Robust Japanese Road Sign Detection and Recognition in Complex Scenes Using Convolutional Neural Networks

Ryo Hasegawa, Yutaro Iwamoto, and Yen-Wei Chen
College of Information Science and Engineering, Ritsumeikan University, Shiga, Japan
Email: is0322ix@ed.ritsumei.ac.jp, yiwamoto@fc.ritsumei.ac.jp, chen@is.ritsumei.ac.jp

***Abstract*—In recent years, the development of image processing technologies used for autonomous driving has been remarkable. Automatic detection and recognition of road signs are required for the practical use of autonomous vehicles. In the detection and recognition of road signs, changes in scale and contrast greatly affect the accuracy. In this study, we solve this problem by learning road signs using a deep learning technique that is robust against scale changes, and thought an experiment, we compare our method with recently proposed deep learning methods. We also show the results using our proposed method for individual Japanese road signs. The proposed method shows higher accuracy in the detection and recognition of road signs than the faster Region-based Convolutional Neural Network (Faster R-CNN) and Single Shot multibox Detector (SSD) methods.**

***Index Terms*—deep learning, detection, recognition, self-driving technology, road sign**

I. Introduction

To achieve autonomous driving, it is necessary to drive an autonomous vehicle in accordance with road regulations. This requires the automatic detection and recognition of road signs. In road sign detection, detecting and classifying road signs of various sizes in an image are important. In the large-scale datasets [1], [2] widely used in object detection, the target object of the object detection occupies most of the target image. However, in road sign images taken while driving, the target object may occupy only a small part of the image. Because these road signs are small but play an important role, they need to be detected and recognized, and compared with conventional object detection, more robust detections are required due to scale changes.

There are several issues in the study of road sign detection and recognition. The first is that the detection and recognition accuracies are low for objects that occupy small areas in an image, such as distant road signs. In autonomous driving, it is necessary to detect road signs in the distance in advance because it takes time to detect road signs and complete control processing for the car. The second is that contrast changes during nighttime and during bad weather affect the detection and recognition accuracies of road signs. For practical use in outdoor autonomous operation, it is necessary to perform high-accuracy detection and recognition without depending on environmental conditions, such as the weather. In this study, we propose a highly accurate detection and recognition method by learning road signs using a deep learning technique that is robust against scale changes. We then verify the effectiveness of the proposed method using original datasets. The contributions of this study are as follows:

1) The high-accuracy detection and recognition of road signs via a deep learning technique that is robust against scale changes;
2) A Comparison and verification of the method with recently proposed high-accuracy object detection algorithm; and
3) A Comparison of the effectiveness of the proposed method for individual Japanese road signs.

The structure of this paper is as follows. Section II introduces related research and recently proposed object detection methods using deep learning. Section III introduces the procedure and details of the construction of the Japanese road sign database. In Section IV, a deep learning training flow that is robust against scale changes is introduced. In Section V, the experimental datasets and experimental methods are explained and the experimental results discussed. Finally, in Section VI, the conclusion of our study is provided.

II. Related Work

In an object detection and recognition method, the tasks of detecting an object and recognizing the detected object are performed, and the object detection network then outputs the position and class of the object *via* these processes.

A. Road Sign Detection

Prior to the adoption of convolutional neural networks, various road signs were detected based on Support Vector Machine (SVM) [3], sparse representation [4], and sliding window [5] techniques. Creusen *et al.* detected road signs by manually performing feature design using color information by means of Histograms of Oriented Gradients (HOG) and SVM [6]. Such conventional object detection methods are primarily feature extraction

Manuscript received January 3, 2020; revised July 27, 2020.

doi: 10.18178/joig.8.3.59-66

algorithms because they depend on the setting, which decides what type of feature is important to extract. Color features and shape features are very important in the detection of road signs [7]. Therefore, it is necessary to design an optimal feature for each road sign. Conventional feature design is performed manually and therefore takes time and effort to properly design the function compared with automatic feature extraction using deep learning. In addition, in the extraction of candidate regions using a sliding window, the object candidate is searched for by moving a window of a preset size; therefore, it depends on the scale factor in the sliding window, and it is not easy to build a model that is robust to scale changes.

