Bronze inscription classification based on rare-class sample generator

Zhanteng Fang, Youguang Chen*
School of Data Science and Engineering, East China Normal University, Shanghai, China

ABSTRACT

Bronze Inscriptions is essential to the study of ancient Chinese history. We have obtained a set of materials containing 102,338 images of bronze texts collected by experts, but the distribution of these acquired images is imbalance, with a large sample of common words but the opposite for rare words. To solve this problem, we have adopted an end-to-end enhanced rare data generation module that can generate feature maps for rare classes during the training phase. The experiment results show that the proposed model can obtain an overall average accuracy of 77.38% and a minority accuracy of 68.24%.

Keywords: CNN, image classification, data augmentation, global attention mechanism

1. INTRODUCTION

Bronze inscriptions are characters written in various Chinese scripts on bronze vessels in the B.C. period, it is important to the study of ancient Chinese history. Traditionally, when conducting research on ancient scripts, scholars and experts need to interpret them against records such as interpretations kept in databases, which is not only inefficient but also has a high error rate. Moreover, modern Chinese characters are divided into commonly used characters and uncommon characters, and the same problem still exists by analogy to ancient scripts, i.e., commonly used characters occupy most of the sample size, while the more diverse types of uncommon characters occupy only a small portion of the sample. If the histogram is drawn by sorting the sample size of each script category from most to least, the distribution of the graph will have a "long tail", which is so called "long-tail distribution".

With the rapid development of deep learning technology, more and more computer vision tasks can be done using neural network such as CNN. We have obtained a dataset of ancient texts from real inscription rubbings organized and marked by experts—BRONZE-SET. However, BRONZE-SET is not the same as other well-designed large-scale datasets where each category has a particularly wide range of variation in sample size. This dataset has more than four hundred classes, where 35% of the images belong to the first 16 ancient Chinese characters, while more than 200 ancient Chinese characters with only 15% of the sample.

In this paper, we designed a classification model using ResNet[1] as the backbone. To solve the sample imbalance problem, we introduce an embeddable module that automatically generates samples during the training process based on RSG[2] method. In addition, we use a different loss function from the traditional softmax. To improve the prediction of the model, we add GAM[3] to the model to improve the performance of deep neural networks by reducing information dispersion and amplifying the global interaction representation. Section 2 focuses on the distribution of the dataset. Section 3 illustrates some of the algorithms used in this paper to solve the long-tail distribution problem. Section 4 introduces our model, followed by a comparison experiment between our model and baseline. Section 5 is the conclusion of our paper.

2. OVERVIEW ON BRONZE-SET

BRONZE-SET has a sample of over 100,000 images labelled by experts, each representing one text. These images are divided into 402 categories, the category with the largest sample size has 4926 images, while the smallest has only 50 images. This dataset also has the following problems:

- One part of the image is white on black, the other part is black on white, resulting in a very different image style.
- Due to the long history of topography, many characters are written in a completely different way from modern Chinese, and there are many different variants of the same character in the dataset.

* ygchen@cc.ecnu.edu.cn

Third International Conference on Computer Science and Communication Technology (ICCSCT 2022)
edited by Yingfa Lu, Changbo Cheng, Proc. of SPIE Vol. 12506, 125063J

- Many images have severe noise, and some images even have only part of the text in them.

