*Journal of Multimedia Information System VOL. 7, NO. 1, March 2020 (pp. 41-44): ISSN 2383-7632 (Online)*
http://doi.org/10.33851/JMIS.2020.7.1.41

# Brief Paper:

# Drivable Area Detection with Region-based CNN Models to Support Autonomous Driving

Hyojin Jeon[1], Soosun Cho[2*]

**Abstract:** In autonomous driving, object recognition based on machine learning is one of the core software technologies. In particular, the object recognition using deep learning becomes an essential element for autonomous driving software to operate. In this paper, we introduce a drivable area detection method based on Region-based CNN model to support autonomous driving. To effectively detect the drivable area, we used the BDD dataset for model training and demonstrated its effectiveness. As a result, our R-CNN model using BDD datasets showed interesting results in training and testing for detection of drivable areas.

**Key Words**: Detection of drivable areas, Region-based CNN, BDD dataset, Autonomous driving software.

## I. INTRODUCTION

Autonomous driving typically relies on sensors, actuators, complex algorithms, machine learning systems, and critical software that requires a powerful processor. Object recognition in particular helps autonomous driving software to follow traffic rules and navigate obstacles. The technique of indicating whether the desired object exists in a given image and then where it is located is divided into object detection and segmentation.

Object detection performs classification and indicates the position of an object through a bounding box. Segmentation divides the image into pixels to determine what object class it is, so it shows the location in more detail than detection. Segmentation includes semantic segmentation and instance segmentation. The former distinguishes between objects and determines what class each object is. The latter distinguishes between different instances of the same class. Semantic segmentation aims at dividing different kinds of objects into classes, and FCN, DeepLab [1] are known as representative models. Instant segmentation displays different colors on the prediction mask when the same class corresponds to different instances. DeepMask and Mask R-CNN [2] are typical models.

These deep learning object recognition models are actively used in autonomous driving research. In particular, the Region-based CNN (R-CNN) models have evolved to output not only object detection but also segmentation in just a few years, and it is expected to be developed more in the future. Recently, research using this model has been actively conducted in many industries as well as autonomous vehicle research.

This paper introduces a case study of extracting drivable areas from road images and analyzing the results. For this purpose, we used the most recently developed Mask R-CNN [2] model among the R-CNN models, and we used the BDD [3] dataset in order to train the model for a situation similar to the actual road environment.

## II. RELATED STUDIES

Mask R-CNN is a model that implements both functions of object detection and partitioning and it connects R-CNN's lineage. Early R-CNN showed that high-performance object detection can be achieved by linking the Convolutional Neural Network, which performs image classification, and the Region Proposal algorithm, which suggests areas where objects exist in an image [4]. However, early R-CNN had a problem that it took a long processing time. To address this problem, Fast R-CNN was developed, which introduced the concept of Region of Image Pooling and solved both time and cost issues by organizing all models into a single network [5].

In addition, Faster R-CNN [6] appeared to save time by creating a faster and more accurate domain proposal by designing the network inside the model. Afterwards, Faster R-CNN has evolved to Mask R-CNN [2], a method of adding a branch to create a mask by placing a Fully

**Manuscript received February 27, 2020; Revised March 09, 2020; Accepted March 15, 2020. (ID No. JMIS-20M-02-005)**

Corresponding Author (*): Soosun Cho, Uiwang-si, Gyeonggi-do, Republic of Korea, 031-460-0584, sscho@ut.ac.kr.
[1]Dep.of Com. Science & Info. Eng., Korea National University of Transportation, Rep. of Korea, hyodg@naver.com
[2]Dep.of Com. Science & Info. Eng., Korea National University of Transportation, Rep. of Korea, sscho@ut.ac.kr

Convolutional Network on top of the CNN feature in order to enable pixel-level segmentation. This enables not only object detection but also segmentation.

Table 1 shows a brief history of CNNs in Image Segmentation [7]. We choose Mask R-CNN, the latest technology that offers the most powerful features.

