

Deep-learning-based Automatic Detection and Classification of Traffic Signs Using Images Collected by Mobile Mapping Systems

Hyeong-Yoon So and Eui-Myoung Kim*

Graduate School of Spatial Information Engineering, Namseoul University,
91, Daehak-ro, Seonghwan-eup, Seobuk-gu, Cheonan-si, Chungcheongnam-do, 31020 Republic of Korea

(Received April 27, 2022; accepted August 3, 2022; online published August 15, 2022)

Keywords: high-definition maps, traffic sign, mask R-CNN, Inception-v3, autonomous driving

As interest in autonomous driving has increased in recent years, various sensors have been developed for use in vehicles to detect and classify traffic signs. When road traffic facilities are not recognized owing to the malfunction of sensors, point cloud data and images collected by mobile mapping systems (MMSs) are used to construct high-definition maps containing road traffic facility information. However, when traffic signs, among the targets of high-definition map construction, are constructed using point cloud data, it becomes difficult to detect and classify traffic signs because they are highly reflective. In this study, we detected and sub-classified traffic signs by combining Mask Regions with Convolutional Neuron Network (Mask R-CNN) and Inception-v3 models based on image data obtained using MMSs. Image data obtained by various types of MMS were used to detect traffic signs and classification results were verified. The detection accuracy of traffic signs was 87.6% and the classification accuracy was 77.5%; thus, the method proposed in this study can be used to automatically construct traffic signs for high-definition maps.

1. Introduction

As interest in autonomous driving has increased, studies are actively underway to detect and classify road traffic facilities on roads using various sensors, such as LiDAR, RADAR, and cameras. However, when sensors are used, there are cases when a malfunction occurs or objects of interest are not detected under adverse circumstances, including long distance, blind spots, or bad weather conditions.⁽¹⁾ Accordingly, the need to construct high-definition maps is imperative so that this information can compensate for the malfunction of the sensors.

In South Korea, mobile mapping systems (MMSs) are currently used to collect various data, such as horizontal images, panoramic images, point clouds, and positional information to construct high-definition maps.⁽²⁾ The process of producing high-definition maps is as follows: the operator uses the collected point cloud to detect and classify objects of interest manually with the naked eye, and then digitizes them to construct maps. However, in the process of detecting and classifying objects of interest, those that are highly reflective to laser beams, such as traffic

*Corresponding author: e-mail: kemyoung@nsu.ac.kr
<https://doi.org/10.18494/SAM3956>

signs, have an insufficient point density in the point cloud, as shown in Figs. 1(b) and 1(d), making it difficult to classify and locate traffic signs with the naked eye. In contrast, classifying traffic signs and determining their locations are improved using images from MMS cameras, as shown in Figs. 1(a) and 1(c). In the case of traffic signs, using images from MMS cameras, rather than using point clouds, is effective when detecting and classifying objects of interest.

In studies on the detection of traffic signs using point clouds collected by MMSs, after classifying the ground and non-ground surfaces, methods such as Euclidean distance clustering, filtering and clustering using reflection intensity, and bag of visual phrases (BoVP) can be used to detect traffic signs.^(3,4) However, there are limitations, namely, the reflection intensity must be normalized for traffic sign detection, the process of preprocessing the point clouds consumes much time, and the results of traffic signs vary depending on the threshold setting determined by the user experience. For the efficient construction of high-definition maps, research is required to detect and classify traffic sign objects using images.

Studies on the detection of traffic signs using images have used image processing algorithms and convolutional neural network (CNN)-based deep learning models. Studies using image processing algorithms include a study using a FAST feature point extraction method and a study using the Viola-Jones algorithm to determine the edge of a traffic sign and then detect the traffic sign region using support vector machine (SVM).^(5,6) These methods can be used to detect the edge of the traffic sign, but their limitation is that the region cannot be detected if no edge is determined for the traffic sign.

