

## Research Article

# Multilayer Hybrid Deep-Learning Method for Waste Classification and Recycling

Yinghao Chu ,<sup>1</sup> Chen Huang,<sup>1</sup> Xiaodan Xie,<sup>2</sup> Bohai Tan,<sup>3</sup> Shyam Kamal ,<sup>4</sup> and Xiaogang Xiong ,<sup>5</sup>

<sup>1</sup>AIATOR Co., Ltd., Block 5, Room 222, Qianwanyilu, Qianhai, Shenzhen, China

<sup>2</sup>Department of Industrial and Systems Engineering, Ohio University, Athens, OH, USA

<sup>3</sup>Sagacity Environment (China) Co. Ltd., A201 Qianwanyilu, Qianhai, Shenzhen, China

<sup>4</sup>Indian Institute of Technology (BHU), Varanasi, Uttar Pradesh 221005, India

<sup>5</sup>Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China

Correspondence should be addressed to Xiaogang Xiong; [xiong@ctrl.mech.kyushu-u.ac.jp](mailto:xiong@ctrl.mech.kyushu-u.ac.jp)

Received 12 August 2018; Accepted 24 September 2018; Published 1 November 2018

Guest Editor: Uma Seeboonruang

Copyright © 2018 Yinghao Chu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study proposes a multilayer hybrid deep-learning system (MHS) to automatically sort waste disposed of by individuals in the urban public area. This system deploys a high-resolution camera to capture waste image and sensors to detect other useful feature information. The MHS uses a CNN-based algorithm to extract image features and a multilayer perceptrons (MLP) method to consolidate image features and other feature information to classify wastes as recyclable or the others. The MHS is trained and validated against the manually labelled items, achieving overall classification accuracy higher than 90% under two different testing scenarios, which significantly outperforms a reference CNN-based method relying on image-only inputs.

## 1. Introduction

Globally, the annual solid waste is expected to reach 2.2 billion tonnes by 2025, which would cost \$375.5 billion in waste management [1]. Improper waste management will have enormous adverse impacts on the economy, the public health, and the environment [1]. Municipal solid waste (MSW) recycling has been recognized as the second “most environmentally sound” strategy for dealing with urban waste by the Environmental Protection Agency (EPA) [2]. Effective waste recycling is both economic and environmentally beneficial. It can help in recovering raw resource, preserving energy, mitigating greenhouse gaseous emission, water pollution, reducing new landfills, etc [1, 3–5].

In developing country, MSW recycling relies on household separation via scavengers and collectors who trade the recyclables for profits [6–8]. In developed countries, communities are more involved in recycling program [9]. Several techniques, such as mechanical sorting and chemical sorting, are available in developed countries for

automatic waste sorting [10]. However, there is huge potential to improve waste recycling even in the developed country. The municipal recycling rates of the USA and European Union are around 34% and 50%, respectively, which are significantly lower than the target recycling rate of 75% [5, 11].

The key obstacles to waste recycling include the following: (1) government plan and budget: insufficient government regulation and budget for MSW management; (2) household education: households are unaware of the importance of self-waste recycling; (3) technology: lack of effective recycling technology; and (4) management expense: the high cost of manual waste classification [1, 8, 9].

The recent progress in deep learning has contributed to unprecedented improvements in computer vision. Convolutional neural network (CNN) is one of the most recognized deep-learning algorithms for its wide application in image classification, segmentation, and detection [12–15]. Therefore in this literature, CNN is proposed to perform waste classification.

Awe et al. [16] propose an experimental project using a Faster R-CNN model to classify waste into three categories: paper, recycling, and landfill. This method achieves a mean average precision of 68%. Thung and Yang [17] deployed support vector machine (SVM) and a convolutional neural network (CNN) to classify waste into six categories. It achieves an accuracy rate of 63% for SVM and 23% for CNN. Rad et al. [18] developed a GoogLeNet-based vision application to localize and classify urban wastes. The study claims to have an accuracy rate ranging from 63% to 77% for different waste types. Donovan [19] proposed to use Google's TensorFlow and camera capturing to automatically sort waste objects as compost and recyclable. However, as a conceptual project, there is no experimental result so far. Mittal et al. [20] designed a project to detect whether an image contains garbage or not. This project employs the pretrain AlexNet model and achieves a mean accuracy of 87.69%. However, this project aims at segmenting garbage in an image without providing functions of waste classification.

