Supplementary Materials to "SGD with Partial Hessian for Deep Neural Networks Optimization"

Ying Sun[1], Hongwei Yong[1], and Lei Zhang[1]✉

The Hong Kong Polytechnic University, Hong Hum, Kowloon, Hong Kong, China.
csysun@comp.polyu.edu.hk, hongwei.yong@polyu.edu.hk, cslzhang@comp.polyu.edu.hk

Abstract. In this supplementary file, we report the numerical results of SGD-PH on person ReID benchmarks Market1501 and DukeMTMC-ReID to comprehensively evaluate the performance of SGD-PH.

1 Results on Person ReID Benchmarks

In this section, we further apply SGD-PH to two person ReID benchmarks Market1501 [3] and DukeMTMC-ReID [2,4]. The dataset Market1501 includes 1501 pedestrians and 32668 detected pedestrian rectangles, where 751 pedestrians with 12936 images of them are in training set and the rest 750 pedestrians with 19732 images are in testing sets, while DukeMTMC-ReID, a subset of DukeMTMC dataset for person ReID task, contains 702 persons with 16522 images for training, other 702 persons with 2228 query images and 17661 gallery images for testing. As well-known and commonly used benchmarks, they have been tackled in many existing works. Here, we adopt the method produced in [1] as our baseline [1].

Unlike image classification tasks that may benefit more from SGD than other optimizers, for person ReID tasks, the most popular optimizer is ADAM, which may be considered to have a more robust performance, and the default optimizer used in the baseline [1] is also ADAM. Tables 1 and 2 lists the results of Rank1 and mAP on the two datasets with ResNet18 and ResNet34 backbones. The experiments are done on eight GeForce RTX 2080Ti GPUs, repeated for 4 times and reported in the "mean $\pm$ std" format. Besides citing the performances of ADAM from [1], we tune the learning rate and weight decay of SGD by grid search with other settings keep the same as the baseline, while we keep the same hyperparameters as SGD for SGD-PH. Specifically, the learning rate and weight decay for SGD and SGD-PH on the two datasets are 0.03 and 0.003, respectively. For SGD-PH, the second order learning rate τ_{SO} takes 0.005 and 0.01 for Market1501 and DukeMTMC, respectively. The experiments on Market1501 is not inferior to SGDM (with slightly 0.1% better for most cases), while on DukeMTMC-ReID it gains 0.3%$\sim$ 0.4% for rank1 accuracy and 0.5%$\sim$ 0.6% for mAP compared with SGDM. Therefore, we illustrate the adaptability of our newly-proposed SGD-PH on person ReID tasks.

[1] The repository can be downloaded via: `https://github.com/michuanhaohao/reid-strong-baseline`.

Table 1. Experiment results (%) on Market1501.

Network	ResNet18		ResNet34	
Indicator	Rank1	mAP	Rank1	mAP
ADAM	91.7	77.8	92.7	82.7
SGDM	92.5 ± .4	81.2 ± .3	**93.5 ± .3**	83.9 ± .1
SGD-PH	**92.6 ± .2**	**81.3** ± .1	**93.5** ± .2	**84.0 ± .4**

Table 2. Experiment results (%) on DukeMTMC-ReID.

Network	ResNet18		ResNet34	
Indicator	Rank1	mAP	Rank1	mAP
ADAM	82.5	68.8	**86.4**	**73.6**
SGDM	84.0 ± .8	69.9 ± .5	85.4 ± .1	71.7 ± .1
SGD-PH	**84.4** ± **.4**	**70.5** ± **.1**	85.7 ± .9	72.2 ± .8

References

1. Luo, H., Gu, Y., Liao, X., Lai, S., Jiang, W.: Bag of tricks and a strong baseline for deep person re-identification. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (June 2019)
2. Ristani, E., Solera, F., Zou, R., Cucchiara, R., Tomasi, C.: Performance measures and a data set for multi-target, multi-camera tracking. In: European conference on computer vision. pp. 17–35. Springer (2016)
3. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: A benchmark. In: Proceedings of the IEEE international conference on computer vision. pp. 1116–1124 (2015)
4. Zheng, Z., Zheng, L., Yang, Y.: Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In: Proceedings of the IEEE international conference on computer vision. pp. 3754–3762 (2017)