A double branches binary neural network with the application for garbage classification

Meijin Lin[*] Senjie Chen Zihan Zhang

School of Mechatronic Engineering and Automation, Foshan University, Foshan 528000, China

Abstract: In order to improve the efficiency of waste classification, automatic garbage classification technology gradually replaces the traditional manual sorting method. Deep neural networks are popular in the field of artificial intelligence, however, it faces the problems of a number of layers, millions of parameters, the heavy computation and storage, which inevitably limits its application for garbage classification in practice. To further improve the efficiency of waste classification, a garbage image classification model based on double branches binary neural network (DBBNN) is proposed in this paper. In DBBNN, an improved network architecture with an extra compensation module is designed to offset the information loss. Based on hinge loss function, an improved network loss function named HP-loss is proposed. Combined with the exponential decreasing learning rate, the DBBNN model is trained to meet the requirements of garbage classification task. In order to illustrate the performance of the proposed model, comparative experiments on CIFAR-10 and GIGO public datasets have been done for seven different models. Then, DBBNN is applied for automatic garbage classification on our dataset of garbage objects. The experimental results illustrate that the proposed DBBNN exceeds other four compared models in terms of classification accuracy.

Keywords: BNN; garbage classification; HP-loss function; Image recognition; Double Branches Structure;

[*]Corresponding author. Tel.: +86 13690675153

E-mail addresses:linmeijin@fosu.edu.cn(Meijin Lin);

2016881971@qq.com(Senjie Chen);

509891212@qq.com(Zihan Zhang)

1.Introduction

1.1 Garbage classification

As an important way to develop circular economy, garbage collection and recycling have social, economic and ecological benefits by improving the garbage treatment capacity and treatment equipment, reducing the treatment cost and the consumption of land resources. With the rapid economic development and fast growth of population, the domestic garbage production volume is increasing substantially. Therefore, garbage recycling and treatment is an arduous task **Error! Reference source not found.**2.

Garbage classification is one of the weaknesses in the development of China's environmental protection industry, and it is also one of the fundamental reasons for air pollution and difficult reuse of resources. Generally, garbage is collected and transported to the plant for unified treatment. Most garbage treatment plants rely on manual operation on the assembly line to sort waste, resulting in high labor intensity and low sorting efficiency, which hardly meet the needs of a large number of waste treatment. In addition, the types of waste sorted manually are limited, and most of the waste cannot be recycled 3. With the development of modern industrial intelligence, automatic waste sorting equipment based on computer vision technology has gradually replaced the traditional manual sorting.

In the early days, most scholars used classical image classification algorithms 456 to classify garbage images, based on manually extracted image features and corresponding classifiers. Wu Jian et al. 7 preliminarily completed waste identification by using color and texture features. HSV and K-Means algorithms were used in 8 for real-time recognition and classification of construction garbage images. Bicheng Wu et al. 9 used inception V3 to extract the features of garbage pictures. Due to the different background, size and quality of different data sets, traditional algorithms need to extract different features from the corresponding data manually. This leads to poor robustness of the algorithm, complex data processing process and long running time, which is difficult to meet the real-time requirements. With the rapid development of convolution neural network (CNN), deep learning is widely used in the field of image recognition. As a data-driven method with a strong fitting ability, CNN can effectively and automatically extract image features at a fast speed. In 2012, AlexNet 10 won the championship of ImageNet image classification competition, which marks the rise of deep learning. In the following years, some algorithms such as M3SPCANet **Error! Reference source not found.**, VGGNet [12] and ResNet 13 were proposed to further improve the accuracy of image classification and successfully applied in the field of face recognition, vehicle detection and so on. At the same time, garbage image classification has made a great breakthrough with the help of deep learning algorithm **Error! Reference source not found.Error! Reference source not found.Error! Reference source not found.Error! Reference source not found.**. In Stanford university, Yang et al. set up a public dataset called TrashNet, which contains 6 classes with 2527 pictures in total. Based on this public dataset, a classification model of recyclable garbage images based on deep learning was tested in 18 and the classification accuracy reaches 95.87%. In [19], an improved MobileNetV2 deep learning model was proposed for garbage detection and classification, which generated 90.7% of the garbage classification accuracy on the "Huawei Cloud" datasets. Besides, an optimized DenseNet121 was developed in 19 and achieved the accuracy of 99.6% on TrashNet.

From the above analysis, the theoretical research of waste classification based on full Convolutional Neural Network (CNN) has achieved great success. However, in fact, the practical application of waste classification lags far behind the theoretical research. Due to the increasing amount of domestic and industrial waste and the insufficient application of automatic sorting, a lot of waste cannot be recycled, which leads to a serious waste of resources. There are two reasons for this. Firstly, there are many different kinds of garbage with complex appearance characteristics. Waste is generally divided into four categories: recyclable waste, hazardous waste, kitchen waste and other waste. Each category listed above includes subcategories. And the same kind of waste sometimes has the problem of low similarity in appearance. According to the unique characteristics of waste, setting up a database with complete features is still in the formation stage. Secondly, the waste classifier based on full

CNN needs harsh operating conditions, such as strong data processing ability and a large amount of storage space. Waste classification should start from the source side. It is a good idea that trash cans in each community are equipped with automatic classifiers. The number of such kind of classifier systems is numerous. Obviously, they will be limited in cost, power consumption, memory and so on. Hence, it is meaningful to develop a garbage classification method suitable for small hardware systems. Based on the above reasons, this paper develops a lightweight model for garbage image classification based on an improved binary neural network (BNN), which is suitable for resource-limited systems. We are committed to reducing the workload of waste classification in waste plants, improving the accuracy of waste classification and the efficiency of resource recovery.

1.2 Binary neural networks

In 2016, Courbariaux and Bengio 21 firstly introduced a Binary Neural Network (BNN) named BinaryNet, a method which trains neural networks with binary weights and activations when computing parameters' gradient. At runtime, BinaryNet greatly reduces memory usage and replaces most multiplication operations with 1-bit exclusive-not-or (XNOR), which has a great impact on general-purpose hardware. However, BinaryNet constrains weights and activations to +1 or -1, which is easy to suffer some loss in classification accuracy. In order to improve the accuracy of BNN in classification, a large number of solutions have emerged in the past few years, such as minimizing quantization error, improving the loss function, reducing the gradient error and so on 21.

At the beginning, researchers considered improving the algorithm from the aspect of quantization error. Rastegari et al 23 proposed Binary-Weight-Networks (BWN) and XNOR-Networks. BWN adopts the setting of binary weight and full precision activation, while XNOR-Networks binarize both weight and activation. Moreover, the floating-point parameters are well approximated by introducing the scale factor of binary parameters. XNOR-Networks offer the possibility of running networks on CPUs (rather than GPUs) in real-time. The comparative experimental results show that the classification accuracy with BWN is the same as the full-precision AlexNet and outperforms BinaryConnect and BinaryNets by large margins on ImageNet. Minimizing the quantization error in a similar way to XNOR-Networks, Mishra et al 24 proposed Wide Reduced-Precision Networks (WRPN), in which the number of filters are increased in each layer. Besides, an adversarial attention mechanism was introduced to refine the binarized kernels based on their real-valued counterparts in 25, which can be effectively implemented on various mainstream backbones for the person re-identification tasks.

Some researchers indicated that the learning of the parameters in binary neural networks can be guided by improving the network loss function. Hou et al [26] proposed Loss-Aware Binarization (LAB-Net), where Quasi-Newton algorithm is used to minimize the total loss related to binary weights. Martinez et al [27] trained strong BNNs by matching the spatial attention maps computed at the output of the binary and real-valued convolutions with a loss function. The experimental results illustrated that the proposed model beats the compared methods by more than 5% top-1 accuracy on ImageNet. By using distribution loss to explicitly regularize the activation flow and developing a framework to systematically formulate the loss, the proposed approach in [28] can significantly improve the accuracy of some networks using 1-bit weights and activations for AlexNet on ImageNet dataset. In BNNs, the model is supposed to learn a soft distribution but not a hard one as in traditional classifier networks. Hence, a distillation loss in [29] was added to guide the training subsidiary component to be more stable.

