An Ensemble Classification Algorithm for Convolutional Neural Network based on AdaBoost

Shuo Yang, Li-Fang Chen, Tao Yan, Yun-Hao Zhao, Ye-Jia Fan
(School of Digital Media, Jiangnan University, Wuxi 214122, China)

***Abstract*—AdaBoost is a classic ensemble learning algorithm with good classifier performance. In the past, it mainly used weak classifier as base classifier, such as KNN. They are simple and easy to train, but the essence of the weak classifier, it is impossible to get very high classification accuracy. In order to improve the correct rate, this paper introduces the AdaBoost ensemble classifier based on convolutional neural network, namely adaBoost-CNN, referred to as ACNN. ACNN design a new training method, it not only gives the weight of base classifier according to the error rate of base classifier in pre-training phase, but also dynamically adjusts this weight and learning coefficient of training sample according to the error rate of each class in ensemble training phase. Finally, through experiments on some public datasets, it was proved that ACNN not only can effectively reduce the classification error rate, but also can solve the problem of class recognition rate imbalanced caused by similar categories or training samples quantity deviation.**

Keywords—deep learning; convolution neural network; ensemble learning; adaBoost; classification

I. Introduction

As a basic problem in machine learning, classification problem experienced a long period of research and development, generated a lot of solutions such as KNN (k-NearestNeighbor), Decision Tree, and Naive Bayesian. Those traditional methods can achieve a generally good classification accuracy rate, but also have several problems: First, they need artificial extraction features that create the uncertainty of features, which affects the accuracy rate. Second, these classifiers cannot get a very high classification accuracy, so called weak classifier. Later, the rise of ensemble learning such as adaBoost which integrates multiple weak classifiers for voting with weights provides conditions to solve these problems, adaBoost improves the classification accuracy rate to a certain extent. But it is limited by the accuracy of weak classifier, adaBoost still cannot get enough accuracy to meet industry requirements.

The emergence of CNN(convolution neural network) solves these problems. LeNet network proposed by Yan LeCun et al. [1] succeed in raising the recognition rate of handwritten fonts to over 98% and was used in the US check handwriting recognition widely. However, due to the limited speed of hardware equipment at that time, and the large computation and over-fitting of CNN, CNN gradually faded out of people's vision. Until 2006, Professor Hinton [2] present two important points in their paper: First, the multi-hidden neural network can learn attributes with the nature of data and is benefit to data visualization and classification. Second, we can overcome the difficulty of deep neural network in training by means of unsupervised "layer-by-layer initialization" strategy, which solves the problem of difficult training in neural network. In 2012, Professor Hinton [3] designed the Alexnet network for image classification. Alexnet use some new layers like ReLU, Dropout and LRN normalization to prevent over-fitting. In 2014, Google team [4] designed a deeper network model GooLenet (22 layers), and used the multi-scale data training methods, got the first place in the ILSVRC 2014 classification project, Top5 test error rate dropped to 6.66%. Since then, image classification has basically become the world of convolution neural network.

In recent years, the study of classification problem with CNN is mainly divided into two directions. One is to design a deeper network structure to study the effect of depth on classification. However, the deeper network means the more severe gradient disappearance. In 2015, He K et al. [5] did an analysis about just deepening network will make the training error raise, they introduced residual to CNN to promote network faster and more stable convergence, and designed a 152-layer network to assess the residual network. In 2016, Gao H et al. [6] made improvements based on previous paper, which increase residual network model to more than 1,200 layers by dropping some layers randomly, these improvements significantly enhance accuracy rate and training speed.

Another direction is to optimize network structure and integrate the output of different network structures or different training methods as the final result. Tang Y et al. [7] aiming at the problem that Support Vector Machine (SVM) will fail on ensemble learning (Bagging), put forward a kind of selective SVM ensemble learning algorithm (SE-SVM). SE-SVM select a set of component SVMs with high generalization performance and high diversity during ensemble through recursive elimination algorithm to classification. That effectively improve correct rate. Cang W et al. [8]aiming the problem that traditional Stochastic Gradient Boosting(SGB) highly correlate with the shrinkage parameters, engender over-fitting and sometimes cannot obtain a satisfactory generalization performance, they use Gaussian process regression(GPR) as base classifier, propose a modified SGB ensemble learning approach, this approach improve accuracy and the over-fitting problem of SGB.

