# MSU-4S - The Michigan State University Four Seasons Dataset

Daniel Kent, Mohammed Alyaqoub, Xiaohu Lu, Hamed Khatounabadi, Kookjin Sung, Cole Scheller, Alexander Dalat, Xinwei Guo, Asma bin Thabit, Roberto Whitley, and Hayder Radha
Michigan State University
{kentdan3, luxiaohu, radha}@msu.edu

## Abstract

*Public datasets, such as KITTI, nuScenes, and Waymo, have played a key role in the research and development of autonomous vehicles and advanced driver assistance systems. However, many of these datasets fail to incorporate a full range of driving conditions; some datasets only contain clear-weather conditions, underrepresenting or entirely missing colder weather conditions such as snow or autumn scenes with bright colorful foliage. In this paper, we present the Michigan State University Four Seasons (MSU-4S) Dataset, which contains real-world collections of autonomous vehicle data from varied types of driving scenarios. These scenarios were recorded throughout a full range of seasons, and capture clear, rainy, snowy, and fall weather conditions, at varying times of day. MSU-4S contains more than 100,000 two- and three-dimensional frames for camera, lidar, and radar data, as well as Global Navigation Satellite System (GNSS), wheel speed, and steering data, all annotated with weather, time-of-day, and time-of-year. Our data includes cluttered scenes that have large numbers of vehicles and pedestrians; and it also captures industrial scenes, busy traffic thoroughfare with traffic lights and numerous signs, and scenes with dense foliage. While providing a diverse set of scenes, our data incorporate an important feature: virtually every scene and its corresponding lidar, camera, and radar frames were captured in four different seasons, enabling unparalleled object detection analysis and testing of the domain shift problem across weather conditions. In that context, we present detailed analyses for 3D and 2D object detection showing a strong domain shift effect among MSU-4S data segments collected across different conditions. MSU-4S will also enable advanced multimodal fusion research including different combinations of camera-lidar-radar fusion, which continues to be of strong interest for the computer vision, autonomous driving and ADAS development communities. The MSU-4S dataset is available online at https://egr.msu.edu/waves/msu4s.*

## 1. Introduction

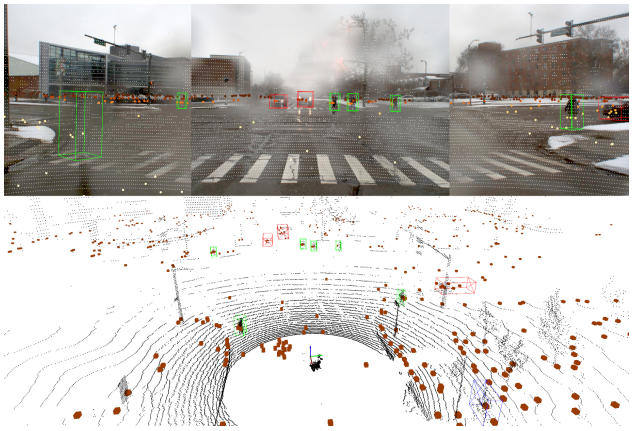

Figure 1. An example of fused output from stitched cameras, lidar (small points), and radar (larger points) from MSU-4S data, with vehicles (red), pedestrians (green), and a cyclist (blue) bounding box labels. Scene challenges include partially occluded camera sensor due to rain, making certain objects (such as the pedestrian on the left) difficult to detect in 2D, despite visibility to lidar and radar. These challenges highlight the importance of multimodal datasets in general, especially radar-inclusive datasets, that are captured under challenging conditions.

The availability of public datasets has been a key factor in the development of many recent innovations in the field of computer vision, and arguably one of the largest beneficiaries of these data are related to autonomous driving. However, in spite of the massive corpus of data of all kinds, there are still key shortfalls in the variety of publicly available datasets. Many commonly used datasets, such as the Karlsruhe Institute of Technology and Toyota Technical Institute (KITTI) dataset [9], only contain clear weather or cloudy condition in daylight scenarios, and do not contain labels for the weather conditions. While other datasets such as Waymo [21] contain more varied lighting and weather conditions, their primary collection locations preclude the

existence of colder weather conditions such as snow and sleet. Very few datasets exist that attempt to address multi-weather domain conditions at all, let alone provide labeled weather conditions combined with scenes collected in the same physical location for computing metrics on how specific weather conditions might affect object detection performance. Of these datasets, the Ithica365 dataset is arguably the current state of the art for multi-weather domain analysis, though it focuses mainly on amodal segmentation, and far less on the critical task of 3D object detection given the dataset's limited number of annotated lidar frames with labeled 3D bounding boxes [8].

Given the aforementioned shortfalls of existing popular autonomous driving datasets, it is clear that there is a need for a high quality dataset that contains multiple weather and lighting conditions for the same set of diverse physical locations. Having such data available would make benchmarking object detection algorithms against varied weather conditions significantly more meaningful, and would help advance robustness in new state-of-the-art perception frameworks. With the emergence of several domain adaptation techniques targeting challenging driving conditions, there is an acute need for datasets that can further validate the performance of these and other state-of-the-art algorithms [6, 11, 13, 14].

