Resources, Conservation & Recycling 164 (2021) 105132

Contents lists available at ScienceDirect

Resources, Conservation & Recycling

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

Full length article

Recycling waste classification using optimized convolutional neural network

Wei-Lung Mao[a], Wei-Chun Chen[b], Chien-Tsung Wang[c], Yu-Hao Lin[b,*]

[a] Department of Electrical Engineering, National Yunlin University of Science and Technology, 123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan
[b] Bachelor Program in Interdisciplinary Studies, National Yunlin University of Science and Technology, 123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan
[c] Department of Chemical and Materials Engineering, National Yunlin University of Science and Technology, 123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan

ARTICLE INFO

Keywords:
Convolutional neural network (CNN)
Genetic algorithm (GA)
DenseNet
Waste classification
Image recognition
Recycling

ABSTRACT

An automatic classification robot based on effective image recognition could help reduce huge labors of recycling tasks. Convolutional neural network (CNN) model, such as DenseNet121, improved the traditional image recognition technology and was the currently dominant approach to image recognition. A famous benchmark dataset, i.e., TrashNet, comprised of a total of 2527 images with six different waste categories was used to evaluate the CNNs' performance. To enhance the accuracy of waste classification driven by CNNs, the data augmentation method could be adopted to do so, but fine-tuning optimally hyper-parameters of CNN's fully-connected-layer was never used. Therefore, besides data augmentation, this study aims to utilize a genetic algorithm (GA) to optimize the fully-connected-layer of DenseNet121 for improving the classification accuracy of DenseNet121 on TrashNet and proposes the optimized DenseNet121.

The results show that the optimized DenseNet121 achieved the highest accuracy of 99.6%, when compared with other studies' CNNs. The data augmentation could perform higher efficiency on accuracy improvement of image classification than optimizing fully-connected-layer of DenseNet121 for TrashNet. To replace the function of the original classifier of DenseNet121 with fully-connected-layer can improve DenseNet121's performance. The optimized DenseNet121 further improved the accuracy and demonstrated the efficiency of using GA to optimize the neuron number and the dropout rate of fully-connected-layer. Gradient-weighted class activation mapping helped highlight the coarse features of the waste image and provide additional insight into the explainability of optimized DenseNet121.

1. Introduction

Recycling tasks is the correct way to reduce waste production, mitigate the environment and improve the whole nation's economy (Elagroudy et al., 2016; Wang et al., 2019). The efficiency and quality of the recycling task is highly dependent on the purity and accuracy of the sorted raw materials (Tachwali et al., 2007). However, recycling tasks usually required a lot of labor cost (Seike et al., 2018), and computer vision and deep learning (DL) techniques help in the automatic detection and classification of waste types for recycling tasks (Vo et al., 2019). DL is one of artificial intelligence section, which could solve visual recognition and classification, and convolutional neural network (CNN) is one of the DL architecture currently referred to as the dominant approach to image recognition. Begin with modified national institute of standards and technology (MNIST) dataset classification (LeCun et al., 1998), the several famous CNN models, such as AlexNet (Krizhevsky et al., 2012), VGG (Simonyan and Zisserman, 2014), ResNet (He et al., 2016), and DenseNet (Huang et al., 2017), were adopted in the previous experimental studies. The derived transfer learning (TL) focuses on transferring the learned parameters of CNN model to the new one, and so this facilitates the new CNN model training (Pan and Yang, 2009) and improves the accuracy of the new CNNs' classification when the training dataset is limited (Chen et al., 2020; Han et al., 2018). The TL of CNNs are transferrable and applicable to traffic object detection (Zhang et al., 2018), and recyclable garbage classification (Aral et al., 2018).

Yang and Thung (2016) collected the waste image dataset with a total of 2527 images, namely TrashNet (Yang and Thung, 2016). The models, including support vector machine (SVM) with scale-invariant feature transform and ResNet50 with SVM, were used to classify the waste image in the TrashNet, and the two models achieved an accuracy of 63% and 87%, respectively (Yang and Thung, 2016). Because previous studies have been conducted to look into the comparisons of the performances with the CNN architectures on the TrashNet (Aral et al.,

* Corresponding author.
E-mail address: 11844@yahoo.com.tw (Y.-H. Lin).

https://doi.org/10.1016/j.resconrec.2020.105132
Received 4 July 2020; Received in revised form 25 August 2020; Accepted 25 August 2020
Available online 08 September 2020

2018; Bircanoğlu et al., 2018; Meng and Chu, 2020), TrashNet became a public benchmark for waste classification. Besides TrashNet, few waste datasets, such as TACO, AquaTrash and VN-trash, were established (Panwar et al., 2020; Vo et al., 2019), and they have some shortcomings such relatively small amount or waste in a specific environment, also the dataset was not open source. In addition, the study used ResNext as a base CNN model combining with several improvements to classify automatically inorganic and medical wastes (Vo et al., 2019). Garbage sorting devices equipped with a CNN-based algorithm and numerical sensors can achieve waste classification accuracies around 90% (Chu et al., 2018; Mittal et al., 2016). Nowakowski and Pamuła used fast region-based CNN to detect the category and size of the e-waste equipment in the images (Nowakowski and Pamuła, 2020). However, the accuracy of waste classification still needs to be improved, even though the existing waste classification driven by CNNs could work efficiently (Huang et al., 2020).

Basically speaking, CNN is composed of a convolutional layer (feature extraction layer) and fully-connected-layer (classifier). The fully-connected-layer consists of huge neuron numbers, and dropout rates could classify effectively. However, this also causes the disadvantage of low computing efficiency and overfitting (Liu et al., 2019). Recently, several architectures, such as SVM, global average pooling, convolutional layers, and deconvolutional layers, have been used as classifiers instead of the fully-connected layers in some deep CNNs. The study directly outputs the spatial average of the feature maps to replace the traditional classification layer on CNN for reducing overfitting (Lin et al., 2013).

