SUGISAKA MASANORI

Journal of Advances in Artificial Life Robotics
Vol. 4(1); June (2023), pp. 6–11
ON LINE ISSN 2435-8061; ISSN-L 2435-8061
https://alife-robotics.org/jallr.html

Research Article

A Real and Synthetic Dataset for Robotic Vision in Outdoor Beach Environment – BCRobo

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

ARTICLE INFO

Article History

Received 02 December 2022
Accepted 13 July 2023

Keywords
Synthetic dataset
Computer vision
Real dataset
Field robotics
Image segmentation

ABSTRACT

Datasets are one of the key elements which determine the performance of a deep learning network. Urban environments datasets receive much attention nowadays due to the rise of autonomous cars but off-road environment on the other hand lacks quality datasets. Offroad environments need equal attention as only 55% of the world's population lives in urban areas. This paper tackles this issue to close the gap of robotic visual perception on the beach, one of the common offroad environments that lack attention by presenting a real and synthetic dataset, namely BCRobo.

1. Introduction

Deep learning is blooming since the start of the 21st century especially in recent years with the generative network like GPT-4 getting popular. On the other hand, computer vision is still the center of deep learning networks as object detection and image segmentation are proven to be extremely useful in every industry. Not to mention that electric cars are getting more and more popular nowadays and slowly replacing the traditional gasoline car. Largely labeled datasets like CityScapes [1] and KITTI [2] dataset contribute to the development of autonomous driving by providing a huge learning dataset to train computer vision-related neural networks. Apart from datasets, optimizer algorithms, and learning policy also influence the performance of a neural network. However, regardless of how well-designed is a neural network, the models will always perform better in environments that are like its training dataset. In other words, trained models tend to skew toward their training dataset.

This is also known as the overfitting issue that all single deep-learning neural networks are facing now. For instance, an image segmentation model trained with urban datasets will have an outstanding performance in urban areas but in turn, have a corresponding substantial performance in rural areas. Depending on the situation, this property of a neural network might be useful because it is specialized in doing one thing at once. In other words, neural networks trade flexibility for accuracy in a specific territory. A simple workaround to tackle this problem of diversity in neural networks is to have different kinds of training datasets for different applications.

Corresponding author's E-mail: tan.jie-chi339@mail.kyutech.jp, m-san@mmcs.mse.kyutech.ac.jp, tominaga@kct.ac.jp, tomokawa.takumi163@mail.kyutech.jp, ogawa.shintaro553@mail.kyutech.jp, haya@mse.kyutech.ac.jp URL: http://www.kyutech.ac.jp/

The primary objective of this paper is to drive progress in image segmentation specifically within offroad settings, with a particular emphasis on beach environments. The exploration and availability of datasets in offroad environments, unlike urban environments, have received less attention, primarily due to the growing emphasis on autonomous driving cars. However, the advancement of autonomous robots in offroad settings, including forest and beach exploration, remains a significant focus for researchers in the field of robotics.

The paper commences by providing an overview of the sensor setup, dataset configuration, and the process of collecting the data. Subsequently, it presents the dataset's statistics and evaluates its performance using three cutting-edge image segmentation networks.

2. Sensor Setup

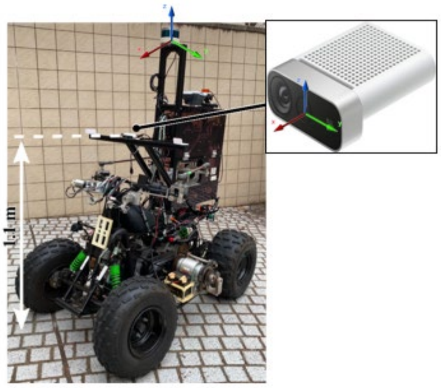

Figure 1: SOMA Robot’s Sensor Setup

Figure 1 illustrates the comprehensive sensor setup employed for capturing the BCRobo dataset. The study utilizes SOMA, an autonomous forest and beach exploration robot developed within Hayashi Laboratory [3]. Originally an All-Terrain Vehicle (ATV), SOMA underwent modifications to transform into an autonomous driving exploration robot. In addition to its innovative fully automated steering mechanism, the robot is equipped with various distinct sensor types, listed as follows:

