Journal of Cleaner Production 271 (2020) 122393

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro

One-dimensional convolutional neural networks for acoustic waste sorting

Gang Lu [1], Yuanbin Wang [1], Huayong Yang , Jun Zou [*]

State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University, Hangzhou 310027, China

ARTICLE INFO

Article history:
Received 10 November 2019
Received in revised form
28 March 2020
Accepted 18 May 2020
Available online 20 June 2020

Handling editor: Jun Bi

Keywords:
Waste sorting
Sound classification
1D convolutional neural networks
Design of orthogonal experiment

ABSTRACT

This paper presents a successful application of one-dimensional convolutional neural networks (1D CNNs) for waste sorting at source based on acoustic data. Most of the existing methods use images for waste classification, which causes a high computational complexity and requires a huge amount of training data. Acoustic data have also been employed due to its high correlation to the waste material type. The traditional approaches usually use fixed and handcrafted features extracted from acoustic data. Finding efficient features is usually time consuming and could be difficult in complex situations. In this paper, the 1D CNN method is proposed to automatically extract features from acoustic data and achieve high classification accuracy. Orthogonal experiment method is applied to optimize three key hyper-parameters of the 1D CNN structure. The experiment shows that the proposed method can achieve 92.40% classification accuracy within a short time. A traditional method using handcrafted features with a shallow classifier is taken as a benchmark and the attained classification accuracy is only 81.92%. In addition, a classification accuracy rate of 68.06% is achieved when using a shallow classifier with raw acoustic data as input. Therefore, the proposed method is promising to be practically applied in the trash bins for automatic waste source separation.

1. Introduction

Product disposal is an important aspect in cleaner production. Accurately classifying these wastes is the key to achieve efficient recycling and other treatment. Recently, the amount of municipal solid waste (MSW) has been dramatically increased due to the rapid growth of the population, urbanization and economy (Nižeti'c et al., 2019; Knickmeyer, 2020). In this situation, efficient waste classification becomes increasingly crucial to minimize the environmental damage of product disposals. Source separation is a crucial step to improve the quality of the classification and influences how the wastes could be processed afterwards. Currently, the source separation is usually conducted by the people who dispose it. This requires a huge effort to educate and regularize the citizens, especially for developing countries whose citizens have not formed a good habit (Chen et al., 2019). Meanwhile, it is difficult to achieve a higher granularity level of the classification. Therefore, automatic source separation techniques become attractive to release people from these burdens in their daily lives and achieve a more accurate classification.

Existing research mainly focuses on image-based approaches to recognize the wastes (Ruiz et al., 2019; Wang et al., 2019). However, there are several drawbacks of using image signals. First, the images usually carry a large amount of data which cost more computation resources to process. Second, the image features including shapes, colors and textures could be dramatically different for the same type of waste. Even for the same waste, it may also have different deformations and damages. This brings a great challenge to prepare sufficient datasets for training classification model.

The sound signal of an object, unlike its image, is highly correlated with its material properties. This is one of the most important aspects for waste classification. Sound signals generated by external actions (e.g., striking, free falling, etc.) can expose the intrinsic properties of waste such as elasticity and internal friction (Klatzky et al., 2000). The elasticity influences the frequency of the generated sound. The internal friction determines how the generated sound decays over time and provides shape-variant acoustic features for waste sorting (Krotkov et al., 1997). The deep learning has become a hot topic in recent years and achieved a huge success mainly for image processing. Some attempts have been conducted

* Corresponding author.
E-mail address: junzou@zju.edu.cn (J. Zou).
[1] Contribution: G. Lu and Y. Wang contributed equally to this work.

https://doi.org/10.1016/j.jclepro.2020.122393
0959-6526/

for sound classification problems (citations). However, how to use deep learning for the waste classification problem has not been studied. In this situation, the features of wastes are different from other sound classification problems such as speech recognition and knocking sound classification. The signal length is short and could vary with different dropping postures, deformations, shapes, mixtures, etc. This paper explores the potential of deep learning for the waste sorting application and proposes a 1D CNN model that achieves a good balance between classification accuracy and computation efficiency. The contributions of this paper are summarized as:

(1) The 1D CNN method using acoustic data is proposed for waste sorting at source. It is able to deal with complex scenarios occurring in reality with high accuracy.
(2) The major hyper-parameters in 1D CNNs are systematically studied and the best combination is found for the highest classification accuracy.

The remainder of this paper is structured as follows. The related work on waste sorting and related methods are reviewed in Section 2. Section 3, introduces the 1D CNN-based impact sound classification method and the whole experimental design. The process of data acquisition is presented in Section 4. In Section 5, the effects of different hyper-parameters on the classification performance are evaluated. In the following section, the comparison between the proposed method and traditional sound classification methods is carried out. Conclusions and the further work are given in Section 7.

2. Related works

2.1. Waste sorting methods

Over the last decade, many research efforts have been put on the automated waste sorting techniques for MSW (Gundupalli et al., 2017a). Different sorting techniques have been developed for MSW sorting, including laser induced breakdown spectroscopy (Jull et al., 2018), spectral imaging (Li et al., 2019) and thermal imaging (Gundupalli et al., 2017b, 2018). However, these systems were complicated and involved expensive apparatus. They were only suitable for source-separated waste streams in waste treatment plants.

In recent years, vision has gained great attentions in waste sorting because its simplicity of deployment. Rahman et al. (2011) introduced a new method for an automated paper sorting system that utilized an image processing technique with the K-nearest neighbor (K-NN) classifier. The proposed system performance for correct paper grade identification is more than 90%. Özkan et al. (2015) used five different feature extraction methods and support vector machine (SVM) for the plastic bottle classification. It could automatically classify the plastic bottle types with approximate 90% recognition accuracy. Srinilta and Kanharattanachai (2019) explored performance of CNN-based waste-type classifiers (VGG-16, ResNet-50, MobileNet-V2 and DenseNet-121) in classifying waste types of 9200 MSW images. The highest waste-type classification accuracy was 94.86% from the ResNet-50 classifier. Nevertheless, vision depended heavily on the surrounding environments and would fail due to the variance of fullness, deformation conditions and illumination changes.

Sound signal is another promising feature of wastes for classification as it contains rich material information. Korucu et al. (2016) first used sound recognition to identify the packing waste. SVM and hidden Markov model (HMM) based classification approaches with Mel-Frequency Cepstral Coefficients (MFCCs) provided a high classification performance. However, only four categories of the packing waste and simple cases were considered with manually defined features. In reality, the situation is much more complex and the manually defined features and shallow classifiers used in this paper may not be able perform well. Following this work, there were no more researches concerning acoustic waste sorting for MSW.