B. Object Detection via Deep Learning

Recently, road signs have been detected through state-of-the-art object detection methods using faster region-based Convolutional Network (Faster R-CNN) [8], Single Shot multibox Detector (SSD) [9], and You Only Look Once (YOLO) [10]. Object detection using deep learning methods has made automatic extraction of the optimum features of a target image possible. Optimal use of information from a large amount of training data enables detection with higher accuracy than conventional machine learning and improves processing speed. There are two detection methods using deep learning: the two-stage method and the one-stage method (Fig. 1). The two-stage method performs object detection and classification using separate networks. First, an object candidate area is searched for in an input image using selective search [11] or Region Proposal Network (RPN). Next, the object candidate region is classified using deep learning. The two-stage method uses two separate networks; therefore, a network suitable for each detection and classification process can be used. However, the processing is complicated and the speed is slow due to the separate networks. The one-stage method performs all of the object candidate area searches and class classifications within a single neural network. The one-stage method achieves the same detection and recognition accuracy as the two-stage method [12] but is faster and has less redundant processing. In this study, Real-time detection of road sign is necessary; therefore, in the proposed method, we used one-stage method, which has a high processing speed as a CNN network. The following describes Faster R-CNN, SSD, and YOLO, which are high precision object detection methods using deep learning that have been proposed so far.

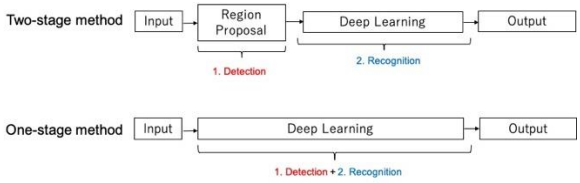


Figure 1. Two different types of object detection methods

1) Faster R-CNN

Faster R-CNN is an object detection and recognition method proposed by Ren *et al.* Faster R-CNN consists of two networks. One network searches for candidate areas of objects in an image, and the other classifies objects in the candidate area. In the search for the object candidate regions, RPN is used. RPN searches for object candidate regions using a method called sliding window that slides on the feature map window by window (Fig. 2). In the window used to search for the object candidate area, the object candidate area is determined by k anchor boxes with predefined aspect ratios. By preparing anchor boxes of various sizes, it is possible to detect object candidates of various scales. In the process of classifying object candidate areas, fixed-size feature maps are extracted using Region-of-Interest (ROI) pooling for the object candidate areas on the feature map. Then, object candidate areas pass through the identification network to determine to which class the object belongs.

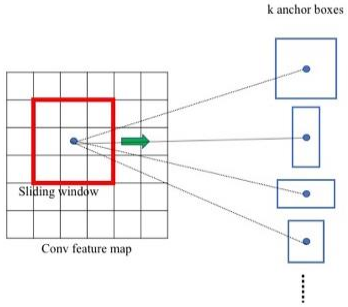


Figure 2. Search for object candidate areas using anchor boxes

2) SSD

SSD is an object detection and recognition method proposed by Liu *et al.* SSD performs the search and classification of the object candidate areas in a single network. In searching for object candidate areas, a default box, which works similarly to the anchor box in Faster R-CNN, is used. Unlike an anchor box, the default box is applied to multiple feature maps of different sizes. A large-sized feature map generated from the first half layer is responsible for detecting small objects, and a small-sized feature map generated from the second half layer is responsible for detecting large objects. However, because this method predicts a large number of object candidate areas, also a large number of erroneous object candidate areas will be generated. Therefore, with SSD, a method called hard negative mining is used to reduce false positives. The network structure of SSD is shown in Fig. 3. In SSD, different feature maps of various layers are used to determine the object candidate areas; therefore, good results can be obtained even for low-resolution images. SSD is a faster object detection method than Faster R-CNN.

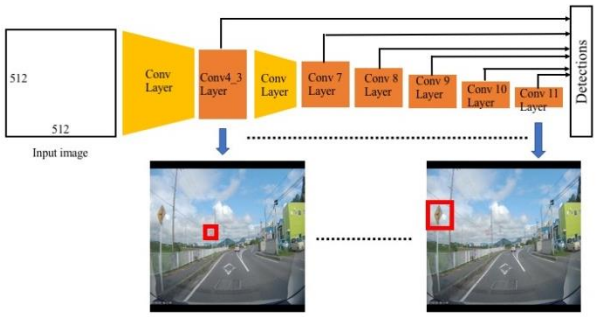


Figure 3. Network structure of SSD

3) YOLO

YOLO is an object detection and recognition method proposed by Redmon *et al*. It performs the search and classification of object candidate areas using a single network. In YOLO, the input image is divided into $S \times S$ grids in advance, and the class probability and the probability that an object candidate area and an object exist are calculated for each grid cell (Fig. 4). At this time, the central coordinates (x, y), width, and height are predicted for the object candidate regions. An object candidate region with a value obtained by multiplying the calculated class probability and the probability that an object exists in the object candidate region that is equal to or greater than a threshold is detected as an object. In SSD and Faster R-CNN, the image size used for learning is a fixed size. However, because YOLO has a Fully Convolutional Network (FCN) structure [13] and does not include full connection layers, it can be applied to images of various sizes. In SSD, features are extracted by the convolution layer, and a large number of object candidate regions are estimated from each layer; therefore, there are many false detections. In Faster R-CNN, because object areas are detected using the RPNs, the background is often erroneously detected as an object. However, YOLO uses all the information in an image and simultaneously learns the surroundings of an object; accordingly, it can suppress false detections of the background more than other methods.

