2020 the 5th International Conference on Control, Robotics and Cybernetics

Research on Computer Vision-Based Waste Sorting System

Haochen Cai, Xinli Cao, Likun Huang, Lianying Zou, Shubin Yang
School of Electrical and Information Engineering
Wuhan Institute of Technology
Wuhan, China
e-mail: chchen27@163.com, xinli_cao@126.com

Abstract—**Nowadays, waste sorting has become a hot topic of society in China. Many cities, such as Beijing and Shanghai, have begun to strictly implement regulations of waste sorting. However, in this process, it still exists that people are not able to distinguish them for residual waste or household food waste. This paper utilizes computer vision approach to recognize two categories of wastes. We find out typical examples from them for image recognition, and collect relevant image dataset through the Internet, with a total of 2800 images. The methods used are support vector machine based on feature extraction and transfer learning based on convolutional neural network. Our experiment shows that the latter has better performance.**

Keywords-waste sorting; computer vision; feature extraction; support vector machine; convolutional neural network; transfer learning

I. INTRODUCTION

In recent years, with the rapid development of global industrialization and urbanization, there is an unprecedented rise in the generation of waste worldwide [1]. “Junk-besieged city” is becoming a global trend, which will doubtless do great damage that would serve as a wake-up call to people. With the improvement of people's living standard, the composition of waste is also complex and diverse. They contain a large number of perishable materials such as vegetable leaves and pericarp, which are mixed with metals, paper, glass, plastics, increasing the difficulty of waste sorting. The classified collection and treatment of waste is particularly important. Efficient recycling can be regarded as an effective solution to the waste of natural resources [2] and environmental pollution. At the stage of waste sorting, source segregation is often performed for a preliminary sorting of recyclables. In fact, all household wastes should be separately collected and treated at the source instead of mixed collection.

Researchers worldwide have been actively exploring automated sorting techniques for efficiently processing increasing quantities of waste. Previously, in the fields related to the automatic waste sorting for recycling [3], the physical characteristics of the objects were often used or a chemical method was used to determine the chemical composition to separate certain type of waste. For example, triboelectrostatic separation technology [4] is used to sort plastic from waste, and low-cost sorting technology [5] is used to separate waste paper. Currently, manual segregation in recycling plant is widely used, and the efficiency cannot reach the required level to have a better recycling rate. It can also cause harm to the health of sorting workers. At the collecting stage, waste sorting also should pay attention to these problems.

The application of computer vision could be an efficient way for waste sorting. Computer vision is to use camera and computer instead of human eye to acquire and analyze the target, and further do image processing, so that the computer can obtain a high-level understanding from the digital image or video. Image classification is one of the most basic tasks of computer vision. This operation can be simple to assign a label to an image, such as cat, dog, or advanced to interpret the content of the image and output a human readable sentence. This field has attracted the attention of a large number of researchers for a long time. The main method of early image recognition is to extract feature vectors from images and classify them by machine learning. The biggest problem is that it is difficult to design corresponding image features to represent new categories when there are more categories.

Since 2010, many image recognition methods using deep learning have been proposed. In some image recognition competitions, the image recognition method using deep learning is much better than that before deep learning. Using the deep learning method, feature extraction can be performed automatically from images directly [6]. Because of this special ability, deep learning has excellent efficiency in vision-based tasks such as pattern recognition and image classification [7]. In many published paper, major directions of research are target detection and image classification. And in the aspect of object detection [8], the image processing is further optimized.

Based on the computer vision technology to achieve waste sorting, the traditional machine learning and deep learning methods are used respectively in this paper. Through experiments and their results, the advantages and disadvantages of garbage image recognition are compared, and the development trend and related problems of waste sorting based on computer vision are introduced.

II. MATERIALS AND METHODS

A. Dataset

In order to realize the ideal garbage classification and recognition, it is necessary to have a garbage image dataset as the basis of the experiment. However, there is currently no such dataset that can be used directly.

