SUGISAKA MASANORI

Journal of Advances in Artificial Life Robotics
Vol. 3(4); March (2023), pp. 238–241
ON LINE ISSN 2435-8061; ISSN-L 2435-8061
https://alife-robotics.org/jallr.html

Research Article

Development of Beach Litter Detection System using Deep Learning on Beach Clean-up

Shintaro Ogawa[1], Tan Chi Jie[2], Takumi Tomokawa[2], Sylvain Geiser[2], Sakmongkon Chumkamon[2], Ayumu Tominaga[3], Eiji Hayashi[2]
[1]*Department of Creative Informatics, Kyushu Institute of Technology, 680-4, Kawazu, Iizuka-city, Fukuoka, 820-8502, Japan*
[2]*Department of Mechanical Information Science and Technology, Kyushu Institute of Technology, 680-4, Kawazu, Iizuka-city, Fukuoka, 820-8502, Japan*
[3]*Department of Creative Engineering Robotics and Mechatronics Course, National Institute of Technology Kitakyushu Colllege, 5-20-1 Shii, Kokuraminamiku, Kitakyushu-city, Fukuoka, 802-0985, Japan*

ARTICLE INFO

Article History

Received 02 December 2022
Accepted 20 July 2023

Keywords

Mask R-CNN
HTC
Field robotics
Object detection
Deep learning
TACO

ABSTRACT

This paper proposed a deep learning-based beach litter detector specifically designed for assessing litter levels on beaches effectively. This litter detector was created utilizing a HTC, also known as the Hybrid Task Cascade network, and its performance was compared to that of the traditional mask R-CNN network in order to judge its effectiveness. The findings uncovered that the HTC network possessed heightened sensitivity towards small and tiny litters taken within the RGB colored images.

© 2022 *The Author*. Published by Sugisaka Masanori at ALife Robotics Corporation Ltd.
This is an open access article distributed under the CC BY-NC 4.0 license
(http://creativecommons.org/licenses/by-nc/4.0/).

1. Introduction

Beach litter, a form of marine debris, has become a significant issue, necessitating manual surveys that are time-consuming. To address this, we aimed to develop a litter detection system that would minimize the labor required for such surveys. Creating an effective litter detector is crucial for conducting accurate and efficient litter assessments. The range of beach litter found on beaches varies from tiny objects like microplastics to larger objects such as fishing goods. Moreover, there are a lot of beaches that are challenging for humans to access. In this research, we designed a litter detector utilizing deep learning techniques for instance segmentation and object detection for images taken on the beach. This detector identifies and classifies beach litter on a pixel-by-pixel basis, accurately determining the type of litter present. We validated the detector's accuracy using a dataset comprising images captured during actual beach cleanup activities.

2. Debris detection system with HTC

In this research, we utilized a deep learning object detection model namely, Hybrid Task Cascade (HTC) for the purpose of detecting beach litter. HTC is a type of instance segmentation neural network model that combines two well-known models, the Cascade R-CNN and the Mask R-CNN. The Cascade R-CNN is an enhancement of the Faster R-CNN, a neural network that measures the extent of overlap between two regions using Intersection over Union (IoU) with a variable threshold (represented by Equation (1)). On the other hand, Mask

Corresponding author's E-mail: ogawa.shintaro553@mail.kyutech.jp, tan.jie-chi339@kyutech.ac.jp, m-san@mmcs.mse.kyutech.ac.jp tominaga@kct.ac.jp, haya@mse.kyutech.ac.jp URL: http://www.kyutech.ac.jp/

R-CNN is an object detection model that extends Faster R-CNN by incorporating segmentation capabilities. Therefore, HTC is formed by merging these two established models.

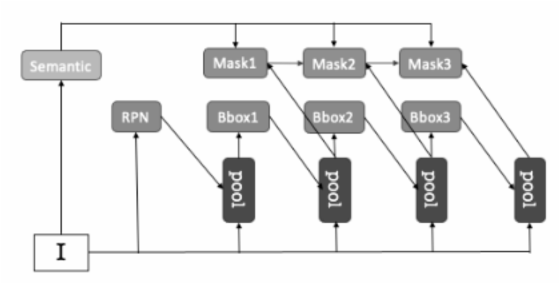

Fig. 1 HTC model structure

2.1. *Dataset*

The dataset utilized in this study was generated by merging a publicly available dataset of general trash called TACO [1] together with images that were captured by ourselves in Hokuto Mizukumi Park, Japan. TACO is a comprehensive dataset of various trash in the wild, encompassing a diverse range of locations beyond just beaches, comprising a total of 1,500 images. The latter portion of the dataset consists of 186 images specifically taken at authentic cleanup sites. Overall, the dataset comprises 1,686 images, with 1,636 images allocated for training purposes and 50 images designated for evaluation. It covers a total of 60 distinct classes. To facilitate a comparative analysis of detection accuracy across different datasets, we formulated and trained two separate datasets: one employing only the TACO dataset, and the other incorporating both TACO and images obtained from actual cleanup sites.