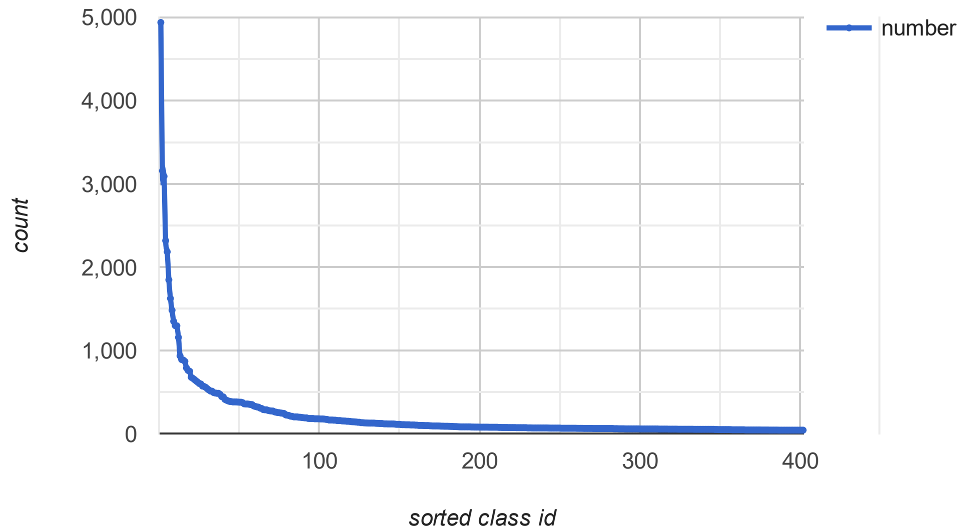

Figure 1. Long-tail distribution of the BRONZE-SET.

As shown in Figure 1, the distribution of the whole dataset is severely imbalanced. We can see that the top 50 classes account for more than half of the total sample, but they only account for about one-tenth of the total number of classes. Similarly, Figure 2 shows a few sampled images, which can be seen in a different style.

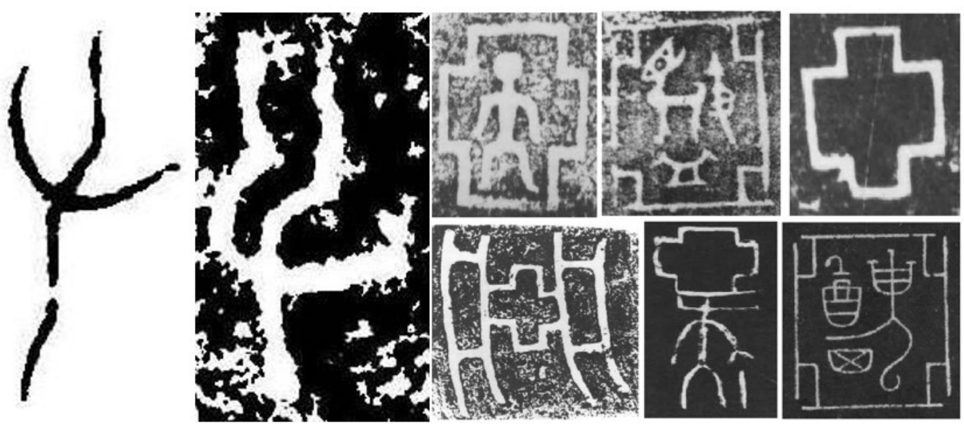

Figure 2. Some samples images of the BRONZE-SET.

3. BRONZE-SET CLASSIFICATION ALGORITHM

As shown in Figure 1, due to the severe sample imbalance problem in the dataset, if the overall data is directly used to train the model, it will inevitably cause the learned model to overfit the head class, resulting in many rare class samples being recognized as common words. Current algorithms to solve such problems are undersampling[4], oversampling[5], re-weight[6], data augmentation[7], logit adjustment[8], transfer learning[9], ensemble learning[10] and so on.

3.1 Overview of enhanced rare-class sample generator

Our paper employs an end-to-end data augmentation method based on rare-class sample generator (RSG) called ERSG (Enhanced Rare-class Sample Generator). In addition, a GAM attention module is added to the model to enhance the feature extraction capability. This module is only used in the training phase to generate rare class samples and is no longer used in the validation phase. Figure 3 shows the combination of this module and a simple ResNet.