2.2. Acoustic classification

Although the study for audition-based waste sorting is very limited, the acoustic classification problems in general has gained great interests in recent years including acoustic frog call classification (Xie et al., 2019), agricultural products classification (Sun et al., 2018) and End-of Life vehicles' plastic classification (Huang et al., 2015, 2017), etc. In the aforementioned works, handcrafted features and shallow classifiers were utilized. In a recent study, Luo et al. (2017) proposed a deep learning-based method for the acoustic object recognition using the raw acoustic data. As a result, an overall classification accuracy of 91.50% was achieved. However, the fully connected networks used in this work had a lot of parameters. Its classification accuracy may decrease with the deepening of the network. Some researchers have tried to use deep 2D CNNs for acoustic classification (Ren et al., 2018). This method firstly converted 1D signal to 2D images as the CNNs for image classification have been widely studied. However, converting 1D signals to 2D actually brings more complexity to the classification problem as it could dramatically increase the size of each simple data. Therefore, the classification network will have more parameters and the processing speed will be decreased.

2.3. 1D CNN applications

1D CNNs have recently been proposed to deal with 1D signals and achieved superior performance with high efficiency. In a relatively short time, 1D CNNs have become popular with a state-of-the-art performance in various signal processing applications such as early arrhythmia detection in electrocardiogram (ECG) beats, structural damage detection and high power engine fault monitoring (Kiranyaz et al., 2019). Comparing 2D CNNs, this type of method directly processes 1D signals while still has the capability of automatically learning complex features from training samples. Comparing traditional sound classification methods, it can directly process the raw acoustic data and achieve an end-to-end approach. In this way, the classification model could be more flexible and less rely on the expert knowledge.

To sum up, acoustic waste sorting is a promising approach to realize automatic waste sorting at source. Among various methods for acoustic classification, 1D CNNs have the potential to be a more flexible and efficient way. However, to the best of the authors' knowledge, the performance of 1D CNNs for acoustic waste sorting has not been investigated so far. The hyper-parameters of CNNs play a critical role for the classification performance. As both classification accuracy and computation efficiency are important, it is necessary to systematically investigate effects of common hyper-parameters on the network performance and find the optimal network structure for waste classification.

3. Methodology

3.1. 1D CNN

The 1D CNN mainly consists of three parts including the convolution layer, the pooling layer and the fully connected layer. The general 1D CNN architecture is illustrated in Fig. 1. A 1D signal is fed into the input layer of the 1D CNN. Convolution operations are

performed between the input signal and corresponding convolution kernels to generate the input feature maps. Then the input feature maps are passed through the activation function to generate the output feature maps of the convolution layer. The output of the convolution layer can be expressed as (Bouvrie, 2006):

$$y_j^l = f(b_j^l + \sum_{i \in M_j} conv1D(\omega_{ij}^{l-1}, x_i^{l-1})) \quad (1)$$

where y_j^l is the output of the j^{th} neuron at layer l, $f(\cdot)$ is a nonlinear function, b_j^l is a scalar bias of the j^{th} neuron at layer l, M_j represents a selection of input maps, x_i^{l-1} is the output of the i^{th} neuron at layer $l-1$, ω_{ij}^{l-1} is the kernel weight from the i^{th} neuron at layer $l-1$ to the j^{th} neuron at layer l.

After the convolution layer, a pooling layer is usually adopted not only to reduce the computational cost by reducing the dimension of features extracted from the upper convolution layer, but also to provide basic translation invariance to the features. Its formula is as below (Bouvrie, 2006):

$$s_j^{l+1} = f(\beta_j^l down(y_j^l) + b_j^{l+1}) \quad (2)$$

where down $(\cdot)$ represents a subsampling function, β_j^l means the weighting coefficient and b_j^{l+1} is bias coefficient.

The output of each neuron of the pooling layer become the input of each neuron of a fully connected layer. Usually, the fully connected layer acts as a classifier in the whole 1D CNN.

3.2. Experimental design

In this study, the 1D CNN method is applied to acoustic waste sorting at source. The main steps of the methodological approach used is shown in Fig. 2. First, the sound acquisition device is designed in consideration of the free falling impact type. Then, sixteen types of the waste with different specifications and conditions are collected for this study. All the sounds generated from the wastes in the sound acquisition device are recorded. Afterwards, the design of experiment method is applied to reveal the effects of different hyper-parameters in the 1D CNNs model on the classification performance and the optimal parameter combination is found. After the optimal CNN structure is generated, the comparison between the proposed method and traditional sound classification methods is carried out. Finally, Conclusions and the further work are given.

The experiments in this study follow the predefined standards and protocols. In the sound recording process, the height for waste dropping in this study is randomly chosen between 801 mm and 954 mm above the impact plate center for each throw to mimic an adult throwing situation. The sound recording starts when the waste is released and ends after the waste impacts the plastic board. After the end of each recording, the waste is removed from the exit at the bottom of the sound acquisition device. The waste used in each sound recording process is randomly adjusted to different postures (e.g., horizontal, vertical, sloping, etc.) and conditions (e.g., random deformation, random filling, etc.) for each throw. In the process of the 1D CNN structure optimization, the 5 cross-validation approach is used to train the model to achieve a reliable and steady model. The orthogonal experiment approach is applied to study the effects of CNN hyper-parameters and find the optimal parameter combination in the model.

4. Data acquisition and pre-processing

4.1. Sound collection equipment and objects

An open-top empty chamber with 400 mm × 400 mm × 800 mm dimensions is designed for sound recording experiments. The chamber is composed of aluminum alloy profiles and acrylic boards and its inner surface is covered by sponge material in order to isolate ambient noises. A microphone is placed in the corner of the chamber. A plastic impact plate made of acrylonitrile-butadiene-styrene (ABS) is set near the bottom of the chamber at an angle 45° downward. The thickness of the plastic plate is 10 mm. All the sound recording processes are carried out in the empty chamber. The sound recordings are obtained by the free falling. The experimental setup used in this study can be found in Fig. 3.