We have collected the garbage image dataset ourselves and labeled each image with the corresponding name. Data collection has always been a time-consuming and labor-

978-1-7281-8624-5/20/$31.00 ©2020 IEEE

consuming work. In order to complete the establishment of garbage image dataset quickly and conveniently, we implement it by searching the image corresponding to the name of garbage in each category in the guide of classification and delivery of domestic garbage issued by several cities in China, such as Beijing and Shanghai.

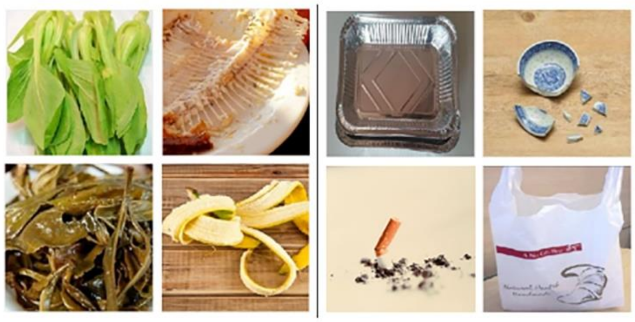


Figure 1. Example images of household food waste and residual waste.

In actual application scenarios, the samples to be classified are often uncontrollable, so the category of "other" is generally added to accommodate various abnormal samples. In garbage sorting, except recyclable, hazardous and kitchen waste, all belong to residual waste, so the waste plays an "other" role. Generally, we have a clear understanding of recyclables and hazardous waste, but it is difficult to distinguish residual waste and household food waste. The household food waste is easy to spoil. Our motivation is to find a way to automatically recognize these garbages.

We find out typical examples from residual waste and household food waste for image recognition. The household food waste image dataset includes four parts: vegetable leaves, fishbone, tea residue and pericarp. And residual waste image dataset, including broken ceramics, cigarette butts, disposable tableware, stained trash bags. There are 350 images in each category in the dataset, with a total of 2800 images. Part of the image of the dataset is shown in Fig. 1.

B. Data Allocation

Before image classification, some representative data are taken from the picture data to be collected as training samples. In addition, a certain number of test samples are required. This work is an important part of the experiment. When the algorithm has no room for improvement, a good training dataset is usually constructed to improve the classification effect. The training dataset should satisfy the following conditions:

- The training dataset should be representative.
- No wrong samples in the training dataset.
- The training dataset should be as complete as possible.
- The images in test set cannot be trained in advance.

C. Method

Human beings are good at identifying a new thing through a small number of samples. For example, children only need to know what is "apple" or "banana" from some pictures in the book, while machines need a lot of data support as the basis for learning.

We use support vector machine as a classifier for feature extraction algorithm [9] and convolutional neural network [10] to identify garbage image dataset, in order to evaluate their performance on garbage classification and recognition problems.

In traditional machine learning, image classification needs a feature extraction process. An algorithm based on researcher's knowledge is used to extract and express image features, which is called hand-crafted feature. Convolutional neural network (CNN) is a special deep learning method, which is used to learn classification and feature extraction from training samples, as shown in Fig. 2.

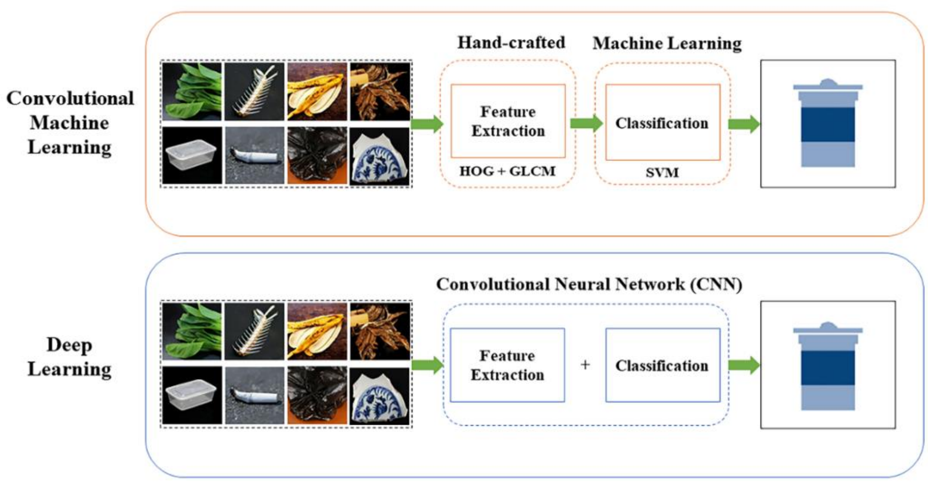


Figure 2. Conventional machine learning and deep learning.

