A Web-Based Dataset for Garbage Classification Based on Shanghai's Rule

Yunyi Liao

***Abstract*—In 2019, Shanghai has published a new regulation of trash management, which obtains a series of achievements in managing trash. In order to implement this regulation further and help citizens understand clearer about trash classification, we decided to develop a deep learning dataset and models to classify waste automatically. The object of this study is to take an image as input and identify the category of trash. We write a web crawler to capture images. After pre-processing, we gain about 14,000 images in total. We compare models including CNN, a ResNet50 model, and a VGG16 model on the dataset. Our experiments show that the ResNet50 model perform better than the others.**

***Index Terms*—Dataset, garbage classification, web crawler, convolutional neural network, transfer learning.**

I. Introduction

Urban littering means that people dispose waste products incorrectly and manage them improperly. As for the size of cities is growing and urban population is increasing, the management of urban littering has become a non-negligible problem. Shanghai is a very large, beautiful and world-renowned metropolitan city. In 2017, Shanghai's families generated a daily average of 24,700 tons of waste, or 1.02 kilograms per person per day, calculated on a number of permanent residents of 24.18 million. Urban cleanliness directly impacts citizen's living experience, economic construction and development of tourism. So urban littering is becoming a huge problem that China government needs to consider. From July 1st, 2019, Shanghai is forced to enact a new trash management regulation in order to solve this serious problem [1]. Based on the new regulations, all individuals and business companies need to divide and sort their trash into four categories, which include recyclable waste, household food waste, residual waste, and hazardous waste. Because people are used to throwing their trash without any rules, it is hard to make them understand how to differentiate which specific category their trash should belong to.

Mobile Internet developed very quickly in the past decade, and people rely on smartphones in their daily life. The data collected from smartphones or the Internet have supported many research projects, e.g., image classification [2], language translation [3], and traffic pattern analysis [4], [5], etc. So the simplest and cheapest way to let people understand how to categorize their trash is to build an application in their phone for this purpose, if people feel confused about the category of a specific waste, they can use this application and find a way to solve it quickly. To make it into reality, a waste dataset is needed to train a machine learning algorithm to support automatic classification. In this paper, we want to demonstrate that this application is feasible by leveraging the state-of-art deep learning models [6], which have been proved to be effective in a variety of problems, e.g., handwritten digit recognition [7], face recognition [8], stock market prediction [9], and traffic forecasting [10]. Furthermore, it can be continuously improved with an increasing dataset and provide more accurate result in the future.

For now, there is no trash dataset accord with Shanghai's rule available in the previous studies. In the first step, we write a crawler program to help us capture images from Google image and Baidu image (Baidu is the most popular search-engine in China), based on the keywords released by Shanghai government. Then we choose and annotate manually part of the images to use in our research. After data cleaning, we use the state-of-the-art deep learning models to build classifiers for garbage classification.

In Section II, we present the related datasets and classification methods. Section III discusses the approach in which we gain our dataset. In Section IV, we compare several deep learning models for garbage classification based on our dataset and discuss the results. We would also give the future directions in Section V. We conclude this paper in Section VI.

II. Related Work

Trash management is a universal problem in different areas, e.g., China, Europe and India. This problem is drawing more and more attention nowadays because of the terrifying realty that cities are being surrounded of a huge amount of trash and people have to migrate to seek clean water and air. In this section, we would give a short review of the previous studies which aimed to solve this problem from two aspects, garbage datasets and garbage classification methods.

A. Garbage Datasets

Garbage Datasets are the basic of training models for automatic garbage classification. Previous studies are conducting experiments on some small and limited datasets, and we want to increase the dataset size and build a larger dataset accord with Shanghai's rule in this paper. The previous datasets are summarized in Table I.

In [11], the authors categorize solid waste into four types including general waste, compostable waste, recyclable

Manuscript received May 12, 2019; revised April 17, 2020.

The author is with the School of Information Science and Technology, Xiamen University Malaysia, Jalan Sunsuria, Bandar Sunsuria, 43900 Sepang, Selangor Darul Ehsan, Malaysia (e-mail: swe1709227@xmu.edu.my).

doi: 10.18178/ijmlc.2020.10.4.979

waste and hazardous waste. They make a list about items of each types and gain 9200 images with a resolution of 224× 224 pixels in total from Food-101 dataset, Cola bottle identification dataset, Home object dataset, Flickr Material Database (FMD), Glasssense-Vision dataset, Glasses and bottles dataset and waste images scraped through Google Search. In [12], the authors divide trash into 6 main classes, consisting of glass, paper, metal, plastic, cardboard, and other trash. They collect data manually in Flicker Material Database and Google Images. They totally gain 400-500 images per class, which have 256×256 pixels' image resolution.