Another study used color, shape, and both color and shape to detect traffic signs.⁽⁷⁾ In the case of the color-based method, the process was fast and simple, but it was sensitive to light. The shape-based method was limited in that the detection result varied depending on the type of edge detector used. When color and shape were used simultaneously, traffic signs could be detected effectively using a maximally stable extremal region (MSER) and high-contrast region extraction (HCRE), but the limitation in using color and shape together is that the detection accuracy of the traffic signs varied depending on the color enhancement method. In a study that detected objects using a machine learning or CNN-based deep learning algorithm, the You Only Look Once (YOLO) V4-Tiny model was trained using approximately 9000 images to detect and classify 45

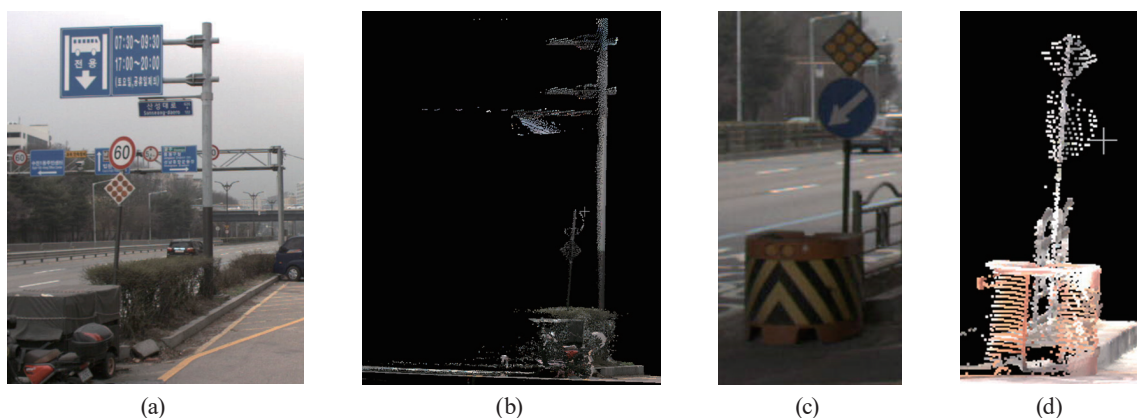


Fig. 1. (Color online) (a), (c) Traffic signs in images from MMS. (b), (d) Traffic signs in point cloud from MMS.

types of traffic sign, but it showed a low accuracy of approximately 52%.⁽⁸⁾ After training a Mask R-CNN model using 6000 images, a high accuracy of 95.2% was obtained after detecting and classifying ten types of traffic sign.^(9,10) Through these research cases, it was confirmed that the Mask R-CNN model can be used to detect and classify traffic signs, and the outlines of the traffic signs can be found through segmentation, a characteristic of the Mask R-CNN model.

The studies on object classification include the use of the MSER method to detect traffic signs in images and Inception-v3 to classify them into three main categories, namely, warning, regulatory, and mandatory signs, and the Belgian traffic sign dataset (BTSD) was used to train an Inception-v3 model and classify 62 traffic signs.^(11–14) Through these studies, it can be seen that in order to classify traffic signs, it is necessary to find the areas of traffic signs to be extracted from images first. Therefore, the purpose of this study is to propose a methodology for automatically detecting and classifying traffic signs by combining a Mask R-CNN model and an Inception-v3 model in images acquired using MMSs.

2. Materials and Methods

The traffic signs defined to construct high-definition maps in South Korea are classified into three main categories (Fig. 2): warning signs (B1_1), regulatory signs (B1_2), and mandatory signs (B1_3), which consist of 47, 27, and 35 classes, respectively, comprising a total of 103 classes as sub-classifications.⁽²⁾ The warning signs have a triangular shape with a red edge and the regulatory signs have a circular shape with a red edge. The mandatory signs are circular, triangular, or square with a blue edge.