In this paper, we introduce the Michigan State University Four Seasons dataset, or MSU-4S, an autonomous vehicle dataset containing more than 100,000 frames. The salient attributes of our dataset are:

- MSU-4S was collected during clear, cloudy, rainy, and snowy weather conditions throughout contiguous four seasons of the year, with an associated label for each frame containing this information.
- Our dataset uniquely strikes a balance between diversity and consistency of scenes. MSU-4S includes cluttered scenes that have large numbers of vehicles and pedestrians; and it also captures industrial scenes, busy traffic thoroughfare with traffic lights and numerous signs, and scenes with dense foliage.
- While providing a diverse set of scenes, our data incorporate an important feature: virtually every scene and its corresponding lidar, camera, and radar frames were captured in four different seasons. This attribute enables unparalleled analyses and testing of the true domain shift problem across four seasons and weather conditions while stationary components, such as underlying roads/pavements, buildings, traffic signs, and other aspects of of the ambient background are consistent across these conditions, which is rare among existing datasets.
- In the context of having consistent scenes across four seasons, we present detailed analyses for 2D and 3D object detection showing a strong domain shift effect among MSU-4S data collected across different conditions.
- MSU-4S is highly multimodal, containing data from cameras, lidars, radars, Real-Time Kinematic (RTK) GNSS, Inertial Measurement Units (IMUs), wheel speed, steering angle, and other modalities. This diversity of data will enable advanced multimodal fusion research including different combinations of camera-lidar-radar fusion, which continues to be of strong interest for the autonomous driving development community. In particular, the inclusion of data captured by six radars makes MSU-4S a compelling dataset for single- and multimodal radar-inclusive fusion methods. A visual example of the multimodal nature of our dataset is shown in Figure 1.
- Trained personnel and engineers manually labeled (and continue to label) our data for training, validation, and testing. To ensure high quality labeling, a large number of instances from the 3D bounding box labeling of our lidar point cloud and 2D bounding box labeling of our camera images have been conducted independently of each other first, and then they are compared to ensure consistency and reliability of both sets of 2D and 3D labels.

The remainder of the paper is organized as follows. Section 2 will discuss related work, namely the popular autonomous vehicle datasets. Section 3 will discuss the MSU-4S dataset specifically, including the collection platform and route, calibration and synchronization, the data labels and labeling process, the format of the dataset, and dataset label statistics. Section 4 evaluates many variations of the MSU-4S domain shift scenarios in the context of 2D and 3D object detection, demonstrating the invaluable contribution of our dataset.

## 2. Related Work

While there are perhaps dozens of public autonomous vehicle related datasets, there are a select few that have seen significant use among researchers. We have compiled a selection of the quantitative and qualitative differences in Table 1.

**KITTI** [9] - The Karlsruhe Institute of Technology and Toyota Technical Institute (KITTI) dataset was the first public dataset introduced for autonomous vehicle algorithm development. In addition, KITTI also provides researchers an outlet to test 2D and 3D perception algorithms against unreleased segments of their dataset, and have their results posted publicly, fostering competition among interested researchers and developers for finding the top algorithms in tasks such as 2D and 3D detection and other perception algorithms.

While KITTI is the basis for a large body of perception research, it unsurprisingly has shortcomings that are being addressed by newer datasets. While the original KITTI's 15,000 samples of labeled data made it the largest dataset of its kind at that time, many datasets, including the more recent KITTI-360 dataset [1, 2, 15], have since surpassed

| | MSU-4S (ours) | KITTI / KITTI360 | nuScenes | Waymo | BDDD100k | Argoverse 2 | Ithica365 |
|---|---|---|---|---|---|---|---|
| **2D+3D Samples** | 100,000+ | 15,000 / 80,000 | 40,157 | 230,000 | 100,000 | 150,000 | 14,750 / 7,000 [1] |
| **Camera** | 3x, front facing, approx. 150 ° FOV | 2 stereo pairs, no FOV specified / 1 stereo pair + 2 side fisheye cameras, 360° FOV | 6x, positioned for 360° FOV | 5x front facing, approx. 205° FOV | 1x front 720p dashcam, no FOV specified | 2x front stereo + 7x ring cameras, 360° FOV | 4x, 60° FOV each |
| **Lidar** | 1x 64 line, 1x 32 line | 1x 64 line / 1x 64 line + 1x directional 1-line | 1x 32-line | 1x mid-range lidar, 4x short-range lidar | No | 2x 32 line | 1x 128-line, 2x 16-line |
| **Radar** | 6x Medium Range/ Long Range Radars with Radar Cross Section (RCS) & SNR Measurements | No | 5x Short/Long Range Radar | No | No | No | No |
| **Position** | RTK GNSS | RTK GNSS + INS | RTK GNSS + INS | Local Coordinates | GNSS Only | City-specific coordinates | Post-processed RTK GNSS |
| **CAN** | Wheel speed, steering angle, throttle, brake | No | Wheel speed and overall speed, steering angle and rate, throttle, signaling, gear, environment sensors | No | No | No | No |
| **Weather** | Clear, Cloudy, Snow, Rain, Autumn Foliage, Fallen Leaves | Clear/Cloudy Only | Clear, Cloudy, Rain | Clear, Rain | Clear, Cloudy, Snow, Rain | Various | Sunny, Rainy, Cloudy, Snow, Night |

[1]: Amodal labeled frames (road surface / semantic segmentation) with corresponding lidar data [8]

Table 1. Comparison of Popular Autonomous Vehicle Datasets with MSU-4S

its size and scope. Meanwhile, the original KITTI dataset, which is still used heavily by the research community, lacks imagery taken under varied lighting or weather conditions beyond clear or cloudy. Furthermore, other datasets have expanded on the amount and types of sensors and other data included in the dataset, including other sensor modalities such as radar; even the most recent KITTI-360 dataset does not include radar data.