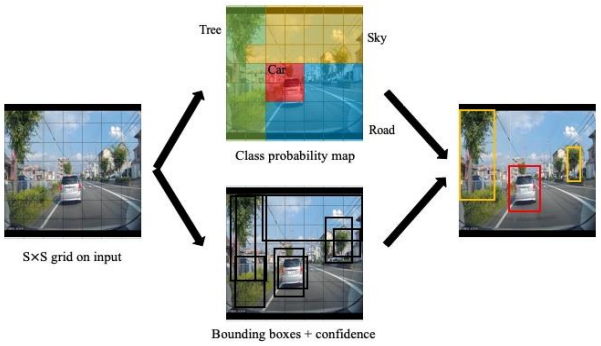


Figure 4. Object detection per grid

III. Japanese Road Sign Database

In this section, we will explain the construction procedure and the details of the Japanese road sign database.

A. Data Annotation

The flow for the road sign database construction is shown in Fig. 5. The construction of the database was done in three steps. The first task was to collect frames containing road signs from videos taken by a drive recorder. The second task was to record the coordinates of the road signs (center coordinates, width, and height) in the collected image. The third task was to assign label numbers corresponding to the coordinates of the collected road signs. All images in the database are taken with the same drive recorder. The maximum angle of view of the drive recorder is 120 ° diagonal, 100 ° horizontal, and 55 ° vertical. The drive recorder is installed horizontally with the vehicle traveling direction.

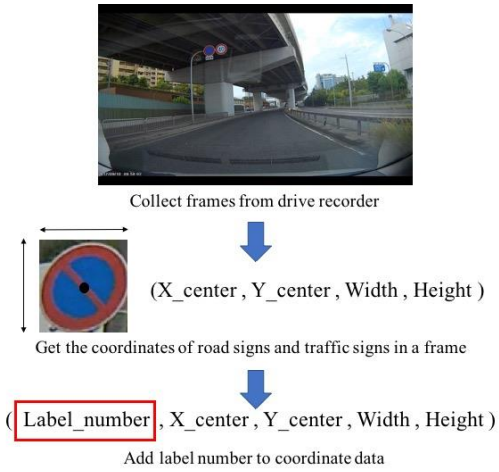


Figure 5. Procedure for the construction of the road sign database

B. Database Details

To detect and recognize road signs with high accuracy using deep learning, we constructed a database targeting 16 out of 102 types of road signs [14] by combining Japanese road signs and traffic signals (Fig. 6). The collected road signs were those that could be found 200 or more times in the video. They are also important road signs for car control. Japanese road signs can be classified into three categories, and the constructed database covers road signs included in all three categories: warning signs, regulation signs, and indication signs. There are three types of traffic signals: blue, red, and yellow. The size of the collected images is 1093×615. The image format is JPEG, and the label data are stored in the order of label number, center coordinates of road signs, width, and height. Fig. 7 shows data for each road sign size in the database. It can be seen from Fig. 7 that many road signs are less than approximately 1% of the size of the image. In addition, a relatively large number of small road signs are included in the dataset. The dataset includes 4477 images, and the total number of labels is 7160 (green light, 932; speed limit, 917; no parking, 840; bicycle and pedestrians only, 625; crossroad, 1, 460; red light, 425; crosswalk 1, 363; straight ahead or left turn permitted, 356; crossroads, 2, 339; crosswalk 2, 337; traffic division, 322; no overtaking, 294; no turns, 282; stop, 256; one-way street, 209; and yellow light, 203). In this experiment, four types of data augmentation (high contrast, low contrast, flip horizontal, and noise) were performed to increase the amount of data and, at the same time, improve the generalization performance. In the experiment, three types of test data, clear weather, night, and small objects, were constructed to compare the accuracies in each situation (Fig. 8). The number of labels included in each of the three types of test data is shown in Table I. The clear weather test data measure the accuracy when the visibility is good, such as when it is sunny, in the morning, or in the daytime. Night test data are for night conditions only. Using the night test data, it is possible to measure the effect on detection accuracy due to contrast changes. The small object test data include small objects (road signs 20 m or farther ahead) in the image. The

small object test data can be used to verify the accuracy for small objects in an image.

