Dual Loss for Manga Character Recognition with Imbalanced Training Data

Yonggang Li, Yafeng Zhou, Yongtao Wang*, Xiaoran Qin and Zhi Tang

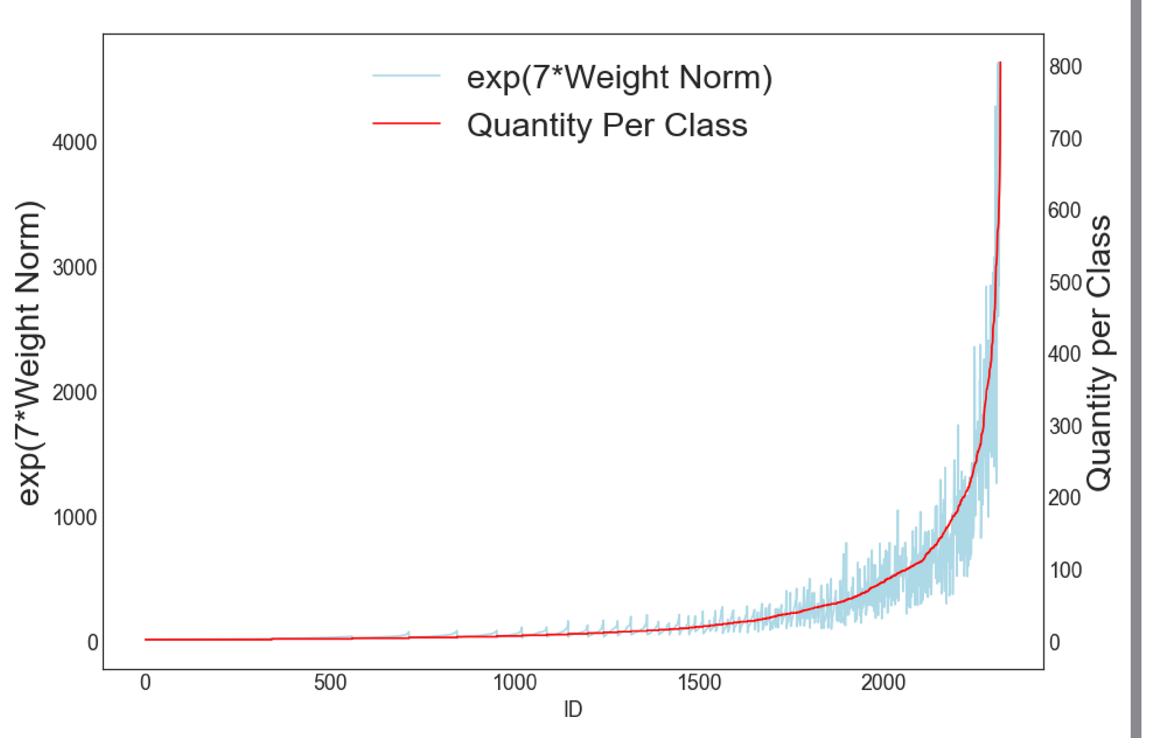
Wangxuan Institute of Computer Technology, Peking University

Email: {liyonggang, ola, wyt, qinxiaoran, tangzhi}@pku.edu.cn

1. Introduction

Manga character recognition is a key technology for manga character retrieval and verification. This task is very challenging since the manga character images have a long-tailed distribution and large quality variations. We classify the data imbalance into the *quantity imbalance* and the *quality imbalance*. The *quantity imbalance* means the wide discrepancies in the number of samples for different classes, which is also denoted as the long tail property [1]. Fig.1(a) illustrates the long-tailed sample quantity distribution of the Manga109 [2] training set that we use in character recognition experiments. The above phenomenon occurs because a manga book usually has a small number of major characters (protagonists) who appear frequently but a large number of minor characters (supporting roles) who appear infrequently. The *quality imbalance* means the sample scale, pose, sharpness, and fineness of the same identity vary greatly in manga, as shown in Fig.1(b). Training models with cross-entropy softmax loss on such imbalanced data would introduce biases to feature and class weight norms.

2. Data Imbalance



(a)

(b)

Figure 1: (a) The distribution of dataset sample quantity per class and the distribution of $e^{7\cdot(weight\ norm)}$ per class. One can see that weight norm is exponentially correlated with the number of samples per class. (b) Illustration of the imbalance of sample quality. This imbalance is caused by the sample scale, pose, sharpness, and fineness.