These researches were selected simple classifier as KNN, SVM and Gauss process regression. This paper constructs an ensemble network model named adaBoost-CNN(ACNN) based on classical CNN model(Lenet). ACNN presents a new characteristic named class adaptability, and design a new base class weight assignment method based on it. This paper also designs a new training method that ACNN not only training each base classifier alone in pre-training phase as adaBoost, but also iterative training ensemble network in ensemble training phase. ACNN will adjust base classifier weight (classifier’s weight when voting, in ACNN, it is especially means the weight of each class) and sample learning coefficient (sample’s coefficient when training) after epoch (each training sample train once) training. That make ACNN’s training constantly tend to those samples that always be classification error and class that has lower accuracy, effectively reduce the classification error rate.

978-1-5090-5507-4/17/$31.00 ©2017 IEEE
ICIS 2017, May 24-26, 2017, Wuhan, China

When certain class lack of distinguishing features from other classes or the number of training samples was significantly less than others, training results will tend to other classes, it leads to a very uneven class recognition rate. This problem is called class recognition rate imbalanced. Zhou Yu et al. [9] propose a data selection method based on shadow set, and raised the concept of kernel data and boundary data. Firstly, they obtained the optimal fuzzy matrices of sample data by fuzzy C-Means Clustering (FCM), and then induced the corresponding shadow sets. Finally, they selected the data using kernel data and boundary data which are composed by sample data and shadow sets. This method effectively retains the typical sample, improve the accuracy of the class with high error rates. Li Y et al. [10] propose a new classification method to solve this problem by feature selection and ensemble learning, called BPSO-Adaboost-KNN algorithm. This algorithm improves the stability of adaBoost by extracting key features, increases accuracy rate by about 20%-40% compared with KNN classifier alone.

AdaBoost can solve this problem by adjusting sample learning coefficient. This is proved in [11][12], but because of CNN's special feature extraction method, it has not been found that this characteristics of adaBoost is also applicable to CNN. Therefore, this paper also explores this problem and proves that ACNN can be used as a solution to this problem in the experimental stage.

II. BASIC THEORY

A. *Convolution Neural Network*

Convolution neural network is a single scale structure composed of multi-layer feature extraction stage and classifier. that is, the input is extracted by multi-layer to get features of higher level, then put these features into classifier [13]. Feature extraction phase mainly includes convolution layer and pooling layer, then connect one or two layers of full-connected layer to classification. Taking Lenet as an example, this model enters image sample with pixel size of 32 * 32, after two convolution layers which have 6 and 16 kernels that size is 5*5 and two pooling layers, extracts to 16 characteristics that size is 5*5 pixel, and then outputs the classification results through two full-connection layers.

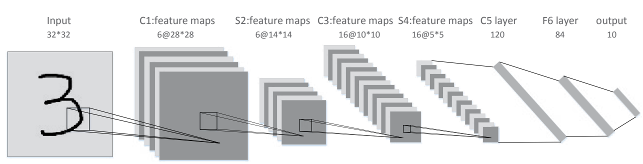


Fig. 1. Lenet network model

In the follow-up study of CNN, scholars have designed more networks such as AlexNet, VGGNet, GoogleNet and ResNet on the basis of Lenet. They have designed many new activation function and regularization in order to solve the problem of gradient disappearance and over-fitting. These methods improved the classification accuracy, and widely used in such as target classification, target detection and semantic segmentation and many other areas.

B. *adaBoost Algorithm*

The basic idea of adaBoost is to obtain classification result by calculating the output of base classifier with weight [14]. The main process is: first, select a classifier and data set($S = \{x_i, y_i\}, i = 1,2,\dots,m$, x_i represent image samples, y_i represents the class label for corresponding sample),then set a weight D and initialize D to 1/m for each sample, then train this classifier for T iterations, each iteration will update the sample learning coefficient based on training results , the basic rule is samples that classification error are given a larger weight, classifier will focus on these samples on next iteration. Each iteration obtains a base classifier trained by different sample learning coefficient. After T iterations, adaBoost obtains T base classifiers, then adaBoost obtains the final output by voting on these base classifiers [15].