In addition, both data augmentation and optimally fine-tuned hyper-parameters of CNN's fully-connected-layer could improve the efficiency of CNNs (Al-Hyari and Areibi, 2017; Frid-Adar et al., 2018). Data augmentation methods mainly include an affine transformation of images, white-box methods, and black-box methods. For example, random vertical and horizontal flip in the method of an affine transformation of images could increase the diversity of images in the dataset for training CNN (Tong et al., 2019). The hyper-parameters of the fully-connected-layer of CNNs are concerned with the number of neurons in the fully-connected-layer. The optimization operators of Genetic Algorithm (GA) could effectively resolve optimization problems such as water quality management, flexible job shop scheduling, and waste generation forecasting (Abdallah et al., 2020; Chen et al., 2019; Piroozfard et al., 2018). The hyper-parameters of a deep CNN can be encoded as a chromosome of GA, but the application of a GA enables both the optimization of architectures and the hyper-parameters of CNNs to occur only in few studies (Rattanavorragant and Jewajinda, 2019). Based on a similar concept, several studies used particle swarm optimized algorithm to optimize the parameters of SVM classifier to better obtain classification performance (Navaneeth and Suchetha, 2019; Raj and Ray, 2017). To date, a few studies have been devoted to fine-tuning optimally hyper-parameters of CNN's fully-connected-layer using optimizer to improve the efficiency of CNNs, and none of these studies have adopted the CNN with the optimized hyper-parameters of fully-connected-layer in the field of waste classification. In addition, no literature was found that also identified the dominant features of the waste images extracted by a CNN.

Therefore, the aim and novelty of this study is to classify waste using an optimized CNN efficiently. First, the famous CNN model, i.e., DenseNet121, was applied to the image of the waste of the TrashNet dataset. Original DenseNet121 used a global average pooling and softmax classifier to replace the traditional fully-connected-layer, but this study still considers traditional fully-connected-layer as the classifier of DenseNet121 to enhance the recognition accuracy of DenseNet121. In addition to data augmentation, this study also utilized GA to optimize the fully-connected-layer of DenseNet121. This was to improve the classification accuracy of DenseNet121 and propose the optimized DenseNet121. Moreover, this study generated gradient-weighted class activation mapping (Grad-CAM) served as interpretation of the optimized DenseNet121′s prediction, and the Grad-CAM could identify visually the dominant features of the waste images extracted by optimized DenseNet121. The remainder of this paper is structured as follows. Section 2 describes the trash dataset and data augmentation setting. In this section, descriptions of a method and basic environment setting are also provided. Experiment results are shown in Section 3. The conclusions are drawn in Section 4.

2. Methodology

2.1. TrashNet dataset and DenseNet121 model

In this study, the dataset source heavily depends on the TrashNet (Yang and Thung, 2016). The TrashNet dataset consisted of a number of 2527 image divided into six categories. The categories of the dataset included cardboard, glass, metal, paper, plastic, and trash. To create a large dataset, this study did augment the original dataset as well. The dataset augmentation gave 2527 pictures of flipping horizontal, 2527 pictures of flipping vertical, and 2527 random 25° rotation then created a total of 10108 waste images. Also, this study used 2527 images and a total of 10108 images to compare the results. Moreover, the dataset was split, and each class was randomly grouped into 90% and 10% for training and testing sets, respectively.

The DenseNet core mainly consisted of a dense block (DB), transition layer, and growth rate. DenseNet was mainly formed by dense connectivity as the input for other convolutional layers, as clearly shown in Fig. 1. DenseNet had four types of architecture with different number of DB. The output of each DB was concatenated into one single input or input tensor. Detail DenseNet121 was adopted by referring to Huang's study (Huang et al., 2017). DenseNet121 had a scheme (6, 12, 24, and 16) of DB. Each DB had two sizes of a convolutional layer, and each convolutional layer had three sequences of batch normalization, ReLu, and convolutional layer. To reduce overfitting on tasks with smaller training set sizes, the original DenseNet121 used a global average pooling and softmax classifier to replace the traditional fully-connected-layer. However, this study still used two fully-connected-layers as the classifier of DenseNet121 and tested this method for enhancing the recognition accuracy of DenseNet121. The stochastic

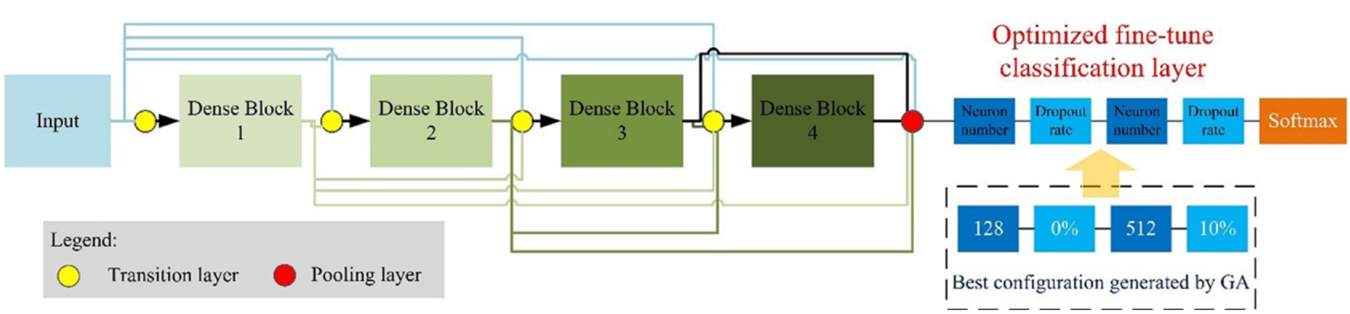

Fig. 1. Optimized DenseNet121 architecture.

gradient descent (SGD) and the Adadelta algorithm could be used to calculate the weights of CNNs (Zeiler, 2012). Our preliminary test showed that Adadelta performed much better than SGD on the weight calculation of DenseNet121. As a result, the Adadelta with a learning rate of 0.1 and decay of 0.001 was used to calculate the weights of DenseNet121-based models.

2.2. Genetic algorithm, optimized DenseNet121, and performance test

An appropriate neuron number and dropout rate of the fully-connected-layer could not only improve the effectiveness of classification but also prevent overfitting of the CNN. Different from the fully-connected-layer of original DenseNet121, this study used a classification layer for classifying objects, and GA was used to optimize the classification layer of optimized DenseNet121, as shown in Fig. 1. GA algorithm begins with a random chromosome of the population; based on the concept of the survival of the fittest, the population of GA implemented standard evolving operators, including evaluation, crossover, and mutation, to further improve populations over a generation. Chromosomes in the population represent decision variables or solutions and could be encoded as substrings of binary digits or real numbers or integers (McKinney and Lin, 1994).

