Object Status Detection in Cluttered Environment for Robot Grasping Using Mask-RCNN

Kasman
Graduated School of Creative Informatics, Kyushu Institute of Technology
Castail Iizuka 318, 268-1, Kawazu, Iizuka-City, Fukuoka, 820-0067, Japan

Eiji Hayashi
Department of Mechanical Information Science and Technology, Kyushu Institute of Technology
680-4, Kawazu, Iizuka-City, Fukuoka, 820-8502, Japan
E-mail: kasman@mmcs.mse.kyutech.ac.jp, haya@mse.kyutech.ac.jp, kasman@umi.ac.id
www.kyutech.ac.jp

Abstract

Detecting object status in cluttered manipulator's robot environment before grasping is quite challenging to recognize the target because of unstructured and uncertainty scenes. Using Mask R-CNN for detecting the status of the object i.e. free for picking, close, overlapping and piling to the other objects is very useful as computer vision before the manipulator doing next procedures to complete its task. This paper provides a systematic summary and analysis target detecting and recognizing object status using Mask R-CNN. Unlike related solution methods that use machine vision and deep learning directly and combine together for doing robot controlling, pushing and grasping, we are doing image processing separately and simply for detecting the object's status before performing like pushing the object for making free and easy grasping. Experiment with this method shows that it has good accuracy, easy to implement for detecting and classification using the algorithm.

Keywords: object detection, Mask RCNN, cluttered environment

1. Introduction

Object detection and classification for robotic applications has emerged as an important goal because object images in cluttered environment provide basic information for several natural image application. object detection should determine the features in image that is contain multi-objectives complex problem considering to classification and localization single or multi-object in image. One of method to approve object detection is using Mask-RCNN. Mask R-CNN is an extension of the Faster R-CNN model for object detection, localization and instance segmentation in natural images.

There are some paper [1], [2], [3] using object classification in cluttered environment combined using grasping strategies to compensate perception error distance from the obstacle or free space around the object, using single perspective [3] and dual perspectives [4] camera for classifying the object in the cluttered environment embedded with reinforcement learning for doing pushing and grasping together with. There are also many research applications in machine learning successful to use Mask RCNN for classifying and identifying the target aims like in smart farming [5], [6] in medical purposes [10], [11], , navigations and remote sensing [8], [9].

This proposed system aims to implement object detection and classification using deep learning Mask RCNN to classify the object status like free or not in the cluttered or uncertainty environment applying in robot manipulation application and then masking the target objects with the boundary mask for determining the borderline mask of the target status. From the propose system we can mark the status of the object ie. complimentary from the other objects or obstacle or not that means the object is close to the other objects or the object over stack to the other so that the object will appear overlapping boundary in images. Output from this propose can be used in manipulator robot to do pick, push or sliding the object based on the object status in advance.

2. Proposed Method

In this paper we introduce an approach called Mask R-CNN which is an extension of Faster RCNN, Faster R-CNN yields two branches class name and bounding box. Mask R-CNN adds extra branch of mask along with the class name and the bounding box

Mask-RCNN [12] was popularized by He et al. in 2018 as a development of Faster RCNN to confess an accurate pixel-based segmentation. This method includes two main steps namely: Feature Pyramid Network (FPN) and Region Proposal Network (RPN). In the feature pyramid network, a various number of proposals was produced about the regions where there might be the background like an object different from the background based on the input image.

In this paper we obtain instance segmentation using Mask R-CNN for masking the object and object status. It consists of (CNN) convolutional backbone, which is a pretrained model like VGGNET,AlexNet, GoogleNet, ResNet etc. Any model we use for masking using RCNN will involve some block like RPN, ROI and some network layers. First, RPN is a Regional Proposal Network which is needed to generate a region of proposal for the given input image. It will produce the feature map of the image. A ROI Align Layer is used for generating fixed size of feature map. A Fully Connected Layer exists for the classification of objects in the image with the bounding box. A Mask Branch is used to generate the mask to the identified object by the fully connected Layer.