Compared with the datasets used for other tasks, which often have a million level of images, the garbage datasets are relatively small for now. This situation is caused by the requirement of the garbage definition. Not every picture contains a class of trash. And for now, the garbage datasets don't have a common rule of different classes. To address this issue, we would follow the Shanghai's rule for garbage classification, which contains a specific and detailed explanation of different categories. The hot discussions of Shanghai's rule online also provide many examples of the garbage, which makes the pictures easier to collect.

TABLE I: A SUMMARY OF EXISTING GARBAGE DATASETS

Paper	Class	Amount	Image Resolution
[11]	4	9,200	224×224
[12]	6	2,400	256×256
[13]	25	6,896	640×480
[14]	3	340	520×380
[15]	3	2,000	256×256
[16]	4	5,000	240×240
[17]	2	450	256×256
[18]	9	681	420×400

B. Garbage Classification Methods

Convolutional neural networks and support vector machines are widely used by researchers for garbage classification. Convolution neural network is an extraordinary model for classifying data. Researchers need to pick and design the network structure manually and then build the classifiers.

CNN is a very powerful architecture for image recognition and identification.

In [2], CNN models using the structure of Siamese network are proposed and evaluated on 7 handwritten digit datasets for recognition. CNN models are not only useful for two-dimensional images, but also applicable for one-dimensional data, e.g., time series. In [19], one-dimensional fully convolutional network and residual neural networks are compared with traditional distance-based classifiers for time series classification problems and the neural networks outperform traditional methods.

In [11], the authors use four different types of architecture and compare their accuracies. But something need to be mentioned here is this research only implements in solid waste, and the range is not wide. Usage scenarios are not extensive enough, either. In [12], the authors use SVM and CNN to classify their dataset. For SVM, the ratio for the training set and testing set is 70% to 30%. They use the SIFT features and get a 63% accuracy eventually. As for CNN, they use Touch framework and Lua language, and implement a 11-layer CNN which has the similar but smaller architecture with the AlexNet architecture due to computational constrains. They separate dataset with 70% for training, 13% for validating and 17% for testing but only achieve 22% accuracy in the end. The poor performance of CNN is attributed to the pool choice of hyper parameters and the lack of a larger dataset.

While training on a larger garbage dataset is unavailable currently, transfer learning can be applied for the garbage classification task. Transfer learning has been proved effective for different domains of tasks, especially those without enough training samples. There are two ways of applying the transferred models. One way is to mix the data from different domains for the collaborative training and the other way is to use the pre-trained models from one domain and fine-tune them in the other domain. For example, we can use the models pre-trained on larger dataset, e.g., ImageNet, for garbage classification, to achieve a better performance.

III. DATASET DESCRIPTION

A. Categories

There are four main categories, as we mentioned earlier. There are 60 sub-classes under the four main categories too, as shown in Table II.

B. Web Crawler

In order to gather adequate training samples from websites, we write a web crawler to help us capture images from Baidu Image and Google Image. For each sub-classes, we set the web crawler to download up to 300 images, both from Baidu and Google. We add filter conditions including the white background and a resolution requirement higher than 200×200. However, some websites are expired, and we totally collected near 42,000 images. While these images cannot be directly used as they have different resolutions and formats and some of them are not closely related to the keywords as garbage types, we have to select and clean the images manually, as we did in the dataset preprocessing step.

C. Dataset Preprocessing

We find out there are many irrelevant images contained in the dataset after browsing the downloaded raw images. To achieve a better training dataset, we deleted some irrelevant or inappropriate images, which are unhelpful in classification and could not convey correct information. Next, we develop a simple program to transform the size of images to a united size of 320×240. Thirdly, we filter the images which are not in a jpg format.

D. Dataset Statistics

After preprocessing, we annotate and collect 17,690 images in total for the purpose of training and testing. Specifically, there are 6,674 images for recyclable waste, 4,577 images for hazardous waste, 1,498 images for household food waste, and 4,941 images for residual waste.

For some specific classes with no or few images, we directly remove this kind of classes. In the next section, we use 234 classes selected from a total number of 240 sub-classes.

