

The Design and Implementation of an Intelligent Paid Waste Recycling Bin based on Computer Vision

Boyang Ren*

Department of Engineering, University of Birmingham, Birmingham, B15 2TT UK

*Corresponding author: Bxr471@student.bham.ac.uk

Abstract. In the context of rapid urbanisation leading to a significant increase in household waste and the low efficiency of traditional recycling methods, this study has designed an intelligent paid rubbish recycling box integrated with computer vision technology. The system uses OpenMV for image acquisition, employs convolutional neural networks (CNN) to achieve automatic rubbish classification, and completes payment through 3D face recognition. The experiments show that CNN achieves a 98.5% recognition accuracy on the MNIST dataset, and data augmentation techniques further enhance the model's generalisation ability. The 3D face recognition algorithm has an accuracy rate of 98.5% which is superior to DeepFace's 97.3%, and the combination of rigid region matching and exclusion algorithm can balance the recognition speed and accuracy, reducing 65.1% of the computational load. The point reward mechanism of the recycling box effectively promotes user participation. This study not only improves the efficiency and classification accuracy of rubbish recycling but also lays a foundation for intelligent rubbish management and the development of the circular economy.

Keywords: OpenMV; CNN; DeepFace.

1. Introduction

Urbanization has intensified around the world, leading to a rapid increase in urban household waste, with the treatment and recycling of waste becoming increasingly problematic. Traditional recycling systems are inefficient and simply do not meet the needs of modern society for sustainability and efficient recycling. Thus, designing an intelligent paid waste recycling box is a salient way to address these problems. This apparatus incorporates advanced computer vision technology providing capabilities such as automatic identification, classification, and billing of waste, which can improve efficiency and sustainability in recycling waste. Advancements in computer vision technology have provided the technical foundation for designing an intelligent paid waste recycling bin. Conversely, this research will utilize image recognition technology to provide automatic identification and classification of waste bin according to type, which can automate the corresponding payment for waste. The introduction of such technology would not only save labor cost, and the probability of human error during classification of waste, but it should improve the rate of waste recycling and utilization of organizational resources. In addition, the deployment of intelligent paid waste recycling bins also has considerable social and economic implications, as it can galvanize the participation of the public in waste recycling behaviour, raise the awareness of residents to sustainable household waste outcomes, and generate revenue for waste treatment companies. Besides such benefits, data collected from the intelligent paid garbage recycling bin can be analyzed for waste generation and waste recycling that can provide better insight for authorities to establish sound waste policies [1]. Intelligent waste management is more widely accepted in developed countries like Japan and Germany, where intelligent public waste recycling is embedded in urban life and citizens have been subject to intelligent identification technologies, to obey garbage recycling requirements and enforce waste reduction behaviours as a function of their existence. Certainly, automatic recycling boxes incorporating image recognition technologies have also begun to emerge in select areas of Japan, where different waste types can be automatically sorted, in support of efficient waste recycling and utility practices [2]. In China, the technological developments and social advancements have also matured deliberation of intelligent automated waste recycling bin implementation. At present, some

domestic research institutions and enterprises are committed to developing garbage classification technology based on computer vision, aiming to improve the accuracy and efficiency of garbage classification through advanced identification algorithms. In recent years, the number of projects using deep learning technology has increased significantly. These projects accurately identify garbage types by training large-scale neural network models [3]. In addition, research on user authentication, payment system integration and user experience optimization is also constantly advancing. By applying technologies such as facial recognition and QR code scanning on the recycling box, combined with mobile payment, the user's operation process is simplified and the convenience of using the recycling box is improved [4].

In a word, at home and abroad, some achievements have been made in the design and implementation of intelligent paid garbage collection boxes. However, there are still many challenges and areas for improvement, especially in terms of improving identification accuracy, improving system stability and optimizing user interaction experience. In the future, with the further development of technologies such as artificial intelligence and the Internet of Things, intelligent paid garbage collection boxes will become more intelligent and convenient. The intelligent trash can system can be divided into four main parts: image acquisition, data processing, executive control and face recognition. OpenMV camera is responsible for capturing garbage images. After CNN processes the image, it controls the servo rotation and rotates the trash can to the corresponding classification slot. Finally, it performs login and payment through face recognition. You can earn points every time you use them, which can be exchanged for money or deducted.

