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

Abstract

How to correctly stress hard samples in metric learning is critical for visual recognition tasks, especially in challenging person re-ID applications. Pedestrians across cameras with significant appearance variations are easily confused, which could bias the learned metric and slow down the convergence rate. In this paper, we propose a novel weighted complete bipartite graph based maximum-value perfect (MVP) matching for mining the hard samples from a batch of samples. It can emphasize the hard positive and negative sample pairs respectively, and thus relieve adverse optimization and sample imbalance problems. We then develop a new batch-wise MVP matching based loss objective and combine it in an end-to-end deep metric learning manner. It leads to significant improvements in both convergence rate and recognition performance. Extensive empirical results on five person re-ID benchmark datasets, i.e., Market-1501, CUHK03-Detected, CUHK03-Labeled, Duke-MTMC, and MSMT17, demonstrate the superiority of the proposed method. It can accelerate the convergence rate significantly while achieving state-of-the-art performance. The source code of our method is available at https://github.com/IAAI-CVResearchGroup/MVP-metric.

1. Introduction

Person re-identification (re-ID) is a hot yet challenging research topic in computer vision [5, 46, 1, 25, 53]. Recently, with the remarkable progresses in deep metric learning [13, 11, 33, 41, 19, 52], many advanced methods [21, 45, 6, 35, 40, 44] have been developed for visual recognition. The joint learned deep feature representation and semantical embedding metric yield significant improvements in the community of person re-ID [26, 55, 19, 24, 22].

The core challenge lies in person re-ID is how to spot the same pedestrian accurately across different disjoint cameras under intensive variations of appearance, such as human poses, illumination conditions, and camera viewpoints. Figure 1.(a) illustrates some typical easily confused *hard samples* in the *Market-1501* dataset. They could be broadly divided into three categories, i.e., the appearance of differ-

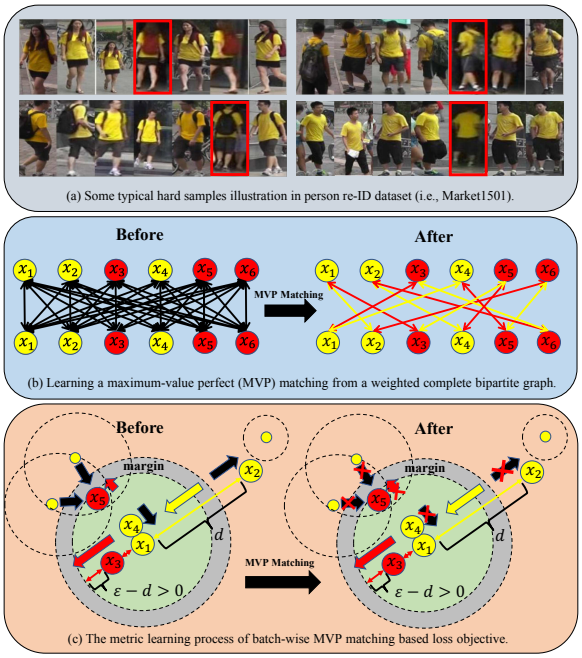

Figure 1: Schematic illustration of learning with the proposed MVP matching for emphasizing hard positive and negative sample pairs. **(a):** Typical hard inter-class and intra-class samples in the *Market-1501* dataset. **(b):** Learning an MVP matching from a weighted complete bipartite graph for mining hard positive and negative sample pairs. The colors of particles represent semantical (or category) information. **(c):** The metric learning process of our batch-wise MVP matching based loss objective. The learned metric is optimized within batches so that positive pairs with large distances and negative pairs with small distances are emphasized as red and yellow arrows shown. Learning with the MVP matching guarantees that only one exclusive hard positive and negative pairs are selected simultaneously. The adverse optimization including overtraining (e.g., x_1 and x_4) and counteracting (e.g., x_2 and x_5) as black arrows shown can be eliminated effectively. The length of arrows indicates weights optimized by the proposed loss objective.

*Contact Author

ent pedestrians may be highly similar, the pose of a person may vary significantly as time and space changed, and the light conditions taken by some cameras are sometimes poor. These hard samples would strongly slow down the convergence rate of the metric learning, which works on pulling similar samples to cluster together while pushing dissimilar ones to widen apart. Or worse of all, the learned embedding metric and feature representation could be heavily biased by these hard samples. Most of the pre-existing metric learning methods still have some limitations on this issue. For instance, seminal contrastive loss [13] or triplet loss [11] learns the semantical information within image pairs or triplets based on the siamese-like networks [9]. They do not make full use of all available information within a batch of samples. Batch all triplet loss [19] and N-pair loss [41] have been developed to remedy this flaw, but they do not attach enough attention to hard samples and require expensive re-sampling techniques to boost the performance. Lifted loss [33] and quadruplet loss [12] only consider hard negative samples mining while ignoring the hard positive samples. Batch hard triplet loss [19] considers the hardest positive and negative mining depended on the distances of features simply. Its performance is easily influenced by some outlier samples (e.g., indistinguishable or mislabeled images in person re-ID datasets), which could be regarded as hardest sample pairs by many other samples simultaneously and lead to oscillation during metric learning process.

