# Adaptive Sparse Pairwise Loss for Object Re-identification
# – Supplementary Materials –

## 1. Proof

Unlike traditional dense sampling methods that may suffer from sampling harmful pairs, the proposed Sparse Pairwise (SP) loss can effectively avoid introducing harmful positive pairs **when combined with our least-hard positive mining strategy**. Here, we provide a simple theoretical proof to demonstrate **the expected percentage of harmful pairs sampled by SP in a mini-batch is lower than that by the traditional dense sampling methods**.

Let us assume that harmful positive pairs always have lower visual similarities than real positive pairs (which is true in general). Given a mini-batch with $N$ IDs and each ID contains $M$ instances, which shape a positive similarity matrix $\mathbf{S} \in \mathbb{R}_+^{M\times M}$, dense methods sample $M$ positive pairs with the lowest similarities in each row of $\mathbf{S}$. While our approach utilizes the pair with the highest similarity from the $M$ positive pairs sampled by dense methods. Suppose there are $K$ harmful positive pairs for each ID, which results in two situations according to the value of $K$:

**Situation I**: $K < M$. SP does not encounter harmful pairs because at least $M-1$ positive pairs with similarity smaller than the pair sampled by SP, while there are only $K(\leqslant M-1)$ harmful pairs. However, dense methods definitely sample harmful pairs and its expectation of sampling harmful pairs is assumed to be $E_I^D \in (0, K]$.

**Situation II**: $K \geqslant M$. SP is possibly sampling harmful pairs and its sampling expectation is assumed to be $E_{II}^{SP} \in [0,1]$ since it only samples one positive pair for each ID. Meanwhile, the expectation of dense methods for sampling harmful pairs is given by:

$$E_{II}^D = ME_{II}^{SP} + E_*^D\left(1 - E_{II}^{SP}\right) \tag{S1}$$

where $E_*^D$ denotes the expectation of dense methods for the ID that SP does not sample a harmful pair.

Assume the numbers of ID with $K < M$ and $K \geqslant M$ in a mini-batch are $U$ and $V$, respectively. The expected percentage of harmful pairs sampled by SP and dense approaches in a mini-batch can be given by:

$$P_{SP} = \frac{VE_{II}^{SP}}{N} \tag{S2}$$

$$\begin{aligned} P_D &= \frac{UE_I^D + V(ME_{II}^S + E_*^D(1 - E_{II}^{SP}))}{NM} \\ &= \frac{UE_I^D}{NM} + \frac{VE_{II}^{SP}}{N} + \frac{VE_*^D\left(1 - E_{II}^{SP}\right)}{NM} \end{aligned} \tag{S3}$$

Therefore, $P_{SP} < P_D$ is concluded. Especially for the ID with $K < M$, SP can completely avoid sampling harmful positive pairs.

Table S1. Statistics of object ReID datasets utilized in this work.

| Person | #ID | #images | Vehicle | #ID | #images |
|---|---|---|---|---|---|
| CUHK03 | 1467 | 13164 | VeRi-776 | 776 | 49357 |
| Market-1501 | 1501 | 32668 | VehicleID | 26328 | 221567 |
| DukeMTMC | 1404 | 36411 | VERI-WILD | 40671 | 416314 |
| MSMT17 | 4101 | 126441 | – | – | – |

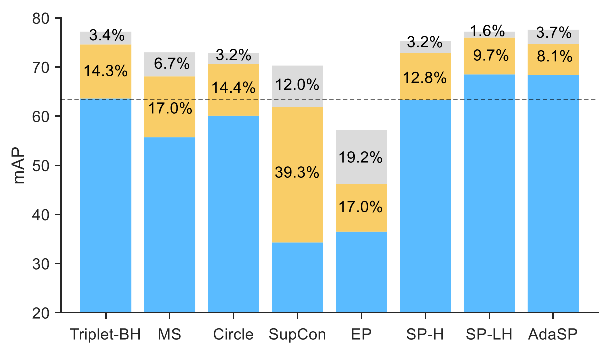


Figure S1. Comparison on the robustness of the performance regarding the number of instances per identity (including 8, 16, and 32) in a mini-batch on VeRi-776 dataset. The numbers in each bar represent the percentage of dropped performance relative to the best performance achieved in 8 instances. The dashed line denotes the mAP performance of SP-H with 32 instances per ID.

