Paper:

The Cars' Type Recognition Algorithm Based on Ensemble Learning and Multi-Features Fusion

Bin,Cao[1] Jiahui,Wang[2] Hongbin,Ma[*3] Ying,Jin[4]

Beijing Institute of Technology, Beijing 100081,P.R.China
E-mail: mathmhb@bit.edu.cn

Abstract. The automobile industry is an important industry in the national economy. It has the characteristics of long industrial chain, high accuracy and integrating all kinds of high and new technologies. It has played a leading role in the national economic and social development and the cars' type recognition plays an important role in automobile intelligent manufacturing. Some scholars are studying the application of deep learning algorithm for cars' type recognition[1–3]. However, the problems of lacking samples and high accuracy make it difficult for a single deep learning technology to meet needs from production. Based on the important industrial scene of cars' type recognition, this paper integrates a variety of image features under the framework of ensemble learning based on the in-depth understanding of automobile intelligent production lines and improves the existing ensemble learning algorithms so as to design a cars' type recognition algorithm based on multi-features fusion and ensemble learning. The algorithm can meet the stringent requirements of industrial applications and has low dependence on samples, which makes the algorithm have a certain value of popularization in automobile manufacturers. The algorithm proposed in this paper has achieved good results in automobile intelligent production line.Compared with algorithms using single feature, the accuracy of proposed algorithm is higher.

Keywords: Cars' type recognition; Multi-features fusion; Ensemble learning

1. Introduction

With the development of artificial intelligence, the automobile production line is becoming more and more intelligent.The cars' type recognition part of automobile intelligent production line in shown in Figure 1. After the cars' type recognition is performed, the robotic arm wil carry out the operation of painting according to the result of cars' type recognition,so the automobile intelligent production line urgently needs a high accuracy and real-time cars' type recognition system, which can accurately identify a variety of cars' types.At present, the most of cars' type recognition systems adopt radio frequency recognition method, laser recognition method, infrared recognition method and visual recognition method. Among them, visual recognition methods can be divided into the traditional feature extraction with machine learning methods and deep learning image classification methods. Due to the high cost of radio frequency recognition method, serious pollution of laser recognition method and poor robustness of infrared recognition method, visual recognition method has become the main method. The cars' type recognition algorithm based on single feature extraction has low accuracy and poor universality. It can only recognize one single specificly car type and is easy to be disturbed by factors such as illumination and location. However, the cars' recognition system based on image classification using deep learning has disadvantages of high dependence on samples, poor interpretability and poor generalization ability, which is difficult to meet the expectation of industry.

Fig. 1. Automobile intelligent production line

Under the framework of ensemble learning, this paper designs a cars' type recognition system based on multi-features fusion using traditional computer vision, which can accurately recognize various types of cars. It has strong universality and has a strong application in automobile intelligent production line. The algorithm mainly takes SVM as the basic classifier and uses the optimized stacking algorithm to fuse features of the image so as to realize the universality of the algorithm and improve the robustness of the algorithm. The effectiveness of the algorithm proposed in this paper has been verified in an automobile intelligent production line.

2. Related Work

2.1. Feature Descriptor

Features are the characteristics of a certain type objects that are different from other types objects. For images, each type of image has its own characteristics that are different from other types images. Some features are observable, such as: brightness, edge, color, etc.; but some features are obtained through complex transformation and processing, such as: SIFT, ORB, BRISIK. SIFT means scale-invariant feature transformation. This method was proposed by David Lowe in ICCV in 1999 and imporved in five years. It was published in IJCV in 2004[4]. SIFT feature detection mainly includes four steps: detecting maximum points in different scale spaces, locating key points, determinating direction and generating descriptor. The SIFT is an algorithm based on some local feature points on the object, which maintains a certain degree of stability for radiation transformation, brightness changes, viewing angle changes, noise, etc.[5]; ORB which means oriented FAST and rotated BRIEF was proposed by Ethan Rublee et al. in 2011 in ICCV[6]. ORB feature descriptor combines FAST feature point detection and BRIEF feature descriptor. ORB feature detection is divided into two steps: extracting FAST key points and generating BRIEF descriptor. The biggest advantage of ORB is that the time required for feature extraction is relatively short. In addition, due to the use of the BRIEF feature descriptor, the ORB descriptor has the characteristics of scale-invariant and rotation-invariant; BRISK means binary robust invariant scalable keypoints. This method was proposed by Stefan Leutenegger and published on ICCV in 2011[7], BRISK feature detection mainly includes four steps: establishing scale space, supressing non-maximum value, locating key points based on subpixel precise and binary coding to generate descriptors. The BRISK feature has good robustness to noise.

