Performance Analysis of Object-based Scene Recognition Algorithm using Mask R-CNN

Min Htet Thar*, Cheol Min Lee*, Dong Myung Lee*
*Dept. of Computer Engineering, Tongmyong University

htatmin332@gmail.com, captaink74282gmail.com, dmlee@tu.ac.kr

Abstract

In this paper, an object-based scene recognition algorithm was proposed by applying the mask region-based convolutional neural network (Mask R-CNN) framework, and the accuracy of the proposed algorithm was analyzed. The proposed algorithm is designed based on a deep learning model consisting of a convolutional neural network (CNN), region proposal network (RPN), and region of interest (RoI) align, and detects objects scattered in the surrounding environment. The average object-based scene recognition accuracy was measured to be about 95% and 93% in 1st and 2nd experiments, respectively.

I. Introduction

Scene recognition is one of the computer vision techniques that can identify environmental scenes in an image or video. It collects information about the surrounding environment of a given scene and learns to recognize what a specific scene is, like the human eye. Most scene recognition algorithms are designed using a simple image classification function that learns a set of scene image data.

In general, since a scene is composed of various objects, it is difficult for most scene recognition algorithms based on an image classification function to completely recognize a scene without detecting an object. Therefore, an algorithm for recognizing a scene with an object-based scene recognition algorithm is known to have excellent recognition performance among many scene recognition technologies currently being presented. In this paper, we designed an object-based scene recognition algorithm using various objects controlled by the surrounding environment on mask region-based convolutional neural network (Mask R-CNN) frame, and analyzed the recognition performance.

II. Related Studies

An indoor positioning system using scene recognition was proposed by [1]. The system uses the received signal strength indicator (RSSI) and a database of fingerprint maps of the scene to estimate the user's location. Since this system performs scene recognition for the purpose of estimating the user's location, the performance of scene recognition is one of the very important factors.

As another related study, Mask R-CNN is one of a series of R-CNN models for object detection [3]. This is an extension of Faster R-CNN, which is a general framework, adding segmentation functionality in the R-CNN model. This model is known to achieve the best performance in instance segmentation, object detection, and human key point detection.

III. Proposed Algorithm

The proposed algorithm is performed in two stages: object detection stage and scene recognition stage. As shown in Fig. 1, the object in the image is 1st detected in the object detection step. Then, in the scene recognition step, objects are collected and constructed using the scene database [5].

3.1 Object Detection Phase

In the object detection phase, as shown in Fig. 2, the input image is first extracted as a feature map, and the location of the object in the image is extracted using region proposal network (RPN) [4]. The detected object is adjusted to the maximum value through region of interest (RoI) alignment, which is classified as a detected object [2]. Upon completion of this process, detected objects are output as categories.

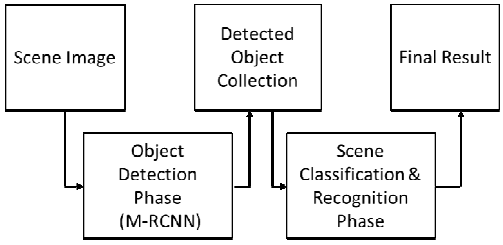

Fig. 1. Architeture of object-based scene recognition algorithm.

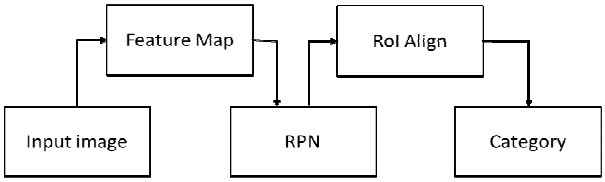

Fig. 2. Object detection phase using Mask R-CNN.

3.2 Scene Recognition Phase

3.2.1 Scene database

In the scene recognition phase, the scene database is important for recognizing objects as scenes. In the scene database, each scene is made up of the most relevant objects, as shown in TABLE I. If all detected objects match the relevant scene according to the scene database, the scene with the highest accuracy is selected and recognized. Therefore, the scene recognition accuracy is significantly affected by the quality of the scene database.

3.2.2 Scene classification and recognition flow

Scene classification and recognition flows are shown in Fig. 3. Each object detected in the category is checked if it matches all scenes in the scene database. If one of the detected objects is related to an object in a scene database, the scene is selected as a pre-available scene. If all objects detected in the category match each scene in the database, all pre-available scenes are collected. Then, from among the pre-available scenes, the scene that most matches the detected object in the category is selected and recognized as the final scene [5].