2. Data and Method

2.1. OpenMV

OpenMV is an open source development platform based on machine vision that could be used for the fast development and deployment of machine vision applications. The OpenMV platform consists of a single-chip computer module with an integrated camera or eyeball and a programming environment to support programming with the Python scripting language which allows programmers to write programs for various functions including image processing, video capture and data collection and analysis. OpenMV provides some of the best characteristics of development platforms, which are ease of use and flexibility. Any novice user in the field of computer vision can quickly get started and develop many intelligent vision projects. The key benefit of the OpenMV platform is the exceptionally strong image processing ability and great algorithm library. Using the built-in OpenMV IDE (Integrated Development Environment), users can run code directly on the device and get real-time feedback to see the image processing results and get faster product development cycles. OpenMV also supports all types of visual recognition technologies such as object detection, color recognition and face detection, which allow the platform to meet the basic to advanced needs of almost all types of computer vision. In practical application, OpenMV is used fundamentally for the fast development of vision applications in areas such as robot navigation, automatic control systems, and home security monitoring due to efficient small form factor and low power consumption. In this study OpenMV is used as one of the core technologies to be able to realize the image collection functions in intelligent pay-as-you-throw rubbish bins and demonstrate its capability to work reliably in complex situations[5].

2.2. CNN Network mode

CNNs and classical neural networks are quite different in image area. The CNN structure allows the network to automatically extract features of the image and CNN works better than classical neural networks in most vision tasks. First of all, in terms of feature extraction ability, CNN uses the convolutional layer to extract features at the local receptive field. This allows the CNN to learn low-level features automatically such as edges and textures in the image and then further extracts more essential information from the data using the pool layer [6]. The pyramidal feature extraction form

provides additional representation ability of multidimensional image processes with better generalization ability when working with complex processes. Additionally, add in parameter sharing mechanism here. Different neurons in the same layer share weights, which means that each neuron only focuses on part of the image area, greatly reducing the number of parameters [7]. In contrast, traditional neural networks usually require a large number of parameters to represent each dimension of input data, which not only increases the computing load, but also easily leads to overfitting [7]. Finally, there is also the ability of parallel processing. Due to the structural characteristics of CNN, they can process different parts of the image in parallel, which makes CNN more efficient in processing large-scale image data.

Regarding the performance comparison between CNN and traditional neural networks (such as multi-layer perceptron, MLP) in image classification tasks, researchers conducted a detailed simulation based on the MNIST handwritten digit dataset (containing 60,000 training samples and 10,000 test samples). The core objective of the experiment was to verify the advantages of CNN in feature automatic extraction and classification accuracy, and to highlight the performance improvement brought about by its structural innovation through comparison with the traditional neural network MLP. The experiment first constructed a simple CNN model, which consisted of an input layer, two convolutional layers and two fully connected layers. The convolutional layers were used to extract spatial features, while the fully connected layers were used to map the features to the classification space. The entire model was implemented using the TensorFlow framework. Specific training parameters included: the initial weights were initialised using a truncated random normal distribution, the learning rate was set to 0.1, the loss function used was cross-entropy, and the weights were optimised using the backpropagation algorithm. The number of iterations was 6,000. During the training process, the model autonomously learned features through the parameter sharing mechanism of the convolutional layers (reducing the risk of overfitting) and the spatial down-sampling of the pooling layers (enhancing scale invariance) without the bottleneck of manual feature extraction in traditional methods. The two figures presented below respectively reveal the core rules of CNN in terms of training dynamics and data scale sensitivity. Figure 1 visually presents the performance evolution curve of the CNN model based on the MNIST dataset after 6,000 iterations: In the early training stage (0-1,000 iterations), the accuracy jumps from a random level to over 90%, highlighting the ability of the convolutional layer to quickly capture the underlying features of handwritten digits; in the middle stage (1,000-3,000 iterations), it enters a stable ascending period, and the fully connected layer gradually learns the spatial combination relationship of features; in the later stage (>3,000 iterations), it stabilizes at a high plateau of 98.5%, confirming the saturated performance of hierarchical feature extraction.