Sixteen types of the waste collected from teaching buildings, administration buildings and convenience stores in Zhejiang University are used in this study, as depicted in Fig. 4. In reality, the

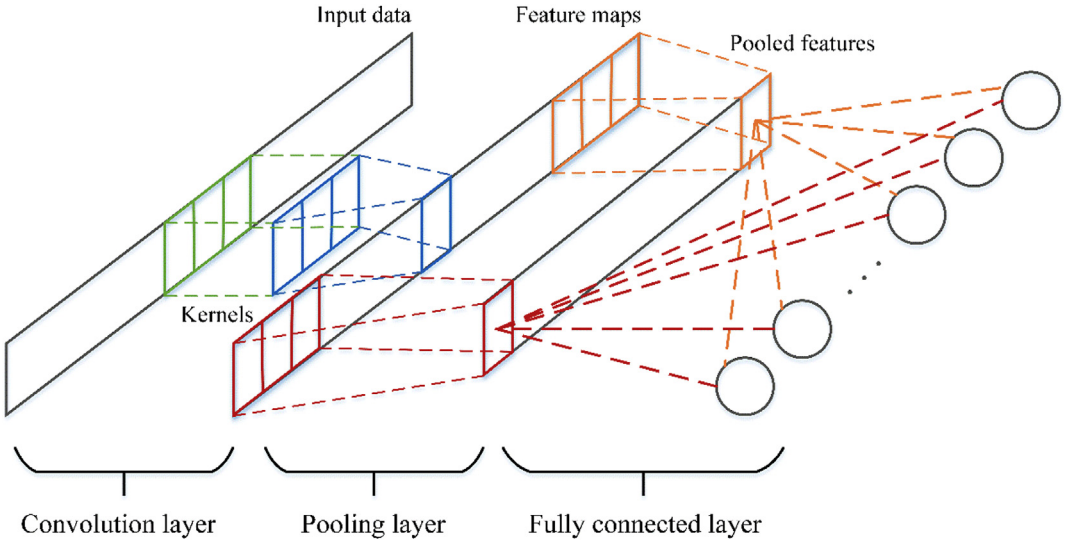

Fig. 1. Structure of proposed 1D CNN.

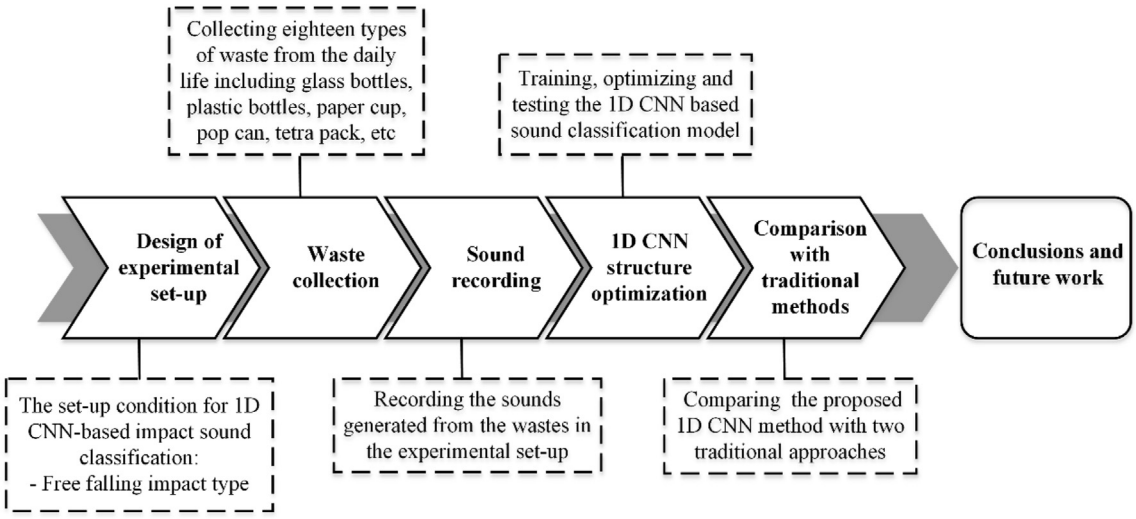

Fig. 2. Methodological approach used in this study. The basic assumptions in this research are as follows.

- The sound recording process is carried out by imitating people standing next to a dustbin and putting a waste in it. Other extreme throwing conditions are not considered.
- The sound is generated when a waste strikes a plastic board in the sound acquisition device with the free falling. Other types of sound generation methods are not considered.
- All the sounds are captured in a sound insulation environment. The influence of noises from the environment is not considered in this study.

status of each waste object could be in huge variety. Take the plastic bottle as an example, different brand may have different shapes and colors and the bottle could be deformed, broken, having residue in it, etc. In this experiment, most conditions are considered and the status of each sound sample is different in one or multiple aspects including brands, deformations, filling and mixture (details are listed in Table 1). All these variations are randomly chosen and different from others.

4.2. Sound recording

The height for waste dropping needs to meet the requirement that tall people don't have to bend and short people don't need to lift their elbows. According to the ergonomics (citation) (Sanders and McCormic, 1993) as shown in Table 2, a standing adult usually put a waste to the dustbin from the height between 801 mm and 954 mm. Therefore, each waste object in the experiment falls from a random height in this range to mimic the real situation. The generated impact sound is recorded on a laptop via an Audio-Technica microphone using Adobe Audition software. The sampling frequency is 44100 Hz and quantization level is 24bits. The schematic of the sound recording process is shown in Fig. 5. The posture of each waste object is also random chosen. Hence, each sound sample obtained is in unique condition and represents a sole situation. For each type of wastes, 360 sound samples are collected in different status and conditions. The total number of sound recordings in the study is 5760 (360 × 16 waste types).

4.3. Data pre-processing

To trim the redundant information in the data, only a 100 ms audio (4410 points) starting from the 20% of the peak value of each sound recording, which can cover the whole impact process for each sound recording trial, was taken as the input for 1D CNNs, as shown in Fig. 6. Before the sound classification, all the samples are divided into six groups. The numbers of samples of each type are identical in each group. A group is chosen randomly as the test set and the remaining five groups as the training set. The 1D CNNs for the acoustic waste sorting constructed here consists of two phases.