In this paper, we propose a novel weighted complete bipartite graph based maximum-value perfect (MVP) matching for mining hard sample pairs within metric learning framework. The primary motivation of the MVP matching is how to correctly capture these inter-class and intra-class hard samples in person re-ID datasets. As illustrated in Figure 1.(b), we first construct a complete bipartite graph [2] from a batch of samples, whose vertices (i.e., samples) can be divided into two disjoint and independent sets such that each edge (i.e., the weights of samples) connects a vertex from a set to one in another set. Then an MVP matching (i.e., a bijection with one-to-one correspondence) could be found in this weighted bipartite graph with the Kuhn-Munkres assignment (KA) algorithm [20]. We learn two MVP matchings as yellow and red bi-directional arrows shown in the weighted bipartite graph to emphasize the hard positive and negative pairs, respectively. We further formulate a batch-wise loss objective based on the proposed MVP matching for deep metric learning. Figure 1.(c) illustrates the metric learning process of the proposed loss objective schematically. As mentioned, conventional batch-wise loss objectives [41, 33, 19, 52] for metric learning can be optimized using all available information within training batches, so that all similar positive pairs with large ground distances and dissimilar negative pairs with small ground distances would be emphasized simultaneously. However, these methods may encounter overtraining. Similar positive pairs with small distances would still be optimized (e.g., x_1 and x_4). Or worse, if we treat the optimization as a whole rather than the individual particle, the metric learning process of these methods is vulnerable to oscillation. Since the hard samples might be emphasized by many particles simultaneously, they could all cancel each other out (e.g., x_2 and x_5). In contrast, metric learning with the MVP matching based loss objective can guarantee that each sample selects one exclusive hard positive and negative pairs. Then the adverse optimization, e.g., overtraining and counteracting yielded by other anchors as black arrows shown in Figure 1.(c) would be effectively eliminated. As a consequence, the convergence rate and recognition performance of metric learning could be improved significantly. We finally evaluate the performance of our proposed method on five widely used benchmark datasets, i.e., *Market-1501* [56], *CUHK03-Detected* [23], *CUHK03-Labeled*[23], *Duke-MTMC* [38], and *MSMT17* [49]. Experimental results demonstrate that the proposed method can accelerate the convergence rate significantly while achieving state-of-the-art performance.

In a nutshell, our main contributions in the present work can be summarized as follows:

(1) We propose a novel weighted complete bipartite graph based maximum-value perfect (MVP) matching for mining hard samples. It can emphasize the hard positive and negative sample pairs respectively and thus relieve adverse optimization and sample imbalance problems.

(2) We develop a new batch-wise MVP matching based loss objective and combine it into an end-to-end fashion for deep metric learning. It leads to significant improvements in both convergence rate and recognition performance.

(3) We verify the superiority of our proposed methods on person re-ID datasets. Our method can achieve state-of-the-art performances with a notably fast convergence rate.

2. Related Work

Person re-ID: The research works on person re-ID mainly focus on the visual feature extraction and similarity metric learning. The traditional feature representations are based on hand-crafted methods, such as color histogram [17], SIFT [29], LOMO [26], etc. Recently, with the developments on deep learning techniques and large-scale person re-ID benchmark datasets, e.g., *Market-1501* [56], *CUHK-03* [23], *Duke-MTMC* [38], *MSMT17* [49] etc., many advanced methods have been proposed. For instance, features of a pedestrian image split into three horizontal parts are extracted by a siamese CNN, and the cosine distance metric measures the similarity of features from different images. FaceNet [40] consists of a batch of images and a deep CNN backbone followed by ℓ_2 normalization, which transforms input to a mensurable Euclidean space. Deep-person [4] is proposed to apply LSTM structure in an end-to-end way to

model pedestrian images, seen as a sequence of body parts from head to foot. BraidNet [47] proposes a deep CNN with specially designed cascaded WConv layers to extract features. PSE network [39] contains 3 channels of RGB information and 14 channels of pose information with acquired camera view and the detected joint locations, which helps to learn an effective representation. A discriminative identity loss combined with the verification loss objective [57] also shows a superior recognition performance.

Deep metric learning: In deep metric learning, deep visual features and semantical embedding metric can be learned jointly [23, 22]. Inspired by the contrastive loss [13] and triplet loss [11], many improved margin-based loss objectives have been proposed for learning, e.g., quadruplet loss [12], margin sample mining loss [51], lifted loss [33], etc. These methods adopt margins as hyper-parameters tuned before training to control the distance between pairs of samples. To make the fixed margin able to adjust feasibly, the average distances of positive and negative pairs in a batch represent margin thresholds adaptively in [12]. Furthermore, a learnable variable is introduced to determine the boundary between positive and negative pairs [50]. Meanwhile, metric learning methods are sensitive to the selection of hard pairs or triplets. Therefore, hard samples mining is an essential and important element of [19, 12, 51]. The semi-hard [40] and batch hard [19] loss objectives select the hard triplets based on the margins directly. Since sampling matters to learn deep embeddings, distance-weighted sampling is proposed in [50] and thus steadily produces informative examples while controlling the variance. Considering weighting samples according to a novel ground distance metric, the batch-wise sample-weighted matrix is learned in the optimal transport problem framework [52]. Generally, these methods pick out dissimilar positive pairs (a.k.a., hard positive samples) and similar negative pairs (a.k.a., hard negative samples) according to similarity scores.