**nuScenes** [5] - The nuScenes dataset was one of four datasets published in 2020, and was explicitly developed to improve the state of autonomous vehicle datasets beyond the standard that the KITTI dataset established. Specifically, nuScenes was designed to be more - more cameras (5, arranged in an omnidirectional configuration instead of front-facing stereo), radar added as an additional sensor modality; more collection locations (Singapore and Boston, MA, USA), additional vector map data, and of course more data frames: over 40,000 compared to the original KITTI dataset's 15,000. With these additions, nuScenes significantly increased the amount of data researchers had access to, offering the opportunity to improve existing algorithms and provide new data to develop novel multimodal fusion algorithms.

**BDD100K** [23] - The Berkley Deep Drive 100k dataset, also published in 2020, is a large dataset that contains 100,000 40-second 720p dashcam videos crowdsourced from drivers in New York City, San Francisco, Berkeley, Oakland, San Jose, and Mountain View. The self-stated goal of the BDD100k dataset is to study domain transfer problems in autonomous vehicle sensor processing. To support this goal, BDD100k contains scenes with varied weather conditions including snow and rain, and were collected at various times during the day and night. Despite both the quality and quantity of data, the BDD100k dataset lacks 3D sensor modalities, which limits its effectiveness for developing and testing 3D object detection algorithms, which are a core component of any autonomous driving platform.

**Waymo** [21] - First released in 2019, the Waymo dataset consists of data collected by their automated vehicles operating mainly in San Francisco, California; Mountain View, California, and Phoenix, Arizona. At 200,000 total frames, Waymo is one of the largest publicly available multimodal autonomous driving datasets. However, due to the locations represented in the dataset, winter weather conditions are not present in the Waymo dataset.

While KITTI, nuScenes, BDD100K, and Waymo are arguably some of the most widely cited autonomous vehicle datasets, there are dozens of other datasets such as **Cityscapes** [7], **Apolloscape** [12], **Argoverse** [21], and **A2D2** [10] that have been released to the public and used in research and development. There has also been a specific interest in datasets focused on or otherwise containing scenery in challenging conditions. For example, the **Canadian Adverse Driving Conditions Dataset (CADS)** contains 7500 labeled lidar and 15000 labeled camera frames, with a specific emphasis on snow [17]. **4seasons** is a multi-weather dataset specifically focused on testing localization algorithms in multiple seasons [22]. Arguably the current state-of-the-art dataset for testing multi-weather perception is **Ithica365**, a camera-lidar dataset of a 15 kilometer route taken over multiple weather and lighting conditions, with 14,750 labeled frames containing amodal road masks and 7000 frames with amodal instance masks [8]. However, this dataset's applicability to 3D object detection is far more limited than other datasets given its relatively small number of labeled 3D bounding boxes.

## 3. The Michigan State University Four Seasons (MSU-4S) Dataset

To augment the current publicly available data for autonomous vehicle research and development, we are introducing the Michigan State University Four Seasons (MSU-4S) Dataset, a collection of data captured during varied times of year and under different weather conditions. These data are taken using fixed routes that cover a highly diverse