To further improve the performance of DenseNet121, this study utilized GA to optimize the full-connected layer of the DenseNet121, and then the proposed optimized DenseNet121 was validated on the TrashNet. In other words, this study used GA to optimally fine-tune the hyper-parameters of the fully-connected-layer of DenseNet121, and both of the neuron number and the dropout rate of the fully-connected-layer were deemed as the main decision variables in the proposed optimized DenseNet121. Fig. 1 shows the original DenseNet121 approach to be optimized using a GA.Table 2 shows the optional code in GA and the corresponding value of the DenseNet121 hyper-parameter, including the number of neurons or the dropout rate. For example, the code of gene equals one, and the number of neurons is corresponding to 128. Same as the number of neurons, the dropout rate optional code from 0-5 is also given the real corresponding value from 0% - 50%.

The total epoch of each DenseNet121 operation was 40 epochs. Fig. 2 shows the flowchart of the optimized DenseNet121 utilized in waste image recognition. This study used a random value of chromosomes in the beginning population. The number of populations each iteration was assigned as 10. Then, the next operations of GA, including crossover, mutation, and fitness calculation, were implemented. The crossover and mutation rate in this task defined as 85% and 3%, respectively. The accuracy result of DenseNet121 was assigned as a fitness score to each chromosome. The genes would evolve in 4 generations from 0 to 5, and the number of generations was chosen to 4 generations, and it was also assigned as the end of GA operation. At the end of the GA operation, the best chromosome that indicated most satisfy neuron number and a dropout rate of the fully-connected-layer would be applied on the fine hyper-parameters of the fully-connected-layer of DenseNet121, and this study defined such architecture as the optimized DenseNet121 for waste classification of TrashNet. In addition, to visually evaluate the performance of CNNs, the confusion matrix displays the number of correct and incorrect predictions made by the CNN model in comparison with those made from the actual classifications in the test image.

2.3. Gradient-weighted class activation mapping

To visually validate where the CNN was indeed looking at the correct patterns in the image, Selvaraju et al. proposed Grad-CAM to produce a coarse localization map highlighting the important regions in the image for predicting the concept (Selvaraju et al., 2017). Grad-CAM generally uses the gradient information flowing into the last convolutional layer of the CNN to understand the importance of each neuron for a decision of interest, and the coarse heatmap of input image generated by Grad-CAM can highlight the features of the images of great importace for predicting a particular abnormality. This study used the technique of Grad-CAM invented by Selvaraju's study (Selvaraju et al., 2017) to generate a coarse heatmap for waste area localization and to explained the results of the waste recognition visually by the optimized DenseNet121. Usually, the regions displayed in yellow to red colors could highlight the spatial location of the features, which activated more intensely the last convolutional layer of CNN before the classification.

All experiments were implemented in Python 3.5.2, using Keras (Chollet, 1997) to train the DensNet121-based models on GeForce RTX 2070 8GB, and they were performed on an Intel Core i7-9700K 3.60GHz, 64GB RAM with Ubuntu 18.04.3 LTS.

3. Discussion

3.1. Performance of the original and optimized DenseNet121

In this study, the DenseNet121 with two fully-connected-layers refers to a DenseNet121 equipped with a fully-connected-layer but where that layer is not optimized by GA. In the study this method was referred to as the unoptimized DenseNet121. Table 2 shows the accuracy of the original DenseNet121, unoptimized DenseNet121, and the optimized DenseNet121 when the augmentation was implemented to create a total of 10108 images. The original DenseNet121 achieved an accuracy of 98.91% when the data augmentation was implemented. Also, unoptimized DenseNet121 achieved at least 99.30% of the accuracy, and this part of the results indicated that using two fully-connected-layers as a classifier of DenseNet121 performed better than the original DenseNet121 equipped with a global average pooling and softmax classifier. However, this did not cause overfitting, as shown in Fig. 3(b), even though it had a slightly longer training time because of the larger number of neurons. After using GA to optimize the neuron number and dropout rate of the two fully-connected-layers, the optimized DenseNet121 could further enhance an accuracy of about 0.3% and so achieve 99.60% accuracy in this regard. Also, the optimized fully-connected-layer improved the performance of DenseNet121; however, this only gained accuracy improvements of 0.7% accuracy. This result demonstrated that the optimized fully-connected-layer employing an optimization algorithm, such as GA, could improve the performance of DenseNet121 regardless of the small accuracy improvements accuracy; this small increment of accuracy could be attributed to the fact that the data augmentation technique already enhanced the performance of the original DenseNet121 to achieve a relatively high accuracy of 98.91%.

As can be seen from Table 2, at the end of GA iteration, i.e., 4 generations evolved, better chromosome in each generation can be obtained. The optimized DenseNet121 that appeared at fourth-

Table 1
The option values of the genes in the GA.

Layer	Hyper-parameters	Optional code of the gene	The corresponding value of hyper-parameters
First	Neuron number	[0, 1, 2, 3, 4, 5, 6, 7, 8]	[0, 128, 256, 384, 512, 640, 768, 896, 1024]
	Dropout rate	[0, 1, 2, 3, 4, 5]	[0%, 10%, 20%, 30%, 40%, 50%]
Second	Neuron number	[0, 1, 2, 3, 4, 5, 6, 7, 8]	[0, 128, 256, 384, 512, 640, 768, 896, 1024]
	Dropout rate	[0, 1, 2, 3, 4, 5]	[0%, 10%, 20%, 30%, 40%, 50%]

Table 2
Better chromosome in each generation of the optimized CNN when the augmentation was implemented on TrashNet, and their performances.

Model	Generation	Accuracy (%)	Training time (s)	Hyperparameter of fully-connected-layer[a]			
				First layer		Second layer	
				Neuron number	Dropout rate (%)[c]	Neuron number	Dropout rate (%)
Original DenseNet121[b]	-	98.91	8309	-[b]			
Unoptimized DenseNet121[c]	-	99.30	9499	1024	30	1024	30
Optimized DenseNet121	1	99.40	5400	256	50	640	10
	2	99.40	5400	256	50	640	10
	3	99.40	5400	256	50	640	10
	4	99.60	5542	128	0	512	10

[a] : two fully-connected-layers replace the function of the original classifier of DenseNet121.
[b] : a global average pooling is performed and then a softmax classifier is attached.
[c] : the fully-connected-layers were not optimized by GA, and the neuron number and dropout rate of DenseNet121 were 1024, 30%, 1024, and 30% in the fully-connected-layer, respective.