Unlike Korean traffic signs, U.S. warning signs have a yellow diamond shape as shown in Fig. 3(a), and regulatory and mandatory signs include squares and letters on a white background as shown in Figs. 3(b) and 3(c).⁽¹⁵⁾

Unlike the Korean warning signs shown in Fig. 4(a), the Czech warning signs in Europe are triangular signs with white backgrounds as shown in Figs. 4(b) and 4(c). In addition, comparing the Korean and Austrian regulatory signs shows that the shapes are similar, but the background color or symbols are different.⁽¹⁶⁾

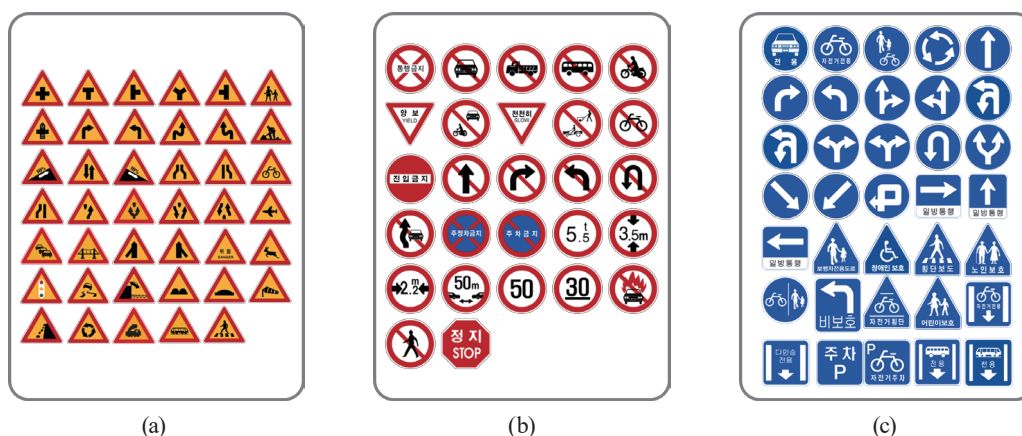


Fig. 2. (Color online) (a) Warning signs defined in high-definition maps, (b) regulatory signs, and (c) mandatory signs.



Fig. 3. (Color online) (a) Warning, (b) regulatory, and (c) mandatory signs in the United States.



Fig. 4. (Color online) (a) Korean warning sign, (b) Czech warning sign, (c) French warning sign, (d) Korean no overtaking regulatory sign, (e) Austrian no overtaking regulatory sign, (f) Korean one-way mandatory sign, and (g) Austrian one-way mandatory sign.

2.1 Methodology

In this study, we propose a methodology that combines a Mask R-CNN model and an Inception-v3 model to detect and sub-classify traffic signs in the images obtained using MMSs (Fig. 5). The first step in automatically constructing traffic signs on high-definition maps is to detect the region of the traffic sign in each image and define its main classification as a warning, regulatory, or mandatory sign.

The Mask R-CNN model was applied to the traffic sign regions in the original images collected using MMSs in order to define their respective main classifications. Then, the image coordinates (x, y) for the edge of the traffic sign region were obtained and recorded in a JavaScript object notation (JSON) format file. The JSON format file includes records of information, such as the main classification code, main classification name, and upper left and lower right image coordinates of the object region.

After automatically detecting the region of the traffic sign in the image and providing a main classification as either a warning, regulatory, or mandatory sign, the Inception-v3 model was required to sub-classify the traffic sign type in the second step. To this end, the Inception-v3 model was applied to read the JSON format file (the output of the first step), and after creating cropped images using the upper left and lower right image coordinates of the object region, the sub-classification class was assigned, attendant to the main classification.