Figure 6. Examples of 16 types of road signs

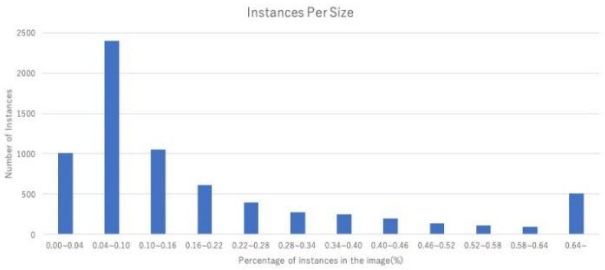


Figure 7. Size distribution of road signs in the database

Figure 8. (left) clear weather test data, (middle) night test data, and (right) small object test data

TABLE I. NUMBER OF LABELS FOR THE TEST DATA

Scene	Label
Clear weather	123
Night	241
Small objects	140

IV. METHOD

A. Problems in Road Sign Detection

In road sign detection, it is necessary to detect a road sign located relatively far away depending on the speed and situation in which the vehicle is traveling; therefore, compared with general object detection, sign detection must be performed, taking into consideration scale change. For example, when a car travels at 80 km/h, it will be approximately 60 m before the car can stop after the detection of a road sign [15]. Therefore, it is necessary to detect and recognize road signs 60 m in advance. Further, because vehicles detect road signs in an outdoor environment, conditions such as cloudy or rainy weather and the decrease in contrast at night greatly affect the results of detection and recognition.

B. Robust Road Sign Detection and Recognition by YOLOv2

1) YOLOv2

Real-time detection of road signs is necessary; therefore, in the proposed method, we used YOLOv2, which has a high processing speed as a CNN network and relatively few false detections. YOLOv2 is an improved version of YOLO, which was described in Section II. In YOLOv2, object candidates are detected for each grid by an anchor box, as in Faster R-CNN. In addition, the anchor box improves the detection accuracy by predetermining the most commonly used anchor box from the training data. When compared with YOLO, the detection speed and detection accuracy of YOLOv2 show better results. The network of the proposed method is shown in Fig. 9, which consists of 22 convolutional layers and 5 pooling layers, and all the layers are composed of convolutional layers; therefore, the position information is retained until output. Because YOLOv2 has an FCN structure and does not include the full connection layers, it has features that can be applied to images of any size. Therefore, in the proposed network, an image size, which is randomly changed for each batch from five image sizes, 640×640, 672×672, 704×704, 736×736, and 768×768, is used as the input image. YOLOv2 divides the input image into $S \times S$ areas and predicts the bounding box B (center coordinates, width, height, and confidence score) of each area and the conditional class probability C. The output of the YOLOv2 network is a feature map, which is $S \times S \times (B \times 5 + C)$ channels. In YOLOv2, the three predictions of the anchor box, confidence, and conditional probability are integrated into one error function. The YOLOv2 error function used in this experiment is shown below.

$$
\begin{aligned}
Loss = {} & \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\
& + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] \\
& + \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} \left(C_i - \hat{C}_i \right)^2 \\
& + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{noobj} \left(C_i - \hat{C}_i \right)^2 \\
& + \sum_{i=0}^{S^2} 1_{i}^{obj} \sum_{c \in classes} \left(p_i(c) - \hat{p}_i(c) \right)^2
\end{aligned}
\tag{1}
$$

here, x is the x-coordinate center of the predicted candidate area, y is the y-coordinate center of the predicted candidate area, w is the width of the predicted candidate area, and h is the height of the predicted candidate area. The variables $\hat{x}, \hat{y}, \hat{w}$, and $\hat{h}$ indicate the position information of the correct corresponding object area variable. $p(c)$ indicates the class probability of each class.

