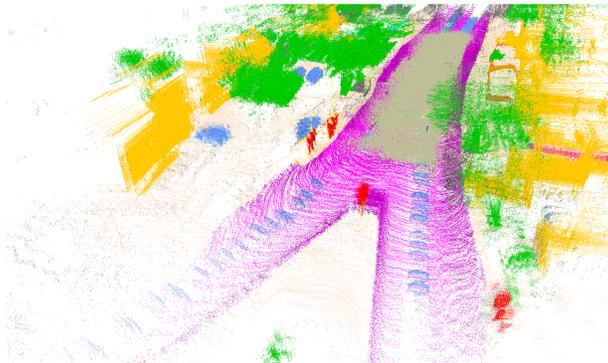


Fig. 19 A labeled scan showing a fork in a residential area in Houghton, Michigan, United States.



Fig. 20 A labeled scan showing a semiurban scene in Houghton, Michigan, United States.

therefore appears to be following the path of the vehicle. Streaks in blue are from moving vehicles.

10 Appendix D: Detailed Description of all Year 3 Files

In Fig. 21 we detail each of rosbag in the WADS Year 3 data set along with the enumerated features listed. Entries without checkmarks may still contain heavy falling snow and heavy traffic. Examples of each of the categories listed in Fig. 21 can be found in Figs. 22–28.

bag file name	driving outside lane boundaries	Truck with plot	Municipal plow or other vehicle	Heavy blowing snow	Snowmobile	White out conditions	Snow on road large snow banks	Pedestrian near roadway
5_Jan_AMK2CLM_1_a.bag	x	x						
5_Jan_AMK2CLM_1_b.bag	x	x				x		
5_Jan_APS2MTU_2_a.bag	x		x					
5_Jan_APS2MTU_2_b.bag			x	x				
5_Jan_APS2MTU_3_a.bag	x		x				x	
5_Jan_APS2MTU_3_b.bag			x	x				
feb_21_mtu2aps_2_a.bag		x		x			x	
feb_21_mtu2aps_2_b.bag	x		x					
feb_22_airport.bag	x					x		
feb_22_aps2clm_1_a.bag	x	x						
feb_22_aps2clm_1_b.bag		x				x		
feb_22_aps2mtu_1_a.bag		x						
feb_22_aps2mtu_2_a.bag			x	x			x	
feb_22_aps2mtu_2_b.bag		x				x	x	
feb_22_cali2aps_1_a.bag				x				
feb_22_cali2aps_1_b.bag	x							
feb_22_cali2mhk_2_a.bag	x			x		x	x	
feb_22_cali2mhk_2_b.bag	x	x		x		x		
feb_22_cali2mhk_3_a.bag		x						
feb_22_cali2mhk_3_b.bag	x		x					
feb_22_h2aps_1_a.bag		x		x			x	
feb_22_h2aps_1_b.bag	x	x						
feb_22_h2h_1_a.bag						x	x	
feb_22_h2h_2_a.bag		x					x	
feb_22_h2h_2_b.bag		x					x	
feb_22_mhk2cali_1_a.bag					x			
feb_22_mhk2cali_2_a.bag					x			
feb_22_mhk2cali_2_b.bag		x				x		
feb_22_mtuloop_1_a.bag				x		x	x	
feb_22_mtuloop_1_b.bag							x	
5_Jan_CLM2APS_1_a.bag		x					x	
5_Jan_CLM2APS_1_b		x						
5_Jan_CLM2H_1_a.bag								
5_Jan_CLM2H_3_a.bag		x						
5_Jan_CLM2H_3_b.bag		x						
5_Jan_CLM2H_4_a.bag								
5_Jan_Airport.bag	x	x				x		
5_Jan_H2AMK_1_a.bag	x				x			
5_Jan_ERLoop2H_4_b.bag	x							
5_Jan_ERLoop2H_4_a.bag	x							
5_Jan_ERLoop2H_3_b.bag	x							
5_Jan_ERLoop2H_3_a.bag	x		x					
5_Jan_COOP2CLM_1_a.bag	x	x	x					
5_Jan_COOP2CLM_2_a.bag			x		x			
5_Jan_COOP2CLM_2_b.bag		x			x			
5_Jan_ERLoop2H.bag	x				x	x		
Total	21	22	11	2	9	7	12	13

Fig. 21 Tabulation of all “bag” files in the WADS year 3 dataset and WADS specific features found in each bag.



Fig. 22 WADS year 3: examples of high snow banks and blown snow over road.



Fig. 23 WADS year 3: examples of following tracks off lane centers or driving in the middle of the roadway.



Fig. 24 WADS year 3: examples of municipal snow plows.



Fig. 25 WADS year 3: examples of other municipal snow removal equipment.



Fig. 26 WADS year 3: examples of pedestrians occluded by snow or walking on roadway.



Fig. 27 WADS year 3: driven and parked snowmobiles adjacent to road.



Fig. 28 WADS year 3: low visibility conditions caused by blowing-snow or whiteout conditions.

Acknowledgments

Portions of this work were made possible by a Michigan Tech Research Excellence Fund, Infrastructure Enhancement grant. Robotics Systems Enterprise students (RSE) Ian Mattson, Alexander Nedvidek, Makayla Miller, Aun Abbas, Jay Sweeney assisted with preparing the year 3 table in the Appendix D and the associated images. Students from RSE also assisted in labeling the LiDAR point clouds scans. Derek Chopp designed and built the IPEA and ADR.