III. WASTE SORTING BASED ON FETURE EXTRACTION

For image feature extraction [11], the combined features of both HOG feature and gray level co-occurrence matrix are used as image features. The schematic block diagram of garbage recognition based on feature extraction is shown in Fig. 3.

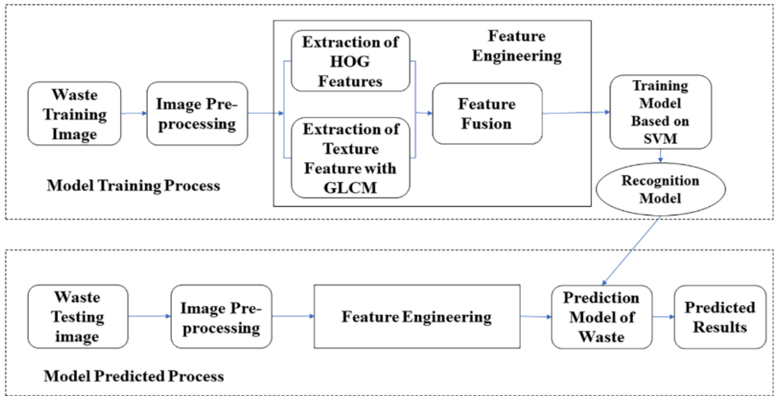


Figure 3. Schematic block diagram of waste sorting process based on feature extraction.

A. *Extraction of HOG Feature*

The HOG+SVM method was proposed by French researcher Dalal [12] at the 2005 CVPR, and its purpose is for pedestrian detection. The HOG feature divides the entire detection window into blocks, and each block is composed of several cells. The image feature is constructed by calculating and counting the gradient direction histogram of the local area of the image. This feature has the ability to describe the structure and outline of the object, and has a strong identification effect on the description of the local area. It can well imitate human vision and accurately describe the appearance edges and structural features of objects in the image.

In this article, the size of the input image is 128×128, each cell is 4×4 pixels, and the cell selects 9-dimensional features. The adjacent 2×2 cell units form a block, and each block corresponds to a 36-dimensional feature vector. The scanning step size is a cell unit, and it is calculated that there are 31 scanning windows in the horizontal and vertical directions respectively. Finally,4×9×31×31=34596 dimensional features can be obtained.

B. *Extraction of Texture Feature with GLCM*

Haralick [13] proposed a texture description method of gray level co-occurrence matrix, which is one of the earliest methods for image texture analysis. The texture is expressed by the gray distribution of pixels and their surrounding spatial neighborhood. It can only reflect the characteristics of the surface of the object and cannot fully reflect the essential attributes of the object, so only by using texture features [14] cannot obtain high-level content of the image. In this paper, the gray level cooccurrence matrix (GLCM) is used to extract the texture features of the image. The direction of co-occurrence matrix can be generated from four directions: 0°, 45°, 90°, and 135°. The four features of contrast, energy, correlation and homogeneity are selected, and then the average value and variance of these features are calculated as the final extracted image texture features.

C. *Support Vector Machine*

SVM is based on optimization theory to deal with machine learning. It is mainly used to solve the problem of two kinds of classification. It also provides a method to deal with nonlinear problems [15]. The "1V1" multi-classification mode of support vector machine, that is, for a k-class problem, a SVM is designed between any two classes of samples. Therefore, k(k-1)/2 SVM should be designed for k-class samples. For the classifier that distinguishes class i and j, the class i sample is defined as positive class sample, and class j sample is defined as negative class sample. The relationship can be expressed as:

$$\begin{gathered} \underset{w^{ij},\, b^{ij}}{min} \; \frac{1}{2}\left\|w^{ij}\right\|^2 + c\sum_{j=1} \xi_t^{ij} \\ s.t.\;\; y_t\left((w^{ij}x_t) + b^{ij}\right) \geq 1 - \xi_t^{ij},\;\; \xi_t^{ij} \geq 0 \end{gathered} \tag{1}$$

In this paper, this mode is used to train SVM classifier. And the images are classified by the template. For the dataset, the ratio of training and test data split is 7:3.

