

# Method for C/N ratio estimation using Mask R-CNN and a depth camera for organic fraction of municipal solid wastes

Jingjing Huang<sup>1,3</sup>, Hongduo Zhang<sup>2</sup>, Xu Xiao<sup>1</sup>, Jingqi Huang<sup>3</sup>, Jingxin Xie<sup>1</sup>, Liang Zhang<sup>1</sup>, Heming Hu<sup>2</sup>, Sihui Dai<sup>4\*</sup>, Ming Li<sup>3\*</sup>, Yongwei Xu<sup>5</sup>

(1. College of Mechanical and Electrical Engineering, Hunan Agricultural University, Changsha 410128, China;

2. Graduate School of Agricultural and Life Sciences, the University of Tokyo, Tokyo 113-0033, Japan;

3. Hunan Agricultural Equipment Research Institute, Changsha 410125, China;

4. College of Horticulture, Hunan Agricultural University, Changsha 410128, China;

5. Center for Spatial Information Science, University of Tokyo, Kashiwa 277-8568, Japan)

**Abstract:** Fast assessment of the initial carbon to nitrogen ratio (C/N) of organic fraction of municipal solid waste (OFMSW) is an important prerequisite for automatic composting control to improve efficiency and stability of the bioconversion process. In this study, a novel approach was proposed to estimate the C/N of OFMSW, where an instance segmentation model was applied to predict the masks for the waste images. Then, by combining the instance segmentation model with the depth-camera-based volume calculation algorithm, the volumes occupied by each type of waste were obtained, therefore the C/N could be estimated based on the properties of each type of waste. First, an instance segmentation dataset including three common classes of OFMSW was built to train mask region-based convolutional neural networks (Mask R-CNN) model. Second, a volume measurement algorithm was proposed, where the measurement result of the object was derived by accumulating the volumes of small rectangular cuboids whose bottom area was calculated with the projection property. Then the calculated volume was corrected with linear regression models. The results showed that the trained instance segmentation model performed well with average precision scores  $AP_{50}=82.9$ ,  $AP_{75}=72.5$ , and mask intersection over unit (Mask IoU)=45.1. A high correlation was found between the estimated C/N and the ground truth with a coefficient of determination  $R^2=0.97$  and root mean square error RMSE=0.10. The relative average error was 0.42% and the maximum error was only 1.71%, which indicated this approach has potential for practical applications.

**Keywords:** carbon to nitrogen ratio, estimation, volume measurement, organic fraction of municipal solid waste, depth camera, instance segmentation

**DOI:** 10.25165/ijabe.20211405.6382

**Citation:** Huang J J, Zhang H D, Xiao X, Huang J Q, Xie J X, Zhang L, et al. Method for C/N ratio estimation using Mask R-CNN and a depth camera for organic fraction of municipal solid wastes. *Int J Agric & Biol Eng*, 2021; 14(5): 222–229.

## 1 Introduction

Urbanization and rapid growth in human population have resulted in generation of vast amounts of wastes, where the degradable waste, i.e., organic fraction of municipal solid waste (OFMSW), normally accounts for more than half of the total

amount<sup>[1,2]</sup>. Among the organic waste treatment technologies, composting is a widely implemented technology owing to its cost-effectiveness and environmental friendliness<sup>[3,4]</sup>. Composting is a bioconversion process that converts biodegradable waste to nutrient-rich products which can be used as biofertilizer to enhance the fertility of the soil<sup>[5,6]</sup>, biopesticides to manage plant diseases, and bioremediation to stabilize heavy metals in the contaminated soil<sup>[7-9]</sup>. During the transformation of organic waste in the composting process, carbon and nitrogen are the most vital nutrients for microorganisms to build cell structures and acquire metabolic energy<sup>[10]</sup>, and the initial Carbon to Nitrogen ratio (C/N) of feedstock is one of the crucial factors influencing degradation rate and nutrients losses, which can be adjusted with different bulking agent<sup>[11-13]</sup>.

To meet the prerequisite of adjusting C/N, analytical technique such as elemental analyzers is often used in many studies to determine carbon content and total nitrogen of the feedstock<sup>[14,15]</sup>; it is accurate but too complicated to be performed automatically. Further, there are some studies on fast assessment of the C/N of materials. For instance, Althaus et al.<sup>[16]</sup> proposed an approach based on near-infrared reflectance spectroscopy (NIRS) for estimation of nitrogen and carbon content in animal feces with a high coefficient of determination ( $R^2$ ) of 0.97. Mikhailova et al.<sup>[17]</sup> proposed a relatively inexpensive approach to predict soil organic carbon and total nitrogen based on color sensors with  $R^2$  of 0.959

**Received date:** 2020-12-28 **Accepted date:** 2021-04-23

**Biographies:** **Jingjing Huang**, MS candidate, research interests: agricultural intelligent equipment, Email: huangjingjing2016@gmail.com; **Hongduo Zhang**, PhD candidate, research interests: agricultural robot, Email: zhdhnr@outlook.com; **Xu Xiao**, PhD candidate, research interests: agricultural robot, Email: 395949410@qq.com; **Jingqi Huang**, MS, research interests: agricultural intelligent equipment, Email: 1251820633@qq.com; **Jingxin Xie**, PhD candidate, research interests: precise crops protection system, Email: xiejingxin1021@163.com; **Liang Zhang**, PhD candidate, research interests: agricultural navigation system, Email: 18810987071@163.com; **Heming Hu**, PhD candidate, research interests: precise spraying system, Email: hu-heming950211@g.ecc.u-tokyo.ac.jp; **Yongwei Xu**, PhD, research interests: remote sensing image analysis, Email: xyw@csis.u-tokyo.ac.jp.