Reducing the gradient error is another efficient way to improve the performance of binary neural networks. In [29], a user-defined ApproxSign function was presented to replace Sign for gradient calculation, which solved the gradient mismatch between the Sign function and the gradient generated by Straight-Through Estimator (STE) to some extent. Desired quantization functions in forward propagation can also reduce the gradient error. In 31, a Differentiable Soft Quantization (DSQ) method was designed to replace the traditional quantization function. In DSQ, to solve the gradient mismatch, the data distribution was adjusted in a steerable way. It is well known that the quantization brings information loss in both forward and backward propagation, which is the bottleneck of training accurate binary neural networks. To address these issues, Qin et al 32 propose an Information Retention Network (IR-Net) to retain the information that consists in the forward activations and backward gradients. In [**Error! Reference source not found.**], Distribution-sensitive Tivo-stage Estimator (DTE) was used to retain the information of gradients by distribution-sensitive soft approximation.

To improve the performance of BNN, the design of network architecture is also an efficient focus 34. In [35], a channel-wise interaction based binary convolutional neural network (CI-BCNN) learning method for efficient inference was proposed. The CI-BCNN mines the channel-wise interactions, where prior knowledge is provided to alleviate inconsistency of signs in binary feature maps and preserves the information of input samples during inference. Experimental results on the CIFAR-10 and ImageNet datasets demonstrate the effectiveness of the proposed approach. Similarly, a channel-wise reshaping and shifting operation on the activation distribution was proposed in 37. The Piecewise Approximation (PA) scheme can lessen accuracy loss by approximating full precision weights and activations efficiently, and maintain parallelism of bitwise operations to learn more representative features. Besides, a method was proposed in [37] by widening the data channel to reduce the information loss of the first convolutional input through the sign function.

From the above analysis, it can be seen that scholars have put forward many different measures and achieved great success in improving the performance of BNN. However, the application of BNN in low-cost and small-scale systems is still in its infancy. Aiming at the task of garbage image classification, this paper proposes an improved model based on BNN. We add an information compensation module to reduce the loss of image information and adjust the distribution of data in the process of binarization, and improve the loss function according to the new network architecture and task characteristics. On the basis of hinge loss function, we add a penalty term to train the model with exponentially decreasing learning rate, so that our model can meet the task requirements of garbage classification.

2 Dataset and analysis

2.1 Dataset

In order to verify the effectiveness of our proposed classification model, we establish a garbage image dataset. This dataset has 1800 garbage images, of which 35% are from the internet, 55% are from Huawei Cloud, and the remainder is taken by a digital camera. In this dataset, four kinds of waste (kitchen waste, recyclables, hazardous waste and other garbage) with 12 categories (fruit peel, bone, eggshell, can, glass bottle, old clothes, plastic hanger, soiled plastic, dry battery, cigarette, plastic bag and surgical mask) are considered. All images in the dataset have been marked with their corresponding categories. To ensure the consistency of data, each image has the same size of 64×64 and format of JPG. An example of garbage image dataset is shown in Figure 1.

2.2 Data analysis

Image recognition technology is important in the field of artificial intelligence. It is a technology of object recognition based on the main features of image. However, the same kind of garbage usually has different shapes and colors due to different experiences, which means the garbage categories are visually heterogeneous with different sizes, origins, materials, and visual appearance of the objects of interest. From the perspective of image recognition technology, the same kind of garbage often shows complex and obviously different features. Compared with other image recognition tasks, such as face recognition, fruit recognition, handwriting recognition and so on, garbage image recognition is more challenging. To fully illustrate this difference, two sets of figures are presented in Figure 2 and Figure 3.

Figure 2 shows two sets of original face images of the same person with different expressions, along with corresponding gradient, gray histogram, and inter-pixel redundancy. Figure 3 shows two sets of original can images of the same kind with different experiences, along with corresponding gradient, gray histogram, and inter-pixel redundancy.

It can be seen that the gradients diagrams, gray histograms and pixel difference of two face images have only small differences. It shows that in the case of the same person, even if the expression is different, the difference between samples is small.

From the original image of Figure 3, we can see that the shapes and colors of two cans are quite different due to their different experiences. They are two common shapes of cans in life. Although they belong to the same category, their gradient diagrams and gray histograms are obviously different. From the perspective of pixel redundancy, it is evident that there are significant differences in the distribution of pixel differences and the magnitude of statistical features between two cans of the same type. Of course, the same situation occurs in other category of garbage. In summary, this requires the network model to pay more attention to the differences between samples and learn the correlation between samples during training.

2.3 Data preprocessing

In real life, garbages can appear in different forms and colors due to various factors, which can be some newly designed products, shape change caused by external force, or color change caused by chemistry and sunlight. In order to ensure that the model can fully learn the different features of the samples, diverse data augmentation strategies which are translate, fip, HSV transformation, shear, rotation and scaling, perspective transformation and so on, have been used to preprocess the severely deformed object images. For translate transformation, random translate range is set as [0, 0.2]. Flip means horizontal flipping with a probability of 0.5 in our paper. Random HSV, random perspective transformation and random shear with the range [0, 30°] have been adopted to increase dataset variability. For rotation and scaling, a 30-degree Random Rotation and a random scale within [0.8, 1.2] add new perspectives. These techniques are intended to increase the size of the dataset and improve the accuracy of class representation for deep learning model training. Data preprocessing for the sample of a deformed can is shown in Fig. 4.

3 Methodology

3.1 Double branches binary neural network (DBBNN)

It is well known that the depth of network has an important impact on its performance. ResNet 13 and VGGNet 12 have shown that the deeper the depth, the better the network performance. They constructed a deep CNN to extract image features and realized high-precision recognition. Deep networks integrate low- level, mediumlevel and high-level features, and multi-layer features can be enriched by network stacking. However, excessively increasing the depth of the network will lead to too many parameters, gradient explosion, over fitting and so on. In 2014, GoogleNet firstly proposed a parallel merging of convolution kernels, named the bottleneck layer. By using convolution kernels of different sizes, multi-branch architectures can extract different sparsity features in the same layer, each layer in the network can learn sparse and non-sparse features, which increases the adaptability of the network to scale. Moreover, features are synthesized by deep concat to obtain the nonlinear attribute.

Due to the quantitative expression of weights and activation values, the demand for storage space of BNN is greatly reduced, which is suitable for embedded systems. In BNN, the weights and activation values are usually set to 1 or - 1, which leads to a lot of information loss in multiple convolution operations. Compared with the traditional full-precision convolutions, binary convolutions express much less information, and the features extracted by the convolution module are relatively single. To overcome this drawback, XiaoFan Lin 38 proposed ABC-Net with three branches and the architecture of the approximate revolution, which is expected to approximate the linear combination of the traditional full precision convolution and the binary convolution.

It is well-known that the same kind of garbage usually has different shapes and colors due to different experiences. From the perspective of image recognition technology, the same kind of garbage often shows complex and obviously different features. In combination with the characteristics of garbage images, a DBBNN is designed to reduce the information loss caused by binarization. The network architecture of DBBNN is shown in Figure 5 below. The model has nonlinear properties and is expected to extract more garbage image features.

As Figure 5 shown, DBBNN includes two branches with three key elements, BFE module (binary feature extraction module), BIC module (binary information compensation module) and BSE block (binary squeeze-and-excitation block). First of all, garbage images enter two different branches for feature extraction and flattening. Then, the binary features from these two branches are handled by concat operation, and finally the prediction results are obtained through the classification module.