TABLE I. ADABOOST ALGORITHM DESCRIPTION

Algorithm Name: adaBoost classification algorithm
Input:datastS $= \{x_i, y_i\}, i = 1,2,\dots,m, y_i \epsilon Y, Y = \{c^1, c^2, \dots, c^k\}, c^k$ is label, T is the number of iterations, I is a classifier, Z_t is normalization factor that make sure D_{t+1} is a distribution. 1.initialize sample learning coefficient $D_1(x) = 1/m$; 2.for t = 1,2,…,T 3.train classifierI(S, D_t), get a weak assumption $h_t = X \to \{c_1, c_2, \dots, c_k\}$; 4.calculate the classification error rate $e_t = \sum_{i=1}^{m} D(i)[y_i \neq h_t(x_i)]$; 5.if $e_t > 0.5$ then break 6.classifier weight $\alpha_t = \frac{1}{2}\ln(\frac{1-e_t}{e_t})$; 7.update sample learning coefficient $D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} \exp(-\alpha_t), & if\ y_i = h_t(x_i) \\ \exp(-\alpha_t), & if\ y_i \neq h_t(x_i) \end{cases}$; 8.end 9.normalized classifier weight $s_t = \frac{\alpha_t}{\sum_1^T \alpha_x}$;
Output:H(x) $= sign(\sum_{t=1}^{T} s_t h_t(x))$

III. ABOOST-CNN ENSEMBLE NETWORK CLASSIFIER

ACNN's structure is as follows (contain 3 base classifiers, 4 classes):

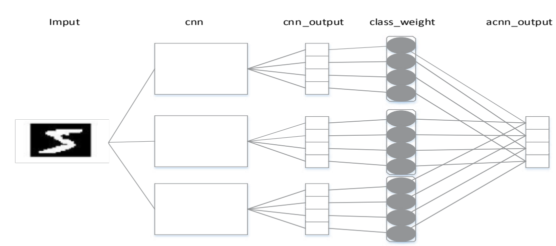


Fig. 2. Structure of ACNN

ACNN obtain different base classifiers by training on different sample learning coefficient, then obtains the final output by voting by them.

$$H = \sum_{1}^{n} s_t \odot h_t \quad (1)$$

H represents the output of the ensemble network, n represents the serial number of base classifiers, s_t represents CNN_t's vote weight for each class, $h_t$represents CNN_t's output, $\odot$ represents dot product operation.

A. *Class Adaptability and New Class Weight Assignment Method.*

In this paper, we find that CNN have a special adaptability to different class for the same dataset. This adaptive performance that some classes' accuracy rate is particularly good, we call it network strong class, but other classes' accuracy rate is not good, we call it network weak class. This may be due to the lack of significant different characteristics between certain classes, or the excessive deviation of the sample number of different classes. Although this adaptability determined by network itself, but it can also be influenced by dataset or training method.

Taking Lenet training on MNIST(experimental section has an introduction to this dataset) to total error rate reach 20%, the classification error rate for each class are as follows:

TABLE II. EACH CLASS ERROR RATE (%)

Class	1	2	3	4	5	6	7	8	9	10
Error Rate	4.39	4.93	19.67	16.83	20.78	48.77	11.17	14.11	27.62	30.53

We can see , 6 class's error rate is significantly higher than other classes (which may be handwritten font 2 and 5 are too similar), much higher than 1,2 class, we think that the network strong class are 1 class and 2 class, the network weak class are 6 class and 10 class.

The first innovation on ACNN is a new class weight assignment method. In view of the class adaptability and the characteristic of CNN, We no longer assign weight to each base classifier as adaBoost, but assign weight to each base classifier's each class. These weights are calculated by each base classifier's error rate for each class, then make the sum of weights of all base classifier for each class become 1 by normalize these weights.

1) Calculate the weight of each base classifier for each class based on the error rate:

$$\alpha_{tc}=\frac{1}{2}\ln(\frac{1-e_{tc}}{e_{tc}}) \quad t=1...n, c=1...m \tag{2}$$

e_{tc} represents the error rate that CNN_t to c class, n is the total number of base classifiers, and m is the total number of class.