\***Corresponding author:** **Sihui Dai**, PhD, Associate Professor, research interests: horticultural crops information application. College of Horticulture, Hunan Agricultural University, No.1 Nongda Road, Furong District, Changsha 410128, China. Tel: +86-15974298578, Email: daisihui@126.com; **Ming Li**, PhD, Professor, research interests: crops sensing information acquisition and agricultural robots. Hunan Agricultural Equipment Research Institute, No.480, Donghu Road, Furong District, Changsha 410125, China. Tel: +86-15974170086, Email: liming@hunau.net.

and 0.912, respectively. A similar idea was presented by Morais et al.<sup>[18]</sup> where histograms of images were used to build a model for the prediction of soil organic carbon with a correlation of 0.93 between estimation and ground truth. However, these techniques only work well on homogeneous materials, thus it remains a challenge to measure C/N of heterogeneous materials, such as OFMSW. To our knowledge, there are no studies for estimation of C/N of OFMSW. In the field of dietary assessment, the nutritional parameters were estimated by the food type and its corresponding volume. Lo et al.<sup>[19]</sup> proposed an approach for food volume estimation where a mask region-based convolutional neural networks (Mask R-CNN) instance segmentation model was trained to generate mask predictions and then incorporated with a depth map obtained from a depth camera. It is possible to estimate C/N of OFMSW by combining the techniques of volume measurement and instance segmentation.

To realize volume measurement, methods based on stereo reconstruction, shape templates, depth camera, and deep learning have been studied<sup>[20]</sup>, in which the depth camera based method is more reliable than other methods. Long et al.<sup>[21]</sup> proposed a depth-camera-based method for volume estimation of potatoes with an error range from 9% to 30%, where the volume was calculated as the accumulation of volumes of rectangular cuboids. The bottom areas of these cuboids were projected areas from pixels on a calibration surface, and the heights of it were the difference between two depth maps before and after the object was placed. The bottom areas of the rectangular cuboids were not well considered but simply applied a fixed value, which could be one of the major factors affecting accuracy. On the other hand, with the great advances of computer vision and deep learning, there is a rising trend in the implementation of convolutional neural networks and object detection models to waste classification<sup>[22-24]</sup>. However, few studies focused on instance segmentation of OFMSW.

In this study, an integrated approach based on depth camera and deep learning was proposed to estimate the initial C/N of OFMSW for automatic composting control. Specifically, we studied the hyper-parameters determination and volume calculation algorithm optimization for better accuracy. Experiments were conducted to evaluate the performance of the proposed approach.

## 2 Materials and methods

The flowchart of the method is shown in Figure 1. Firstly, image and depth data were acquired. Next, the instance segmentation model was trained, and the volume measurement algorithm was constructed. Finally, the C/N estimation algorithm was built by combining the instance segmentation model with the volume measurement algorithm.

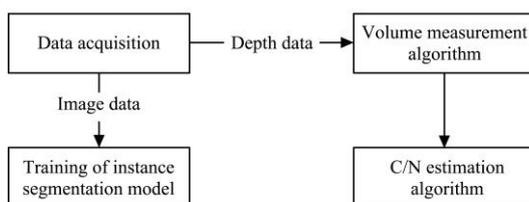


Figure 1 Flowchart of the C/N estimation method

### 2.1 Samples and data acquisition

#### 2.1.1 Waste samples

The composition of OFMSW varies greatly by region, season, and socio-economic effect, mainly including food waste, kitchen waste, and yard waste. Among them, wasted food accounts for a

large portion of the total amount. Food waste in China consists mainly of vegetables (54% wet base), rice (13%), fruit (13%), and other items (20%)<sup>[25]</sup>. To evaluate the effectiveness of the C/N estimation method proposed in this paper, three types of organic waste, namely lettuce, steamed rice, and banana, were chosen as representations for vegetables, rice and fruit respectively. The steamed rice was recycled leftovers from local restaurants, while the lettuce and bananas were collected from local markets. The density and the moisture content of each sample were measured using the drainage and drying methods, respectively. The measurement of each sample was repeated three times and then averaged. Table 1 shows the basic properties of each category.

**Table 1 Basic physical and chemical properties of samples**

Name	$\rho/\text{g}\cdot\text{cm}^{-3}$	MC/%	C/%	N/%	C/N
lettuce <sup>[26]</sup>	0.92	95.8	49.84	2.59	19.24
Steamed rice <sup>[27]</sup>	1.147	54.5	42.5	1.8	23.61
banana <sup>[28]</sup>	0.96	80.3	31.52	1.05	30.02

Note:  $\rho$  represents specific weight or unit weight; MC represents the moisture content of the sample; C, N, C/N represent carbon content, nitrogen content, and carbon to nitrogen ratio, respectively.

#### 2.1.2 Image and depth data acquisition

The dataset for training the instance segmentation model includes two parts, as shown in Figure 2. The first part consists of 75 images taken by the data acquisition system, and each of which was cropped to a resolution of  $1024\times 1024$  pixels. While another part contained 1500 images, which were downloaded from search engines such as Baidu and Bing and were screened manually. These two parts were combined into one dataset of 1575 images, named OrganicWaste-3.



Figure 2 Image samples for model training

A data acquisition system was built, as shown in Figure 3, consisting of a computer, an imaging platform inside a dark box, a ring-shaped LED light source, and a binocular camera ZED produced by Stereolabs Inc. ZED can output multiple resolutions up to  $1242\times 2208$  with a depth range from 0.3 m to 25 m. When acquiring the highest resolution, the frame rate was 15 fps, and the field of view along the horizontal axis and vertical axis were  $77.13^\circ$  and  $48.31^\circ$  which were obtained by the application programming interface (API) of the camera. ZED was set 0.65 m above the imaging platform. Unlike depth-sensing systems based on red green blue depth (RGB-D) sensors such as Kinect V2 and Intel Real Sense 435, ZED acquires depth only based on the disparity between the images captured by the left and right eye of the binocular camera. Therefore, a closed dark room was used to prevent sudden ambient light changes affecting operation stability of the camera, and the light source was provided by a ring-shaped light-emitting diode (LED) array. The camera was calibrated by ZED Calibration software automatically. Parameters of the

camera were set as auto exposure, depth acquisition mode was selected as FULL, and both confidences of depth and texture were set as 100 while remaining parameters were set as default. The depth camera acquires depth map based on disparity which can be calculated by epipolar geometry<sup>[29,30]</sup>. The data format provided for storage was a matrix with a shape of  $h \times w \times 3$ , where  $h$  and  $w$  were pixel numbers of the sides along height and width of the matrix, and 3 meant 3 channels. For better accuracy, we selected the highest resolution. So, the shape of the acquired data was  $1242 \times 2208 \times 3$ . To match the input size of the image segmentation model applied in this paper, all data were cropped to a resolution of  $1024 \times 1024 \times 3$ .