Matching in a graph: In the mathematical discipline of graph theory [8], a matching within a graph is a set of edges without common vertices. Finding the maximum weighted matching [30] in a weighted bipartite graph is one of the fundamental combinatorial optimization problems [37]. It is crucial both in theoretical and practical. On one hand, it is a special case of more complex problems, such as the generalized assignment problem [31], minimum cost flow, and network flow problem [15]. On the other hand, many real-world problems can be categorized in a matching problem, such as worker assignment problem [32].

3. Our Method

3.1. Adaptive Weighting for Positives and Negatives

We construct a complete bipartite graph [2] $G(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{E})$ based on a batch of samples. As illustrated in Figure 2, the vertices (i.e., samples) in the graph can be divided into two disjoint and independent sets $\boldsymbol{U}$ and $\boldsymbol{V}$ such that each edge (i.e., the weights of sample) in set $\boldsymbol{E}$ connects a vertex in $\boldsymbol{U}$ to one in $\boldsymbol{V}$. Considering the fact that the number of dissimilar negative pairs is often much more than similar positive pairs within a training batch (a.k.a., sample imbalance problem [48]), we further divide the graph G into two bipartite graphs G_P and G_N for positive and negative pairs respectively. We also define an adaptive weight $\boldsymbol{M}_{ij}^{+}$ and $\boldsymbol{M}_{ij}^{-}$ of each edge (i, j) in G_P and G_N. Specifically, the edge weight for positive pairs can be defined as

$$\boldsymbol{M}_{ij}^{+}(\boldsymbol{x}_i, \boldsymbol{x}_j; f) = \max\{0, ||f(\boldsymbol{x}_i) - f(\boldsymbol{x}_j)||_2^2 - \alpha\}, \quad (3.1)$$

where α is a learnable variable that denotes the margin of similar positive samples. The hinge loss function $\max\{0, ||f(\boldsymbol{x}_i) - f(\boldsymbol{x}_j)||_2^2 - \alpha\}$ penalizes similar samples beyond the margin α. It would prevent the overtraining for positive samples in contrastive loss, which demands similar pairs gather as close as possible. For dissimilar negative samples, we define the edge weight correspondingly as

$$\boldsymbol{M}_{ij}^{-}(\boldsymbol{x}_i, \boldsymbol{x}_j; f) = \max\{0, \beta - ||f(\boldsymbol{x}_i) - f(\boldsymbol{x}_j)||_2^2\}, \quad (3.2)$$

where $\beta = \varepsilon + \alpha$ determines the margin of negative pairs and the hyper-parameter ε controls the relative distance between positive and negative margins. The hinge loss $\max\{0, \beta - ||f(\boldsymbol{x}_i) - f(\boldsymbol{x}_j)||_2^2\}$ penalizes the dissimilar pairs within the margin β and ignores the others.

3.2. MVP Matching for Hard Samples Pairs

We define the matching variables $\boldsymbol{T}_{ij}^{+}$ and $\boldsymbol{T}_{ij}^{-}$ for the edge (i, j) on these two weighted bipartite graphs G_P and G_N. The element $t_{ij} \in \boldsymbol{T}^{+(or-)} = 1$ indicates a maching pair, while $t_{ij} = 0$ means unmatching one. Thus the total weights of positive and negative pairs are $\sum_{ij} \boldsymbol{T}_{ij}^{+} \boldsymbol{M}_{ij}^{+}$ and $\sum_{ij} \boldsymbol{T}_{ij}^{-} \boldsymbol{M}_{ij}^{-}$. Our goal is to find a maximum-value perfect (MVP) matching for assigning positive and negative pairs, respectively. Obviously, each vertex is adjacent to one edge in the matching exactly, which can be formulated as linear constraints, i.e., $\sum_j \boldsymbol{T}_{ij} = 1$ for $i \in \boldsymbol{U}$ and $\sum_i \boldsymbol{T}_{ij} = 1$ for $j \in \boldsymbol{V}$. Then, the MVP matching in these two weighted bipartite graphs G_P and G_N can be formulated as

$$\begin{aligned} &\max_{\boldsymbol{T}_{ij} \geq 0} \sum\nolimits_{i,j=1}^{n} \boldsymbol{T}_{ij}^{+(\text{or}-)} \boldsymbol{M}_{ij}^{+(\text{or}-)} \\ &\text{s.t.} \quad \sum\nolimits_{j=1}^{n} \boldsymbol{T}_{ij}^{+(\text{or}-)} = 1, \quad \sum\nolimits_{i=1}^{n} \boldsymbol{T}_{ij}^{+(\text{or}-)} = 1 \end{aligned} \quad (3.3)$$

The process of optimizing this MVP matching is also called as an *assignment problem* [20, 31], which is a fundamental combinatorial optimization problem [10, 60]. It consists of finding a perfect matching (i.e., one-to-one matching) where the sum of edge weights is maximum in a weighted