3.1.1 BFE module

BFE module has four different binary convolutional submodules which works as a backbone network to extract main features of garbage images. As can be seen from Figure 6, the first submodule consists of a binary convolution layer, a BN layer and an activation layer. The second one consists of a binary convolution layer, a maximum pooling layer, a BN layer and an activation layer. The third one consists of a binary convolution layer, a BN layer and an activation layer. The fourth one consists of a binary convolution layer, a maximum pooling layer, a BN layer and an activation layer. In BFE module, different combinations of convolution kernels, pooling layers and BN layers are used to reduce the amount of computation and obtain more image features.

3.1.2 BIC module

As we know, the same garbage often shows complex and obviously different characteristics. Multiple binary convolution operations will lead to a lot of information loss, resulting in network over fitting and low recognition accuracy. In our proposed DBBNN, BFE module is used as the backbone network to extract the main features of garbage images, and BIC module is used as a supplement to provide as much context information as possible. In order to extract image features as much as possible and improve the recognition accuracy of the model, BIC module is designed consisting of a maximum pool layer, three convolution blocks and a BN layer, as can be seen in Figure 7. First, the image data can be dimensionally reduced through the maximum pooling layer. Then, the three spatially separable convolutional layers and residual network architecture are used to extract features. The residuals can help deepen the depth of training, ensure that low dimensional features can be retained, and help preserve image information. Finally, BN layer normalizes the feature data and adjusts the data distribution.

3.1.3 BSE block

To further extract features and improve the accuracy of the model, a binary squeeze and extraction block (BSE block) is developed and placed behind the BIC module. In the BSE block, the weights of SE-Net (squeeze and extraction networks) 39 are binarized, and a BatchNorm layer is added. SE-Net focuses on the channel relationship. In SE-Net, channel-wise feature responses by explicitly modelling interdependencies between channels are adaptively recalibrated. Due to the characteristics of ReLU activation function and binary neural network, more and more weights fall into the hard saturation region in the training process. To solve this problem, the BatchNorm layer is introduced before the sigmoid activation function. In the BatchNorm layer, the input data is normalized, which is conducive to reducing over fitting. However, the BatchNorm layer inevitably increases the amount of calculation, so we reset the scaling factor in the BSE Block. The structure of BSE Block is shown in the Figure 8 below.

In Figure 8, C represents the number of channels, H represents the height of feature map, W represents the width of feature map, $W_1 \in R^{\frac{C}{r} \times C}$, $W_2 \in R^{\frac{C}{r} \times C}$ and r is scale factor. In general, on the second branch of the DBBNN, the garbage image is dimensionally reduced and feature extracted through the BIC module, and then some important features are selectively enhanced, while the unimportant features are compressed through the BSE module.

3.1.4 Concat operation

In this paper, the image features extracted from the two branches are fused through concat operation 39. The concat operation is expressed in the following:

$$Z_{concat} = \sum_{i=1}^{C} X_i * K_i + \sum_{i=1}^{C} Y_i * K_{i+c} \qquad (1)$$

Where X_i and Y_i represent the input of the i^{th} channel of two branches, respectively, $*$ denotes the convolution operation and C represents the number of channels. It is known that after concat operation, the information in the feature map does not change, but the number of channels increases 41.

3.2 Loss function

Loss function plays an important role in the training of artificial neural networks [42]. It is the same for BNN. In order to train the BNN through learning with noisy supervision, Kai Han 43 added a penalty item to the cross-entropy loss function to help the training of the binarization mapping. To achieve higher compression rate and recognition accuracy, two alternative methods in [44] are provided to calculate the prediction loss.

Many quantized convolutional neural networks including binarized models [45][46][47] determined the optimal quantizer by minimizing quantization errors:

$$\min J\left[Q_x\left(x\right)\right]=\left\|x-Q_x(x)\right\|^2 \tag{2}$$

where x indicates the full-precision parameters, $Q_x(x)$ denotes the quantized parameters and $J\left[Q_x\left(x\right)\right]$ denotes the quantization error between full-precision and binary parameters. If we only focus on the minimization of quantization error, in extreme cases, the information entropy of quantization parameters may be close to zero. To solve this problem, Libra Parameter Binarization (Libra-PB) of IR-Net was proposed in 32. IR-Net adopts the error decay estimator to calculate gradients and minimizes the information loss by better approximating the Sign function, which ensures sufficient updating at the beginning and accurate gradients at the end of training. The loss function of IR-Net combines the quantization error and information entropy of quantized values:

$$L_{IR}=\min J\left[Q_x\left(x\right)\right]-\lambda\mathcal{H}[Q_x\left(x\right)] \tag{3}$$

where x represents a full precision parameter, $Q_x\left(x\right)$ is binary quantized function, J denotes quantization error, $\mathcal{H}$ represents information entropy, and λ is a factor.

The parameters of the binary network model can only represent 0 or 1, so that the amount of information expressed by neurons is very limited. Garbage images have an important feature, that is, there may be large differences between images of the same kind. Different from Qin's concept of loss function design based on quantization error, we focus on the vector difference between the predicted value and the actual value in the training process.

To minimize the information loss in forward propagation, hinge with a penalty was adopted that jointly considers both vector distance error and information loss.

In 48, a penalty on the coefficients was added to hinge to solve the problems of over fitting and gradient explosion. In this paper, to improve the hinge function, the vector difference between the predicted value and the real value of the garbage image is focused, and a penalty is designed and added. The mathematical expression of the Hp loss function is listed as follows.

$$L_{Hp} = \max(0, \frac{1}{m}\sum_{i=1}^{m} \mu - a_i y_i + m\varphi) \tag{4}$$

Where m represents the sample number, y denotes the label value of the sample, a represents the predicted value, μ is a constant value in $(0,1]$, m is the multiplier coefficient of the penalty term, φ is a penalty with the expression as follows.

$$\varphi = 1 - \left(\frac{a_i y_i}{\sqrt{\sum_{j=1}^{n} a_{ij}^{2}}\sqrt{\sum_{j=1}^{n} y_{ij}^{2}}}\right)^2 \tag{5}$$

In order to better mine the correlation characteristics between garbage images of the same kind, we design a penalty term φ in the loss function. It represents the difference between the prediction vector and the label vector. And it can be inferred from the expression above, the value of φ is in [0,1].

Gradient disappearance is a well-known disadvantage of typical binary neural network, that is, the model training process converges too fast, and over fitting occurs. In order to better mine the correlation features between similar garbage images, a penalty term φ is designed for the loss function. φ with the value range [0,1] represents the difference between the prediction vector and the entity vector. When the predicted result deviates from the actual result, the penalty term plays a role. The greater the difference between the predicted value and the actual value, the penalty item value becomes larger. So that the loss function value with the penalty term will decline more smoothly during the training process, which can effectively reduce the over fitting.

4 Experiment

In this part, all experiments are conducted on a computer with a Nvidia GTX-2080

GPU, an Intel Core i5-8300H and 32G RAM. Our proposed DBBNN with the HP loss function (DBBNN-Hp) is implemented in Python3.7 with Keras. The Learning rate in our model is initially set to 0.001 and decreased to 0.0000001 over the training epoch. Straight-Through Estimator (STE) [49] is used to train our model, with the optimization method as Adam [50]. In HP loss function, the λ set as 0.6 and μ set as 1.

In order to verify the effectiveness of the proposed model, a comparative experiment was conducted on the common dataset CIFAR-10 50. Then, to illustrate the capability of the proposed method in detecting garbage/non-garbage, another experiment was performed on the publicly available “Garbage In, Garbage Out” (GIGO) dataset [52]. Finally, the garbage classification and ablation study were carried out on our garbage dataset.

4.1 Comparison experiment on CIFAR-10

In order to reflect the superiority of the proposed DBBNN-Hp, six typical quantitative neural networks were selected for comparison in this experiment, including DoReFa-Ne [53], LQ-Net [54], DSQ, LAB-Net, XNOR-Net and BNN. The classification results of the seven algorithms on CIFAR-10 are listed in Table 1, where "W" and "A" represent the weight and active bit width of the model respectively. The results in Table 1 show that our proposed model DBBNN-Hp achieves the highest classification mean average precision (CMAP) for CIFAR-10. It is worth mentioning that when the weight and activation in the model are both set to 1, our method is obviously better than the three methods LAB-Net, XNOR-Net and BNN. Compared with other three methods (DoReFa-Ne, LQ-Net and DSQ) of 1W/32A setting, our method is also slightly better.

