The 33rd Youth Academic Annual Conference of Chinese Association of Automation (YAC)
May 18-20, 2018, Nanjing, China

Image Classification with Category Centers in Class Imbalance Situation

Yulu Zhang, Liguo Shuai, Yali Ren and Huilin Chen
School of Mechanical Engineering, Southeast University

Nanjing, China
220150271@seu.edu.cn

***Abstract*—In recent years, deep convolutional networks have made a milestone progress in the field of image recognition. However, the recognition ability of the deep convolution network declines in the case of unbalanced training data, in terms of categories with fewer training images, the recognition ability of the deep convolution network declines more seriously. Aiming at this kind of problem, this paper presents a new classification method which recognizes a query image by comparing distances between category centers of CNN features of the whole training dataset and the corresponding CNN feature of this query image. The experimental results on Cifar-10 and Cifar-100 show that the claimed method can more accurately identify images whose categories have only a few training samples and that the mean precision of the recognition can be improved effectively.**

Keywords—category centers; class imbalance; CNN feature; image recognition.

I. INTRODUCTION

In the past decade, image recognition methods have experienced rapid development, and the convolutional neural network(CNN) has made breakthroughs in various application fields. In 2012, Alex Krizhevsky designed a deep convolutional network model raising the accuracy of image classification to a new height in ILSVRC-2012 [1]. Since then, researchers put forward new convolutional network models, such as Res-Net [2], Inception [3] and VGG [4], these models bring not only better image recognition results, but also further better performance in the object detection and semantic segmentation. Although these carefully designed convolutional structures have excellent ability of recognition, these models would still come across performance degradation due to the problem of class imbalance of training data sets. In general, class imbalance is a problem with which most practical applications of image recognition model must face. Researchers have explored many coping skills to avoid recognition performance degradation of convolutional neural network models in situations of class imbalance, these methods can be classified into two aspects: data level and model level. At the data level, artificial methods can make the unbalanced subsets of training data meet the number of balance, and make the recognition results be improved in some unbalanced situations, common methods include over-sampling and under-sampling. At the model level, there are a number of techniques that allow the model to better deal with class imbalance, such as threshold sliding and cost sensitivity operation.

The method proposed in this paper is to optimize the recognition performance at the model level. Class imbalance is rooted in that the amounts of training samples of some categories are smaller than others' amounts in training step, which lead to less adaption alteration of parameters in the same model for these categories than that for categories with more training samples. From this point of view, decreasing the portion of parameter adjustment process based on the unbalanced data in the whole mode learning process can reduce the degradation of recognition performance to some extent. Based on this deduction, this paper claims to select a pre-trained model on large-scale data set, then fine tune it on unbalanced data set, finally calculate all categories' centers of high level features of convolutional model with training data set and predict query image's label by the nearest category center to this query image's corresponding convolutional feature. This method utilizes a complete convolutional neural network model in training stage, and abandons the classification layer and turn to use nonparametric categories' centers in the recognition stage, leading great reduction of the number of parameters of the final model. So the recognition ability of categories with less training samples can be greatly improved and well generalized.

The organization and structure of this paper is as below: the second part of this paper gives a simple analysis on recent researches of image recognition in unbalanced dataset situation and feature distribution of convolutional neural network models, the third part of this paper introduces the detail method of image classification with category centers in class imbalance situation, the forth part of this paper gives experimental implementations to test category centers' recognition ability in class imbalance situation, the last part gives the final conclusion.

II. RELATED WORK

The deep convolution network is usually the first choice for image recognition tasks. Reference [5] used a linear support vector machine to recognize images with features extracted

978-1-5386-7255-6/18/$31.00 ©2018 IEEE

from OverFeat [6] and achieved good results. The method proposed in this paper is also based on CNN features, but it is to construct the category centers of CNN features of all training images first, then determine a query image's category by the nearest neighbor of this query's corresponding CNN feature, and the total number of parameters in test step is much less than the original classification layer. Reference [7] systematically summarized the methods to deal with class imbalance about CNN model for image recognition, pointing out that over sampling is always the most effective method. So, the method of over sampling is comprehensively compared with our method in this paper.

