MVP Matching: A Maximum-value Perfect Matching for Mining Hard Samples, with Application to Person Re-identification

Han Sun[1,2], Zhiyuan Chen[1,2], Shiyang Yan[3], Lin Xu[1,2*]
[1]Nanjing Institute of Advanced Artificial Intelligence [2]Horizon Robotics [3]Queen's University Belfast
{han.sun1102, zhiyuan.chen01, elyotyan, lin.xu5470}@gmail.com

In this supplementary material, we first present a detailed analysis of relationships between our proposed batch-wise maximum-value perfect (MVP) matching based loss and other related loss objectives for metric learning. Then, we show a full batch-wise (i.e., 32×32) visualization of the exclusive hard positive and negative pairs selected via the MVP matching.

1. Relationships with Related Loss Objectives

In this section, we theoretically discuss the relationships between the proposed MVP matching based loss and other recently developed metric learning loss objectives. We simplify all well-known losses into a unified batch-wise form for fair comparison and clear demonstration.

We assume that each batch has n samples from p random classes (i.e., person identities), and samples c images from each category. Thus, each batch contains $p \times c$ images.

1.1. Contrastive Loss

The batch-wise version of contrastive loss [3] can be reformulated as:

$$\begin{aligned} L_{contrastive} &= \sum_{i,j}^{n} \left[y_{ij} d_{ij} + (1 - y_{ij}) \max\left(0, \varepsilon_{diff} - d_{ij}\right) \right] \\ &= \sum_{i,j}^{n} \boldsymbol{T}_{ij} \left(\boldsymbol{Y}_{ij} \boldsymbol{D}_{ij} + (1 - \boldsymbol{Y}_{ij}) \boldsymbol{D}_{ij} \right) \\ &= \sum_{i,j}^{n} \boldsymbol{T}_{ij} \boldsymbol{D}^{+}_{ij} \;\; + \;\; \sum_{i,j}^{n} \boldsymbol{T}_{ij} \boldsymbol{D}^{-}_{ij}, \end{aligned} \tag{1}$$

where the label $\boldsymbol{y}_{ij} \in \{0, 1\}$ indicates whether the pair of $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is from the same class. The margin parameter ε_{diff} imposes a margin between dissimilar samples. The weighted matrix $\boldsymbol{T}_{ij}$ for mining hard samples is an all-ones matrix in the contrastive loss. The $\boldsymbol{D}^{+}_{ij}$ and $\boldsymbol{D}^{-}_{ij}$ are the distances between similar and dissimilar pairs, respectively.

1.2. Triplet Loss

The batch-wise triplet loss [1] can be reformulated as following:

$$\begin{aligned} L_{trp} &= \sum_{i,j,k|s_i=s_j,s_i\neq s_k}^{n} \left[\boldsymbol{D}_{ij} - \boldsymbol{D}_{ik} + \varepsilon_{trp} \right]_+ \\ &= \sum_{i,j,k|s_i=s_j,s_i\neq s_k}^{n} \left(\left[\boldsymbol{D}_{ij} - \varepsilon^{i}_{same} \right]_+ + \left[\varepsilon^{i}_{diff} - \boldsymbol{D}_{ik} \right]_+ \right) \\ &= \sum_{i,j|s_i=s_j}^{n} (n - c) \left[\boldsymbol{D}_{ij} - \varepsilon^{i}_{same} \right]_+ \\ &+ \sum_{i,j|s_i\neq s_j}^{n} c \left[\varepsilon^{i}_{diff} - \boldsymbol{D}_{ij} \right]_+ \\ &= \sum_{i,j}^{n} \boldsymbol{T}_{ij} \boldsymbol{D}^{+}_{ij} \;\; + \;\; \sum_{i,j}^{n} \boldsymbol{T}_{ij} \boldsymbol{D}^{-}_{ij}, \end{aligned} \tag{2}$$

where $[\cdot]_+ = max(\cdot, 0)$ the margin parameter ε_{trp} denotes a margin threshold in the triplet loss. It can be regarded as the relative distance which is the difference between the absolute distance of negative samples ε^{i}_{diff} and positive samples ε^{i}_{same}. Therefore, the constraint condition is $\varepsilon^{i}_{diff} - \varepsilon^{i}_{same} = \varepsilon_{trp}$. For each positive pair, it has $(n - c)$ choices to select the corresponding negative pair. Thus, if each sample in the batch is regarded as an anchor, $c(n - c)$ triplets could be structured. Hence, the sample imbalance ratio is $\frac{c}{n-c}$ in this type of loss objective.