2.1. *RGB-D sensor*

Positioned at around 1.1m above the ground, an RGB-D sensor is mounted in the frontal area of SOMA. The RGB-D camera employed is the Microsoft Azure Kinect DK [4]. The depth camera is set to operate in Wide Front of View (WFOV) mode, generating 1024x1024 depth images at a rate of 15 frames per second (fps). Simultaneously, the RGB camera is configured to capture a 1280x720 image stream in MJPEG format, maintaining a frame rate of 30fps.

2.2. *Global Positioning System (GPS)*

SOMA is also equipped with an Emlid Reach RS+ device, allowing for the inclusion of GPS data in the form of NMEA, also known as National Marine Electronics Association messages. The GPS data is acquired with a precision of up to +/- 5cm, utilizing Real-time Kinematic Positioning (RTK) technology.

2.3. *Lidar sensor*

As depicted in Figure 1, a Velodyne VLP-16 rotating 3D laser scanner is mounted on the upper part of SOMA. This sensor has the ability to capture 3D point clouds within a 360-degree field of view, covering a range of 1m to 100m. It operates at a rotation rate of 5Hz and maintains an accuracy of +/- 3cm [5].

3. Structure of Dataset

The BCRobo dataset comprises both real and synthetic beach environment data, including image sequences and LiDAR 3D point clouds recorded by SOMA Robot. The robot's operations were controlled manually by a human operator to travel and capture the scenes of various beaches around Kitakyushu City and Munakata City in Japan. In total, 6850 color images were recorded, and manually annotated ground truth images were provided for every tenth frame of a video sequence. In cases where the tenth frame appeared blurred, the preceding or subsequent frame may have been used as a substitute. The exploration data can be categorized based on the following locations:

- Jinoshima Island – An island in Munakata City with a port area and rock bed environment.
- Agawa Hosenguri Seaside Park – A typical beach for vacation and sea bathing.

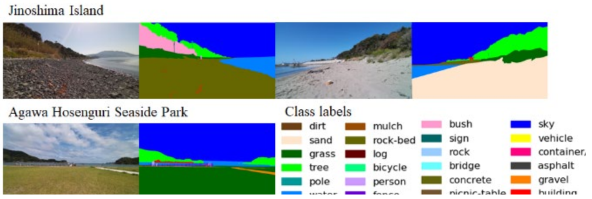

Figure 2: Class labels, Video sequences frame and annotated sample

Figure 2 depicts sample RGB images captured in Agawa Hosenguri Seaside Park and Jinoshima, along with their corresponding manually annotated images with 24 class labels. Approximately 10 minutes of data were recorded for each location, with a video frame rate of 15Hz as well as a LiDAR point cloud rate of 1Hz. It should be noted that not every single image in the video sequence is included in this dataset. Also, the annotated images within this dataset encompass 24 classes derived from the RUGD and KITTI datasets.

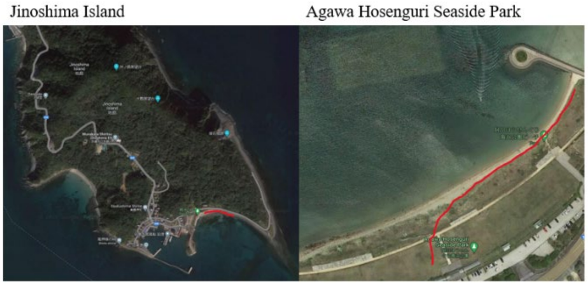

Figure 3: Data collection route predicted with GPS

Thanks to the RTK GPS sensor, the precise location of the robot is accurately tracked throughout the entire recording process. However, there were instances when the GPS connection with the base station was weak, leading to the prediction of robot routes using the available GPS coordinates, as illustrated in Figure 3. In Figure 3, SOMA is controlled manually according to the red route in a back-and-forth manner.