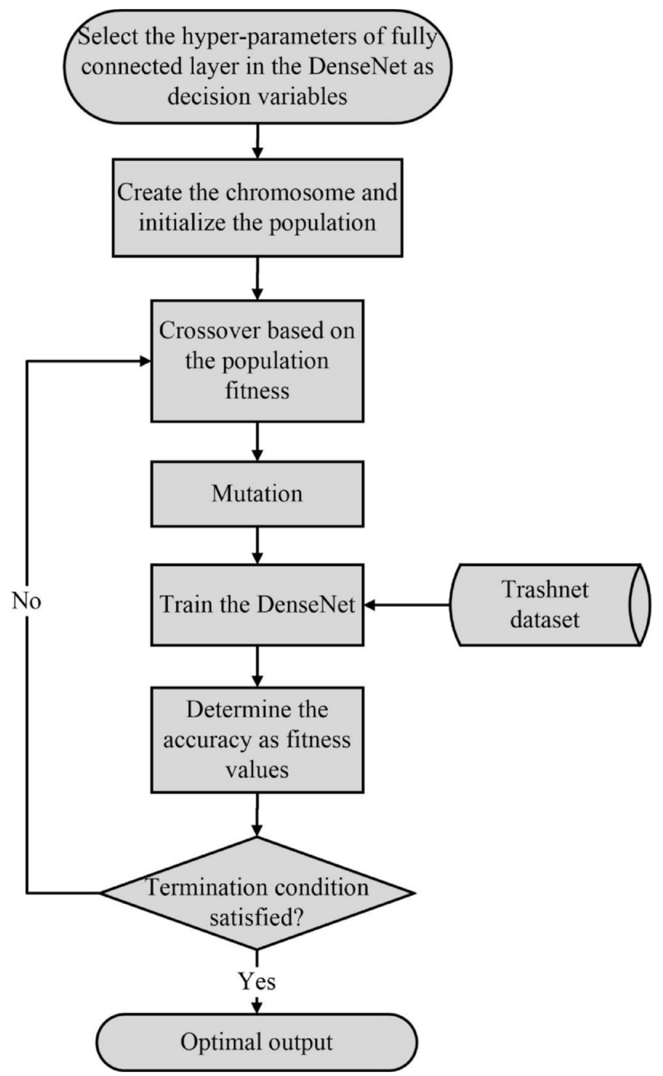

Fig. 2. A flowchart of the optimized DenseNet121 utilized in waste image recognition.

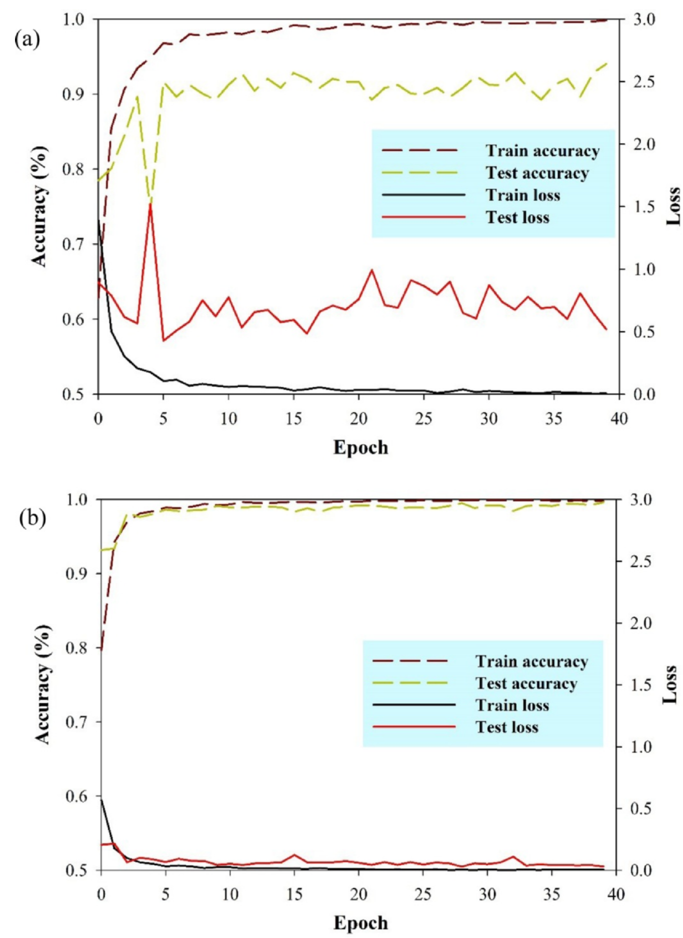

Fig. 3. The accuracy and cross-entropy loss of optimized DenseNet121 using (a) 2527 and (b) 10108 images.

generation had optimal optional code 1, 0,| 4, 1; the corresponding value of the first fully-connected-layer of the optimized DenseNet121 consisted of 128 neuron numbers with a 0% dropout rate. The second fully-connected-layer of optimized DenseNet121 consisted of 512 neurons with a 10% dropout rate. Dropout is a random procedure. This procedure could drastically improve CNN's performance through the random omission of the neuron numbers, and so help reduce the computational time of CNNs. During the evolution of 4 generations, the trend of accuracy gradually improved; training time of the optimized DenseNet121 was 5542 seconds. This time was lower than that observed from the original DenseNet121 (8309 seconds). Similarly, Table 3 shows that the training time of the unoptimized and optimized DenseNet121 were 1259 and 1058 seconds, respectively, when data augmentation was not conducted; meanwhile, the reported accuracies were 89.24 % and 94.02%, respectively. When data augmentation was conducted, training times of the unoptimized and optimized DenseNet121 were 9949 and 5542 seconds, and their accuracies were 99.3 % and 99.6%, respectively.