Table 1 presents the experimental results on the flowers dataset using the ResNet-50 model, which clearly demonstrate how the scale of training samples significantly affects the training dynamics and performance metrics of the CNN, including the number of convergence iterations, convergence loss values, training accuracy, and test accuracy. This reveals the profound effect of data volume on the generalisation ability of the model. The experiment selected four sample scale scenarios: 1500 samples (random reduction), 2500 samples (further reduction), 3320 samples (original dataset), and 13280 samples (expanded from the original dataset through data augmentation techniques). The experimental results show that an increase in sample size generally improves the model's generalisation ability on the test set, but the convergence behaviour and fitting degree also exhibit interesting changes during the training process. Specifically, as the sample number increases from 1500 to 3320 (without augmentation), the number of convergence iterations increases from 20,000 to 100,000, indicating that a larger-scale dataset requires more training rounds to converge stably. Meanwhile, the convergence loss value remains at a low level (0 to 0.3), the training accuracy decreases slightly from 99.67% to 97.40%, which may be due to the increased complexity of the data making it more difficult for the model to fit perfectly, while the test accuracy increases from 79.83% to 87.75%, clearly demonstrating that increasing the sample size can effectively reduce the risk of overfitting and enhance the model's ability to recognize unknown data. However, when the sample

size is expanded to 13280 through data augmentation techniques, the number of convergence iterations unexpectedly decreases to 55,000 which is much lower than the original 3320 samples' 100,000 iterations, the convergence loss value remains stable at 0.3, and the training accuracy rebounds to 99.26%, while the test accuracy reaches 88.75% (the best among all configurations). This highlights that data augmentation not only expands the dataset by introducing noise but also accelerates training convergence and improves generalisation performance by simulating real data variations.

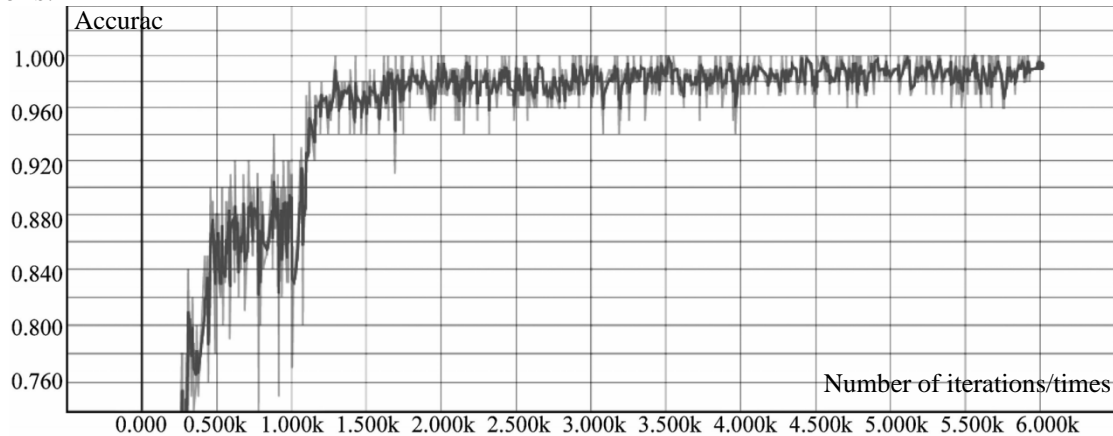


Fig. 1 The accuracy rate changes with the number of iterations.

