Data Description

Data Number (will be filled by the Commons Center)	DPPSCdbp_2025-07en
Title	Training Data for Measuring Truck Traffic Volume on the Joban Expressway Using Satellite Imagery and Satellite-Based Estimation of Truck Traffic Volume
Creator	Hisahiro Naito Meiyi Zhuang
Contact information	naito@dpipe.tsukuba.ac.jp
Summary (Subject and/or Methodology for generarion)	There are five folders: 1_ground_truth_video: contains the videos we collected. 2_sentinel2_dataset: contains Sentinel-2 images masked by road AOIs in Ibaraki Prefecture. 3_annotated_images: contains the training dataset used for the YOLO keypoint detection model, along with annotation files. 4_prediction_images: contains the predicted images and CSV files generated by the YOLO model trained in step 3. 5_traffic_data: contains the constructed datasets (tables) and corresponding visualization figures. Here is the detailed information for 1_ground_truth_video folder: This folder contains ground truth videos collected in Tsukuba, Ibaraki, Japan. We collected ground truth videos at two locations. One is the Kise region (with four bridge candidates), with GPS coordinates (latitude = 36.0843909, longitude = 140.1554314). The other is the Heights bridge, with GPS coordinates (latitude = 36.1059684, longitude = 140.1764893). In the Kise region, there are four bridge candidates, labeled as 1, 2, 3, and 4. The videos are named according to the bridge ID (1-4), followed by “kise”. In the video filenames, “south” indicates that the camera is facing Tokyo, while “north” indicates that it is facing Mito. Here is the detailed information for 2_sentinel2_dataset folder: This folder contains Sentinel-2 images covering expressway segments in Japan,

	organized by area of interest (AOI). We first extract the Japanese expressway network and divide it into 1 km × 1 km AOIs, each with a 50-meter buffer. Each AOI is assigned a unique ID. Then, we use the openEO API (We accessed openEO via the endpoint https://openeo.dataspace.copernicus.eu/openeo/1.2/) to download road-buffered Sentinel-2 images for Ibaraki Prefecture. We download the data using the following python code: datacube = (connection .load_collection("SENTINEL2_L2A", temporal_extent=TIME_RANGE, bands=BANDS, max_cloud_cover=CLOUD_MAX) .filter_spatial(aoi_geojson)) datacube.download(NC_PATH) In total, there are 388 AOIs in Ibaraki, covering the period from 2016 to 2024. Here is the detailed information for 3_annotated_images folder: This folder contains two annotated datasets in different formats but with identical content: keypoint_sentinel2.v2i.coco: COCO format, compatible with both macOS and Windows systems. Our workflow: From the downloaded images, we selected AOIs corresponding to high-bridge regions where ground truth videos were collected. We used Roboflow for annotation. We annotated images covering the entire calendar year of 2024. The annotated dataset is organized into train, test, and valid folders, each containing image files and corresponding JSON annotation files (e.g., COCO JSON). To reproduce the annotation and training setup: During the upload process, Roboflow automatically matches each image with its corresponding annotation file based on the file names. Once uploaded, the dataset can be directly used for training a YOLO keypoint detection model within the Roboflow platform. In detail, to replicate, users should do: Log in to Roboflow and create a new project. Select Keypoint Detection as the model type. Define one class named “rainbow”, and create three keypoints labeled “red”, “green”, and “blue”. Upload the unzipped dataset folder (either COCO). Ensure that image and annotation files are uploaded together. Then click "Save and Continue." Click "Train Model" in the upper-right corner. Select "Custom Training" and choose YOLOv11 (Fast) as the model. Proceed to start training.

	The model training process typically takes approximately 15 minutes. Here is the detailed information for 4_prediction_images folder: We selected AOIs in southern and northern Ibaraki Prefecture and generated predictions for the period from 2016 to 2024. We used the YOLO keypoint detection model for prediction, with parameters and weights trained in step 3. This folder contains three CSV files: per_day_box_vs_kpts_filtered: contains time-series information for all images within each AOI. pred_keypoint_long_all: contains all prediction results, including keypoint locations and confidence scores for all images within each AOI. predictions_all_batches: records batch-level information for each AOI, including the number of detected “rainbows”. This folder also contains several subfolders. Each subfolder is named by its AOI ID. Within each AOI subfolder, there are multiple sub-subfolders named as {AOI_ID}_b01, {AOI_ID}_b02, etc. Here, “b” denotes “batch”. Due to the large number of images, predictions are processed in batches, with each batch containing 128 images. The indices 01, 02, 03 indicate the first, second, and third batches, respectively. Inside each batch folder, there is a labels folder that records the detected “rainbow” locations for each image in that batch. Each AOI subfolder also contains two CSV files, which correspond to the main CSV files at the folder level. The only difference is that these CSV files include information specific to the corresponding batch. Here is the detailed information for 5_traffic_data folder: This folder contains four CSV files: per_day_aoi: dataset at the AOI-date level. per_day: dataset aggregated at the date level, summing detection counts across all AOIs. This folder also contains four figures: by_month: monthly aggregation based on the date-level dataset. by_year: yearly aggregation based on the date-level dataset.

Update History (Versions, File names, Dates)	Created at April 25, 2026
Data formats	The video files are in MP4 format. The Sentinel-2 files are in TIFF format.and png format The annotated images are in COCO format. The prediction results are provided as CSV and JPEG files. The traffic data are provided as CSV and png format.
Data sizes	1.9 giga in compressed format(zip)
Things to be aware of when using the data (e.g., required metadata)	
Related reports and/or papers	
Other remarks	