2.2 Construction of training data

In this study, the AI-Hub data managed by the National Information Society Agency (NIA) was used to train the Mask R-CNN and Inception-v3 models.⁽¹⁷⁾ The “Panorama Image Set of

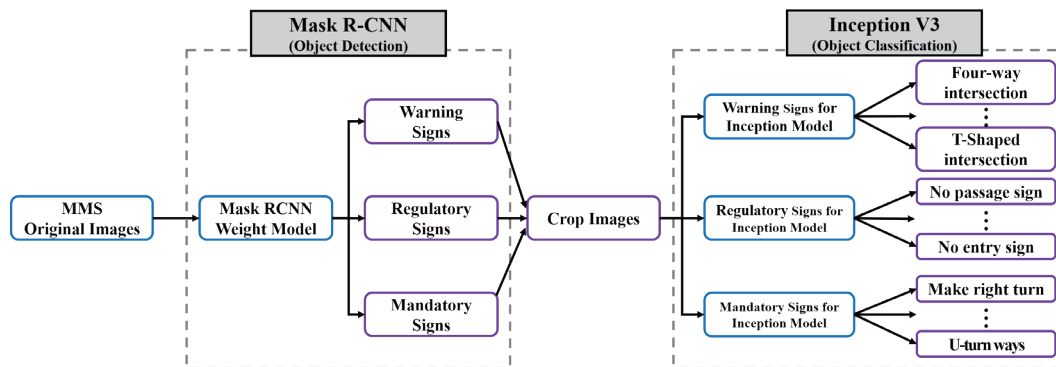


Fig. 5. (Color online) Methodology for detecting and classifying traffic signs.

Road Environments (v. 1.0)” provided by AI-Hub is a dataset that provides eight-direction images (.jpg) captured in the daytime using a horizontal camera mounted on an MMS and the label information (.json), such as traffic lights, traffic signs, and road surface marking (Table 1). The label information has records of the image name, class code and name, and upper left and lower right image coordinates.

AI-Hub provides approximately 330000 images of training data, but in some cases, the label was missing as shown in Fig. 6(a), or the labeling standard was not met as shown in Fig. 6(b). Therefore, we manually labeled 10000 images to develop the training data for traffic signs.

To proceed with the training of the Mask R-CNN model, we divided the 10000 manually constructed images in a ratio of 8:2 to create training and validation data. Table 2 details the construction of the training and validation data.

To proceed with the training of the Inception-v3 model, we used 86400 images provided by the AI-Hub and the object region information in the label file (.json) to create cropped images as shown in Fig. 7. The created cropped images were separated into three folders [warning (B1_1), regulatory (B1_2), and mandatory (B1_3)] through the information provided in the label file (.json). Then, three training datasets were constructed so that each folder would have sub-folders consisting of sub-classification codes.


3. Results and Discussion

3.1 Training and validation of Mask R-CNN model

The training was performed for 600 epochs using the constructed training data, and 100 steps were performed in one epoch. Among the 600 weighted models created through the epochs, a weighted model that had a low loss rate indicating the difference between the values predicted using the model and the true values of the problem and a high mean average precision (mAP) using the validation data was used to detect the traffic sign. The mAP represents an index that can consider both detection rate (recall) and accuracy (precision) at the same time.

Table 3 shows the loss and mAP for 200, 400, 568, and 600 epochs performed to detect the traffic signs, and Fig. 8 shows a graph of the loss rate for the validation data. Using Table 3 and

Table 1
(Color online) Sample images and label information provided by AI-Hub.

Category	Description	Format	Example
Planar split images	8 split images	.jpg	
Annotation information	Object location and type information	.json	<pre> { "interface": { "id": "028770", "filename": "028770_Panorama_crop_2.jpg", "path": "02/crop_image/", "resolution": [1000, 1000], "location": "Gangnam-gu", "datetime": "2020-01-09 11:28:18.0", "annotations": [{ "annotation_id": 161305, "annotation_type": "bbox", "class_code": "0201022600000", "class_name": "Traffic sign/Regulatory sign/Slow", "memo": "", "coord_xy": [[863, 940], [373, 443]] }, { "annotation_id": 161314, "annotation_type": "polygon", "class_code": "0204001300000", "class_name": "Road sign/Crosswalk warning", "memo": "", "coord_xy": [[613, 660, 700, 646], [562, 555, 591, 599]] }] } } </pre>