In this paper, we used a pre-trained architecture on Coco (80 class) dataset but we customized to 10 classes residue. Generally, the size of the recent model is substantially smaller due to the usage of global average pooling rather than fully-connected layers. We choose ResNet50 as a feature extractor network which encodes input image into 32x32x2048 feature map. The FPN (feature pyramid network) extracts regions of interest from features of different levels according to the size of the feature which feeds as input to Next stage (RPN). In Region Proposal Network (RPN), the regions scanned individually and predicted whether or not an object is present. The actual input image is never scanned by RPN instead RPN network scans the feature map, making it much faster. Next, each of regions of interest proposed by the RPN as inputs and outputs a classification (SoftMax) and a bounding box (regressor). Finally, Mask- RCNN adds a new branch to output a binary mask that indicates whether the given pixel is or not part of an object. This added branch is a Fully Convolutional Network on top of the backbone architecture. The proposed method consists of two main steps: Training and testing steps as illustrated in Figure 1.

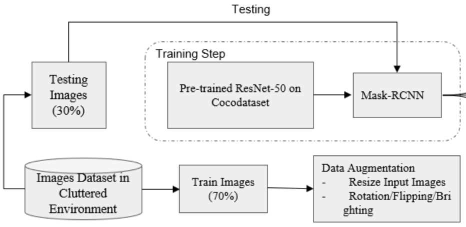


Figure 1. The proposed Status Object detection Architecture

.

The below steps will show the working of the Mask R-CNN

- The input image is given to the Mask R- CNN then the pertained model will give feature map as the output to the RPN.
- The RPN (Regional Proposal Network) will take output of the CNN and gives the multiple regions of interest using light weight binary classifiers.
- The feature Map along with the ROI of the image is given to the ROI Align layer which gives the fixed size feature map as the output, by wrapping the multiple bounded boxes into fixed dimensions.
- Then the fixed size feature map from the ROI Align is given to the fully connected layers and Mask Branch.
- The fully connected Layer will classify the objects in the image with bounded boxes.
- Then the mask branch outputs a binary mask for each ROI of the image.

2.1. Loss Function

Mask R-CNN utilized a multi-task loss function that combined the loss of classification, localization and segmentation mask as illustrated in Eq. (1).

$$L = L_{cls} + L_{bbox} + L_{mask} \quad (1)$$

Where L_{cls} , L_{bbox} are same as in Faster R-CNN [13]. The added mask L_{mask} illustrated in Eq. (2). as the average binary cross-entropy that only includes *k th* mask if the region is associated with the ground truth class k .

$$L_{mask} = -\frac{1}{m^2}\sum_{lsi,j\,sm} y_{ij} \log \hat{y}_{ij}^{k} + (1 - y_{ij})\log(1 - \hat{y}_{ij}^{k}) \quad (2)$$

Where the mask branch generates a mask of dimension m x m for each *RoI* and each class y_{ij} and *k, ŷ, ij* are cell *(i, j)* label of the true mask and the predicted value respectively.

2.2. Training Step

Mask-RCNN requires a large number of annotated data for training to cover up the overfitting. To overcome the problem of limited annotated dataset in cluttered environment in robot application, we adopted transfer learning by selected the pre-trained network weights of the resnet50 model, which was successfully trained with the coco dataset. We utilized the pre-trained resnet50 and fine-tuned the network weights to the Coco dataset. we use 30 epochs with learning rate 0.001. We used different augmentation methods such as horizontal flip, vertical flip, image rotation, and image translation to enlarge the training data. One can observe that this domain-specific fine-tuning allows learning good network weights for a high-capacity CNN for coco dataset.

2.3 Testing step

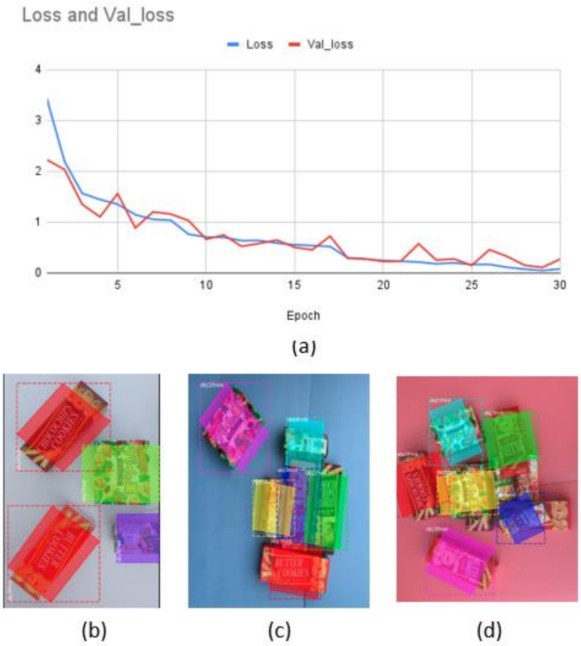
The learned model used directly to predict class label, boundary box, and masked segment for each image in testing data. To evaluate the learned model performance, the predicted labels and boundary box is matched with those in the dataset.