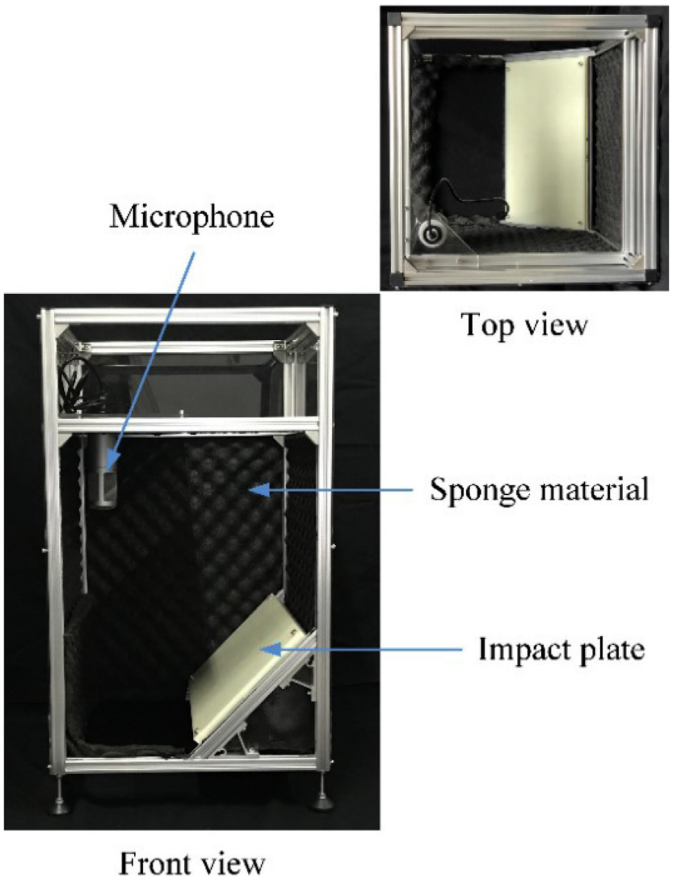

Fig. 3. Sound recording experimental setup.

Firstly, each raw acoustic data in the training set is fed into the input layer nodes of the 1D CNNs and the output nodes are waste classes. The entire network is then fine-tuned in a supervised manner to minimize the error in predicting the waste labels using back

Fig. 4. The waste used for the experiments and they are labeled from 0 to 15 marked at the upper right of the picture of each waste. 0. Adhesive tape 1. Binder clip 2. Circuit board 3. Empty glass bottle 4. Empty packing box 5. Empty plastic bottle 6. Empty plastic storage case 7. Empty tetra pack 8. Nail clipper 9. Empty paper cup 10. Plastic bottle with the filling 11. Plastic storage case mixed with other waste 12. Empty pop can 13. Rollerball pen 14. Tetra pack with the filling 15. Tweezer.

Table 1
Specifications of the waste types used in this study.

Label	Waste type	Specifications	Actual conditions
0	Adhesive tape	4	Different usage amounts
1	Binder clip	5	Folding or flattening tail handles
2	Circuit board	5	–
3	Empty glass bottle	8	–
4	Empty packing box	10	–
5	Empty plastic bottle	8	No or random deformation
6	Empty plastic storage case	6	–
7	Empty tetra pack	6	No or random deformation
8	Nail clipper	5	Folding or flattening pressing handles
9	Paper cup	6	No or random deformation
10	Plastic bottle with the filling	8	Random filling
11	Plastic storage case mixed with other waste	6	–
12	Pop can	5	No or random deformation
13	Rollerball pen	12	–
14	Tetra pack with the filling	6	Random filling
15	Tweezer	6	–

Table 2
Human body dimensions with the standing posture.

Dimension (mm)	Man			Woman		
	5%	50%	95%	5%	50%	95%
Elbow height	954	1024	1096	899	960	1023
Functional hand height	680	741	801	650	704	757

propagation. Subsequently, the 1D CNNs model is evaluated in the test set and the classification accuracy achieved is used as the final evaluation index of the model.

5. 1D CNN structure optimization

In this study, how different hyper-parameters in the 1D CNNs model will affect the classification performance including the network depth, the convolution kernel size and the learning rate is investigated. The 5-fold cross validation setup is employed to train

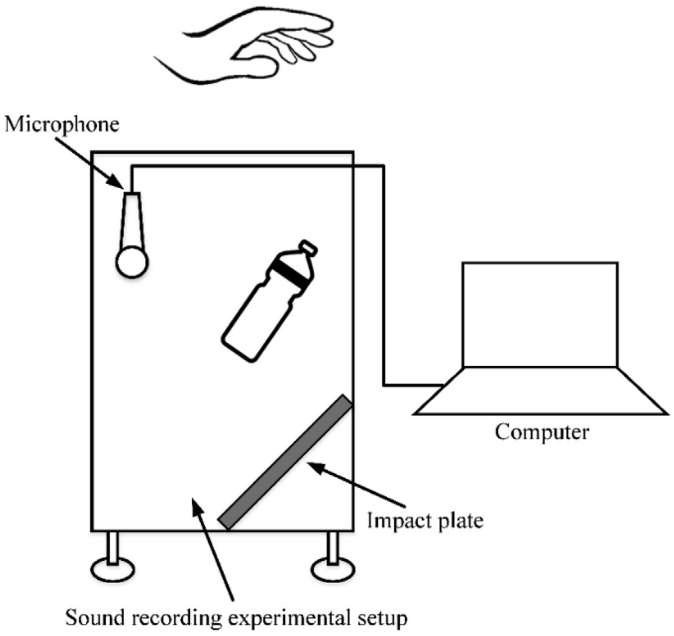

Fig. 5. Experimental sound acquisition system.

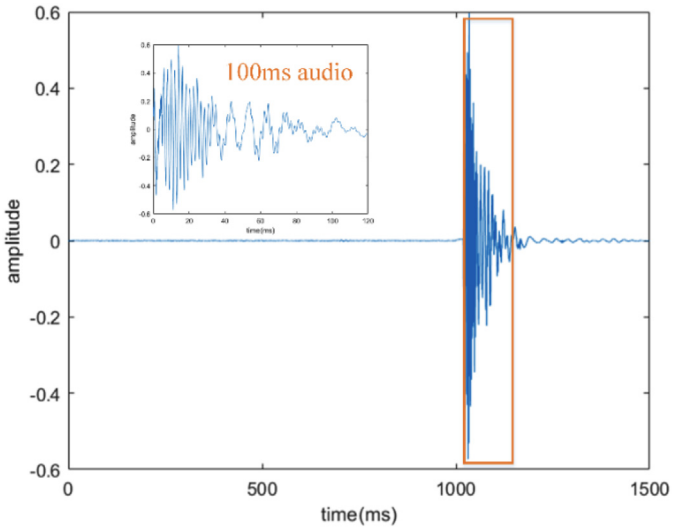

Fig. 6. The extracted audio.

the model in the training set and the average accuracy rate (AAR) is used as the evaluation index of each set of hyper-parameters in the 1D CNNs model. There are 4410 nodes in the input layer and 16 nodes in the output layer representing the 16 classes shown in Fig. 4. Besides, a convolution layer and a pooling layer are considered as a group. The number of groups in the 1D CNNs represents the network depth. In the first group, the number of convolution kernel is 8. Subsequently, the number of convolution kernel in the latter group is twice as much as that of the former group.