The synthetic part of this dataset shown in Figure 4 is generated through a beach environment simulated in Unreal Engine 4. An unreal plugin, UnrealCV [13] is utilized to perform automatic annotation which generates the annotated images based on the objects' material type. This dataset consists of 300 synthetic images and its corresponding annotated images which is approximately around 30% of the whole dataset.

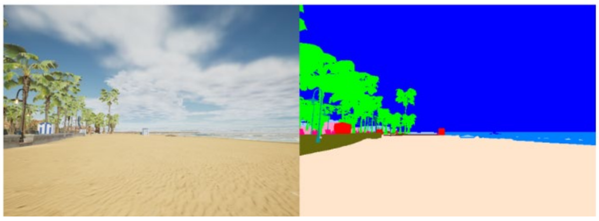

Figure 4: Synthetic RGB and Annotated Images

3.1. *Statistical Analysis of BCRobo Dataset*

The distribution of class annotations in the BCRobo dataset is illustrated in Figure 5. As anticipated, the dataset exhibits a bias towards the sky, grass, sand, and water labels. This outcome aligns with the dataset's specific purpose of addressing the scarcity of image segmentation datasets for beach environments. Given the requirement for autonomous robots to navigate accurately in beach environments, it becomes essential for them to distinguish between various types of traversable terrain such as sand, grass, and mulch, while identifying untraversable areas like water and rock beds.

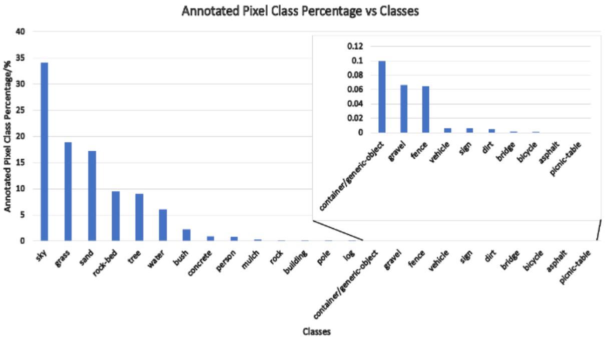

Figure 5: Annotated Class Pixel Percentages

4. Experiment and Evaluation

To assess the dataset's quality and practical applicability, three state-of-the-art semantic segmentation methods were chosen to be trained using the BCRobo dataset. Specifically, the selection of these three models was based on their ResNet50 backbone structure [6] within this experiment. ResNet50 was selected as the constant variable in this experiment due to its significance as the pioneering successful deep feedforward neural network. Despite being introduced as the earliest working model of its kind, ResNet50 continues to be widely employed as a backbone and remains one of the most frequently referenced neural networks in image segmentation methods since its victory in the ImageNet competition of 2015. These semantic segmentation approaches selected for this experiment are as follows:

- PSPnet [7] – ResNet50 – d8 backbone
- OCRnet [8] – ResNet50 – d8 backbone
- UPerNet [9] – ResNet50

PSPnet, an early image segmentation approach that incorporates global context information in scene parsing, emerged as the winner in the PASCAL VOC and Cityscapes benchmark back in 2016. The inclusion of global scene category analysis has proven to be valuable in intricate scene parsing scenarios, particularly in beach environments, where careful attention is required to distinguish various sub-regions containing significantly small or large objects.

Among the three selected approaches for this experiment, OCRnet stands out as the most recent one. Its notable feature is the capability to discriminate between contextual pixels belonging to the same object class and those belonging to different object classes. Furthermore, the utilization of dilated convolutions for multi-scale context enables OCRnet to leverage the high-resolution and extensive contextual information in this dataset.

In contrast, UPerNet aims to integrate multiple tasks such as texture recognition, object classification, pixel-level scene parsing, and scene recognition within a single neural network, utilizing an unprecedented learning approach namely unified perceptual parsing. Like PSPnet, UPerNet incorporates the Pyramid Pooling Module (PPM), applying one PPM for scene, object, part, and material recognition.