2) Then normalize these weights:

$$s_{tc}=\frac{\alpha_{tc}}{\sum_1^n \alpha_{tc}} \quad t=1...n, \ c=1..m \tag{3}$$

α_{tc} represents the weight of CNN_t to c class, s_{tc} is the result of normalization. Thus, each base classifier get a set of weights of voting for each class.

B. Change Sample Learning Coefficient

When samples or classes are always classified error, we hope that training can appropriate bias these, which can be achieved by changing sample learning coefficient. This change is similar to adaBoost. The difference is weight is no longer determined by the base classifier's unique weight, but by base classifier on each class's weight.

$$\mathrm{D}_{t+1}(i)=\frac{\mathrm{D}_t(i)}{Z_t}\times\begin{cases}\exp(-\alpha_{tc}) & if\ h(i)=f(i)\\ \exp(\alpha_{tc}) & if\ h(i)\neq f(i)\end{cases} \tag{4}$$

$\mathrm{D}_t(i)$ represents the learning coefficient of i sample in t iteration, $\mathrm{D}_{t+1}(i)$represents this coefficient in next iteration, h(i) represents the classification result of classifier, f(i)represents this sample's label, $Z_t$is the normalization factor to ensure that D_{t+1} is a distribution, just like divided by average.

Network weak class get more fully learning. After changing the sample learning coefficient, do the experiment again (training to the same total error rate). The error rate for each class is shown below:

TABLE III. RESULT ON SAMPLE LEARNING COEFFICIENT CHANGED(%)

Class	1	2	3	4	5	6	7	8	9	10
Error Rate	11.02	24.3	13.76	17.43	16.09	22.87	5.32	14.30	11.91	17.44

Compared with Table 2, the classification error rate for 6 class and 10 class are significantly reduced. Although 6 class is still the highest error rate, but it is clearly that these error rates are more balanced. ACNN will change the sample learning coefficient again according to the result of this time. In the following training, 6 class or other classes' simple with high error rate will get higher weight to reduce error rate. And the sample learning coefficient is calculated by the current state of the network (error rate of ACNN for each class), which is adaptively adjusted and more appropriate than the artificial threshold or simply increasing or reducing the number of samples per class.

AdaBoost's makes training tend to samples that always be classified error. Our paper introduces class adaptability and a new class weight assignment method, that makes training not only tend to samples that always be classified error, but also tend to classes with high classification error rate.

C. New Training Method

Because of CNN's slow convergence speed and easy to over-fitting, traditional train method is not suitable for CNN. It have several problems: First, the training process is linear, only after getting KNN_i will start training KNN_{i+1},and KNN_i won't change once it get, if KNN_i no complete convergence, it must increase the ensemble network's error rate. Second, base classifier's training is independent, the information transfer is one-way, training result of previous classifier affects sample learning coefficient of next classifier, but the next classifier's result cannot affect the previous base classifier, that's why adaBoost should ensemble a large number of base classifiers, adaBoost need enough base classifier to change sample's learning coefficient, this change makes next classifier's train have more purposeful (samples that always be classification error); Third, training method is not the same as the final vote method(each base classifier is independent when it is trained, but they are not independent when they voting). Actually, adaBoost only train each base classifier, but not ensemble network, this method cannot guarantee that ensemble classifier's error rate will be reduced when base classifier error rate is reduce by training constantly.

This paper design a new training method. Training is divided into pre-training phase and ensemble training phase.

1) Pre-training Phase

The purpose of pre-training phase is to obtain some base classifier with different vote weights. We hope that each base classifier has a different strong class by introducing class weights assignment method. The end condition of each training is each class's error rate less than 0.5. That can reach after several epochs.

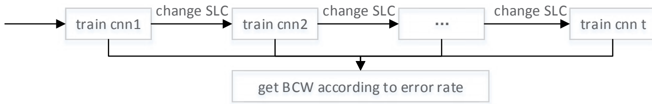


Fig. 3. Pre-training phase

SLC represent sample learning coefficient, BCW represent base classifier weight, same as the following.

2) Ensemble Training Phase

This phase make an iterative training of ACNN, each iteration(each iteration is an epoch) training change SLC by ACNN's error rate of each class and BCW by each base classifiers error rate of each class once that used on next iteration. When finish the specified number of iteration, over the training.

