NUNI-Waste: Novel Semi-Supervised Semantic Segmentation for Waste Classification with Non-Uniform Data Augmentation

Jianchun Qi[*], Minh Nguyen, Wei Qi Yan

Auckland University of Technology, Auckland 1010 New Zealand

Abstract. Waste categorization and recycling are critical approaches for converting waste into valuable and functional materials, thereby significantly aiding in land preservation, reducing pollution, and optimizing resource usages. However, real-world classification and identification of recyclable waste face substantial hurdles due to the intricate and unpredictable nature of wastes, as well as the limited availability of comprehensive waste datasets. These factors limit efficacy of the existing research work in the domain of waste management. In this paper, we utilize semantic segmentation at individual pixel level and introduce a semi-supervised metod for authentic waste classification scenarios, leveraging the Zerowaste dataset. We devise a non-standard data augmentation strategy that mimics the ever-changing conditions of real-world waste environments. Additionally, we introduce an adaptive weighted loss function and dynamically adjust the ratio of positive to negative samples through a masking method, ensuring the model learns from relevant samples. Lastly, to maintain consistency between predictions made on data-augmented images and the original counterparts, we remove input perturbations. Our method proves to be effective, as verified by an array of standard experiments and ablation studies, achieved an accuracy improvement of 3.74% over the baseline Zerowaste method.

Keywords: Semi-supervised semantic segmentation · Data augmentation · Waste classification · Adaptive weighted loss function

1   Introduction

With the advancement of industrialization and socialization, it is estimated that total volume of waste will increase by 20% to 2.6 billion tons per year post-2030 [21]. Environmental pollution caused by inadequate waste management mechanisms poses a global challenge requiring immediate attention. For instance, around 6 trillion tons of plastic waste litter the global oceans, with the difficulty of plastic degradation contributing to long-term pollution hazards. Moreover, soil and water pollutions are also globally prevalent [11, 32]. By 2020, waste recyclability rates in a few of countries remained unsatisfactory, barely reaching 35% [40].

At present, waste treatment plants primarily employ semi-manual and semi-automated waste classification methods. These methods, however, are not only

inefficient but also pose significant health risks to workers, such as inhaling excessive dust and harmful chemical gases emanating from waste. Therefore, there is an urgent need to incorporate deep learning and computer vision into waste classification and recycling [15], which promises a healthier approach by significantly enhancing waste classification efficiency and recyclability and ultimately fostering environmental protection and economic growth.

While waste classification models [10, 15, 29, 39, 48] have demonstrated remarkable classification accuracy, explored waste classification and detection remain limited and insufficiently researched for real waste classification scenarios. These scenarios grapple with a wide variety of waste types, intricate classification backgrounds, diverse waste forms, and the impact of different real-world light sources. Most waste classification models were refined and improved based on the Labeled Waste in the Wild [38], Trash Annotation in Context (TACO) [35], and TrashNet [45] datasets. These datasets share three common characteristics: A small amount of waste data (due to the higher cost of manual labeling), pristine backgrounds, and a rather simplistic representation of waste scene. Despite the high accuracy of these waste classification models, they largely overlook the complexities present in real waste classification scenarios, which could potentially result in model overfitting.

Therefore, in this paper, we propose a semi-supervised semantic segmentation model for the Zerowaste dataset (the first real-world industrial-grade waste dataset) [1], amalgamated with deep learning, specifically designed for highly complex real waste classification scenarios. The choice of semantic segmentation over the enhancement of related visual object detection methods is motivated after our observations of the extreme clutter in real waste classification scenarios. If there is a significant quantity and variety of waste, waste items can become obscured by one another. At this point, semantic segmentation can segment the entire image without gaps into each region, with each region belonging to a category. This approach is sensitive to the size, shape, and state of objects. Our method not only improves waste classification and mitigates the extensive pixel-level manual annotation work but also alleviates the data-hunger problem [28] during model training. From collaborative training [6, 36] and self-training [18, 44], to generative adversarial networks [28] and the currently popular consistent regularization [7, 20], semi-supervised learning methods have been in development for a considerable duration. How can we offset the scarcity of waste datasets, the shortcomings of waste datasets, and the complexity of waste dataset scenarios? How can we employ semi-supervised learning to effectuate improvements in waste classification efficiency and accuracy? How can we ameliorate the extreme lack of pixel-level annotation in semi-supervised semantic segmentation? These questions interest us greatly and are addressed in this paper.

In this paper, we concentrate on the consistent regularization framework to tackle the aforementioned issues by introducing non-uniform data augmentation tailored to the dataset characteristics of real waste classification scenarios. Through our exploration, we found that the existing data augmentation methods, such as Cutout [9], Cutmix [46], and Mixup [16], have limitations. For instance, Cutout omits portions of the image and cannot utilize the full image information, while the data post-Mixup is often unnatural due to local blurring. In contrast, our non-uniform data augmentation method neither adds nor removes image information, which is uniform and continuous, each changed pixel point maintains correlation, avoiding the direct addition of noise to the image that ensures the generalization ability of this model. Moreover, we discovered that in the Zerowaste dataset, two categories with low data volume lead to substantial differences.

To counteract the escalating misclassification costs, we have designed an adaptive weighted loss function by assigning different weight values according to the data size characteristics of different categories. We have added a mask to make adaptive adjustments to the number of positive and negative samples, compelling the model to filter out features that have been particularly well learned without imposing a limit. This improvement is based on the waste dataset that can be applied to other categories of unbalanced datasets. Our contributions are summarized as follows:

(1) We construct a semi-supervised semantic segmentation waste detection model applied to real waste classification scenarios. Our experimental results visually demonstrate its state-of-the-art performance.

(2) A novel non-uniform data augmentation method is proposed. It simulates natural light by making the model better adapted to the waste classification scenario. Additionally, it increases the amount of data and reduces network overfitting. Our method is smooth and continuous, ensuring the generalization ability of our proposed models.

(3) An adaptive weighted loss function is designed to counter the model non-robustness resulting from severe data imbalance.

In the rest of the paper, our related work is shown in Section 2. Section 3 illustrates our methodology. The experiments and ablation studies are detailed in Section 4. Finally, our conclusions are presented in Section 5.