As reviewed, the automatic classification methodologies available in the literature solely deploy image-based CNN and result in limited accuracy. In this work, we propose a multilayer hybrid method (MHS) to perform waste classification in public areas. Waste images associated with other numerical information measured by sensors are fed into the system. The system can automatically sort the waste item as recyclable or the others. The proposed MHS achieves a mean accuracy higher than 90%, which significantly outperforms reference image-based method.

The specific contributions of the paper are:

- (i) Firstly, this study achieves an excellent accuracy that is useful for in-field applications: the experimental results indicate that MHS achieves an overall accuracy higher than 90%, which outperforms all reference waste classification methods in the literature.
- (ii) Secondly, this study proposes an innovative architecture to simulate the sensory and intellectual process of human inspections. While most of current waste classification methods take images as the sole input, the proposed method makes use of an AlexNet CNN to act as "human eyes" to visualize and extract key image features from its last dense layer. The system also utilizes sensors to act as "ears" and "nose" to detect other numerical feature information, which are barely discussed in the literatures. Ultimately, multilayer perceptrons (MLP) act as a response center (the "human brain") to classify the waste object by consolidating information collected from diverse channels.

The paper is organized as follows: Section 2 introduces the inspected waste items, hardware, and data; Section 3 presents the proposed methods, including CNN, MLP, the multilayer hybrid system, and evaluation metrics; Section 4 presents the results and corresponding discussions; and conclusion are given in Section 5.

## 2. Hardware and Data

This study focuses on wastes found in urban public areas, including parks, street cleaning, landscaping, and other recreational areas. These wastes are mostly disposed of by individual visitors, pedestrians, commuters, and occasionally from commercial events. Unlike industry or household waste, the majority of municipal solid wastes are separated singular items e.g., a singular bottle or a singular lunch box [1].

This study analyzes a total of 50 different waste items that are commonly found in the investigated area [1]. Among which, 40 are recyclable and 10 are the others. The recyclable wastes are grouped into 4 main categories: paper, plastic, metal, and glass; the "others" class consists of fruit/vegetable/plant, kitchen waste, and others (Table 1). Each group is made up of representative items: the paper group, for instance, consists of books, magazines, cups, boxes, etc. Table 2 presents the detailed information of waste items' quantity, corresponded group, and class.

**2.1. Hardware.** The proposed system consists of a high-resolution camera (model 0V9712), a bridge sensor (model HX711AD), and an inductor (model TL-W3MB1 PNP) (Table 3). The system hardware is selected based on availability, low-cost, effectiveness, and easy installation.

The camera captures images of study objects, and the images will be transferred to a PC end via USB 2.0. The bridge sensor is used to measure the weight of study object, and the inductor can detect whether the waste is made from metal or not. To our knowledge, few studies employ sensing systems for waste classification in the literature of MSW, such as applications of medical waste and wastewater sorting using weight, density, and texture detection [3]. Therefore, we propose to deploy the bridge sensor and inductor to facilitate classification of solid waste, particularly for MSW waste. Digital information measured by sensors are received and processed via an Arduino board first. The Arduino board serves as a microcontroller, which can read the outputs from the sensors, convert them into proper numerical form, and then transfer information to the PC end.

In the experiment, the investigated waste items are placed in an enclosed box with a dark grey background. The camera is placed at the upper front-right of the experiment box to maximize the marginal angle of view. Waste objects are rotated for the camera to capture views from different angles, so to simulate a three-dimension effect. The bridge sensor and the inductor are placed directly under the study object to measure their corresponding feature information.

**2.2. Data.** A total of 100 RGB images are captured for each investigated item, and 5000 (50 × 100) images in total are collected in JPG format. Each waste image is grouped with its counterpart numerical feature information as a data

TABLE 1: Representative waste images.



instance, which is then manually labelled as either recyclable or not for training/testing purpose.