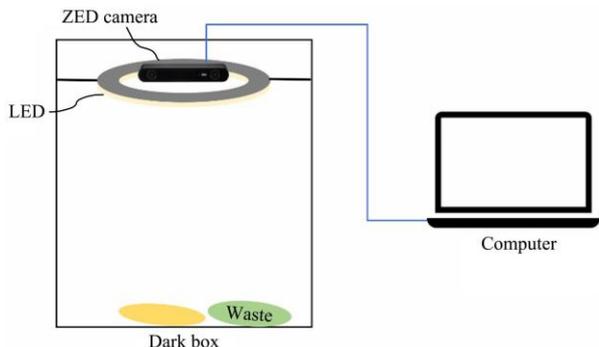


Figure 3 Schematic diagram of the system

Previous studies have shown that some depth cameras have a period of warm-up time before outputting reliable depth measurements, which causes a time-shift phenomenon<sup>[31]</sup>. To investigate such characteristics of ZED, the camera was mounted perpendicular to a plane, and depth measurements were taken every 5 seconds by averaging  $3 \times 3$  pixels at the center of the depth map.

The depth measurement results of ZED, as shown in Figure 4, exhibited an obvious time-shift phenomenon and reached a steady

state in 25 min. Therefore, in this paper, depth data acquisition was only performed after 25 min to reduce the error caused by the time-shift phenomenon.

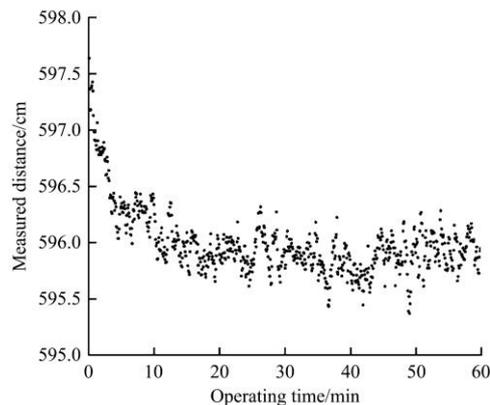


Figure 4 Time-shift phenomenon of ZED

2.2 Instance segmentation

Since the application of region-based convolutional neural network (R-CNN)<sup>[32]</sup> to object detection, several modified versions have emerged that significantly improved performances. Mask R-CNN is one of these branches, which has a strong instance segmentation capability<sup>[33]</sup>. The loss function of Mask R-CNN model is a multi-task loss function containing class loss, bounding box loss, and mask loss<sup>[33]</sup>. The architecture of Mask R-CNN model is shown in Figure 5, where the region proposal network (RPN) is implemented to generate bounding box proposals. To improve the accuracy, bilinear interpolation is used when mapping the proposed bounding box to a size-fixed region of interest (ROI), namely ROIAlign layer. Finally, the model output with three branches for classification, bounding box regression, and mask generation.

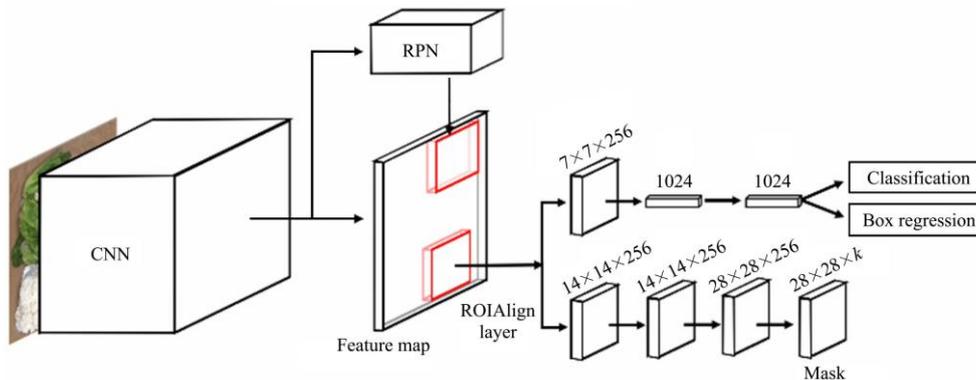


Figure 5 Architecture of Mask R-CNN model

The training of the Mask R-CNN model was performed with Tensorflow<sup>[34]</sup>. The computer used for training was equipped with an Intel i9-10920 CPU and an RTX TITAN graphic card. The model was trained with the OrganicWaste-3 dataset using mini-batch stochastic gradient descent algorithms<sup>[35]</sup> and the initial weights were pre-trained with a public dataset (COCO)<sup>[36]</sup> which is a large-scale dataset consisting of 91 classes of common objects.

The main hyperparameter settings were as follows: Batch size was 2, momentum was 0.9, the initial learning rate was 0.001, and the learning rate decay regularization was 0.0001. ResNet-101 was used as the backbone. The step sizes for each layer in feature pyramid networks (FPN) were (4, 8, 16, 32, 64) and the lengths of anchor edge were (32, 64, 128, 128, 256, 512). Width to height ratios when generating anchors were (0.5, 1, 2). The number of anchors generated from each image was 256, and the weight of

each loss function was 1.

The training was conducted in three steps. (i) Train the heads layer at a large learning rate of 0.001 during 1-14 epochs. (ii) Train all layers of Mask R-CNN with the same learning rate during 15-20 epochs. (iii) Train all layers at a lower learning rate of 0.0001 during 21-60 epochs.

To ensure that the available training samples were as many as possible, the share of the training set, validation set, and test set were set as 8:1:1. Normally, image augmentation is a suitable way to tackle the problem of data insufficiency for training of image recognition or object detection models by generating variants of original samples with operations, such as blurring, flipping, and affine transformations. To find out the best operations of augmentation, three sets of models were trained using image augmentation operations A, B, and a control C respectively.