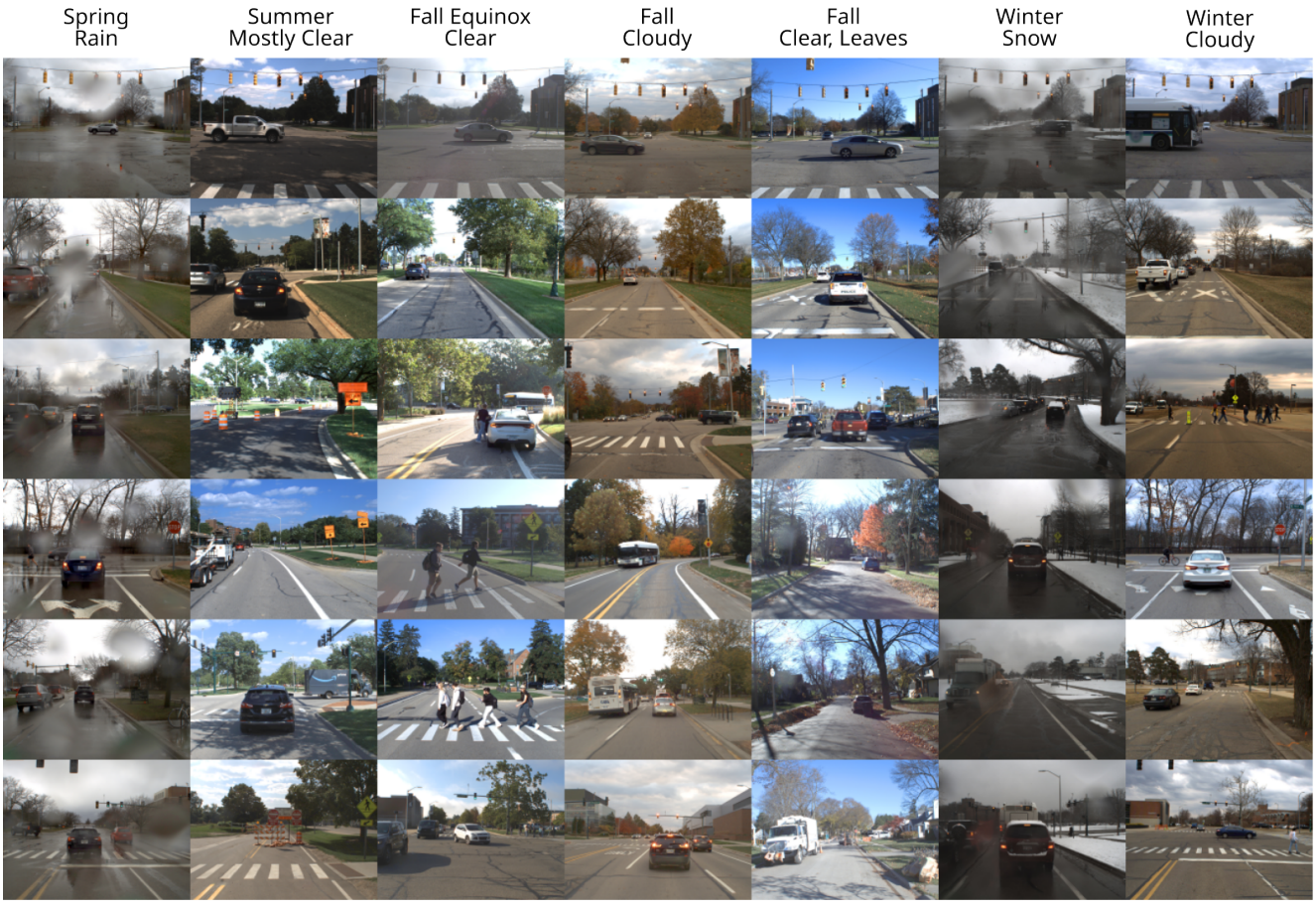


Figure 2. Selection of scenes from the Michigan State University Four Seasons Dataset. Top two rows show samples taken from the same physical location over four seasons under different weather conditions.

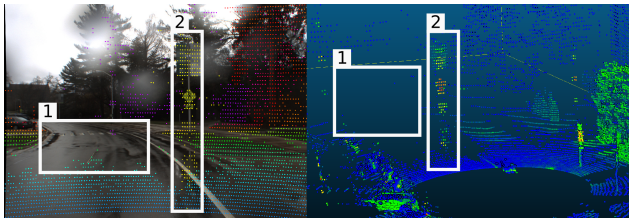


Figure 3. Lidar artifacts present in lidar during wet weather conditions, including (1) dead zones and object (2) reflections.

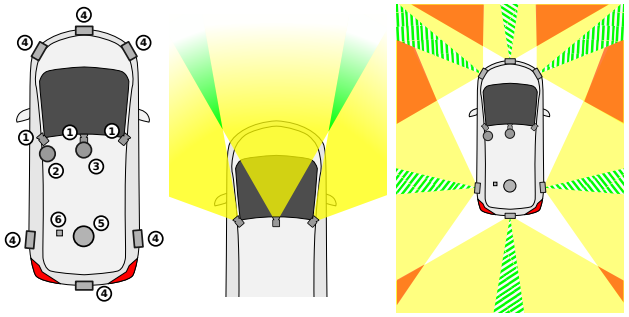

Figure 4. Left: Sensor layout of data collection vehicle platform. Sensors cameras (1), 64-line lidar (2), 32-line lidar (3), radar (4), GNSS antenna (5), and IMU (6). Middle: Camera FOV overlap. Right: Short/Long Range Radar FOV

set of scenes, enabling testing for robustness across diverse scenarios. This section will explain the data collection platform, the overall collection route, the 2D and 3D labeling process, and the final data format that will be available to the broader computer vision and autonomous vehicle development community.

The collection platform used to collect our data for the MSU-4S is based on a modified 2017 Chevrolet Bolt EV. The sensors on our platform that were used for MSU-4S are:

- Three FLIR Blackfly BFS-PGE-31S4C-C cameras with Sony IMX265 sensors running at 2048x1536 resolution at their maximum 34 frames per second, each with a C-mount lens with a horizontal field of view of 60 degrees,

angled to capture approximately 150 degrees of front-facing imagery
- One Ouster OS-1 64-line 3D lidar using 10Hz data rate,
- One Velodyne VLP-32C 3D lidar using 10Hz data rate,
- Six Continental ARS430 medium-long range radars, equipped to return individual detections including Radar Cross Section (RCS) and Signal-to Noise Ratio (SNR),

Our camera configuration was chosen to maximize the effective front-facing field of view without using fisheye or other high-field of view (FOV) lenses. Using such a lens would severely distort the imagery, which would effectively reduce the effective resolution at the edges of the camera's FOV. Our lidar configuration was chosen to provide data for training algorithms on low density lidar, high density lidar, or fused data from both lidars.

The compute system on our platform is an Intel Xeon dual-socket server with 44 total cores, giving it plenty of compute capability for ingesting data including minimal on-vehicle processing such as image debayering. This system is connected to most of the sensors through shielded Category 6a networking cables, connected to the compute system through a 10Gb link aggregation switch. The compute system is also connected to the vehicle's controller area network (CAN) busses. The IMU is the only sensor connected through USB. In total, we estimate the overall sensor suite to produce approximately 5 gigabits of raw data per second. Each data frame has an associated timestamp in Unix format with nanosecond-level precision. This timestamp is based on the system clock, which is first synchronized to GPS time, then distributed among supported sensors using Precision Time Protocol (PTP). For sensors that do not support PTP, we use a software timestamp recorded when the packet was collected as the basis for synchronization.

The collection platform uses the Robot Operating System (ROS) [18] middleware to interface with all the sensors, and provide time synchronization and data storage functions. Data is stored using the native ROS storage format, known as *rosbag*, which is a serialized and timestamped format, with added quality of life features such as optional data compression and encryption. This process ensures that data integrity, including both sensor-specific timestamp and data receive timestamp, is maintained, which we analyze in Subsection 3.2.