4.2 Comparison experiment on GIGO

To illustrate the capability of the proposed method in detecting garbage/non-garbage, GIGO dataset was selected here. This dataset contains 25000 images, of which 9352 are marked as garbage images and 15648 are non-garbage. To ensure compatibility with our model, after data cleaning and normalization preprocessing, we selected 15000 images as simulation data, including 7000 garbage image data and 8000 non-garbage images. The dataset is divided into 70%, 10% and 20% for training, validation and testing, respectively. To evaluate the performance of our model on the entire dataset and reduce random sampling variance, a 5-fold cross validation method was adopted.

To evaluate the performance of our model, five well-established evaluation metrics [16] such as accuracy, precision, recall, specificity, and F1-score are used in this study. In order to reflect the superiority of the proposed DBBNN-Hp, six typical quantitative neural networks including DoReFa-Ne, LQ-Net, DSQ, LAB-Net, XNOR-Net and BNN were selected for comparison in this experiment. The simulation results are listed in Table 2.

The results in Table 2 indicate that in the comparative experiment of garbage/non garbage classification, the proposed DBBNN-Hp model outperformed other comparison models and achieved the highest f1 score. Specifically, the Accuracy value of DBBNN-Hp is 83.49%, Precision is 82.36%, Recall is 82.23%, Specificity is 84.55%, and f1-Score is 82.32%. These results demonstrate that the DBBNN-Hp model for garbage/non garbage classification can exhibit better ability than other six typical networks.

4.3 Comparison experiment on garbage classification

In order to verify the effectiveness of DBBNN-Hp for garbage classification, a garbage image dataset has been established in subsection 2.1 with 1800 garbage images, of which 1620 are for training and 180 for testing. Besides, six common lightweight models are selected as comparison methods, such as EffNet [55], ShuffleNet [56], MobileNet [57], DenseNet [58], RecycleNet[59] and HOG CNN[60]. The experimental results of seven different models for garbage classification are listed in Table 3.

It shows that in the comparative experiment of garbage classification, the proposed model DBBNN-Hp achieves the highest CMAP. Then comes EffNet, which has a CMAP of 93.89%. DenseNet ranks the third while ShuffleNet ranks last. RecycleNet and HOG CNN achieve comparable results, ranking the fifth and sixth respectively. In conclusion, compared with the other six algorithms, DBBNN-Hp has better fitting, stronger robustness and higher accuracy in garbage classification.

4.4 Ablation study

In this section, we investigate effects of different network structures combined with different loss functions on BNN performance. This ablation study is conducted on the garbage image dataset established in Section 2.1. The proposed DBBNN and baseline BNN are combined with Hp loss function and Hinge loss function respectively to obtain four different models, which are DBBNN-Hp, DBBNN-Hinge, BNN-HP and BNN-Hinge. The network structures of baseline BNN and DBBNN are set with the same number of convolution modules and the same optimization method Adam. These four different algorithms perform garbage classification independently on the garbage database, and change processes of loss function value and CMAP in 50 iterations have been shown in Figure 9 and Figure 10 respectively. The results at the 50th iteration have been listed in Table 4.

Figure 9 shows that the loss function values of all the four algorithms decrease rapidly at the beginning of the iteration, and gradually converge at the end of the iteration. Compared with the other three algorithms, the Loss curve of DBBNN-Hp is smoother, and the Loss value of DBBNN-Hp at the same iteration number is smaller.

Figure 10 shows that the CMAP values of all the four algorithms rise rapidly in 10 iterations. From the 10th to the 38th iteration, all four CMAP curves fluctuate slightly. After the 38th iteration, DBBNN-Hp converges and gets the highest CMAP value among all four algorithms.

Table 4 shows that in the comparative experiment of garbage classification, the proposed model DBBNN-Hp achieves the highest CMAP value 94.12% with the lowest Loss value 0.0751. Then comes BNN-Hp, which has a CMAP of 92.84% and a loss function value of 0.1786. DBBNN-Hinge ranks the second in term of Loss value, while ranks the third in term of CMAP. BNN-Hinge performs worst in terms of CMAP and Loss value among all the four models.

In general, the proposed DBBNN-Hp has better robustness and higher accuracy for garbage classification.

5. Conclusions

In this paper, we have designed a double branches binary neural network (DBBNN) for the intelligent garbage sorting device with limited power and memory. In DBBNN, the double branch structure can not only solve the problem of multiple features of garbage images, but also effectively reduces the loss of image information in the process of binarization. Besides, by adding a penalty term to the hinge loss function, the designed loss function Hp improves the efficiency of network training.

In order to verify the effectiveness of the proposed model, a comparative experiment has been conducted on the common dataset CIFAR-10. The results show that our proposed model DBBNN-Hp achieves the highest CMAP among the seven different models. Then, to illustrate the capability of the proposed model in detecting garbage/non-garbage, a comparative experiment on GIGO dataset was carried out with the result that our model outperformed other six compared methods.

Finally, the garbage classification and ablation study has been carried out on our garbage dataset. The results show that, DBBNN-Hp has better fitting, stronger robustness and higher CMAP than the other four models. Furthermore, the network structure DBBNN combined with the loss function Hp have proved to be an effective tool for garbage image classification.

Acknowledgments

This work was supported by Projects of Guangdong Province Department of Education (No. 2019KZDZX1034, 2021KCXTD027 and 2018KTSCX237) and Natural Science Foundation of Guangdong, China (No. 2022B1515120003 and 2019A1515110180) and the National Fund Project No. 62303116 . The authors are grateful to all the editors and the referees for their hard works.

References

1. Huang, L., Li, M., Xu, T., et al. "A waste classification method based on a capsule network", *Environmental Science and Pollution Research*, 30, pp. 86454-86462 (2023). DOI: 10.1007/s11356-023-27970-7.
2. Mollica, G.J.G. and Balestieri, J.A.P. "Is it worth generating energy with garbage?

Defining a carbon tax to encourage waste-to-energy cycles", *Applied Thermal Engineering*, 173 (2020). DOI: 10.1016/j.applthermaleng.2020.115195.

3. Zhang, Y., Shang, X., Li, K., et al. "Present situation and management countermeasures of municipal solid waste treatment technology", *Ecology and Environmental Sciences,* 20(2), pp. 389-396 (2011). DOI: 10.16258/j.cnki.1674-5906.2011.02.028.
4. Abeywickrama, T., Cheema, M. A. and Taniar, D. "K-Nearest neighbors on road networks: A journey in experimentation and in-memory implementati on", *Proceedings of the VLDB Endowment*, 9(6) (2016). DOI: 10.14778/290 4121.2904125.
5. Lowe, D.G. "Distinctive image features from scaleinvariant keypoints", *International Journal of Computer Vision*, 60(2), pp. 91-110 (2004). DOI: 10.1023/B:VISI.0000029664.99615.94.
6. Harri, C. and Stephens, M. "A combined corner and edge detector", *Alvey Vision Conference. Manchester*, pp. 207-217 (1988). DOI: 10.5244/C.2.23.
7. Wu, J., Chen, H. and Fang, W. "Research on waste analysis and recognition based on computer vision", *Information Technology and Informatization*, 10, pp. 81-83 (2016). DOI: 10.3969/j.issn.1672-9528.2016.10.020.
8. Huang, H., Han, J., Wu, F., et al. "Research on color feature extraction and classification of construction waste", *Optical and Optoelectronic Technology*, 16(1), pp. 53-57 (2018). DOI: 10.19519/j.cnki.1672-3392.2018.01.010.
9. Wu, B., Deng, X., Zhang, Z., et al. "Intelligent garbage classification system based on convolutional neural network", *Physical Experiment*, 39(11), pp. 44-49 (2019). DOI: 10.19655/j.cnki.1005-4642.2019.11.009.
10. Krizhevsky, A., Sutskever, I. and Hinton, G. "ImageNet classification with deep convolutional neural networks", *Advances in Neural Information Processing Systems*, 25(2), pp. 1106-1114 (2012). DOI: 10.1145/3065386.
11. Zhou, D. and Feng, S. "M3SPCANet: A simple and effective ConvNets with unsupervised predefined filters for face recognition", *Engineering Applications of Artificial Intelligence*, 113 (2022). DOI: 10.1016/j.engappai.2022.104936.