The operations were selected abided by the principle that the augmented image is recognizable. Operation A was performed with the following operations with a probability of 0.5: horizontal flipping, vertical flipping, gaussian blurring with a standard deviation of 0 to 3, and some affine transformations. The affine operations included scaling along axes with a ratio from 0.8 to 1.2 randomly, translation along the axis directions with 10% of the side length, and random rotation with an angle of  $-45^\circ$  to  $45^\circ$ . Operation B did not include affine transformations and its remaining operations were identical to operation A. There are no augmentation operations applied in the control operation C.

The average precision (AP)<sup>[36]</sup> is a common metric for accuracy evaluation of object detection models, which can be calculated as the following equation

$$AP = \int_0^1 p(r) dr \quad (1)$$

where,  $p(r)$  is the precision-recall function in which the precision and the recall can be calculated as the following equations:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

where, TP is the number of correctly predicted positive samples; FP is the number of negative samples incorrectly predicted as positive and FN is the number of positive samples incorrectly predicted as negative.

Whether the prediction is a TP, an FP, or an FN was classed by a predefined intersection over union (IoU) threshold. And 0.5 and 0.75 are common IoU thresholds when calculating the AP metric, namely AP<sub>50</sub> and AP<sub>75</sub>.

The averaged mask IoU was used as a metric to evaluate the instance segmentation performance of the trained model. Mask IoU was defined as the following equation

$$\text{Mask IoU} = \frac{1}{t} \sum_{i=1}^t \frac{A(\text{mGT}_i \cap \text{mPre}_i)}{A(\text{mGT}_i \cup \text{mPre}_i)} \quad (4)$$

where,  $t$  is the number of the sample;  $\text{mGT}_i$  and  $\text{mPre}_i$  are total ground truth mask and total predicted mask of the  $i$ -th sample respectively, in which the total mask is obtained by performing a logical AND operation on the masks generated.  $A(\cdot)$  represents the area calculation function of the mask.

### 2.3 Volume measurement algorithm

A volume measurement algorithm based on depth map was proposed, where the object in the depth map was considered as a combination of small rectangular cuboids, as shown in Figure 6. The volume of the object was derived by accumulating the volumes of these tiny rectangular cuboids.

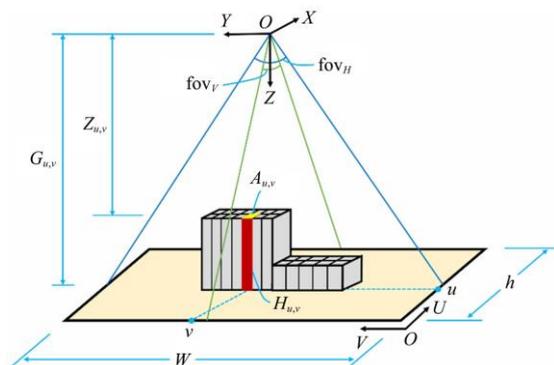


Figure 6 Schematic diagram of volume calculation

The shape of the depth map acquired is  $h \times w$  pixels, in which

each pixel contains depth values in the camera coordinate system  $OXYZ$ . Therefore, the volume of rectangular cuboid at any point  $(u, v)$  on the depth map can be obtained by calculating the bottom area  $A_{u,v}$  and the height  $H_{u,v}$ .

To calculate the heights of these rectangular cuboids, two depth maps before and after placing the object were acquired. Since the data are indexed by the pixel coordinate system, named  $OUV$ , the  $X$  axis and  $Y$  axis data are redundant. So, the depth map that only retained  $Z$  axis data was used for further calculation.

The depth map after placing the object is converted to a height matrix  $H$  by the following equation

$$H_{u,v} = \begin{cases} G_{u,v} - Z_{u,v} & \text{when } \sigma_{\min} \leq Z_{u,v} \leq \sigma_{\max} \\ G_{u,v} & \text{other wise} \end{cases} \quad (5)$$

where,  $u$  and  $v$  are pixel indexes in the ranges of  $[0, h-1]$  and  $[0, w-1]$  respectively;  $\sigma_{\min}$  and  $\sigma_{\max}$  are thresholds for screening outliers, and  $G_{u,v}$  is the reference for heights calculation, which is calculated as the average of the depth map of the background.

$$G_{u,v} = \frac{1}{h \times w} \sum_{u=0}^{h-1} \sum_{v=0}^{w-1} Z'_{u,v} \quad (6)$$

where  $Z'$  is the depth map of the imaging background.

For the calculation of the bottom areas, Long et al.<sup>[21]</sup> implemented a calibration method, where a cube with known side lengths was used to calculate the bottom areas at a certain sensing distance. However, this approach resulted in large error since the variation of sensing distance caused by the irregular shape of object. Thus, we proposed a new method where the projection characteristics of the camera were modeled to calculate the bottom area at different depths. The lengths and the widths are calculated by the following equations:

$$m_{u,v} = \frac{2Z_{u,v} \tan\left(\frac{\text{fov}_V}{2}\right)}{h} \quad (7)$$

$$n_{u,v} = \frac{2Z_{u,v} \tan\left(\frac{\text{fov}_H}{2}\right)}{w} \quad (8)$$

where,  $\text{fov}_V$  and  $\text{fov}_H$  are field of view angles along the vertical and horizontal axes of the image, respectively. Consequently, the bottom area of rectangular cuboids can be calculated by the following equation:

$$A_{u,v} = m_{u,v} n_{u,v} \quad (9)$$

Finally, the volume of an object can be calculated by the following equation:

$$V = \sum_{u=0}^{h-1} \sum_{v=0}^{w-1} A_{u,v} H_{u,v} \quad (10)$$

Volume  $V_i$  for the waste class of  $i$  can be measured by the combination of instance segmentation model and volume measurement algorithm proposed, as shown below:

$$V_i = \sum_{u=0}^{h-1} \sum_{v=0}^{w-1} A_{u,v} H_{u,v} M_i \quad (i = 1, 2, \dots, k) \quad (11)$$

where,  $k$  is the number of the classes of the waste;  $M_i$  is a mask, which is a matrix generated by Mask R-CNN model for the waste class of  $i$  with a shape of  $h \times w$  and the values of its pixels are 0 or 1.