Code, Data, and Materials Availability

Our labeled dataset is publicly available at Ref. 24. For the raw data, please reach out to the authors.

References

1. H. Rakha et al., "Inclement weather impacts on freeway traffic stream behavior," *Transport. Res. Rec.* **2071**(1), 8–18 (2008).
2. S. Chen, Y. Leng, and S. Labi, "A deep learning algorithm for simulating autonomous driving considering prior knowledge and temporal information," *Comput.-Aid. Civ. Infrastruct. Eng.* **35**(4), 305–321 (2020).
3. D. J. Fremont et al., "Formal scenario-based testing of autonomous vehicles: from simulation to the real world," in *IEEE 23rd Int. Conf. Intell. Transport. Syst. (ITSC)*, pp. 1–8 (2020).
4. W. G. Hatcher and W. Yu, "A survey of deep learning: platforms, applications and emerging research trends," *IEEE Access* **6**, 24411–24432 (2018).
5. S. Abrecht et al., "Testing deep learning-based visual perception for automated driving," *ACM Trans. Cyber-Phys. Syst.* **5**(4), 1–28 (2021).
6. Q. Xu et al., "SPG: unsupervised domain adaptation for 3D object detection via semantic point generation," (2021).
7. J.-I. Park, J. Park, and K.-S. Kim, "Fast and accurate desnowing algorithm for LiDAR point clouds," *IEEE Access* **8**, 160202–160212 (2020).
8. J. P. Bos et al., "Autonomy at the end of the earth: an inclement weather autonomous driving data set," *Proc. SPIE* **11415**, 1141507 (2020).

9. J. P. Bos et al., "The Michigan Tech autonomous winter driving data set: year two," *Proc. SPIE* **11748**, 1174809 (2021).
10. Y. Xie, J. Tian, and X. X. Zhu, "Linking points with labels in 3D: a review of point cloud semantic segmentation," *IEEE Geosci. Remote Sens. Mag.* **8**(4), 38–59 (2020).
11. J. Behley et al., "Semantickitti: a dataset for semantic scene understanding of LiDAR sequences," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, pp. 9297–9307 (2019).
12. H. Caesar et al., "nuScenes: a multimodal dataset for autonomous driving," in *Proc. IEEE/CVF Conf. Comput. Vis. and Pattern Recognit.*, pp. 11621–11631 (2020).
13. X. Huang et al., "The apolloscape open dataset for autonomous driving and its application," *IEEE Trans. Pattern Anal. Mach. Intell.* **42**, 2702–2719 (2020).
14. M. Bijelic et al., "Seeing through fog without seeing fog: deep multimodal sensor fusion in unseen adverse weather," in *IEEE/CVF Conf. Comput. Vis. and Pattern Recognit. (CVPR)* (2020).
15. M. Pitropov et al., "Canadian adverse driving conditions dataset," *Int. J. Robot. Res.* **40**(4–5), 681–690 (2020).
16. A. Pfeuffer and K. Dietmayer, "Optimal sensor data fusion architecture for object detection in adverse weather conditions," (2018).
17. A. Geiger et al., "Vision meets robotics: the KITTI dataset," *Int. J. Robot. Res.* **32**(11), 1231–1237 (2013).
18. C. Sakaridis, D. Dai, and L. Van Gool, "Semantic foggy scene understanding with synthetic data," *Int. J. Comput. Vis.* **126**, 973–992 (2018).
19. G. Roy et al., "Physical model of snow precipitation interaction with a 3D LiDAR scanner," *Appl. Opt.* **59**, 7660–7669 (2020).
20. R. Heinzler et al., "CNN-based LiDAR point cloud de-noising in adverse weather," *IEEE Robot. Autom. Lett.* **5**, 2514–2521 (2020).
21. A. Kurup and J. Bos, "Winter adverse driving dataset (WADS): year three," *Proc. SPIE* **12115**, 121150H (2022).
22. K. Bansal et al., "Pointillism: accurate 3D bounding box estimation with multi-radars," in *Proc. 18th Conf. Embedded Netw. Sens. Syst.*, pp. 340–353 (2020).
23. Keweenaw Research Center, Michigan Technological University, "KRC snow data," http://mtukrc.org/met/weather_snow_data.htm (accessed 15 February 2022).
24. "The michigan tech winter adverse driving dataset (WADS)," <https://bitbucket.org/autonomytmu/wads> (2021).

Akhil M. Kurup received his PhD and MS degrees from Michigan Tech in 2022 and 2018, respectively. His research interests are in perception systems for robotics and autonomous vehicles. He is a member of SPIE, IEEE, and SAE, where he has authored scholarly contributions on using multimodal sensors and machine learning to further autonomous tasks such as perception in inclement weather, simultaneous localization and mapping and object detection and tracking.

Jeremy P. Bos is an associate professor of Electrical and Computer Engineering at Michigan Technological University. He received his PhD and BS degrees from Michigan Tech in 2012 and 2000, respectively, and his MS degree from Villanova University in 2003. He is a senior member of Optica, SPIE, and IEEE, and an author on over 100 scholarly contributions. His research interests are in the areas of imaging and light propagation in random media, signal processing, and sensor fusion.