From the perspective of the distribution of CNN features, similar images' high-level features of convolution network are relatively concentrated [8], showing that semantically relevant images' high-level features of convolution network tend to cluster in feature space. On the bases of this distribution, center point of each category is an excellent choice to descript a category whatever the number of training samples is. Therefore, the distance from each center point to the query image's feature at corresponding layer of CNN model can accurately determine the image semantic properties and the classification process can minimize the impact of the number of training samples.

III. METHOD

In this paper, the category centers of CNN features are used to recognize images, thus, constructing this kind of center point will be the first step of classification. After finishing training or fine-turning of a deep convolutional neural network with target dataset, CNN features of all images in each layer can be extracted through the forward propagation operation. High-level features of CNN model tend to be more close to the semantic description of image [9] and have more concentrated distributions [8], it is considered in this paper that the appropriate method is to construct the category center point by image feature at the last convolution layer or its consequent full connection layer.

A given training data set consists of N labeled image data, the number of categories is K. The k-th category contains N_k training samples. As depicted in Fig. 1, the set of high-level features of training data of the k-th category is F_k, the set of center points of all categories is C, each category center of CNN features is denoted by C_k. Suppose q is the query image, and f_q is the corresponding high-level features at the same model, its label is calculated as following:

$$C_k = \frac{1}{N_k}\sum_{i=1}^{N_k}\frac{f_i}{\| f_i \|_2} \tag{1}$$

$$label_q = \arg\min_k \| \frac{f_q}{\| f_q \|_2} - C_k \|_2 , \quad k \in [1, K] \tag{2}$$

Each high-level feature is scale to a normalized length here. The query image is fed to CNN model to extract f_q, and European distance is used to determine nearest neighbor.

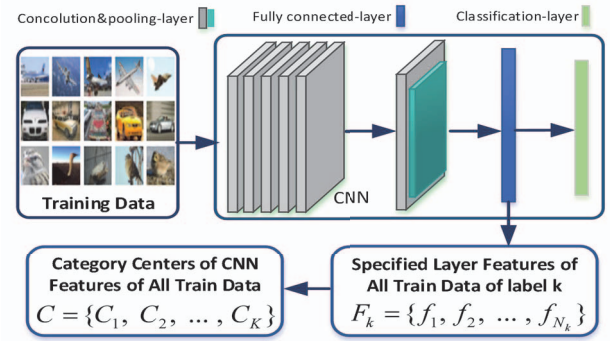

Fig. 1. Calculate all training images' CNN features and corresponding category centers at a specified layer. These category centers provide an excellent classification ability in class imbalance situations.

When the data sample size is large enough, the boundary of the sample points of a single category is able to approximate the real boundary between the categories well. If there is a decrease in the number of samples in a certain class, the boundary related to this category will be different from the real boundary, as shown in Fig. 2.

The classification layer of the CNN model is precisely to fitting the boundary of the high-level features for classification, therefore there is a large error between the fitted boundaries and the real ones in imbalance situation. However, regardless of the number of samples, the category center of CNN features is always the best approximation of sample distribution. The central point of a category is derived from all known samples, and the boundary is determined by part of the samples. In terms of probability, the stability of the category center is significantly better than that of the boundary, which is shown in Fig. 3.

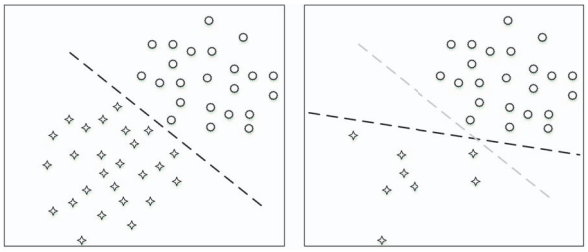

Fig. 2. As the number of samples decreases, the boundary changes. The left picture represents a balance situation, the right one represents a class imbalance situation and the dotted line represents the boundary.