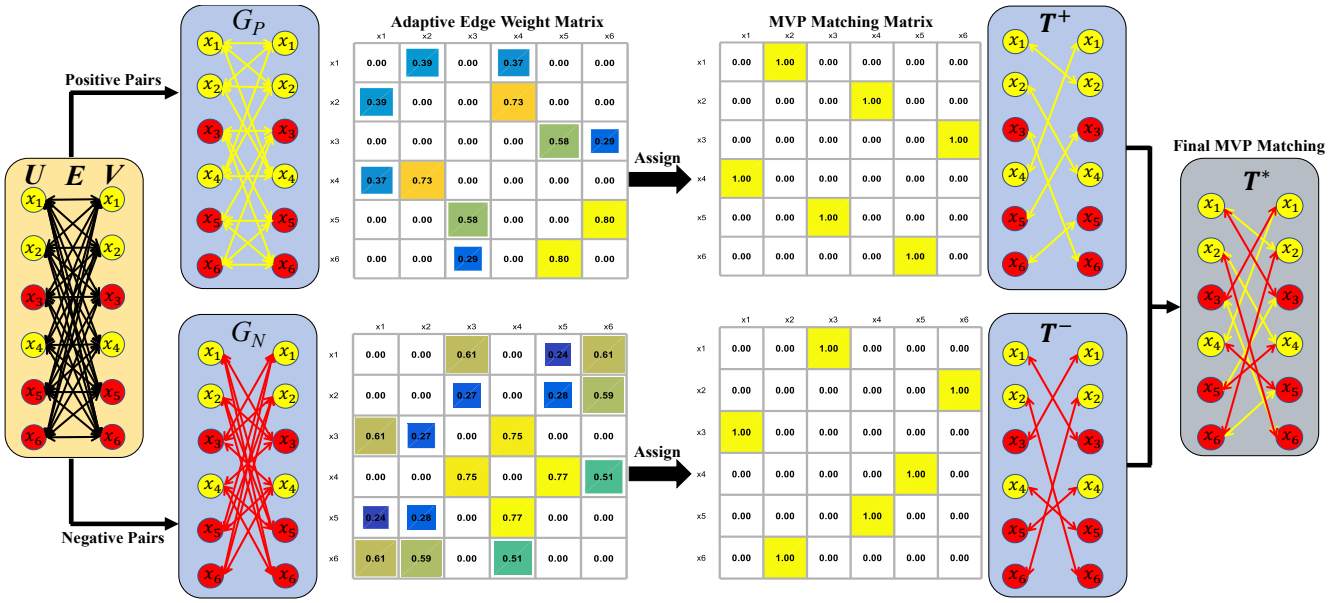

Figure 2: Schematic illustration of our proposed MVP matching for mining hard positive and negative sample pairs. The two sets $\boldsymbol{U}$ and $\boldsymbol{V}$ could be thought as samples with two categories (i.e., yellow or red). A complete bipartite graph can divide into subgraphs G_P and G_N for positive and negative pairs. The positive and negative MVP matching matrices $\boldsymbol{T}^+$ and $\boldsymbol{T}^-$ are assigned according to the corresponding pre-defined adaptive edge weight matrices $\boldsymbol{M}^+$ and $\boldsymbol{M}^-$. Each row and column of $\boldsymbol{T}^+$ and $\boldsymbol{T}^-$ has only one non-zero element, which is used to indicate the position of the hard positive and negative pairs with maximum-value of edge weights. The final MVP matching solution $\boldsymbol{T}^*$ combines $\boldsymbol{T}^+$ and $\boldsymbol{T}^-$ to guarantee that one exclusive hard positive and negative pairs are selected simultaneously.

bipartite graph. The learned matching could emphasize the hard positive and negative sample pairs respectively.

A naive solution for the MVP matching problem in Equation (3.3) is to check all possible assignments and calculate the cost of each one. However, this is very inefficient, since there are $n!$ (i.e., factorial of n) different assignments with n-pair of samples. Many algorithms [34, 7] have been developed for solving the assignment problem in a polynomial time bounded of n. The Kuhn-Munkres assignment (KA) algorithm [20] is one of the most popular algorithms [10, 60], which can find the global optimal matching in Equation 3.3 with the complexity $\mathcal{O}(n^3)$. Considering a complete weighted bipartite graph G, where the weight of $edge\,(i,j) \in \boldsymbol{E}$ is denoted as $\boldsymbol{M}_{ij}$, we define a labeling function $\ell : \boldsymbol{U} \cup \boldsymbol{V} \rightarrow \mathcal{R}$ for each vertex. The detailed procedures of KA algorithm are described in Algotithm 1. For explaining KA algorithm how to find an MVP matching in a bipartite graph, we first introduce definitions of the feasible labeling and equality graph.

Definition 3.1. *Feasible labeling: For each vertex in the graph G, a labeling $\ell : \forall x_i, x_j \in \boldsymbol{U} \cup \boldsymbol{V} \rightarrow \mathcal{R}$ is defined to compute the vertex labeling value. The feasible labeling demands the weight $\boldsymbol{M}_{ij}$ of any edge (i, j) to satisfy*

$$\boldsymbol{M}_{ij} \leq \ell(x_i) + \ell(x_j). \tag{3.4}$$

Definition 3.2. *Equality graph: The summation of labeling values between any two vertexes equals to the weight for the corresponding edge in the equality graph G_ℓ (with respect to ℓ), which is represented as*

$$G_\ell = \{\forall (i,j) : \boldsymbol{M}_{ij} = \ell(x_i) + \ell(x_j)\}. \tag{3.5}$$