To enhance image features and to remove unwanted noise, images captured by the camera are preprocessed under the Keras framework: <http://keras.io/>. The original images (e.g., Figure 1) are  $640 \times 480$  pixels in resolution while the processed images are  $240 \times 240$ . During training, 9 augmented images (Table 4), including image rotation, height/width shifting, size rescaling, zooming etc., are generated for each data instance to enhance the universality of the training model [21].

The training model has been tested twice to validate the system performance. Firstly, each waste item is placed into the system with predefined position and each item is tested for 3 times. A total of 150 ( $50 \times 3$ ) test set are generated for the first test. Secondly, each item is placed randomly in the

system for 3 times and another set of 150 data is generated. In sum, 300 test data are created, and the system classifies each of them as recyclable or the others.

### 3. Method

A multilayer hybrid method (MHS), which consists of several subsystems, is proposed to perform waste classifications. The core of this system includes a convolutional neural network (CNN) and multilayer perceptrons (MLP). Evaluation metrics to access system performance are also discussed in this section.

**3.1. CNN.** Convolutional neural networks (CNN) are widely applied in analyzing visual image [12–15]. Generally, CNN

TABLE 2: Waste item.

Class	Group	Item	Quantity
Recyclable	Paper	Books	5
		Cups	5
		Boxes	4
		General bottles	6
		Shampoo bottles	4
	Plastic	Pen	1
		Watch	1
		Cans	7
		Key	1
	Metal	Scissor	1
		Beer cap	1
		Bottle	4
<i>Sum</i>	<i>4</i>	<b>12</b>	<b>40</b>
Others	Fruit/vegetable/plant	Apple	1
		Banana	1
		Carrot	1
		Cabbage	1
		Rose	1
	Kitchen waste	Others	1
		Egg	1
		Lunch box	1
	Others	Trash bag	1
		Bowl	1
<i>Sum</i>	<i>3</i>	<b>10</b>	<b>10</b>

takes images containing investigated items as inputs and classify images into different categories.

CNN is unique in its 3D volumes of a neuron: width, height, and depth. The CNN consists of a series of convolutional layers, polling layers, fully connected layers, and normalization layers [12–15]. The neurons in the convolutional layer will only connect to a small region of the previous layer. In fully connected layers, the activation neurons of the layer are fully connected to all activation neurons in the previous layer. The fully connected function can be expressed as the following forward and backward propagation rules in mathematical form:

$$\begin{aligned} X_i^{L+1} &= \sum_i W_{j,i}^{L+1} X_j^L, \\ g_i^L &= \sum_i W_{j,i}^{L+1} g_j^{L+1}, \end{aligned} \quad (1)$$

where  $X_i^L$  and  $g_i^L$  represent the activation and the gradient of neurons  $i$  at layer  $L$  and  $W_{j,i}^{L+1}$  is the weight connecting neurons  $i$  at layer  $L$  to neurons  $j$  at layer  $L + 1$ .

The capability of CNN can be controlled by varying dimensional parameters and local architecture structure [12]. In recent years, different CNN architecture variations emerge [12, 15]. In considering the computational cost and in-field application limitations, AlexNet [12] is employed in this work.

**3.2. AlexNet.** AlexNet [12] came to the spot in the 2012 ImageNet Challenge (ILSVRC) by significantly reducing the image classification top-5 error from 26% to 15.3%. It is well-recognized for its highly capable architecture.

AlexNet contains 8 learning layers: the first five convolutional followed by three fully connected layers. The output of the last layer is fed into the 1000-way softmax which can create 1000 class labels. The kernels of the second, fourth, and fifth layers are connected to those kernels in previous layers sharing the same GPU. Kernels in the third layer, however, are fully connected to the kernels in the second layer. Response-normalization layer is associated with the first and second layers. Max-pooling layers are placed after both response-normalization layers and the fifth layer. The ReLu nonlinearity is associated with each learning layer. The neurons in the fully connected layers are connected to all neurons in the previous layer, with 4096 neurons each [12].