### 2.4 C/N estimation of the waste

The total carbon to nitrogen ratios of the waste samples are calculated by the following equation

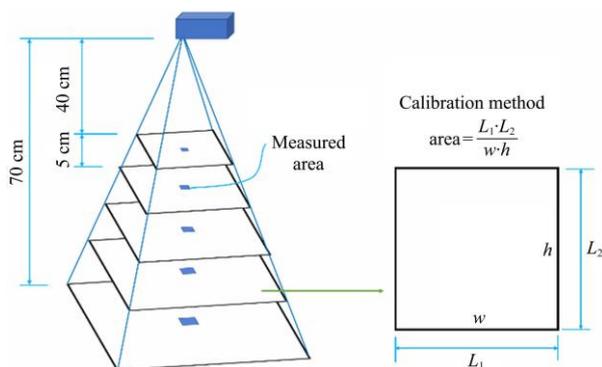
$$\frac{C}{N} = \frac{\sum_{i=1}^k V_i \rho_i (1 - MC_i) C_i}{\sum_{i=1}^k V_i \rho_i (1 - MC_i) N_i}, \quad (i = 1, 2, \dots, k) \quad (12)$$

where, C/N represents the carbon to nitrogen ratio;  $V_i$  represents the volume of the  $i$ -th type of waste sample;  $\rho_i$ ,  $MC_i$ ,  $C_i$ , and  $N_i$  are specific weight, moisture content, carbon content, and nitrogen content for the  $i$ -th type of waste, respectively.

## 2.5 Experiments for evaluation

### 2.5.1 Evaluation of the bottom area measurement algorithm

To evaluate the accuracy of measurement of the bottom areas, the depth maps of an empty plane perpendicular to the optical axis were captured at 7 distances distributed from 0.4 to 0.7 m with a spacing of about 5 cm, as shown in Figure 7. The bottom area at the center of each depth map was calculated by Equation (9). Meanwhile, the ground truth values of the bottom areas were obtained by applying the calibration method at each sensing distance, in which the side lengths of the bottom areas were calculated by dividing manually measured lengths with the numbers of pixels along the sides of the depth map.



Note:  $L_1$  and  $L_2$  are manually measured side lengths of the square that has  $h \times w$  pixels.

Figure 7 Evaluation of the accuracy of bottom area calculation

Coefficient of determination ( $R^2$ ) and root mean square error (RMSE) were used for performance evaluation of the regression between measured bottom areas and the ground truth values.

Relative average error (RAE) and maximum relative error (MRE) were computed to express the error.

$$RAE = \frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{y_i} \times 100\% \quad (13)$$

$$MRE = \max_{1 \leq i \leq n} \left( \frac{|\hat{y}_i - y_i|}{y_i} \right) \times 100\% \quad (14)$$

where  $n$  is the number of samples;  $\hat{y}_i$  is estimated value and  $y_i$  is ground truth value and  $i$  is sample index.

### 2.5.2 Evaluation of the volume measurement algorithm

To evaluate the performance of the algorithm for volume measurement, a 3D printed object, a 10 cm×10 cm×10 cm cube whose volume is 1000 cm<sup>3</sup>, was used for testing. The heights of rectangular cuboids were calculated by Equation (5). The bottom areas were calculated using Equation (9). For comparison, we also calculated the bottom areas by applying the calibration method<sup>[21]</sup> at the background plane. The result for each method was the average of 3 replications, in which the location of the cube was changed for each testing. The RAE was used to evaluate the performances.

### 2.5.3 Evaluation of volume calculation

The accuracy of volume measurement is affected by the huge difference in porosity of the materials due to different structural properties. To tackle this problem, linear regression models were established for each type of waste between estimated volumes and measured volumes.  $R^2$ , RMSE, RAE, and MRE were used for the performance evaluation of the regression.

8 samples were selected for each waste class, whose volumes were linearly distributed. The volume measurements were calculated by Equation (11) and repeated 5 times, in which the positions of the samples were changed for each testing. While the ground truth volume, namely  $V'$ , was reference value derived from the sample's weighted weight and its specific weight, as the following equation:

$$V' = \frac{\text{weight}}{\rho} \quad (15)$$

where  $\rho$  is the specific weight of samples as shown in Table 1.

### 2.5.4 Evaluation of C/N estimation

To evaluate the capability and accuracy of the C/N estimation algorithm proposed in this paper, 27 samples were used for the experiment, each of which was a combination of samples from 9 individual samples, in which each class of waste has 3 different weights.

The estimated total carbon to nitrogen ratios of the waste samples was calculated by Equation (12) where the measured volumes  $V$  were calculated by Equation (11). While the ground truth values were calculated by Equation (12) using volumes  $V'$  calculated by Equation (15). The RAE,  $R^2$ , and RMSE were used as metrics for evaluation.

## 3 Results

### 3.1 Performance of instance segmentation

The examples of instance segmentation were shown in Figure 8. The predicted masks of steamed rice were close to the ground truth, but bananas and lettuces had different results when applying different augmentation operations. The models trained with augmentation operations performed much better than those trained with control C. The augmentation operations reduced detection errors of lettuces and enhanced the accuracy when generating masks for the bananas occluded. The model trained with operation A performed better than that trained with operation B.

The performance metrics of the trained models were shown in Table 2, where the scores of the models trained with operation A and operation B were higher than that trained with control C. These results showed that the applied augmentation operations are suitable for the images involved in this study. The results of the model trained with operation A ( $AP_{50}=82.9$ ,  $AP_{75}=72.5$ , Mask IoU = 45.1) were better than that of the model trained with operation B. Hence, the model trained with augmentation operation A was selected for further usage.