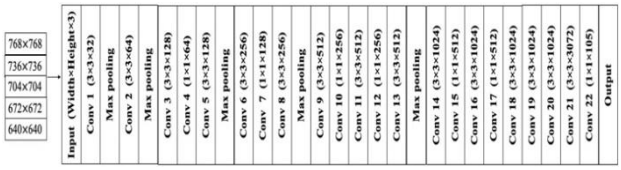


Figure 9. Network structure of YOLOv2

2) Robustness improvement with multi-scale inputs

In road sign detection, the detection must be performed robustly despite scale changes. In related object detection research, the image size used for training is often limited

to a fixed size; however, the size of an object changes depending on the distance to and the nature of the object. Therefore, it is necessary to collect abundant training data or perform trimming and enlargement processing on the object to generate a model that is robust to scale changes of the object, hence the creation of a model with the above characteristic *via* training with multi-scale inputs. Training with multi-scale inputs can result in a model that is more robust to scale changes than that resulting from training with only a single image size. Fig. 10 shows multi-scale image inputs. As mentioned previously, YOLOv2 is a network that can account for variable input image sizes because it has an FCN structure and does not include the full connection layers. Therefore, by preparing one image of a certain size, it is possible to simulate the object contained in the image at various sizes and to train the model based on these images. In the proposed method, training is performed using an image size in which one image is randomly changed from batch to batch between five image sizes. Therefore, in the proposed method, it is possible to train road signs with different scales even in the same image. As shown in Fig. 10, larger-scale images can be used to train larger-scale road signs, and smaller-scale images can be used to train smaller-scale road signs.

Figure 10. Training with multi-scale inputs

C. Evaluation

In this study, precision and recall are used as evaluation criteria for the detection accuracy. The Precision and recall are obtained via the following formulas:

$$\text{Precision}=\frac{TP}{TP+FP} \tag{2}$$

$$\text{Recall}=\frac{TP}{TP+FN} \tag{3}$$

here, *TP*, *FP*, and *FN* indicate true positive, false positive, and false negative, respectively. The higher the precision and recall, the better the accuracy. In this study, as a definition of "detection," when the Intersection over the Union (IoU) between the predicted candidate area and the correct object area is 0.3 or greater, it is regarded as a detection success. The formula for IoU is

$$\text{IoU}=\frac{(Object \cap Detected\ box)}{(Object \cup Detected\ box)} \tag{4}$$

where *Object* is the area of the correct object, *Detected box* is the predicted candidate area, and *Object* ∩ *Detected box* is the area where the two overlap. IoU represents the percentage of the image overlap and ranges from 0 to 1. Larger values indicate a higher amount of match between regions, and smaller values indicate that the regions are less consistent.

V. Experiment

A. Experimental Method

We detected and recognized 16 road signs using the proposed method. The experiment used the data set described in Section III. The division of test data used holdout method. We used Python to build the network used in the proposed method and Chainer [16] as the library. The Graphics Processing Unit (GPU) used was a GeForce GTX 1080 Ti [17]. The input image size at training in the conventional YOLOv2 was 704 × 704 pixels. In the comparison experiments with the proposed method, SSD and Faster R-CNN, the experimental conditions, such as the data augmentation in each method, were the same.

B. Single-Scale (Conventional YOLOv2) vs. Multi-scale (Proposed Method)

The experimental results of conventional YOLOv2 and the proposed method are shown in Table II. The resulting images are shown in Fig. 11. The detection results for each road sign are shown in Table III. As shown in Table II, in the model trained using multi-scale image inputs, the accuracy of the precision and recall improved for all test data compared with the model trained using image input of a single size. For the clear weather test data, the proposed method showed high accuracy. For the night and small object test data, the accuracy in the proposed method was improved for all test data compared with the model trained using single-size image input. However, it can be seen that the accuracies of the two datasets were reduced by approximately 10% or more compared with the case of clear weather. In the results in Fig. 11(b), one road sign could not be detected using conventional YOLOv2. Conversely, with the proposed method, the two road signs were correctly detected and recognized. Even for the night and small object datasets, the proposed method correctly detected and recognized the road signs as well as it did in the case of clear weather (Figs. 11(d) and 11(f)). Road signs that could not be detected correctly included those with complex backgrounds, such as buildings and forests.

TABLE II. Experimental Results in Each Environment

		Conventional YOLOv2 (Single-scale)	Proposed Method (Multi-scale)
Clear weather	Precision	78.50%	**94.64%**
	Recall	68.29%	**84.12%**
Night	Precision	52.12%	**83.76%**
	Recall	35.68%	**66.39%**
Small objects	Precision	66.66%	**77.06%**
	Recall	31.42%	**60.00%**

