The BCRobo dataset for Robotic Vision and Autonomous Path Planning in Outdoor Beach Environment

Tan Chi Jie
Department of Mechanical Information Science and Technology, Kyushu Institute of Technology
680-4, Kawazu, Iizuka-City, Fukuoka, 820-8502, Japan

Takumi Tomokawa[1], Sylvain Geiser[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 Colllege, 5-20-1 Shii, Kokuraminamiku, Kitakyushu, Fukuoka, 802-0985, Japan

E-mail: tan.jie-chi339@mail.kyutech.jp, m-san@mmcs.mse.kyutech.ac.jp, tominaga@kct.ac.jp, tomokawa.takumi163@mail.kyutech.jp, geiser.nathan-sylvain@mail.kyutech.jp, ogawa.shintaro553@mail.kyutech.jp, haya@mse.kyutech.ac.jp,
http://www.kyutech.ac.jp/

Abstract

Along with the universalization of autonomous driving and image segmentation, various datasets are available freely for anyone to use to train their own neural network which speeds up the growth of deep learning technology. However, most of the datasets target only urban environments and other offroad environments are still lacking in datasets. This paper presents a beach environment dataset, BCRobo with the aim to contribute to closing the gap of robotic visual perception in offroad environment, especially in beach.

Keywords: Dataset, Image segmentation, Deep Learning, Field Robotics, Computer Vision

1. Introduction

Deep learning is advancing at an extremely fast pace, especially in the field of image segmentation and object detection. This is because of the curation of largely labeled datasets such as CityScapes [1] and KITTI [2] dataset targeted to push forward the development of autonomous driving using computer vision and neural network. As said, dataset plays a huge part in training a neural network for deep learning regardless of how good or efficient is the algorithm, the characteristic of the trained network mainly depends on the dataset itself. This is the sole reason that there are a few varieties in the types of datasets such as TACO that focus on garbage detection, COCO dataset for indoor detection and KITTI or Cityscapes for outdoor detection.

The current challenge of deep learning is that every neural network tends to become less flexible after they are trained. Regardless of the problem of overfitting, neural networks are usually specialized at one thing only such as a network that is good in indoor detection will in

turn have a substandard performance in outdoor detection. This is something that no neural network could escape from. In simpler words, with the tradeoff of flexibility, neural networks gain accuracy in specific territory. Although the problem of diversity in neural network is yet to be tackled but a simple workaround is just to have several types of datasets matching its applications.

This paper aims to push forward the advancement of image segmentation in offroad environment especially in beach environment. Offroad environment is one of the environments that still lacks exploration and dataset as urban environment is given much more focus due to the recent surge of autonomous driving cars. Nevertheless, the development of autonomous robots in offroad environments such as forest and beach exploration remain as one of the main targets of robotic researchers.

This paper begins with the description of the sensor setup, dataset setup and collection, statistics and evaluation of the dataset using 3 types of current state-of-art image segmentation network.

2. Sensor Setup

Figure 1: SOMA Sensor setup

Figure 1 shows the overall sensor setup used to capture BCRobo dataset. An autonomous forest and beach exploration robot that was built in Hayashi Laboratory, named SOMA [3] is used. SOMA is originally an All-Terrain Vehicle (ATV) which was then modified to be an autonomous driving exploration robot. On top of a creative fully automated steering mechanism, the robot is also equipped with distinct types of sensors as below:

2.1. *RGB-D sensor*

At the height of approximately 1.1m from the ground, an RGB-D sensor is installed in front of SOMA. The RGB-D camera used is Azure Kinect DK produced by Microsoft [4]. Wide Front of View (WFOV) mode is chosen for the depth camera which produces 1024x1024 depth images at the rate of 15 frames per second (fps). On the other hand, the color camera is set to produce 1280x720 MJPEG video stream at 30fps.

2.2. *Lidar sensor*

A Velodyne VLP-16 rotating 3D laser scanner is installed on top of SOMA as shown in Figure 1. This sensor is capable of recording 3D point clouds with a field of view of 360 degrees at the range from 1m to 100m. The rotation rate and accuracy are set at 5Hz and +/- 3cm [5].

2.3. *Global Positioning System (GPS)*

Emlid Reach RS+ is installed on SOMA as well which means that GPS data is also included in the form of National Marine Electronics Association (NMEA) messages. The GPS is taken with the precision up to +/- 5cm with Real-time Kinematic Positioning (RTK).