4.1. *Experimental Setup*

Similar to other semantic datasets, the recorded images in this experiment are divided into the train, validation, and test sets without considering the synthetic images generated using Unreal Engine. Specifically, 80% of the manually annotated images are allocated to the train set, while the remaining 20% is evenly distributed between the validation and test sets. Since the dataset encompasses two distinct beach sceneries, it is essential to ensure that the image segmentation models are trained equally with a similar splitting ratio. To achieve this, the aforementioned splitting ratio is applied separately to each beach environment, and the resulting subsets are combined as the final train set, validation set, and test set, as shown in Table 1.

Table 1. Train set, Validation set, and Test set.

	Jinoshima	Agawa	Total	%
	1. 31			
Train	5	233	548	80.00
Validation	39	30	69	10.07
Test	39	29	68	9.93

The training processes of this experiment are conducted in the environment as below:

- Nvidia RTX 3090 – 3 units
- AMD Ryzen Threadripper 3960X 24-Core
- MMSegmentation v0.29.1 [10]
- Ubuntu LTS 20.04

The images are initially resized to 688x550 before being fed into the deep learning network for training. A crop size of 300x375 is set for the training process. The batch size is configured at 6 per GPU, totaling 18 batches since 3 GPUs are utilized. For optimization, the Stochastic Gradient Descent (SGD) optimizer with momentum [11] is chosen, with a learning rate of 0.015 and a momentum of 0.9. A weight decay of 0.0004 is applied. To prevent overtraining at the beginning of the training process, the "Polynomial learning rate" policy with warmup is employed. This policy involves linearly increasing the learning rate for 1000 iterations before it reaches 0.015, followed by a polynomial decay until it reaches the minimum learning rate of 0.0001 for the entire training duration. Using MMSegmentation, the models are trained for 2000 epochs which are approximately 60000 iterations for all three models mentioned before.

4.2. *Result Evaluation*

The efficiency and performance of the models are assessed using the standard semantic segmentation metrics, including mean pixel-wise classification accuracy (mAcc) and mean Intersection-over-Union (mIoU). mIoU is computed as the mean Intersection-over-Union of each class, where IoU is defined as TP/(TP+FP+FN) [12]. Here, TP represents true positives, FP stands for false positives, and FN denotes false negatives. Additionally, mAcc is calculated as the mean pixel classification accuracy (aAcc) across all classes. This evaluation process involves passing the test and validation sets into the trained models, also known as the inferring process. The results are presented in Table 2. Furthermore, an additional evaluation is conducted on all three sets combined, as indicated in Table 3.

- Table 2. Testing on Test set + Validation set

	PSPnet, %	OCRnet, %	UPerNet, %
	2. 73		
mIoU	.90	74.64	75.34
mAcc	81.74	83.26	84.22
aAcc	98.09	98.06	97.83

- Table 3. Testing on the whole dataset including the training set

	PSPnet, %	OCRnet, %	UPerNet, %
	3. 71		
mIoU	.76	72.32	71.70
mAcc	78.22	79.71	79.14
aAcc	98.20	98.12	97.86

In general, the evaluation results indicate that regardless of the sets used, all models achieve a commendable mIoU rate above ~70%, demonstrating their ability to correctly learn the visual classes. The pixel-wise classification accuracy (aAcc) also exhibits high values, approximately ~98%. However, some degradation is observed in terms of mAcc. This degradation is likely attributed to the presence of irregular boundaries, which are common in beach environments due to factors such as fluctuating water tides, shifting sand formations, and the movement of tree branches and leaves caused by windy conditions.

5. Summary

In summary, the BCRobo dataset is a specialized collection of high-resolution images of beach environments recorded by the SOMA field exploration robot. Consequently, every image segmentation model that is trained with the BCRobo dataset would likely achieve better mIoU values if compared to other prominent datasets. This bias towards major class labels in the beach environments, including sky, water, sand, and trees, within the dataset contributes to this trend.