In this work, the network is constructed with following details:

- (i) Layer 0: input image of size  $240 \times 240$
- (ii) Layer 1: convolution with 96 filters, size  $11 \times 11$ , stride 4
- (iii) Layer 2: max-pooling with a size  $3 \times 3$  filter, stride 2
- (iv) Layer 3: convolution with 256 filters, size  $5 \times 5$ , stride 1
- (v) Layer 4: max-pooling with a size  $3 \times 3$  filter, stride 2
- (vi) Layer 5: convolution with 384 filters, size  $3 \times 3$ , stride 1
- (vii) Layer 6: convolution with 384 filters, size  $3 \times 3$ , stride 1
- (viii) Layer 7: convolution with 256 filters, size  $3 \times 3$ , stride 1
- (ix) Layer 8: max-pooling with a size  $3 \times 3$  filter, stride 2
- (x) Layer 9: fully connected with 512 neurons
- (xi) Layer 10: fully connected with 512 neurons
- (xii) Layer 11: fully connected with 22 neurons

The number of the neurons in the last layers is set to 22 to equalize the number of waste categories discussed in Section 2. An additional Layer 12, which contains 1 neuron with sigmoid activation function (recyclable as 1 and others as 0), is used during training. This layer is then removed when integrated into the multilayer hybrid system that ingests the 22 outputs from Layer 11 as image features. To do so, the CNN reduces the high dimension image to low-dimensional representation with robust features only. Also, it reserves the information of important identity features which may be wrapped by final classification [22]. This application is extensively adopted in human face verification problem and face recognition as a binary classification problem.

**3.3. MLP.** Multilayer perceptrons (MLP), one of the most established deep-learning structures for nonlinear classification and regression, are frequently used for modeling and forecasting [23–26].

Neurons, which are placed in layers, are the basic processing elements of MLP. The layers between the first layer (inputs) and the last layer (outputs) are called hidden layers. The MLP employed in this work has 1 hidden layer with 10 neurons as suggested in [25] that this network is able

TABLE 3: Experiment sensors.

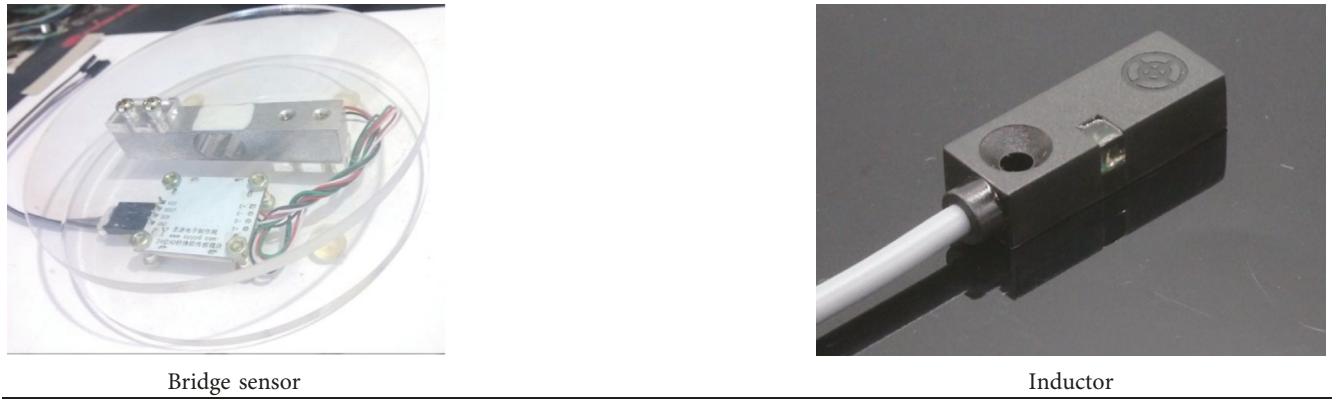


FIGURE 1: Example of the original image.

to fit any continuous function. Neurons on each layer sum the weighted inputs, add a bias to the sum, and then apply an activation function to process the sum and compute the outputs. The signal processing of neurons can be mathematically expressed as

$$Y_i = f \left( \sum_{j=1}^M (W_{ij} X_j + \beta_{ij}) \right), \quad (2)$$

where  $Y_i$  is the output of the  $i^{th}$  neuron on current layer,  $w_{ij}$  and  $\beta_{ij}$  are the weight and bias of the  $j^{th}$  input on the  $i^{th}$  neuron,  $M$  is the number of inputs,  $X_j$  is  $j^{th}$  output from the previous layer, and  $f$  is the activation function, which is a sigmoid function in this work.

$$f(y) = \frac{1}{1 + e^{-y}}. \quad (3)$$

**3.4. Multilayer Hybrid System.** The multilayer hybrid system (MHS) developed in this work simulates human sensory and intelligence process system. This MHS is a combination of several interdependent subsystems, including: (1) image system (2) sensor system and (3) the central back-end classification system.