5.1. The design of orthogonal experiment

5.1.1. Factors and levels

To investigate the effect of the 1D CNNs model for the classification performance, three factors are considered, including (A) the network depth, (B) the convolution kernel size and (C) the learning rate. Each factor includes five levels and the range of variation is chosen according to previous experiences. The levels of the experimental factors are shown in Table 3.

Table 3
The experiment factors and levels.

Levels	Factors		
	Network depth	Convolution kernel size	Learning rate
1	3	3	5×10^{-5}
2	4	5	1×10^{-4}
3	5	7	5×10^{-4}
4	6	9	1×10^{-3}
5	7	11	5×10^{-3}

5.1.2. Orthogonal table

The format of the orthogonal table is $L_n(a^b)$, where L represents the name of the orthogonal table, n indicates the number of rows in the orthogonal table (that is the number of trials that need to be carried out), b denotes the number of columns in the orthogonal table (that is the maximum number of factors that can be arranged), and a represents the number of levels of each factor (Cai et al., 2019). In this paper, there are three factors and five levels. According to the principle for selecting the orthogonal table, the orthogonal table noted as $L_{25}(5^6)$ is adopted without considering the interaction between the factors.

6. Result and discussion

As shown in Table 4, the number from 1 to 25 in the first column represents 25 experiments. Second to fourth columns denote three different factors. The row in the table is corresponded to an instance. For instance, the fifth experiment indicates that the network depth is 3, the convolution kernel size is 11 and the learning rate is 5e-3. The fifth column is the value of the evaluation index, which shows the classification performance of the model using given hyper-parameters. The rightmost column is the standard deviation of the 25 experiments respectively. All values of the standard deviation are distributed in the range of 0–2%, which indicates that the results are robust.

Through the range analysis of data in Table 4, Table 5 is obtained. In Table 5, $\overline{K}_{ij}$ is the average evaluation index of each experimental factor and can be expressed as (Deng et al., 2019):

$$\overline{K}_{ij} = K_{ij}/K_i \tag{3}$$

Where i (i=A, B, C) and j (j=1, 2, 3, 4, 5) are the factor and the level number respectively. K_{ij} is the sum of the evaluation indexes of all levels in each factor i. K_i is the total levels of the corresponding factor. The range of factor i can be calculated as (Deng et al., 2019):

$$R_i = \max\{\overline{K}_{ij}\} - min\{\overline{K}_{ij}\} \tag{4}$$

According to the R_i value of each factor in Table 5, the order of the factors' impact is A > C > B. R_A is the maximum among all these ranges, which indicates that the network depth is the dominant factor influencing the classification performance. The followers are the learning rate and the convolution kernel size. The variation trend for each evaluation index caused by the factors is shown in Fig. 7. As shown in Fig. 7(a), the AAR increases when the network depth increases from 3 to 6. However, the AAR decreases when the network goes deeper. This shows that too many parameters in the network may cause overfitting and impact the generality. The AAR increases dramatically and then decreases with the increase of convolution kernel size. The highest AAR emerges when the

Table 4
The orthogonal designed table and its results.

Experiment number	Factors			Evaluation index	Standard deviation
	A	B	C	AAR (%)	
1	3	3	5×10^{-5}	85.83	0.69
2	3	5	1×10^{-4}	87.71	0.45
3	3	7	5×10^{-4}	88.50	0.59
4	3	9	1×10^{-3}	89.33	0.86
5	3	11	5×10^{-3}	80.90	1.51
6	4	3	1×10^{-4}	88.45	0.61
7	4	5	5×10^{-4}	89.69	0.78
8	4	7	1×10^{-3}	90.86	0.67
9	4	9	5×10^{-3}	85.81	0.46
10	4	11	5×10^{-5}	90.46	0.54
11	5	3	5×10^{-4}	89.22	1.05
12	5	5	1×10^{-3}	92.64	0.50
13	5	7	5×10^{-3}	88.50	1.74
14	5	9	5×10^{-5}	91.47	0.40
15	5	11	1×10^{-4}	92.14	0.45
16	6	3	1×10^{-3}	91.58	0.61
17	6	5	5×10^{-3}	89.11	0.72
18	6	7	5×10^{-5}	92.31	0.66
19	6	9	1×10^{-4}	92.39	1.06
20	6	11	5×10^{-4}	93.58	0.44
21	7	3	5×10^{-3}	88.72	1.88
22	7	5	5×10^{-5}	92.11	1.02
23	7	7	1×10^{-4}	91.72	1.12
24	7	9	5×10^{-4}	93.17	0.72
25	7	11	1×10^{-3}	92.49	0.56

Table 5
Range analysis of the evaluation index.

Range analysis	Factors		
	A	B	C
$\boldsymbol{K_{i1}}$	432.27	443.82	452.19
$\boldsymbol{K_{i2}}$	445.28	451.26	452.41
$\boldsymbol{K_{i3}}$	453.97	451.88	454.16
$\boldsymbol{K_{i4}}$	458.97	452.17	456.91
$\boldsymbol{K_{i5}}$	458.22	449.57	433.04
$\boldsymbol{\overline{K}_{i1}}$	86.45	88.76	90.44
$\boldsymbol{\overline{K}_{i2}}$	89.06	90.25	90.48
$\boldsymbol{\overline{K}_{i3}}$	90.79	90.38	90.83
$\boldsymbol{\overline{K}_{i4}}$	91.79	90.43	91.38
$\boldsymbol{\overline{K}_{i5}}$	91.64	89.91	86.61
$\boldsymbol{R_i}$	5.339	1.670	4.773

convolution kernel size is 9, shown in Fig. 7(b). In Fig. 7(c), the AAR increases when the learning rate increases from 5×10^{-5} to 1×10^{-3}. Then the AAR decreases sharply when the learning rate exceeds 1×10^{-3}.