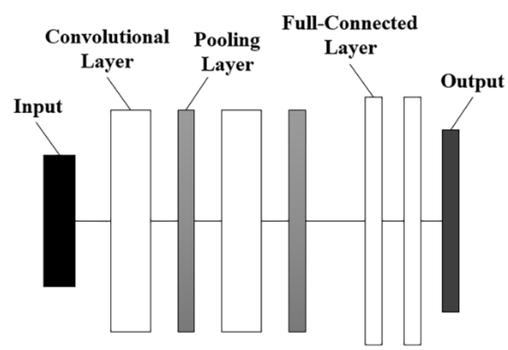


Figure 4. The main structures of convolutional neural networks for image classification problems.

IV. WASTE SORTING BASED ON CNN

CNN is a deep learning model or multilayer perceptron similar to artificial neural network, which is often used to analyze visual images. The main structures [16] of convolutional neural network for image classification are input, convolution layer, pooling layer, full connection layer and output, as shown in Fig. 4. A convolutional neural network is formed by stacking these hierarchies.

This paper uses two models, AlexNet and VGG19, and compares the advantages and disadvantages of the two in the garbage image classification.

A. AlexNet

The model is a deep network model proposed by Alex Krizhevsky [17]. As the winner of ILSVRC 2012, AlexNet has an 8-layer structure, the first 5 layers are convolution layers, and the last 3 layers are fully connected layers. The last fully connected layer is the Softmax classification layer with 1000 output. In addition, it adds a Relu activation function after each convolutional layer and applies a dropout layer, which alleviates the overfitting of the model.

B. VGGNet

VGGNet [18] is a deep convolutional neural network developed by the Visual Geometry Group at Oxford University and researchers at Google DeepMind. VGGNet contains a multi-layer network structure with a depth ranging from 11 to 19 layers. The more commonly used are VGG16 and VGG19. It inherits the idea of AlexNet, and based on AlexNet, has established a network with more layers and deeper structure. VGGNet all uses 3×3 convolution kernels and 2×2 pooled kernels to improve performance by continuously deepening the network structure.

Although the network has satisfactory performance in multiple transfer learning tasks, it consumes more computing resources and uses more parameters, which makes its computing cost higher and needs more memory when optimizing learning parameters.

C. Transfer Learning

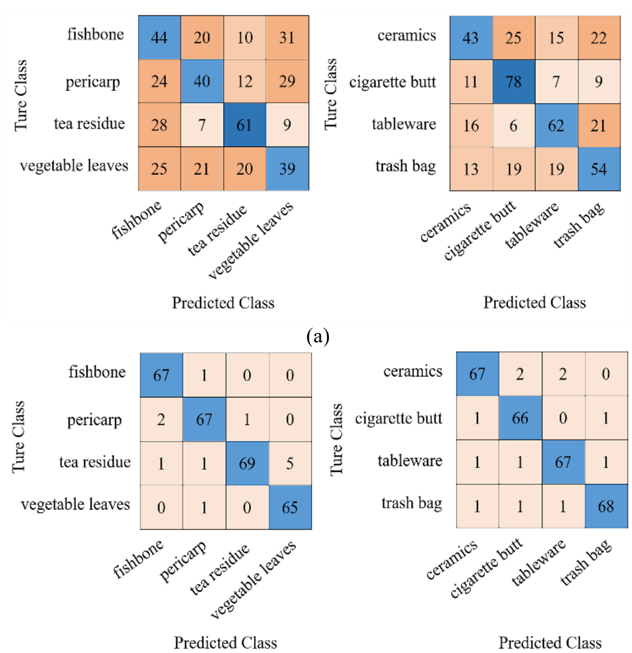
For image classification using deep learning, the rule of thumb is that each classification requires at least 1000 images. Due to insufficient data images collected, transfer learning method [19] is introduced to solve this problem. Transfer learning is generally aimed at the case where the target task has less training data. By loading a pre-trained model and combining less data for retraining, it can effectively transfer to such a visual task with insufficient training data, so as to optimize the classification effect.