Figure 2 illustrates the multilayer hybrid system with three interacting subsystems and their associated components. The arrows indicate the processing flow and interaction of subsystems.

When a waste item is received into the hybrid system, the camera and sensors are activated to observe the item. The imaging system consists of a camera to capture an image, which is analyzed by the CNN. The sensor system, simultaneously, functions to obtain numerical information from the objects. The ultimate results (binary output) are obtained using MLP system, whose inputs are the 22 outputs from the CNN and the numerical information from sensors. In this respect, MLP system can be trained independently from CNN model, and the weight and bias parameters of CNN model can be kept unaffected. On the other hand, as the outputs of CNN are the inputs for MLP, these two models actually function simultaneously to generate the binary classification results.

**3.5. Evaluation Metrics.** Each classification prediction by the proposed system is compared to the manually classified label, which is set as the “truth”. The confusion matrix shown in Table 5 quantifies the hits and misses of the automatic classification system. A CNN-only model with exactly the same structure discussed above is trained and evaluated as a reference model.

The system performance is evaluated with accuracy, precision, and recall. The accuracy of classification is defined as the percentage of images that are correctly classified:

$$\text{accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \times 100\%. \quad (4)$$

Precision, therefore, represents the correctness of classification prediction systems.

$$\text{Precision} = \frac{TP}{(TP + FP)} \times 100\%. \quad (5)$$

Recall represents the effectiveness of classification prediction systems.

$$\text{Recall} = \frac{TP}{(TP + FN)} \times 100\%. \quad (6)$$

Incorporating precision and recall can reduce the bias forecasting caused by the unbalanced dataset: the minority class is harder to learn and the model tends to over forecast the majority class with highly skew data [26].

TABLE 4: Example of images generated by the augmentation algorithm.

