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

Supplementary Material

Yi Xie\textsuperscript{1} Huaidong Zhang\textsuperscript{1*} Xuemiao Xu\textsuperscript{1,4,5,6*} Jianqing Zhu\textsuperscript{2} Shengfeng He\textsuperscript{3}

\textsuperscript{1}South China University of Technology \textsuperscript{2}Huaqiao University \textsuperscript{3}Singapore Management University

\textsuperscript{4}State Key Laboratory of Subtropical Building Science

\textsuperscript{5}Ministry of Education Key Laboratory of Big Data and Intelligent Robot

\textsuperscript{6}Guangdong Provincial Key Lab of Computational Intelligence and Cyberspace Information

ftyxie@mail.scut.edu.cn, {huaidongz, xuemx}@scut.edu.cn

jqzhu@hqu.edu.cn, shengfenghe@smu.edu.sg

---

<table>
<thead>
<tr>
<th>Methods</th>
<th>MP (M)</th>
<th>FLOPs (G)</th>
<th>mAP(%)</th>
<th>R1(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>standard retrieval strategy</td>
<td>14.56</td>
<td>4.56</td>
<td>80.28</td>
<td>96.13</td>
</tr>
<tr>
<td>role-cross retrieval strategy</td>
<td>14.30</td>
<td>4.46</td>
<td>80.67</td>
<td>96.66</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Methods</th>
<th>MP (M)</th>
<th>FLOPs (G)</th>
<th>mAP(%)</th>
<th>R1(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDD</td>
<td>18.59</td>
<td>5.77</td>
<td>80.36</td>
<td>96.48</td>
</tr>
<tr>
<td>CDD+CWGR [1]</td>
<td>14.35</td>
<td>4.46</td>
<td>79.41</td>
<td>95.71</td>
</tr>
<tr>
<td>CDD+RGGR</td>
<td>14.30</td>
<td>4.46</td>
<td>80.67</td>
<td>96.66</td>
</tr>
</tbody>
</table>

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

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 $F_r$ instead of $F_s$ as the query set to retrieve the gallery set $G$ because $F_s$ is not invariant and semantic enough for presenting the image information during early training. Thus, we analyze the influence 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., $F_s$ as the query set) by 0.39% mAP and 0.53% R1, demonstrating that using $F_r$ 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 gra-

---

*Corresponding authors

---

1.3. Hyper-parameter Analysis

The G-LASSO weight (i.e., $\alpha$ in Eq. (3)) The hyper-parameter $\alpha$ is crucial to CDD+RGGR to control the parameter sparsity of the student network. Specifically, as $\alpha$ 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 $\alpha$ value increases from $3 \times 10^{-3}$ to $6 \times 10^{-3}$ 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.

References

[1] Xiaohai Ding, Tianxiang Hao, Jianchao Tan, Ji Liu, Jungong Han, Yuchen Guo, and Guiguang Ding. Restrep: Lossless cnn
Figure 1. The influences of $\alpha$ values. (a) on mAP, (b) on MP, (c) on FLOPs. As $\alpha$ 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.