Most automotive intelligent production lines are in a closed environment without sunlight, but there still exists interference from electric lights and other light sources.In addition, due to the difference of sensors used in automobile intelligent production line, there is an error of 1-2 cm about the position of automobile in the views every time. Considering the interference in automobile intelligent production lines, SIFT, ORB and BRISK are selected as the basic features of the multi-features fusion for cars' type recognition algorithm based on ensemble learning[8]. The advantages and disadvantages of the three features are shown in Table 1.

Table 1. Advantages and Disadvantages of Features.

Feature	Advantages	Disadvantages
SIFT	Stability for changing viewing angle and affine transformation and brightnes	Dependence on local pixel gradient and slow computing
ORB	Scale invariance, low feature dimension and fast computing speed	Recognition accuracy is not as good as SIFT
BRISK	Strong robustness to noise	High feature dimension

2.2. Ensemble Learning

Ensemble learning refers to a machine learning algorithm that completes a learning task by combining multiple classifiers. Ensemble learning can be used for classification problem, regression problem, abnormal point detection etc. The shadow of ensemble learning can be seen in many machine learning fields. The ensemble learning architecture is shown in Figure 2. Ensemble learning algorithms can be roughly divided into three categories: Boosting, Bagging and Stacking.

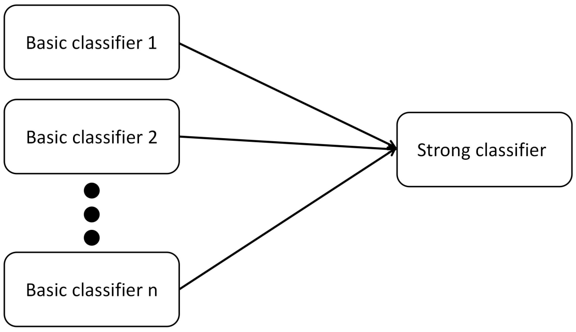

Fig. 2. Ensemble learning architecture

The basic idea of the Boosting algorithm is to firstly train a basic classifier from the initial training set, and then assign weight to the basic classifier according to the error rate of the basic classifier, and adjust the sample weight, so that the next classifier "pays more attention to" the samples that the previous classifier misclassified. This process is repeated until all classifiers are trained. The most famous algorithm in the Boosting algorithms is the AdaBoost algorithm. The original Adaboost algorithm can only be used for binary classification problems. Although Hui Zou later improved the Adaboost algorithm and extended it to multi-classification problems[9], because the Boosting algorithm works in a serial type, it is difficult to apply in actual industrial scenarios. It is difficlt to meet the real-time requirements and is very sensitive to abnormal samples which will get a higher weight in the iterative process, which will eventually affect the accuracy of the strong classifier.

The basic idea of the Bagging algorithm is to extract the same number of samples from the data set every time, using different samples to independently train classifiers, and finally basic classifiers votes to get the final result. If the basic classifier

Table 2. Advantages and Disadvantages of Features.

Method	Advantages	Disadvantages
Boosting	Assign weight to each classifier and make training error tend zeros	Work serially and require a lot of time and computational power combined with deep learning
Bagging	Fast and train parallelly	No emphasis between classifiers
Stacking	Assign weight to each classifier and train parallelly	Weights are constant

uses a decision tree, it is the famous random forest algorithm[10]. The advantage of Bagging algorithm is that it can work in the parallel type and make a decision using multiple classifiers. It has good real-time performance, but the combination method adopts simple voting or averaging method. There is no weight amgong basic classifiers. It can only rely stacking on the number of basic classifiers to achieve high accuracy.

The basic idea of the Stacking algorithm is to use the data set to train multiple classifiers independently, and then use the output of each basic classifier as the input to train the secondary classifier to obtain the final output of the model[11]. Generally speaking, logistic regression is usually used as a combination strategy. Although the Stacking method has the parallel characteristics of the Bagging method and use logistic regression can be used to integrate the results of each basic classifier,the weight given to the each basic classifier by logistic regression is a constant, and the model is relatively simple, which can not effectively integrate the classification results of the basic classifiers.