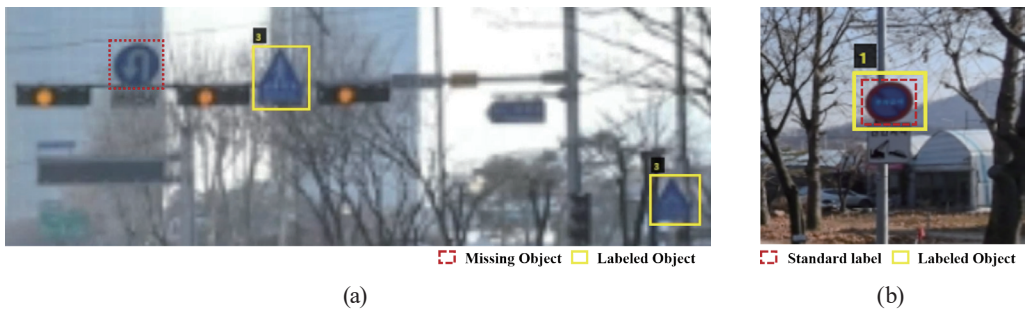


Fig. 6. (Color online) (a) Missing label contained in AI-hub dataset and (b) non-standard label.

Table 2
Detailed information of training and validation data.

ID	CODE	Description	Training set (8000 images)		Validation set (2000 images)	
			Number of labels	Ratio (%)	Number of labels	Ratio (%)
1	B1_1	Warning	446	9.7	116	10.0
2	B1_2	Regulatory	2162	46.9	535	46.2
3	B1_3	Mandatory	2000	43.4	508	43.8
Total			4608		1159	

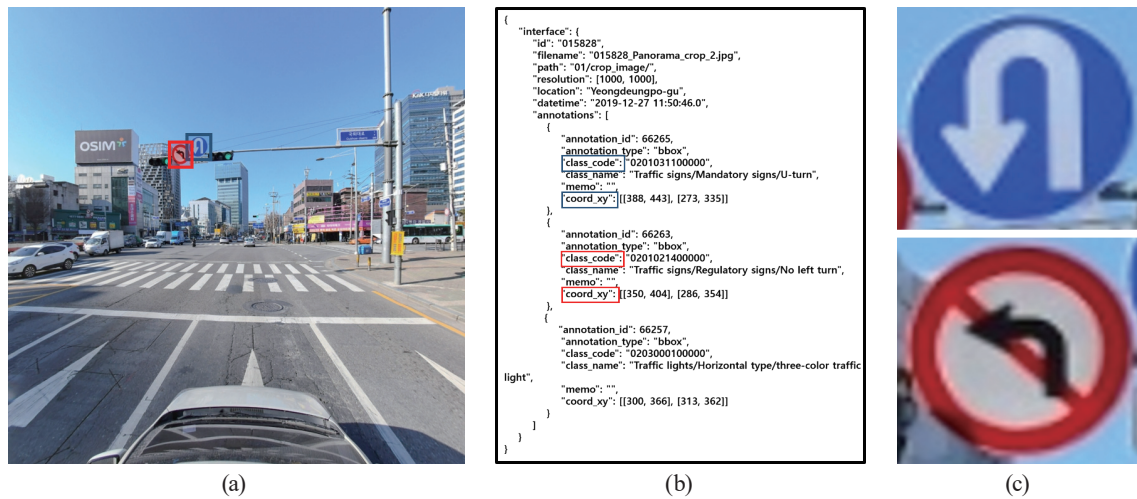


Fig. 7. (Color online) (a) AI-Hub original image (.jpg), (b) AI-Hub labeling data (.json), and (c) cropped images.

Table 3
Mean average precision and loss by number of epochs observed during training.