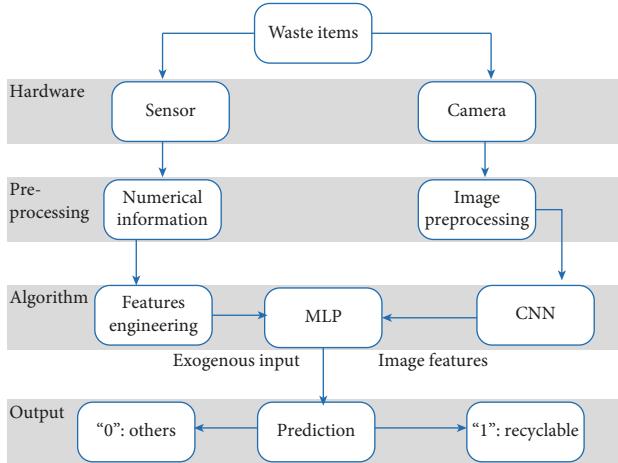
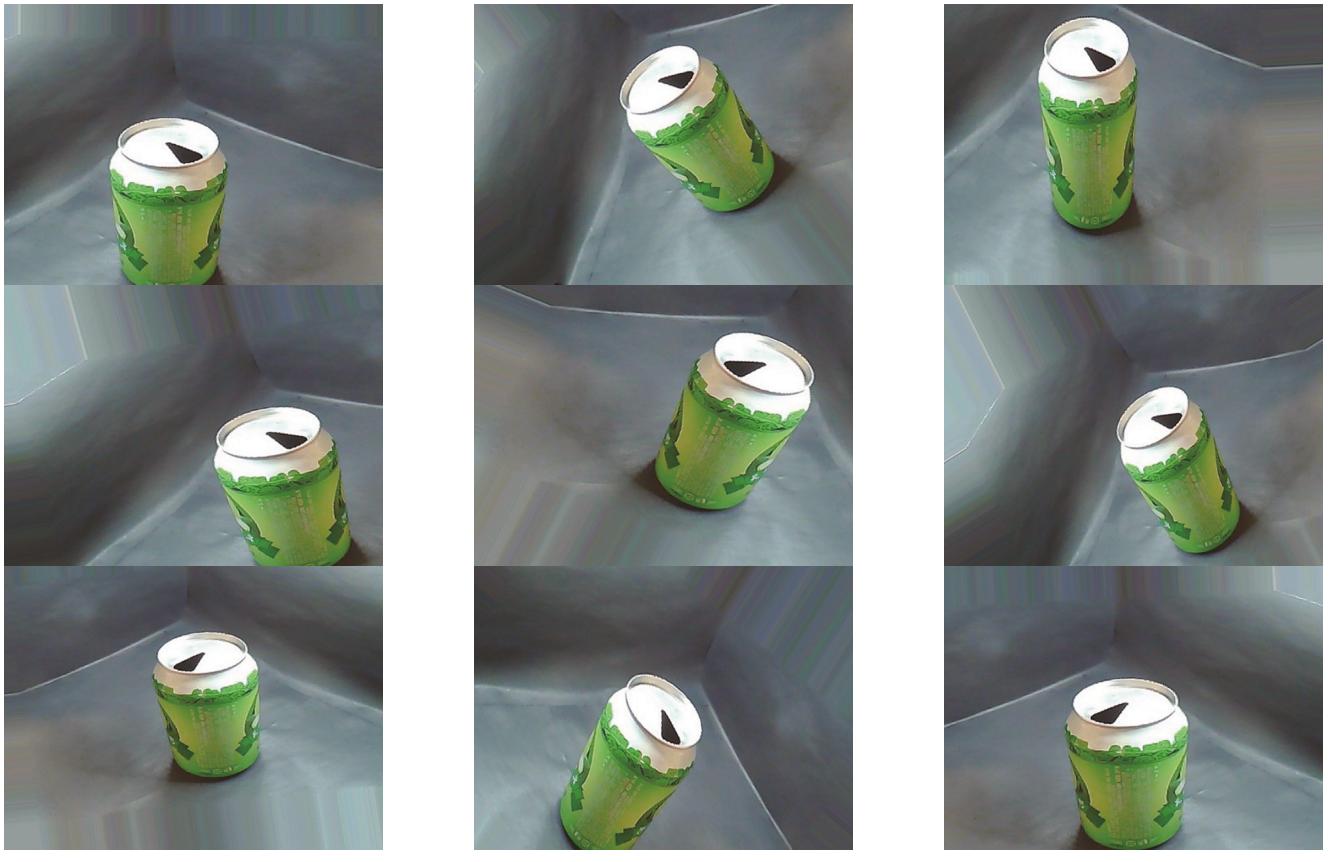


FIGURE 2: Multilayer hybrid system (MHS).

## 4. Result and Discussion

The MHS model, which is trained using 5000 data instances, is evaluated under two different scenarios: the item is placed with fixed and random orientations. The model performances are compared with a CNN model that takes the images as the only input. The classification results from each model are presented in Table 6.

TABLE 5: Confusion matrix of automatic waste classification for each category.

Automatic classification	Manual classification	
	Recyclable	Others
Recyclable	True positive (TP)	False positive (FP)
Others	False negative (FN)	True negative (TN)

The evaluation results presented in Table 6 indicate that MHS significantly outperforms CNN-only model in terms of all three matrices (accuracy, precision, and recall), particularly for the “others” category.

MHS achieves accuracy rates above 90% for both the first and second tests, which are 10% higher than the CNN-only model (Table 6). The MHS model also achieves higher precision rate of 98.5%, 97.1% to 88.6%, 85.9%, respectively, indicating MHS model’s effectiveness in predicting recyclable items. In addition, the MHS model shows great performance in recall (99% and 92%, respectively, for the first and second tests), showing that MHS is highly sensitive in identifying dedicated recyclable waste items.

The following tables display three sets of representative items returning from testing results. Table 7 represents items that are correctly classified by both MHS and CNN; Table 8 represents items that are correctly classified by the MHS but incorrectly classified by the CNN. Table 9 consists of items with low classification accuracy in both models.

TABLE 6: Confusion matrices for different classification models.

