Recognition of Plastic Bottles Region Using Improved DeepLab v3+

Yusuke Murata
Department of Mechanical and Control Engineering, Kyushu Institute of Technology,
1-1 Sensui-cho, Tobata-ku, Kitakyushu, Fukuoka, 804-0015, Japan

Tohru Kamiya
Department of Mechanical and Control Engineering, Kyushu Institute of Technology,
1-1 Sensui-cho, Tobata-ku, Kitakyushu, Fukuoka, 804-0015, Japan
E-mail: murata.yusuke973@mail.kyutech.jp, kamiya@cntl.kyutech.jp

Abstract

Factory automation is one solution to the labor shortage. We focus on the sorting of plastic bottles in waste disposal plants and try to automate the process using robotic arms. In this paper, we propose an image analysis method for the recognition of plastic bottles limited to 500ml capacity. The method is semantic segmentation, and the deep learning model is DeepLab v3+. Modifications using ECA Block and Mish function show improvements at the points of misrecognition with the base model.

Keywords: Deep Learning, Semantic Segmentation, Convolutional Neural Network (CNN), DeepLab v3+, Efficient Channel Attention Block (ECA Block), Mish function

1. Introduction

The labor shortage in Japan is becoming more serious every year [1]. This study focuses on labor shortage in factories. The causes include the decline in the labor force due to the low birth rate and aging population, and the negative image of repetitive and simple work. As solutions to these problems, the employment of foreign and female workers, the training of workers who can perform multiple tasks, and the enhancement of the factory's image can be considered. However, this study focuses on factory automation to take advantage of technology.

Factory automation refers to the use of machines to replace tasks traditionally performed by humans. Many factories have already automated some of their equipment, and an increasing number of factories are investing in equipment with more advanced technologies. Factory automation is expected to not only solve the labor shortage, but also reduce the labor cost, stabilize the product quality, increase the operation rate, and improve the work environment.

This study focuses on the sorting of plastic bottles at waste treatment facilities. The processes required for recycling vary from material to material, and each resource must be sorted. Currently, cans are sorted using magnets, but plastic bottles are sorted manually. To automate this process, we are trying a method to recognize plastic bottles from images and pick them up with a robotic arm. In this study, we propose an image analysis method specifically for 500ml plastic bottles. As basic research, we conducted experiments to investigate how accurately individual plastic bottles can be recognized in images. Semantic segmentation was used as the method, and DeepLab v3+ [2], which is a powerful deep learning model in this field, was employed. As the backbone, we used ResNet-101 [3], and evaluated the performance of each model with improvements by combining the Efficient Channel Attention Block (ECA Block) [4] and the Mish function [5].

2. Method

In this section, we describe an image analysis method for the recognition of plastic bottles. The overall flow is to collect input images and manually annotate them. Then, training and inference are performed on this dataset using a deep learning model. In this way, plastic bottles separated into their bodies and caps can be recognized from unknown images. Semantic segmentation is used as the recognition method, and the details of the deep learning model with this method are described below.

2.1. *DeepLab v3+*

We used DeepLab v3+ [2] as our deep learning model. DeepLab v3+ is one of the best performing Convolutional Neural Networks (CNNs) for semantic segmentation, which was proposed in 2018. There are two main CNN structures used in this method. The first is a network that uses Spatial Pyramid Pooling (SPP) [6], and the second is a network that consists of encoders and decoders. DeepLab v3+ combines these two structures to obtain a wide range of information and sharp object contours, and achieves high performance.

In this study, the body and cap of a plastic bottle are recognized by semantic segmentation. In particular, the cap is small compared to the body and needs to be recognized accurately. DeepLab v3+ enables highly accurate recognition of small objects due to the advantages described above, and this is the reason why DeepLab v3+ was chosen.

2.2. *ResNet-101*

We adopted ResNet-101 [3] as the backbone of DeepLab v3+. ResNet is a CNN that enables the construction of deep layers by introducing Residual Blocks, which was proposed in 2015. The performance of CNNs improves with the depth of the layers because they extract higher-dimensional global features. However, simply increasing the depth of the layers leads to poor performance due to the degradation problem, etc. To solve this problem, the network that introduces the skip connection is ResNet.