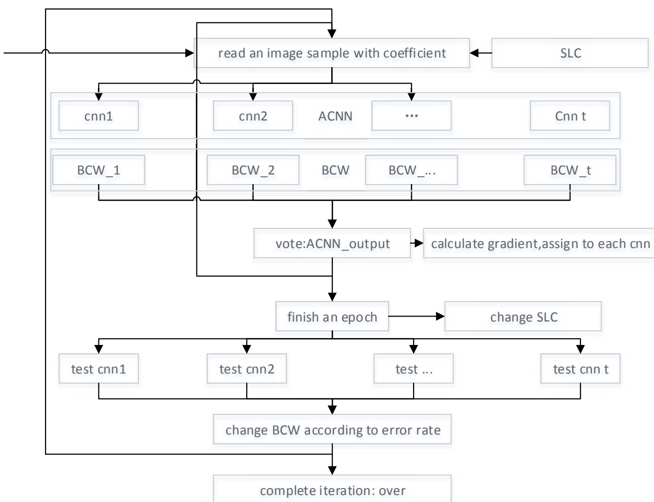


Fig. 4. Ensemble Training phase

There are several advantages to this type of training: First, base classifiers are no longer determined by once training but constant dynamic adjustment, which makes base classifiers and samples obtain a dynamic weight, training process making these weights more reasonable; Second, in ensemble training phase, classifier's training is non-independent, each base classifier has affected gradient acquirement of other classifiers, the following base classifier's training results also affect the previous base classifier in the next iteration training. Third, because of the training of each classifier is iterative constantly, only after the end of the ensemble classifier training, base classifier training will end, so we don't need to decide the number of training iterations before training; Fourth, ACNN has trained the ensemble classifier, although it is still trained by each base classifier's BP, but the ensemble classifier determines the gradient distribution and guides the training tend to lower error rate of ensemble classifier, not lower error rate of base classifier. This method ensures that ensemble classifier can be getting lower and lower error rates by training.

IV. EXPERIMENTS AND RESULT ANALYSIS

A. Experimental Dataset

Experiments select two common handwritten digital datasets (MNIST and USPS) and a physical dataset (ETH-80). Each dataset is described as follows:

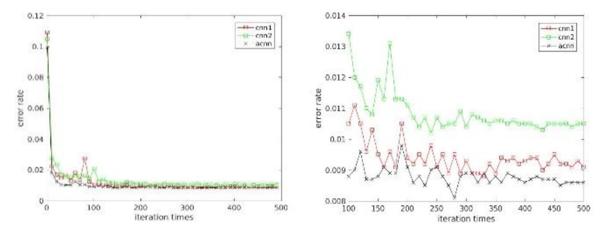
TABLE IV. DISTRIBUTION OF DATASET

Name	Class number	Training/Test samples number	Training / Test sample distribution
MNIST	10	60000/10000	5923 6742 5958 6131 5842 5421 5918 6265 5851 5949/980 1135 1032 1010 982 892 958 1028 974 1009
ETH-80	8	2400/800	300 300 300 300 300 300 300 300/100 100 100 100 100 100 100 100
USPS	10	7200/2000	1174 994 727 636 646 549 655 643 538 638/357 263 198 165 198 160 170 147 166 176

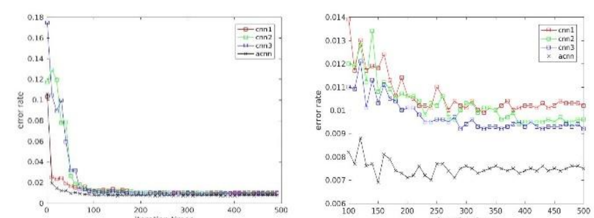
B. Verification of Validity of ACNN

Because CNN's accuracy rate is high enough but the convergence rate is too slow, these experiments cannot and doesn't need to integrate a large number of base classifiers as [11]. Experiments first design different ACNN with 2, 3, 4, 6, 8 and 10 base classifiers (Lenet) training on MNIST dataset, the result is shown in figure 5. Polyline cnn1, cnn2, …, cnn* represents the error rate of * base classifier, polyline acnn represents the error rate of the ACNN.