By comparing the average value of each evaluation index, the $\overline{K}_{A1}<\overline{K}_{A2}<\overline{K}_{A3}<\overline{K}_{A4}>\overline{K}_{A5}$ is got. That is to say $A_4$is the optimal value in the A levels. Similarly, B_4 is the optimal value in the B levels and C_4 is the optimal value in the C levels. Therefore, the best combination is A_4, B_4 and C_4. Based on the above works, the optimized parameters are obtained as listed in Table 6. A new 1D CNNs model is trained by using these hyper-parameters. The classification accuracy of the model is 94.17%, which only demonstrates that this set of hyper-parameters is optimal. The final classification accuracy is 92.40% by evaluating the model in the test set.

7. Comparison with traditional methods

In this section, two traditional approaches are tested and compared with the proposed 1D CNNs method. All algorithms written in python 3.6 are implemented in TensorFlow 1.9 and executed on a laptop with a GTX1050Ti graphics card, an 8th Intel Core i5 processor and a 512 GB SSD.

As has been mentioned above, the traditional sound classification is achieved by using handcrafted features and shallow classifiers. In the first test, MFCCs and its first and second differential are used as features because it is used frequently in sound classification. It is based on the mechanism of human auditory characteristics and can well describe the nonlinear characteristics of human ear frequency. In the classification process, 12th-order MFCC together with its first and second temporal derivatives as features extracted from the acoustic signal for each 23 ms time window with 11 ms overlapping. As a result, each feature was a 36 dimensional vectors. Then the features are implemented with a SVM classifier.

In the second test, the raw acoustic data is utilized as the input of the SVM classifier. The classification results using different methods are listed in Table 7. It is remarkable that a high accuracy rate of 92.40% is achieved better than other two methods. The result shows that the proposed 1D CNNs can exploit the latent feature representations of the raw acoustic data and the waste can be classified accurately. The classification performance of the SVM-based classifier applied to raw data is much inferior to that of the SVM-based classifier applied to MFCCs. Therefore, the handcraft features seem to outperform raw data when using shallow classifiers.

The confusion matrixes of waste sorting with three methods is displayed in Fig. 8. They all perform well on some waste types including adhesive tapes (waste 0), empty glass bottles (waste 3) and pop cans (waste 12), due to the obvious distinction in the material sound.

Compared to the 1D CNNs method using raw data, the SVM-based method with MFCC features has two disadvantages in two aspects, including the ease of operation and the classification accuracy. First, the pre-defined features need to be extracted manually, which is time-consuming. Second, six types of waste are assigned to wrong labels such as circuit boards (waste 2), empty paper cups (waste 9) and plastic bottles with the filling (waste 10).

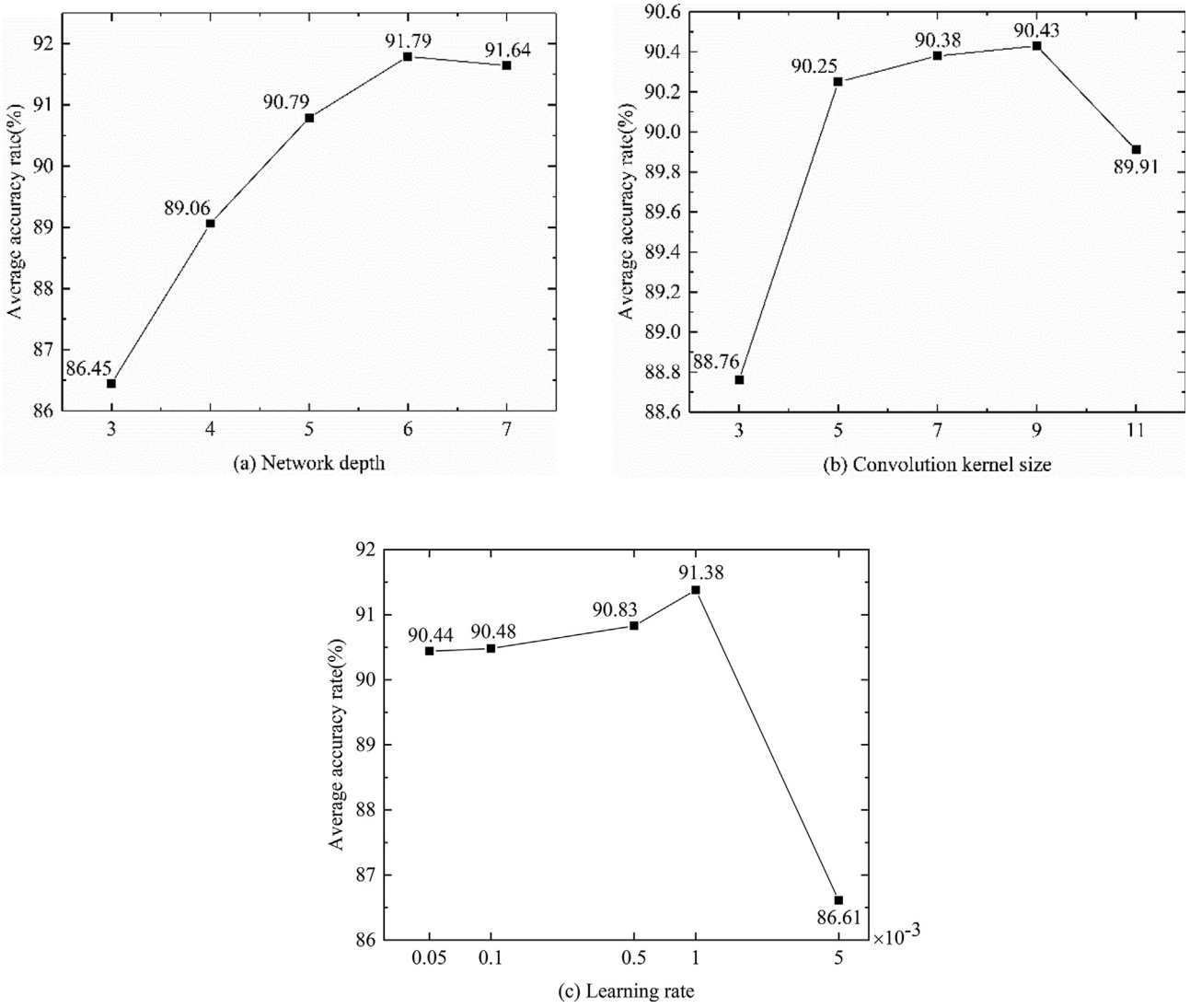

Fig. 7. The variation trend of each index (a) Network depth (b) Convolution kernel size (c) Learning rate.