In terms of computational cost, the training time of the optimized

Table 3
A comparison of performance with different CNN models based on TrashNet dataset

CNN model	Accuracy (%)	Total Images	Training time (s)	Epochs
Unoptimized DenseNet121[a]	89.24	2276 train images, 251 test images[b]	1259	40
Optimized DenseNet121	94.02	2276 train images, 251 test images	1058	40
Unoptimized DenseNet121	99.30	9095 train images, 1013 test images[c]	9945	40
Optimized DenseNet121	99.60	9095 train images, 1013 test images	5542	40
Resnet50 (Meng and Chu, 2020)	95.35	9095 train images, 1013 test images	-	40
	91.40	2276 train images, 251 test images	-	40
HOG CNN (Meng and Chu, 2020)	93.56	9095 train images, 1013 test images	-	40
	81.53	2276 train images, 251 test images	-	40
Simple CNN (Meng and Chu, 2020)	93.75	9095 train images, 1013 test images	-	40
	79.49	2276 train images, 251 test images	-	40
SVM + HOG (Meng and Chu, 2020)	47.25	9095 train images, 1013 test images	-	-
	23.51	2276 train images, 251 test images	-	-
DenseNet121 (Bircanoğlu et al., 2018)	95	Verticaland horizontal flip, 15-degree rotation with fine tuning	-	200+10
RecycleNet (Bircanoğlu et al., 2018)	81	Verticaland horizontal flip, 15-degree rotation	-	200
DenseNet121 (Aral et al., 2018)	95	2527 (70% training, 17% testing, 13% valid)	-	10+100
DenseNet169 (Aral et al., 2018)	95	2527 (70% training, 17% testing, 13% valid)	-	7+120
Inception V4 (Aral et al., 2018)	94	2527 (70% training, 17% testing, 13% valid)	-	7+120
	89		-	10+200
MobileNet (Aral et al., 2018)	84	2527 (70% training, 17% testing, 13% valid)	-	10+200
SIFT + SVM (Yang and Thung, 2016)	63	1769 train images, 758 test images	-	-

[a] :the fully-connected-layers were not optimized by GA.
[b] : the data augmentation were not conducted.
[c] : the data augmentation were conducted.

DenseNet121 sets were both lower than those observed in the unoptimized DenseNet121, regardless of whether the data augmentation had been conducted or not. That was because the optimized DenseNet121 had better the hyperparameter of fully-connected-layer, including fewer neurons and proper dropout procedures. So, this reduced the computational costs. In addition, the accuracy of optimized DenseNet121 sets were better than those derived from the unoptimized DenseNet121, regardless of whether the data augmentation was conducted or not. Only accuracy improvements of 0.7% were achieved when data augmentation was conducted, this small increment could be attributed to the fact that the data augmentation already enhanced the performance of the original DenseNet121 and unoptimized DenseNet121 to achieve a relatively high accuracy. Overall, these results highlighted the efficiency of using GA to optimize the neuron numbers and the dropout rates of the fully-connected-layer. In other words, the optimized fully-connected-layers of CNNs could identify image features extracted from the convolutional layer more efficiently in this regard, and GA could be used as tools to optimally fine-tune the hyperparameters of fully-connected-layer of CNNs for improving the accuracy.

Fig. 3 shows the accuracy and cross-entropy loss in the training and test processes of the optimized DenseNet121 for 2527 and 10108 waste images. Overfitting was not deemed as an issue at a point for the optimized DenseNet121 because both the training and test losses followed the same decreasing curve before 40 epochs. In Fig. 3(a), the optimized DenseNet121 with 2527 images revealed stable results after 5 epochs and indicated that the model could started to converge and achieve 99.91% and 94.02% for training and test accuracy, respectively. Compared with the accuracy result that optimized DenseNet121 was trained and tested in 2527 images, Fig. 3(b) shows that the optimized DenseNet121 with 10108 images plateaued train and test accuracy after about 3 epochs, and indicates that the model could started to converge and achieve 99.95% and 99.60% for training and test accuracy, respectively; this part of the results shows that data augmentation could not only improve the accuracy of optimized DenseNet121 but also reduce the cross-entropy loss of optimized DenseNet121.

3.2. Performance comparison and interface of Windows Application development

This study confronts the results of optimized DenseNet121 with other CNNs for highlighting the effectiveness of optimized DenseNet121. Table 3 displays the accuracies of different CNNs on TrashNet dataset, and the different sizes of the pictures revealed different results of accuracy. The higher number of training tests was, the higher result this study could get. The original owner of the dataset was built with the SVM classifier, which only achieves 63% of accuracy. Other related studies constructed four CNNs using the TrashNet dataset (Meng and Chu, 2020), and the optimized DenseNet121 overcame other CNNs regarding the test accuracy. In comparison with other CNNs, the optimized DenseNet121 with 10108 images achieved the highest accuracy, and the misclassification rate was only 0.40% from 10108 images i.e., the four images shown in Fig. 5. Moreover, DenseNet121 with 10108 images attained an accuracy improvement of 4.6% over that of the CNN proposed by Bircanoğlu's study (Bircanoğlu et al., 2018), conducted under similar conditions. However, it gained an obvious accuracy improvement of 36.6% over the CNN proposed by Yang's study (Yang and Thung, 2016). Also, when the data augmentation technique was not implemented, and the original 2527 waste images were used to train and test DenseNet-based models, unoptimized DenseNet121 and the optimized DenseNet121 achieved 89.42 and 94.02%, respectively. This result indicated that optimizing fully-connected-layer of DenseNet121 could increase 4.78% accuracy of unoptimized DenseNet121. This highlighted the efficiency of using GA to optimize the neuron numbers and the dropout rates of the fully-connected-layer.

When the data augmentation technique was implemented, the waste images increased from a total of 2527 images to a total of 10108 images, and the accuracies of unoptimized DenseNet121 obviously increased by 10.06%; this result showed that the data augmentation technique used in this study could obtain higher efficiency and accuracy of the improvement in the image classification than that obtained from optimizing the fully-connected-layer of DenseNet121 for TrashNet data. The optimized DenseNet121 with data augmentation achieved the highest accuracy of 99.6%, compared with other studies' CNN models because, under the same condition, the study used Resnet50 performed the 95.35 % of accuracy (Meng and Chu, 2020).