Table 1. Influence of data sample size on CNN [8]

Number of training samples	Convergence iteration count / 10,000 times	Converging loss value	Accuracy Rate (Training, %)
1500	2	0	99.67
2500	7	0.3	96.92
3320	10	0.2	97.40

2.3. Face Recognition

When discussing the differences between the three-dimensional face recognition algorithm and DeepFace, we focused on their performances in terms of recognition accuracy, processing speed, and application scenarios. Through a detailed comparison of the two algorithms, we can better understand their respective advantages and limitations.

The three-dimensional face recognition technology mainly relies on the three-dimensional structure information of the face, which makes it more robust in cases of changes in lighting and expressions. While DeepFace is a face recognition method based on deep learning, it obtains a model that can effectively recognise facial features through extensive data training [9]. Table 2 presents a comparison chart of 3D face recognition and Deep Face in terms of accuracy, processing speed and application scenarios.

Table 2. Comparison of recognition accuracy, processing speed and application scenarios

Evaluation indicators	Three-dimensional face recognition algorithm	DeepFace	Evaluation indicators
Recognition accuracy (%)	98.5	97.3	Recognition accuracy (%)
Processing speed (ms)	120	80	Processing speed (ms)
Application scenario	Safety monitoring, identity verification	Social media, mobile payment	Application scenario

From the table, it can be seen that the 3D face recognition algorithm has slightly higher recognition accuracy than DeepFace, which is attributed to its utilisation of the three-dimensional structure of the face. However, in terms of processing speed, DeepFace performs better, possibly because its

algorithm design is more efficient and it can complete calculations quickly when dealing with large-scale datasets.

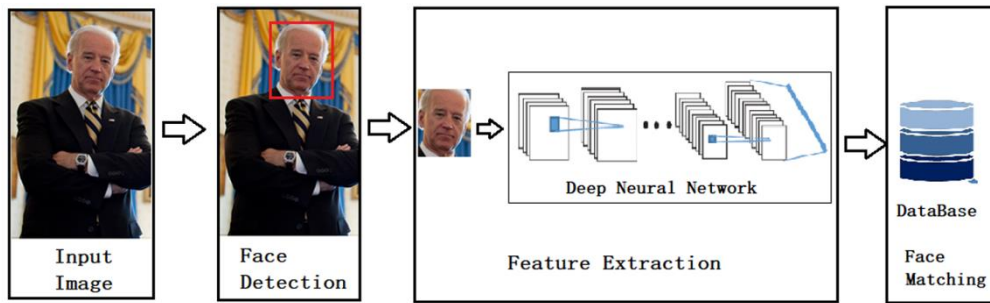


Fig. 2 Facial recognition system process with deep learning model.

Figure 2 fully illustrates the workflow of the facial recognition system based on deep learning: Firstly, data is collected through the camera, obtaining facial images (including 2D, 3D or thermal imaging types); then the system uses algorithms to precisely locate the face area and mark key feature points in the image; subsequently, the deep learning model (such as convolutional neural network) converts the face into high-dimensional feature vectors, quantifying biological features such as eye distance and cheek shape; finally, by comparing the feature vectors with the records in the database, the similarity scoring mechanism (such as cosine distance) is used to determine the identity[10].

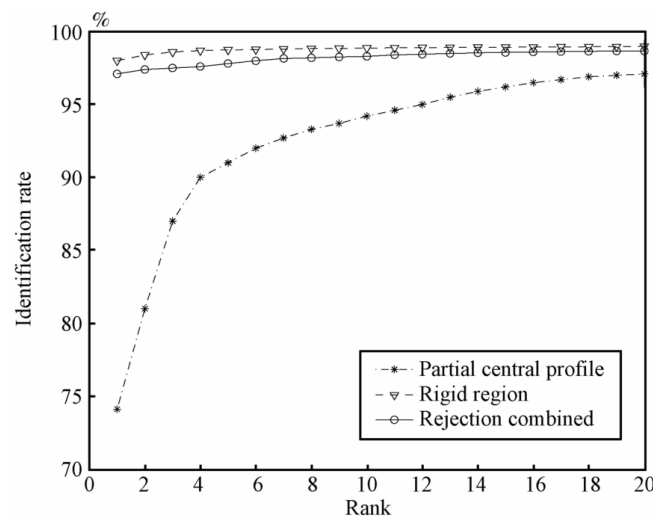


Fig. 3 CMC curve [11].