Table 6
Optimized parameters.

Parameter	Value
Network depth	6
Convolution kernel size	9
Learning rate	1×10^{-3}

Table 7
Classification accuracy rates and time taken for classifying the waste in test set with different methods.

Method	Accuracy rate	Time/s
1D CNNs	92.40%	1.49
SVM with MFCCs features	81.92%	0.39
SVM with raw data	68.06%	15.11

They are likely to be recognized as empty plastic storage cases (waste 6), empty packing boxes (waste 4) and tetra packs with the filling (waste 14). The reason could be that the pre-defined features lack the capability to accurately represent the characteristics of waste types.

Compared to the 1D CNNs method using raw data, the SVM method using raw data performs much worse. A lot of wastes are difficult to distinguish by using the SVM-based method with raw data, e.g., binder clips and empty paper cups (waste 1 and 9), empty plastic bottles and pop cans (waste 5 and 12), circuit boards and tweezers (waste 2 and 15). The reason could be that the shallow classifier lacks the capability to identify higher level features in the raw data. On the contrary, the SVM-based method with MFCC features is more effective than using raw data. The proposed method has the most accurate results due to its capability to automatically discover and extract different level of features from the raw data.

The results also show that the mixed wastes are the most difficult scenario to deal with. The classification accuracies are lower for waste 6 and 11 in all the three methods. This shows that the limitation of using sound signal is that it is difficult to accurately classify the objects with very similar sounds. However, even in these situations the proposed method still can achieve around 80% accuracy.

The computational efficiency is another important requirement in reality. Therefore, the time consumed (excluding time for training models) for classifying the waste in test set by using three

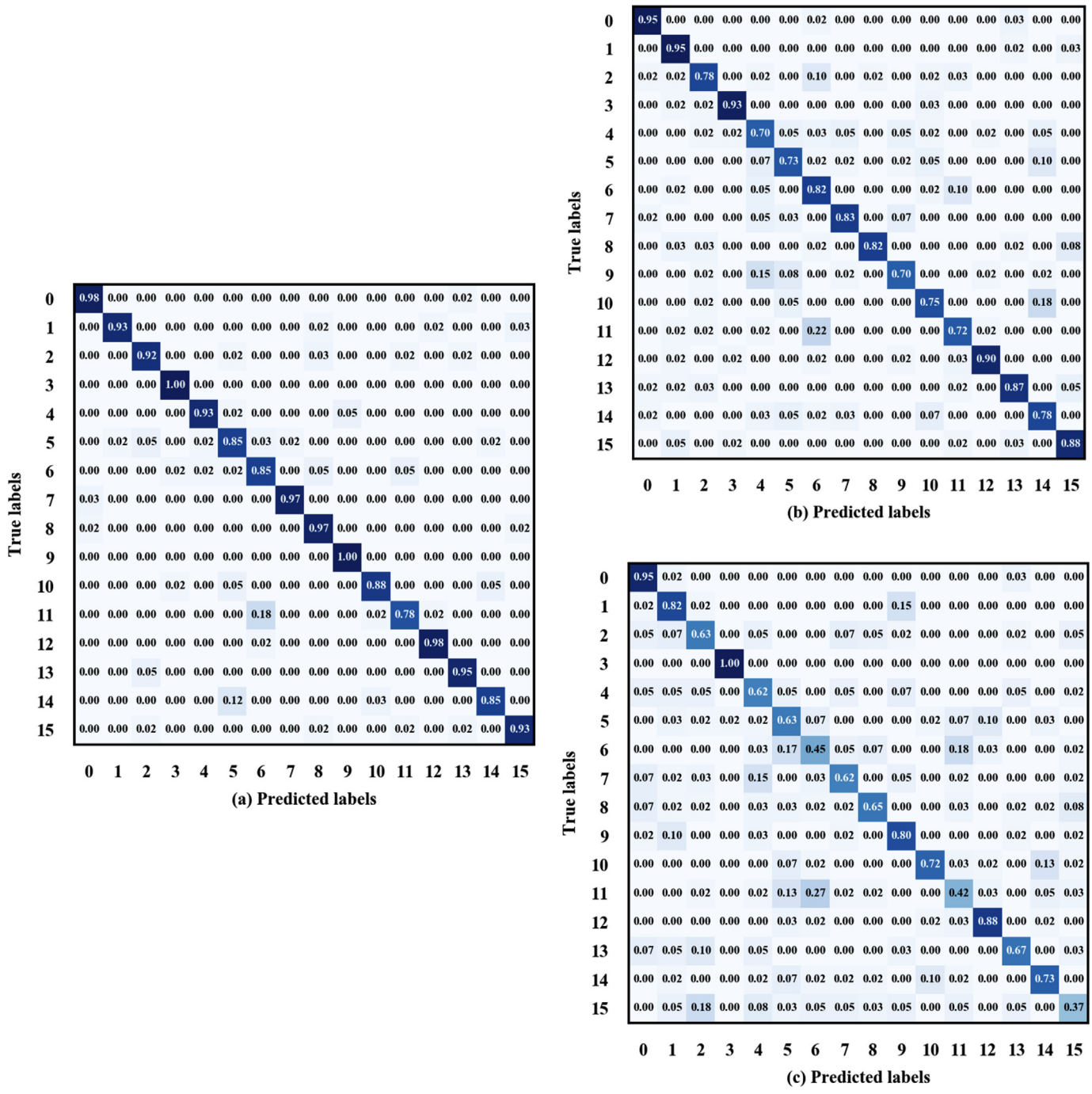

Fig. 8. Confusion matrixes of waste sorting with three methods. The ground truths of the waste labels are listed in the vertical axis while estimations are listed in the horizontal axis. The waste labels are consistent with the ones in Fig. 4. (a) 1D CNNs (b) SVM with MFCCs features (c) SVM with raw data.

methods respectively is compared, as shown in Table 7. For classifying the sixteen categories of the waste, the average time taken is 0.39s using SVM with MFCCs and the average time taken is 15.11s by using SVM with raw data. However, the time taken using our proposed 1D CNNs framework is 1.49s, which is also fast and suitable for real-time applications.

8. Conclusions and future work

Waste sorting at source is one of the most important steps for product disposal for cleaner production. This paper proposes a 1D CNN-based method to achieve intelligent waste sorting. A multi-layer nonlinear mapping structure of 1D CNNs is designed to automatically extract features from raw acoustic data. Three key hyper-parameters in the 1D CNNs are investigated based on orthogonal experiment method to optimize the network performance. It is also proved that the proposed method can achieve better classification performance compared to the traditional method using handcrafted features and shallow classifiers.