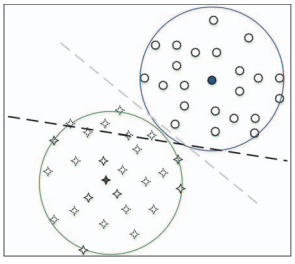

Fig. 3. The category center is similar to the center of a hypersphere consisted of all the samples with same label. Stars represented by dotted lines and solid lines are two subsets which have similar centers but different boundaries.

IV. Experiments and Discussion

A. Datasets and Model

In order to illustrate the applicability of the proposed method in class imbalance situations, three kinds of unbalanced datasets are sampled from Cifar-10 and Cifar-100, which are shown in table 1 and table 2. Each imbalance type is produced through random sampling in the original balance training data and validation data. The labels of Cifar-10 and Cifar-100 are translated into numbers, ranging from 0 to 9 and from 0 to 99 respectively. In each imbalance situation, r denotes the ratio of the maximum number of one label's training samples to the minimum number, m denotes a linear changing of the ratio, u denotes the number of labels which are under-sampled from original dataset. VGG-16 [4] is selected as experimental model in this paper. Besides, All-CNN [10] is added to consideration to test the generalization ability of the proposed method. In VGG-16, the initial learning rate is 0.001 and the batch size is 50. A random shuffle operation is done before each epoch. The classification loss and L2-normalization loss are optimized by SGD and the learning rate exponentially goes down by $0.5e^{0.1s}$ in which s is the number of recessions of validation accuracy. Layers in bold type in table 3 are the specified layers to construct category centers. All training processes are fine-turned on pre-trained model based on ImageNet.

TABLE I. Three Imbalance Datasets Sampled from Cifar-10

Label	Imbalance A r=10, u=5		Imbalance B r=20, u=9		Imbalance C r=m, u=9		Test Set
	Train	*Valid*	*Train*	*Valid*	*Train*	*Valid*	
0	400	100	200	50	200	50	1000
1	400	100	200	50	600	150	1000
2	400	100	200	50	1000	250	1000
3	400	100	400	100	1400	350	1000
4	400	100	400	100	1800	450	1000
5	4000	1000	400	100	2200	550	1000
6	4000	1000	2000	500	2600	650	1000
7	4000	1000	2000	500	3000	750	1000
8	4000	1000	2000	500	3500	870	1000
9	4000	1000	4000	1000	4000	1000	1000

TABLE II. Three Imbalance Datasets Sampled from Cifar-100

Label Range [a]	Imbalance A r=10, u=5		Imbalance B r=20, u=9		Imbalance C r=m, u=9		Test Set
	Train	*Valid*	*Train*	*Valid*	*Train*	*Valid*	
0~9	40	10	20	5	20	5	100
10~19	40	10	20	5	60	15	100
20~29	40	10	20	5	100	25	100
30~39	40	10	40	10	140	35	100
40~49	40	10	40	10	180	45	100
50~59	400	100	40	10	220	55	100
60~69	400	100	200	50	260	65	100
70~79	400	100	200	50	300	75	100
80~89	400	100	200	50	350	87	100
90~99	400	100	400	100	400	100	100

[a] Labels in each range share same numbers of training samples, validation samples and test samples.

TABLE III. Extracting High-Level Features with VGG-16

Layer	Data Dimension
Input	(224, 224, 3)
2 * (Convolution+Relu) + Maxpool	(112, 112, 64)
2 * (Convolution+Relu) + Maxpool	(56, 56, 128)
3 * (Convolution+Relu) + Maxpool	(28, 28, 56)
3 * (Convolution+Relu) + Maxpool	(14, 14, 512)
3 * (Convolution+Relu) + Maxpool	(7, 7, 512)
Convolution+Relu	**(1, 1, 4096)**
Dropout(0.5)	(1, 1, 4096)
FC-4096	**(4096,)**
Dropout(0.5)	(4096,)
FC-100	(100 or 10,)
Softmax	(100 or 10,)