TABLE I. SCENE DATABASE

Scene				
Bathroom		Classroom	Computer Room	Library
Objects	Toilet	Table	Table	Books
	Sink	Chair	Chair	Chair
	People	Monitor	Monitor	Table
	-	People	Keyboard	People
	-	-	People	-

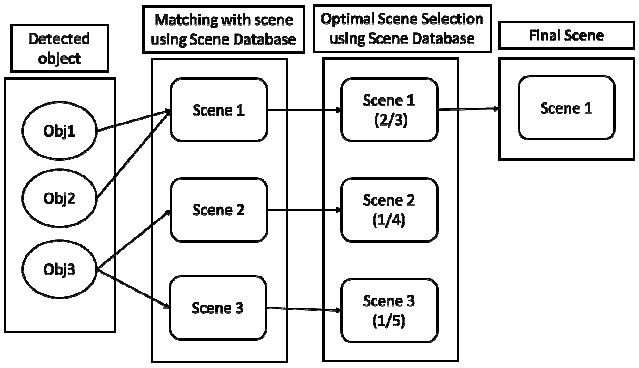

Fig. 3. Scene classification and recognition flow.

IV. EXPERIMENT AND RESULTS ANALYSIS

4.1 Experiments Environment

In order to test the accuracy of scene recognition of the proposed algorithm, Experiments were performed twice using images of 4 scenes: bathroom, classroom, computer room, and library with 20 and 40 images in the 1st and 2nd experiments, respectively. First, the object detection phase was performed by applying the Mask R-CNN framework, and then the scene recognition phase was performed using the scene database.

4.2 Result Analysis

The accuracy of scene recognition of the experiment was as shown in Fig. 4, and the average object-based scene recognition accuracy was measured to be about 95% and 93% in 1st and 2nd experiment, respectively. In the 1st experiment, the recognition accuracy for both the library and computer room scenes were 90%, and the recognition accuracy for both the classroom and bathroom scenes were 100%, respectively. In the 2nd experiment, the recognition accuracy for bathroom scenes and remaining scenes was 100% and 90%, respectively.

As result analysis, the reason that the recognition accuracy for bathroom scene was 100% on two experiments is the bathroom scene has few objects similar to other scenes, so even if only one or two objects will be recognized, the accuracy might be higher. In the case of computer, classroom, or library scenes, it is important to recognize images of all objects in the scene, as objects that are similar to each other may exist. Otherwise, the scene recognition accuracy may decrease slightly.

However, in this experiment, although the scene recognition accuracy of 95% and 93% in the 1st and 2nd experiments were obtained, if the experiment is performed in a more complex environment, the recognition accuracy is expected to be very low compared to the results of two experiments.

V. CONCLUSION

In this paper, an object-based scene recognition algorithm was proposed by applying the Mask R-CNN framework, and as a result of performing experiments, it was confirmed that the average object-based scene recognition accuracy was 95%. However, since there are many complex scenes in the real environment, more experiments and algorithm upgrades are constantly required to obtain the higher accuracy. In the future, we plan to upgrade the proposed algorithm by constructing a new data model and dataset suitable for various situations.

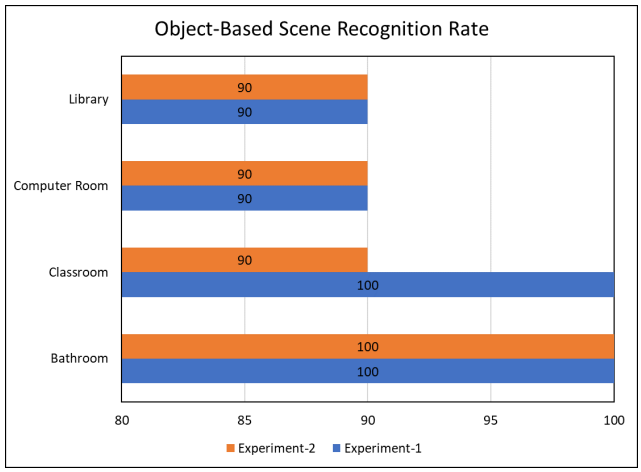

Fig. 4. Scene recognition accuracy on proposed algorithm.

ACKNOWLEDGMENT

This work was supported by the BB21+ Project in 2020.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2019R1F1A1062670).

REFERENCES

[1] B. A. Labinghisa, “A Study of Indoor Localization System using Reliable Deep Learning based Scene Recognition,” Thesis for Degree of Doctor of Engineering, The Graduate School of Tongmyong University, pp.1-87, Dec. 2020.

[2] S. Ren, K. He, R. Girshick and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” in Proc. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, Issue 39, pp.1137-1149, Jun. 2016.

[3] K. He, G. Gkioxari, P. Dollar and R. Girshick, “Mask r-CNN,” in Proc. IEEE International Conference on Computer Vision, pp.2961-2969, Mar. 2017.

[4] R. Girshick, “Fast r-cnn,” in Proc. IEEE International Conference on Computer Vision, pp.1440-1448. Sep. 2015.

[5] M. H. Thar and D.M. Lee, “A Study on Object-based Scene Recognition Algorithm using Mask Region-based Convolutional Neural Network,” in Proc. 31st Joint Conference on Communications and Information (JCCI 2021), pp.311-312. 28-30 Apr. 2021.