Table 2 Performance scores of models trained with different augmentation operations

Operations	AP <sub>50</sub>	AP <sub>75</sub>	Mask IOU
A	82.9	72.5	45.1
B	79.9	70.4	44.6
C	80.3	68.9	44.2

### 3.2 Performance of the bottom area measurement algorithm

A good correlation was found between measured bottom areas and ground truth values as shown in Figure 9. The  $R^2$  was larger than 0.99 and the RMSE was 0.0046. In addition, the relative average error (RAE) was 2.43% and the maximum relative error (MRE) was 3.12%. This result shows that the proposed algorithm performs well in calculating the bottom areas of tiny rectangular cuboids at various distances.

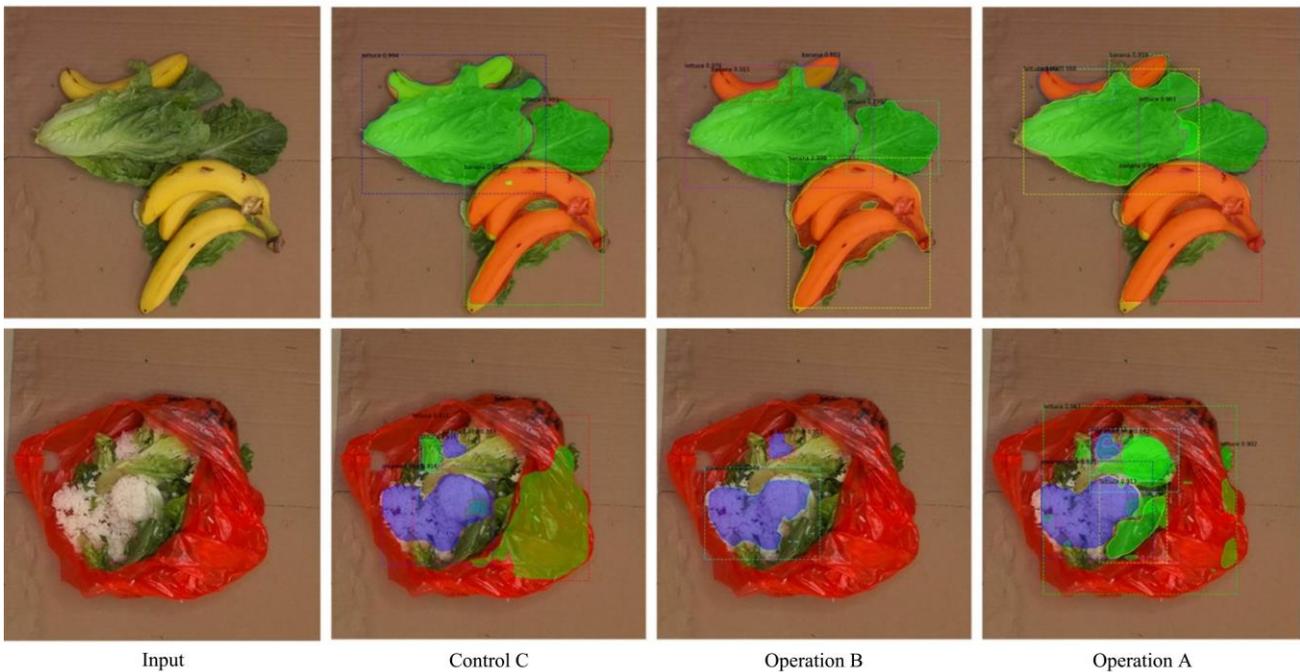


Figure 8 Instance segmentation results obtained by models that trained with different image augmentation processes

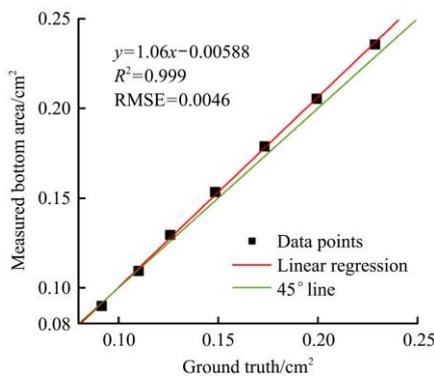


Figure 9 Linear regression plot for the measured bottom areas against the ground truth

**3.3 Performance of the volume measurement algorithm**

The average volume of the testing cube calculated by the

proposed algorithm was closer to the ground truth than the result obtained by the calibration method, in Table 3. The results showed that our approach was able to calculate volume of cube with a relative average error of 2.77%.

**Table 3 Volume measurement results**

Method for calculating the bottom area	Average volume/cm <sup>3</sup>	Standard error	RAE/%
Calibration	1274.08	10.8	27.41
Proposed algorithm	972.32	7.82	2.77

**3.4 Performance of volume calculation**

The regression results showed that ground truth and measured volumes of banana, lettuce, and rice showed good linear correlations with  $R^2=0.985$ ,  $R^2=0.955$ , and  $R^2=0.970$ , the RAE of 11.24%, 38.24%, and 17.34% as well as the MRE of 26.8%, 46.34%, and 28.03%, respectively (Figure 10).

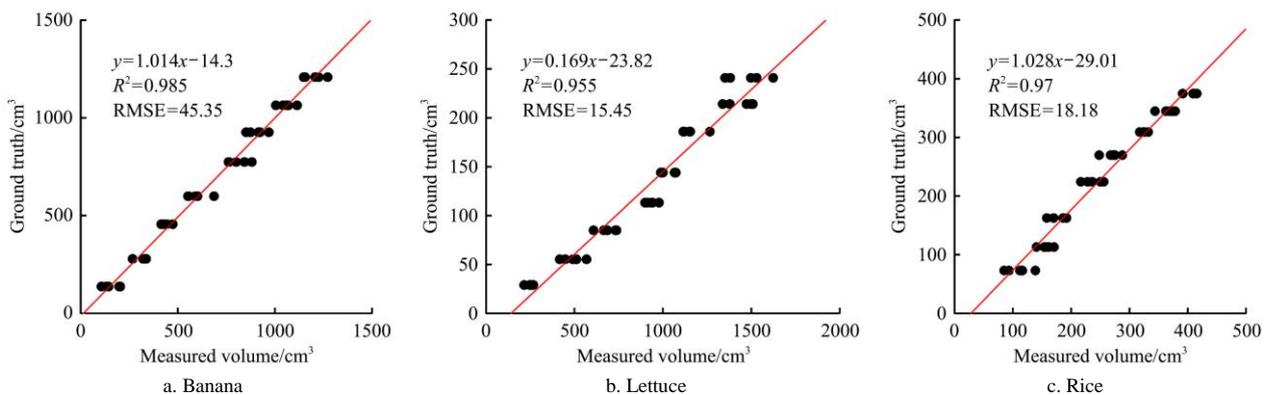


Figure 10 Linear regression plots for the measured volume against the ground truth volume