TABLE IV. Extracting High-Level Features with All-CNN

Layer	Data Dimension
Input+Dropout(0.1)	(32, 32, 3)
2*(Convolution+Relu+Batch normalization)	(32, 32, 96)
Convolution+Relu+Batch normalization	(16, 16, 96)
Dropout(0.5)	(16, 16, 96)
2*(Convolution+Relu+Batch normalization)	(16, 16, 192)
Convolution+Relu+Batch normalization	(8, 8, 192)
Dropout(0.5)	(8, 8, 192)
Convolution+Relu+Batch normalization	(6, 6, 192)
Convolution+Relu+Batch normalization	(6, 6, 192)
Convolution+Relu+Batch normalization	**(6, 6, 10)**
Maxpool	(1, 1, 10)
Softmax	(10,)

In All-CNN, layer in bold type in table 4 is the specified layer to construct category centers and the feature is flattened here. The initial learning rate is 0.005 and the batch size is 256. The rest setting is same as that in VGG-16.

B. Classification in Imbalance Situations

In each imbalance situation, it is necessary to extract high-level features of all training images including training set and validation set before calculating centers of all categories. After training and feature extracting, category centers are gained according (1). Then these category centers are used in classification tasks in all unbalanced datasets in table 1 and table 2. All test images' features at corresponding layers are L2-normalized, then the normalized features' nearest category centers give the results of recognition. Each label's recognition precision with VGG-16 on test data is depicted as Fig. 4, and all mean precisions are shown in table 5 and table 6. The results in All-CNN are shown as Fig. 5 and table 7. In this section, D denotes classification with the original classification layer, C denotes classification with category centers of the last convolutional layer, F denotes classification with category centers of the consequent full connection layer, the prefix of O

means that the unbalanced training data gets an over-sampling operation.

As depicted in Fig. 4, Fig. 5, table 5, table 6 and table7, category centers whether based on the last convolutional layer or the full connection layer, generally perform better than both the original classification layer and over-sampling, especially on labels of less training samples. Although there is a slightly decline of recognition ability on labels of more training samples, the mean precision is still improved obviously.

Compared with classification with only the category centers, combining category centers with over-sampling can provide further improvement in Cifar-10, which is opposite in Cifar-100. The only difference between Cifar-10 and Cifar-100 is that the total number of samples of each label in Cifar-10 is 10 times as that of Cifar-100. Thus, it can be inferred that the absolute number of training samples affects the best choice of solution while using category centers.

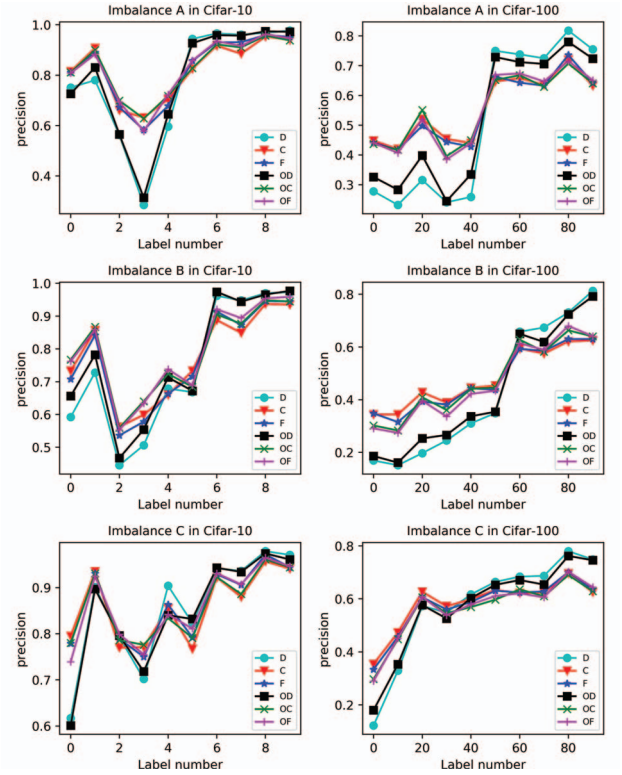

Fig. 4. Category centers' classification ability in class imbalance situations with the model of VGG-16.