2.3. SENet

SENet was proposed by Momenta company and published on CVPR in 2017[12]. The network won the champion of the last Imagenet image recognition competition. Traditional deep neural networks usually aggregate the information of spatial dimension and feature dimension on the local receptive field to get the global information.SENet starts with the relationship between channels and hopes to model the dependencies between channels.Specifically, the importance of each feature map is automatically obtained through learning, so that useful features are enhanced and features that are not useful to the current task are suppressed so as to realize the adaptive calibration of between feature maps.

The main idea of the paper is to design the SE-block, as shown in Figure 3.SE-block contains two operations, Squeeze and Excitation. The Squeeze operation uses global average pooling to compress each feature map after obtaining multiple feature maps so as to compresses C feature maps into a $1 \times 1 \times C$ tensor.

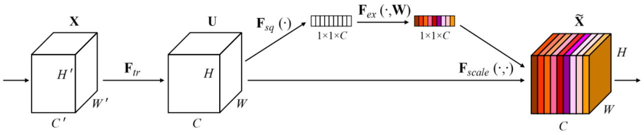

Fig. 3. SE-block structure

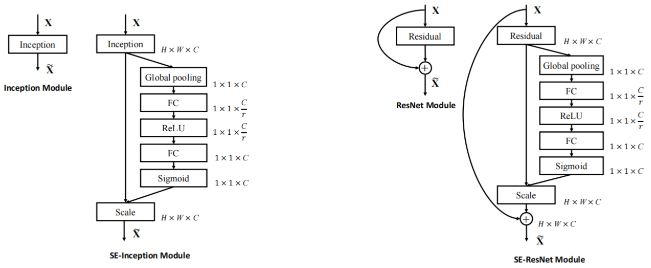

Fig. 4. SE-GoogleNet and SE-ResNet structure

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j), z_c \in R^C \qquad (1)$$

The Excitation operation obtains the weight of each channel. The paper introduces two fully connected layers, which is the dimensional reduction layer with the parameter W_1 and the dimensional increase layer with the parameter W_2. The activation function uses the ReLU function, and then the weight is normalized by the sigmoid function.

$$s = F_{ex}(z,W) = \sigma(g(z,W)) = \sigma(W_2(RELU(W_1 z)))(2)$$

Finally, the weight of each channel learned is multiplied by the corresponding feature map.

$$\widetilde{x}_c = F_{scale}(u_c, s_c) = u_c \bullet s_c \quad . \; . \; . \; . \; . \; . \; . \; . \qquad (3)$$

The whole block has learned the weight of each channel, so that each channel will be emphasized in subsequent operations, which can be considered as a kind of attention mechanism. SE-block can be flexibly applied to existing networks. Figure 4 shows the application of SE-block to the ResNet and GoogleNet.

2.4. SVM

SVM means support vector machine which is a linear classifier initially that performs classification task.It was proposed by V.N. Vapnik A.Y. Chervonenkis C. Cortes etc[13]. in 1964.The learning strategy of SVM is to maximize the interval, which can be formalized as a problem of solving convex quadratic programming, which is also equivalent to the problem of minimizing the regularized hinge loss function. The learning algorithm of SVM is the optimal algorithm for solving convex quadratic programming.Before deep learning, SVM was considered the best classifier.

Although ensemble learning combines multiple “weak classifiers” to obtain a “strong classifier” and many studies on ensemble learning are based on weak classifiers,in actual industrial applications, due to the constraints of cost, time, computing power and other conditions, it is still necessary to select a “strong classifier” as the basic classifiers for integration, so as to meet the high accuracy requirements of industry. As the optimal algorithm before the deep learning, SVM is suitable for lacking sanples, non-linear, high-dimensional classification problems and also has a wide range of applications in the industry. Therefore, SVM is selected as the basic classifier.

3. Model Design

Inspired by SE-block and traditional ensemble learning algorithms, a multi-features fusion algorithm based on ensemble learning is designed. The algorithm uses SE-block in the Stacking algorithm to improve the existing model of using logistic regression as a secondary learner,which is too simple to obtain enough information from the results of the basic classifiers. Compared with a single feature and using Logistic regression as a secondary classifier, this algorithm can effectively improve the classification accuracy.

