Image recognition using convolutional neural network combined with ensemble learning algorithm

Weilong MO, Xiaoshu LUO, Yexiu Zhong, Wenjie Jiang
College of Electronic Engineering, Guangxi Normal University, Guilin 541004, China
E-mail:1915138054@qq.com

Abstract: An image recognition algorithm based on ensemble learning algorithm and convolution neural network structure (ELA-CNN) is proposed to solve the problem that a single convolution neural network (CNN) classifier may be more prone to error or unreliable prediction. In order to improve the effect of ensemble learning, enhance the transfer of features, extract deeper features and multi-scale features, the network structure uses various model structure of the mainstream algorithms. Bagging training method is used in the training process, that is, different learners use different data sets to ensure the learning differences. Finally, the prediction result of all classifiers is used to get the final image target recognition according to the decision algorithm. The algorithm is simulated with of the open data set of cifar-10. The experimental results show that the proposed algorithm has a high recognition accuracy. The recognition rate of the test set reaches 98.89% and the recognition result is secure and reliable.

1. Introduction
With the rapid development of deep learning technology theory and the improvement of computer performance and hardware cost, people require that the accuracy of machine vision recognition of image targets to be higher and higher. Especially in the fields of precision strike, road monitoring, product quality monitoring, and automatic driving, there is a high requirement for recognition accuracy, because misidentification may cause unpredictable consequences.

Convolutional neural networks have good advantages in image classification, and have achieved excellent results in many object recognition tasks, mainly because their network structure itself can extract multi-level features of images.

Krizhevsky et al. designed the AlexNet[1] network and proposed using data amplification methods to increase the number of data sets. The Zhang X team designed ShuffleNe[2], which uses point-by-point group convolution to reduce computational complexity and channel shuffle to help flow information. The effect is faster than AlexNet in ARM-based embedded devices; Yan Shuicheng proposed the DPN[3] model, which combines ResNeXt and DenseNet into a network, which makes the model low cost and more fully utilized features and achieving better results.

This paper combines an integrated learning algorithm with a deep convolutional neural network. The model uses the Baging integrated classifier to maintain an independent, parallel structure between the learners. The network structure combines ResNet, DenseNet, DenseNet-BC and Inception-Resnet-v2. Because of the differences between individuals, the decisions made by each learner when making decisions are not completely consistent. Finally, the decision method of the Bagging classifier is used to make a judgment.
2. Convolutional neural networks and integrated learning algorithms

2.1. Principle of Convolutional Neural Networks
Convolutional neural networks have become a research hotspot in the field of computational vision such as image recognition, detection and segmentation. In order to identify 2D images, the deep convolutional neural network has designed a multi-layer perceptron, which can automatically learn the distribution law of data from big data, and has powerful learning and feature expression ability, which can greatly improve the accuracy of image recognition.

2.1.1. Basic network structure and function
The convolutional layer is a key part of the convolutional neural network for feature extraction. The output of the convolutional layer is shown in Equation 1.

\[ x_j^l = y(\sum_{i \in M_j} x_i^{l-1} * w_{ij}^l + b_j^l) \]  

The pooling layer can be used for dimensionality reduction, nonlinearity, and the ability to expand the perception field. The pooling layer is shown in Equation 2.

\[ x_j^l = y(\beta^l \downarrow \text{down}(x_i^{l-1}) + b_j^l) \]  

Each neuron in the fully connected layer will be fully connected to all neurons in the upper layer, the main purpose of which is dimensional transformation, turning high-dimensional features into low-dimensional sample markers. Each neuron can use an activation function such as Relu, Sigmoid, or Tanh, and the final output is the classification result.

2.1.2. Back propagation
After the network obtains the actual output through forward propagation, the loss function is obtained. Finally, the parameters of the network are updated by backpropagation to correct the error. When the network training is completed, a nonlinear mapping relationship between input and output can be obtained. The loss function used in this article is defined as Equation 3:

\[ J = -\frac{1}{n} \sum_{ij} [y_j \ln a_j^l + (1 - y_j) \ln(1 - a_j^l)] + \lambda \sum_w w^2 \]  

The first term on the right side of the formula is the cross-entropy expression, and the second term is the L2 regularization of the weight, which is used to avoid overfitting problems after training.

2.2. Integrated learning algorithm
Integrated learning[4] is developed from machine learning[5], which refers to the use of multiple classifiers to predict the data set, thereby improving the generalization ability of the overall classifier. When classifying new data instances, by training multiple classifiers, the classification results of these classifiers are combined (such as voting) to determine the classification results to achieve better results. Figure 1 shows the structure of an integrated learning classifier.

