

AI Image Generation for Safety Evaluation of Automated Driving Systems

Nobutoshi Ozaki ¹⁾ Tetsuya Niikuni ¹⁾

*1) National Traffic Safety and Environment Laboratory (NTSEL),
7-42-27, Jindaijihigashimachi, Chofu, Tokyo 182-0012, Japan (E-mail: n-ozaki@ntsel.go.jp)*

KEY WORDS: Software and its underlying, Test framework, AI and data analysis, Generative AI [E3]

With the recent advancement of automated driving and advanced driver-assistance systems, evaluating the safety of machine-learning-based in-vehicle AI has become an important issue. In 2025, the UNECE World Forum for Harmonization of Vehicle Regulations established the Informal Working Group on Artificial Intelligence to consider how such AI should be treated under existing vehicle regulations. The group has not yet developed new vehicle regulations; instead, it is organizing safety-evaluation concepts and methods that consider AI-specific characteristics such as data quality and predictability. Evaluation must confirm required performance under specified conditions while identifying conditions that reduce performance and the AI's applicable range. However, real-world driving data depend on situations actually encountered, making it difficult to intentionally set object positions, appearances, and other conditions for a particular evaluation purpose. A document presented to the group noted the difficulty of securing sufficient test coverage and described synthetic data, including simulation data, as a means of supplementing real-world data.

As a first step toward evaluating and validating object-detection AI, this study developed a method for generating road images from the viewpoint of a monocular camera. Individual-image object detection was selected to separate the effects of object position and appearance from those of time-series processing and sensor fusion. Monocular-camera perception also remained in use for advanced driver assistance in 2026. Evaluation images require both realistic appearance and ground-truth labels indicating each object's class and position. Although methods such as MagicDrive generate images from road maps and three-dimensional bounding boxes, individually specifying object appearance or state is difficult. Therefore, this study used Seg2Any to generate images from short object captions and pixel-level instance masks created from three-dimensional bounding boxes. Seg2Any was fine-tuned using ZOD Frames, a subset of the Zenseact Open Dataset containing 100,000 road images collected in 14 European countries. Each road object has a three-dimensional bounding box with information about its class and state. Areas showing the ego vehicle were excluded because they were unrelated to road-object detection. The training-data construction process is shown in Fig. 1. Each bounding box was projected onto the image coordinate system using camera parameters to create a rough instance mask. HQ-SAM then produced a pixel-level mask following the object shape. A caption was created from the recorded class and state information and the bounding box's position and orientation relative to the camera. For pedestrians, RTMPose estimated skeletal keypoints used to classify poses, including standing and walking, for addition to the captions. The masks and corresponding bounding boxes and object information were stored in a database.

During image generation, three-dimensional bounding boxes were manually placed according to the intended traffic situation, and each object's class and, when necessary, state were specified. The database item closest in dimensions, orientation, and object information was automatically selected, and its mask was transformed to match the bounding box's image position and size. Captions were produced as in training. The fine-tuned Seg2Any generated a reasonable road image consistent with a specified traffic situation, including one considered rare in reality. It also generated two pedestrians ahead of the ego vehicle, with the left pedestrian standing and the right pedestrian walking, as specified in Fig. 2. These examples show the possibility of generating reasonable images from specified object conditions. Future work will quantitatively evaluate resemblance to real images and develop a method for evaluating object-detection AI using images generated while varying object conditions.

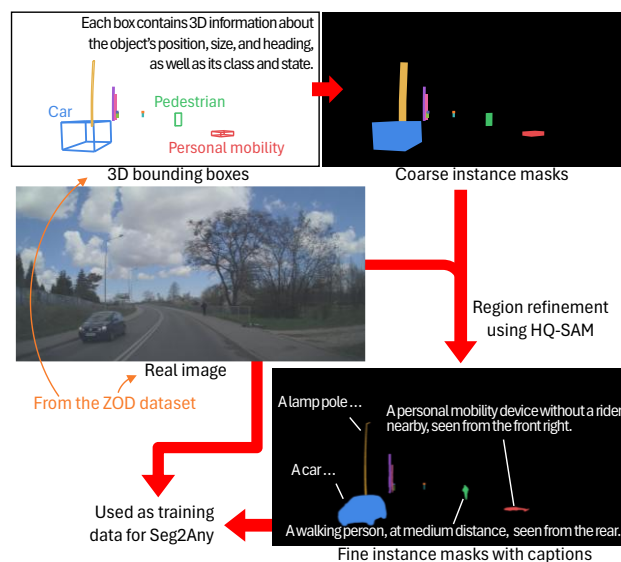


Fig. 1 Training data construction method.

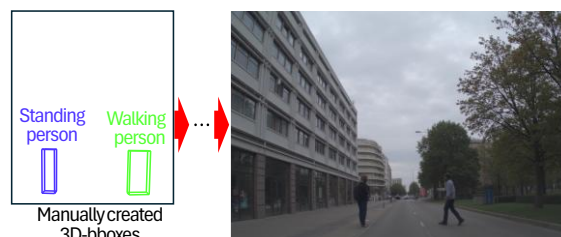


Fig. 2 Generated image with specified pedestrian states.