Table 1. A brief history of CNNs in Image Segmentation.

| | |
|---|---|
| 2014:<br>R-CNN | An Early Application of CNNs to Object Detection **:** propose a bunch of boxes in the image and see if any of them actually correspond to an object |
| 2015:<br>Fast R-CNN | Speeding up and Simplifying R-CNN<br>Insight 1: RoI (Region of Interest) Pooling<br>Insight 2: Combine All Models into One Network |
| 2016:<br>Faster R-CNN | Speeding Up Region Proposal : only one CNN needs to be trained |
| 2017:<br>Mask R-CNN | Extending Faster R-CNN for Pixel Level Segmentation : RoiAlign - Realigning RoIPool to be More Accurate |

## III. PROPOSED METHOD

The purpose of our study is to develop a case study of learning Mask R-CNN by using dataset which contains information on driving area under various environmental conditions including various weather conditions. For this purpose, BDD dataset [3] was used. In the paper of Mask R-CNN [2], there are no experiments with BDD datasets. So the results of our experiment will be compared with the results of this paper.

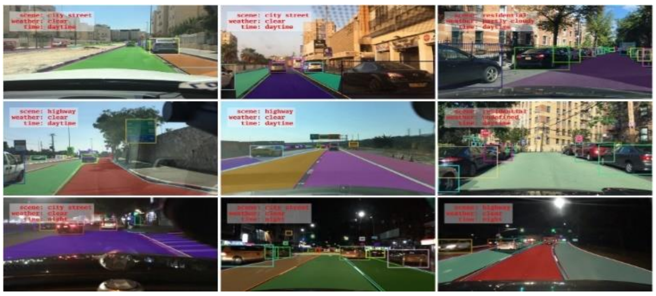

Fig. 1. BDD dataset represents drivable area and road condition.

BDD dataset is a database related to road and driving called BDD100K, published by the UC Berkeley Artificial Intelligence Lab. The BDD100K stands for Berkeley Deep Drive and consists of 70,000 training images, 10,000 validation images and 20,000 test images. All images are 1280x720 size. The dataset, which consists of 100,000 images, contains annotated images of buses, traffic lights, traffic signs, people, and cars as 2D bounding boxes. As shown in Fig. 1, the information about the driving area in the form of image segmentation and polygons is included. In addition, various weather conditions are recorded, such as rainy, cloudy, foggy, or sunny weather. Dataset and image segmentation labels and annotations can be downloaded from the BDD website [8].

In this experiment, we implement a Mask R-CNN deep learning model that extracts the driving area by using the BDD dataset as the training and validation image for the model. And the recognition result and accuracy through the implemented model are compared with the existing results.

In addition, after testing using the best performing learning weights obtained through the experiment, the driving area is extracted, and the result is visually inspected. Through this, we will check the case of incorrect classification in our model.

## IV. EXPERIMENTAL RESULTS

The system specifications and software environment for model training and testing are as follows; GPU computation was performed using CUDA 9.0 and cuDNN 7.5.0, both architecture and software to speed up learning in deep learning.

Mask R-CNN provides a guide to learning using the MS COCO dataset [9]. They do not provide a learning guide for the BDD dataset used in this experiment, so we use this guide to create a learning drive file. The information of each image in the BDD dataset is stored as a json file, which includes information about the detection object and drivable area provided by the dataset. Json file including original image and information can be used as input data of Mask R-CNN. In learning, only the information whose object category is marked as “drivable area” of the json file is extracted, and then the learning file is modified and used to learn the model.

The number of learning was performed 20,000 times by performing 200 epochs at 100 times per step. In order to check the difference according to the number of layers of the neural network, the layers are designated as 'heads' from the beginning to 40 epochs, '4+' from 41 epochs to 120 epochs, and 'all' from 121 epochs to 200 epochs. We confirmed how the number of layers in a neural network affects performance. The neural network of the model was tested in both cases of ResNet-50 and ResNet-101 to compare the performance according to the depth of the neural network. In addition, only 90% of reliability is adopted to consider the risk of incorrect results in extracting drivable areas. And to check whether resetting the image size mentioned in Mask R-CNN has a significant effect on the performance improvement of the model, we tested by changing the image size of the dataset in various ways.