To demonstrate the robustness of the proposed optimized DenseNet121, the optimized DenseNet121 also classified the waste images from a public waste dataset proposed by Kim's study (Kim et al., 2019). However, this dataset was only used in Kim's study, and a modified LeNet model was utilized to classify the images. The modified

Table 4
The Confusion matrix

PredictedReal	Cardboard	Glass	Metal	Paper	Plastic	Trash
Cardboard	162	0	0	0	0	0
Glass	0	200	0	0	**1**[a]	0
Metal	0	**1**[b]	163	0	0	0
Paper	**1** [c]	0	0	236	**1** [d]	0
Plastic	0	0	0	0	193	0
Trash	0	0	0	0	0	55

a, b, c, and d: the corresponding image were shown in Fig. 5.

LeNet model exhibited accuracy of about 95%, and the optimized DenseNet121 outperformed the modified LeNet model, achieving 100% success under similar test conditions. This result indicates that the proposed optimized DenseNet121 would display similar robustness across other waste datasets. In addition, this study developed Windows Application based on C# language for convenient use. Keras.Net was used as an interpreter from Python language to C# under Visual Studio platform. The model file of the optimized DenseNet121 was employed to predict the images. The interface of Windows Application is shown in Fig. 6. The image could be picked in the windows directory or an image from the camera could be captured.

3.3. Feature identification of waste classification

Confusion matrix summarized the number of correct and incorrect predictions about count values broken down by each class, and the column matrices present the number of real classifications in the test images and the row matrices present the number of predicted classifications made by CNN. Table 4 shows the confusion matrix result of DenseNet121 with 10108 images, and the model preformed an accuracy of 99.6% (1009/1013) in terms of a total of 1013 test images. In other words, there are four predictions in three waste classes found to be slightly inaccurate, including a glass was misclassified into the plastic class, a metal was misclassified into the glass, and two papers were misclassified into the cardboard and plastic, respectively. To explain the results derived from the confusion matrix, Grad-CAM was implemented to generate the coarse heatmap to visually highlight the important patterns observed in these six correctly predicted and four misclassified images. The coarse heatmap generated by Grad-CAM could provide a visualization for the parts of the image that are most influential for a given classification. In the coarse heatmap, the red, orange, and yellow regions were strong, medium, and weak class features, respectively. In other words, those areas with red color could be identified as locations of the image with the greatest effect on the final classification. Based on the entire confusion matrix, some correctly predicted images and the four misclassified images recognized by optimized DenseNet121 with 10108 pictures, together with their Grad-CAM images are depicted in Figs. 4 and 5, respectively.

Fig. 4 shows that the optimized DenseNet121 could accurately locate the area of waste classes and then recognize the waste classes. For example, necks are referred to as the main features of glass and plastic bottles; the optimized DenseNet121 could identify whether the object in the image was metal or trash according to the feature of the entire object; the optimized DenseNet121 could also identify whether the object in the image was metal or trash possibly according to the entire shape and color of the object; the main features of paper and cardboards classified by the optimized DenseNet121would appear in the edges and corners of objects. Such Grad-CAM results also demonstrated that the optimized DenseNet121 indeed had a good ability to identify different classes of waste.

In Fig. 5, only a metal image was misclassified as glass, and Grad-CAM indicated that this incorrect prediction could be attributed to an incorrect feature of the shadows around the glass item; however, the number of features obtained from the top view was too small, and even humans were not easy to recognize this picture, not to mention the identification of items on machine. Only a glass image was misclassified as plastic, and it was possibly because the image did not fully show the glass neck referred to as a strong feature of glass items. Also, only two papers were misclassified as plastic and cardboards, respectively, even though Grad-CAM could still accurately located their area. Interestingly, Grad-CAM indicated that the optimized DenseNet121 specifically focused on recognizing the texture of the paper, but these causes of misclassification need to be further studied.

4. Conclusions

Recycling helps minimize the amount of waste, and using automatic classification machine of waste could further reduce labor costs of recycling tasks. An efficient vision-based image recognition model, such as deep CNN, was a vital technology to the waste sorter machine. Several CNN, such as DenseNet121, have been applied in waste classification, and have achieved an accuracy of 95%. Besides the data augmentation technique, to further improve the effectiveness of DenseNet121, this study restored the classifier of DenseNet121 to two fully-connected-layers, and then used GA to optimize the hyperparameters of the fully-connected-layer, i.e., to optimze the neuron numbers and dropout rates in order to have the optimized DenseNet121 proposed. This optimized DenseNet121 was validated on the TrashNet and compared with other CNNs.

The results show that the data augmentation method used in this study, including pictures of horizontal flipping, vertical flipping, and random 25° rotation, could achieve greater levels of efficiency and accuracy in terms of the improvement in image classification than optimizing fully-connected-layer of DenseNet121 for TrashNet dataset. Using two fully-connected-layers, the replacement of the original classifier of DenseNet121 could improve DenseNet121's performance. The optimized DenseNet121 further improved the accuracy and demonstrated the efficiency of using GA to optimize the neuron number and the dropout rate of fully-connected-layers. Therefore, the optimized DenseNet121 performed the highest accuracy of 99.6%, when compared with other CNN models. To demonstrate the robustness of optimized DenseNet121, grad-CAM highlighted the coarse features of the waste images recognized by an optimized DenseNet121. For example, both the strong features of glass and plastic materials were in the neck of the bottles; the optimized DenseNet121 could classify the paper and cardboard items based on the objects' edges and corners. Therefore, the optimized DenseNet121 indeed had a good ability to identify the class of waste. The incorrect predictions could be possibly attributed to incorrect features of shadows around the objects, the small number of features obtained from the top view, or pictures' lack of important features of the objects. In other words, novelty of research is of prime importance, and so the major contributions made by this current research proved the feasibility of the proposed optimized DenseNet121 because the proposed optimized DenseNet121 was found to have the potential to offer an efficient solution to the waste detection in waste sorting machine equipped with a digital camera system. A further contribution of this study was the use of the Grad-CAM approach to show the dominant features of the waste images extracted by the optimized DenseNet121. Further studies should be conducted to look into the causes of misclassification, and multi-objects detection of the waste images would be conducted by referring to the TrashNet. Interestingly, because different countries generate various waste and recycling classes, our future research will continue to collect waste images of our country to test the robustness of the optimized DenseNet121.