Figure 3 presents a comparison chart of the cumulative matching characteristics (CMC) curves of different matching methods obtained in the identity recognition experiment. This figure is a key visualisation tool for evaluating the performance of the recognition system, with the horizontal axis (X-axis) representing the number of the most similar candidate objects returned by the system (Rank-n), and the vertical axis (Y-axis) representing the cumulative probability or recognition rate of correctly identifying the face under these Rank-n conditions. Specifically, the three curves depicted in the figure represent: the curve representing "side profile line matching" shows the weakest performance, with the Rank-1 recognition rate significantly lower than 75%, indicating that relying solely on profile line features, the system's probability of correctly identifying the correct face in the most prioritized single result is less than three quarters, and as the number of returned candidate objects increases, such as Rank-5, Rank-10, the improvement in recognition rate is relatively flat, suggesting that the representational ability of this feature is limited and the ability to distinguish similar individuals is insufficient; in sharp contrast is the curve representing "overall rigid region matching", which demonstrates higher recognition performance, with the Rank-1 recognition rate reaching 97.6%, indicating that directly using rigid regions of the face such as the nose and the area between the eyes to perform precise matching can achieve very high first-round recognition accuracy. The curve starts at a high point at Rank-1 and then rises gradually, meaning that when the system

fails to recognize in the first round, even if the candidate range is expanded, the improvement in recognition rate is relatively limited; the most core curve in the figure is the rigid region matching curve combined with the exclusion algorithm, with a Rank-1 recognition rate of 97.1%, although it is slightly lower by 0.5 percentage points than the overall rigid region matching, this is mainly due to the exclusion algorithm possibly mistakenly excluding the correct model in a very few cases during the preprocessing stage. However, this curve clearly reveals the core advantage of this method - with only a slight sacrifice in recognition accuracy, it quickly eliminates approximately 65.1% of the library models through side profile line matching, significantly reducing the computational load of the subsequent time-consuming high-precision rigid region matching, thereby achieving a qualitative leap in recognition efficiency. The overall trend of the curve is similar to that of the overall rigid region matching, and it also stabilises after Rank-1, proving the effectiveness of the exclusion strategy in maintaining the overall recognition capability of the system. This figure intuitively confirms the superiority of rapid exclusion and precise matching in balancing speed and accuracy in the three-dimensional face recognition task.

3. Results and Discussion

The accuracy curve of the CNN model on the MNIST dataset shows that it quickly reaches over 90% accuracy within 1,000 iterations and stabilises at 98.5% after 3,000 iterations. This highlights the advantage of CNN over traditional neural networks in automatic feature learning, characterised by higher classification accuracy and lower parameter complexity. From Table 2, these results collectively indicate that in CNN training, sample size is a key factor in balancing training efficiency and model accuracy. When data volume is insufficient, the model fits the training set quickly but performs poorly on the test set, while simply increasing the sample size improves generalisation but prolongs training time. However, combining data augmentation techniques can efficiently overcome this bottleneck, achieving a better balance between accuracy and efficiency, providing empirical guidance for practical applications such as data set construction in resource-constrained scenarios. In face recognition, the accuracy rate of the three-dimensional algorithm reaches 98.5%, which is higher than DeepFace's 97.3%, but DeepFace has a faster processing speed (80 milliseconds compared to 120 milliseconds). The cumulative matching characteristic (CMC) curve further indicates that combining rigid region matching with exclusion algorithms can balance recognition speed and accuracy, reducing the computational load by 65.1% while maintaining high precision.