ResNet is divided into five types according to the number of residual blocks. In this study, we chose ResNet-101 to avoid overlearning due to the complexity of the network while incorporating global information.

2.3. *ECA Block*

ECA Block [4] is a channel-directed attention mechanism proposed in 2019. Channel-directed attention mechanisms emphasize important information and have been shown to improve the performance of deep CNNs. The advantage of the ECA Block is that it effectively captures inter-channel dependencies while keeping the complexity of the mechanism low. This is achieved by incorporating a single 1D convolutional layer instead of multiple fully connected layers. Fig. 1. shows the structure of ECA Block.

In this study, the goal is to automate simple tasks in factories, and it is necessary to improve the recognition accuracy while keeping the processing speed low. ECA Block is suitable to achieve this. Because it emphasizes important information without increasing the complexity of the model. We believe that this characteristic improves the overall recognition rate without increasing the computation time.

2.4. *Mish function*

Mish function [5] is an activation function proposed in 2019, and it is represented by Eq. (1).

$$f(x) = x \tanh(\ln(1 + e^x)) \quad (1)$$

The Mish function has the property that its output changes smoothly and continuously with respect to the input and is differentiable over the entire range. This property improves gradient stability and mitigates problems such as gradient loss and gradient explosion during the learning process. In addition, The Mish function can smoothly use information for negative-valued inputs, thus capturing a wider variety of features than activation functions that completely truncate negative values to zero. Because of these properties, The Mish function is effective for detecting fine patterns in images, and its performance improvement has been reported in various fields.

In this study, plastic bottles are recognized by dividing them into the body and the cap. Since the cap is smaller than the body, it is necessary for the model to accurately recognize small features. Therefore, we believe that the introduction of the MIsh function is effective.

2.5. *Proposed network*

In this study, we used DeepLab v3+ as the base model and ResNet-101 as the backbone. The activation function was replaced with the MIsh function, and the ECA Block was added at appropriate locations. The overall structure of the proposed network is shown in Fig. 2.

We placed the ECA Block in two locations. The first is before the addition of skip connection in the Residual Block. ECA Block is a module that is added after the convolutional layer, and by placing it after the last convolutional layer in Residual Block, it is possible to select important information from features extracted up to the previous layers. Therefore, we thought it would be possible to clarify important information throughout ResNet. The second is after the encoder and decoder are combined. The part combines the high-dimensional features from the encoder output and the low-dimensional features from the skip connection. By placing the ECA Block after this, we thought it would be possible to focus on the important channels from the combined features, resulting to improved performance.

3. Experiment

The dataset used in this study was created independently and consists of 200 images of plastic bottles. The annotation was done manually. In addition, we performed five-fold cross-validation as a validation method and employed Intersection over Union (IoU) and mean Intersection over Union (mIoU) as an evaluation index.