CRediT authorship contribution statement

Wei-Lung Mao: Conceptualization, Methodology, Resources, Formal analysis, Investigation, Data curation, Project administration, Writing - review & editing. **Wei-Chun Chen:** Validation, Data curation,

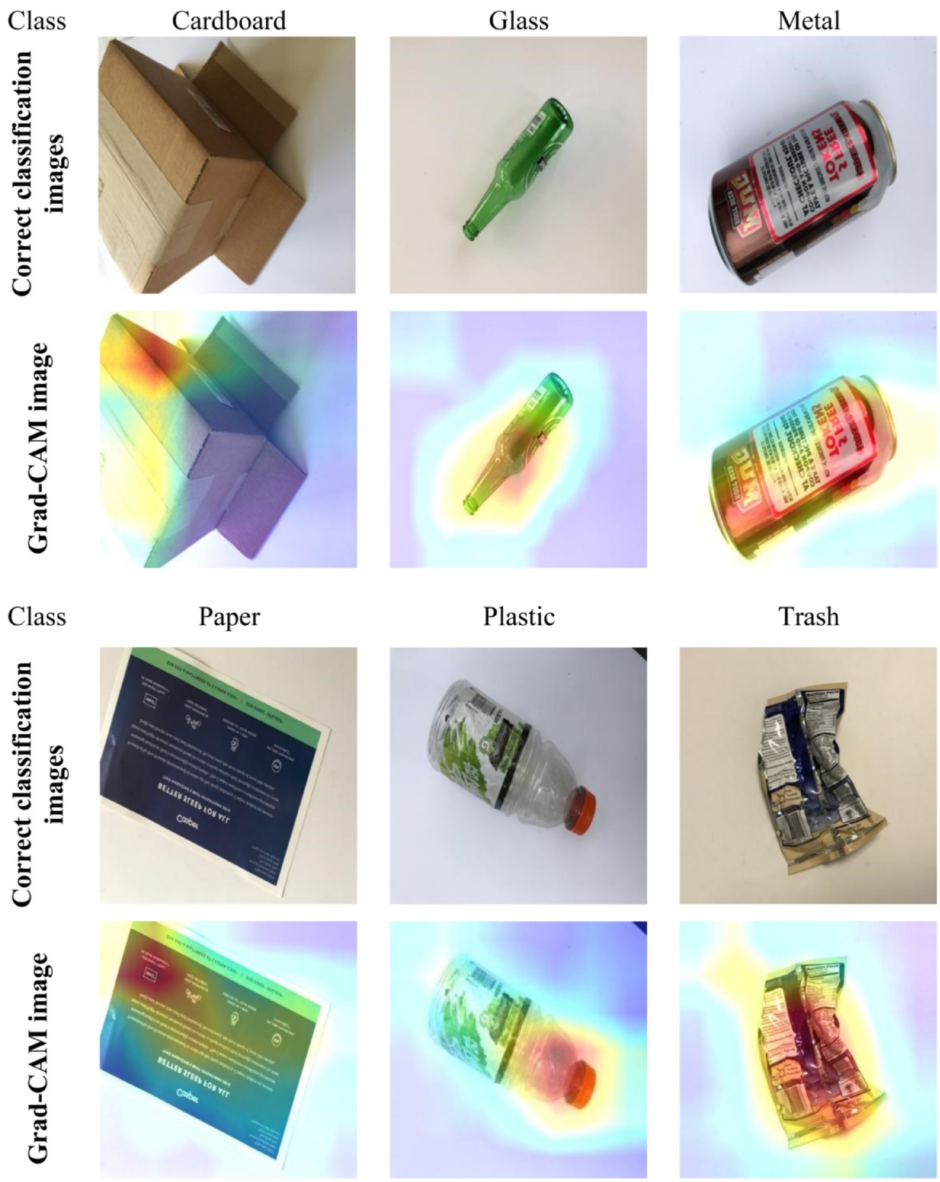

Fig. 4. The six types of waste images and their Grad-CAM images. Best viewed in color

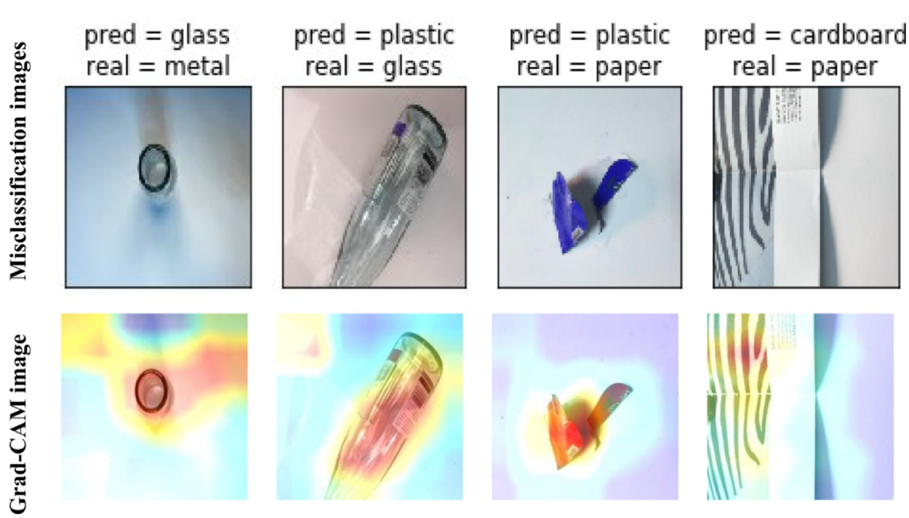

Fig. 5. Four misclassified images by optimized DenseNet121 with 10108 pictures, and their Grad-CAM images

Fig. 6. The user-friendly interface of Windows Application under Visual Studio platform

Formal analysis. **Chien-Tsung Wang:** Writing - review & editing. **Yu-Hao Lin:** Conceptualization, Supervision, Project administration, Writing - original draft, Funding acquisition.

Declaration of Competing Interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors would like to extend special thanks to National Science Council of Taiwan for the partial financial support of this research under Project MOST 108-2218-E-224 -004 -MY3. I also owe a debt of gratitude and thanks to Dr. Ping-Yu Liu, who provided me with the help related to the editorial assistance and the English language editing. I also thank Haris Imam Karim Fathurrahman for his assistance in collecting waste pictures and programming.