TABLE II: FOUR MAIN CATEGORIES AND THEIR SUB-CLASSES

Main categories	Sub-classes
household food waste	egg shell, shrimp shell, cookies, bread, etc.
recyclable waste	milk box, plastic, glass bottle, clothes, etc.
hazardous waste	medicine, battery, roll firm, spray, etc.
residual waste	hair, gum, tissue, lib-stick, etc.

We show an example of the household food waste in Fig. 1., which is the egg shell.

Fig. 1. The example of the household food waste.

We show an example of the recyclable waste in Fig. 2., which is the milk box.

Fig. 2. The example of the recyclable waste.

We show an example of the household food waste in Fig. 3., which is the battery.

Fig. 3. The example of the hazardous waste.

IV. METHOD

A. Models

In this section, we implement classifiers based on our dataset. We build a CNN model from scratch and compare it with pre-trained models from the transfer learning approach. We design two classification problems, classifying the four main classes and classifying the 234 sub-classes. We want to evaluate the accuracies on both problems.

Convolutional neural network is a feedforward neural network, whose artificial neuron can respond to the surrounding units in a partial coverage range, and has excellent performance for large-scale image processing, with the convolutional and pooling operations. And CNN possesses three main advantages. Firstly, the size of the convolution kernel is generally equal or smaller than the size of the input image, so the features extracted by the convolution layer will pay more attention to local areas which is very consistent with the image processing we are exposed to in daily life. In fact, each neuron does not need to perceive the global image, just only needs to focus on the local information, and then integration of the local information will be constructed at the higher level to obtain the global information. Secondly, CNN can reduce the computation though parameter sharing. Thirdly, we will not use only one convolution operation to filter the input images generally, because the parameters of a kernel are fixed, and the extracted features will be simplified. It's kind of like how we look at things physically in the really world, you have to look at things from multiple perspectives to try to avoid bias as much as possible. We also need multiple convolution checks to convolve the input images. So CNN is good at feature extraction and classification. As for this problem, we divided dataset with a ratio of 64%:16%:20% for training, validation, and testing. Then we normalization data to the range between 0 to 1, to train the model easily. After normalization, we build a classic CNN architecture for this research in Fig. 4. The specific layers include:

Layer 0: Input images of size 240×320 with 3 color channels

Layer 1: Convolution with 32 filters which have size as 3 ×3 and relu as the activation function, stride and padding as default value.

Layer 2: Convolution with 64 filters which have size as 3 ×3 and relu as the activation function.

Layer 3: Max-Pooling with a size 2×2 filter.

Layer 4: As for the output pooling layer, probability equal to 0.25 Dropout is adopted.

Layer 5: Flatter all the pixels.

Layer 6: Fully connected with 128 neurons in which the nonlinear activation function is used with the relu function.

Layer7: As for the fully connection layer, probability equal to 0.5 Dropout is adopted.

Layer 8: The last layer is a fully connection layer, the amount of neurons depends on the number of classes, and we use non-normalized log softmax scores as activation function.

We also use the transfer learning approach. Transfer learning is a machine learning technique that a model trained on one task which is re-purposed on other related task. Based on the extraordinary portability of transfer learning, it has been widely used in deep learning tasks which do not have a large amount of training dataset and need to be supported by the pre-trained models. In this study, the images we collected are not enough for training a deeper model with large numbers of parameters. Instead, we use

two models, namely, ResNet50 [20] and VGG16 [21], trained on the ImageNet dataset and fine-tune these models on our dataset.

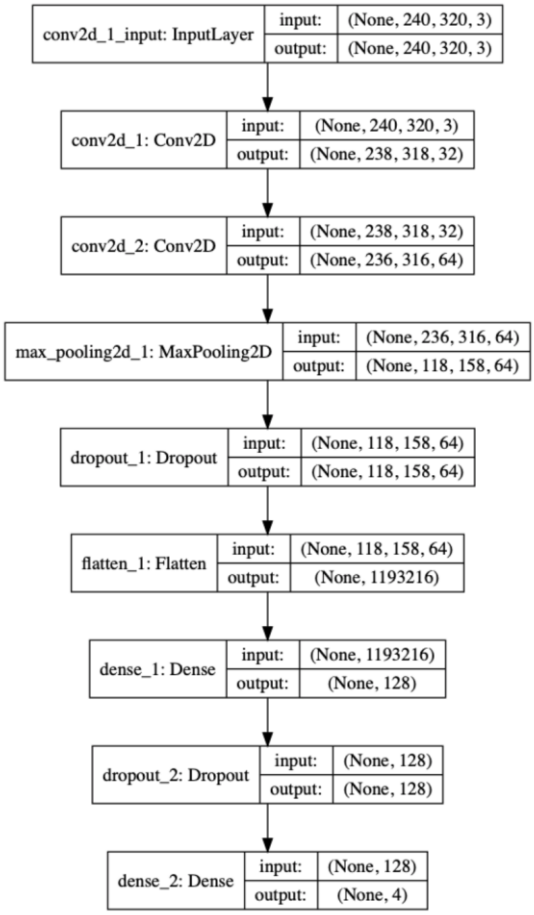

Fig. 4. The structure of the CNN model.