## 2. Additional Results

**Robustness to increasing intra-class variations.** To demonstrate the robustness of SP loss towards increasing intra-class variations, we also evaluate various metric losses on the VeRi-776 dataset. The experimental results are ex-

Table S2. Experimental results on person ReID datasets using the backbones of ResNet50-IBN, ResNet-152 and MGN with the input size of $256 \times 128$. The loss function consists of a cross-entropy loss and a metric loss (AdaSP for our approach, Triplet-BH for comparison).

| Backbone | Loss | MSMT17 | | Market1501 | | DukeMTMC | | CUHK03-L | | CUHK03-D | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | mAP | R1 | mAP | R1 | mAP | R1 | mAP | R1 | mAP | R1 |
| ResNet50-IBN | Triplet-BH | 57.3 | 80.0 | 83.3 | 93.1 | 75.0 | 86.6 | 65.5 | 67.6 | 62.1 | 65.0 |
| | AdaSP | **61.7** | **83.7** | **87.7** | **95.4** | **79.8** | **90.3** | **69.7** | **71.4** | **67.6** | **70.3** |
| ResNet-152 | Triplet-BH | 57.9 | 80.2 | 84.8 | 93.7 | 77.1 | 88.1 | 69.3 | 72.1 | 64.8 | 68.0 |
| | AdaSP | **60.1** | **82.4** | **88.1** | **95.0** | **80.0** | **89.4** | **71.6** | **73.2** | **69.2** | **79.4** |
| MGN | Triplet-BH | 59.2 | 80.8 | 88.1 | 95.0 | 79.1 | 88.6 | 74.9 | 76.9 | 72.2 | 75.1 |
| | AdaSP | **60.6** | **82.1** | **88.5** | **95.5** | **80.3** | **90.1** | **77.7** | **79.7** | **74.5** | **77.3** |

Table S3. Experimental results of the ablation study on the weight for metric loss $\lambda$ in VeRi-776 dataset.

| $\lambda$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 |
|---|---|---|---|---|---|
| mAP | 82.0 | 82.5 | **82.7** | 82.4 | 82.2 |
| R1 | **97.1** | 97.0 | 96.7 | 96.7 | 96.5 |

Table S4. Experimental results of the ablation study on the temperature $\tau$ in VeRi-776 dataset.

| $\tau$ | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|
| mAP | 67.6 | 76.0 | 79.9 | 82.6 | **82.7** | 82.1 | 81.6 | 78.9 | 77.1 |
| R1 | 90.5 | 94.3 | 95.5 | **96.8** | 96.7 | 97.0 | 96.4 | 96.0 | 95.1 |

hibited in Fig. S1. Consistent with the results on the MSMT17 dataset, shown in the main text, all metric losses still suffer from a performance decline as the intra-class variations increase. It can be observed that the EP loss achieves very limited mAP even though $8$ instances per identity are utilized. Besides, the performance of SupCon loss drops sharply and the MS loss loses more than $23\%$ of mAP with a growing number of instances. The triplet loss still outperforms other dense pairwise losses and achieves comparable performance with our SP-H loss. However, it can be seen that SP loss variants with the least-hard and the adaptive positive mining strategies lose less than $12\%$ of mAP when the instance number increases to $32$, indicating considerable robustness toward increasing intra-class variations.

**Generalizing to different networks.** We also test SP loss on other networks, including ResNet50-IBN, ResNet-152 and MGN, with the input size of $256 \times 128$. The experimental results are shown in Tab. S2. Compared to ResNet-50, the performance of Triplet-BH on each dataset is significantly increased by these well-designed networks. Nevertheless, our approach AdaSP can still facilitate the ReID performance on each dataset, especially the hard ones, such as MSMT17 and CUHK03.

**Impact of hyper-parameters.** We adopt the VeRi-776 dataset to explore the impact of hyper-parameters on Vehicle ReID tasks. The experimental results are exhibited in Tab. S3 and S4. The best mAP is achieved at a temperature of $0.05$ and a weight of $0.5$. In addition, it can be seen that all the mAPs are higher than $82.0$ when the weight of SP loss varies from $0.1$ to $0.9$, suggesting that the vehicle ReID performance is not sensitive to the weight of SP loss.