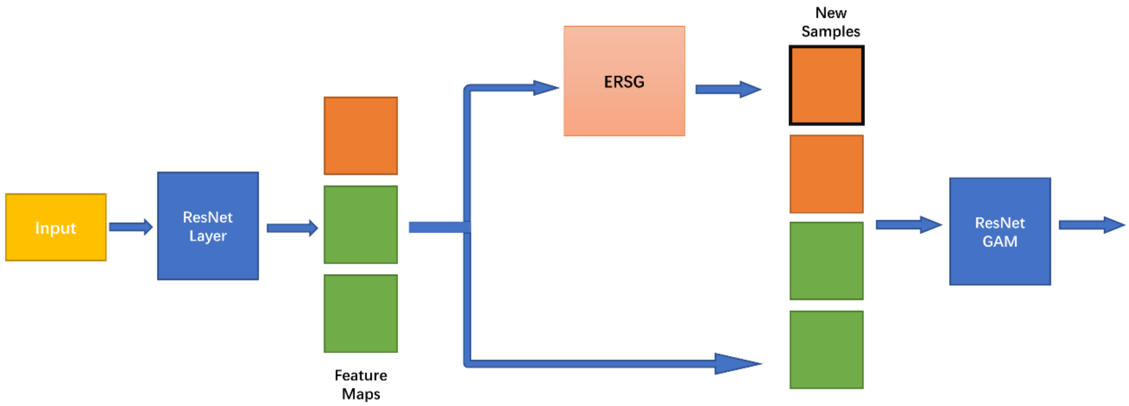

Figure 3. Combination of ERSG and ResNet.

3.2 Design of the algorithm

In detail, our BRONZE-SET dataset satisfies the assumptions of RSG, that is, all samples from the same class follow a unimodal distribution or a multi-modal distribution. Because these data are derived from several bronzes, the same bronze is written in a very similar way and the image noise too. First, we will input a batch of feature maps containing frequent and rare classes to our model, and they will go through a central estimation module, this module will estimate one or several centres of this samples. These centroids will be used as the base points for generating new samples. We further add a probability parameter to filter the suitable centroids. Unlike RSG, we discard its contrastive module and instead used a 5×5 channel convolutional layer interleaved with a Leaky ReLU[11] activation layer. This also generates a probability to represent whether two samples from the same class. This ensures that no information related to frequent classes is included in the sample generation process. Then, a vector transformation module calculates the feature displacement, which represents the displacement of a sample in a class with respect to its corresponding centre, caused by the same object under different conditions for each frequent class. This allows the generation of a new minority class samples by virtue of the feature displacement of the majority class samples, which is a good solution to the model misjudgement caused by the lack of variation in a few classes of samples.

4. EXPERIMENT

In our experiments, we used ResNet as the backbone, because ResNet incorporates a residual module, it can prevent performance degradation as the network gets deeper and deeper. Specifically, it includes a constant mapping connection that ensures that features relevant to classification are superimposed without losing the features already learned.

Global attention mechanism is abbreviated as GAM, unlike CBAM[12] which uses space and channel attention mechanisms respectively, GAM is an attention mechanism across the spatial-channel dimension that retains information to amplify "global" cross-dimensional interactions and can capture important features of all three dimensions.

To accurately evaluate the performance of the model, we designed multiple controlled experiments, each controlled to change only one condition. Through our previous experiments and experiences, the best results can be achieved by resize the image to 32×32. So, we first resize all the images to 32×32, and then put them into the model for training. In addition, we sort the classes according to the number of samples and divide the head and tail classes in a ratio of 1:5. In the testing phase, we classify those with more than 200 samples as majority classes, those with less than 50 as minority classes, and the rest as intermediate classes. The baseline experiment called ResNet-Only is to train ResNet18 with the dataset. Control experiment 1 adds the CBAM module to the baseline and is called ResNet-CBAM. For control experiment 2, a combination of ResNet18 and GAM was used. Control experiment 3 used ResNet18 and the unmodified RSG module with the same parameter configuration and is called ResNet-RSG. The final control experiment used the enhanced RSG algorithm mentioned above. In addition to the overall average accuracy, we also tracked the average accuracy of the three categories by number, which was used to better measure the model's predictive performance for the head and tail classes.