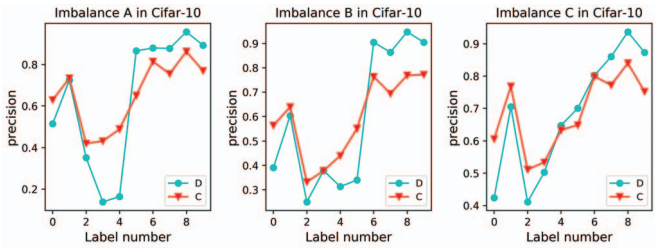

Fig. 5. Category centers' classification ability in class imbalance situations with the model of All-CNN.

TABLE V. MEAN PRECISION IN CIFAR-10 WITH VGG-16

Classification	Mean Precision in Cifar-10		
	Imbalance A r=10, u=5	*Imbalance B r=20, u=9*	*Imbalance C r=m, u=9*
D	0.779	0.747	0.857
C	0.824	0.775	0.859
F	0.826	0.772	0.865
OD	0.787	0.770	0.850
OC	0.831	0.792	0.861
OF	0.830	0.796	0.862

TABLE VI. MEAN PRECISION IN CIFAR-100 WITH VGG-16

Classification	Mean Precision in Cifar-100		
	Imbalance A r=10, u=50	*Imbalance B r=20, u=90*	*Imbalance C r=m, u=90*
D	0.511	0.430	0.575
C	0.558	0.482	0.583
F	0.555	0.476	0.577
OD	0.524	0.434	0.573
OC	0.555	0.476	0.563
OF	0.555	0.468	0.565

TABLE VII. MEAN PRECISION IN CIFAR-10 WITH ALL-CNN

Classification	Mean Precision in Cifar-10		
	Imbalance A r=10, u=50	*Imbalance B r=20, u=90*	*Imbalance C r=m, u=90*
D	0.637	0.589	0.682
C	0.656	0.590	0.686

V. CONCLUSION

In image recognition tasks of class imbalance situations, it can effectively improve the recognition results to determine image's label by category centers of CNN features, and the improvement is more obvious than conventional over-sampling. While combined with over-sampling, the category centers can provide further recognition improvement if the number of each label's training samples grows greater.

ACKNOWLEDGMENT

Thanks for the support by the Research Fund of Science and Technology of Shenzhen (No.JCYJ20170817165149850).

REFERENCES

[1] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in NIPS, Lake Tahoe, Nevada, USA, 2012, pp. 1097-1105.

[2] S. Zagoruyko and N. Komodakis, "Wide residual networks," in arXiv:1605.07146, 2016.

[3] C. Szegedy, V. Vanhoucke, S. Ioffe and J. Shlens, "Rethinking the inception architecture for computer vision," in arXiv:1512.00567v3, 2015.

[4] K, Simonyan and A, Zisserman, "Very deep convolutional networks for large-scale image recognition," in arXiv:1409.1556, 2014.

[5] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: an astounding baseline for recognition," in CVPRW, Columbus, OH, USA, 2014, pp. 512-519.

[6] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus and Y. LeCun, "Overfeat: integrated recognition, localization and detection using convolutional networks," in arXiv:1312.6229, 2013.

[7] M. Buda, A. Maki and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," in arXiv:1710.05381v1, 2017.

[8] A. Babenko and V. Lempitsky, "Aggregating deep convolutional features for image retrieval," in ICCV, Santiago, Chile, 2015, pp. 1269-1277.

[9] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in ECCV, Zurich, Switzerland, 2014, pp. 818-833.

[10] J. T. Springenberg, A. Dosovitskiy, T. Brox and M. Riedmiller, "Striving for simplicity: the all convolutional net," in arXiv:1412.6806v3, 2015.