12. Simonyan, K. and Zisserman, A. “Very deep convolutional cetworks for large-scale image recognition”, *Computer Vision and Pattern Recognition* (2014). DOI: 10.48550/arXiv.1409.1556.
13. He, K., Zhang, X., Ren, S., et al. “Deep residual learning for image recognition”, *Conference on Computer Vision and Pattern Recognition*, pp. 770-778 (2016). DOI: 10.1109/CVPR.2016.90.
14. Chen, Y., Luo, A., Cheng, M., et al. “Classification and recycling of recyclable garbage based on deep learning”, *Journal of Cleaner Production*, 414 (2023). DOI: 10.1016/j.jclepro.2023.137558.
15. Hossen, M. M., Ashraf, A., Hasan, M., et al. “GCDN-Net: Garbage classifier deep neural network for recyclable urban waste management”, *Waste Management*, 174, pp. 439-450 (2024). DOI: 10.1016/j.wasman.2023.12.014.
16. Song, Y., He, X., Tang, X., et al. “DEEPBIN: Deep learning based garbage classification for households using sustainable natural technologies”, *Journal of Grid Computing*, 22(2) (2024). DOI: 10.1007/s10723-023-09722-6.
17. Qin, J., Wang, C., Ran, X., et al. “A robust framework combined saliency detection and image recognition for garbage classification”, *Waste Management*, 140, pp. 193-203 (2022). DOI: 10.1016/j.wasman.2021.11.027.
18. Zhang, Q., Zhang, X., Mu, X., et al. “Recyclable waste image recognition based on deep learning”, *Resources Conservation and Recycling*, 171(99) (2021). DOI: 10.1016/j.resconrec.2021.105636.
19. Jin, S.F., Yang, Z.X, Krolczykg, G., et al. “Garbage detection and classification using a new deep learning-based machine vision system as a tool for sustainable waste recycling”, *Waste Management*, 162, pp. 123-130 (2023). DOI: 10.1016/j.wasman.2023.02.014.
20. Mao, W.L., Chen, W.C., Wang, C.T., et al. “Recycling waste classification using optimized convolutional neural network”, *Resources, Conservation and Recycling*, 164 (2021). DOI: 10.1016/j.resconrec.2020.105132.
21. Courbariaux, M. and Bengio, Y. “BinaryNet: Training deep neural networks with weights and activations constrained to +1 or -1”, *Machine Learning* (2016). DOI:

10.48550/arXiv.1602.02830.

22. Qin, H., Gong R., Liu, X., et al. “Binary neural networks: A survey”, *Pattern Recognition*, 105 (2020). DOI: 10.1016/j.patcog.2020.107281.
23. Rastegari, M., Ordonez, V., Redmon, J., et al. “XNOR-Net: ImageNet classification using binary convolutional neural networks”, *European Conference on Computer Vision*, pp. 525-542 (2016). DOI: 10.1007/978-3-319-46493-0_32.
24. Mishra A., Nurvitadhi E., Cook J J., et al. “WRPN: Wide reduced-precisio n networks”, *Computer Vision and Pattern Recognition* (2017). DOI: 10.48 550/arXiv.1709.01134.
25. Xu, S., Liu, C., Zhang, B., et al. “BiRe-ID: Binary neural network for efficient person Re-ID”, *ACM Transactions on Multimedia Computing, Communications, and Applications*, 18(1), pp. 1-22 (2022). DOI: 10.1145/3473340.
26. Hou, L., Yao, Q. and Kwok, J.T. “Loss-aware binarization of deep networks”, *Neural and Evolutionary Computing* (2018). DOI: 10.48550/arXiv.1611.01600.
27. Martinez, B., Yang, J., Bulat, A., et al. “Training binary Neural networks with real-to-binary convolutions”, *Computer Vision and Pattern Recognition* (2020). DOI: 10.48550/arXiv.2003.11535.
28. Ding, R., Chin, T.W., Liu Z., et al. “Regularizing activation distribution for training binarized deep networks”, *Computer Vision and Pattern Recognition* (2019). DOI: 10.48550/arXiv.1904.02823.
29. Xu, Y., Dong, X., Li, Y., et al. “A main/subsidiary network framework for simplifying binary neural network”, *Computer Vision and Pattern Recognition* (2018). DOI: 10.1109/CVPR.2019.00732.
30. Bethge, J., Yang, H., Bornstein, M., et al. “BinaryDenseNet: Developing an architecture for binary neural networks”, *International Conference on Computer Vision Workshop*, pp. 1951-1960 (2019). DOI: 10.1109/ICCVW.2019.00244.
31. Gong, R., Liu, X., Jiang, S., et al. “Differentiable soft quantization: Bridging full-precision and low-bit neural networks”, *Computer Vision and Pattern Recognition*, pp. 4852-4861 (2019). DOI: 10.48550/arXiv.1908.05033.
32. Qin H., Gong R., Liu, X., et al. “Forward and backward information retention for

accurate binary neural networks", *Computer Vision and Pattern Recognition*, pp. 2247-2256 (2020). DOI: 10.1109/CVPR42600.2020.00232.

33. Qin, H., Zhang, X., Gong, R., et al. "Distribution-sensitive information retention for accurate binary neural network", *International Journal of Computer Vision* (2023). DOI: 10.1007/s11263-022-01687-5.
34. Zhao, Z., Xu, K., Ma, Y., et al. "A gradient optimization and manifold preserving based binary neural network for point cloud", *Pattern Recognition*, 139 (2023). DOI: 10.1016/j.patcog.2023.109445.
35. Wang, Z., Lu, J. and Zhou, J. "Learning channel-wise interactions for binary convolutional neural networks", *Computer Vision and Pattern Recognition*, 43(10), pp. 3432-3445 (2021). DOI: 10.1109/CVPR.2019.00066.
36. Zhu, B., Zaid, A.A. and Wei, P. "Towards lossless binary convolutional neural networks using piecewise approximation", *Computer Vision and Pattern Recognition* (2020). DOI: 10.48550/arXiv.2008.03520.
37. Zhang, Q., Sun, L., Yang, G., et al. "TBNN: totally-binary neural network for image classification", *Optoelectronics Letters*, 19(2), pp. 117-122 (2023). DOI:
38. Lin, X., Zhao, C. and Pan, W. "Towards accurate binary convolutional neural network", *Machine Learning* (2017). DOI: 10.1007/s11801-023-2113-2.
39. Hu, J., Shen, L. and Sun, G.G. "Squeeze-and-excitation networks", *Computer Vision and Pattern Recognition* (2018). DOI: 10.1109/CVPR.2018.00745.
40. Ma, W., Wa, N.X. and Yu, J. "A lightweight feature fusion single shot multibox detector for garbage detection", *IEEE Access*, 8, pp. 188577-188586 (2020). DOI: 10.1109/ACCESS.2020.3031990.
41. Lu, W., Liang, L., Wu, X., et al. "An adaptive multiscale fusion network based on regional attention for remote sensing images", *IEEE Access*, 8, pp. 107802-107813 (2020). DOI: 10.1109/ACCESS.2020.3000425.
42. Abbasi, J. and Andersen, S. "Simulation and prediction of counter current spontaneous imbibition at early and late time using physics-informed neural networks", Energy and Fuels, 37(18): 13721-13733 (2023). DOI: 10.1021/acs.energyfuels.3c02271.