1.3. Batch Hard Triplet Loss

The batch hard triplet loss [4] can be reformulated as:

$$\begin{aligned} L_{BHT} &= \sum_{i}^{n} \left[\max_{j=1\ldots n, s_i=s_j} \boldsymbol{D}_{ij} - \min_{k=1\ldots n, s_i\neq s_k} \boldsymbol{D}_{ik} + \varepsilon_{trp} \right]_+ \\ &= \sum_{i}^{n} \max_{j=1\ldots n, s_i=s_j} \left[\boldsymbol{D}_{ij} - \varepsilon^{i}_{same} \right]_+ \\ &+ \sum_{i}^{n} \max_{j=1\ldots n, s_i\neq s_j} \left[\varepsilon^{i}_{diff} - \boldsymbol{D}_{ij} \right]_+ \\ &= \sum_{i,j}^{n} \boldsymbol{T}_{ij} \boldsymbol{D}^{+}_{ij} \;\; + \;\; \sum_{i,j}^{n} \boldsymbol{T}_{ij} \boldsymbol{D}^{-}_{ij}, \end{aligned} \tag{3}$$

Figure 1: Visualization of an exclusive hard positive and negative pair selected via the batch-wise (i.e., 32×32 correspondence) MVP matching from a batch of samples. The leftmost image is the anchor sample, and the right is a batch of samples (with batch size 32). For each anchor image, the hard similar positive and dissimilar negative images selected by the MVP matching are marked with yellow and red borders, respectively. Please also refers to the electronically edition (2×2 enlargement in PDF) for better visual effect.

where the term of batch hard means to select the hardest positive and negative pairs in a batch. Therefore, the weighted matrix $\boldsymbol{T}_{ij}$ is a sparse matrix. Each row of the matrix has only two non-zero elements with the value of 1. But each column may appear many ones.

1.4. Quadruplet Loss

The batch-wise form of quadruplet loss [2] could be reformulated as:

$$\begin{aligned} L_{quad} &= \sum_{i,j,k|s_i=s_j,s_i\neq s_k}^{n} [\boldsymbol{D}_{ij} - \boldsymbol{D}_{ik} + \varepsilon_{trp}]_+ \\ &+ \sum_{i,j,k,l|quad}^{n} [\boldsymbol{D}_{ij} - \boldsymbol{D}_{lk} + \varepsilon_{quad}]_+ \\ &= L_{triplet} + \sum_{i,j|s_i=s_j}^{n} X\left[\boldsymbol{D}_{ij} - \varepsilon^i_{same}\right]_+ \\ &+ \sum_{i,j|s_i\neq s_j}^{n} Y\left[\varepsilon^{ij}_{diff} - \boldsymbol{D}_{ij}\right]_+ \\ &= L_{triplet} + \sum_{i,j}^{n} \boldsymbol{T}_{ij}\boldsymbol{D}^+_{ij} + \sum_{i,j}^{n} \boldsymbol{T}_{ij}\boldsymbol{D}^-_{ij} \\ X &= \boldsymbol{C}^2_{n-c} - (p-1)\boldsymbol{C}^2_c \\ Y &= (p-2)\boldsymbol{C}^2_c, \end{aligned} \tag{4}$$

where the margin parameter ε^{ij}_{diff} is a threshold of distances among dissimilar pairs of the sample i and j in the quadruplet loss. The function $\boldsymbol{C}$ is a combination function. According to combinatorial mathematics, each negative pair has $(p-2)\boldsymbol{C}^2_c$ positive pairs and each positive pair has $\boldsymbol{C}^2_{n-c} - (p-1)\boldsymbol{C}^2_c$ corresponding negative pairs to construct a quadruplet. This type of loss can be considered as fine-tuning on $L_{triplet}$ with the help of second and third term in Equation (4) which provide re-margin and re-weight. However, the re-weight matrix $\boldsymbol{T}_{ij}$ does not consider hard positive mining and sample balance. The imbalance ratio is $\frac{X}{Y}$.

1.5. Lifted Loss

The lifted loss [5] without smooth max function could be transformed as:

$$\begin{aligned} L_{lifted} &= \sum_{i,j|s_i=s_j}^{n} \left[\boldsymbol{D}_{ij} - \min_{s_k\neq s_i,s_j} (\boldsymbol{D}_{ik}, \boldsymbol{D}_{jk}) + \varepsilon_{trp}\right]_+ \\ &= \sum_{i,j|s_i=s_j}^{n} \left[\boldsymbol{D}_{ij} - \varepsilon^i_{same}\right]_+ \\ &+ \sum_{i,j|s_i=s_j}^{n} \left[\max_{s_k\neq s_i,s_j} \left(\varepsilon^i_{diff} - \boldsymbol{D}_{ik}, \varepsilon^i_{diff} - \boldsymbol{D}_{jk}\right)\right]_+ \\ &= \sum_{i,j}^{n} \boldsymbol{T}_{ij}\boldsymbol{D}^+_{ij} + \sum_{i,j}^{n} \boldsymbol{T}_{ij}\boldsymbol{D}^-_{ij}. \end{aligned} \tag{5}$$