The residual learning block of ResNet50 used in this study is shown in Fig. 5. Theoretically better results can be obtained when the deep learning model is deeper. However, it is found through experiments that the deep network has a degradation problem due to the potential problems of vanishing and exploding gradients. When the layer increases, the accuracy of the network saturates, or even decreases. ResNet addresses this problem by introducing residual learning. Its structure is inspired by the VGG network, which is modified on the basis of which the residual unit is added through the shortcut connections mechanism. The bottleneck design is one of the implementations of residual learning which is used in ResNet50 to implement deep as well as residual learning while reducing the number of parameters. As the name indicated, ResNet50 has a structure of 50 layers.

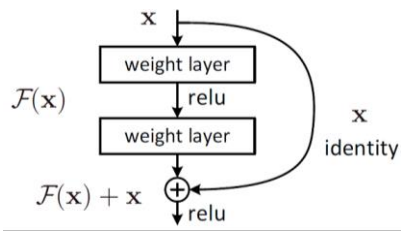

Fig. 5. The residual learning block of the ResNet50 model.

The structure of VGG16 used in this study is shown in Fig. 6. Proposed by the Oxford visual geometry group, VGG demonstrates that increasing network depth can optimize network performance to a certain extent. VGG has two variants, namely, VGG16 and VGG19. The former has 13 convolutional layers and the latter has 16 convolutional layers, both of which contain three fully connected layers. For a given receiving field, it is better to use multiple small convolutional kernels than a single large convolutional kernel because multiple nonlinear layers can increase the network depth to ensure more complex learning patterns with fewer parameters.

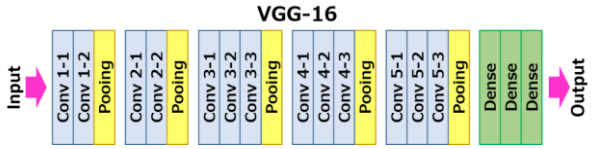

Fig. 6. The structure of the VGG16 model.

B. Implementation

The implementation of the models is conducted on a computer with Windows 10 OS, which is equipped with an Intel i5-9600K CPU (which has six 3.7GHz processors), and a NVIDIA GeForce GTX 2070 GPU with 16GB DDR4 RAM for acceleration of the convolutional operations in the deep learning models.

The loss function is the categorical cross entropy, and Adam is used as the optimizer. The batch training size is set to be 16. For each model, we train for 40 epochs. The best parameter with the highest validation accuracy is saved and further used for the evaluation on the testing set.

C. Result

We show the experiment results in Table III and Table IV.

TABLE III: THE RESULT FOR SPLITTING DATASET TO 4 MAIN CATEGORIES

Model	Training Accuracy	Validation Accuracy	Testing Accuracy
CNN	0.3758	0.3763	0.3763
VGG16	0.3543	0.3763	0.3762
ResNet50	0.5594	0.4968	0.4752

TABLE IV: THE RESULT FOR SPLITTING DATASET TO 234 SUB-CLASSES

Model	Training Accuracy	Validation Accuracy	Testing Accuracy
CNN	0.0080	0.0113	0.0116
VGG16	0.0147	0.0162	0.0161
ResNet50	0.9594	0.1783	0.1843

From Table III and Table IV, we can compare the models both horizontally and vertically. We discovered that the accuracy of pre-trained models is better than using a CNN directly and ResNet50 is even better than VGG16 horizontally. We believe the main reason leading to this result is that the size of dataset is too small, and we cannot obtain an accurate parameter from such a simple dataset. And the results are strongly affected by the number of layers, CNN model only has 8 layers, VGG16 has 16 layers and ResNet50 has 50 layers. With the vertical comparison, we

found that model barely learned from architecture in the case of CNN and transfer learning base on VGG16 model, because the difference among training accuracy, validation accuracy and testing accuracy are too small. And the model has over-fitted on the training set in the case of transfer learning base on ResNet50 model when classifying the 234 sub-classes, because the validation accuracy and testing accuracy are obviously smaller than the training accuracy. Based on the analysis, we believe that expanding the size of dataset and increasing the training layer will be very helpful for the problem of garbage classification.