Epoch	200	400	568	600
Train_mAP (%)	72.5	93.8	95.5	95.4
Train_loss (%)	30.3	19.8	12.7	13.3
Val_mAP (%)	88.6	89.9	90.7	90.1
Val_loss (%)	47.8	47.3	37.9	46.0

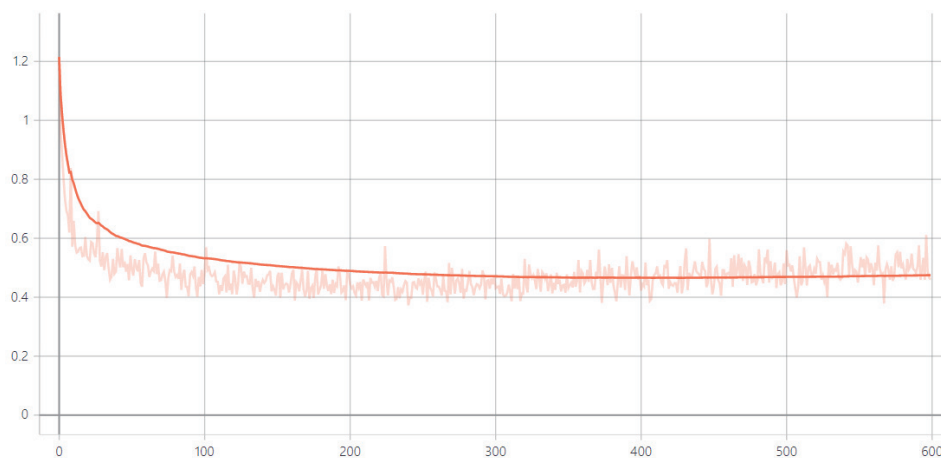


Fig. 8. (Color online) Loss graph for validation data.

Fig. 8, we found that in the case of 568 epochs, the training accuracy (95.5%) and the validation accuracy (90.7%) were the highest, and the loss was the lowest. On the basis of these results, a weighted model of 568 epochs was used for the detection of traffic signs to classify the traffic signs into three main categories, namely, warning, regulatory, and mandatory signs.

3.2 Assessment of Mask R-CNN model

Three types of MMS test set were used to objectively assess the performance of the training results of the Mask R-CNN model, which was intended to detect and classify traffic signs into either warning, regulatory, or mandatory signs. In test set 1, we used the images not used in the training of the Mask R-CNN model from the data provided by AI-Hub. For test set 2, we used 'Pegasus: Two Ultimate', Leica's MMS. Lastly, for test set 3, we used R1, an MMS of NaverLabs. Table 4 provides detailed information and shows the mAP results for the test sets.

In assessing the accuracy of detecting the traffic signs and determining their main classification, high accuracies of 91.3, 81.2, and 90.2% were obtained for the test datasets of test sets 1, 2, and 3, respectively. Figure 9 shows sample results of the detection of the traffic signs and their main classification in the three test sets using the weighted Mast R-CNN model.

3.3 Training and validation of Inception-v3 model

The Inception-v3 model was used to sub-classify the traffic signs, which were already classified as warning, regulatory, or mandatory signs. We created sub-classifiers based on the main categories to train the Inception-v3 model. The three sub-classifiers were trained by setting the epochs to 100000 times, the learning rate to 0.001, and the batch size to 100. As a result of the training, we found that the training accuracy was 83.1% for the warning sign sub-classifier, 88.9% for the regulatory sign sub-classifier, and 86.2% for the mandatory sign sub-classifier.

That the training accuracy was not higher than 90% was determined to be due to the imbalance in the sub-classifier data used for the training. Therefore, the amount of the training data was increased by repeating the processes of rotating the image, changing the image brightness, and changing the image size using the Image Data Generate (IDG) function provided by TensorFlow version 2.4.^(18,19) As a result, the training accuracy improved to 90.0% for the warning signs, 95.3% for the regulatory signs, and 91.5% for the mandatory signs (Table 5).

Table 4
Detailed information on the test sets and the accuracy of their classification using Mask R-CNN model.