In the training process of transfer learning, the weight of CNN with millions of images previously trained is assigned as the initial value. In order to further improve the stability and accuracy of the network for waste sorting, it is necessary to fine tune the network model, while preserving the feature expression in the pre-training dataset, and combining the garbage image dataset to further fit the extracted features.

To compensate for the lack of image data, a method of enhancing image data is used, that is, to perform an enhancement operation on the training image: the training image is randomly flipped along the vertical axis and 30-pixel units are randomly moved horizontally and vertically on the image. During the fine-tuning process, the learning rate is set to 10^{-5}, and the stochastic gradient descent method is used for optimization. It needs 84 iterations per cycle and 8 cycles of training, with a total of 672 iterations. To verify our results, we used proportional stratified random sampling and divided the training data into two subsets, training and verification. The ratio of training set, verification set and test set is 6:2:2.

V. CLASSIFICATION PERFORMANCE DISCUSSION

Confusion matrix [20] is a commonly used form of expression in the field of pattern recognition. It depicts the relationship between the true attributes of sample data and the type of recognition results, and is a common method for evaluating the performance of classifiers. We tend to use accuracy because it is familiar with its definition, not because it is the best tool for evaluating models. It is very basic to measure the performance of machine learning models with indicators such as precision and recall. Accuracy refers to how many judgments are correct among all judgments. Precision is relative to the prediction result, and the recall is relative to the sample. The performance parameters of the confusion matrix from the SVM run on the test data and the confusion matrix are given in Fig. 5 and Table I, respectively.



(b)

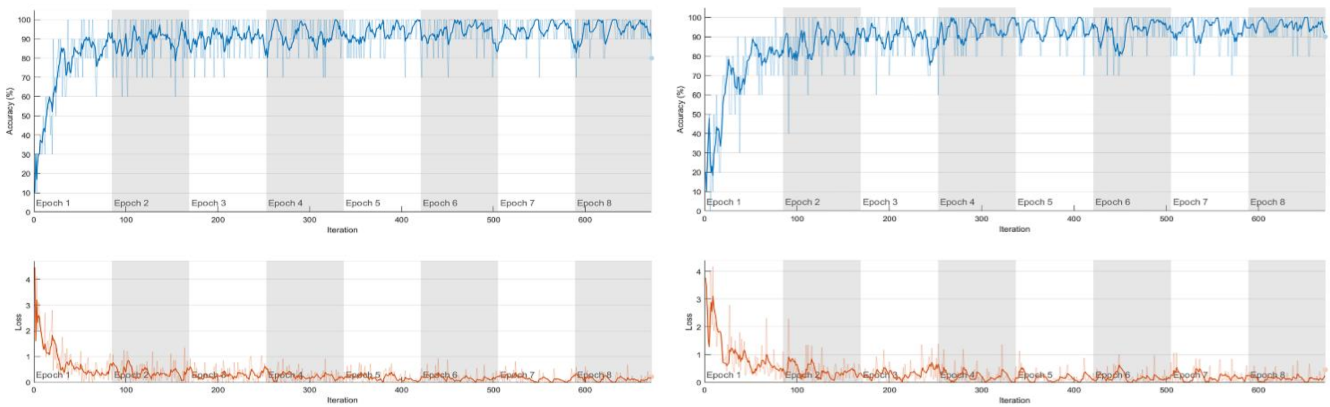
Figure 5. Confusion matrix from the SVM run on the test data (a) and Confusion matrix from the CNN run on the test set (b).