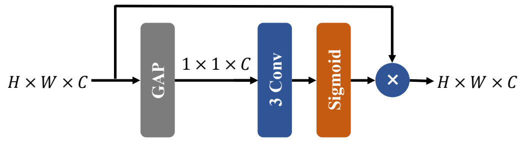


Fig. 1. ECA Block structure

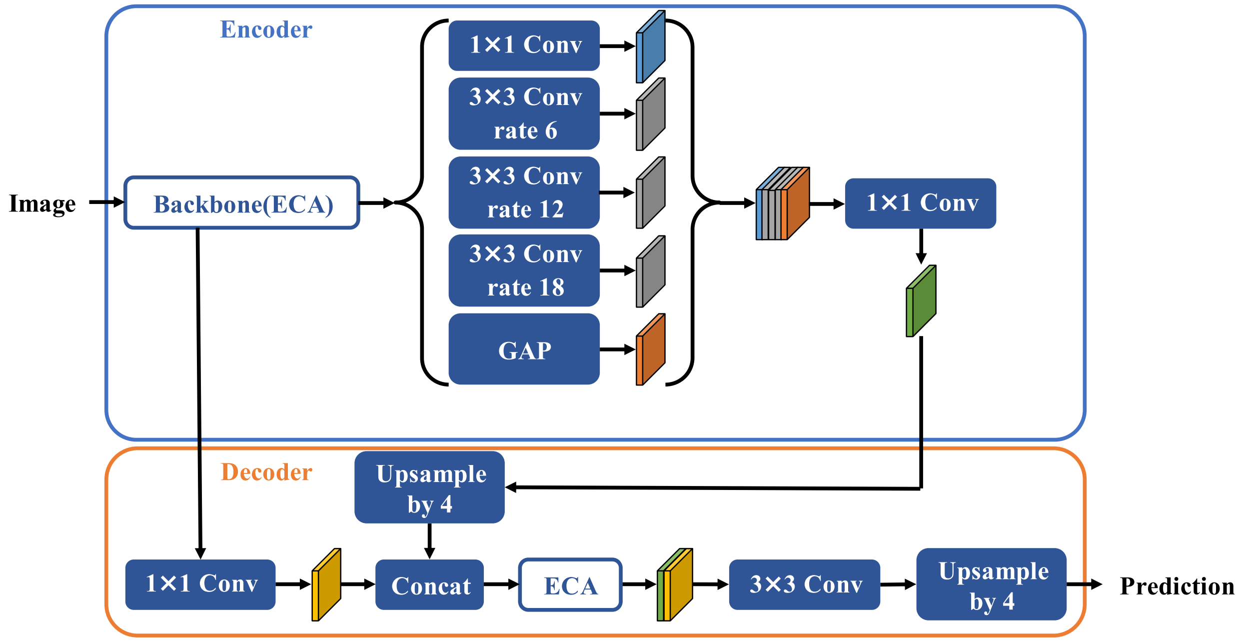


Fig. 2. Proposed network structure

3.1. *Result*

The experimental results are shown in Table 1. In this study, we calculated the IoU of the plastic bottle body, IoU of the cap part, mIoU, and fps for the five models. Note that the IoU of the body is denoted by "body" and the IoU of the cap part is denoted by "cap". In addition, the input image and the result image are shown in Fig. 3. Red is the mask of the body and green is the mask of the cap part. As shown in the results, the proposed method did not show any improvement over the base model in terms of numerical performance. However, there were changes in the trend of the captured features, the causes of which will be discussed later.

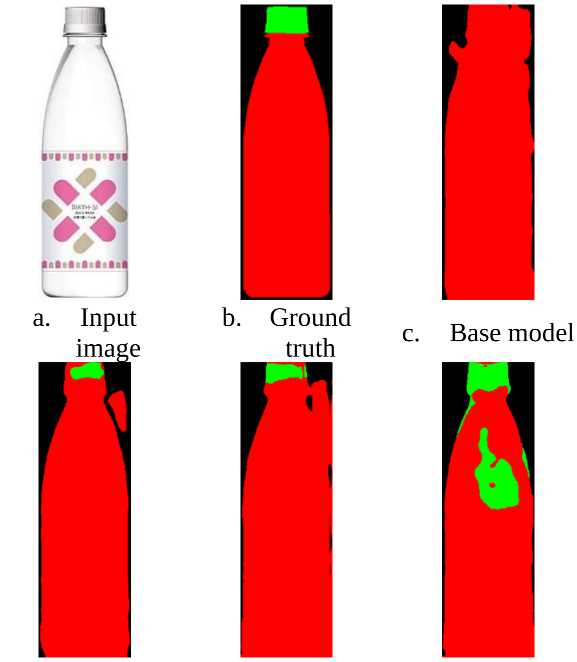
3.2. *Discussion*

Compare the individual images in Fig. 3. When the ECA Block was added to the base model, the number of false detections of the body and non-detections of the cap part decreased, probably because the ECA Block emphasized the important information. When the Mish function was added to the base model, the detection of the cap part increased. This could be because the Mish function improved the detection performance for small features. We thought that the combination of these characteristics would enable accurate recognition of both the body and the cap part, and the result of adding the ECA Block and Mish function is shown in Fig. 3. f. As expected, both the body and the cap part showed improvement over the base model. However, new false detection of the cap part appeared in the center of the body. A possible reason for this is that the cap part was made easier to detect by the Mish function and emphasized by the ECA Block, and the range of the features detected as the cap part became wider. In addition, even when the ECA Block and Mish function are added independently, there are many data in which their characteristics have negative effects, which is believed to lead to lower accuracy.