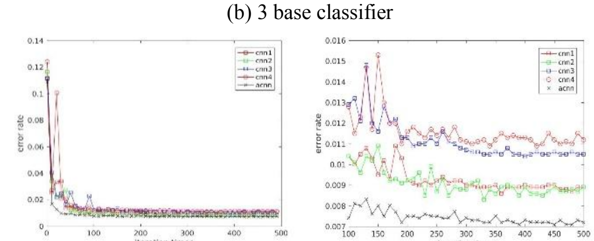
Since MNIST is very simple, CNN can achieve a very low error rate after several iterations. To clarify the difference between these classifiers, we intercept iterative results of 100 to 500 times to shown on the right side of the figure below:



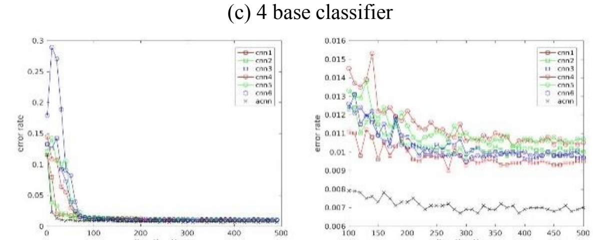
(a) 2 base classifier



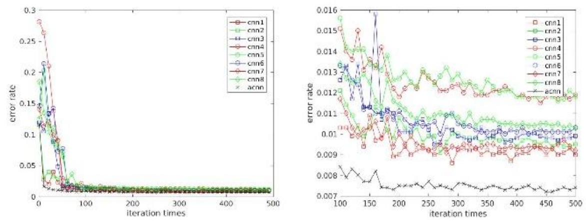
(b) 3 base classifier



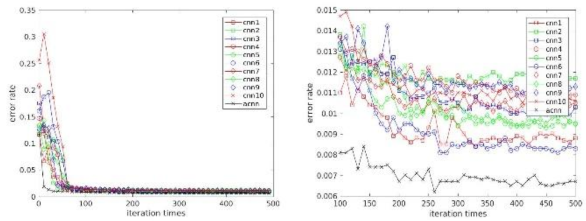
(c) 4 base classifier



(d) 6 base classifier



(e) 8 base classifier



(f) 10 base classifier

Fig. 5. Error rate of ACNN with different number of base classifiers

As shown in above figures, error rate of ACNN (both black line) is always lower than all base classifiers whether it has several base classifiers, it is proved that ACNN can effectively reduce the error rate.

C. The Suitable Number of Base Classifier

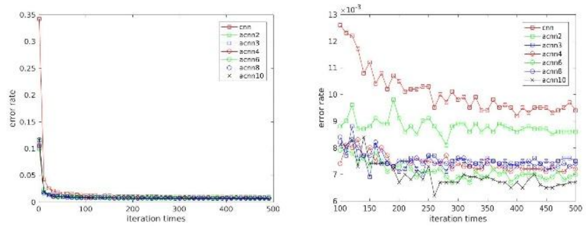


Fig. 6. Comparison of ACNN that have Different number of base Classifiers

As shown in above figure (acnn* represents an ACNN that have * base classifiers), When the number of base classifier is increasing, the error rate is declining. However, when the number reaches 6, it is not always possible to achieve better results by increasing number of base classifiers. acnn8's error rate is higher than acnn6, which may be due to ACNN's training method. In ensemble training phase, ACNN assigns gradient to each base classifier according to each base classifier's vote weight. When the number of base classifier increases, each base classifier will get a less gradient, even when the number of base classifier is more than the number of class, if a base classifier cannot gain an advantage in the classification of any class, that is, there is no greater weight, it will get a very low gradient and inadequate training. Anyway, it will affect the final output.