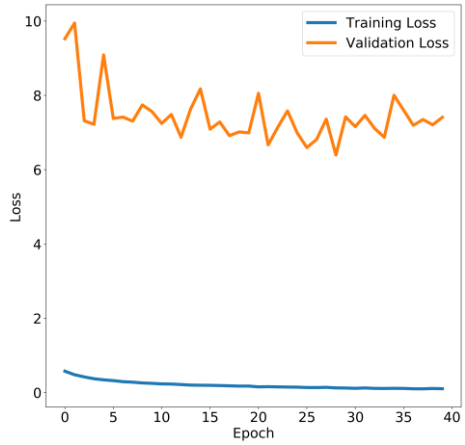

Fig. 7. Loss of the ResNet50 model for 234 sub-classes.

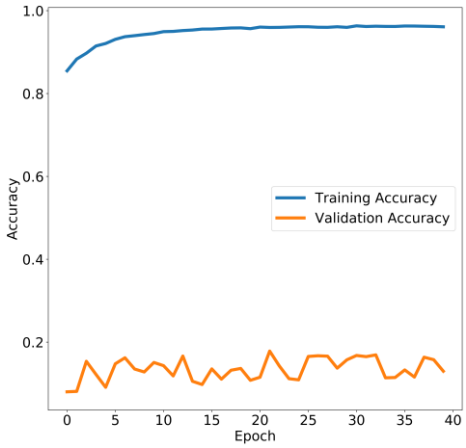

Fig. 8. Accuracy of the ResNet50 model for 234 sub-classes.

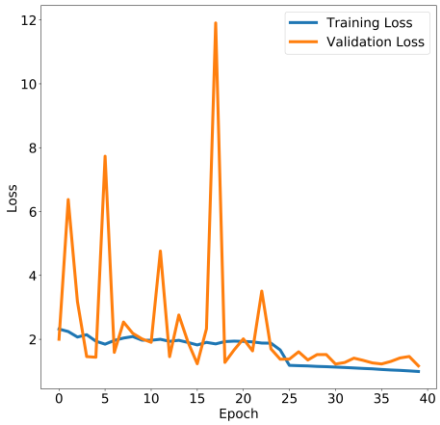

Fig. 9. Loss of the ResNet50 model for 4 main categories.

To further explain our results, we show the training/validation loss for 234 sub-classes of ResNet50 in Fig. 7. As we can tell from Fig. 7, both the training and validation losses keep decreasing with the epochs. However, the change of the training loss is more smooth because we are fitting on the training data.

We show the training/validation accuracy for 234 sub-classes of ResNet50 in Fig. 8. Similarly, we can observe that both the training and validation accuracies are increasing.

We would also observe the situation for 4 main categories. We show the training/validation loss for 4 main categories of ResNet50 in Fig. 9.

We show the training/validation accuracy for 4 main categories of ResNet50 in Fig. 10.

Fig. 10. Accuracy of the ResNet50 model for 4 main categories.

V. Future Directions

Our study is preliminary and has several future improvement directions.

A. Larger Training Dataset

Even though we aim to build a large-scale dataset for garbage classification, both the size and quality are far from satisfactory. More effort should be put for image collection and cleaning. However, our work contributes a good start for this time-consuming long-term work. The following research can also combine our dataset with previous garbage datasets and build a larger dataset, with reshaping the image sizes and relabeling. More data sources should be investigated, including those collected by human beings, e.g., crowdsourcing platforms.

B. Better Models

Better and lighter models are needed, especially when we want to deploy a garbage classification application on our smartphones, where computing and storage resources are limited.

Instead of training and inferencing locally, new paradigms including cloud computing and edge computing can also be explored and applied, which would lower the requirements for mobile devices.

C. System Integration

An isolated system for garbage classification is not the end of the story. For better regulation and recycling of wastes, the integration of garbage classification and other systems are necessary. For example, when some garbage is classified as recyclable, more instructions for proper actions

could be given and the information can be collected for environmental protection administration to give rewards for the users, e.g., cash or credits.

VI. Conclusion

In this study, we write a crawler and use it to capture images from the Internet for garbage classification. We also investigate the possibility of implementing deep learning models for automatic waste classification. We implement three different models in this research including a CNN model and VGG16 and ResNet50 models, which are pre-trained on the ImageNet dataset. We find the superiority of the ResNet50, which we attribute to the contribution of the pre-trained parameters and the complex structure with deeper layers. We plan to release our dataset for future research and expand our dataset for achieving a better classification accuracy. We also pointed out several future research directions that would inspire the following research.

