東京大学　関本研究室 / Sekimoto Lab. IIS, the University of Tokyo.

Vison-based Traffic Sign Detection and Localization in Tokyo Metropolitan Area

Zhehui Yang, Hiroya Maeda, Yoshihide SEKIMOTO

Background

Being a world-class metropolis, Tokyo has a significantly complex transportation system. A system to automatic detect and localize traffic signs is needed to be developed for the digital management of transportation systems and automated mapping platforms. Owing to the rapid development of computer vision algorithms in recent years, more accurate and efficient ways to achieve these tasks with minimal labor costs have been proposed. In this study, we propose a vision-based pipeline to detect, classify, and locate traffic signs in different road scenarios in Tokyo metropolitan area. Detection was performed using YOLOv4. Classification was performed by comparing the VGG19, ResNet101, and Swin-B networks. Because collecting driving data with ground truth distance data is a challenging task for supervised learning algorithms, we originally tested a self-supervised deep estimation network to perform localization. To make our study more plausible, we conducted a driving experiment and created a custom dataset to test our concept. The results showed that the proposed system is capable of detecting and locating traffic signs, with a 7.5% error rate cap up to 20 m ahead, from a single image of the Tokyo road area.

Contribution

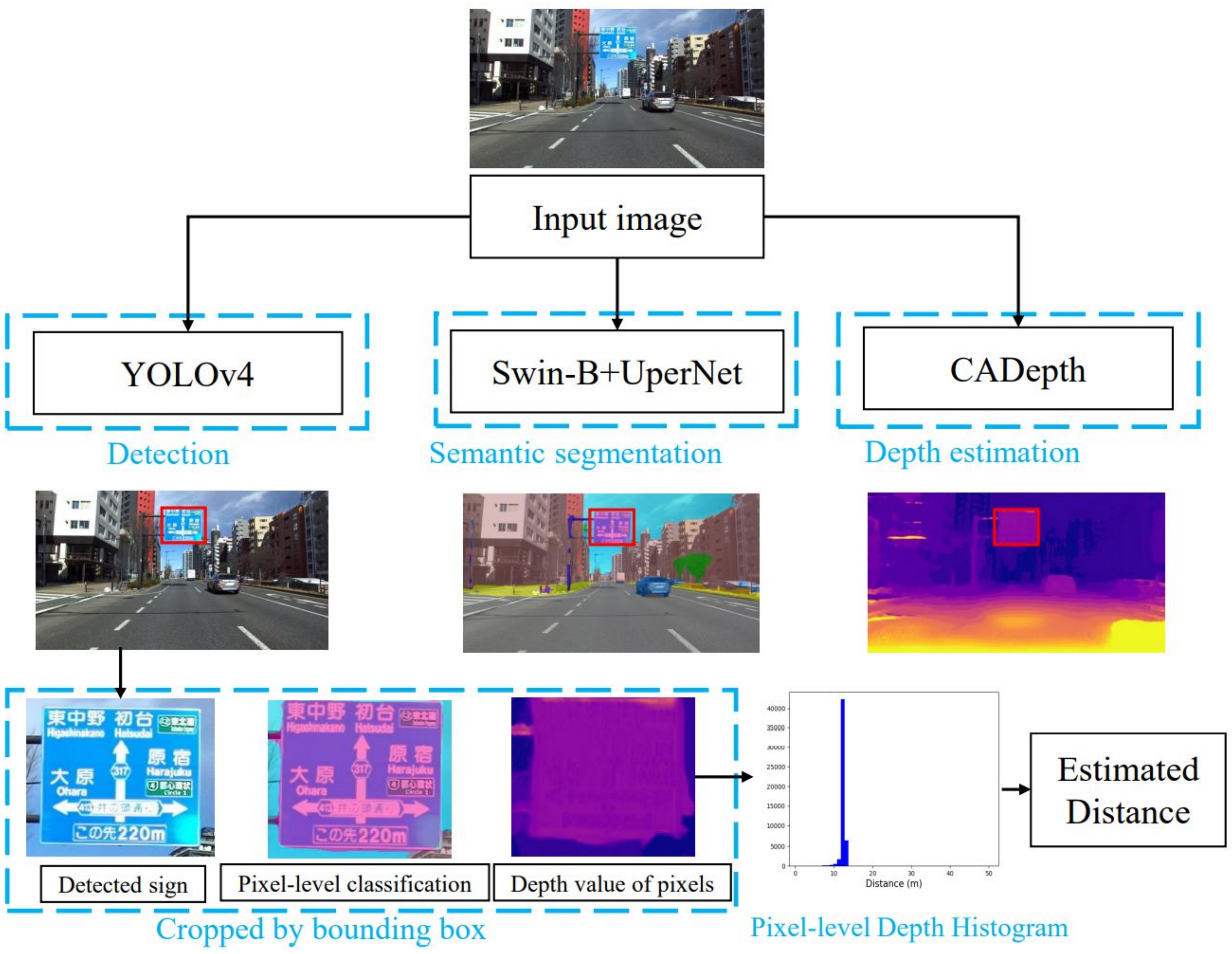
1. we proposed a vision-based pipeline to automatically detect and locate traffic signs from our custom driving image dataset of the Tokyo road environment.
2. We collected and developed a high-quality stereo driving images dataset of Tokyo road scene.
3. We also tested several deep neural networks for traffic sign classification

Dataset

(a) Experimental setup and driving route

(b)Examples of left and right view images taken by ZED stereo camera

Methodology



Combined with the detected bounding box from YOLOv4, the sign detection model made it easier to filter out the pixel region belonging to the target traffic signs because the pixels of the target sign occupy the principal part of the bounding box. Subsequently, (1) is applied to each pixel belonging to the target sign. A histogram was used to count the number of pixels of the target sign at the corresponding depth value. The final distance was obtained by calculating the average depth values in the depth histogram peak range.

Results

TABLE I. Yolov4 Detection Results (IoU= Threshold=50%)

AP	Precision	recall	F1-score	mAP
94%	89%	92%	91%	94%

*1)*The YOLOv4 detection model had high precision and recall for traffic sign detection, even when trained with a small amount of dataset. It can provide robust bounding boxes for the pixel-level depth value estimation.

2) We used the ZED stereo camera and BOSCH GLM50C (Bosch, Germany) data transfer laser rangefinder (measurement error within 1.5 mm) to collect 50 street images with different traffic signs as the ground truth distance data. The ground truth data ranged from approximately 4 to 20 m. From the calculation, the average distance error rate of the 50 test images was 7.5%, and the mean distance error was 0.6 m.