Since ACNN's training process is actually done by each base classifier, the training time of ACNN is proportional to the number of base classifier. Although the error rate of acnn10 is slightly lower than acnn6, but acnn10 need nearly two times as much as acnn6. We think that spending double time does not match the effect of the improvement, so we compare the 3 constructed classifiers with several DBN models in the next experiment to verify ACNN's performance in several major datasets. Introduction is as follows:

TABLE V. 3 CONSTRUCTED CLASSIFIERS

Name	Introduction
cnn	Lenet network trained in traditional training method
acnn4	An ensemble classifier with 4 Lenet base classifiers
acnn6	An ensemble classifier with 6 Lenet base classifiers

In addition, in order to more intuitively show that ACNN's class recognition rate is more balanced, experiments added class variance(CV)that represents the relationship between recognition rate of different classes. The variance is smaller, the class recognition rate is more balanced, the mathematical representation of the following:

$$\mathrm{CV} = \sum_{t=1}^{T} (e_t - \frac{\sum_{t=1}^{T} e_t}{T})^2 \quad (5)$$

D. ACNN Performance on Each Dataset

We design the same number of iterations and the learning rate descent function to several classifier structures mentioned above. The recognition error rate is as follows (The data is the mean of the last 5 iterations, same as the following).

TABLE VI. ERROR RATE OF 3 CLASSIFIERS ON DIFFERENT DATASETS(%)

Model	MNIST	ETH-80	USPS
cnn	0.997	10.913	4.6
acnn4	0.721	6.875	4.2
acnn6	0.696	6.25	4.125

ACNN obtains base networks with different network strong classes by using new training method of this paper, achieves a significant decrease in error rate. Because CNN has been able to get enough good recognition accuracy, so the decreased on MNIST and USPS is not big. In contrast, samples in ETH-80 dataset have three-dimensional rotation of space, more different color, posture and illumination in the same kind, which is more complicated than the handwritten digital dataset. The recognition rate of the traditional CNN is not high, ACNN decrease the error rate by more than 4%.

In summary, we can consider that ensemble learning method like adaBoost is not only applicable to weak classifiers such as KNN, but also can integrate strong classifiers such as CNN into stronger classifiers. Even dataset like MNIST has been able to get a very high accuracy rate by CNN, ACNN is still able to bring better results. In addition, because of CNN's slow training speed and low error rate, integrating a large number of base classifier is unreasonable, generally only need 4 or 6 base classifier can get a good upgrade effect.

E. Class Recognition Rate Imbalanced Problem

Class recognition rate imbalanced problem refers to the fact that certain classes have a high degree of similarity, or the number of samples of different classes in training set has a large deviation, which cause the result tend to those classes that have more significant features or more train samples, leads to a very uneven class recognition rate. It is mainly manifested that the recognition rate of some special classes is still very bad when the overall recognition rate is good. This is very common in engineering applications, such as in the road sign recognition, due to some type of sign appear less, the number of samples of these classes is small; or deep learning for medical systems like case analysis, the results tend to those common diseases. However, it cannot be said that those rare diseases are not important.

ACNN's new weight assignment method makes training tend to those classes have low accuracy rate, new training method makes training results more balanced, and the recognition rate of each class is closer. After fully training on MNIST, ETH-80, and USPS, different structures' final error rate of each class are as follows:

TABLE VII. ERROR RATE FOR EACH CLASS ON DIFFERENT DATASETS(%)

(a) MNIST

Model	1	2	3	4	5	6	7	8	9	10	ER	CV
cnn	0.306	0.449	0.824	1.119	1.059	1.547	1.514	0.875	0.852	1.566	0.997	0.1756
acnn4	0.306	0.308	0.494	0.693	0.815	0.874	1.044	0.953	0.739	1.06	0.721	0.0587
acnn6	0.306	0.467	0.484	0.693	0.642	0.886	0.093	0.875	0.606	1.11	0.696	0.0863

(b) ETH-80

Model	1	2	3	4	5	6	7	8	ER	CV
cnn	7	9	12	7	19	33	0	0	10.913	103.3608
acnn4	1	7	5	7	9	26	1	1	6.875	59.6719
acnn6	1	4	5	6	9	25	1	1	6.25	56.0625

(c) USPS

Model	1	2	3	4	5	6	7	8	9	10	ER	CV
cnn	1.681	3.042	6.061	7.879	6.566	6.25	4.706	6.803	4.819	2.273	4.6	4.0749
acnn4	1.681	3.422	4.545	7.879	6.061	6.25	3.529	6.122	4.217	1.705	4.2	3.8640
acnn6	1.961	3.042	5.051	9.697	6.566	5	3.529	4.762	2.711	1.705	4.125	5.3101

It can be seen that ACNN has achieved a more balanced class error rate on MNIST, ETH-80 and USPS, especially in MNIST (6 class due to lack distinguishing features from other classes, still have a high error rate) and USPS (class 1 has 1174, class 7 has only 538 samples).