TABLE I. THE PERFORMANCE PARAMETERS OF THE CONFUSION MATRIX FROM THE SVM RUN ON THE TEST DATA

	Fishbone	Pericarp	Tea residue	Vegetable leaves	Ceramics	Cigarette butt	Tableware	Trash bag
Precision	41.9%	38.1%	58.1%	37.1%	41.0%	74.3%	59.0%	51.4%
Recall	36.4%	45.5%	59.2%	36.1%	51.8%	60.9%	60.2%	50.9%

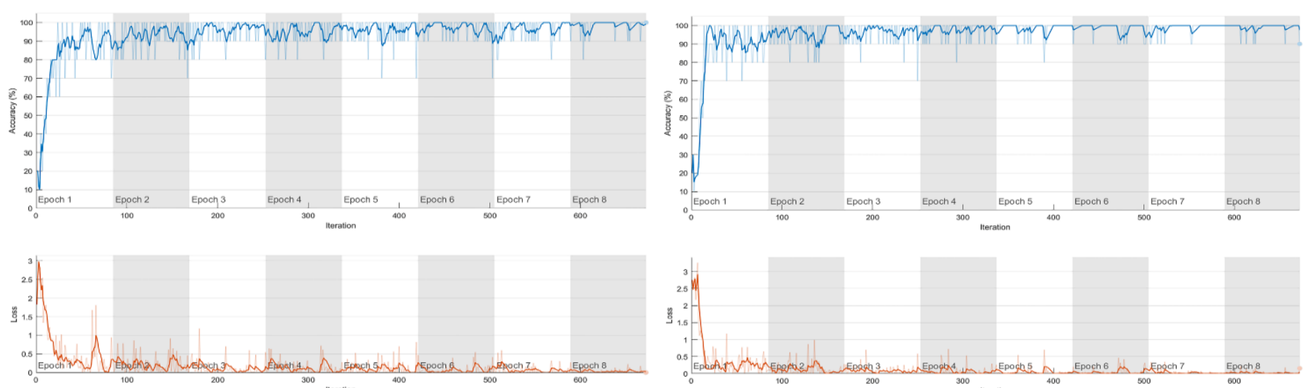
All experimental results adopt the method of averaging the values of repeated tests. It can be seen that the accuracy of the garbage image classification based on feature extraction and support vector machine is not very high. In the sample test data, only 56% and 44% of the test accuracy for recognition of residual waste and household food waste. However, we can find that for specific garbage such as tea residue and cigarette butt, the recognition rate of this method has a better effect, which is also related to the image dataset collected on the Internet. As mentioned before that when the algorithm has no room for improvement, building a good training set can improve the classification effect. If we subdivide the pericarp in household food waste and the broken ceramics in residual waste, it may have a good effect. But this will add more SVM classifiers. We will improve it in future work.

The training loss and model accuracy of the two pre-training models during training are shown in the Fig. 6 and Fig. 7. As the number of iterations increasing, the loss function gradually converges, and the classification accuracy gradually increases. The accuracy of the test set using VGGNet is up to 97%, and classification effect of VGGNet is superior to AlexNet.



(a) (b)

Figure 6. The training loss and model accuracy of the AlexNet in household food waste (a) and residual waste (b) image classification.



(a) (b)

Figure 7. The training loss and model accuracy of the VGG19 in household food waste (a) and residual waste (b) image classification.

But in the same state, the training time of the two models is different. Table II shows the average training time for each iteration of batch processing size 20, and the average test time for each image classification and recognition result on the test dataset, the time unit is seconds. AlexNet has less training and testing time. This is caused by the different depth structure of convolutional neural network.

TABLE II. AVERAGE TRAINING TIME PER ITERATION AND TEST TIME

	AlexNet	**VGG19**
Training time	50	455
Test time	2.78	38.36

Compared with the classification methods based on feature extraction and based on convolutional neural network, the latter has batter performance. This is a trend discovered by computer vision, and it is also the direction of our further research in the future.

VI. CONCLUSION

With the rapid development of computer vision technology, intelligent image and video processing system has gradually become the common demand of people. We can use computer vision to achieve the task of waste sorting. Although the classification effect of traditional machine learning is not very ideal compared with deep learning, it is also worth studying. The feature extraction should be more diversified and adapted to the target task. At the same time, we have also used the deep learning.