This type of loss can be viewed as a t-triplet, where t is the number of positive samples in a batch. Each row of the weighted matrix $\boldsymbol{T}_{ij}$ in L_{lifted} has $2t$ non-zero elements with value 1. The half of 1-elements is located according to positive distance matrix $\boldsymbol{D}^+_{ij}$, and the remaining t 1-elements is determined by negative distance matrix $\boldsymbol{D}^-_{ij}$. Actually, each mini batch has t positive pairs and $(\boldsymbol{C}^2_n - t)$ negative pairs. Therefore, the lifted loss just considers sample balance but does not contain hard positives mining.

1.6. N-pair Loss

N-pair loss [6] proposes an (N+1)-tuplet, where N is #classes in a tuplet. It needs pairs with a unique class label to build such a tuplet and thus has limitations in practice. It does not do hard-mining, and $\boldsymbol{T}$ is an all-one matrix.

1.7. Batch-wise Optimal Transport Loss

The batch-wise optimal transport loss [8] could be formulated as

$$\begin{aligned} L_{OT} &= \sum_{i,j|s_i=s_j}^{n} \boldsymbol{T}_{ij}\boldsymbol{D}_{ij} + \sum_{i,j|s_i\neq s_j}^{n} \boldsymbol{T}_{ij}\left[\varepsilon^i_{diff} - \boldsymbol{D}_{ij}\right]_+ \\ &= \sum_{i,j}^{n} \boldsymbol{T}_{ij}\boldsymbol{D}^+_{ij} + \sum_{i,j}^{n} \boldsymbol{T}_{ij}\boldsymbol{D}^-_{ij}. \end{aligned} \tag{6}$$

The batch-wise optimal transport loss is inspired by the optimal transport programming [7]. It can utilize all available semantical information within training batches and learn an importance weighted matrix $\boldsymbol{T}_{ij}$ via the Sinkhorn's algorithm. The optimal weighted matrix is a probability distribution. However, this loss may encounter that the similar positive pairs with small distances would still be optimized (i.e., overtraining). Moreover, it does not consider the sample balance issue, and the imbalance ratio is $c/(n-c)$.

1.8. Brief Summary

The innovations of above mentioned metric learning loss objectives could be summarized in the following aspects including the definition of distance metric or margin, re-weighting hard samples, and sampling balance between positives and negatives.

Our proposed batch-wise MVP matching based loss objective take these three points into consideration comprehensively. Unlike the batch hard triplet loss, the optimal MVP matching $\boldsymbol{T}^*$ is a re-weighting matrix, where each row and column has only two non-zero values. This means the MVP matching will not favor any sample and can avoid the outliers to dominate the training process. Conversely, the batch hard triplet loss just guarantee each row has only two non-zero values. However, lots of non-zero values may appear in the same column (i.e., pick the same hard sample) when encountering outliers. These outliers may dominate

gradient and lead to model collapse during training. Thus, the MVP loss makes the training process more stable and hard samples mining more balanced.

2. Visualization of Hard Sample Pairs

In our proposed MVP matching based loss objective, 4 images (i.e., c images) resized to 256×128 from each of 8 (i.e., p categories) persons are picked randomly to construct a 32-size batch. The visualization of results is shown in Figure 1. An interesting note is that the hard similar positive pairs selected via the MVP matching within a batch are often with the intensive appearance variations, e.g., human poses, scale, and viewpoints, while the hard dissimilar negative pairs are usually with the similar appearance.

References

[1] Gal Chechik, Varun Sharma, Uri Shalit, and Samy Bengio. Large scale online learning of image similarity through ranking. *Journal of Machine Learning Research*, 11(Mar):1109–1135, 2010. 1

[2] Weihua Chen, Xiaotang Chen, Jianguo Zhang, and Kaiqi Huang. Beyond triplet loss: a deep quadruplet network for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 403–412, 2017. 3

[3] Sumit Chopra, Raia Hadsell, and Yann LeCun. Learning a similarity metric discriminatively, with application to face verification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 539–546. IEEE, 2005. 1

[4] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017. 1

[5] Hyun Oh Song, Yu Xiang, Stefanie Jegelka, and Silvio Savarese. Deep metric learning via lifted structured feature embedding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4004–4012, 2016. 3

[6] Kihyuk Sohn. Improved deep metric learning with multi-class n-pair loss objective. In *Advances in Neural Information Processing Systems*, pages 1857–1865, 2016. 3

[7] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008. 3

[8] Lin Xu, Han Sun, and Yuai Liu. Learning with batch-wise optimal transport loss for 3d shape recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3333–3342, 2019. 3