Theorem 3.3. *If ℓ is a feasible labeling function and $\boldsymbol{T}$ is*

Algorithm 1 Kuhn-Munkres Assignment (KA) Algorithm

Input: A complete weighted bipartite graph and a weighted matrix denoted as $G\,(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{E})$ and $\boldsymbol{M}_{ij}$.
Output: An optimal maximum-value perfect (MVP) matching $\boldsymbol{T}^*$.
Step 1. Initialization:
Generate initial values of labeling function l. We set the initial labeling values of vertexes in the set $\boldsymbol{U}$ and $\boldsymbol{V}$ as $+\infty$ and 0, respectively.
Step 2. Checking:
If $\boldsymbol{T}$ is an arbitrary perfect matching (i.e., one-to-one matching) in G_ℓ, assignment terminates. Otherwise pick unmatching vertex $x_i \in \boldsymbol{U}$.
Step 3. Labeling:
If augment from x_i by the Hungarian method [20] unsuccessfully, update the labeling value greedily as

$$\min\{l(x_i) + l(x_j) - \boldsymbol{M}_{ij}\,(x_i, x_j)\}.$$

Step 4. Iteration:
Augment from x_i successfully, update $\boldsymbol{T}$, go to Step 2.

an arbitrary perfect matching in the equality graph G_ℓ, then $\boldsymbol{T}$ must be a maximum-value perfect (MVP) matching $\boldsymbol{T}^*$.

Proof. For an arbitrary perfect (i.e., one-to-one) matching $\boldsymbol{T}$ in a weighted graph $G(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{E})$, each vertex is covered only once. Thus, the edge weight $\boldsymbol{M}_{ij}$ of each matching (i, j) satisfies the condition in Equation 3.4. The equality holds only when $edge(x_i, x_j) \in G_\ell$. We denote the summation of all labeling values as K. Therefore, the summation of all edge weights in this matching $\boldsymbol{T}$ satisfies

$$\sum\nolimits^{(x_i,x_j)\in \boldsymbol{T}} \boldsymbol{M}_{ij} \leq \sum\nolimits_i^{x_i\in \boldsymbol{U}} \ell(x_i) + \sum\nolimits_j^{x_j\in \boldsymbol{V}} \ell(x_j) = K. \tag{3.6}$$

Only when all matching edges are in G_ℓ, the equality in Equation 3.6 is obtained. Hence, when $\boldsymbol{T}$ is an arbitrary perfect matching in G_ℓ, the summation of all edge weights reaches the maximum value K and the maximum-value perfect (MVP) matching $\boldsymbol{T}^*$ is received as

$$\sum\nolimits^{(x_i,x_j)\in \boldsymbol{T}^*} \boldsymbol{M}_{ij} = \sum\nolimits_i^{x_i\in \boldsymbol{U}} \ell(x_i) + \sum\nolimits_j^{x_j\in \boldsymbol{V}} \ell(x_j) = K. \tag{3.7}$$

The two matrices $\boldsymbol{T}^+$ and $\boldsymbol{T}^-$ on graphs G_P and G_N as shown in Figure 2 are the MVP matchings for hard positive and negative sample pairs, which have a value of 1 in each column and row. The combination $\boldsymbol{T}^*$ ensures each sample selects one exclusive hard positive and negative samples. □

3.3. Batch-wise MVP Matching based Loss

Finally, we formulate a batch-wise loss objective based on the proposed MVP matching for metric learning as

$$\begin{aligned} &\mathcal{L}(\boldsymbol{x}_i, \boldsymbol{x}_j; f) = \mathcal{L}^+ + \mathcal{L}^- \\ &= \sum\nolimits_{ij}^n y_{ij} \boldsymbol{T}_{ij}^+ \boldsymbol{M}_{ij}^+ + \sum\nolimits_{ij}^n (1 - y_{ij}) \boldsymbol{T}_{ij}^- \boldsymbol{M}_{ij}^-, \end{aligned} \tag{3.8}$$

where y_{ij} is a binary label assigned to a pair of samples. Let $y_{ij} = 1$ if samples x_i and x_j are deemed similar, and $y_{ij} = 0$ otherwise. The $\boldsymbol{M}_{ij}^+$ and $\boldsymbol{M}_{ij}^-$ are adaptive edge weights. The MVP matching $\boldsymbol{T}_{ij}^+$ and $\boldsymbol{T}_{ij}^-$ can be learned for emphasizing hard positive and negative pairs.

We minimize the proposed loss objective with batch gradient descent. The gradients of loss function $\mathcal{L}(\boldsymbol{x}_i, \boldsymbol{x}_j; f)$ with respect to the input feature embedding representations $f(\boldsymbol{x}_i)$ and $f(\boldsymbol{x}_j)$ at each update are computed as

$$\frac{\partial \mathcal{L}}{\partial f(\boldsymbol{x}_i)} = \sum_{j=1}^n 2(f(\boldsymbol{x}_i) - f(\boldsymbol{x}_j))(\delta_{ij}^+ y_{ij} \boldsymbol{T}_{ij}^+ - \delta_{ij}^- (1 - y_{ij}) \boldsymbol{T}_{ij}^-),$$

$$\frac{\partial \mathcal{L}}{\partial f(\boldsymbol{x}_j)} = -\sum_{i=1}^n 2(f(\boldsymbol{x}_i) - f(\boldsymbol{x}_j))(\delta_{ij}^+ y_{ij} \boldsymbol{T}_{ij}^+ - \delta_{ij}^- (1 - y_{ij}) \boldsymbol{T}_{ij}^-),$$

where δ_{ij}^+ and δ_{ij}^- are binary indicators assigned to the pairs of samples. If the edge weight for similar pairs in Equation (3.1) is larger than 0, δ_{ij}^+ outputs 1. The indicator of the edge weight for dissimilar pairs in Equation (3.2) is denoted by δ_{ij}^-. During optimization, $\boldsymbol{T}_{ij}^+$ and $\boldsymbol{T}_{ij}^-$ are solved by the KA algorithm and not considered as variables to compute gradients. In conventional double-margin contrastive loss [27], the margins are manually selected. However, it is feasible to learn the positive margin by the variable α in our loss objective. The gradient with α could be computed easily as