2 Related Work

2.1 Waste Detection

Over the past few years, waste datasets have been updated continuously. The TrashNet dataset for recyclable waste [45] was proposed, which contains more than 2,000

annotated images, including six classes, i.e., paper shell, plastic, paper, metal, glass, and trash, all of which were collected by shooting under natural light illumination. This dataset is characterized by the simplicity of containing only one object in each image, the background of the object is mostly white. Subsequently, UAVVaste [22] was publicly used, which is a collection of outdoor waste collected by using aerial photography from drones in urban environments. It has about 770 waste images and over 3,700 waste objects, all labeled in the "waste" category.

Finally, a more popular waste dataset is TACO [35]. The background of trash images is intricate, including indoor, beach, street, vegetation, and water, with 1,500 images in 28 categories and about 4,800 labeled objects. However, the annotations in TACO do not belong to the waste category, which can interfere with the learning of the model. Although these three typical waste datasets have their characteristics, they do not play all the advantages in real waste classification scenarios and are more suitable for waste detection in the field.

On the other hand, the performance of waste detection using deep learning models has also been gradually improving. A genetic algorithm was employed to optimize the traditional CNN model, DenseNet121, which achieved a high accuracy of 99.6% on the TrashNet dataset [29]. Thus, the effectiveness of the proposed model and the RGBD-based MJU-waste segmentation dataset were all verified by performing a rough segmentation operation on the data first and then selecting the target region for meticulous segmentation. If the backbone network adopted ResNet-101 [17] and the baseline employed DeepLabv3 [5], the IOU mean was 97.14 [39]. Ultimately, applied to the waste classification task, the one-stage deep learning model YOLO achieved 94.79% of the mAP value after a lightweight improvement. All the three waste classification models achieved an accuracy rate of more than 90% [10]. However, they were applied to simple dataset backgrounds, monolithic waste items, and less stacked and obscured waste states, which made the detection results potentially uneven. The improved algorithms are lack of relevance and less consideration of the real form and characteristics of the waste. Therefore, we improve the model performance by performing data augmentation and loss function improvement for real waste plant classification scenarios based on the Zerowaste dataset.

2.2 Semi-supervised Learning

The core idea of semi-supervised learning is how to maximize the use of unlabeled data to advance model learning. Consistency regularization [12, 28, 34, 41] and entropy minimization [3, 4, 13] are two popular semi-supervised learning paradigms. Consistency regularization serves to reduce the phenomenon of overfitting. If unlabeled data is perturbed, its prediction result should not change significantly, i.e., it has

predictive consistency. Entropy minimization combines unlabeled data, labeled data, and pseudo-labeled data to make the network predictions more confident. Other training strategies, data augmentation, such as Cutout [9], Cutmix [46], and Mixup [16], are applied in semi-supervised learning to increase the utilization of unlabeled data to improve the performance of the model. Nevertheless, these methods may not maintain the localization and generalization ability of the data. In contrast, the focus of our non-uniform data augmentation is on continuous improvement of smoothing at the pixel level, which ensures the generalization ability of the proposed model.

2.3 Semi-supervised Semantic Segmentation

Different from object detection, semantic segmentation goes down to the exact pixel point in the image, the semantic information assigned to each pixel is the position of that pixel point in the image. However, pixel-level annotated data is much difficult to obtain and suffers from severe overfitting problems, so the development of semi-supervised semantic segmentation is inevitable. In the early stage, typical Generative Adversarial Networks (GANs) [14] models extract valid training signals from unlabeled data [19, 33]. Later, semi-supervised semantic segmentation follows a semi-supervised learning paradigm for model training [3, 4, 12, 13, 31, 34, 41], such as entropy minimization and consistency regularization, etc. Similarly, data augmentation methods are crucial for semi-supervised semantic segmentation has also been revealed [12]. Therefore, our approach greatly reduces the labor cost required for pixel-level labeling, improves the performance of the semi-supervised semantic segmentation model, and further solves the problem of lack of data for waste classification tasks.

3 Our Method

3.1 The Structure of Our Framework

In this paper, we utilize U-Net as the baseline and ResNet-50 as the backbone network to validate our proposed non-uniform data augmentation and adaptive weighted loss function. The network structure is illustrated in Fig. 1. While model training with unlabeled data, data without non-uniform data augmentation and data with non-uniform data augmentation produced L_1 loss, which is to ensure the consistency of the results predicted by these two types of data, and the better the model results if the L_1 loss is approximately small. Finally, drop perturbation is added to both the original input channel and the feature channel to improve the generalization ability of the model.

3.2 Non-uniform Data Augmentation

Our non-uniform data augmentation encompasses not only lighting tasks to simulate real scenes, but also polymorphic tasks to emulate waste objects. These correspond respectively to non-uniform color data augmentation and non-uniform offset data augmentation.

n× : the size of the feature map is 1/n of the original image, n=2, 4, 8, 16, 32

Fig. 1 The structure of our network

Fig. 2 The typical samples of non-uniform color data augmentation, which is applied to a completely random strategy

Non-Uniform Color Data Augmentation. Non-uniform color data augmentation randomly simulates real-world natural lighting, which is different from random brightness data augmentation [42]. Random brightness simply increases or decreases the same pixel value randomly for all pixels, while our data augmentation makes the pixel points all completely random, as shown in Fig. 2. From Fig. 2(a) to Fig. 2(d),