### 3.1. Collection Routes

Our overall data collection zone is split into five zones based on GNSS coordinates, which include:

**`central_hub`** - A densely populated area with numerous pedestrians, cyclists, cars, buses, and signs. In addition to being a viable representative of a cluttered environment, this area also contains some unusual road types, including a long one-way loop and a non-perpendicular four way stop intersection, all featuring varied overhead foliage.

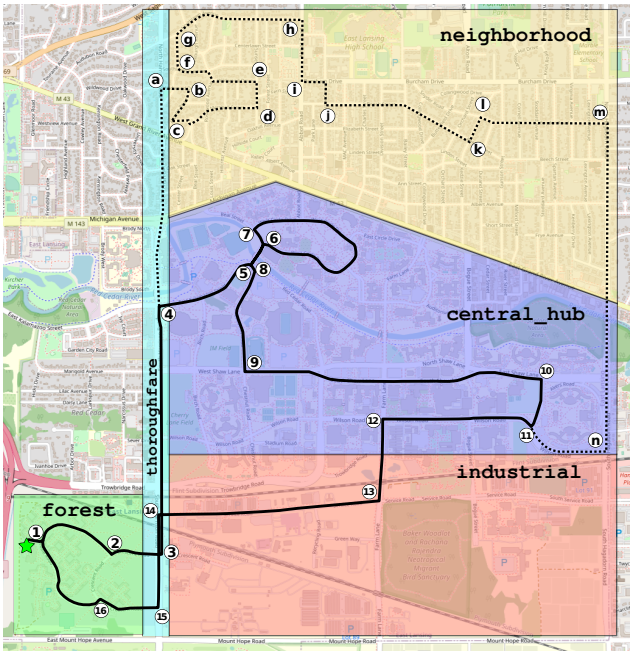


Figure 5. General routes for data collection. Solid line indicates standard route; dashed line indicates auxilliary route that branches off at point 4 and rejoins at point 11. Underlying street map sourced from OpenStreetMap [16].

**`thoroughfare`** - A busy north-south road containing several traffic lights, numerous signs, and two railroad crossings.

**`industrial`** - An area surrounding a power plant; contains more industrial buildings and parking lots. While there are fewer pedestrians, there are more commercial vehicles in this area.

**`forest`** - An area with dense foliage, minimal traffic, and some apartment buildings.

**`neighborhood`** - A residential zone containing narrow streets with parked cars.

Our typical route for MSU-4S consists of a 11.2km loop through the forest, thoroughfare, central hub, and industrial zones, depicted as the solid line in Figure 5. A subset of the overall data consists of a longer, 14.8km alternate route through the neighborhood to capture more unusual scenarios such as large piles of leaves and parked cars at the sides of narrow roads; this route is depicted as a dashed line in Figure 5.

We believe that the diverse attributes of the above locations make our dataset an invaluable asset for training and testing the robustness of autonomous driving perception algorithms. Because these same diverse set of locations were traversed by our vehicle over four different seasons with different weather conditions, MSU-4S provides a rare dataset for analyses and testing of the domain shift problem, and for developing robust domain adaptation algorithms. The most

similar existing dataset, Ithica365, is constructed similarly, but does not contain radar and CAN data.

It is important to note that some of the dataset segments have slightly different routes to account for construction detours, operator error, or other variations when collecting data. As GNSS tracks are available on a per-frame basis, we believe that there should be little to no problem incorporating variances in the route when using MSU-4S.

Each frame contains both the specific GNSS coordinate, as well as the zone. These labels are intended to help users filter the dataset for their intended use case, and to perform cross-domain analysis on algorithms trained in other scenarios. For instance, the `forest` area contains far fewer objects, making it an ideal subset to test for false positives, especially for algorithms trained on urban data.

### 3.2. Calibration and Synchronization

For instrinsic camera calibration, we utilized the classical Perspective n-Point (PNP), using OpenCV's chessboard corner detection methods for automatically detecting 2D-3D keypoints [4]. We used professionally precision manufactured chessboards, ensuring there are no errors from either incorrect scaling or non-flat boards. These calibration parameters are made available alongside each frame in the dataset.

For extrinsic sensor calibration, we first started by defining a common reference point for the vehicle, known as the *base link*. For ground vehicle robots with Ackermann steering geometry, this point is usually defined as the center of the rear axle. From this point, we can easily measure the distance to the ground, which we can then use as a reference for all sensors. Then, we precisely measured the position of the Velodyne VLP-32 lidar (chosen due to its centralized location) relative to the base link, confirmed using both the manufacturer specifications of the vehicle as well as a visualization of the point cloud. For the Ouster lidar, we hand measured the offset between it and the Velodyne lidar, and then used software tools to first hand adjust the two pointclouds to maximize overlap, then used software tools to compute a registration using Normal Distribution Transform (NDT) [3]. For the cameras, we similarly hand measured the offset and rotation for each camera, then used automated methods to fine tune the final transformation. This transformation was verified visually during the early stages of the 3D labeling process. A visualization of the lidar-camera projection using this calibration is shown in Figure 6.