Evaluation metrics	MHS model		CNN model	
	1 <sup>st</sup> test	2 <sup>nd</sup> test	1 <sup>st</sup> test	2 <sup>nd</sup> test
Accuracy (%)	98.2	91.6	87.7	80.0
Precision (%)	98.5	97.1	88.6	85.9
Recall (%)	99.3	92.3	96.8	89.2

TABLE 7: Representative waste items that are correctly classified by both MHS and CNN.

MHS	CNN		
✓	✓		
			

It can be noticed that both MHS and CNN perform well when investigated items have strong image features (Table 7). However, CNN performs poorly when waste items lack distinctive image features, especially for “other” waste.

For instance, the images of beer cap and the transparent box (Table 8) are weak to be distinguished from the experiment background. It is difficult for the CNN to extract their imagery features in training, thus fail in testing. The cabbage item is irregular in appearance showing different figures for different orientations of placement. The CNN model itself is not sufficient enough to construct the feature patterns for accurate classification. The egg’s figure, on the other hand, is too simple to transfer sufficient information for training model resulting in a limited performance.

CNN relies on image information only, and if the study items are weak in imagery features, its classification performance will be adversely affected. MHS can address the problem by integrating both image and other feature information. In situations where the image information is insufficient, MHS is able to take advantages of other useful feature to make the most appropriate classification decision. Nevertheless, for items whose image features and other

numerical features are weak, their MHS classification error may increase. For instance, the MHS accuracy rate of the cup is about 60% (Table 9). After inspecting each individual misclassification case, we found that the major reason to cause these errors is that these objects have cylinder shapes, which are usually misclassified as bottles that are recyclable.

## 5. Conclusion

An automatic classification system based on multilayer hybrid deep learning (MHS) is proposed to classify disposal of waste in the urban public area. The system simulates human sensory and intelligence process system by deploying a high-resolution camera together with multiple functional sensors. The multi-layer hybrid method consists of three interdependent subsystems, including an image processing system, a numerical sensor system, and a multilayer perceptrons (MLP) system. The image processing system deploys AlexNet CNN to extract waste imagery information as inputs for the MLP. The sensor system aims at measuring other waste features as numerical input for MLP. The MHS is used to automatically classify the waste item as either recyclable or the others by consolidating information from both image and sensory channels.

TABLE 8: Representative waste item that are correctly classified by the MHS but incorrectly classified by the CNN.

MHS	CNN	
✓	X	

TABLE 9: Representative waste item with low accuracy in MHS and CNN.

MHS	CNN	
X	X	

A total of 50 waste items are used to evaluate the performance of MHS, which is also compared with a CNN-only model that only takes images as input. The result indicates that the MHS achieves a significantly higher classification performance: the overall performance accuracies are 98.2% and 91.6%, (the accuracy of the reference model is 87.7% and 80.0%) under two different testing scenarios.

This study demonstrates the potential of the proposed MHS in improving waste classification's efficiency and effectiveness. In considering the continually increased volume of waste globally and the urgent requirements for environmentally friendly waste processing, the proposed MHS is both economically and environmentally beneficial.

## Data Availability

The image and numerical data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The authors gratefully acknowledge the support provided by Shiwei Qin, Xin Zhao, and Xianyue Feng who assisted in the

collection of waste items, hardware preparation, data analysis, and algorithm optimization. The authors also gratefully acknowledge the technique advice from Yuanhui Tang and the laboratory and platform support from Dongguan University of Technology. This work was partly financially supported by the National Natural Science Foundation of China (Grant no. 11702073), Shenzhen Key Lab Fund of Mechanisms and Control in Aerospace (Grant no. ZDSYS201703031002066), and the Basic Research Plan of Shenzhen (Grant nos. JCYJ20170413112645981 and JCYJ20170811160440239).