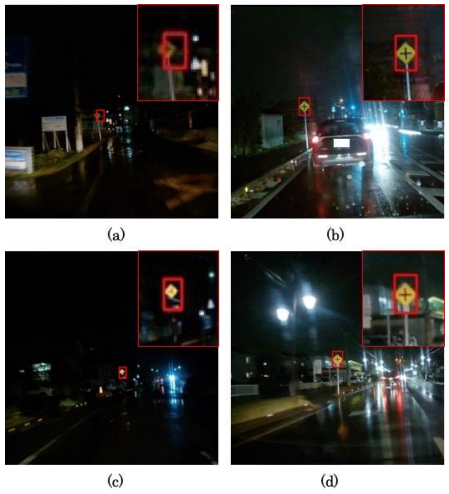
(a) (b) (c) (d) (e) (f)

Figure 11. Experimental result images in each environment: (a): conventional YOLOv2 (clear weather); (b): proposed method (clear weather); (c): conventional YOLOv2 (night); (d): proposed method (night); (e): conventional YOLOv2 (small objects); and (f): proposed method (small objects)

TABLE III. DETECTION RESULTS FOR EACH ROAD SIGN

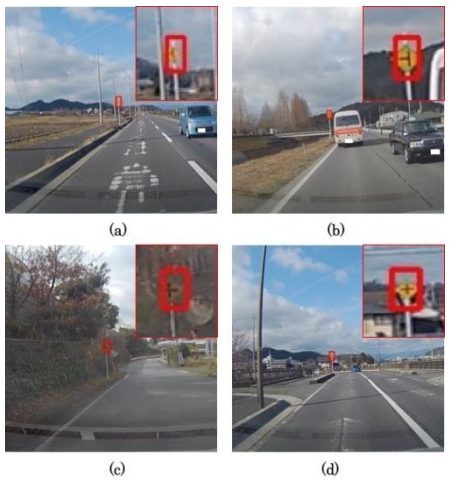
Clear weather	Conventional YOLOv2 (Single-scale)	70.0%	78.5%	72.7%	89.4%	83.3%	62.5%	37.5%	55.5%	25.0%	28.5%	57.1%	63.6%	42.8%	46.6%	55.5%	42.8%
	Proposed method (Multi-scale)	80.0%	85.7%	81.8%	100%	91.6%	87.5%	87.5%	90.0%	50.0%	85.7%	85.7%	90.9%	85.7%	86.6%	88.8%	85.7%
Night	Conventional YOLOv2 (Single-scale)	16.6%	22.2%	30.0%	30.6%	77.2%	33.3%	51.8%	17.2%	23.0%	14.2%	18.7%	28.5%	29.4%	12.5%	27.7%	30.0%
	Proposed method (Multi-scale)	66.6%	55.5%	70.0%	48.3%	86.3%	66.6%	70.3%	82.7%	53.8%	38.0%	75.0%	57.1%	47.0%	87.5%	77.7%	50.0%
Small objects	Conventional YOLOv2 (Single-scale)	65.2%	61.5%	57.1%	31.2%	30.0%	33.3%	25.0%	11.1%	28.5%	18.1%	16.6%	20.0%	35.0%	27.2%	14.2%	21.0%
	Proposed method (Multi-scale)	86.9%	84.6%	71.4%	81.2%	90.0%	66.6%	56.2%	50.0%	57.1%	54.5%	58.3%	60.0%	65.0%	54.5%	42.8%	52.6%

We evaluate the number of detections of each road sign when using conventional YOLOv2 and the proposed method. The evaluation index uses the recall. As shown in Table III, for the clear weather test data, road signs with high recall in conventional YOLOv2 include speed limit and no parking signs. In the proposed method, it can be confirmed that all 16 road signs show high recall. Results from the small object test data show that training with multi-scale image inputs is effective for all road signs. From the results of the night and small object test data, when using conventional YOLOv2, many of the road signs have low recall values. However, in the proposed method, there is an improvement in the recall value for each road sign and the recall value for the road signs as a whole is improved. In addition, in the night and small object test data, the yellow road signs have a low recall value. Even though these yellow road signs can be detected, only a few can be recognized, resulting in a low recall value. In the experiment, it was confirmed that yellow road signs were often confused for each other. Fig. 12 is an example of detecting a yellow road sign in the night test data. In Fig. 12(a), the road sign was misrecognized because it couldn't be detected at the correct position. In Fig. 12(c), the road sign is blurred, and it is difficult for us to recognize correctly even if we actually see and judge. Therefore, in Fig. 12(c), the road sign was misrecognized. On the other hand, in Figs. 12(b) and 12(d), since the detection position of the road sign was accurate and the image quality was relatively clear, correct recognition could be performed. Fig. 13 is an example of detecting a yellow road sign in the small object test data. In Fig. 13(a), correct recognition is difficult because the road sign is blurred and unclear because the road sign is too small. In Fig. 13(c), the contrast between the road sign and its background is similar and the contrast difference is small. Because of that, recognition of the road sign was very difficult and made false recognition. On the other hand, in Figs. 13(b) and 13(d), since the contrast difference between the road sign and its background is easy to distinguish and the image quality is relatively clear, the road sign was correctly recognized.



