Automatic Classification Method for Plastic Bottles and Caps Using Multi Attention Eff-UNet

Shunsuke Moritsuka, Tohru Kamiya
Kyushu Institute of Technology, 1-1 Sensui-cho, Tobata-ku, Kitakyushu-shi, Fukuoka 804-8550, Japan
E-mail: kamiya@cntl.kyutech.ac.jp

Abstract

In Japan, increasing amounts of waste are becoming a social problem. One of approaches to solve the problem is recycle of the plastic bottles. However, they are thrown away with their caps still attached, and it should be removed by hand. To solve this problem, we developed a method for automatic identification of plastic bottles and caps using deep learning technique. In this paper, we propose a method that combines different numbers of Efficient blocks and adds an attention structure and verify its usefulness through experiments.

Keywords: Plastic bottle recycling, Convolutional neural network, Segmentation, Eff-UNet, Attention

1. Introduction

In Japan, increasing amounts of general waste are becoming a social problem. It is increasing year by year and is expected to be as large as 41.67 million tons in 2020[1], which will fill up landfills soon. Therefore, it is necessary to reduce the amount of general waste as much1as possible and to devise ways to prevent landfills from becoming full. One solution to reduce the amount of general waste is to reuse plastic bottles. For this reason, some communities have special bags for disposing of plastic bottles. However, many bottles are thrown away with the caps still attached to bags, which are manually sorted at the landfill after checking with the human eye to see if the caps are still attached. This involves standing for long periods of time doing the same work repeatedly, which causes fatigue and oversight.

In recent years, digital cameras and other devices have become more precise and sensitive, and AI (Artificial Intelligence) technology has been developed in analysis software. Therefore, it was thought that by effectively utilizing deep learning, automatic detection of plastic bottles and caps could be realized, automatic sorting could be performed, and the problem could be solved.

Based on an above mentioned, this paper describes a development of a new model that combines Eff-UNet structure, which specializes in semantic segmentation, with an attention structure that removes irrelevant regions such as the background, among Convolutional Neural Networks (CNN), one of deep learning methods, using images of plastic bottles obtained from the Internet and the evaluation of its performance.

2. Method

This section describes an image analysis technique that automatically extracts regions of a plastic bottle and its cap from an input image. Although various methods exist for image extraction, semantic segmentation is employed in this study because it is sufficient to identify regions of the plastic bottle and cap. The proposed method is described below.

2.1. *U-Net*

U-Net[2] is a model which is developed for biomedicine by Olaf et al. and consists of an encoder and a decoder (Fig. 1). In the encoder section, an input image is convolved multiple times to extract image features. In the decoder section, up-sampling is performed to obtain a probability map of the same size as an input image based on the features extracted by the encoder. Since the feature values are lost in this process, a skip connection structure is used to convey the information of the large feature map on the encoder side to the decoder side, facilitating the acquisition of object location information during it.

2.2. *Eff-UNet*

Eff-UNet[3] is a model proposed by Baheti, in which the encoder part of U-Net is changed to EfficientNet[4] as shown in Fig. 2. It is a model created by combining squeeze and excitation-optimized mobile inverted bottleneck convolution (MBConv)[5]. Fig.3 shows EfficientNet and Fig. 4 shows MBConv respectively. It is characterized by its ability to balance three parameters of model depth, breadth, and resolution. Therefore, an improvement of accuracy by scaling up the model has already been implemented. It also saves time compared to conventional models because the model is trained with optimal parameters.

Based on an above, it is a model that is expected to improve the overall performance of the algorithm by incorporating a strong CNN like EfficientNet as an encoder and U-Net as a decoder with a skip-connection structure that can minimize information, and This method was also considered to be effective in this study.

2.3. *Attention Gate (AG)*

In this paper, we adopt the Attention Gate (AG) used in Attention U-Net[6] devised by Oktay et al. Fig. 5 shows the AG architecture. It consists of a flow where features obtained by up-sampling are passed through AG and multiplied with the original output, and a skip-connection structure. by using Sigmoid function as last activation function of AG, original features are Since the value of 0 to 1 is multiplied, the value of the region of interest remains unchanged, while the values of other regions become close to 0, enabling learning that is specialized to the region of interest. Therefore, while background information unrelated to recognition may have a negative impact in normal image recognition, the use of attention makes it possible to focus attention on relevant regions.

2.4. *Proposed Method (MAE-UNet)*

We propose MAE-UNet (Multi Attention Eff-UNet). The architecture of the proposed method is shown in Fig. 6. There are two key points. First, AG is used in the decoder part of Eff-UNet like Attention U-net to remove regions not related to recognition and reduce over-detected regions. Second, inspired by the DDU-Net[7] proposed by Tang et al., we combine Eff-UNet up to block6, which is dedicated to global segmentation, and Eff-UNet up to block4, which is dedicated to local segmentation. Conventional Eff-UNet has a skip connection structure specialized for global segmentation, which causes a problem of information loss during learning. Therefore, by using Eff-UNet with a small number of MBConv blocks, local segmentation can be performed, and the loss of information can be minimized, leading to improved accuracy.