3. Dataset setup

BCRobo dataset is a beach environment dataset that consists of video sequences and lidar point clouds that were recorded by SOMA in Figure 1. The operation of the robot is manually controlled by human to explore and record the environment of several beaches in Munakata City and Kitakyushu City, Japan. A total of 6850 video frames are captured while ground truth annotations are provided for every tenth frame of a video sequence. If the tenth frame is blurred, the frame before or after might be used instead. The exploration data could be categorized based on the location below:

- 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.

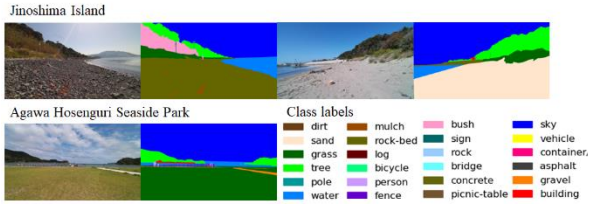

Figure 2: Example of video sequences frame and its corresponding ground truth image with class labels

Figure 2 shows the sample images taken in Jinoshima and Agawa Hosenguri Seaside Park with its corresponding ground truth image. For every location, approximately 10 minutes of data are recorded where the video frame rate is 15Hz and lidar rate is 1Hz. Note that not all video sequences frames are included in this dataset. SOMA moves at 1.6m/s at all times except during start up or climbing across some obstacles. The ground truth images in this dataset consists of 24 classes derived from KITTI and RUGD dataset.

Figure 3: Predicted route of SOMA based on GPS data

With the RTK GPS sensor, we are able to pinpoint the location of the robot throughout the whole recording but there are times with bad connection with the base of the GPS and hence, the routes taken by the robot are predicted using the GPS coordinates as shown in Figure 3. SOMA is controlled and follows the red route in Figure 3 back and forth.

3.1. *BCRobo Class Labels and Statistical Analysis*

Figure 4 shows the breakdown of class annotations for BCRobo dataset. As expected, this dataset is skewed towards the sky, grass, sand and water labels. This is the expected outcome as this dataset is made specially to tackle the issue of lack of dataset for image segmentation in beach environment. For autonomous robot to transverse in a beach environment, the robot needs to have the ability to recognize the terrain accurately which is to differentiate between the traversable terrain such as sand, grass, gravel and mulch with the untraversable terrain like water and rock bed depending on the dynamic ability of the robot whether it is a wheeled or tracked

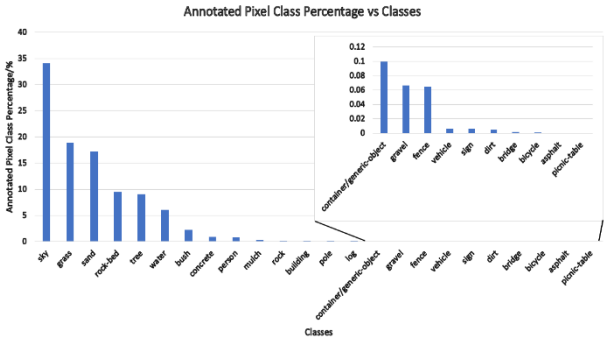

Figure 4: Annotated Class Pixel Percentages

robot.

4. Experiment and Evaluation

In order to evaluate the quality and actual usage of the dataset, three types of state-of-the-art semantic segmentation approaches are selected to be trained with the BCRobo dataset. In this experiment, the three selected semantic segmentation models are chosen based on their backbone structure which is ResNet50 [6]. The reason that ResNet50 is chosen as the fixed variable in this experiment is that despite being the very first working very deep feedforward neural network, ResNet50 still remains as the most used backbone and one of the most cited neural networks in image segmentation approaches since winning the ImageNet competition in 2015. The three selected semantic segmentation approaches are:

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

PSPnet is one of the earliest approaches to include global context information in scene parsing for image segmentation which eventually become the winner for PASCAL VOC and Cityscapes benchmark in 2016. This ability of examining global scene category is proven to be useful in complex scene parsing scenarios especially in beach environments where much attention is needed to

different sub-regions that contain some remarkably small or large objects.

OCRnet is the most recent approach among the three approaches chosen for this experiment due to its ability to differentiate not only the same-object-class contextual pixels but also the different-object-class contextual pixels. In addition, the multi-scale context using dilated convolutions also benefits from high-resolution, large-scale contexts of this dataset.

On the other hand, UPerNet is trying to combine object classification, scene recognition, pixel-level scene parsing and texture recognition in a single neural network along with a novel learning method. Having the same Pyramid Pooling Module (PPM) as PSPnet, UPerNet applies one PPM on scene, object, part and material recognition each to achieve the said unified perceptual parsing.