![Fig.1 Bagging Integrated learning classifier structure](image)

The integrated learning classifier requires the following conditions:
1. There must be differences between different classifiers.
2. The classification accuracy of each classifier must exceed 0.5.
For the final combination part of Figure 1, there are three ways to decide:

1. The direct use of the highest number of votes is the classification result, that is, the minority obeys the majority (Hard Voting). The following table shows the detailed rules of Hard Voting:

| Decision method | Rule |
|-----------------|------|
| HV_1            | The minority obeys the majority. When the votes are inconsistent, they are randomly selected. |
| HV_n            | The minority obeys the majority. At least n votes, not enough to refuse to predict[6]. |

2. Weight Vote: the higher the correct rate, the greater the weight.
3. Soft Vote: Some classifiers have probabilistic information, so they can vote with probability.

The reason why integrated learning is effective is that diverse base learners can learn from each other in different models. Each base learner makes different mistakes, and it is unlikely to make mistakes in combination. In theory, the error rate can eventually become zero.

### 3. Convolutional neural network model design based on integrated learning algorithm (ELA-CNN)

#### 3.1. Overview of the overall design and principles of the network model

The structure of ELA-CNN designed in this paper is shown in Figure 2. The structure is divided into three parts: Stage1, Stage2 and Stage3. Stage1 is a three-layer convolutional layer that is then output to the three networks in the Stage2 section. Each network in the Stage 2 section is then output to the three branch of new network structures in the Stage 3 section. The network in the Stage2 and Stage3 sections maintain parameter and structural differences.

#### 3.2. The specific design of the network structure model

Table 2 is a specific parameter corresponding to the structure of Fig. 2 one by one. The table integrates a variety of mainstream networks with different characteristics. Table 2 uses the ResBlock structure of literature [7]. The structure consists of a residual branch and a shortcut branch. When the network is deep, it can transmit lowlevel information so that the network does not disappear. DensesBlock and DenseBC_Block are derived from the literature [8], each block connects all the layers to achieve the effect of feature reuse, especially in the back propagation is more conducive to the spread of the gradient. Block35 and Block8 are derived from the paper [9], which combines the Inception module with the Residual Connection, resulting in faster convergence and improved performance.
The entire network is mixed with various modules of Resnet, Densenet and Inception-Resnet-v2, and the parameters and layers of each classifier are different, which meets the requirements of the integrated learning algorithm to ensure that each classifier guarantees differentiation.

| State | State 2 | State 3 |
|-------|---------|---------|
| 3*Cnn | 32*ResBlock + 10*Block35 | 5 * DenseBC Block |
|       |         | 4 * DenseBC Block |
|       |         | 6 * DenseBC Block |
|       | 32*ResBlock | 6 * DensesBlock |
|       |         | 5 * DensesBlock |
|       |         | 4 * DensesBlock |
| 32*ResBlock + 9*Block8 | 10 * Block8 |
|       | 9 * Block8 |
|       | 8 * Block8 |

4. Simulation experiment

4.1. Experimental condition setting

The computer configuration used in this experiment is a dual E5-2637 v4 CPU, the operating system is Ubuntu 16.04, and also uses GTX1080Ti graphics card, 32GB memory to accelerate training. The platform used is the machine learning framework tensorflow1.9 developed by Google.

The experiment mainly compares the performance of ELA-CNN of different decision algorithms in the cifar10 data set. The data set used in this experiment was cifar-10. Among them, there are 60,000 color images in the cifar-10 dataset. These images are 32*32 in size and have 10 categories. The categories include airplanes, cars, birds, cats, deer, dogs, frogs, horses, boats and trucks. There are 6000 pictures for each category. There are 50,000 pieces for training, divided into 5 training batches. Each picture is clipped to a 24*24 size from 32*32 randomly, and then trained by random flip, zoom, color transform, noise disturbance and so on.

The training also uses the learning rate of change, as well as weight decay, adding Dropout[10], batch standardization and other optimization techniques to improve training accuracy.

The training method of the integrated learning algorithm Bagging is used in the training process, that is, different learners use different data sets one by one, so that each classifier learns different contents to ensure learning differences.

Each 64 images are trained for one batch, one batch is one step, the learning rate is set to the first 30,000 steps using a learning rate of 0.01, and the 30,000 to 60,000 to step learning rate is 0.005, 60,000 to 80,000 steps. The learning rate used after 0.001, 80,000 steps is 0.0005.