(a) (b) (c) (d)

Figure 12. Examples of detecting a yellow road sign in the night test data. (a) and (c) are examples of misrecognition; and (b) and (d) are correctly recognized examples



(a) (b) (c) (d)

Figure 13. Examples of detecting a yellow road sign in the small object test data. (a) and (c) are examples of misrecognition; and (b) and (d) are correctly recognized examples

C. Comparison with Other Methods Using Deep Learning

Comparison results between the proposed method and other methods using deep learning are shown in Table IV and Fig. 14. Fig. 14(a)-(c) are the results for clear weather, (d)-(f) for night conditions, and (g)-(i) for small objects. Table II and Table IV indicate that conventional YOLOv2 has a lower accuracy than Faster R-CNN and SSD. However, in the case of the proposed method in which training is performed with multi-scale image inputs, Table IV indicates that the accuracy is better than Faster R-CNN and SSD. In Faster R-CNN and SSD, the value of the precision is low because there are more false positives compared with the proposed method. In the test data results for clear weather shown Fig. 14, the proposed method correctly detects and recognizes all the road signs that were not by Faster R-CNN and SSD. In the result for the night test data, Faster R-CNN did not detect one road sign. In SSD, false detections caused by car lights occurred, and the crosswalk road sign was misidentified. However, in the proposed method, all the road signs were correctly detected and recognized, and there were no false detections. In the test results for small objects, two road signs serving as target objects were too small and could not be detected by Faster R-CNN and SSD. However, the proposed method correctly detected and recognized all road signs.

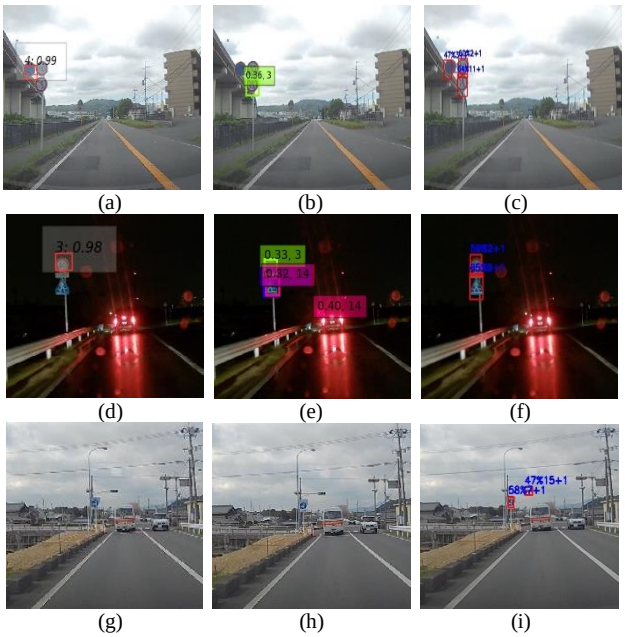


Figure 14. Comparison of image results with other methods using deep learning for clear weather (top row), night condition (middle row) and small objects (bottom row). (a), (d), and (g) are detection results for Faster R-CNN; (b), (e), and (h) are detection results for SSD; and (c), (f), and (i) are detection results for our proposed method

TABLE IV. COMPARISON RESULTS WITH OTHER METHODS

		Faster R-CNN	SSD	Proposed Method
Clear weather	Precision	83.33%	75.43%	**94.64%**
	Recall	77.23%	69.91%	**84.12%**
Night	Precision	70.04%	61.69%	**83.76%**
	Recall	63.07%	51.45%	**66.39%**
Small objects	Precision	67.32%	69.15%	**77.06%**
	Recall	48.57%	52.85%	**60.00%**

VI. CONCLUSION

In this study, we proposed a high-accuracy detection and recognition method by training road signs using a deep learning technique that is robust against scale changes. We then verified the effectiveness of the proposed method. In the proposed method, training with multi-scale image inputs had an effect on the scale changes of road signs, and four types of data augmentation were performed on the training data to improve the robustness of the method. The proposed method showed higher accuracy in detection and recognition than Faster R-CNN and SSD. Future issues include the improvement of the training data shortage and the increase of the detection number in each environment.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

In this research, software development, experiments and paper writing, Hasegawa; methodology and analysis, Iwamoto; conceptualization and validation, Chen.