these figures showcase the light being shaded from the top, bottom, left, and right of the object. Fig. 2(e) has darker light near the edges of the image and brighter light near the middle of the image, while Fig. 2(f) is exactly the opposite of Fig. 2(e). After that, Fig. 2(g) and Fig. 2(h) represent the case of intensely brighter and deeply darker light, respectively. Fig. 2 only shows eight images with typical representative non-uniform color data augmentation methods, similar but not identical samples, as shown in Fig. 3.
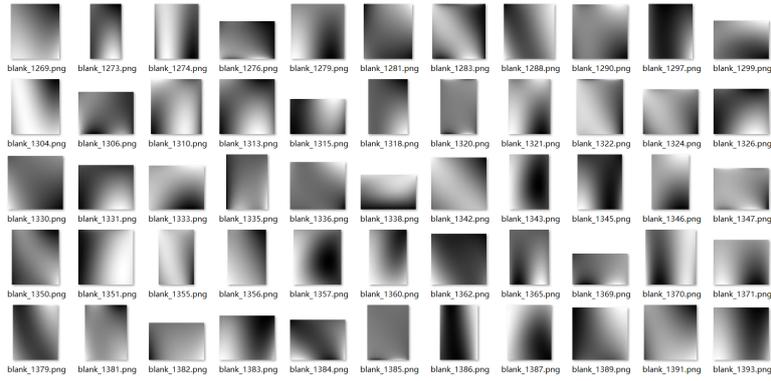
Fig. 3 The selected samples from atypical non-uniform color data augmentation
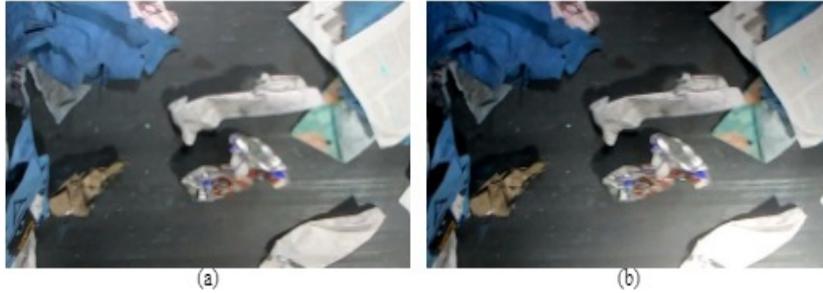
Fig. 4 Comparison of the original image and the image after data augmentation. (a) is the original image in the Zerowaste dataset, we see that in the original image, the light is also blocked. The color of the lower left corner is darker than the upper right corner because the light is blocked on the left side. (b) is the image after non-uniform color data augmentation. At this point, we see that the light of the image is blocked from the upper left angle, and the lower right corner is brighter.

Applying non-uniform color data augmentation to the Zerowaste dataset, the position of the pixel values in the image is left unchanged, the samples after non-uniform color data augmentation are obtained by randomly and continuously changing the values of the pixels. As shown in Fig. 4(b), the light from the top left of the image

is blocked, while the light from the bottom right of the image is enhanced after waste classification is carried out. Compared with the Fig. 4(a) in the dataset, the upper left part of Fig. 4(b) is darkened, and the lower right part is brightened, simulating the situation where the waste samples are disturbed by different light for the same number, the same object, and the same shooting angle.

Non-Uniform Offset Data Augmentation. Fig. 5(b) is shown as a compressed version of Fig. 5(a) or the whole pixel is shown an upward translation trend. Fig .5(d) is displayed as an enlarged version of Fig .5(c) or the overall pixels are panned to the right. This is our non-uniform offset data augmentation, which is different from the simple scale and aspect-ratio-based transformations for image augmentation. Specifically, each pixel in the image is given a different, random offset, such that one pixel is shifted by 1 pixel and another pixel is shifted by 3 pixels, and this offset is gradually reduced if it reaches a maximum peak value.
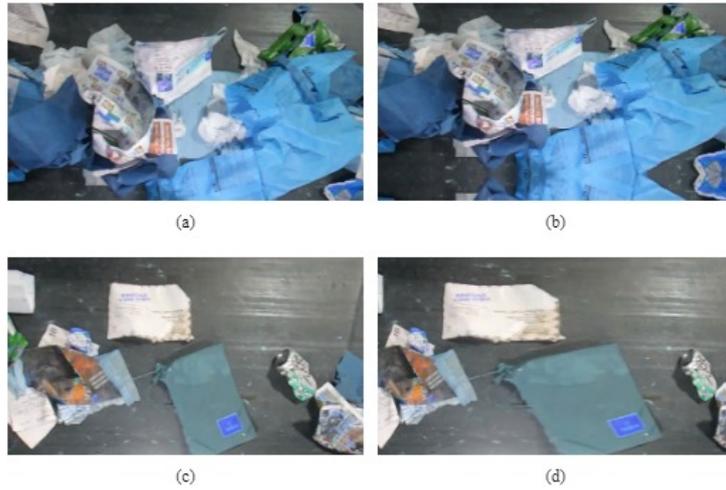
Fig. 5 Comparison of the original image with the image after data augmentation. (a) and (c) are original images in the Zerowaste dataset, and almost all the waste exists distorted and folded in different forms. Therefore, after non-uniform offset data augmentation, the morphology of the waste objects in (b) and (d) is smoothly variegated.

Let $i \in I^{W \times H \times C}$, i denotes a training sample, and this sample consists of two dimensions x and y. The purpose of non-uniform offset data augmentation is to generate a new training image $\tilde{\imath}(x,y)$. We define the formula $\tilde{\imath}(x,y)$ for as follows:

$$\tilde{\imath}(x, y) = \Delta i(x, y) * \sin(2\pi * r / v_c) + i(x, y), \tag{1}$$

and

$$v_c = 1200 + 200 * ((r - 0.5) * 2) \tag{2}$$

where x and y are the intensity of each pixel at x-axis direction and y-axis direction, respectively. $\Delta i(x, y)$ is the initial value of the offset that we randomly assigned to the pixels, after which it is multiplied with the sine function to obtain the final pixel offset. r represents a random number that conforms to a uniform distribution in the range of 0 to 1.0, which serves to make the Sine function value also conform to the random criterion. Finally, v_c indicates the random peak of the sine wave. The equation of $\Delta i(x,y)$ is shown as follows:

$$x \sim Unif(0, W), \qquad y \sim Unif(0, H), \tag{3}$$

and

$$\Delta i(x,y) = \begin{cases} \Delta x = v_o + 15 * ((r - 0.5) * 2) & r < 0.5 \\ \Delta y = v_o + 15 * ((r - 0.5) * 2) & r \geq 0.5 \end{cases} \tag{4}$$

where v_o means that we define a fixed value 70 for the pixel and perform a random addition and subtraction calculation on this fixed value to get the final value of $\Delta i(x, y)$. Then, Eq. 4 means that $\Delta i = \Delta x$ when $r < 0.5$ and $\Delta i = \Delta y$ when $r >= 0.5$. Thus, we keep the offset within a range. In this experiment, the initial value of the offset is taken to be between 55 and 85 (this initial value can be adjusted according to the scene of different datasets). The initial value 70 and the range of randomly selected offsets are the optimal choices determined through our experiments. As illustrated in Eq. 1, we calculate the final offset by using the initial value and a sine function (other computations such as cosine, cubic function, and quadratic function can also be employed, contingent on the characteristics of different dataset scenes).