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \alpha} = \sum_{i,j=1}^n &-y_{ij} \mathbf{1}\{||f(\boldsymbol{x}_i) - f(\boldsymbol{x}_j)||_2^2 - \alpha > 0\} + \\ &(1 - y_{ij}) \mathbf{1}\{\varepsilon + \alpha - ||f(\boldsymbol{x}_i) - f(\boldsymbol{x}_j)||_2^2 > 0\} \end{aligned}$$

As far as $\frac{\partial \mathcal{L}}{\partial \alpha}$, $\frac{\partial \mathcal{L}}{\partial f(\boldsymbol{x}_i)}$ and $\frac{\partial \mathcal{L}}{\partial f(\boldsymbol{x}_j)}$ are derived, the gradient of network parameters could be easily computed with the back-propagation method. Therefore, the whole network can be trained end-to-end discriminatively.

4. Experiments

In this section, we evaluated the performance of our proposed methods with applications to person re-ID tasks. Five widely used benchmark datasets were employed in our experiments, including *Market-1501*, *CUHK03-Labeled*, *CUHK03-Detected*, *Duke-MTMC* and *MSMT17*.

4.1. Datasets

CUHK03 [23] is a large person re-ID dataset from the Chinese University of Hong Kong (CUHK). The whole dataset including 13,164 images of 1,360 pedestrians is captured with six surveillance cameras. Images are recorded from these cameras for several months. The dataset has manually cropped pedestrian images and also provided samples detected with a state-of-the-art pedestrian detector.

Market-1501 [56] is another dataset provided by the Tsinghua University. It totally contains 32,643 bounding boxes grouped 1,501 identities. Images are also recorded from six cameras, including five 1280×1080 HD cameras and one 720×576 SD camera. Since these cameras are placed in an open environment, each identity may have multiple images under the same camera.

Duke-MTMC [38] is a manually annotated and multi-camera video-based dataset. It consists of 8×85 minutes of 1080p video recorded at 60 frames per second from 8 static cameras deployed on the Duke University during periods between lectures. The data has about ten hours of video, more than 2 million frames and 6,791 trajectories for 2,384 different identities. There are 2.5 single-camera trajectories per identity and up to 7 in some case on average.

MSMT17 [49] is a large-scale multi-scene multi-time person re-ID dataset provided by the Peking University. The dataset utilizes a 15-camera network including 12 outdoor and 3 indoor cameras. The complex scenes and backgrounds make this dataset more appealing and challenging. Four days with different weather and lighting conditions are selected during a month for video recording. Finally, $126,441$ bounding boxes of $4,101$ different identities are detected and annotated by Faster *RCNN* [16].

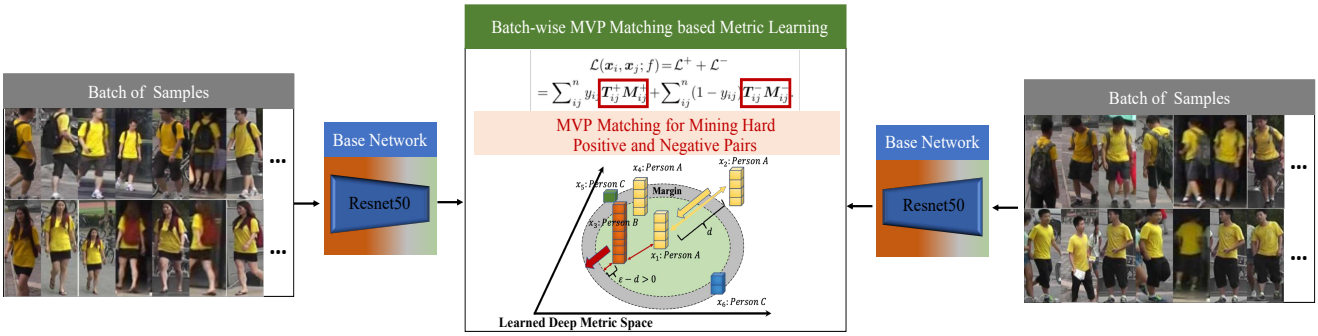

Figure 3: An overview of our proposed batch-wise MVP matching based metric learning framework. Given a batch of samples, we use a siamese-like architecture with *Resnet-50* to transform input images into deep CNN embeddings. The whole network could be trained end-to-end discriminatively with the proposed batch-wise MVP matching based loss objective. The highlighted perfect matching $\boldsymbol{T}_{ij}^{+}\boldsymbol{M}_{ij}^{+}$ and $\boldsymbol{T}_{ij}^{-}\boldsymbol{M}_{ij}^{-}$ are used for emphasizing hard positive and negative sample pairs. For instance, the MVP matching could find the hard positive x_2 and negative x_3 samples with maximum-value weights for an anchor x_1.