2.2. *Data annotation*

In supervised learning, the creation of accurate reference data, known as teacher data, is essential. This process involves mapping teacher labels to the respective target objects within images and is referred to as annotation. For this research, we employed the coco-annotator [2] tool to annotate each litter class, thereby generating the required teacher data. Figure 2 illustrates an example, with the left image representing the original image prior to annotation and the right image displaying the annotated version. In this particular example, a plastic drinking bottle is manually annotated and split into distinct parts, such as the bottle and the cap.

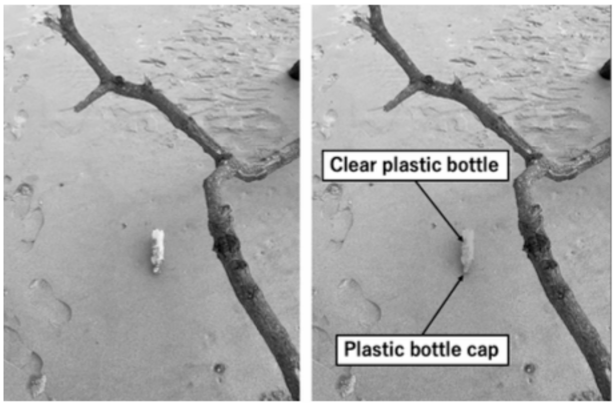

Fig. 2. Sample of Dataset

2.3. *Discriminator creation and methods of evaluation*

The discriminator was trained for a total of 2000 iterations, employing a learning rate of 0.002. The loss function used was multi-class cross-entropy. Besides of HTC, Mask R-CNN was also utilized during training process to enable real-time performance comparison. The discriminators were tested with untrained samples. To assess the performance of the discriminator, a set of 50 untrained images was prepared, and both training models were evaluated using these images. The accuracy was determined by evaluating the F-value, recall, and goodness of fit within the IoU range of 0.5 and above. In object detection, IoU serves as an accuracy measurement metric of the expected output region, while the F-value represents the harmonic mean of the goodness of fit and recall. Equations (1) to (3) provided below illustrate the formulas for IoU, precision ratio, and recall ratio.

$$IoU = \frac{Area\ of\ Intersection}{Area\ of\ Union} \tag{1}$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{2}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{3}$$

3. Experimental results

Figure 3.1 and Figure 3.2 show the discrimination result of different model trained using the datasets.

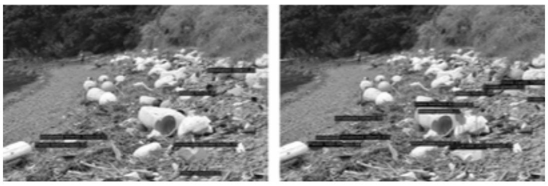

Fig. 3.1. Sample of discrimination Result
(Dataset: TACO, Left: HTC, Right: Mask R-CNN)

Fig. 3.2. Sample of discrimination Result
(Dataset: TACO+Real env, Left: HTC, Right: Mask R-CNN)

The experimental results were assessed based on the overall fit rate, F-value, and repeatability for the entire class. Additionally, the fit rate specifically for plastic drinking bottles was examined, as indicated in Table 1. Furthermore, Table 2 presents the label's precision performance of the plastic bottle class.