Table 1 The offset values

Pixel numbers	Offset values
000	0.0000000000000000
001	0.4715737324604586
002	0.9431290236200252
003	1.4146474328989704
……	……
250	75.40823542253278
251	75.41026314065878
252	75.40934187486734
253	75.40547166118532
……	……
297	72.33678257778097

9

298	72.20210897214199
299	72.06461184024413
……	……

Table 1 presents the offsets for *x*-axis. Our data augmentation method is continuous and naturally smooth in transition. In the Zerowaste dataset, approximately 6,212 unlabeled images are collected. The deficiency in data volume can be substantially addressed by using non-uniform data augmentation. Given that our method operates by mapping each pixel of the image to the original image, creating variations in the pixel value and position stands distinct from noise addition which merely darkens a pixel and incites abrupt pixel changes. Our non-uniform data augmentation is neither independent nor random, but continuous with a natural smooth transition. This enhances the generalization ability of our proposed model and brings it closer to actual waste classification scenarios. Furthermore, since non-uniform data augmentation neither deletes nor adds pixels, it circumvents the problem of erroneous positive and negative sample assignment due to image resizing.

Adaptive weighted loss function. While training the model, we observe that two classes (Metal and Rigid Plastic) in the dataset had low numbers of waste images, which result in an extreme imbalance of the data distribution, and the model did not extract the features of these two categories well. To solve this problem, we designed a new adaptive weighted loss function, as seen in Eq. 5, Eq. 6, and Eq. 7.

$$L_{cls} = -\frac{1}{Z}\left\{\sum_{i=1}^{N} loss(p_i)\right\} \tag{5}$$

$$loss(p_i) = \begin{cases} e^{w-p_i} \cdot \ln p_i, & p_i < \eta \\ 0, & p_i \geq \eta \end{cases} \tag{6}$$

$$Z = \sum_{i=1}^{N}[p_i < \eta] \tag{7}$$

where *i* refers to a pixel, Z represents the number of e in the mask, which is equivalent to a planning coefficient, η is a hyperparameter set to 0.99. Moreover, w denotes the weight and p_i shows the prediction probability of a pixel. e_w refers to the overall weight assigned to each waste class, and w is set by us in the experiments.

Since there are four waste classes and one background class, we assign w as 3 to the Metal class and Rigid Plastic class based on the feature of less data volume in these two classes. All other classes have $w = 1.0$, which is the best value for training results after we have conducted a number of ablation experiments. e_p is to justify a weight assignment of this pixel itself. If e_p is stable, *p* should be large, so that the value of e_p

10

becomes small. Conversely, if the value p is too small, the weight assigned to it will be larger. In this way, it is straightforward to bring the data to a balanced state.

Furthermore, we set a mask as 0 if p_i is greater than or equal to η, as provided in Eq. 6. By the way, the number of positive and negative samples is self-adjusted by mask, which filters out the features that are particularly well learned in the model and without limiting the number. Finally, the weights are then multiplied by the cross entropy to obtain the final loss function value, not as same as OhemCELoss [37], a typical weighting loss method, which restricts the proportion of positive and negative samples, while we do not restrict them, the positive samples also acquire valuable features by achieving a simple and effective way to improve the performance of the model.

4   Our Experiments

4.1   Datasets

We propose a novel non-uniform data augmentation and adaptive weighted loss function to train the semantic segmentation of the Zerowaste dataset for the purpose to improve the accuracy and efficiency of real waste factories in performing waste classification. The Zerowaste dataset contains three branches, namely Zerowaste-*f*, Zerowaste-*s*, and Zerowaste-*w*. We mainly take use of the Zerowaste-*s* dataset, which is available for semi-supervised tasks and contains 6,212 unlabeled images. Additionally, the Zerowaste dataset has four waste categories, i.e., Cardboard, Soft plastic, Rigid plastic, and Metal. We evaluate the model with the metric mean(IoU) and perform ablation experiments to verify the effectiveness of our proposed method.

4.2   Implementation Details

Our experiments were based on a server with RTX A5000 GPU and AMD EPYC 7543 CPU. Installed code editors VSCode, neural network framework PyTorch, CUDA, CUDNN, and, OpenCV. The specific experimental parameters are reflected in Table 2.

Table 2 Training parameters of our experiments

Classes	Parameters
Initial learning rate	$1e^{-4}$
Optimizer	Adamw
Weight decay	$1e^{-3}$
Batch size	26
Epoch	100

11

4.3   Results

Our method is based on the Zerowaste dataset to improve the model of real waste factory classification scenarios, so we follow the result comparison rules of the Zerowaste baseline to ensure the fairness and validity of the experiments. In Table 3, the mean(IoU) of the test set from our Unet method (semi) of ResNet-50 is 55.37%, which is higher than the experimental results obtained in the Zerowaste baseline of about 3.74%. If we do not utilize ResNet-50 but with EfficientNet, the IoU is also reduced by 0.39%. Besides, to obtain visual comparison results, we applied our method to DeepLabv3+ [5] (the same baseline method as Zerowaste, whose backbone network is ResNet-101) and obtained a mean(IoU) value of 54.77%, which is higher than the original value of 3.14%. It proves that our method is effective that can noticeably improve the value of the mean(IoU). After that, following the experimental approach of Zerowaste, we compared it with UniMatch [43], AugSeg [47], ReCo [26], CCT [33], and EPS [23]. The mean(IoU) of the semi test sets of the UniMatch, AugSeg, and ReCo method is 54.65%, 53.88%, and 44.12%, respectively, which are lower than that of our method. Finally, regarding CCT and EPS, all their mean(IoU) values are below 33%, while our method is up to 55.37%. Moreover, since the Transformer model has improved the performance of the classification model and segmentation model, we also tested the CLUSTERFORMER [24] and Swin Transformer [27], the obtained mAP values were 52.76% and 53.21%, respectively. We see that our method is also applicable in the Transformer models.