## References

- [1] D. Hoornweg and P. Bhada-Tata, *What a Waste: A Global Review of Solid Waste Management*, World Bank, Washington, DC, USA, 2012.
- [2] R. E. Sanderson, *Environmental Protection Agency Office of Federal Activities' Guidance on Incorporating EPA's Pollution Prevention Strategy into the Environmental Review Process*, EPA, Washington, DC, USA, 1993.
- [3] P. T. Williams, *Waste Treatment and Disposal*, Wiley, West Sussex, UK, 2005.
- [4] T. H. Christensen, E. Gentil, A. Boldrin, A. W. Larsen, B. P. Weidema, and M. Hauschild, "C balance, carbon dioxide emissions and global warming potentials in LCA-modelling of waste management systems," *Waste Management & Research*, vol. 27, no. 8, pp. 707–715, 2009.
- [5] US EPA, *Facts and Figures about Materials, Waste and Recycling*, EPA, Washington, DC, USA, 2018, <https://www.epa.gov/facts-and-figures-about-materials-waste-and-recycling/advancing-sustainable-materials-management-0>.
- [6] O. Kofoworola, "Recovery and recycling practices in municipal solid waste management in Lagos, Nigeria," *Waste Management*, vol. 27, no. 9, pp. 1139–1143, 2007.
- [7] J. Tai, W. Zhang, Y. Che, and D. Feng, "Municipal solid waste source-separated collection in China: a comparative analysis," *Waste Management*, vol. 31, no. 8, pp. 1673–1682, 2011.
- [8] A. M. Troschinetz, *Twelve Factors Influencing Sustainable Recycling of Municipal Solid Waste in Developing Countries*, Michigan Technological University, Houghton, MI, USA, 2005.
- [9] N. Kollikkathara, H. Feng, and E. Stern, "A purview of waste management evolution: special emphasis on USA," *Waste Management*, vol. 29, no. 2, pp. 974–985, 2009.
- [10] S. Al-Salem, P. Lettieri, and J. Baeyens, "Recycling and recovery routes of plastic solid waste (PSW): a review," *Waste Management*, vol. 29, no. 10, pp. 2625–2643, 2009.
- [11] European Environment Agency, *Waste Recycling*, 2018, <https://www.eea.europa.eu/data-and-maps/indicators/waste-recycling-1/assessment>.
- [12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proceedings of Neural Information Processing System Conference*, Lake Tahoe, CA, USA, December 2012.
- [13] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. Lecun, "OverFeat: integrated recognition, localization and detection using convolutional networks," in *Proceedings of International Conference on Learning Representations*, Banff, Canada, April 2014.
- [14] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional neural networks," in *Proceedings of 13th European Conference on Computer Vision (ECCV)*, Zurich, Switzerland, September 2014.
- [15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of Conference on Computer Vision and Pattern Recognition*, Las Vegas Valley, NV, USA, June 2016.
- [16] O. Awe, R. Mengistu, and V. Sreedhar, "Smart trash net: waste localization and classification," arXiv Preprint, 2017.
- [17] G. Thung and M. Yang, "Classification of trash for recyclability status," arXiv Preprint, 2016.
- [18] M. S. Rad, A. V. Kaenel, A. Droux et al., "A computer vision system to localize and classify wastes on the streets," in *Computer Vision System*, pp. 195–204, Springer, Cham, Switzerland, 2017.
- [19] J. Donovan, *Auto-Trash Sorts Garbage Automatically at the TechCrunch Disrupt Hackathon*, 2018, <https://techcrunch.com/2016/09/13/auto-trashsorts-garbage-automatically-at-the-techcrunch-disrupt-hackathon/>.
- [20] G. Mittal, K. B. Yagnik, M. Garg, and N. C. Krishnan, "SpotGarbage: smartphone app to detect garbage using deep learning," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, New York, NY, USA, 2016.
- [21] S.-T. Bow, *Pattern Recognition and Image Preprocessing*, Taylor & Francis, Abingdon, UK, 2002.
- [22] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: closing the gap to human-level performance in face verification," in *Proceedings of Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference*, Columbus, OH, USA, 2014.
- [23] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2008.
- [24] C. M. Bishop, "Neural networks and their applications," *Review of Scientific Instruments*, vol. 65, no. 6, pp. 1803–1832, 1994.
- [25] M. Nielsen, *Neural Networks and Deep Learning*, Determination Press, 2015.
- [26] F. Provost, "Machine learning from imbalanced data sets 101," in *Proceedings of AAAI'2000 Workshop on Imbalanced Data Sets*, Menlo Park, CA, USA, 2000.