**3.5 Accuracy of C/N estimation**

The results showed that the RAE of the volume estimations for banana, lettuce, and rice were 11.5%, 38.2%, and 17.3%, with MRE of 59.8%, 55.8%, and 36.6%, respectively.

As shown in Figure 11, a high correlation was found between the estimated C/N results and the ground truth with  $R^2=0.97$ ,  $RMSE=0.10$ . The RAE of C/N estimation was 0.42%, and the

MRE was 1.71%.

The errors of volume calculation were mainly caused by the error of instance segmentation, such as the predicted masks were smaller than the ground truth contour. Despite the error obtained in volume calculation, the result showed a good performance for estimation of C/N as the C/N depended on the shares of each type of waste.

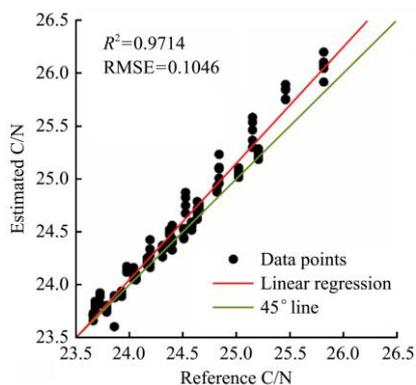


Figure 11 Linear regression plot for the C/N estimation results against the reference C/N values

## 4 Discussion

A novel method was proposed to estimate the C/N of OFMSW for automatic composting control and some good results were achieved. The proposed volume measurement algorithm was optimized by introducing the projection properties to the calculation of the bottom areas of rectangular cuboids representing the 3D shapes within the depth map. The trained instance segmentation model performed well with scores of AP<sub>50</sub>, AP<sub>75</sub>, and Mask IoU reached 82.9, 72.5, and 45.1, respectively. Then, the proposed C/N estimation method achieved a rather low RAE of 0.42% and MRE of 1.71%, which meant that the proposed approach could meet the technical demand for automatic composting control.

However, there are some concerns when implementing the method to practical applications. Since the estimated C/N values of this approach were derived from the measured volumes and their properties, the results were inevitably affected by the following factors that require further research.

First, though the instance segmentation model achieved good scores, the mask predictions were not robust enough. The reason could trace back to the structure of the Mask R-CNN model, in which the mask loss was associated with the class predicted by the classification branch which was not easy to recognize the waste with variations in shape and size or with occlusions. To obtain higher mask prediction accuracy, further study should try to optimize the classification branch of the Mask R-CNN model or apply panoptic segmentation model<sup>[37]</sup>.

Furthermore, the waste occlusions will inevitably cause errors in the calculation of the volumes. To solve this problem, there is a compromise solution which is to build an imaging system combined with a mechanical device. And integration of multiple sensors is worth trying.

On the other hand, there are situations that the feedstocks may be stored before composting, which might cause changes in its structure, appearance, and property. Therefore, to improve the capability in practical applications, the properties and the volume correction models of the waste at certain levels of rot should be considered in further studies.

## 5 Conclusions

A novel approach for estimation of the C/N of organic fraction of municipal solid waste (OFMSW) was proposed, which consisted of instance segmentation model, volume measurement algorithm as well as volume correction models. The instance segmentation model was trained to predict masks for three types of common

OFMSW, including banana, lettuce and rice, and the volume measurement algorithm was optimized resulted in higher accuracy. The results showed that the approach proposed could effectively estimate the initial C/N of feedstocks without occlusions.

## Acknowledgements

This research was funded by the National Key Research and Development Program of China (Grant No. 2018YFD0200800); Key Research and Development Program of Hunan Province (Grant No. 2018GK2013); Hunan Modern Agricultural Industry Technology Program (Grant No. 201926); Innovation and Entrepreneurship Training Program of Hunan Agricultural University (Grant No. 2019062x).