Table 3 Comparisons of mean(IoU) results of different methods

Method	Supervision	Validation	Test
Ours(Unet+Resnet-50)	semi	49.27	55.37
Ours(Unet+EfficientNet)	semi	49.01	54.98
Ours(DeepLabv3+)	semi	48.89	54.77
CLUSTERFORMER	semi	47.95	53.21
Swin Transformer	semi	46.98	52.76
Unet	full	46.02	51.88
DeepLabv3+	full	45.61	52.30
DeepLabv3+	semi	46.13	51.63
UniMatch	semi	48.53	54.65
AugSeg	semi	47.12	53.88
Reco	full	51.30	52.28
Reco	semi	49.49	44.12
CCT	full	30.79	29.32
CCT	semi	28.70	32.49
EPS	weak-f	13.75	13.91

After the U-Net method was implemented, the value of mean(IoU) is only 0.6% higher than using the DeepLabv3+ method, but we still choose to apply U-Net model as our semi-supervised waste semantic segmentation network. This is because U-Net model consists of an encoder and a decoder that not only obtains the global features of the image but also restores the spatial information of the image, and has remarkable performance in processing smaller datasets, while DeepLabv3+ is suitable for processing large datasets. With the experimental results, we see that our method yields impressive results. Fig. 6(a) and Fig. 6(b) illustrate the mean(IoU) and loss plots of our method, respectively. This shows all the mean(IoU) and loss values during the training process. Another, darker-colored, less undulating fold is the approximate tendency of the mean(IoU) curve and the loss curve, it is present to highlight the focused changes in the curve.

Furthermore, we adopt the SGD optimizer, because though the SGD optimizer is less efficient than the Adamw optimizer, it has more stable training results. We are surprised to see that the SGD optimizer utilized exhibits unsatisfied performance in our model, leading to a negative effect on the training outcomes, as shown in Fig. 7. Its loss values are on average higher than those of the Adamw optimizer.

(a)   (b)
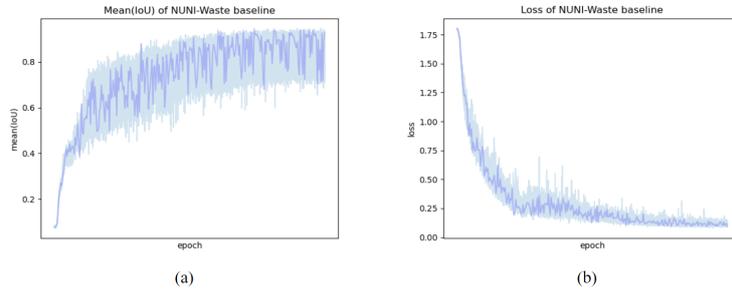
Fig. 6 The mean(IoU) and loss values of our method
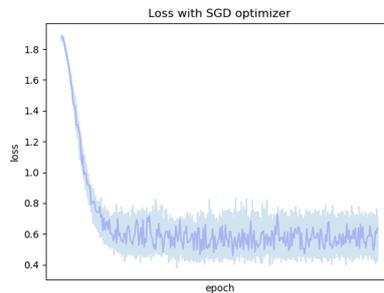
Fig. 7 The loss value of our method when using SGD optimizer

13

5 Ablation Study

5.1 Analysis of the Adaptive Weighted Loss Function

Table 4 Comparisons of mean(IoU) with the application of adaptive weighted loss function

Adaptive weighted loss function	Mean (IoU, %)
w/o	54.12
w/	55.37

The loss function is harnessed to evaluate the deviation of model prediction from the true value, the selection of loss function will impact the training results of the model. In this paper, we propose an adaptive weighted loss function and conduct experiments, as shown in Table 4. The mean (IoU) using our loss function is higher than the mean(IoU) without our loss function by 1.25%. We improve the performance of the model by a simple weighting method, effectively and efficiently. There is still room for improvement in the adaptive weighted loss function, and we believe that exploring more dynamic adaptive weighting schemes is an interesting direction for future work.

5.2 Analysis of w values

In this paper, the Zerowaste dataset has four categories, namely Cardboard, Soft plastic, Rigid plastic, and Metal. But in the semantic segmentation, it is also necessary to add the background category. Therefore, our experiment contains five classes. However, after observation, the data volume of the Metal classs and Rigid Plastic class only accounted for 3.6% and 16.6% of the total dataset, respectively, which show the extreme imbalance of the dataset. After the initial training, the mean(IoU) of these two categories was only about 20.00%, which was much lower than that of other categories, resulting in lowering the overall training of the model results. Our adaptive weighted loss function improves this problem by adding the same weight values for both categories. According to the design of our loss function, we perform ablation experiments on the w values in the adaptive weighted loss function in this subsection.

There are five classes to be weighted, the initial w value is set as 1.0. The model is trained to obtain a mean(IoU) of 52.69%. We envision that the value of mean(IoU) will not be positively correlated with the value of w. If only the Metal class and Rigid Plastic class are given greater weight values, it would lead to overfitting the model. Therefore, we did not assign large values to these two categories, which is detailed in Table 5. The best training result is that if the value of w is 3, and the mean(IoU) is as high as 55.37%. The lowest mean(IoU) value of 52.60% is when the value of w is set to 5. Besides, the mean(IoU) with w equal to 1 is 0.09 higher than 52.60%. Finally, the second highest mean(IoU) value is 54.02%, which is 1.35% lower than 55.37% (when the w value is 3.5). This verifies our conjecture that w value cannot be set too large.