In the experiment, we used the criterion called Intersection over Union (IoU) [10], which is used as an index to evaluate the accuracy of the predicted bounding box in object detection. The higher the percentage of overlap between the prediction and the actual bounding box, the more accurate the detection. Figure 2 represents the concept of IoU.

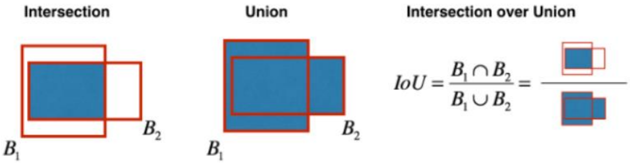

Fig. 2. Concept of Intersection over Union (IoU).

The performance of the model trained using the BDD dataset was verified by comparing IoU with the best performance in the paper of Mask R-CNN model[2]. In addition, the best learning weight was obtained by comparing the performance of models using various learning methods. Using the acquired learning weights, we extracted the drivable area from the image obtained from the driving vehicle and confirmed whether it is effective in general data. In the experiment, the minimum size of the image was changed to 800 and the maximum size was changed to 1024, and the learning reliability was set to adopt only 0.9 or more detection reliability.

To see how well the Mask R-CNN using BDD can detect the drivable area, we compared the results of the model trained with the MS COCO dataset [9], which was the best in work [2]. A comparison of the experimental results is shown in Table 2.

Table 2. comparison of the experimental results in two datasets.

| Datasets | Layers | Loss | Time(ms/step) | IoU |
|---|---|---|---|---|
| MS COCO | heads | 1.1820 | 536 | 0.6563 |
| | 4+ | 1.1020 | 696 | |
| | all | 0.6098 | 818 | |
| BDD Dataset | heads | 1.1796 | 498 | 0.6833 |
| | 4+ | 0.7375 | 661 | |
| | all | 0.4372 | 778 | |

In learning with the BDD dataset, Loss was 0.0024, 0.3645, 0.1726 lower for each layer, and the time required for one step was about 1.07 times faster than the learning with the MS COCO dataset. The IoU was 6.03% higher, indicating better accuracy. Compared to the MS COCO dataset with 81 classes, the BDD dataset appears to have a low loss because it consists of only two classes: background and drivable area. In addition, the sum of the training and validation images of the dataset was 80,000, which is less than the 123,287 MS COCO dataset, which contributed to the reduction of the time required. In any case, Mask R-CNN showed relatively good performance in detecting drivable areas as well as in detecting objects.

By changing the image size of the dataset in various ways, we confirmed that it greatly affected the performance of the model. The comparison of the experimental results is shown in Table 3.

Table 3. comparison of the results in various image sizes.

| Image Size | Layers | Loss | Time(ms/step) | IoU |
|---|---|---|---|---|
| 128 x 100 | heads | 2.0558 | 249 | 0.5433 |
| | 4+ | 0.9717 | 311 | |
| | all | 0.5253 | 325 | |
| 256 x 200 | heads | 1.1101 | 255 | 0.8 |
| | 4+ | 0.5753 | 328 | |
| | all | 0.3575 | 341 | |
| 512 x 400 | heads | 0.7318 | 304 | 0.7250 |
| | 4+ | 0.8299 | 390 | |
| | all | 0.4250 | 427 | |
| 1024 x 800 | heads | 1.1796 | 498 | 0.6833 |
| | 4+ | 0.7375 | 661 | |
| | all | 0.4372 | 778 | |
| 1280 x 720 | heads | 1.8078 | 575 | 0.785 |
| | 4+ | 0.8160 | 653 | |
| | all | 0.5697 | 824 | |

The best performance was found when the image was reduced to 256x200. The IoU value is similar to the case of learning with the same size as the original image, but the loss is lowered, and the learning time is reduced by more than half.

Also, in order to compare the performance according to the neural network depth, we experimented by setting the neural network of the model in two cases, ResNet-50 and ResNet-101. Better performance was achieved using a model trained with ResNet-101. Loss was 0.8358, 0.3353, 0.1095 lower for each layer compared to the ResNet-50 and the IoU was 11.67% higher, indicating better accuracy. But the time required for one step was about 0.85 times slower. This shows that ResNet-101 has a deeper neural network than ResNet-50, resulting in slower time but better performance.