Each frame in our dataset is collected from data that are temporally close to the Velodyne VLP-32 lidar, again selected as the most central sensor on the vehicle. Our vehicle utilizes GNSS-backed PTP on supported sensors to ensure on-device timestamps are as accurate as possible; PTP is enabled on the Velodyne VLP-32, Ouster OS-1, and the

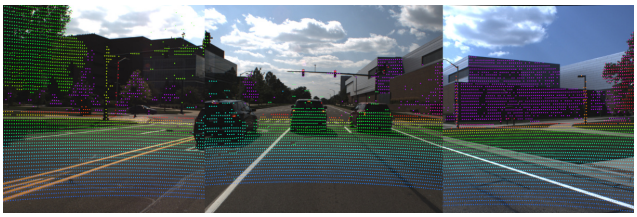

Figure 6. Post-calibration camera-lidar projection results

FLIR cameras. For sensors that do not support PTP, the ROS header and/or packet receive timestamp is used as the data synchronization source.

To verify that the difference device timestamp and packet timestamp is not significantly different, we compared the difference between these timestamps for all Velodyne VLP-32 packets in our dataset. We found that the average absolute difference was approximately 63 nanoseconds (variance $1.49 \times 10^{-14}$), showing that the data transmission time provides a negligible difference in timestamp, and thus using packet receipt timestamps is sufficient for data sources that do not support PTP. We also analyzed the average absolute timestamp difference between the Velodyne VLP-32 and the top central camera. Overall, we found the vast majority of data fell within +/-16ms, with an average absolute timing difference of 7.6ms, with a variance of $8.16 \times 10^{-5}$. A histogram of this timing data is shown in Figure 7.

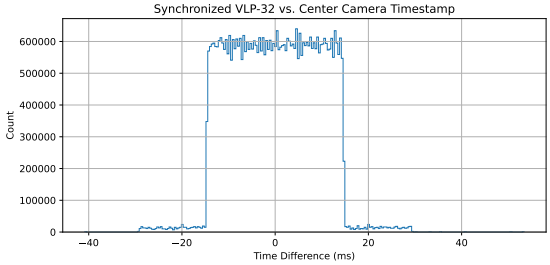


Figure 7. Synchronization timing between VLP-32 and center camera

### 3.3. Labeling Process

The first stage of our data pre-labeling process is the extraction of the dataset into individual, time-synchronized frames to minimize cross-frame temporal error. Camera imagery and lidar point-cloud frames are extracted to separate files, with the file name prepended with the integer nanosecond timestamp of the frame, and the suffix corresponding to the sensor source. For all other sources, as well as intrinsic and extrinsic calibration information, season, and weather condition, a YAML-formatted file with the suffix `_misc.yaml` is created to contain these data. Once data are extracted, the start and end of the dataset are

trimmed to remove irrelevant data collected during data collection startup and shutdown. Finally, the camera data is processed to redact facial and license plate imagery using privacy-preserving tools such as EgoBlur [19].

Our labeling process focuses on providing highly accurate 3D and 2D bounding boxes within our lidar and camera datasets, respectively. To ensure high accuracy of our labeling process, 3D labeling of a large instances of the lidar point-cloud frames and 2D labeling of the corresponding instances of camera images are first performed manually and independently of each other. Both labeling pipelines are conducted by experienced personnel who have been trained with sufficiently large test examples prior to performing the actual labeling. For both labeling processes, 2D and 3D, two independent advanced labeling tools are employed. Meanwhile, for the 3D labeling process, the software tool used provides the views in both the lidat point-cloud and the corresponding camera image with projected 3D boxes within the 2D camera images.

For our 3D point-cloud lidar data, we provide labels for multiple object categories including but not limited to vehicles, pedestrians, and cyclists. Vehicles include all cars, trucks, buses, and construction vehicles. For our camera data, more expansive 2D labels are provided in order to provide additional contextual information that may not be obvious in the lidar frame, such as the content and context of road signs, specific vehicle types, and traffic control devices such as traffic lights. As such, we both expanded the number of classes compared with the 3D labels, as well as split labels such as vehicles and signs into separate labels. Signs are split into nine subcategories matching the legal definition of sign types specified by the state of Michigan, eight types of which are captured in the MSU-4S dataset. The shapes of these signs are shown in Figure 8. As far as we are aware, the MSU-4S dataset is the first to classify signs using their legal definition. While we have labeled for the entirety of the MSU-4S dataset, we have reserved a portion of the labels for future dataset development, similar to other datasets like KITTI, Waymo, and nuScenes.

### 3.4. Data Format and Description

MSU-4S is formatted in an attempt to ease both casual use as well as large scale automated processing of its data. For camera imagery, we provide the compressed imagery in their native JPEG format to avoid further reencoding loss. Lidar data is provided in PCD format, for which several utilities and libraries exist for reading and writing of these data. Labeled frame data are available in plaintext format. All other collected data are provided in the accompanying `_misc.yaml` file.

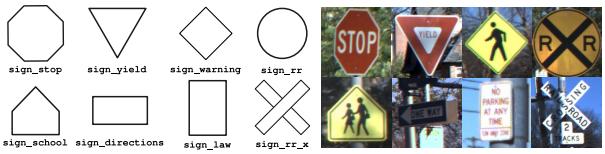


Figure 8. Left: Sign classifications used for MSU-4S labeling. Right: Examples for all sign classes that appear in MSU-4S.

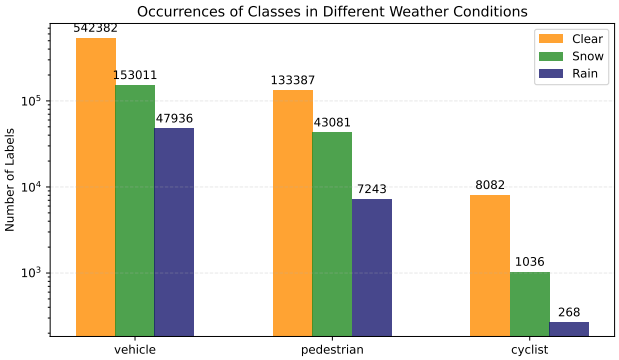


Figure 9. Count of 3D labeled objects per class, separated by labeled weather conditions, using a trained object detector.