ID	Description	Test set 1 (1000 images)			Test set 2 (1000 images)			Test set 3 (175 images)		
		Number of labels	Ratio (%)	Average precision AP (%)	Number of labels	Ratio (%)	Average precision AP (%)	Number of labels	Ratio (%)	Average precision AP (%)
1	Warning	74	12.1	95.0	156	24.6	80.6	3	1.4	100.0
2	Regulatory	330	54.0	92.3	358	56.4	86.5	134	63.2	94.7
3	Mandatory	207	33.9	86.7	121	19.1	76.3	75	35.4	76.0
Total		611	mAP	91.3	635	mAP	81.2	212	mAP	90.2



Fig. 9. (Color online) (a) Sample results of detection and classification using Mask R-CNN model in test sets 1, (b) 2, and (c) 3.

Table 5
Training results of Inception-v3 model.

	Dataset		AI-Hub Dataset		IMAGE	86400	
Epoch	100000		Learning rate		0.001	Batch size	100
	Warning		Regulatory		Mandatory		
	Original	Using_IDG	Original	Using_IDG	Original	Using_IDG	
mAP (%)	83.1	90.0	88.9	95.3	86.2	91.5	
Loss (%)	84.9	21.1	38.1	20.4	56.3	17.5	

3.4 Assessment of Inception-v3 model

To assess the performance of the trained warning, regulatory, and mandatory sign classifier models, we created and used cropped images based on the label data of the test datasets used for the performance assessment of the Mask R-CNN model.

According to the performance assessment results of the three sub-classifiers, the accuracies (mAP) were 77.3, 78.6, and 76.7% shown in Table 6. All three datasets were able to partially increase the evaluation accuracy by increasing the diversity of training data using IDG, but the

Table 6
(Color online) Classification results of symmetric traffic signs.

Code	Keep Left (313)			Keep Right (314)		
	Code	Amount	Ratio (%)	Code	Amount	Ratio (%)
Classification result	313	232	96.27	314	34	75.56
	309	1	0.41	313	11	24.44
	306	1	0.41			
	308	1	0.41			
	307	1	0.41			
	314	4	1.66			
	311	1	0.41			
	Total	241		Total	45	

high evaluation accuracy was not recorded. This means that the classification weight model calculated from the training data constructed using IDG has not been generalized. To solve this problem, it can be seen that it is necessary to increase the diversity of forms for training traffic signs by additionally constructing new data in training data.

Table 6 shows the classification results of symmetrical traffic signs. It was found that left-side traffic was misclassified as right-side traffic and vice versa. For this reason, it can be seen that the accuracy of classification is rather low in the case of symmetrical traffic signs.

4. Conclusions

In this study, a methodology that combines Mask R-CNN and Inception-v3 models to detect and sub-classify 103 traffic signs to produce high-definition maps of South Korea using images obtained by MMSs is described. The Mask R-CNN model was used to determine the regions of the objects, and the main categories were defined as warning, regulatory, or mandatory signs. These categories were then used in the Inception-v3 model to sub-classify the traffic signs into 103 classes.

The weighted Mask R-CNN model was trained to detect the traffic signs defined in the high-definition maps using the images captured by MMSs and to classify them into three main categories. The trained and weighted model was assessed using the test datasets obtained from various MMSs, and an average accuracy of 87.6% was obtained.

The weighted model was trained using Inception-v3 to create sub-classification categories for the traffic signs. The imbalance of the data between the classification targets in the training process of the weighted model could be partially solved by using the IDG function. Combining these models enabled the classification of 103 types of traffic sign in high-definition maps with an average accuracy of 77.5%.

The methodology proposed in this study can be used to detect and classify traffic signs systematically to produce the high-definition maps required for autonomous driving. It will be necessary to obtain additional training data to improve the accuracy of the models.