Table 1. Performance of trained model

	Precision	Recall	F-score
HTC(TACO)	0.034	0.074	0.047
Mask R-CNN(TACO)	0.025	0.03	0.027
HTC(TACO+Real env.)	0.041	0.053	0.046
Mask R-CNN (TACO+Real env.)	0.038	0.053	0.044

Table 2. Plastic bottle label's precision

	Precision
HTC(TACO)	0.164
Mask R-CNN(TACO)	0.127
HTC(TACO+Real env.)	0.141
Mask R-CNN (TACO+Real env.)	0.167

,,,Consideration

Although Table 1 indicates a relatively low precision score, the individual class analysis for plastic bottles in Table 2 reveals better precision results. Specifically, when considering the identification of PET bottles, Mask R-CNN demonstrates a higher conformance rate when real-world environment data is added, while HTC achieves a higher conformance rate without the inclusion of such data. Surprisingly, the overall F-measure for the class does not experience significant improvement with the addition of the dataset in the case of HTC. This raises the question of why HTC fails to yield satisfactory results even with the augmented dataset. One possible explanation is the compatibility between the HTC's model structure and the added real-world dataset. The real environment dataset possesses distinct characteristics, containing a greater abundance of litter, including numerous small objects, in comparison to the TACO dataset. As depicted in Figure 1, the model structure of HTC optimizes the features of the Region Proposal Network (RPN) through the pooling layer. Consequently, the feature optimization process may have inadvertently omitted the detection of small debris due to the introduction of the real-world dataset. Additionally, the classification of small objects is inherently challenging, and misclassification may be a contributing factor. To address these issues, several prospects can be explored. Firstly, increasing the input resolution of the model and elevating the resolution of input images can enhance the detection of small objects by allocating more pixels to them. Secondly, preventing annotation of excessively tiny garbage and minimizing the number of classes during the training data creation process may help mitigate misclassification. These approaches hold promise for improving the performance of HTC in detecting and classifying small litter objects.

4. Summary

The objective of this research was to utilize deep learning for the detection of debris at a beach cleanup site. Through our investigations, we discovered that the number of classes in the dataset plays a crucial role. Furthermore, we identified the need to develop effective techniques for detecting small objects within the dataset. Moving forward, our future endeavors will focus on enhancing the accuracy of debris detection. We aim to refine the detection accuracy and develop a comprehensive system that can be of greater utility in real-world cleaning scenarios. Our ultimate goal is to build a system that not only improves the detection accuracy but also enhances its practicality and usefulness in real beach cleaning sites.

References

1. "TACO Dataset ", http://tacodataset.org/
2. Justine Brooks, https://github.com/jsbroks/coco-annotator

Authors Introduction

Mr. Shintaro Ogawa

He received bachelor degree in Engineering in 2022 from intelligent and Control Systems, Kyushu Institute of Technology in Japan. He is currently a Master student at Kyushu Institute of Technology and conducts research at Hayashi Laboratory.

Dr. Sakmongkon

He received Doctor of Engineering degree from Kyushu Institute of Technology in 2017. He was a postdoctoral Guangdong Technology Presently he is a postdoctoral researcher in Kyushu Institute of Technology since 2019. His research interests include factory automation robots and social robots.

Projected Assist. Prof. Ayumu Tominaga

Projected Assist. Prof. Ayumu Tominaga is a professor in Department of Creative Engineering Robotics and Mechatronics Course at National Institute of Technology Kitakyushu College. He received the Ph.D. (Dr. Eng.) degree from Kyushu Institute of Technology in 2021. His research interests include Intelligent mechanics, Mechanical systems and Perceptual information processing

Mr. Takumi Tomokawa

He received bachelor degree in Engineering in 2021 from mechanical system engineering, Kyushu Institute of Technology in Japan. He is currently a Master student at Kyushu Institute of Technology and conducts research at Hayashi Laboratory.

Mr. Tan Chi Jie

He received his Bachelor of Engineering Electronics Majoring in Robotics and Automation from the Faculty of Engineering, Multimedia University, Malaysia in 2020. He is currently a Master student at Kyushu Institute of Technology and conducts research at Hayashi Laboratory.

Mr. Sylvain Geiser

He studied Engineering and Computer Science in Ecole des Mines de Nancy, France, between September 2018 and June 2020. He is currently a Master student at Kyushu Institute of Technology, Japan, and conducts research at Hayashi Laboratory.

Prof. Eiji Hayashi

Prof. Eiji Hayashi is a professor in the Department of Intelligent and Control Systems at Kyushu Institute of Technology. He received the Ph.D. (Dr. Eng.) degree from Waseda University in 1996. His research interests include Intelligent mechanics, Mechanical systems and Perceptual information processing. He is a member of The Institute of Electrical and Electronics Engineers (IEEE) and The Japan Society of Mechanical Engineers (JSME).