43. Han, K., Wang, Y., Xu, Y., et al. “Training binary neural networks through learning with noisy supervision”, *International Conference on Machine Learning* (2020). DOI: 10.48550/arXiv.2010.04871.
44. Ma C., Guo, Y., Lei, Y., et al. “Binary volumetric convolutional neural networks for 3-D object recognition”, *IEEE Transactions on Instrumentation and Measurement,* 68(1) (2019). DOI: 10.1109/TIM.2018.2840598.
45. Say, B. and Sanner, S. “Planning with learned binarized neural network transition models in factored state and action spaces”, *Artificial Intelligence* (2020). DOI: 10.5555/3304652.3304678.
46. Chen, H., Zhuo, L., Zhang, B., et al. “Binarized neural architecture search for efficient object recognition”, *International Journal of Computer Vision,* 129(2) (2021). DOI: 10.48550/arXiv.2009.04247.
47. Du, G., Chen, B., Li, Z., et al. “A BNN accelerator based on edge-skip-calculation strategy and consolidation compressed tree”, ACM Transactions on Reconfigurable Technology and Systems, 15, pp. 1-20 (2022). DOI: 10.1145/3494569.
48. Tarigan, B., and Geer, S.A. “Classifiers of support vector machine type with \ell1 Complexity regularization”, Bernoulli, 12(6), pp. 1045-1076 (2006). DOI: 10.3150/bj/1165269150.
49. Bengio, Y., Nicholas, L. and Courville, A. “Estimating or propagating gradients through stochastic neurons for conditional computation”, Machine Learning (2013). DOI: 10.48550/arXiv.1308.3432.
50. Kingma, D. and Ba, J. “Adam: A method for stochastic optimization”, *Machine Learning,* (2014). DOI: 10.48550/arXiv.1412.6980.
51. Krizhevsky, A., Nair, V. and Hinton, G. “The cifar-10 dataset”, online: http://www. cs. toronto. edu/kriz/cifar. html (2014). DOI: 10.24432/C5889J.
52. Sukel, M., Rudinac, s. and Worring, M. “GIGO, Garbage In, Garbage Out: An Urban Garbage Classification Dataset”, *International Conference on Multimedia Modeling,* pp.527-538 (2023). DOI: 10.1007/978-3-031-27077-2_41.
53. Zhou, S., Wu, Y., Ni, Z., et al. “DoReFa-Net: Training low bitwidth convolutional

neural networks with low bitwidth gradients", *Neural and Evolutionary Computing* (2016). DOI: 10.48550/arXiv.1606.06160.

54. Zhang, D., Yang, J., Ye, D., et al. "LQ-Nets: Learned quantization for highly accurate and compact deep neural networks", *Computer Vision and Pattern Recognition* (2018). DOI: 10.48550/arXiv.1807.10029.
55. Freeman, I., Lutz, R.K. and Kummert, A. "Effnet: An efficient structure for convolutional neural networks", *International Conference on Image Processing* (2018). DOI: 10.48550/arXiv.1801.06434.
56. Zhang, X., Zhou, X., Lin, M., et al. "ShuffleNet: An extremely efficient c onvolutional neural network for mobile devices", *Computer Vision and Patt ern Recognition*, pp. 6848-6856 (2018). DOI: 10.48550/arXiv.1707.01083.
57. Howard, A.G., Zhu, M., Chen, B., et al. "MobileNets: Efficient convolutional neural networks for mobile vision applications", *Computer Vision and Pattern Recognition* (2017). DOI: 10.48550/arXiv.1704.04861.
58. Huang, G., Liu, Z., Pleiss, G., et al. "Convolutional networks with dense connectivity", *Transactions on Pattern Analysis and Machine Intelligence*, 44(12), pp. 8704-8716 (2022). DOI: 10.1109/TPAMI.2019.2918284.
59. Bircanoglu, C., Atay, M., Beser, F., et al. "RecycleNet: Intelligent waste so rting using deep neural networks", *International Conference on Innovations in Intelligent Systems and Applications*, pp. 1-7 (2018). DOI: 10.1109/INIS TA.2018.8466276.
60. Ahamed, H., Alam, I. and Islam, M.M. "HOG-CNN based real time face recognition", *International Conference on Advancement in Electrical and Electronic Engineering*, pp. 1-4 (2019). DOI: 10.1109/ICAEEE.2018.8642989.

Biographies

Meijin Lin received the B.E. degree from North China Electric Power University, Heibei, China, in 2003. She also received the M.E. and Ph.D. degrees from South China University of Technology, Guangdong, China, in 2006 and 2015, respectively. She is currently an Associate Professor with Department of Automation, Foshan

University, Guangdong, China. Her current research interests include intelligent control.

Senjie Chen is a master's student of Foshan University. His major is artificial intelligence and his research is primarily focused on Classification algorithm. The algorithms he studied are mainly applied in the classification of garbage images and residual current classification prediction.

Zihan Zhang holds a Master's degree in Control Engineering from Foshan University. His research is primarily focused on artificial intelligence, with a particular emphasis on image algorithms and the development and application of multimodal large-scale models.