Reference

Abdallah, M., Abu Talib, M., Feroz, S., Nasir, Q., Abdalla, H., Mahfood, B., 2020. Artificial intelligence applications in solid waste management: a systematic research review. Waste Manage. (Oxford) 109, 231–246.
Al-Hyari, A., Areibi, S., 2017. Design space exploration of convolutional neural networks based on evolutionary algorithms. J. Comput. Vision Imag. Syst. 3 (1).
Aral, R.A., Keskin, Ş.R., Kaya, M., Hacıömeroğlu, M., 2018. Classification of trashnet dataset based on deep learning models. In: 2018 IEEE International Conference on Big Data (Big Data). IEEE. pp. 2058–2062.
Bircanoğlu, C., Atay, M., Beşer, F., Genç, Ö., Kızrak, M.A., 2018. RecycleNet: intelligent waste sorting using deep neural networks. In: 2018 Innovations in Intelligent Systems and Applications (INISTA). IEEE. pp. 1–7.
Chen, H.-W., Chen, W.-Y., Wang, C.-T., Lin, Y.-H., Deng, M.-J., Chiang, C.-Y., 2019. Managing water quality in a river basin with uncertainty. Int. J. Environ. Sci. Technol. 1–12.
Chen, Y., Tong, Z., Zheng, Y., Samuelson, H., Norford, L., 2020. Transfer learning with deep neural networks for model predictive control of HVAC and natural ventilation in smart buildings. J. Cleaner Prod. 254, 119866.
Chollet, F., 1997. Keras: The python deep learning library. https://keras.io/.
Chu, Y., Huang, C., Xie, X., Tan, B., Kamal, S., Xiong, X., 2018. Multilayer hybrid deep-learning method for waste classification and recycling. Comput. Intell. Neurosci. 2018.
Elagroudy, S., Warith, M.A., El Zayat, M., 2016. Municipal solid waste management and green economy. Glob. Young Acad. 10–15.
Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., Greenspan, H., 2018. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing 321, 321–331.
Han, D., Liu, Q., Fan, W., 2018. A new image classification method using CNN transfer learning and web data augmentation. Expert Syst. Appl. 95, 43–56.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.
Huang, G.-L., He, J., Xu, Z., Huang, G., 2020. A combination model based on transfer learning for waste classification. Concurr. Comput.: Practice Exp. n/a(n/a) e5751.
Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708.
Kim, J., Nocentini, O., Scafuro, M., Limosani, R., Manzi, A., Dario, P., Cavallo, F., 2019. An innovative automated robotic system based on deep learning approach for recycling objects. ICINCO (2), 613–622.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 1097–1105.
LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. In: Proceedings of the IEEE. 86. pp. 2278–2324.
Lin, M., Chen, Q., Yan, S., 2013. Network in network. arXiv preprint arXiv 1312, 4400.
Liu, X., Zhou, Q., Zhao, J., Shen, H., Xiong, X., 2019. Fault diagnosis of rotating machinery under noisy environment conditions based on a 1-D convolutional auto-encoder and 1-D convolutional neural network. Sensors 19 (4), 972.
McKinney, D.C., Lin, M.D., 1994. Genetic algorithm solution of groundwater management models. Water Resour. Res. 30 (6), 1897–1906.
Meng, S., Chu, W.-T., 2020. A study of garbage classification with convolutional neural networks. In: ICAN 2020: international conference on computing, Analytics and Networks. Taiwan.
Mittal, G., Yagnik, K.B., Garg, M., Krishnan, N.C., 2016. Spotgarbage: smartphone app to detect garbage using deep learning. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 940–945.
Navaneeth, B., Suchetha, M., 2019. PSO optimized 1-D CNN-SVM architecture for real-time detection and classification applications. Comput. Biol. Med. 108, 85–92.
Nowakowski, P., Pamuła, T., 2020. Application of deep learning object classifier to improve e-waste collection planning. Waste Manage. (Oxford) 109, 1–9.
Pan, S.J., Yang, Q., 2009. A survey on transfer learning. In: IEEE Transactions on knowledge and data engineering. 22. pp. 1345–1359.
Panwar, H., Gupta, P.K., Siddiqui, M.K., Morales-Menendez, R., Bhardwaj, P., Sharma, S., Sarker, I.H., 2020. AquaVision: automating the detection of waste in water bodies using deep transfer learning. Case Stud. Chem. Environ. Eng., 100026.
Piroozfard, H., Wong, K.Y., Wong, W.P., 2018. Minimizing total carbon footprint and total late work criterion in flexible job shop scheduling by using an improved multi-objective genetic algorithm. Resour. Conserv. Recycl. 128, 267–283.
Raj, S., Ray, K.C., 2017. ECG signal analysis using DCT-based DOST and PSO optimized SVM. In: IEEE Transactions on instrumentation and measurement. 66. pp. 470–478.
Rattanavorragant, R., Jewajinda, Y., 2019. A hyper-parameter optimization for deep neural network using an island-based genetic algorithm. In: 2019 16th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON). IEEE. pp. 73–76.
Seike, T., Isobe, T., Harada, Y., Kim, Y., Shimura, M., 2018. Analysis of the efficacy and

feasibility of recycling PVC sashes in Japan. Resour. Conserv. Recycl. 131, 41–53.
Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-cam: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision, pp. 618–626.
Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
Tachwali, Y., Al-Assaf, Y., Al-Ali, A.R., 2007. Automatic multistage classification system for plastic bottles recycling. Resour. Conserv. Recycl. 52 (2), 266–285.
Tong, X., Sun, S., Fu, M., 2019. Data augmentation and second-order pooling for facial expression recognition. IEEE Access 7, 86821–86828.
Vo, A.H., Vo, M.T., Le, T., 2019. A novel framework for trash classification using deep transfer learning. IEEE Access 7, 178631–178639.
Wang, Z., Peng, B., Huang, Y., Sun, G., 2019. Classification for plastic bottles recycling based on image recognition. Waste Manage. (Oxford) 88, 170–181.
Yang, M., Thung, G., 2016. Classification of trash for recyclability status. CS229 Project Rep. 2016.
Zeiler, M.D., 2012. ADADELTA: an adaptive learning rate method. arXiv preprint arXiv 1212, 5701.
Zhang, H., Wang, K., Tian, Y., Gou, C., Wang, F.-Y., 2018. MFR-CNN: incorporating multi-scale features and global information for traffic object detection. In: IEEE Transactions on Vehicular Technology. 67. pp. 8019–8030.