REFERENCES

[1] M. Everingham, L. V. Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal Visual Object Classes (VOC) challenge," *IJCV*, pp. 303-338, 2010.

[2] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *IJCV*, pp. 1-42, Apr. 2015.

[3] S. Maldonado-Bascon, S. Lafuente-Arroyo, P. Gil-Jimenez, H. Gomez-Moreno, and F. Lopez-Ferreras, "Road-sign detection and recognition based on support vector machines," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 2, pp. 264-278, June 2007.

[4] K. Lu, Z. Ding, and S. Ge, "Sparse-Representation-Based graph embedding for traffic sign recognition," *IEEE Transaction on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1515-1524, 2012.

[5] G. Wang, G. Ren, Z. Wu, Y. Zhao, and L. Jiang, "A robust, coarse-to-fine traffic sign detection method," *IJCNN*, pp. 1-5, 2013.

[6] I. M. Creusen, R. G. J. Wijnhoven, E. Herbschleb, and P. H. N. D. With, "Color exploitation in hog-based traffic sign detection," in *Proc. IEEE International Conference on Image Processing*, 2010, pp. 2669-2672.

[7] K. Doman, D. Deguchi, T. Takahashi, Y. Mekada, I. Ide, and H. Murase, "Construction of a traffic sign detector using generative learning method considering color variation," *IEICE Transactions on Information and Systems*, vol. 93, no. 8, pp. 1375-1385, 2010.

[8] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. NIPS*, 2015.

[9] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. European Conference on Computer Vision*, Sep. 2016, pp. 21-37.

[10] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. CVPR*, 2016, pp. 779-788.

[11] J. R. Uijlings, K. E. Sande, T. Gevers, and A. W. Smeulders, "Selective search for object recognition," *IJCV*, vol. 104, no. 2, pp. 154-171, Sep. 2013.

[12] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. CVPR*, 2017, pp. 6517-6525.

[13] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. CVPR*, 2015, pp. 3431-3440.
[14] Ministry of Land, Infrastructure, Transport and Tourism. (2008). Road sign list. [Online]. Available: http://www.mlit.go.jp/road/sign/sign/douro/ichiran.pdf
[15] Itarda Information. (2015). Traffic accident research and data analysis. [Online]. Available: https://www.itarda.or.jp/itardainfomation/info111.pdf
[16] Chainer. (2015). Chainer: A flexible framework for neural networks. [Online]. Available: https://chainer.org
[17] Nvidia. (2017). GTX 1080 Ti Graphics card –GeForce. [Online]. Available: https://www.nvidia.com/ja-jp/geforce/products/10series/geforce-gtx-1080-ti/

Ryo Hasegawa received the B.E. degree from Ritsumeikan University, Kusatsu, Japan in 2019. He is currently working toward the M.E. degree at the Graduate School of Information and Engineering, Ritsumeikan University.

Yutaro Iwamoto received the B.E. and M.E., and D.E. degree from Ritsumeikan University, Kusatsu, Japan in 2011 and 2013, and 2017, respectively. He is currently an Assistant Professor at Ritsumeikan University, Kusatsu, Japan. His current research interests include image restoration, segmentation, classification of medical imaging and deep learning.

Yen-Wei Chen received the B.E. degree in 1985 from Kobe Univ., Kobe, Japan, the M.E. degree in 1987, and the D.E. degree in 1990, both from Osaka Univ., Osaka, Japan. He was a research fellow with the Institute for Laser Technology, Osaka, from 1991 to 1994. From Oct. 1994 to Mar. 2004, he was an associate Professor and a professor with the Department of Electrical and Electronic Engineering, Univ. of the Ryukyus, Okinawa, Japan. He is currently a professor with the college of Information Science and Engineering, Ritsumeikan University, Japan. He is also a visiting professor with the College of Computer Science, Zhejiang University and Zhejiang Lab, Hangzhou, China. He was a visiting professor with the Oxford University, Oxford, UK in 2003 and a visiting professor with Pennsylvania State University, USA in 2010. His research interests include medical image analysis, computer vision and computational intelligence. He has published more than 300 research papers in a number of leading journals and leading conferences including IEEE Trans. Image Processing, IEEE Trans. Cybernetics, Pattern Recognition. He has received many distinguished awards including ICPR2012 Best Scientific Paper Award, 2014 JAMIT Best Paper Award, Outstanding Chinese Oversea Scholar Fund of Chinese Academy of Science. He is/was a leader of numerous national and industrial research projects.