	Market-1501			*CUHK03-Detected*			*CUHK03-Labeled*			*Duke-MTMC*			*MSMT17*		
	mAP	Top-1	Top-5	Top-1	Top-5	Top-10	Top-1	Top-5	Top-10	mAP	Top-1	Top-5	mAP	Top-1	Top-5
ID Loss [51]	70.4%	87.9%	92.9%	73.8%	91.4%	95.0%	75.8%	91.8%	95.2%	58.9%	78.3%	88.3%	33.9%	63.2%	74.0%
BA Triplet Loss [40]	72.3%	86.7%	95.8%	84.5%	97.5%	98.8%	87.6%	99.3%	99.7%	62.4%	77.0%	89.5%	27.1%	48.1%	68.5%
BH Triplet Loss [40]	76.9%	89.5%	95.9%	86.1%	97.5%	98.8%	89.1%	99.1%	99.6%	65.8%	79.8%	90.6%	44.5%	69.4%	83.3%
Lifted Loss [33]	75.5%	87.9%	95.4%	86.1%	97.8%	99.0%	90.2%	99.4%	99.7%	64.4%	79.7%	90.4%	41.7%	66.1%	81.3%
Batch-wise OT Loss [52]	76.7%	88.8%	95.5%	89.6%	98.8%	99.1%	92.5%	**99.6%**	99.7%	66.8%	79.6%	91.3%	44.5%	69.5%	82.6%
MVP Loss	**80.5%**	**91.4%**	**96.9%**	**91.8%**	**98.8%**	**99.4%**	**93.7%**	99.4%	99.6%	**70.0%**	**83.4%**	**91.9%**	**46.3%**	**71.3%**	**84.7%**

Table 1: Comparison results with different loss objectives for deep metric learning on five person re-ID benchmark datasets.

4.2. Experimental Settings

Architecture: Figure 3 illustrates the network architecture for our MVP matching learning scheme. The backbone network is *Resnet-50* [18] with 512-dimensional embeddings. For each sample, the MVP matching could find only one exclusive positive and negative pair to compute loss.

Evaluations: Since the person re-ID task can be considered as a sub-problem of image retrieval [28, 55], we used Euclidean distance to measure the similarity of images based on their embedding representations. Given a query identity, a ranked list of the remaining test samples is returned according to their distances to the query. Then, we can calculate evaluation metrics including the mean value of average precision (mAP) and the cumulated matching characteristics (CMC) curve. The mAP for all queries is a common evaluation metric for retrieval, which considers both precision and recall [56]. The CMC curve is another widely used evaluation metric, which shows the probability that a query identity appears in different sized candidate lists. CMC top-k accuracy is 1 if top-k ranked gallery samples contain the query identity, otherwise accuracy is 0.

Parameters settings: The learning rate and batch size were set as 0.0001 and 64, respectively. Our learning rate would decay by factor of 0.1 adaptively and the Adam optimizer had 0.0005 weight decay rate. The relative distance ε in Equation 3.2 was 200, while we initialized the learnable positive margin variable β equaling 200. More detailed parameters settings please refer our released GitHub code [1].

4.3. Comparison with the Different Loss Objectives

Firstly, we empirically compared the performance of our proposed the maximum-value perfect (MVP) matching based loss with other state-of-the-art loss objectives for deep metric learning, e.g., batch all (BA) triplet loss [40], batch hard (BH) triplet loss [40], lifted loss [33], and batch-wise optimal transport (OT) loss [52]. Identification (ID) loss [51] denotes the cross-entropy loss for classification, which is also a baseline experiment provided in the project [2]. The comparison results are summarized in Table 1. All metric learning loss objectives could improve mAP and accuracy significantly. Among the compared methods, our MVP matching loss based method achieves the best retrieval (i.e., mAP) performance on *Market-1501*, *Duke-MTMC*, and *MSMT17*. Meanwhile, the proposed MVP loss also achieves the best CMC top-1 and top-5 accuracy on both *CUHK03-Detected* and *CUHK03-Labeled* datasets.

We then evaluated the convergence rate of the proposed MVP loss compared with the other top three metric learning

[1] https://github.com/IAAI-CVResearchGroup/MVP-metric
[2] https://github.com/KaiyangZhou/deep-person-reid/