14

Table 5 Comparisons of mean(IoU) with different w values

w value	mean(IoU)
1	52.69
2	53.07
2.5	53.55
3	55.37
3.5	54.02
4	53.18
5	52.60

Next, since Metal class and Rigid Plastic class also have different amounts of data samples, we assign two different sets to these two categories. As an example, let w values be 3 and 2 (or 4 and 2), corresponding to the Metal category and Rigid Plastic category, respectively. However, the mean(IoU) did not increase due to the finer differentiation of the w values, both about 53% (which is different from our conjecture). We speculate that this maybe due to the difference in data volume between the Metal class and the Rigid Plastic class which is small. Moreover, another reason is that the simplified assigning w values do not raise the mean(IoU) values if the difference in data volume is small. In future work, we will further investigate the results with a variety of w values. Fig. 8 is the typical losses with different weights.
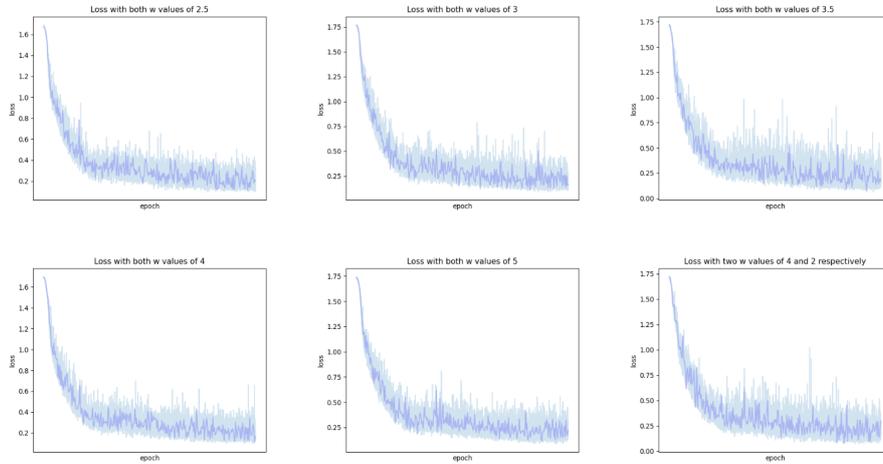
Fig. 8 The typical loss values with different w value

5.3 Analysis of Non-uniform Data Augmentation

Table 6 Comparisons of mean(IoU) with various data augmentation methods

	Data augmentation strategies	

Methods	Zerowaste baseline	Non-uniform data augmentation	Cutout	Cutmix	Mixup	mean(IoU)
DeepLabv3+	√					51.63
		√				52.88
Ours (DeepLabv3+)	√					53.52
		√				54.77
Ours (Unet)	√					52.88
		√				55.37
			√			52.21
				√		53.86
					√	54.08

In this paper, we conduct an in-depth study of non-uniform data augmentation and compare them with other data augmentation methods. In Table 6, we list the results of our experiments. The best result is our non-uniform data augmentation, which is up to 55.37%. Then, we keep U-Net model and employed Cutout, Cutmix, and Mixup for replacement, and obtained mean(IoU) values are 52.21%, 53.86%, and 54.08%, respectively, which are all lower than our method. For a fair comparison with the Zerowaste baseline method, we also conducted the experiments in the case of using DeepLabv3+ as the segmentation network. However, it was only for comparing our method with the method in the Zerowaste, so Cutout, Cutmix, and Mixup were not tested. The results are remarkable in that our non-uniform data augmentation can increase the mean(IoU) value of the model by about 2.49% on average.

5.4   Analysis of the Initial Offset Value

In Section 3.2, we narrated setting the initial value for non-uniform offset data augmentation, 70 as the best solution, which is what we have obtained after the ablation experiment. Table 7 shows that if the initial value is set to 30, the mean(IoU) will reach the lowest value of 51.97%. In addition, if the initial value is set to 100, it also causes a decrease in the mean(IoU) value. Fig. 9 illustrates the comparison plot of the mean(IoU) with the initial value set to 30 and 100. We suppose that the pixel values of the images range from 0 to 255, if the initial value of data augmentation is too large or too small, it cannot get the global information of the images and will cause the network to ignore image features. Therefore, in this experiment, we choose 70 as our initial offset value of data augmentation.
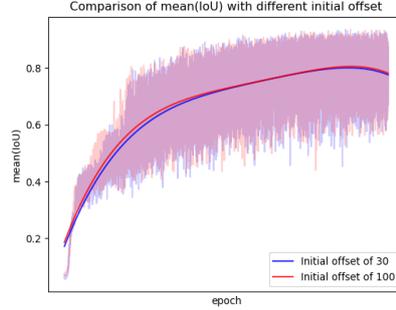
Fig. 9 Comparison of the mean(IoU) with the initial value set to 30 and 100

Table 7 Comparison of mean(IoU) with different initial offset value

Initial offset value	mean(IoU)
30	51.97
50	53.12
70	55.37
100	53.93

5.5   Analysis of the Application of Non-uniform Data Augmentation

Since our non-uniform data augmentation includes the transformation of pixel colors and pixel positions, we compared these two methods. If we only use non-uniform color data augmentation, the value of mean(IoU) is 54.39%, which improves the Zerowaste baseline by 1.41%. In contrast, applying non-uniform offset data augmentation improves the mean(IoU) value of the baseline method by 1.08%. Table 8 shows that our methods are effective and data augmentation for simulated light is more useful than data augmentation for simulated polymorphism. The reason for the difference needs to be explored in depth. We speculate that it is due to the hyperparameter settings of these two data augmentation methods. Therefore, in future, we will further improve our model.

Afterwards, to verify the comprehensiveness of the proposed method, we used the ResNet-101 model to compare with other data augmentation methods [25] on the CIFAR-10 dataset, the results are given in Table 9. We see that non-uniform data augmentation also enhances classification performance by about 0.12%. In our experiments, our method shows that robustness to image disturbances such as illumination variation, scale variation, rotation and tilt. This implies that it may also be resistant to some adversarial attacks. However, an assessment of specific attacks may needed, such as Adversarial Patch [2] and Physical-object-oriented MDE Attack [8]. Therefore, we plan to specifically test our method's resistance to these advanced

17

adversarial attacks in future work and further adjust and improve it to address these challenges.