References

- 1 Y. S. Na, S. K. Kim, Y. S. Kim, J. Y. Park, J. M. Jeong, K. C. Jo, S. J. Lee, S. J. Cho, M. H. Sunwoo, and J. M. Oh: J. Korean Soc. Automot. Eng. **28** (2020) 797. <https://doi.org/10.7467/KSAE.2020.28.11.797>
- 2 National Geographic Information: HD Map Quality Crafting Manual, https://www.ngii.go.kr/kor/contents/view.do?sq=1241&board_code=contents_data (accessed June 2022).
- 3 Y. Yu, J. Li, C. Wen, H. Guan, H. Luo, and C. Wang: J. Photogramm. Remote Sens. **113** (2016) 106. <https://doi.org/10.1016/j.isprsjprs.2016.01.005>
- 4 A. Alvaro, S. Mario, A. Juan, A. Garcia, and R. Belen: Expert Syst. Appl. **89** (2017) 286. <https://doi.org/10.1016/j.eswa.2017.07.042>
- 5 M. J. Choi, J. K. Suhr, K. Choi, and H. G. Jung: IEEE **7** (2019) 149846. <https://doi.org/10.1109/ACCESS.2019.2947287>
- 6 P. Viola and M. J. Jones: Int. J. Comput. Vis. **57** (2004) 137. <https://doi.org/10.1023/B:VISI.0000013087.49260.fb>
- 7 C. Liu, S. Li, F. Chang, and Y. Wang: IEEE **7** (2019) 86578. <https://doi.org/10.1109/ACCESS.2019.2924947>
- 8 L. Wang, K. Zhou, A. Chu, G. Wang, and L. Wang: IEEE **9** (2021) 124963. <https://doi.org/10.1109/ACCESS.2021.3109798>
- 9 D. Tabernik and D. Skočaj: IEEE Trans. Intell. Transp. Syst. **21** (2020) 1427. <https://doi.org/10.1109/TITS.2019.2913588>
- 10 K. He, G. Gkioxari, P. Dollar, and R. Girshick: Proc. IEEE Int. Conf. Comput. Vis. (2017) 2980. <https://doi.org/10.48550/arXiv.1703.06870>
- 11 Y. Yang, H. Luo, H. Xu, and F. Wu: IEEE Trans. Intell. Transp. Syst. **17** (2016) 2022. <https://doi.org/10.1109/TITS.2015.2482461>
- 12 C. Lin, L. Li, W. Luo, K. Wang, and J. Guo: Period. Polytech. Transp. Eng. **47** (2018) 242. <https://doi.org/10.3311/PPtr.11480>
- 13 C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (2016) 2818. <https://doi.org/10.1109/CVPR.2016.308>
- 14 D. G. Kim, Y. G. Kang, Y. R. Park, N. Y. Kim, and J. Y. Lee: Spat. Inf. Res. **28** (2020) 241. <https://doi.org/10.1007/s41324-019-00285-x>
- 15 United States Road Symbol Signs: <https://mutcd.fhwa.dot.gov/services/publications/fhwaop02084/index.htm> (accessed June 2022).
- 16 Comparison of European Road Signs: https://en.wikipedia.org/wiki/Comparison_of_European_road_signs (accessed June 2022).
- 17 The Ministry of Science and ICT in Korea: <https://aihub.or.kr> (accessed June 2022).
- 18 J. S. Kim and I. Y. Hong: J. Korean Soc. Surv. Geod. **37** (2019) 119. <https://doi.org/10.7848/ksgpc.2019.37.3.119>
- 19 Image Data Generator of Tensorflow Core v2.0: https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/image/ImageDataGenerator (accessed June 2022).

About the Authors



Hyeong-Yoon So received his B.S. degree from Namseoul University, Korea, in 2021. He has been enrolled in the master's program in spatial information engineering at Namseoul University, Korea, since 2021. His research interests are in photogrammetry, artificial intelligence, and GIS. (ssoss95@naver.com)



Eui-Myoung Kim received his B.S. and M.S. degrees from Gyeongsang National University, Korea, in 1994 and 1996, respectively, and his Ph.D. degree from Yonsei University, Korea, in 2000. From 2000 to 2002, he was a senior researcher at the Korea Institute of Civil Engineering and Building Technology (KICT). He worked as a postdoctoral fellow at the University of Calgary, Canada, from 2003 to 2005. He was with the Korea Geospatial Information and Communication (KSIC) from 2005 to 2006. Since 2007, he has been a professor at Namseoul University, Korea. His research interests are in photogrammetry and GIS. (kemyoung@nsu.ac.kr)