3. Experiment

This section describes data set, evaluation method, and its experimental results. Specifically, one plastic bottle is placed in the center of each image.

3.1. *Data Set*

The data set used in this study consists of 210 images, 200 obtained from the Internet and 10 images taken by us. During training, data augmentation was performed by randomly flipping horizontally, scaling, smoothing, and normalizing. An annotation work was done by our own hands and divided into three classes: plastic bottles, caps, and backgrounds.

3.2. *Evaluation Method*

We evaluate the accuracy by 5 cross validations. It means that all data is randomly divided into five sets. The method is to use one set of them for testing (42 images) and the remaining four sets (168 images) for learning, evaluate them, and determine the average value for the five sets to obtain many results. The accuracy of the segmentation is determined by the Intersection over Union (IoU), which is the ratio of the product set of predictions and grand tours computed for each class to the sum set, and the average Intersection over Union (mIoU), which is obtained by averaging the IoU for each class. The equation for IoU is shown in Eq. (1). The Intersection region in the equation is the product set of the annotation and output results, and the Union region is the union set.

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

3.3. *Experimental Result*

The experimental results are shown in Table. 1 for comparison with three methods: U-Net, Eff-UNet, and MAEU-Net proposed in this paper.

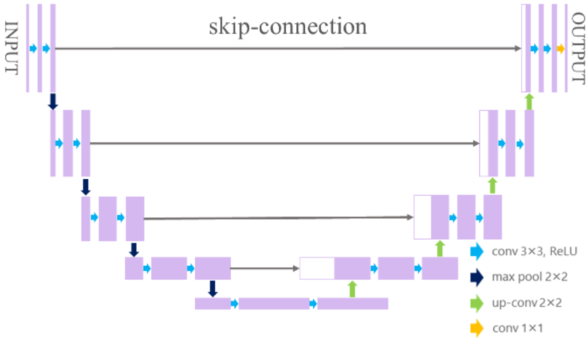

Fig.1. U-Net Architecture

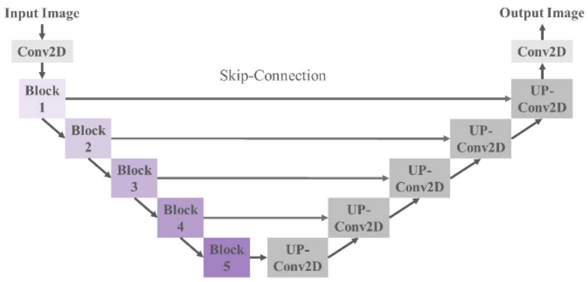

Fig.2. Eff-UNet Architecture

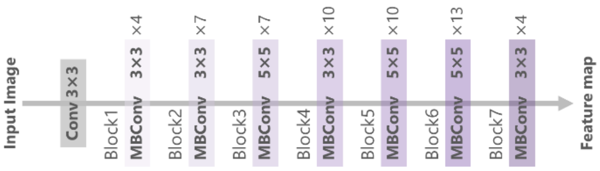

Fig.3. EfficientNet Architecture

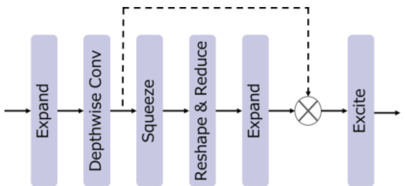

Fig.4. MBConv Architecture

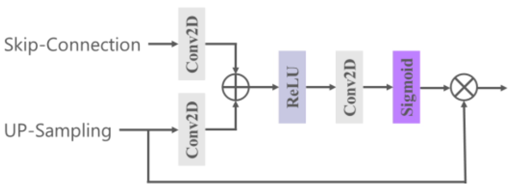

Fig.5. Attention Gate (AG) Architecture

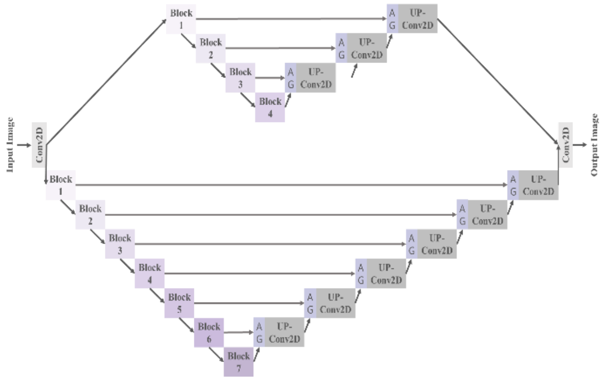

Fig.6. Propose Method Architecture

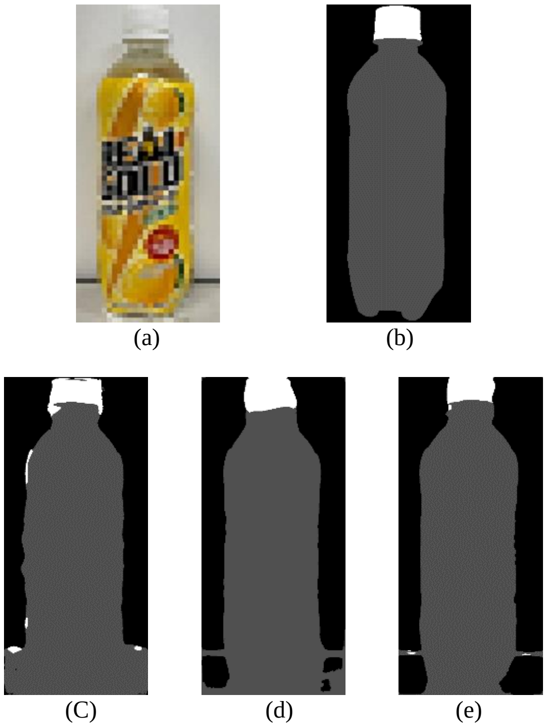

Fig. 7. Experimental Result (Gray:Body, White:Cap); (a)Input Image (Blur image) (b)CorrectLlabel (c)Segmentation Result (U-Net) (d)Segmentation Result (Eff-UNet) (e)Segmentation Result (MAE-UNet)

Table 1. Experiment Result

Model	Bottle	Cap	mIoU
U-Net	0.974	0.817	0.895
Eff-UNet	0.976	0.837	0.906
MAE-UNet	0.980	0.878	0.923

4. Discussion

In this paper, experiments were conducted using images obtained from an Internet as well as images of plastic bottles photographed. Table 1. shows that the proposed method gave the best results, improving accuracy by 2.8% over U-Net and 1.7% over Eff-UNet. The reason for improved accuracy can be attributed to the fact that the attention structure suppressed an over detection of background regions. Experimental result shown in Fig. 7 also indicate that a lower over detection is suppressed. In addition, the use of a model that combines different numbers of Eff-UNet blocks is considered to have improved accuracy compared to the conventional method. From Fig. 7, the roundness of a cap can be sensed when the proposed method is used. However, since the accuracy of the caps does not exceed 90%, further improvement in accuracy is required to achieve practical application. As a solution, it was considered that the area of the cap was so small for the plastic bottle that there might be a loss of features. Therefore, it was thought that the loss of information could be reduced by complicating the skip connection structure as in Unet++[8]. Also, since the number of images is very small, increasing the number of images is also effective.