3. Method

Softmax loss, also known as cross-entropy softmax loss, is fundamental in the recognition task and formulated as:

$$p_{i,j} = \frac{e^{W_j^T x_i + b_j}}{\sum_{k=1}^{M} e^{W_k^T x_i + b_k}}, \tag{1}$$

$$\mathcal{L}_s = -\frac{1}{N}\sum_{i=1}^{N} \log p_{i,y_i}. \tag{2}$$

Dual Ring Loss (DRL): To alleviate the imbalance of feature and weight norm, we combine the ring loss [3] and the under-represented term [4].

$$\mathcal{L}_{dr} = \frac{\lambda_1}{2N}\sum_{i=1}^{N}\left(\|x_i\|_2 - \alpha\ \right)^2 + \frac{\lambda_2}{2M}\sum_{i=1}^{M}\left(\|W_i\|_2 - \beta\ \right)^2, \tag{3}$$

$$\mathcal{L} = \mathcal{L}_s + \mathcal{L}_{dr}. \tag{4}$$

Dual Adaptive Re-Weighting Loss (DARL): To further improve the performance of the deep model on the imbalanced data, we propose the dual adaptive re-weighting loss. It assigns different weights to the softmax loss of different samples or different categories.

$$w_w = 1 - \lambda_3 \frac{\|W_{y_i}\|_2 - \min\|W_j\|_2}{\max\|W_j\|_2 - \min\|W_j\|_2}, j \in [1, M], \tag{5}$$

$$w_x = 1 - \lambda_4 \frac{\|x_i\|_2 - \min\|x_j\|_2}{\max\|x_j\|_2 - \min\|x_j\|_2}, j \in [1, N], \tag{6}$$

$$\mathcal{L}_{dar} = -\frac{1}{N}\sum_{i=1}^{N} w_w w_x \mathcal{L}_s(x_i). \tag{7}$$

Dual Loss: A Combination of DRL and DARL: The above two losses are both forcing the norm of the learned feature or the norm of class weight to be similar. Therefore, they can reinforce the learning of each other. We combine them to form the dual loss $\mathcal{L}_d$ and use it to supervise the training process of the deep model, as given by Equation (8).

$$\mathcal{L}_d = \mathcal{L}_{dr} + \mathcal{L}_{dar}. \tag{8}$$

4. Experiments

Datasets Manga109 [2] is the biggest manga dataset with character identity annotations in this community. It consists of 109 manga volumes, 29845 image pages in total. We crop the head boxes of character from images following the annotations in [5] and get 134005 head images of 3173 identities, which include 93 distract identities.

Table 1: Experimental results using Manga109.

Methods	Retrieval			Verification
	rank-1(%)	rank-5(%)	mAP(%)	Accuracy(%)
Softmax	66.60	81.95	35.72	87.00
L2-Constrained	64.45	**81.35**	**35.10**	87.70
NormFace	64.00	80.25	33.66	**87.90**
CosFace	64.25	79.35	33.22	86.80
ArcFace	60.45	76.65	30.30	87.00
Am_Softmax	**64.50**	80.25	33.14	87.00
Range Loss	68.70	82.50	36.01	86.20
CB Loss	68.30	**83.50**	36.10	87.30
Focal Loss	67.85	83.35	36.70	87.50
CB Focal	68.65	83.25	36.58	87.60
Softmax + RL	68.70	**83.50**	**37.03**	87.40
Softmax + UP	**68.80**	83.00	36.23	**87.80**
Softmax + DRL	69.65	83.55	37.26	87.90
DARL	69.00	83.80	37.66	87.80
Dual Loss	**70.55**	**84.30**	**38.88**	**88.50**

5. References

[1] X. Zhang, Z. Fang, Y. Wen, Z. Li, and Y. Qiao, "Range loss for deep face recognition with long-tailed training data," in *ICCV*, 2017.

[2] Y. Matsui, K. Ito, Y. Aramaki, A. Fujimoto, T. Ogawa, T. Yamasaki, and K. Aizawa, "Sketch-based manga retrieval using manga109 dataset," *Multimedia Tools and Applications*, vol. 76, no. 20, pp. 21 811–21 838, 2016.

[3] Y. Zheng, D. K. Pal, and M. Savvides, "Ring loss: Convex feature normalization for face recognition," in *CVPR*, 2018.

[4] Y. Guo and L. Zhang, "One-shot face recognition by promoting underrepresented classes," *arXiv preprint arXiv:1707.05574*, 2017.

[5] X. Qin, Y. Zhou, Y. Li, S. Wang, Y. Wang, and Z. Tang, "Progressive deep feature learning for manga character recognition via unlabeled training data," in *ACM TUR-C*, 2019.