Figure 1 An example of garbage image dataset

Figure 2 Face images and corresponding gradient, gray and inter-pixel redundancy

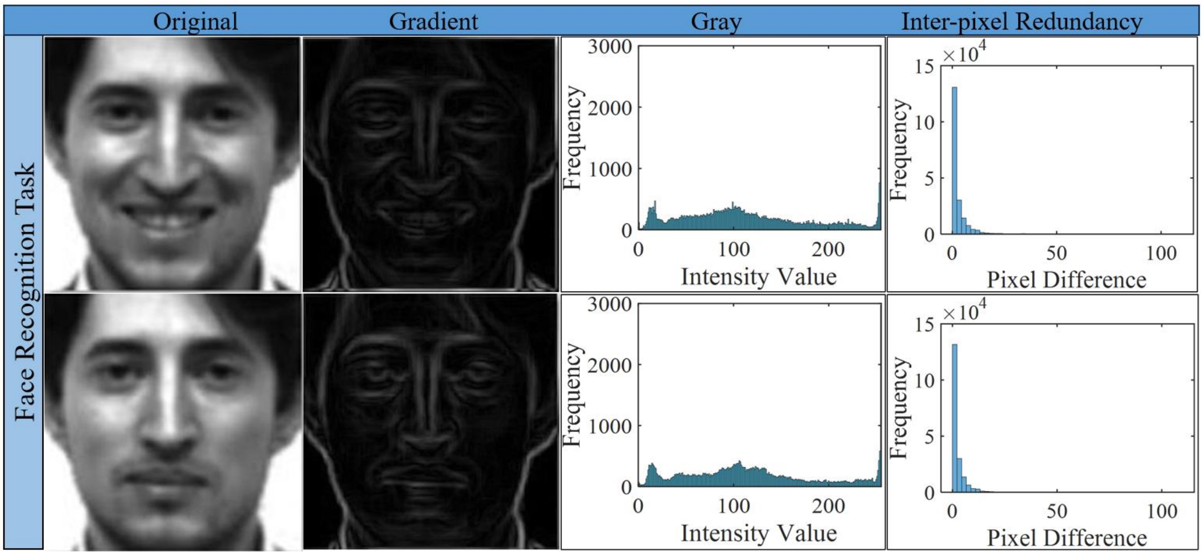

Figure 3 Can images and corresponding gradient, gray and inter-pixel redundancy

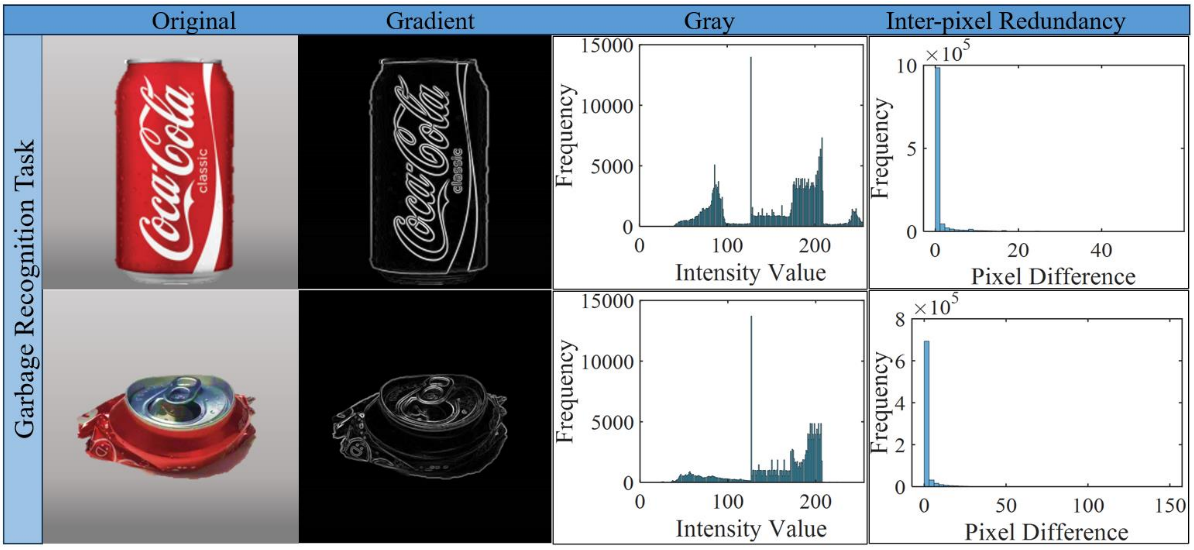

Figure 4 Data preprocessing for the sample of a deformed can

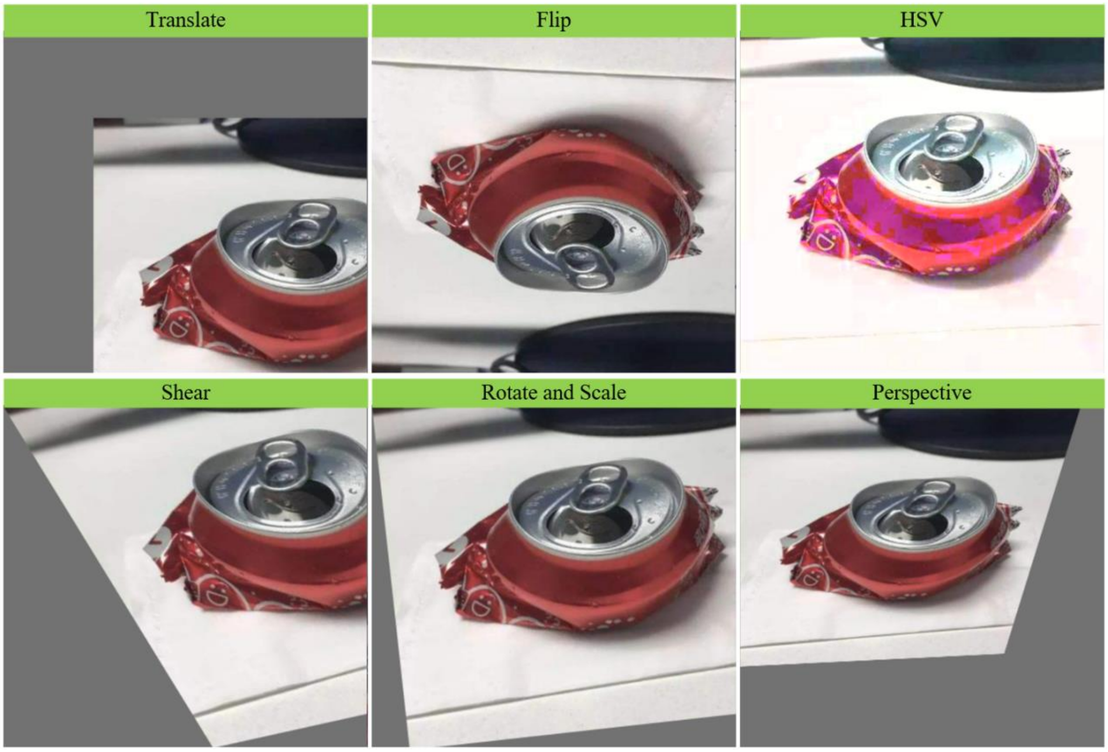

Figure 5 The architecture of DBBNN

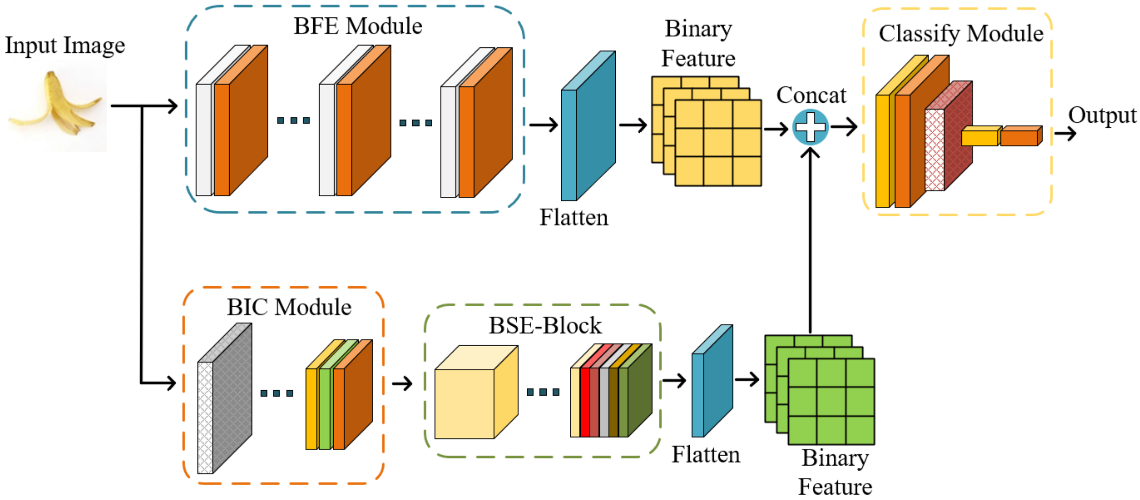

Figure 6 The structure of BFE module

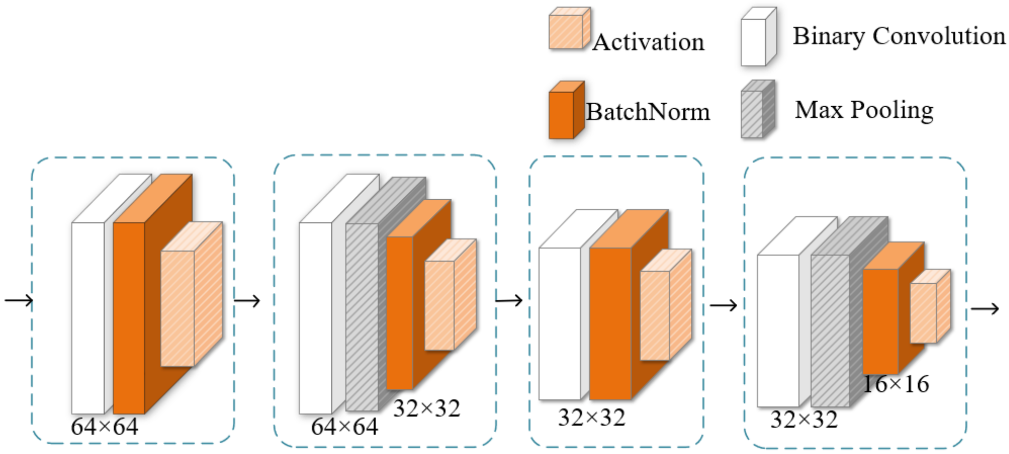

Figure 7 The structure of BIC module

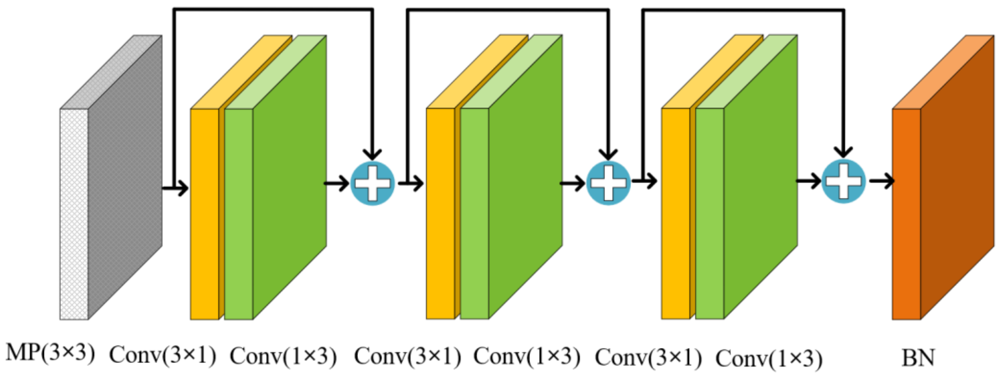

Figure 8 The structure of BSE-Block

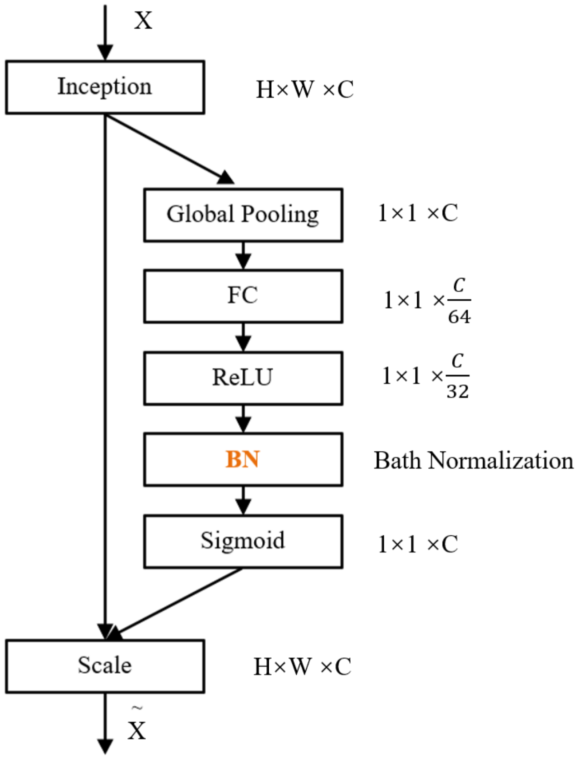

Figure 9 Loss curves of the four different models

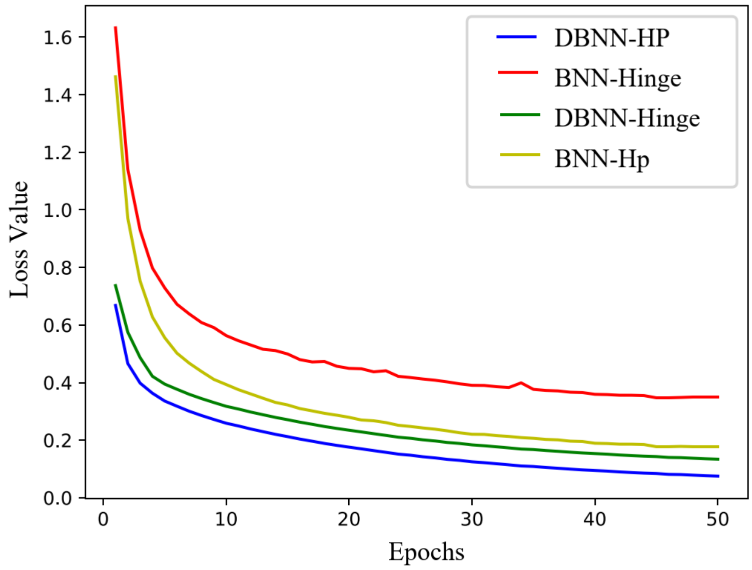

Figure 10 CMAP curves of the four different models

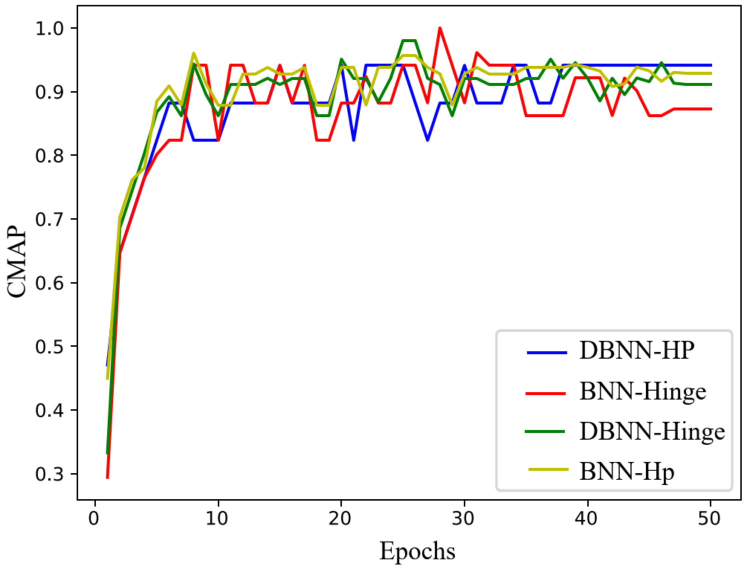

Table 1 Performance comparison of 7 different models on CIFAR-10

Method	Bit/Width	CMAP(%)
DoReFa-Net	1/32	90.0
LQ-Net	1/32	90.1
DSQ	1/32	90.2
LAB-Net	1/1	87.7
XNOR-Net	1/1	89.8
BNN	1/1	89.9
DBBNN-Hp	1/1	**90.3**

Table 2 Performance comparison of seven different models on GIGO