3.1. Training Basic Classifier

The training process of basic classifiers is shown in Figure 5. When training the basic classifier, the input of the whole model is a training image and then using multiple feature extraction methods to extract image features from multiple angles to obtain different vectors.Finally, putting them into basic classifier, considering the detected features which mainly includes geometric shape and texture features, and at the same time, the methods should the have the characteristics of translation and illumination changes suppression. Therefore, three image features extraction methods SIFT, ORB, and BRISK are selected. The three vectors are separately sent to three SVMs for training, and each classifier is trained independently without interfering with each other.

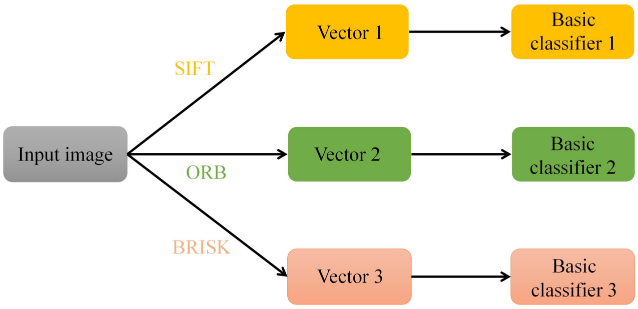

Fig. 5. Basic classifiers training structure

Due to the external lighting conditions and the positions of the car changes, the number of feature points extracted is not a constant. During the training of the basic classifiers, a stable number of feature points is selected, for example : we select 20 feature points which are sent to basic classifiers. If the number of detected feature points exceeds 20, the vector composed of the descriptors corresponding to the first 20 feature points is taken as the input of the basic classifiers; if the number of detected feature points is less than 20 ,the other positions of the descriptor are filled with zeros so as to make the length of the descriptor vector equal to the length of the descriptor vector corresponding to 20 feature points, and then it is used as the input vector of the basic classifier.

3.2. Training Secondary Classifiers

The training process of basic classifiers and the training process of the secondary classifier are serial. After the training process of basic classifiers is finished, the training process of secondary classifier begins. The input of the secondary classifier is the output of the basic classifiers and the output of basic classifiers is the label corresponding to each image. Therefore, the input of the secondary classifier is the label corresponding to each image. Before the process of training the secondary classifier, we save results of basic classifiers as a ”pt” file to the loacl device.When training the secondary classifier, we load the results of basic classifiers as the input of secondary classifier.

The traditional ensemble learning uses logistic regression as a secondary classifier, but the logistic regression is a relatively simple model, so the information obtained from the basic classifiers’ output is very limited. In order to obtain more information in the classification results from the basic classifiers, a more complex model must be used as the secondary classifier which can better mine the information from the results of basic classifiers.Influenced by SE-block to integrate the information of each channel by introducing a fully connected neural network , the fully connected neural network is introduced into the model as a secondary classifier for training, which integrates the results of each basic classifier and outputs the classification probability.The architecture of the secondary classifier are shown in Figure 6.The secondary classifier which is a fully connected neural network has five layers.The first layer has three nodes because we integrate three kinds of image feature which are SFIT,ORB and BRISK.The secondary,third,fourth layer have five,eight and five nodes respectively.The final layer has three nodes because our task is to recognition three different type cars.The dimension of input to secondary classifier is usually low because we need not so much features to

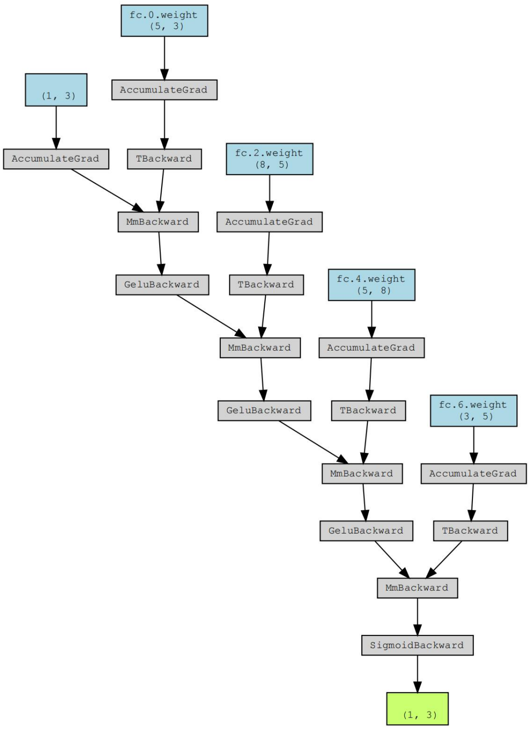

Fig. 6. Architecture of secondary classifier

describ the image and the dimension of output from secondary classifier is also low because the task of classification is usually simple in the industry. Besides,in theory, a neural network with two or more layers can fit arbitrarily complex functions,so we need not a lot of hidden layers to fit functions.Above all, the architecture of secondary classifier is simple and has a small amout of calculation,so the time required to train the secondary classifier is very short.