Table 8 Comparison of mean(IoU) of the application of non-uniform data augmentation

Non-uniform data augmentation	mAP
Non-uniform color data ugmentation	54.29
Non-uniform offset data ugmentation	53.96
Both	55.37

Table 9 Comparison of mean(IoU) of the application of non-uniform data augmentation applied to image classification task

Data augmentation methods	Top-1 Error
Cutout	3.10
Mixup	3.09
CutMix	2.88
OHL-Auto-Aug	2.79
Non-uniform data ugmentation	2.67

6 Conclusion

We present a pioneering approach to waste classification that incorporates a novel non-uniform data augmentation technique. This method excels in simulating various environmental conditions such as natural lighting and object polymorphism, significantly enhancing the model's robustness. Coupled with an adaptive weighted loss function, our method, when applied to the U-Net architecture, elevates the mean Intersection over Union (IoU) by 3.74%. This underscores the efficacy and simplicity of our method in handling real-world waste classification challenges.

While addressing the scarcity of large labeled datasets in waste classification, we see our approach leverages semi-supervised semantic segmentation algorithms. This not only mitigates the financial burden of manual annotation but also enriches the Zerowaste dataset, thereby amplifying the model's efficiency and reliability in waste classification.

Our methodology, though tailored for waste classification, offers transferable insights that could be adapted to other domains, including traffic detection tasks on datasets like COCO and Pascal. Specifically, our non-uniform offset data augmentation methods employs sinusoidal computations, but it can also accommodate other mathematical functions like cosine or quadratic equations for broader applications.

Looking ahead, we aspire to extend our method to additional deep learning challenges, continually seeking enhancements. Moreover, there's potential for further refinements in non-uniform color data augmentation. For instance, beyond altering light and shadow, we plan to explore how manipulating object colors could further improve the model's performance. This serves as a fertile ground for future research. Finally, we suspect that the simplicity of adaptive weighted loss function is a limitation of our research work. Although it improved the accuracy of the model in this paper, there is still room for improvement. In the future, we will conduct further research projects on adaptive weighted loss functions to design more efficient and effective loss function.

Declarations

Competing Interests The authors have no financial or other conflicts of interests to declare.

Authors contribution statement Each author actively analyzed, drafted sections of the manuscript, edited, and approved the final, submitted version.

Ethical and informed consent for data used The dataset used has no ethical risk and is public dataset.

Data availability and access The dataset is accessible and available.

References

[1]   Bashkirova, D., Abdelfattah, M., Zhu, Z., Akl, J., Alladkani, F., Hu, P., ... & Saenko, K.: Zerowaste dataset: Towards deformable object segmentation in cluttered scenes. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21147-21157, (2022).

[2]   Brown, T. B., Mané, D., Roy, A., Abadi, M., & Gilmer, J.: Adversarial patch. *arXiv preprint arXiv:1712.09665* (2017).

[3]   Cascante-Bonilla, P., Tan, F., Qi, Y., & Ordonez, V.: Curriculum labeling: Revisiting pseudo-labeling for semi-supervised learning. In *AAAI Conference on Artificial Intelligence*, 35(8), pp. 6912-6920, (2021).

[4]   Chen, H., Jin, Y., Jin, G., Zhu, C., & Chen, E.: Semisupervised semantic segmentation by improving prediction confidence. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(9), 4991-5003 (2021).

[5]   Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In *European Conference on Computer Vision (ECCV)*, pp. 801-818, (2018).

[6]   Chen, M., Du, Y., Zhang, Y., Qian, S., & Wang, C.: Semi-supervised learning with multi-head co-training. In *AAAI Conference on Artificial Intelligence, 36(6)*, pp. 6278-6286, (2022).

19

[7] Chen, X., Yuan, Y., Zeng, G., & Wang, J.: Semi-supervised semantic segmentation with cross pseudo supervision. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2613-2622 (2021).

[8] Cheng, Z., Liang, J., Choi, H., Tao, G., Cao, Z., Liu, D., & Zhang, X.: Physical attack on monocular depth estimation with optimal adversarial patches. In *European Conference on Computer Vision (ECCV)*, pp. 514-532, (2022).

[9] DeVries, T., & Taylor, G. W.: Improved regularization of convolutional neural networks with cutout. *arXiv preprint arXiv:1708.04552* (2017).

[10] Fan, J., Cui, L., & Fei, S.: Waste detection system based on data augmentation and YOLO_EC. *Sensors*, *23*(7), 3646 (2023).

[11] Ferronato, N., & Torretta, V.: Waste mismanagement in developing countries: A review of global issues. *International Journal of Environmental Research and Public Health*, *16*(6), 1060 (2019).

[12] French, G., Laine, S., Aila, T., Mackiewicz, M., & Finlayson, G.: Semi-supervised semantic segmentation needs strong, varied perturbations. *arXiv preprint arXiv:1906.01916* (2019).

[13] Gong, C., Wang, D., & Liu, Q.: Alphamatch: Improving consistency for semi-supervised learning with alpha-divergence. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13683-13692 (2021).

[14] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y.: Generative adversarial networks. *Communications of the ACM*, *63*(11), 139-144 (2020).

[15] Gundupalli, S. P., Hait, S., & Thakur, A.: Multi-material classification of dry recyclables from municipal solid waste based on thermal imaging. *Waste Management*, *70*, 13-21 (2017).

[16] Guo, H., Mao, Y., & Zhang, R.: Mixup as locally linear out-of-manifold regularization. In *AAAI Conference on Artificial Intelligence* 33(1), pp. 3714-3722 (2019).

[17] He, K., Zhang, X., Ren, S., & Sun, J.: Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778 (2016).

[18] He, R., Yang, J., & Qi, X.: Re-distributing biased pseudo labels for semi-supervised semantic segmentation: A baseline investigation. In *IEEE/CVF International Conference on Computer Vision*, pp. 6930-6940 (2021).

[19] Hung, W. C., Tsai, Y. H., Liou, Y. T., Lin, Y. Y., & Yang, M. H.: Adversarial learning for semi-supervised semantic segmentation. *arXiv preprint arXiv:1802.07934* (2018).

[20] Huo, X., Xie, L., He, J., Yang, Z., Zhou, W., Li, H., & Tian, Q.: ATSO: Asynchronous teacher-student optimization for semi-supervised image segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1235-1244 (2021).

[21] Kaza, S., Yao, L., Bhada-Tata, P., & Van Woerden, F.: *What a waste 2.0: A global snapshot of solid waste management to 2050*. World Bank Publications (2018).