Method	Accuracy(%)	Precision(%)	Recall(%)	Specificity(%)	f1-Score(%)
DoReFa-Net	82.35±0.58	82.36±0.58	80.33±0.59	84.21±0.54	81.33±0.59
LQ-Net	83.31±0.37	82.09±0.39	82.13±0.38	84.33±0.32	82.11±0.38
DSQ	83.11±0.46	81.66±0.46	82.01±0.47	84.06±0.44	81.84±0.46
LAB-Net	80.93±0.52	81.43±0.52	78.51±0.53	83.20±0.51	79.94±0.52
XNOR-Net	80.67±0.45	80.16±0.45	78.78±0.44	82.37±0.43	79.46±0.45
BNN	79.33±0.60	77.31±0.63	78.16±0.61	80.34±0.60	77.74±0.63
DBBNN-Hp	**83.49**±0.51	**82.36**±0.51	**82.23**±0.50	**84.55**±0.52	**82.32**±0.52

Table 3 Performance comparison of 5 different models on garbage classification

Method	kitchen	recyclables	hazardous	Others	CMAP(%)
EffNet	87.09	96.35	97.36	94.76	93.89
ShuffleNet	79.03	95.14	90.96	89.39	88.63
MobileNet	85.38	93.44	98.50	93.88	92.80
DenseNet	86.01	95.90	94.66	95.55	93.03
RecycleNet[61]	84.73	94.36	92.90	92.13	91.03
HOG CNN[62]	87.93	91.51	92.49	91.63	90.89
DBBNN-Hp	**88.07**	**96.44**	**98.63**	**95.82**	**94.74**

Table 4 Ablation study results for DBBNN-Hp at the 50^{th} epoch

Method	Loss	CMAP(%)
BNN-Hinge	0.3486	87.24
BNN-Hp	0.1786	92.84
DBBNN-Hinge	0.1339	91.14
DBBNN-Hp	**0.0751**	**94.12**