In the design of the fully connected neural network, the RELU function which is commonly used is not adopted as the activation function, because using the RELU function as the activation function will lose part of the information and the network is not easy to converge. Compared with the RELU function, the GELU function can retain some information and is easier for the network to converge, so the GELU function is used as the activation function in the network. The curve of GELU function is shown in Figure 7.

4. Experiment and Analysis

4.1. Data Base

In order to verify the validity of the model, the experiment is based on the problem of cars' type recognition and uses pictures collected in an automobile manufacturer as a data base for training and testing. The goal of the experiment is to distinguish the three types of car which are shown in Figure 8 and label them with type 0, type 1, and type 2 respectively.

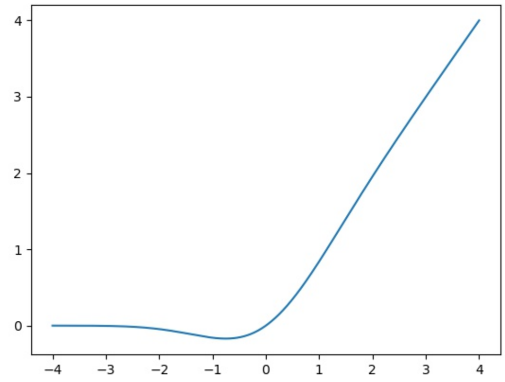

Fig. 7. Curve of GELU function

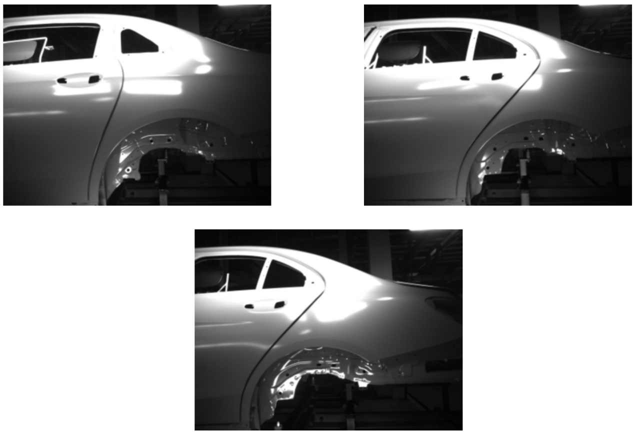

Fig. 8. Car example

In the model training process, the data base is divided into three equal parts. The first part is used to train the basic classifiers, the second part is used to train the secondary classifier and the third part is used as the test set for testing.

4.2. Experiment settings

Considering that SIFT has robustness to changing of lighting,ORB has scale invariance and BRISK has strong robustness to noise,so we choose them as the basic features.About experiments settings,we select SIFT wtit SVM,ORB with SVM and BRISK with SVM as baseline algorithms because we want to prove the accuracy using multi-features is higher than the accuracy using a single feature,we select SFIT, ORB ,BRISK with SVM and logistic regression as the another baseline algorithm beacuse we want to prove the accuracy of using a fully connected neural network as a secondary classifier is higher than using a logistic regression as a secondary classifier.

Table 3. Experiment results

Indx	Method	Classifier	Accuracy		
			Type 0	Type 1	Type 2
1	SIFT	SVM	66.51%(838/1260)	95.49%(1505/1576)	91.57%(532/581)
2	ORB	SVM	94.52%(1191/1260)	84.58%(1333/1576)	91.57%(532/581)
3	BRISK	SVM	42.98%(534/1260)	1.4%(22/1576)	80.21%(466/581)
4	SIFT ORB BIRSK	SVM LR	64.76%(816/1260)	95.62%(1507/1576)	89.85%(522/581)
5	SIFT ORB BIRSK	SVM MLP	98.25%(1238/1260)	98.48%(1552/1576)	100%(581/581)

4.3. Results and Analysis

According to the previously described model, using the three kind of features of SIFT, ORB, and BRISK as the basic features for feature fusion.The proposed multi-features fusion model is compared with the model which uses single feature directly sent to the basic classifier for classification. The experimental results is shown in Table 3.