The most accurate of the conventional methods [7] was DeepLab v3+ with fine tuning. Compared to the methods tested in this study, the accuracy of the conventional method was significantly higher. The presence or absence of fine tuning is a major cause of this difference. Therefore, we would like to repeat the experiment by adding fine tuning to the method used in this study. In addition, although we worked on image recognition using simple images as basic research, in actual waste treatment facilities, it is necessary to detect plastic bottles from complex images. Therefore, the conventional method may not be accurate enough. We believe that the fact that we found changes in the trend of feature extraction with the addition of each module in this study is meaningful for practical use.

Table 1. Results for each model

	cap	body	mIoU	fps
Base model	0.673	0.955	0.814	14.74
With ECA	0.661	0.943	0.802	14.28
With Mish	0.638	0.950	0.794	14.70
Proposed model	0.611	0.941	0.776	14.06
Conventional model	0.932	0.980	0.956	7.42



a. Input image b. Ground truth c. Base model

d. With ECA e. With Mish f. Proposed model

Fig. 3. Segmentation result

In this study, we were able to extract features that were not captured by the base model. However, the effect of the modification was detrimental, resulting in false detections in the cap part. To solve this problem, we believe that emphasizing the positional information of the cap part is effective, so we consider introducing a spatial attention mechanism in the future.

4. Conclusion

In this paper, we conducted basic research on a plastic bottle sorting method for factory automation. Semantic segmentation was applied to images containing individual plastic bottles, and the body and the cap part were recognized separately. We used DeepLab v3+ as a deep learning model and modified it with the ECA Block and Mish function. As a result, improvements were observed in areas of misrecognition with the base model.

References

1. JMA Management Center Inc., "[2024] What are the root causes of labour shortages in the manufacturing industry? A Thorough Explanation of the Actual Situation and Countermeasures", https://www.jmam.co.jp/hrm/column/0136-manufacturing.html (in Japanese) (access 2024/11/28).
2. L. C. Chen et al., "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation", Proceedings of the European Conference on Computer Vision (ECCV), pp.801-818, 2018.
3. K. He et al., "Deep Residual Learning for Image Recognition", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.770-778, 2016.
4. Q. Wang et al., "ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.11534-11542, 2020.
5. D. Misra, "Mish: A Self Regularized Non-Monotonic Activation Function", arXiv preprint arXiv:1908.08681, 2019.
6. K. He et al., "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition", in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.37, no.9, pp.1904-1916, 2015.
7. D. Ideta, T. Kamiya, "Recogniton of Specific Parts of Plastic Bottles Using Improved DeepLab v3+", 2023 23rd International Conference on Control, Automation and Systems (ICCAS), pp.1758-1761, 2023

Authors Introduction

Yusuke Murata

He received his Bachelor's degree in Engineering in 2024 from the Faculty of Engineering, Kyushu Institute of Technology in Japan. He is currently a master student in Kyushu Institute of Technology, Japan

Tohru Kamiya, Prof. Ph.D

He received his B.A. degree in Electrical Engineering from Kyushu Institute of Technology in 1994, the Master and Ph.D. degree from Kyushu Institute of Technology in 1996 and 2001, respectively. He is a professor in the Department of Mechanical and Control Engineering at Kyushu Institute of Technology. His research interests are focused on image processing, medical application of image analysis. He is currently working on computer aided diagnosis based on CT, MR imaging, fluorescence microscope imaging, and automatic classification of respiratory sound.