# Supplementary: Recall@k Surrogate Loss with Large Batches and Similarity Mixup

Yash Patel Giorgos Tolias Jiří Matas
Visual Recognition Group, Czech Technical University in Prague
{patelyas,toliageo,matas}@fel.cvut.cz

## 1. Additional experiments

### 1.1. Additional comparisons with SAP

Note that in Table 2 of the main paper, the batch size for SAP [1] was set to 384, following the original paper. In Table I, we provide additional comparisons with SAP with the batch size of $\min(4000, 4 \times \#\text{classes})$. It is observed that the proposed RS@k outperforms SAP even when SAP is trained with large batches.

| **Dataset** | **# Training Samples** | **SAP**$^\dagger$ | **RS@k**$^\dagger$ | **RS@k**$^\dagger$ + SiMix |
|---|---|---|---|---|
| iNaturalist | $325,846$ | $70.7$ | $71.2$ | $71.8$ |
| VehicleID | $110,178$ | $95.5$ | $95.7$ | $95.3$ |
| SOP | $59,551$ | $81.3$ | $82.8$ | $82.1$ |
| Cars196 | $8,054$ | $79.5$ | $80.7$ | $88.2$ |
| CUB200 | $5,864$ | $63.6$ | $63.8$ | $69.5$ |
| $\mathcal{R}$Oxf & $\mathcal{R}$Par (1m) | $1,060,709$ | $40.6$ | $41.0$ | $41.8$ |

Table I. Recall@1 (in %) with batch size of $\min(4000, 4 \times \#\text{classes})$ for iNaturalist, VehicleID, SOP, Cars196 and CUB200. mAP (in %) with batch size of 4096 for $\mathcal{R}$Oxford and $\mathcal{R}$Paris with 1 million distractor samples.

### 1.2. Impact of SiMix

Our results suggest that SiMix leads to a larger performance gain on smaller datasets, where batch size is restricted by the total number of classes. Results are summarized in Table I, where we additionally report results on CUB which has small (100) number of training classes. On Cars196 dataset, RS@k attains a r@1 of 80.7% without and 88.2% with SiMix (an absolute improvement for 7.5%). Similarly on CUB200, RS@k attains a r@1 of 63.8% without and 69.5% with SiMix (an absolute improvement of 5.7%).

### 1.3. Effect of hyper-parameters

**Values for *k*.** The study for the set of values of *k* used for RS@k loss can be found in Table II. The results RS@$\{1\}$, RS@$\{1, 2\}$, RS@$\{1, 2, 4\}$ and RS@$\{1, 2, 4, 8\}$ suggest that adding larger values of *k* leads to decline in the performance. However, RS@$\{1, 2, 4, 8, 16\}$ gives on an average the same results as RS@$\{1\}$, with higher performance on larger *k* values. Comparing the entries RS@$\{4, 8, 16\}$ with RS@$\{1, 2, 4, 8, 16\}$ suggests that the use of small values, such as $k = 1$ or $k = 2$, is crucial as the performance drops significantly when these values are removed. Further removing $k = 4$ (RS@$\{8, 16\}$) does not change the performance. However, removing $k = 8$ (RS@$\{16\}$) leads to a significant decline in the performance.

| Method | r@1 | r@2 | r@4 | r@8 | r@16 | Avg |
|---|---|---|---|---|---|---|
| RS@$\{1\}^\dagger$ | 81.1 | 87.7 | 92.0 | 95.0 | 96.9 | 90.5 |
| RS@$\{1, 2\}^\dagger$ | 80.2 | 87.2 | 91.9 | 95.0 | 97.2 | 90.3 |
| RS@$\{1, 2, 4\}^\dagger$ | 79.6 | 86.5 | 91.2 | 94.5 | 96.8 | 89.7 |
| RS@$\{1, 2, 4, 8\}^\dagger$ | 79.3 | 86.3 | 91.0 | 94.5 | 96.9 | 89.6 |
| RS@$\{1, 2, 4, 8, 16\}^\dagger$ | 80.8 | 87.6 | 92.2 | 95.0 | 97.1 | 90.5 |
| RS@$\{2, 4, 8, 16\}^\dagger$ | 80.3 | 87.5 | 92.3 | 95.4 | 97.5 | 90.6 |
| RS@$\{4, 8, 16\}^\dagger$ | 79.6 | 87.1 | 91.7 | 95.0 | 97.3 | 90.1 |
| RS@$\{8, 16\}^\dagger$ | 79.6 | 87.1 | 91.7 | 95.0 | 97.3 | 90.1 |
| RS@$\{16\}^\dagger$ | 75.8 | 83.9 | 89.8 | 93.6 | 96.4 | 87.9 |

Table II. Varying the set of values of *k*. Results on Cars196 [2]. In all experiments, $\tau_1 = 1$ and $\tau_2 = 0.01$.

**Batch sizes beyond 4k.** The batch size in our experiments is set to $\min(4000, 4 \times \#\text{classes})$. In Figure 4 (main paper), the study was conducted on Cars196 with 98 training classes. Batch-size higher than 392 requires sampling more than 4 samples per class and that forms a different, not directly comparable setup. Therefore, we now present additional analysis on SOP in Figure I where the batch size is varied from $2^4$ to $2^{13}$. Performance starts to saturate for very large batch size. Additionally, we increased batch size from 4096 to 8192 on iNaturalist and this leads to a slight loss in performance (0.2% r@1).

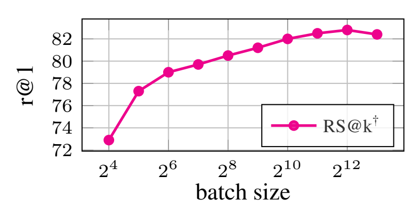


Figure I. Effect of batch size on SOP dataset [3].

## References


[1] Andrew Brown, Weidi Xie, Vicky Kalogeiton, and Andrew Zisserman. Smooth-ap: Smoothing the path towards large-scale image retrieval. In *ECCV*, 2020. 1

[2] Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3d object representations for fine-grained categorization. In *ICCV workshops*, 2013. 1

[3] Eng-Jon Ong, Sameed Husain, and Miroslaw Bober. Siamese network of deep fisher-vector descriptors for image retrieval. In *arXiv*, 2017. 1