5. Conclusion

In this paper, to improve the efficiency of plastic bottle sorting, we constructed a model combining different numbers of blocks of Eff-UNet, which incorporates the Attention structure in the encoder part, and extracted the regions of PET bottle bodies and caps. As a result, we succeeded in improving the accuracy of the model compared to the other two methods (U-Net and Eff-UNet). In the future, we plan to improve the accuracy by improving the partitioning network and increasing the number of data sets.

References

1. Ministry of the Environment, “Waste Management in Jap an (2020 Edition)”.
2. O. Ronneberger et al., “U-Net: Convolutional Networks for Biomedical Image Segmentation”, International Conference on Medical Image Computing and Computer-assisted Intervention, pp.234-241, 2015.
3. B. Baheti. et al., “Eff-UNet: A Novel Architecture for Semantic Segmentation in Unstructured Environment”, Proceedings 2020 Institute of Electrical and Electronics Engineers/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp.358-359, 2020.
4. M. Tan. et al., “Effcientnet: Rethinking Model Scaling for Convolutional Neural Networks”, International Conference on Machine Learning, pp.6105-6114, 2019.
5. Y. Lyu. et al., “Factorizing and Reconstituting Large-kernel MBConv for Lightweight Face Recognition”, Proceedings of the Institute of Electrical and Electronics Engineers/CVF International Conference on Computer Vision Workshops, 2019.
6. O. Oktay. et al., “Attention U-Net: Learning Where to Look for the Pancreas”, Accepted to Publisher in Medical Imaging with Deep Learning 2018, arXiv: 1804.03999, 2018. (Access 2022/11/09)
7. M. Jiang. et al., “A Novel Deep Learning Model DDU-net Using Edge Features to Enhance Brain Tumor Sgmentation on MR Images”, Artificial Intelligence in Medicine 121, 2021.
8. Z. Zhou. et al., “UNet++: ANested U-Net Architecture for Medical Image Segmentation”, 4th Deep Learning in Medical Image Analysis (DLMIA) Workshop, pp.3-11, 2018.

Authors Introduction

Mr. Shunsuke Moritsuka

Shunsuke Moritsuka graduated from the Department of Mechanical and Control Engineering, Faculty of Engineering, Kyushu Institute of Technology, Japan in 2021. He is currently a Master course student in Kyushu Institute of Technology, Japan.

Dr. Tohru Kamiya

Tohru Kamiya received his B.A. degree in electrical engineering from Kyushu Institute of Technology in 1994, the Masters and Ph.D. degree from Kyushu Institute of Technology in 1996 and 2001, respectively. He is a professor in the department of control engineering at Kyushu Institute of Technology. His research interests are focused on medical application of image analysis. He is currently working on automatic segmentation of multi-organ of abdominal CT image, and temporal subtraction of thoracic MDCT images