4.2. Experimental results

4.2.1. Experimental results and analysis

The number of network predictions is 12, of which Stage2 has 3 prediction outputs and Stage3 has 9 prediction outputs. The final training result is shown in Figure 3. It can be seen from the experimental results that the accuracy of the prediction based on ELA-CNN structure is much higher than that of a single network. From the curve, the decision modes hv4, hv6, hv8, hv10, hv12 are improved in turn. The curve avg_12_predict is the average of the prediction accuracy of 12 individual networks, and the correct rate is the lowest. The correct rate is the highest using the hv12 decision method.
Fig.3 ELA-CNN accuracy curve

Table 3 shows the test results of 10,000 test pictures, which gives the correct rate of different decisions of ELA-CNN. The SV_8 in the table is the probability voting decision method. After all the predicted probability vectors of the 12 classifiers are added together, the minimum value in the vector is 0, the maximum value is 12, and the largest probability in the prediction vector is the final prediction. The value, the sum of the probabilities does not exceed 8, and rejects the prediction.

| Decision making method | accuracy/% | Reject numbers |
|------------------------|------------|----------------|
| HV_2                   | 92.18      | 0              |
| HV_4                   | 92.18      | 0              |
| HV_6                   | 93.06      | 137            |
| HV_8                   | 94.15      | 380            |
| HV_10                  | 98.13      | 1257           |
| HV_12                  | 98.89      | 1658           |
| SV_8                   | 96.06      | 872            |

It can be concluded from the experimental results that ELA-CNN combines a variety of different characteristics of the structure, the stricter decision-making method, the number of rejects will also increase, which can effectively eliminate invalid image categories and improve the reliability of prediction.

4.3. Comparison of experimental results

Table 4 Comparison of image recognition accuracy of different CNN models

| method               | accuracy/% |
|----------------------|------------|
| paper[11]            | 82.90      |
| paper[11]            | 89.29      |
| paper[12]            | 90.37      |
| paper[13]            | 91.20      |
| paper[13]            | 92.05      |
| ResNet 101           | 93.75      |
| DenseNet 121         | 95.04      |
| DPN 92               | 95.16      |
| This article ELA-CNN (HV_12) | 98.89      |
5. Conclusion
This paper proposes a model based on integrated learning algorithm combined with convolutional neural network, and uses the cifar10 data set to simulate. The research results show that the convolutional neural network based on integrated learning algorithm can improve the target recognition rate and improve the recognition reliability under the premise of increasing the number of parameters. However, there are still many problems to be solved, such as how to reduce the number of valid images rejected by prediction or how to choose a better decision method.

References
[1] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012:1097-1105.
[2] Zhang X, Zhou X, Lin M, et al. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices[J]. 2017.
[3] Chen Y, Li J, Xiao H, et al. Dual Path Networks[J]. 2017.
[4] Song Xiangfa, Jiao Lichen. Multi-example multi-label image classification method based on sparse coding and integrated learning[J]. Journal of Electronics and Information, 2013, 35(3):622-626.
[5] Yang Shanlin, Ni Zhiwei. Machine learning and intelligent decision support system[M]. Science Press, 2004.
[6] Zhou Zhilhua, YangQiang. Machine learning and its application[M]. Tsinghua University Press, 2011.
[7] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[J]. 2015:770-778.
[8] Huang G, Liu Z, Maaten L V D, et al. Densely Connected Convolutional Networks[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2017:2261-2269.
[9] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning[J]. 2016.
[10] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[J]. 2015:448-456.
[11] Lin Z, Quantisation and pooling method for lowinference latency spiking neural networks[J]. Electronics Letters, 2017, 53(20):1347-1348.
[12] Yan Yuanfei, Zeng shangyou, FenYanyan, et al. Research on Deep Convolutional Network Design Based on Parallel and Slice[J]. Microelectronics and Computer, 2018(3).
[13] Chang xiang, Yanming. Image classification performance based on improved convolutional neural network[J]. Journal of Chongqing University of Technology, 2017, 31(3):110-115.
[14] Peng Zhanjun,Peng Yanbin,Chen Guang.Image classification model optimization based on CIFAR-10[J].Computer applications and software,2018,35(03):177-181.
[15] Hou Congcong,He Yuqing,Jiang Xiao Heng, et al. Deep convolutional neural network based on two-branch convolution unit [J].Progress in laser and optoelectronics,2018,55(02):192-198.