Through experiments, we obtained the best learning weight and then visually examined the result of drivable area extraction using the test image of the BDD dataset. The test set consisted of 100 randomly selected test images from the BDD dataset. It is designed to cover different driving environments such as cities, highways, intersections, curves, day and night, rainy weather and snowy weather.

The test results with the best learning weights are shown in Table 4. Of the total 100 images, Mask R-CNN correctly extracted the drivable area from 83 test images, extracted the wrong drivable area from 17 test images, and failed to extract the drivable area from 4 test images. The precision was 83%, the recall was 95%.

Table 4. test result: visually examined drivable area extraction.

| | | | |
|---|---|---|---|
| True Positive | 83 | | |
| **False Positive** | 17 | Opposite Side | 5 |
| | | Shoulder | 8 |
| | | Parking Area | 3 |
| | | Center Line | 1 |
| True Negative | 4 | | |

Among the 17 failed recognitions, there were 5 images that recognized the opposite lane, 8 images that detected the shoulder, 3 images that recognized the parking area, and 1 image that recognized the center line. Fig 3. represents the failed recognition examples

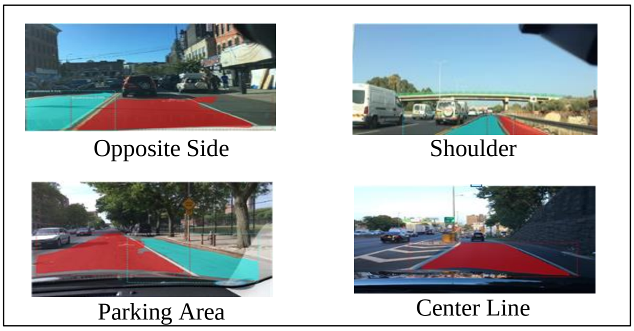
Opposite Side
Shoulder
Parking Area
Center Line

Fig. 3. The failed recognition examples.

## V. CONCLUSION

In this paper, we implemented a model that extracts drivable area using BDD dataset in Mask R-CNN model. Mask R-CNN is one of the most famous of the image segmentation models and is being actively researched recently. It is attracting attention in the research field that needs the location information of the object in the image like autonomous driving.

To check the training accuracy of the model using the BDD dataset, we evaluated the accuracy based on IoU. We can see that the accuracy is higher than that from training with the existing dataset, MS COCO.

In order to verify that the trained model works well, the test was conducted using the best learning weight. The 100 test images were visually confirmed, and the accuracy was 83%. In conclusion, the mask R-CNN showed better performance in extracting drivable areas when learning using the BDD data set.

In our future research, we will detect not only drivable area using the object detection function of Mask R-CNN but also objects of 10 classes provided by BDD data set. By providing them with drivable areas, the quality of information that can be obtained in driving environments will be improved.

## REFERENCES

[1] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation,” *Lecture Notes in Computer Science*, pp. 833–851, 2018

[2] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask R-CNN,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, October 2017.

[3] F. Yu, W. Xian, Y. Chen, F. Liu, M. Liao, V. Madhavan, T. Darrell, “BDD100K: A Diverse Driving Video Database with Scalable Annotation Tooling,” arXiv:1805.04687, 2018.

[4] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation”, in *Proceedings of CVPR*, Columbus, Ohio, USA, June 2014.

[5] Ross Girshick, “Fast R-CNN,” in *Proceedings of ICCV*, Santiago, Chile, Dec. 2015.

[6] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks,” in *Proceedings of NIPS*, 2015.

[7] A Brief History of CNNs in Image Segmentation: From R-CNN to Mask R-CNN, https://blog.athelas.com/a-brief-history-of-cnns-in-image-segmentation-from-r-cnn-to-mask-r-cnn-34ea83205de4, 2019.

[8] Berkeley DeepDrive, bdd-data.berkeley.edu, 2019.

[9] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, “Microsoft COCO: Common Objects in Context,” in *Proceedings of the European conference on computer vision (ECCV)*, Zurich, Switzerland, pp. 740-755, Sep. 2014.

[10] Intersection of Union, www.pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/, 2019.