In conclusion, the experimental evaluation using PSPnet, OCRnet, and UPerNet demonstrates the effectiveness of this dataset for performing image segmentation in beach environments, with mIoU scores exceeding ~70%. However, it should be noted that the high mIoU achieved by these models implies that they may not perform as well in environments other than the beach. This limitation is a common trade-off in current neural network models, where accuracy often comes at the expense of diversity. Therefore, it is advisable to combine this dataset with others when conducting image segmentation in scenes that extend beyond beach environments.

Currently, the BCRobo dataset primarily focuses on beaches located in the southern region of Japan, specifically the Kyushu area. However, there are plans to further enhance the dataset by incorporating additional images and manually annotated images from various beach environments across other parts of Japan and Asia. The dataset can be accessed and downloaded from https://github.com/chijie1998/BCRobo-dataset, providing 685 real manually annotated images as well as 300 synthetic annotated images along with its original RGB images. Due to their large sizes, complete video sequences and point clouds are available upon request.

6. References

1. M.O.S.R.T.R.M.E.R.B.U.F.S.R.a.B.S.M. Cordts, "The cityscapes dataset for semantic urban scene understanding," in *IEEE Computer Vision and Pattern Recognition*, 2016.
2. P.L.C.S.a.R.U.A.Geiger, "Vision meets robotics: The kitti dataset," in *International Journal of Robotics Research*, 2013.
3. E.H.R.F.N. Takegami, "Environment map generation in the forest using field robot," in *Proceedings of International Symposium on Applied Science*, 2019.
4. S.M.B.T.T.E.S.W.O.A.A.P.J.G.M.F.V.R.e.a.C.S. Bamji, "Impixel 65nm bsi 320mhz demodulated of image sensor with 3μm global shutter pixels and analog binning", in *IEEE International Solid-State Circuits Conference - (ISSCC)*, 2018.
5. J. R. Kidd, "Performance Evaluation of the Velodyne VLP-16 System for Surface Feature Surveying," University of New Hampshire, 2017.
6. X.Z.S.R.a.J.S.K.He, "Deep residual learning for image recognition.," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
7. J.S.X.Q.X.W.J.J. Hengshuang Zhao, "Pyramid Scene Parsing Network," in *Computer Vision and Pattern Recognition*, 2017.
8. A.G.N.A.a.J.G.V. Gupta, "OCRNet - Light-weighted and Efficient Neural Network for Optical Character Recognition," in *IEEE Bombay Section Signature Conference (IBSSC) 2021*, 2021.
9. T. a. L. Y. a. Z. B. a. J. Y. a. S. J. Xiao, "Unified perceptual parsing for scene understanding," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.

10. M. Contributors, "OpenMMLab Semantic Segmentation Toolbox and Benchmark," 10 7 2020. [Online]. Available: https://github.com/open-mmlab/mmsegmentation
11. Y. G. Y. Yanli Liu, "An Improved Analysis of Stochastic Gradient Descent," in *34th Conference on Neural Information Processing Systems (NeurIPS)*, Vancouver, Canada, 202.
12. A. Rosebrock, "Intersection over Union (IoU) for object detection," pyiamgeserach, 7 Novemeber 2016. [Online]. Available: https://pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/
13. Weichao Qiu, "UnrealCV: Virtual Worlds for Computer Vision," pyiamgeserach, 27 October 2017. [Online]. Available: https://dl.acm.org/doi/pdf/10.1145/3123266.3129396

Authors Introduction

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's student at Kyushu Institute of Technology and conducts research at Hayashi Laboratory.

Mr. Takumi Tomokawa

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

Mr. Shintaro Ogawa

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

Projected Assist. Prof. Ayumu Tominaga

He is a professor in the Department of Creative Engineering Robotics and Mechatronics Course at the 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.

Dr. Sakmongkon Chumkamon

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

Prof. Eiji Hayashi

He 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).