## [References]

- [1] Bölükbaş A, Akıncı G. Solid waste composition and the properties of biodegradable fractions in Izmir City, Turkey: an investigation on the influencing factors. *Journal of Environmental Health Science and Engineering*, 2018; 16(2): 299–311.
- [2] Ayilara M S, Olanrewaju O S, Babalola O O, Odeyemi O. Waste management through composting: Challenges and potentials. *Sustainability*, 2020; 12(11): 4456. Doi: 10.3390/su12114456.
- [3] Qian X, Shen G, Wang Z, Guo C, Liu Y, Lei Z, et al. Co-composting of livestock manure with rice straw: characterization and establishment of maturity evaluation system. *Waste Management*, 2014; 34(2): 530–535.
- [4] Onwosi C O, Ndukwe J K, Aliyu G O, Chukwu K O, Ezugworie F N, Igbokwe V C. Composting: An eco-friendly technology for sustainable agriculture. *Ecological and Practical Applications for Sustainable Agriculture*. Springer, 2020; pp.179–206.
- [5] Nakasaki K, Hirai H, Mimoto H, Quyen T N M, Koyama M, Takeda K. Succession of microbial community during vigorous organic matter degradation in the primary fermentation stage of food waste composting. *Science of The Total Environment*, 2019; 671: 1237–1244.
- [6] Pergola M, Persiani A, Palese A M, Di Meo V, Pastore V, D'Adamo C, et al. Composting: The way for a sustainable agriculture. *Applied Soil Ecology*, 2018; 123: 744–750.
- [7] Kulikowska D, Gusiatiń Z M, Bułkowska K, Kierklo K. Humic substances from sewage sludge compost as washing agent effectively remove Cu and Cd from soil. *Chemosphere*, 2015; 136: 42–49.
- [8] Hestmark K, Fernández-Bayo J, Harold D, Randall T, Achmon Y, Stapleton J, et al. Compost induces the accumulation of biopesticidal organic acids during soil biosolarization. *Resources, Conservation and Recycling*, 2019; 143: 27–35.
- [9] Zhou H, Shen Y, Li R, Meng H, Zhang X, Wang J, et al. Heavy metals and community structure of microorganism changes during livestock manure composting with inoculation of effective microorganisms. *Int J Agric & Biol Eng*, 2020; 13(6): 125–132.
- [10] Iqbal M, Nadeem A, Sherazi F, Khan R. Optimization of process parameters for kitchen waste composting by response surface methodology. *International Journal of Environmental Science and Technology*, 2015; 12(5): 1759–1768.
- [11] Ekinci K, Tosun İ, Kumbul B S, Şevik F, Sülük B, Bitrak N B. Effect of initial C/N ratio on composting of two-phase olive mill pomace, dairy manure, and straw. *Environmental Progress & Sustainable Energy*, 2021; 40(2): e13517. doi: 10.1002/ep.13517.
- [12] Ekinci K, Tosun İ, Bitrak B, Kumbul B, Şevik F, Sülük K. Effects of initial C/N ratio on organic matter degradation of composting of rose oil processing solid wastes. *International Journal of Environmental Science and Technology*, 2019; 16(9): 5131–5140.
- [13] Zhang L, Sun X. Improving green waste composting by addition of sugarcane bagasse and exhausted grape marc. *Bioresource technology*, 2016; 218: 335–343.
- [14] Kumar M, Ou Y-L, Lin J-G. Co-composting of green waste and food waste at low C/N ratio. *Waste management*, 2010; 30(4): 602–609.
- [15] Vázquez M, Soto M. The efficiency of home composting programmes and compost quality. *Waste Management*, 2017; 64: 39–50.
- [16] Althaus B, Papke G, Sundrum A. Technical note: Use of near infrared reflectance spectroscopy to assess nitrogen and carbon fractions in dairy

- cow feces. *Animal Feed Science and Technology*, 2013; 185(1): 53–59.
- [17] Mikhailova E A, Stiglitz R Y, Post C J, Schlautman M A, Sharp J L, Gerard P D. Predicting soil organic carbon and total nitrogen in the Russian Chernozem from depth and wireless color sensor measurements. *Eurasian Soil Science*, 2017; 50(12): 1414–1419.
- [18] de Oliveira Morais P A, de Souza D M, Madari B E, Soares A D S, de Oliveira A E. Using image analysis to estimate the soil organic carbon content. *Microchemical Journal*, 2019; 147: 775–781.
- [19] Lo F P W, Sun Y, Qiu J, Lo B. Food volume estimation based on deep learning view synthesis from a single depth map. *Nutrients*, 2018; 10(12).
- [20] Lo F P W, Sun Y, Qiu J, Lo B. Image-based food classification and volume estimation for dietary assessment: A review. *IEEE Journal of Biomedical and Health Informatics*, 2020; 24(7): 1926–1939.
- [21] Long Y, Wang Y, Zhai Z, Wu L, Li M, Sun H, et al. Potato volume measurement based on RGB-D camera. *IFAC-PapersOnLine*, 2018; 51(17): 515–520.
- [22] Adedeji O, Wang Z. Intelligent waste classification system using deep learning convolutional neural network. *Procedia Manufacturing*, 2019; 35: 607–612.
- [23] Wen C H, Li J, Dong X. Intelligent domestic garbage recognition based on Faster RCNN. *Laser & Optoelectronics Progress*, 2020; 57(20): 139–145. (in Chinese)
- [24] Li H B, Liao X G, Chen L. Automatic plastic detection system based on one-stage machine learning algorithm on object detection. *Plastics Science and Technology*, 2020; 48(12): 86–89. (in Chinese)
- [25] Negri C, Ricci M, Zilio M, D'Imporzano G, Qiao W, Dong R, et al. Anaerobic digestion of food waste for bio-energy production in China and Southeast Asia: A review. *Renewable and Sustainable Energy Reviews*, 2020; 133: 110–138.
- [26] Sudharsan V V, Kalamdhad A S. Evolution of chemical and biological characterization during thermophilic composting of vegetable waste using rotary drum composter. *International Journal of Environmental Science and Technology*, 2015; 12(6): 2015–2024.
- [27] Chang J I, Hsu T-E. Effects of compositions on food waste composting. *Bioresource Technology*, 2008; 99(17): 8068–74.
- [28] Ghinea C, Apostol L C, Prisacaru A E, Leahu A. Development of a model for food waste composting. *Environmental Science and Pollution Research*, 2019; 26(4): 4056–4069.
- [29] Faugeras O, Luong Q-T, Papadopoulos T. The geometry of multiple images: the laws that govern the formation of multiple images of a scene and some of their applications. MIT Press, 2001.
- [30] Hartley R, Zisserman A. Multiple view geometry in computer vision. Cambridge University Press, 2003.
- [31] Lachat E, Macher H, Mittet M A, Landes T, Grussenmeyer P. First experiences with Kinect V2 sensor for close range 3d modelling. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2015; 40(5): 93–100.
- [32] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, USA, 2014; 580–587.
- [33] He K, Gkioxari G, Dollár P, Girshick R. Mask r-cnn. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, 2017; 2980–2988.
- [34] Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. Tensorflow: A system for large-scale machine learning. *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*; 2016.
- [35] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998; 86(11): 2278–324.
- [36] Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. Microsoft coco: Common objects in context. *European Conference on Computer vision*, 2014.
- [37] Kirillov A, He K, Girshick R, Rother C, Dollár P. Panoptic segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.