In summary, ACNN can be used as a solution to the class recognition rate imbalanced problem, but ACNN is designed to reduce classification error rate more. So when we need to solve the problem only for this, we do not recommend to choose ACNN. Its training time is much more than a single CNN, but the effect is limited.

V. CONCLUSION

This paper design an ensemble network classifier named ACNN that have new class weight assignment method and training method. ACNN can effectively reduce the classification error rate. In addition, this paper discuss the most reasonable number of base classifiers which constitute ACNN. According to the performance of ACNN on MNIST with 2, 3, 4, 6, 8, and 10 base CNN classifiers, come to conclusion that the most suitable number is 4 or 6 combining error rate and training time. In addition, this paper designs a new weight assignment method makes training tend to those classes that have higher error rate, this tendency makes ACNN can solve the problem of the class recognition rate imbalanced to a certain extent and proves it in experiments.

This paper also have several problems, one is due to the huge computational complexity of CNN and limit of personal computer performance (which is the most important factor once hindered the development of CNN). We choose several simple datasets, the size of sample (image) is 32 * 32 pixels, the number of class is 8 or 10. CNN has a good effect on them, so promotion by ACNN is not obvious. Second, in this paper, the purpose of ACNN is to reduce the classification error rate, not specifically the class recognition rate imbalanced problem. Although ACNN can achieve a more balanced result, but it is not recommended as the main solution, because it's training time is much larger than a single CNN.

ACKNOWLEDGMENT

This work is supported by National Science and technology support program of China (No.2015BAH54F01).

REFERENCES

[1] Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.

[2] Hinton G E, Salakhutdinov R R. Reducing the Dimensionality of Data with Neural Networks[J]. Science, 2006, 313(5786):504-7.

[3] Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks [J]. Advances in Neural Information Processing Systems, 2012, 25(2):2012.

[4] Szegedy C, Liu W, Jia Y, Sermanet P. Going deeper with convolutions[J]. 2014:1-9.

[5] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition[J]. Computer Science, 2015.

[6] Huang G, Sun Y, Liu Z, Sedra D, Weinberger K Q. Deep Networks with Stochastic Depth[J]. 2016.

[7] Tang Y, Gao J, Bao Q. Novel Selective Support Vertor Machine Ensemble Learning Algorithm[J]. Journal of Xian Jiaotong University. 2008, 42(10):1221-1225.

[8] Cang W, Yang H. A Soft Sensor Modeling Method based on Modified Stochastic Gradient Boosting. CIESC Journal, 2016.

[9] Zhou Y, Zhu. Sample data selection method for neural network classifiers[J]. Journal of Huazhong University of Science & Technology, 2012.

[10] Li Y,Guo H,Liu X. A boosting based ensemble learning algorithm in imbalanced data classification[J]. Systems Engineering-Theory&Pratice, 2016(1):189-199.

[11] Li Q, Mao Y, Wang Z. Research on Boosting-based Imbalanced Data Classification[J]. Computer Science, 2011, 38(12):224-228.

[12] Qin M, Qiu J, Lu Z, Chen L, Zhao W. AdaBoost-based class imbalabce learing algorithm[J]. Application Research of Computers, 2017,(11):1-8.

[13] Jiao L C,Yang S Y,Liu F, Wang S G, Feng Z X. Seventy Years Beyongd Neural Networks:Retrospect and Prospect[J]. Chinese Journal of Conputers, 2016, 39(8):1697-1716.

[14] Kearns M, Valiant L. Cryptographic limitations on learning boolean formulae and finite automata. Journal of the ACM, 1994, 41(1): 67-95

[15] CAO Ying, MIAO Qi-Guang, LIU Jia-Chen, GAO Lin. Advance and Prospects of AdaBoost Algorithm[j]. acta automatica sinica, 2013,(06):745-758.