3. Result and Discussion

This section illustrates the results of the target status detection and classification on the cluttered environment. Figure 2 depicts the output of Mask RCNN algorithm in fig. (a) output loss and validation loss recorded in 30 epochs while doing training step, (b) target status detection is free, and not free (c) because close and the distance below the threshold and (d) the object detection overlapping to the other objects. The target status from the picture shows the segmentation of the mask for each object detection and the boundary line for detecting how close the object to the other objects

4. Conclusion

The proposed model will overcome the limitations of the object detection for the status of target in the cluttered environment which is implemented in robot manipulator. In the cluttered environment there are many objects that are close each other, pilled to the other up. the proposed method can determine and classify the status of the target weather free or not. The boundary line along with to each object indicate the limitation and the border threshold for each object. when the target is too close each other, the boundary line of the objects in the mask will cross. If the target is classified free, the manipulator robot can do pick the object from target place and places to the destination place. nonetheless if the status of object is not free the boundary line will show cross and the cross will indicate how close the object to the other and the point which is the manipulator can do next step like pushing or separating the object from the other so that the robot can do grasping in advanced.



(a)

(b) (c) (d)

Figure 2. The output of Mask RCNN.
(a). Loss and Validation Loss
(b), (c), (d). Object status detection and Classification.

References

1. I. Sarantopoulos, K. Marios, “Total Singulation with Modular Reinforcement Learning”, IEEE Robotics and Automation Letters, Vol. 6, No. 2, April 2021.
2. I. Sarantopoulos, K. Marios, “Split Deep Q-Learning for Robust Object Singulation|”, IEEE International Conference on Robotics and Automation (ICRA), 31 May – 31 August 2020 Paris France.
3. K. Marios, M Sotiris “Robust Object Grasping in clutter via singulation|”, IEEE International Conference on Robotics and Automation (ICRA), 31 May – 31 August 2020 Paris France.
4. P. Gan, L. Jinhu, G. Shangbin, “A Pushing-Grasping Collaborative Method Based On Deep Q-Network Algorithm in Dual Perspectives”,
5. T. Pallpothu, M. Singh, Riya S, “Cotton leaf disease detection using mask RCNN”, AIP Conference Proceedings, May 2022.
6. T. Pallpothu, M. Singh, Riya S, “Cotton leaf disease detection using mask RCNN”, AIP Conference Proceedings, May 2022.

7. C. Hsien Hsia, T. William Chang, C. Chiang." Mask R-CNN with New Data Augmentation Features for Smart Detection of Retail Products", MDPI Journals, March 2022, Volume 12 Issue 6.
8. Y. Gan, S. You, Z. Luo, K. Liu, T. Zhang, "Object Detection in Remote Sensing Images with Mask R-CNN", Journal of Physics: Conference series, Volume 1673, International Confrecence on Computer Science and Applications 25-27 September 2020, China.
9. K. Zhao, J. Kang, J. Jung, and G.Soh, "Object Detection and Instance Segmentation in Remote Sensing Imagery Based on Precise Mask R-CNN", In: Proc. of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp.247-251, 2018.
10. T. Padma, C. Usha Kumari, D. Yamini, K. Pravalika, "Image Segmentation Using Mask R-CNN for Tumor Detection from Medical Images", International Conference on Electronics and Renewable Systems (ICEARS), March 2022.
11. K. Zhao, J. Kang, J. Jung, and G. Soh, " An Improved Mask R-CNN Model for Multiorgan Segmentation", In: Proc. Of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 247-251, 2018
12. Kaiming He, Gergia G, Piots D, Ross G, "Mask R-CNN", Computer Vision and Pattern Recognition, 2017

Authors Introduction

Mr. Kasman

He received his Bachelor and Master Degree from the Department of Electrical Engineering, Moslem University of Indonesia 1998 and Sepuluh Nopember Surabaya. Indonesia 2009, respectively. Currently, he is a student of Doctoral Degree Program in Kyushu Institute of Technology, Japan

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).