From the comparison of experimental data in Table 3,comparing experiments 1,2,3,5, it can be seen that after using a variety of features fusion, whether the car type is 0, 1, or 2, the accuracy of the recognition accuracy with three features is higher than the recognition accuracy with any single feature.This shows that using the three features at the same time will not cause conflicts in the model. The three features complement each other and jointly improve the performance of the model.Comparing experiments 4 and 5,it can be seen that the accuracy of recognition has been improved using fully connected neural network as the secondary classifier ,whether the car type is 0,1 or 2,compared with using logistic regression as the classifier.This result proves that using fully connected neural network as the secondary classifier can obtin more information from results of basic classifiers to improve the recognition accuracy.

4.4. Activation Function

In this model, the activation function used is the GELU function instead of the traditional RELU function. In the model training process, the corresponding loss curve when using the GELU and RELU functions respectively is shown in the Figure 9.

By comparing the loss curve when the GELU function and the RELU function are used as the activation function, it can be seen that under the premise of the same number of training epoches, although the use of the RELU function as the activation function can temporarily make the loss curve of the model drop faster, but compared with the GELU function, the GELU function can further reduce the loss of the model, avoiding loss oscillations and non-convergence of the model.But why GELU function help the network to converge,this job wil be done in the future.

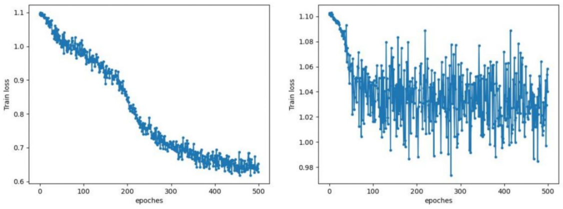

Fig. 9. Loss curve using GELU function(left) and RELU function(right)

5. Conclusion

Based on ensemble learning, this paper improves the stacking technology ,which introduces neural network as a secondary classifier into traditional ensemble learning to integrate the output of basic classifiers so as to propose a multi-features fusion algorithm based on ensemble learning and selectes cars' type recognition as the background to carry out model verification.Through comparative experiments,it can be seen that after using multi-features fusion,the classification accuracy of each category has been improved.The experimental results fully prove the rationality and effectiveness of the model.

In the future,we will further use the fully connected neural network for the interaction in the feature to improve the accuracy of cars' type recognition and explore the effect of activation function for network converge.

References:

[1] Heikki Huttunen, Car Type Recognition with Deep Neural Networks[J]. IEEE Intelligent Vehicles Symposium,2016

[2] Erxi Zhu, Vehicle Type Recognition Algorithm Based on Improved Network in Network[J]. Hindawi,2021

[3] Jitian Wang, Vehicle Type Recognition in Surveillance Images From Labeled Web-Nature Data Using Deep Transfer Learning[J].IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS,2018.

[4] David G. Lowe, Object Recognition from Local Scale-Invariant Features[J].ICCV,1999

[5] Mingyu Qiao, Improved SIFT algorithm based on image filtering[J].ICAACE,2021

[6] Ethan Rublee, ORB: an efficient alternative to SIFT or SURF[J].ICCV,2011.

[7] Stefan Leutenegger, BRISK: Binary Robust Invariant Scalable Keypoints[J].ICCV,2011.

[8] A Setiawan, Comparison of Speeded-Up Robust Feature (SURF) and Oriented FAST and Rotated BRIEF (ORB) Methods in Identifying Museum Objects Using Low Light Intensity Images[J].ICoSITeR,2019.

[9] Hui Zou, Multi-class AdaBoost[J].Statistics and Its Interface,2006.

[10] Breiman L, Random Forests[J].Machine Learning, 2001.

[11] David H. Wolpert, Stacked Generalization[J].Neural Networks,1992.

[12] Jie Hu, Squeeze-and-Excitation Networks[J].CVPR,2017.

[13] Johan A.K. Suykens, Support Vector Machines: A Nonlinear Modelling and Control Perspective[J]. European Journal of Control,2001.