### 3.5. Object Statistics

As mentioned earlier, our dataset consists of more than 100,0000 frames of 2D camera images, 3D lidar point-clouds, and corresponding radar data. Given that our labeling process is currently ongoing, here we provide a conservative estimate for the total objects included in our dataset. To estimate the number of objects captured in over 100,000 frames of MSU-4S data, we used the the PV-RCNN++ [20] detector pretrained on Waymo data. We also used a trained PV-RCNN++ model using our already labelled 3D dataset. In our experience, the Waymo data pretrained model produces a more conservative object count with fewer false positives compared to other pretrained models or models trained specifically on our data. We subsequently used the Waymo pretrained PV-RCNN++ detector. This detector estimated a total of 936,426 objects across all frames, including 743,329 vehicles, 183,711 pedestrians, and 9,386 cyclists. Figure 9 shows a breakdown of each object detection based on labeled weather condition. When comparing the relative ratio of object types detected using the Waymo pretrained model versus our hand-annotated labels, we found that the relative ratio of labels were within 5% for each category, and for cyclists the difference was less than 1 percentage point, which provides stronger confidence in our estimates.

| Test Data → Training Data ↓ | Clear IoU=0.5 (0.7) | Rain IoU=0.5 (0.7) | Snow IoU=0.5 (0.7) |
|---|---|---|---|
| Clear | 67.99 (54.22) | -5.55 (-10.72) | -12.49 (-18.45) |
| Rain | -6.50 (-3.93) | 70.09 (50.55) | -7.29 (-5.16) |
| Snow | -3.86 (-3.24) | -0.21 (-0.92) | 68.15 (54.22) |

Table 2. Average Precision (AP) 3D object detection performance and the changes in its values due to the domain shift problem under different weather conditions using the PV-RCNN++ 3D object detector. All results are based on training and testing using segments of our MSU-4S manually labelled 3D lidar point-cloud with 3D bounding boxes for vehicles and an IoU=0.5 and IoU=0.7.

| Test Data → Training Data ↓ | Spring Rain | Fall Clear | Summer Clear | Winter Snow |
|---|---|---|---|---|
| Spring Rain | 40.2 | -13.30 | -11.10 | -10.40 |
| Fall Clear | -6.20 | 39.4 | -7.30 | -6.80 |
| Summer Clear | -5.60 | -9.20 | 40.3 | -13.40 |
| Winter Snow | -9.40 | -19.30 | -15.30 | 47.3 |

Table 3. Mean Average Precision (mAP) 2D object detection performance and the changes in its values due to the domain shift problem under different weather conditions using the YOLOv5 2D object detector. All results are based on training and testing using segments of our MSU-4S manually labelled 2D camera data with 2D bounding boxes for the detection of all 2D classes covered by our dataset.

## 4. Domain Shift Analysis

One of our key objectives for introducing MSU-4S is to provide the 3D and 2D object detection research communities with a new and compelling dataset for the analysis, training, and testing of algorithms that focus on the domain shift problem across different seasons and under different weather conditions. Some effects are immediately obvious, such as wet areas causing lidar reflections and dead zones as shown in Figure 3, while others may not be as immediately obvious or otherwise quantifiable. Therefore, in this section, we demonstrate that MSU-4S provides arguably unprecedented data with a strong domain shift attributes across different seasons and weather conditions for both lidar-based 3D and camera-based 2D object detection.

In that context, we performed multiple experiments by employing both 3D and 2D popular object detectors trained using our currently-available labeled 3D and 2D data, respectively. For 3D object detection, we trained PV-RCNN++ [20] using three different segments of our labeled datasets representing *clear*, *rainy*, and *snowy* conditions. These three 3D object detector models were then applied to corresponding three different test segments of our labeled data. These results are shown in Table 2. It is clear from the table that MSU-4S exhibits a very strong domain shift across different weather conditions. In particular, the 3D object detector model trained using clear-weather data experienced significant drop in performance when tested on rainy and snowy data segments. Another interesting observation is that a 3D object detector model trained on challenging weather condition seems to be more robust when tested under less challenging conditions. For instance, the *Clear* weather model suffers a more than 18 point drop in AP performance when tested under snowy conditions, while a 3D detector trained using the snow data segment suffers somewhere between 3-4 points reduction in AP when tested on the more favorable clear weather conditions. We can also observe that the domain shift between rain and snow is clearly smaller than the domain shift between clear and snowy conditions.

One of the main attributes of our dataset is the ability to test the domain shift problem due to multi-domain shifts. For example, we can assess the domain shift between clear summer weather and fall weather with vibrant autumn color foliage in the background. Hence, for 2D object detection, we trained YOLOv5 using four different data segments from our manually-labelled 2D camera dataset: *Spring Rain*, *Fall Clear*, *Summer Clear*, and *Winter Snow* conditions. These results are shown in Table 3. It is clear from the table that there is a significant drop in mean Average Precision (mAP) among different multi-domain scenarios. In particular, the data clearly shows the importance of evaluating the domain shift problem due to the fall season. Despite the similarity in weather and lighting conditions between the tested summer and fall segments, we determined there was a significant domain shift in our data.