This paper presents our initial investigation of 1D CNN-based method for waste sorting. There are still a lot of work to be done in the future. Firstly, the dataset should be increased, in terms of not

only the number of samples for each category, but also the number of types of wastes. In this way, the method could be tested in a more complex scenario and the results can gain more confidence. Secondly, the network structures and hyper-parameters investigated in this research are still limited. More complex structures could be explored for better performance. Finally, how to integrate it into the waste bins and create smart bins is another topic to be studied. It will be much more efficient if the bins are smart enough to help people classify their wastes and can communicate with other agents via the internet for higher level waste management planning.

CRediT authorship contribution statement

Gang Lu: Methodology, Software, Investigation, Writing - original draft. **Yuanbin Wang:** Conceptualization, Validation, Writing - review & editing. **Huayong Yang:** Supervision. **Jun Zou:** Project administration, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China [Grant No. 51875507].

References

Bouvrie, J., 2006. Notes on convolutional neural networks. MIT CBCL Tech. Rep., pp. 38–44
Cai, S., Bao, G., Ma, X., Wu, W., Bian, G., Rodrigues, J.J.P.C., de Albuquerque, V.H.C., 2019. Parameters optimization of the dust absorbing structure for photovoltaic panel cleaning robot based on orthogonal experiment method. J. Clean. Prod. 217, 724–731.
Chen, F., Chen, H., Wu, M., Li, S., Long, R., 2019. Research on the Driving Mechanism of Waste Separation Behavior: Based on Qualitative Analysis of Chinese Urban Residents. Int. J. Environ. Res. Public Health 16 (10), 1859.
Deng, W.Y., Ma, J.C., Xiao, J.M., Wang, L., Su, Y.X., 2019. Orthogonal experimental study on hydrothermal treatment of municipal sewage sludge for mechanical dewatering followed by thermal drying. J. Clean. Prod. 219, 236–249.
Gundupalli, S.P., Hait, S., Thakur, A., 2017a. A review on automated sorting of source-separated municipal solid waste for recycling. Waste Manag. 60, 56–74.
Gundupalli, S.P., Hait, S., Thakur, A., 2017b. Multi-material classification of dry recyclables from municipal solid waste based on thermal imaging. Waste Manag. 70, 13–21.
Gundupalli, S.P., Hait, S., Thakur, A., 2018. Classification of metallic and non-metallic fractions of e-waste using thermal imaging-based technique. Process Saf. Environ. Protect. 118, 32–39.
Huang, J., Bian, Z., Lei, S., 2015. Feasibility study of sensor aided impact acoustic sorting of plastic materials from end-of-life vehicles (ELVs). Appl. Sci. 5, 1699–1714.
Huang, J., Tian, C., Ren, J., Bian, Z., 2017. Study on impact acoustic—visual sensor-based sorting of ELV plastic materials. Sensors 17, 1325.
Jull, H., Bier, J., Kunnemeyer, R., Schaare, P., 2018. Classification of recyclables using laser-induced.
Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., Inman, D.J., 2019. 1D Convolutional Neural Networks and Applications: A Survey, 03554 arXiv: 1905.
Klatzky, R.L., Pai, D.K., Krotkov, E.P., 2000. Perception of material from contact sounds. Presence: teleoperat. Vir. Environ. Times 9 (4), 399–410.
Knickmeyer, D., 2020. Social factors influencing household waste separation A literature review on good practices to improve the recycling performance of urban areas. J. Clean. Prod. 245, 118605.
Korucu, M.K., Kaplan, Ö., Büyük, O., Güllü, M.K., 2016. An investigation of the usability of sound recognition for source separation of packaging wastes in reverse vending machines. Waste Manag. 56, 46–52.
Krotkov, E., Klatzky, R., Zumel, N., 1997. Robotic perception of material: experiments with shape-invariant acoustic measures of material type. In: Experimental Robotics, vol. IV, pp. 204–211.
Li, J., Li, C., Liao, Q., Xu, Z., 2019. Environmentally-friendly technology for rapid online recycling of acrylonitrile-butadiene-styrene, polystyrene and polypropylene using near-infrared spectroscopy. J. Clean. Prod. 213, 838–844.
Luo, S., Zhu, L., Althoefer, K., Liu, H., 2017. Knock-Knock: acoustic object recognition by using stacked denoising autoencoders. Neurocomputing 267, 18–24.
Nizetic, S., Djilali, N., Papadopoulos, A., Rodrigues, J.J.P.C., 2019. Smart technologies for promotion of energy efficiency, utilization of sustainable resources and waste management. J. Clean. Prod. 231, 565–591.
Özkan, K., Ergin, S., Işık, S., Işıklı, İ., 2015. A new classification scheme of plastic wastes based upon recycling labels. Waste Manag. 35, 29–35.
Rahman, M.O., Hussain, A., Scavino, E., Basri, H., Hannan, M.A., 2011. Intelligent computer vision system for segregating recyclable waste papers. Expert Syst. Appl. 38 (8), 10398–10407.
Ren, Z., Qian, K., Wang, Y., Zhang, Z., Pandit, V., Baird, A., Schuller, B., 2018. Deep scalogram representations for acoustic scene classification. IEEE/CAA J. Autom. Sinica 5 (3), 662–669.
Ruiz, V., Sanchez, A., Velez, J.F., Raducanu, B., 2019. Automatic Image-Based Waste Classification. International Work-Conference on the Interplay between Natural and Artificial Computation, pp. 422–431.
Sanders, M.S., McCormic, E.J., 1993. Human Factors in Engineering and Design. McGraw-Hill Education, Publisher.
Srinilta, C., Kanharattanachai, S., 2019. Municipal Solid Waste Segregation with CNN. IEEE, pp. 1–4.
Sun, X., Guo, M., Ma, M., Mankin, R.W., 2018. Identification and classification of damaged corn kernels with impact acoustics multi-domain patterns. Comput. Electron. Agric. 150, 152–161.
Wang, Z., Peng, B., Huang, Y., Sun, G., 2019. Classification for plastic bottles recycling based on image recognition. Waste Manag. 88, 170–181.
Xie, J., Towsey, M., Zhang, J., Roe, P., 2019. Investigation of acoustic and visual features for frog call classification. J. Signal Process. Sys 1–14.