[22] Kraft, M., Piechocki, M., Ptak, B., & Walas, K.: Autonomous, onboard vision-based trash and litter detection in low altitude aerial images collected by an unmanned aerial vehicle. *Remote Sensing*, *13*(5), 965 (2021).

[23] Lee, S., Lee, M., Lee, J., & Shim, H.: Railroad is not a train: Saliency as pseudo-pixel supervision for weakly supervised semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5495-5505 (2021).

[24] Liang, J. C., Cui, Y., Wang, Q., Geng, T., Wang, W., & Liu, D.: ClusterFormer: Clustering as a universal visual learner. In *Advances in Neural Information Processing Systems*, *37*, pp. 1-14 (2023).

[25]   Lin, C., Guo, M., Li, C., Yuan, X., Wu, W., Yan, J., ... & Ouyang, W.: Online hyper-parameter learning for auto-augmentation strategy. In *IEEE/CVF International Conference on Computer Vision*, pp. 6579-6588 (2019).

[26]   Liu, J., Pan, C., Yan, W.: Litter detection from digital images using deep learning. Springer Nature Computer Science (2022)

[27]   Liu, S., Zhi, S., Johns, E., & Davison, A. J.: Bootstrapping semantic segmentation with regional contrast. *arXiv preprint arXiv:2104.04465* (2021).

[28]   Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In *IEEE/CVF International Conference on Computer Vision*, pp. 10012-10022 (2021).

[29]   Mahajan, D., Girshick, R., Ramanathan, V., He, K., Paluri, M., Li, Y., ... & Van Der Maaten, L.: Exploring the limits of weakly supervised pretraining. In *European Conference on Computer Vision (ECCV)*, pp. 181-196 (2018).

[30]   Mao, W. L., Chen, W. C., Wang, C. T., & Lin, Y. H.: Recycling waste classification using optimized convolutional neural network. *Resources, Conservation and Recycling*, *164*, 105132 (2021).

[31]   Mittal, S., Tatarchenko, M., & Brox, T.: Semi-supervised semantic segmentation with high-and low-level consistency. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(4), 1369-1379 (2019).

[32]   Miyato, T., Maeda, S. I., Koyama, M., & Ishii, S.: Virtual adversarial training: A regularization method for supervised and semi-supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(8), 1979-1993 (2018).

[33]   Mohanraj, C., Senthilkumar, T., Chandrasekar, M., & Arulmozhi, M.: Conversion of waste plastics into sustainable fuel. In *Waste to Profit* (pp. 41-52). CRC Press.

[34]   Ouali, Y., Hudelot, C., & Tami, M.: Semi-supervised semantic segmentation with cross-consistency training. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12674-12684 (2020).

[35]   Pham, H., Dai, Z., Xie, Q., & Le, Q. V.: Meta pseudo labels. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11557-11568 (2021).

[36]   Proença, P. F., & Simoes, P.: TACO: Trash annotations in context for litter detection. *arXiv preprint arXiv:2003.06975* (2020).

[37]   Qi, J., Nguyen, M., Yan, W.: Small visual object detection in smart waste classification using Transformers with deep learning. International Conference on Image and Vision Computing New Zealand (2022).

[38]   Qi, J., Nguyen, M., Yan, W.: Waste classification from digital images using ConvNeXt. Pacific-Rim Symposium on Image and Video Technology (2022).

[39]   Qi, J., Nguyen, M., Yan, W.: CISO: Co-iteration semi-supervised learning for visual object detection. Multimedia Tools and Applications (2023).

[40]   Qiao, S., Shen, W., Zhang, Z., Wang, B., & Yuille, A.: Deep co-training for semi-supervised image recognition. In *European Conference on Computer Vision,* pp. 135-152 (2018).

[41]   Shrivastava, A., Gupta, A., & Girshick, R.: Training region-based object detectors with online hard example mining. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 761-769 (2016).

[42]   Sousa, J., Rebelo, A., & Cardoso, J. S.: Automation of waste sorting with deep learning. In *XV Workshop de Visão Computacional (WVC)*, pp. 43-48 (2019).

[43]   Wang, T., Cai, Y., Liang, L., & Ye, D.: A multi-level approach to waste object segmentation. *Sensors*, *20*(14), 3816 (2020).

[44]   Xiao, S., Dong, H., Geng, Y., & Brander, M.: An overview of China's recyclable waste recycling and recommendations for integrated solutions. *Resources, Conservation and Recycling*, *134*, 112-120 (2018).

[45]   Xie, Q., Dai, Z., Hovy, E., Luong, T., & Le, Q.: Unsupervised data augmentation for consistency training. In *Advances in Neural Information Processing Systems*, *33*, 6256-6268 (2020).

[46]   Yan, W. Introduction to Intelligent Surveillance. Springer (2019)

[47]   Yan, W. Computational Methods for Deep Learning. Springer (2023)

[48]   Yang, B., Xu, K., Wang, H., & Zhang, H.: Random Transformation of image brightness for adversarial attack. *Journal of Intelligent & Fuzzy Systems*, *42*(3), 1693-1704 (2022).

[49]   Yang, L., Qi, L., Feng, L., Zhang, W., & Shi, Y.: Revisiting weak-to-strong consistency in semi-supervised semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7236-7246 (2023).

[50]   Yang, L., Zhuo, W., Qi, L., Shi, Y., & Gao, Y.: ST++: Make self-training work better for semi-supervised semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4268-4277 (2022).

[51]   Yang, M., & Thung, G.: Classification of trash for recyclability status. *Project Report*, Standord University (2016).

[52]   Yun, S., Han, D., Oh, S. J., Chun, S., Choe, J., & Yoo, Y.: CutMix: Regularization strategy to train strong classifiers with localizable features. In *IEEE/CVF International Conference on Computer Vision,* pp. 6023-6032 (2019).

[53]   Zhao, Z., Yang, L., Long, S., Pi, J., Zhou, L., & Wang, J.: Augmentation matters: A simple-yet-effective approach to semi-supervised semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11350-11359 (2023).

[54]   Zhang, Q., Zhang, X., Mu, X., Wang, Z., Tian, R., Wang, X., & Liu, X.: Recyclable waste image recognition based on deep learning. *Resources, Conservation and Recycling*, *171*, 105636 (2021).