However, this study has certain limitations: the current dataset lacks diversity, and the CNN model may have difficulty in recognising rare or deformed rubbish types; environmental factors such as lighting and noise can affect the image acquisition and 3D face recognition performance of OpenMV; the payment and points redemption system also needs to be optimised to improve operational smoothness. In the future, the rubbish dataset can be expanded by synthesizing data and sampling from real scenarios to improve the recognition of rare rubbish; integrating attention mechanisms into CNN can enhance the feature extraction ability for complex rubbish, and developing lightweight models suitable for edge computing can reduce processing latency; combining IoT technology to achieve real-time monitoring of rubbish bin capacity and logistics scheduling, and incorporating multi-modal recognition such as voice interaction and infrared sensing to enhance user convenience; exploring energy-saving designs such as solar power, and collaborating with rubbish processing plants to optimize the recycling workflow, promoting the development of circular economy practices.

4. Conclusion

This study designed an intelligent paid rubbish recycling box integrating computer vision technology, which realised the functions of automatic rubbish recognition, classification and billing. The CNN model achieved high recognition accuracy on the MNIST dataset, stabilising at 98.5% after 3000 iterations. Data augmentation techniques further enhanced the generalisation ability of the

model. The three-dimensional face recognition algorithm had an accuracy rate of 98.5%, which was superior to the DeepFace algorithm. The combination of rigid region matching and the exclusion algorithm could effectively balance the recognition speed and accuracy. The system also promoted user participation through an integrated reward mechanism. This research not only improved the efficiency and classification accuracy of rubbish recycling but also enhanced public environmental awareness, laying a foundation for intelligent rubbish management and the development of the circular economy.

References

- [1] Liu Xudong, Zhang Yanfen, Zhao Junjie. Design of an Intelligent Waste Classification System Based on ROS and Computer Vision [J]. *Electronic Making*, 2020, (23): 45-46 + 60. DOI: 10.16589/j.cnki.cn11-3571/tn.2020.23.014.
- [2] Gao Shiya, Hu Xiaoying, Dong Jingfeng. Design of a Comprehensive Recycling and Disposal System for Household Waste and an Intelligent Waste Sorting Bin [J]. *Logistics Technology*, 2021, (07): 110-114 + 135.
- [3] Chen Jiaxin, Zhong Zixin. Research on the Design of a Home Intelligent Trash Can Based on Machine Vision [J]. *Network Security Technology and Application*, 2021, (07): 152-153.
- [4] Zhou Xuhua, Wu Yimei, Huang Xiaoru, Weng Jiaxin. Design of an Intelligent Garbage Sorting Robot Based on Machine Vision [J]. *Manufacturing Technology*, 2021, (08): 22-24 + 42.
- [5] Abdelkader, I., Y. El-Sonbaty, and M. El-Habrouk, Openmv: A python powered, extensible machine vision camera. arXiv preprint arXiv:1711.10464, 2017.
- [6] Chauhan, R., K.K. Ghanshala, and R.C. Joshi. Convolutional neural network (CNN) for image detection and recognition. in 2018 first international conference on secure cyber computing and communication (ICSCCC). 2018. IEEE.
- [7] Deng Wanyu, et al., Research on Ultra-Fast Learning Methods of Neural Networks. *Journal of Computer Science*, 2010. 33(2): p. 279-287.
- [8] Dong Ziyuan and Han Weiguang, Garbage Image Classification Algorithm Based on Convolutional Neural Network. *Application of Computer Systems*, 2020. 29(8): p. 199-204.
- [9] Wang, M. and W. Deng, Deep face recognition: A survey. *Neurocomputing*, 2021. 429: p. 215-244.
- [10] Qinjun, L., et al., Facial recognition technology: a comprehensive overview. *Academic Journal of Computing & Information Science*, 2023. 6(7): p. 15-26.
- [11] Li Xiaoli and Da Feipeng, a Fast 3D Face Recognition Method Based on Exclusion Algorithm. *Acta Automatica Sinica*, 2010. 36(1): p. 153-158.