Conflict of Interest

The author claims there is no conflict of interest.

Author Contributions

Mr. Yunyi Liao proposed the idea, prepared and tested the programming code, conducted the data analysis, and wrote the paper.

References

[1] M. H. Zhou, S. L. Shen, Y. S. Xu *et al*., "New policy and implementation of municipal solid waste classification in Shanghai, China," *International Journal of Environmental Research and Public Health*, vol. 16, no. 17, p. 3099, 2019.

[2] J. Deng, W. Dong, R. Socher *et al.*, "Imagenet: A large-scale hierarchical image database," in *Proc. 2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248-255.

[3] V. Ambati, S. Vogel, and J. G. Carbonell, *Active Learning and Crowd-Sourcing for Machine Translation*, 2010.

[4] W. Jiang, J. Lian, M. Shen et al., "A multi-period analysis of taxi drivers' behaviors based on GPS trajectories," in *Proc. 2017 IEEE 20th International Conference on Intelligent Transportation Systems*, 2017, pp. 1-6.

[5] W. Jiang and L. Zhang, "The impact of the transportation network companies on the taxi industry: Evidence from Beijing's GPS taxi trajectory data," *IEEE Access*, vol. 6, pp. 12438-12450, 2018.

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

[7] W. Jiang and L. Zhang, "Edge-SiamNet and Edge-TripleNet: New deep learning models for handwritten numeral recognition," *IEICE Transactions on Information and Systems*, 2020, vol. 103, no. 3, pp. 720-723.

[8] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proc. BMVC*, 2015, vol. 1, no. 3, p. 6.

[9] W. Jiang, "Applications of deep learning in stock market prediction: recent progress," arXiv preprint arXiv:2003.01859, 2020.

[10] W. Jiang and L. Zhang, "Geospatial data to images: A deep-learning framework for traffic forecasting," *Tsinghua Science and Technology*, vol. 24, no. 1, pp. 52-64, 2018.

[11] C. Srinilta and S. Kanharattanachai, "Municipal solid waste segregation with CNN," in *Proc. 2019 5th International Conference on Engineering, Applied Sciences and Technology*, 2019, pp. 1-4.

[12] M. Yang and G. Thung, "Classification of trash for recyclability status," CS229 Project Report, 2016.

[13] M. S. Rad, A. von Kaenel, A. Droux *et al*., "A computer vision system to localize and classify wastes on the streets," in *Proc. International Conference on Computer Vision Systems*, Springer, Cham, 2017, pp. 195-204.

[14] L. J. C. Brinez, A. Rengifo, and M. Escobar, "Automatic waste classification using computer vision as an application in colombian high schools," in *Proc. 6th Latin-American Conference on Networked and Electronic Media*, 2015, pp. 1-5.

[15] G. E. Sakr, M. Mokbel, A. Darwich *et al*., "Comparing deep learning and support vector machines for autonomous waste sorting," in *Proc.* 2016 *IEEE International Multidisciplinary Conference on Engineering Technology*, 2016, pp. 207-212.

[16] Y. Chu, C. Huang, X. Xie *et al*., "Multilayer hybrid deep-learning method for waste classification and recycling," *Computational Intelligence and Neuroscience*, 2018.

[17] G. Mittal, K. B. Yagnik, M. Garg *et al*., "Spotgarbage: Smartphone app to detect garbage using deep learning," in *Proc. the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016, pp. 940-945.

[18] P. Zhang, Q. Zhao, J. Gao *et al*., "Urban street cleanliness assessment using mobile edge computing and deep learning," *IEEE Access*, vol. 7, pp. 63550-63563, 2019.

[19] W. Jiang, "Time series classification: Nearest neighbor versus deep learning models," *SN Applied Sciences*, vol. 2, no. 4, pp. 1-17, 2020.

[20] K. He, X. Zhang, S. Ren *et al*., "Deep residual learning for image recognition," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778.

[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

Yunyi Liao was born in Kunming, Yunnan, China on February 25, 1999. He is a senior student majoring in software engineering from the School of Information Science and Technology, Xiamen University Malaysia, Jalan Sunsuria, Bandar Sunsuria, 43900 Sepang, Selangor Darul Ehsan, Malaysia.

He did research in Tsinghua University, China during September 2019. And he is interested in AI, machine learning, deep learning and website contribution.