Table 1. Different average accuracy of the test set.

	Overall accuracy	Majority accuracy	Medium accuracy	Minority accuracy
Baseline	72.47	76.03	69.38	58.24
ResNet-CBAM	75.63	78.42	72.10	63.91
ResNet-GAM	75.89	80.34	71.32	61.28
ResNet-RSG	74.51	77.23	68.41	64.31
ResNet-ERSG	77.38	81.29	72.67	68.24

Table 1 demonstrates the average accuracy of the overall and the average accuracy of different classes by sample size. All three existing models have been improved compared to baseline, ResNet-GAM has the largest improvement over baseline in 4 metrics: 3.42%, 4.31%, 1.94% and 3.24% respectively. But our ResNet-ERSG algorithm outperforms all of them, not only in terms of average accuracy, but more importantly, the average accuracy of the tail classes has improved significantly.

5. SUMMARY

In large-scale ancient character classification tasks, the sample imbalance problem often affects the performance of the model. To solve this problem, many data augmentation algorithms have been proposed for generating samples of few classes to influence the decision boundary of the model. This paper uses a module for automatic sample generation during the training phase. The experimental results not only illustrate the overall improvement of the algorithm on the model, but also show the excellent performance on rare samples classes. The algorithm can meet the needs of normal ancient text classification tasks.

REFERENCES

[1] He, K., Zhang, X., Ren, S. and Sun, J., "Deep residual learning for image recognition," Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 770-778 (2016).

[2] Wang, J., Lukasiewicz, T., Hu, X., Cai, J. and Xu, Z., "RSG: A simple but effective module for learning imbalanced datasets," Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 3784-3793 (2021).

[3] Liu, Y., Shao, Z. and Hoffmann, N., "Global attention mechanism: Retain information to enhance channel-spatial interactions," arXiv preprint arXiv:2112.05561, (2021).

[4] Liu, X. Y., Wu, J. and Zhou, Z. H., "Exploratory undersampling for class-imbalance learning," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 39(2), 539-550 (2008).

[5] Park, S., Hong, Y., Heo, B., Yun, S. and Choi, J. Y., "The majority can help the minority: Context-rich minority oversampling for long-tailed classification," Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 6887-6896 (2022).

[6] Zhang, Z. and Pfister, T., "Learning fast sample re-weighting without reward data," Proc. of the IEEE/CVF Inter. Conf. on Computer Vision, 725-734 (2021).

[7] Perez, L. and Wang, J., "The effectiveness of data augmentation in image classification using deep learning," arXiv preprint arXiv:1712.04621, (2017).

[8] Hong, Y., Han, S., Choi, K., Seo, S., Kim, B. and Chang, B., "Disentangling label distribution for long-tailed visual recognition," Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 6626-6636 (2021).

[9] Zhang, Y., Cheng, D. Z., Yao, T., Yi, X., Hong, L. and Chi, E. H., "A model of two tales: Dual transfer learning framework for improved long-tail item recommendation," Proc. of the Web Conf. 2021, 2220-2231 (2021).

[10] Zhou, B., Cui, Q., Wei, X. S. and Chen, Z. M., "BBN: Bilateral-branch network with cumulative learning for long-tailed visual recognition," Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 9719-9728 (2020).

[11] Xu, B., Wang, N., Chen, T. and Li, M., "Empirical evaluation of rectified activations in convolutional network," arXiv preprint arXiv:1505.00853, (2015).

[12] Woo, S., Park, J., Lee, J. Y. and Kweon, I. S., "CBAM: Convolutional block attention module," Proc. of the European Conf. on Computer Vision (ECCV), 3-19 (2018).