Towards a Smaller Student: Capacity Dynamic Distillation for Efficient Image Retrieval Supplementary Material

Yi Xie[1] Huaidong Zhang[1*] Xuemiao Xu[1,4,5,6*] Jianqing Zhu[2] Shengfeng He[3]
[1]South China University of Technology [2]Huaqiao University [3]Singapore Management University
[4]State Key Laboratory of Subtropical Building Science
[5]Ministry of Education Key Laboratory of Big Data and Intelligent Robot
[6]Guangdong Provincial Key Lab of Computational Intelligence and Cyberspace Information
ftyixie@mail.scut.edu.cn, {huaidongz, xuemx}@scut.edu.cn
jqzhu@hqu.edu.cn, shengfenghe@smu.edu.sg

Table 1. The result of different retrieval strategies on VeRi776 [2].

Methods	MP (M)	FLOPs (G)	mAP(%)	R1(%)
standard retrieval strategy	14.56	4.56	80.28	96.13
role-cross retrieval strategy	**14.30**	**4.46**	**80.67**	**96.66**

Table 2. The result of RGGR *vs* CWGR [1] on VeRi776 [2].

Methods	MP (M)	FLOPs (G)	mAP(%)	R1(%)
CDD	18.59	5.77	80.36	96.48
CDD+CWGR [1]	14.35	4.46	79.41	95.71
CDD+RGGR	**14.30**	**4.46**	**80.67**	**96.66**

1. Experiments

1.1. RGGR's Retrieval Strategy Analysis

In pipeline of RGGR, we adopt another retrieval strategy (i.e., role-cross retrieval) to acquire accurate retrieval results. Specifically, we use $\mathcal{F}_t$ instead of $\mathcal{F}_s$ as the query set to retrieve the gallery set G because $\mathcal{F}_s$ is not invariant and semantic enough for presenting the image information during early training. Thus, we analyze the influent of different retrieval strategies on RGGR performance, as shown in Table 1. From the table, we can find that the role-cross retrieval strategy outperforms the standard retrieval strategy (i.e., $\mathcal{F}_s$ as the query set) by 0.39% mAP and 0.53% R1, demonstrating that using $\mathcal{F}_t$ as the query set can acquire more accurate retrieval results

1.2. RGGR *vs* CWGR

On the capacity dynamic distillation framework (CDD), we compared RGGR with a convolutional layer weight gradient resetting (CWGR) [1], as shown in Table 2. From the table, we can find that CWGR can further improve the inference performance of the student network but reduce the accuracy performance because output channels with small weight values of DGC may be necessary for retrieval results. Thus, RGGR outperforms CWGR by 1.27% mAP and 0.95% R1.

1.3. Hyper-parameter Analysis

The G-LASSO weight (i.e., α in Eq. (3) The hyper-parameter α is crucial to CDD+RGGR to control the parameter sparsity of the student network. Specifically, as α value increases, the accuracy (i.e., mAP) performance decreases slightly, but the computational performance (i.e., MP and FLOPs) improves significantly. For example, on In-shop [3], as the α value increases from 3×10^{-3} to 6×10^{-3} the mAP performance just drops from 81.5% to 81.1%, while the MP performance obviously improves from 16.2M to 14.0M. It demonstrates that our method has good insensitivity to parameter sparsity, which can accelerate a student network under the premise of preserving the accuracy performance.

The top-K retrieval result (i.e. K, in Eq. (7)) As K value increases, RGGR has more reference information when zeroing the learning gradient of unimportant channels. Fig. 2 exhibits the influence of K on CDD+RGGR mAP. Specifically, CDD+RGGR has good robustness on mAP and the mAP at $K > 1$ outperforms at $K = 1$ on Veri776 [2]. Besides, K value hardly affects inference performance.

*Corresponding authors

References

[1] Xiaohan Ding, Tianxiang Hao, Jianchao Tan, Ji Liu, Jungong Han, Yuchen Guo, and Guiguang Ding. Resrep: Lossless cnn

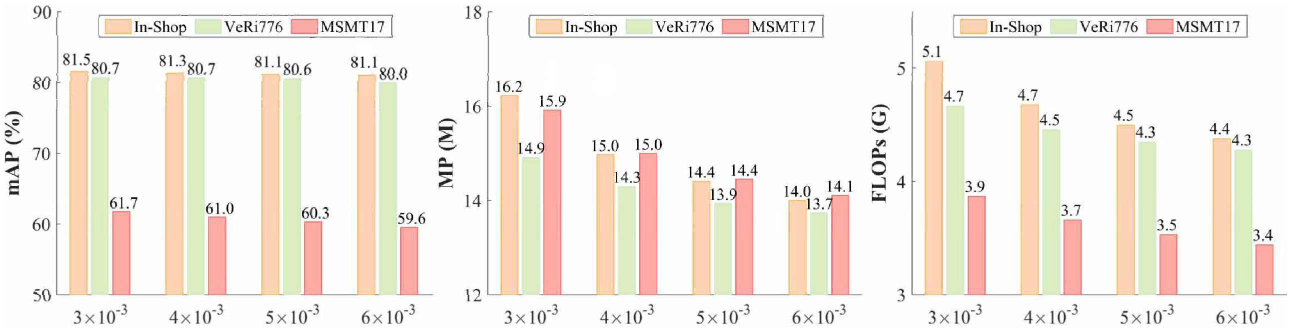

Figure 1. The influences of α values. (a) on mAP. (b) on MP. (c) on FLOPs. As α value increases, the mAP performance decreases slightly, but the computational performance improves significantly.

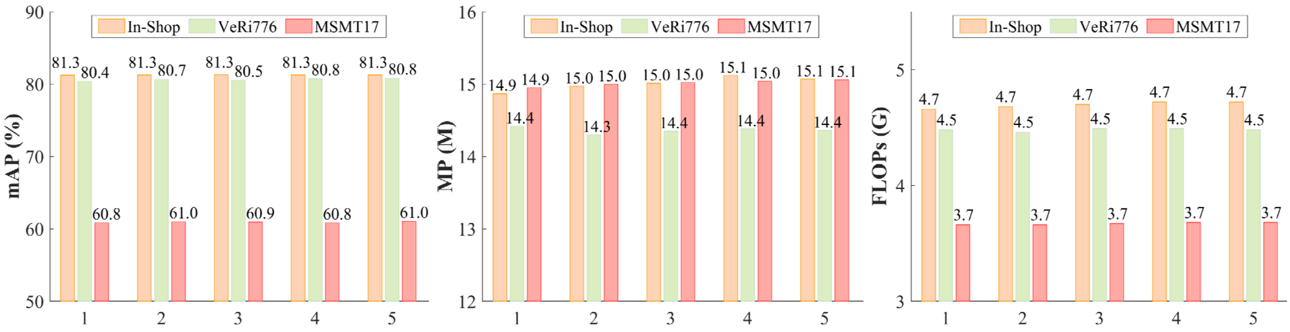

Figure 2. The influences of K values. (a) on mAP. (b) on MP. (c) on FLOPs. As K value increases, the mAP performance of the student network will fluctuate slightly, and the inference performance will be stable.

pruning via decoupling remembering and forgetting. In *ICCV*, pages 4510–4520, 2021. 1

[2] Xinchen Liu, Wu Liu, Huadong Ma, and Huiyuan Fu. Large-scale vehicle re-identification in urban surveillance videos. In *ICME*, pages 1–6, 2016. 1

[3] Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In *CVPR*, pages 1096–1104, 2016. 1