## 5. Conclusions

In this paper, we presented the Michigan State University Four Seasons (MSU-4S) Dataset, which contains real-world collections of autonomous vehicle data. MSU-4S was recorded through a full range of seasons, and capture clear, rainy, snowy, and fall weather conditions, at varying times of day. Our data contains more than 100,000 frames of camera, lidar, and radar data, as well as GNSS, and CAN data, all annotated with weather, time-of-day, and time-of-year. Our data includes cluttered scenes that have large numbers of vehicles and pedestrians, captured in different locations. The diverse scenery also incorporates an important feature: virtually every scene and its corresponding lidar, camera, and radar frames were captured in four different seasons, enabling unparalleled object detection analysis and testing of the domain shift problem across weather conditions while keeping other aspects of the ambient background consistent. We also presented detailed analyses for 3D and 2D object detection showing a strong domain shift effect among data segments collected across different conditions. MSU-4S is available for download online at https://egr.msu.edu/waves/msu4s.

## References


[1] Hassan Abu Alhaija, Siva Karthik Mustikovela, Lars Mescheder, Andreas Geiger, and Carsten Rother. Augmented reality meets computer vision: Efficient data generation for urban driving scenes. *International Journal of Computer Vision*, 126:961–972, 2018. 2

[2] Hassan Abu Alhaija, Siva Karthik Mustikovela, Lars Mescheder, Andreas Geiger, and Carsten Rother. Augmented reality meets deep learning for car instance segmentation in urban scenes. In *British machine vision conference*, 2017. 2

[3] Peter Biber and Wolfgang Straßer. The normal distributions transform: A new approach to laser scan matching. In *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)(Cat. No. 03CH37453)*, pages 2743–2748. IEEE, 2003. 6

[4] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000. 6

[5] 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11621–11631, 2020. 3

[6] Yuhua Chen, Wen Li, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Domain adaptive faster r-cnn for object detection in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3339–3348, 2018. 2

[7] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Scharwächter, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset. In *CVPR Workshop on the Future of Datasets in Vision*. sn, 2015. 3

[8] Carlos A. Diaz-Ruiz, Youya Xia, Yurong You, Jose Nino, Junan Chen, Josephine Monica, Xiangyu Chen, Katie Luo, Yan Wang, Marc Emond, Wei-Lun Chao, Bharath Hariharan, Kilian Q. Weinberger, and Mark Campbell. Ithaca365: Dataset and driving perception under repeated and challenging weather conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21383–21392, 2022. 2, 3

[9] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The KITTI dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, 2013. 1, 2

[10] Jakob Geyer, Yohannes Kassahun, Mentar Mahmudi, Xavier Ricou, Rupesh Durgesh, Andrew S Chung, Lorenz Hauswald, Viet Hoang Pham, Maximilian Mühlegg, Sebastian Dorn, et al. A2D2: Audi autonomous driving dataset. *arXiv preprint arXiv:2004.06320*, 2020. 3

[11] Mazin Hnewa and Hayder Radha. Integrated multiscale domain adaptive yolo. *IEEE Transactions on Image Processing*, 32:1857–1867, 2023. 2

[12] Xinyu Huang, Xinjing Cheng, Qichuan Geng, Binbin Cao, Dingfu Zhou, Peng Wang, Yuanqing Lin, and Ruigang Yang. The apolloscape dataset for autonomous driving. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 954–960, 2018. 3

[13] Sanket Kalwar, Dhruv Patel, Aakash Aanegola, Krishna Reddy Konda, Sourav Garg, and K Madhava Krishna. Gdip: gated differentiable image processing for object detection in adverse conditions. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 7083–7089. IEEE, 2023. 2

[14] Jinlong Li, Runsheng Xu, Jin Ma, Qin Zou, Jiaqi Ma, and Hongkai Yu. Domain adaptive object detection for autonomous driving under foggy weather. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 612–622, 2023. 2

[15] Yiyi Liao, Jun Xie, and Andreas Geiger. Kitti-360: A novel dataset and benchmarks for urban scene understanding in 2d and 3d. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3292–3310, 2022. 2

[16] OpenStreetMap contributors. Planet dump retrieved from https://planet.osm.org . https://www.openstreetmap.org, 2017. 5

[17] 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*, 40 (4-5):681–690, 2021. 3

[18] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, Andrew Y Ng, et al. Ros: an open-source robot operating system. In *ICRA workshop on open source software*, page 5. Kobe, Japan, 2009. 5

[19] Nikhil Raina, Guruprasad Somasundaram, Kang Zheng, Steve Saarinen, Jeff Messiner, Mark Schwesinger, Luis Pesqueira, Ishita Prasad, Edward Miller, Prince Gupta, et al. Egoblur: Responsible innovation in aria. *arXiv preprint arXiv:2308.13093*, 2023. 7

[20] Shaoshuai Shi, Li Jiang, Jiajun Deng, Zhe Wang, Chaoxu Guo, Jianping Shi, Xiaogang Wang, and Hongsheng Li. Pv-rcnn++: Point-voxel feature set abstraction with local vector representation for 3d object detection. *International Journal of Computer Vision*, 131(2):531–551, 2023. 7, 8

[21] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2446–2454, 2020. 1, 3

[22] 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 *Pattern Recognition: 42nd DAGM German Conference, DAGM GCPR 2020, Tübingen, Germany, September 28–October 1, 2020, Proceedings 42*, pages 404–417. Springer, 2021. 3

[23] 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2636–2645, 2020. 3