4.1. *Experimental Setup*

Like every other semantic dataset, the video frames are separated into train, validation and test sets for this experiment. 80% of the annotated ground truth video frames are partitioned into train set while the leftover 20% split evenly between validation and test sets. This dataset includes two different beach environments and to ensure that the semantic segmentation models are trained on both environments, the splitting ratio mentioned before this is applied on the two beach environments individually and then combined to form the final train, validation and test sets as in Table 1.

Table 1. Train, Validation and Test sets.

	Jinoshima	Agawa	Total	%
Train	315	233	548	80.00
Validation	39	30	69	10.07
Test	39	29	68	9.93

The training environment for all three models are setup as below:

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

The images are first downsized to 688x550 before passing into the neural network for training and the crop size is set at 300x375. Batch size is set at 6 per GPU making it 18 since 3 GPUs are used. Stochastic Gradient Descent (SGD) optimizer with momentum [11] is selected with the parameter of learning rate 0.015 and momentum 0.9. The weight decay is set at 0.0004. "Polynomial learning rate" policy with warmup is chosen to avoid over training at the start of training. In other words, the learning rate will increase linearly for 1000 iterations until it reaches 0.015 then decay in a polynomial fashion until it reaches the minimum learning of 0.0001 for the whole training. The modals are trained for 2000 epochs which is around 60000 iterations using MMSegmentation which is an open-source semantic segmentation toolbox.

4.2. *Experimental Evaluation*

The performance of all three models is evaluated using the standard semantic segmentation metrics which are mean Intersection-over-Union (mIoU) and mean pixel-wise classification accuracy (mAcc). mIoU is the mean of IoU of each class given that, IoU is TP/(TP+FP+FN) [12] where TP is true positive, FP is false positive, and FN is false negative. Besides, mAcc is just the mean of pixel classification accuracy (aAcc) of all classes. The evaluation is first performed by inferring the test and validation sets using the trained modals as shown in Table 2. Another evaluation is also performed on all three sets as well as shown in Table 3.

Table 2. Evaluation on Test + Validation sets

	PSPnet, %	OCRnet, %	UPerNet, %
mIoU	73.90	74.64	75.34
mAcc	81.74	83.26	84.22
aAcc	98.09	98.06	97.83

Table 3. Evaluation on all sets

	PSPnet, %	OCRnet, %	UPerNet, %
mIoU	71.76	72.32	71.70
mAcc	78.22	79.71	79.14
aAcc	98.20	98.12	97.86

Overall, it is observed that despite the sets used in evaluation, mIoU for all modals achieves a reasonably good rate above ~70% which means that all the models are learning the visual classes correctly. The aAcc are high as well around ~98% but we do observe some degree of degrade at mAcc. This is probably due to the irregular boundaries which is very common in beach

environments due to the constant changing water tide, shape of sands, changing position of tree branches and leaves due to the windy condition in beach.

5. Conclusion and future work

In the nutshell, BCRobo dataset is a highly specialized dataset that contains high resolution beach environment images captured by a field exploration robot, SOMA. For this reason, any image segmentation model trained with this dataset would expect higher mIoU compared to other major dataset as this dataset is skew towards the major class labels which is sky, sand, water and tree.

As a conclusion, this dataset is proven helpful in performance image segmentation for beach environment from the experimental data evaluated using PSPnet, OCRnet and UPerNet that achieve over ~70% mIoU. However, the high mIoU would also mean that the models trained using this dataset might perform badly in other environments besides beach. Nevertheless, this is the common downside of current neural network models which lose diversity as they gain accuracy. Therefore, it is recommended to use this dataset with other dataset if the scene for image segmentation is not limited to beach only.

As for now, BCRobo dataset is focus only on beaches around the Southern part of Japan, namely Kyushu area. In the future, we will continue to expand this dataset by adding more images and its annotated ground truth images of different beach environments in other part of Japan. The dataset is available for download at https://github.com/chijie1998/BCRobo-dataset with the 685 annotated ground truth images and its corresponding video frames. Full video sequences and lidar point cloud would be provided upon request due to its large sizes.

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 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 tof 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/

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

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. 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.

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.

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.

Dr. Sakmongkon Chumkamon

Dr. Sakmongkon Chumkamon received Doctor of Engineering degree from 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 in Kyushu Institute of Technology since 2019. His research interests include factory automation robots and social robots.

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).