The network used is based on the pre-trained models of AlexNet and VGG19. We evaluated two models on the garbage image dataset and used different network parameters and data enhancements. In the paper, we gave the parameters of the best classification prediction, thereby improving the system performance. Considering the time cost and performance improvement, an appropriate network model can better improve the efficiency of the classification task. We hope and believe that the observations collected in this work will provide some inspiration for other similar visual recognition problems.

In the future, we plan to use the latest progress of convolutional neural network research to try to classify multiple kinds of garbages in more complex background [21], to realize in the same image to distinguish different garbages, which will be a challenging work. To achieve this goal, a large and growing garbage dataset needs to be built.

ACKNOWLEDGMENT

This research was supported by Cooperative education project between production and education of the Ministry of Education of China (201801244007 and 201801244017).

REFERENCES

[1] I. A. Al-Khatib, M. Monou, A. S. F. A. Zahra, H. Q. Shaheen, D. Kassinos, "Solid waste characterization, quantification and management practices in developing countries. A case study: Nablus district – Palestine," Journal of Environmental Management, 91(5), 2010, pp.1131-1138.

[2] A. D. Read, "A Weekly Doorstep Recycling Collection, I had no Idea We Could Overcoming the Local Barriers to Participation," Resources Conservation & Recycling, vol. 26, 1999, pp. 217 -249.

[3] S. P. Gundupalli, S. Hait , and A. Thakur, "A review on automated sorting of source-separated municipal solid waste for recycling," Waste Management, 60(feb.), 2017, pp.56-74.

[4] Guiqing Wu, Jia Li, and Zhenming Xu, "Triboelectrostatic separation for granular plastic waste recycling: A review," Waste Management, 33(3), 2013, pp.585-597.

[5] M. O. Rahman, A. Hussain, and H. Basri, "A critical review on waste paper sorting techniques," International Journal of Environmental Science & Technology, 11(2), 2014, pp. 551-564.

[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521,no. 7553, 2015, pp. 436–444.

[7] J. Amara, B. Bouaziz, A. Algergawy, "A deep learning-based approach for Banana leaf diseases classification," BTW (Workshops), 2017.

[8] Hanyu Hong, Restoration method and application for multi-spectral image in object detection, Beijing, China: National Defense Industry Press, 2017.

[9] D. Decoste, and B. Scholkopf, "Training Invariant Support Vector Machines," Machine Learning, 46(1/2/3), 2002, pp.161-190.

[10] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition," Proc. IEEE 86, 1998, pp.2278–2324.

[11] S. K. Kim, Y. J. Park, K. A. Toh, & S. Lee, "SVM-based feature extraction for face recognition," Pattern Recognition, 43(8), 2010, pp.2871-2881.

[12] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in: IEEE Comp. Soc. Conf. Comp. Vis. Patt. Recog., Vol. 1, 2005, pp. 886–893.

[13] R. M. Haralick, "Statistical and Structural 12 Approaches to Texture," Photogrammetric Engineering and Remote Sensing, 67(5), 1978.

[14] Li Liu, and Gangyao Kuang, "Overview of image textural feature extraction methods," Journal of Image and Graphics, 2009.

[15] J. A. K. Suykens, Support vector machines: A nonlinear modelling and control perspective, Eur. J. Control 2001, 7, 311-327.

[16] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, "Imagenet large scale visual recognition challenge," Int. J. Comput. Vis. 115 (3), 2015, pp.211–252.

[17] A. Krizhevsky, I. Sutskever, G.E. Hinton,"ImageNet classification with deep convolutional neural networks," in: Advances in neural information processing systems (NIPS), 2012, pp. 1106–1114.

[18] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Computer Science, 2014.

[19] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," Advances in Neural Information Processing Systems, MIT Press, 2014.

[20] M. Sokolova, and G. Lapalme, "A systematic analysis of performance measures for classification tasks," Information Processing & Management, 45(4), 2009, pp.427-437.

[21] Hanyu Hong, Xiangyun Guo and Xiuhua Zhang, "An improved segmentation algorithm of color image in complex background based on graph cuts," 2011 IEEE International Conference on Computer Science and Automation Engineering, Shanghai, 2011, pp. 642-645, doi: 10.1109/CSAE.2011.5952551.