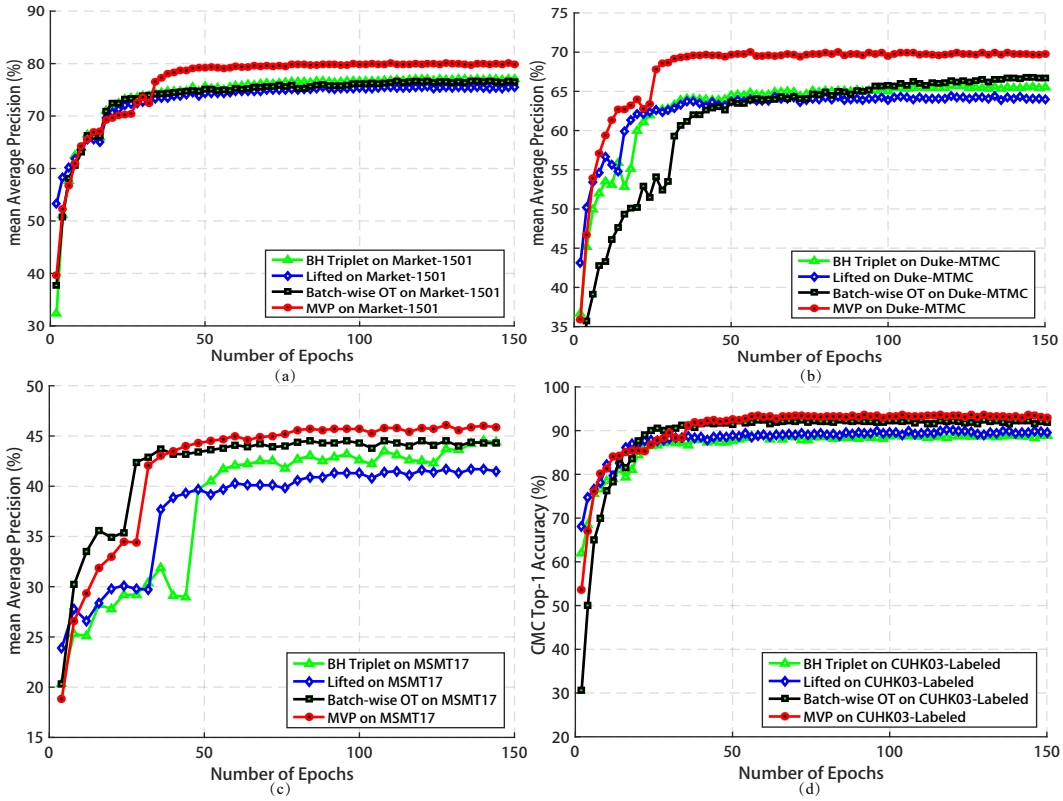

Figure 4: Evaluation metrics curves with respect to the number of epochs restrained by various loss objectives. Subfigure (a), (b) and (c) illustrate the mAP curve on *Market-1501*, *Duke-MTMC* and *MSMT17* respectively. Subfigure (d) represents the CMC top-1 accuracy curve on *CUHK03-Labeled*. The metric is observed each 2 epochs on Duke-MTMC and CUHK03-Labeled, and each 4 epochs on others.

Methods	***Market-1501***		***CUHK03 (Detected)***		***CUHK03 (Labeled)***		***Duke-MTMC***	
	mAP	Top-1	Top-1	Top-5	Top-1	Top-5	mAP	Top-1
BoW + Kissme [56]	20.8%	44.4%	11.7%	33.3%	-	-	12.2%	25.1%
LOMO + XQDA [26]	-	-	46.3%	-	52.2%	-	17.0%	30.8%
Verification + Identification [57]	59.9%	79.5%	83.4%	97.1%	-	-	49.3%	68.9%
MSCAN [22]	57.5%	80.3%	68.0%	91.2%	74.2%	94.3%	-	-
SVDNet [43]	62.1%	82.3%	81.8%	95.2%	-	-	56.8%	76.7%
DPLAR [54]	63.4%	81.0%	-	-	85.4%	97.6%	-	-
PDC [42]	63.4%	84.1%	78.3%	94.8%	88.7%	98.6%	-	-
JLMT [24]	65.5%	85.1%	80.6%	96.9%	83.2%	98.0%	-	-
SSM [3]	68.8%	82.2%	72.7%	92.4%	76.6%	94.6%	-	-
MuDeep [36]	-	-	75.6%	94.4%	-	76.9%	-	-
FMN [14]	67.1%	86.0%	42.6%	56.2%	40.7%	54.5%	56.9%	74.5%
PAN [58]	63.4%	82.8%	36.3%	55.5%	36.9%	56.9%	51.5%	71.6%
D-person [4]	79.6%	92.3%	89.4%	98.2%	91.5%	99.0%	64.8%	80.9%
FMN + Rerank [14]	80.6%	87.9%	47.5%	-	46.0%	-	72.8%	79.5%
PAN + Rerank [58]	81.5%	88.6%	41.9%	-	43.9%	-	66.7%	75.9%
MVP Loss	80.5%	91.4%	91.8%	98.8%	93.7%	99.4%	70.0%	83.4%
MVP Loss + Rerank	**90.9%**	**93.3%**	**96.4%**	**99.4%**	**97.7%**	**99.8%**	**83.9%**	**86.3%**

Table 2: Retrieval results on the *Market-1501*, *CUHK03*, and *Duke-MTMC* datasets.

loss objectives, i.e., BH triplet loss, lifted loss, and batch-wise OT loss according to Table 1. The mAP curves on *Market-1501*, *Duke-MTMC* and *MSMT17*, and the CMC top-1 accuracy on *CUHK03-Labeled* of the comparison results are demonstrated in Figure 4. As illustrated, the convergence rate of our method outperforms the other three loss objectives significantly while tending to achieve better and more stable recognition performance. This indicates that the MVP matching can effectively stress hard positive and negative pairs during the training process.

Figure 5: Visualization of an exclusive hard positive and negative pair selected via the 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. We only chose 6 anchor images (i.e., 6×32) for illustration. The full batch-wise correspondence (i.e., 32×32) can be found in our supplementary materials. Please also refers to the electronically edition for better visual effect.

4.4. Comparison with the State-of-the-Art Methods

We finally compared our method with the other state-of-the-art methods on four person re-ID benchmark datasets, i.e., *Market-1501*, *CUHK03-Detected*, *CUHK03-Labeled*, and *Duke-MTMC* datasets. The detailed comparison results are summarized in Table 2. On the *Market-1501* dataset, a baseline method with MVP loss can achieve 80.5% mAP and 91.4% CMC top-1 accuracy. The results outperform the majority of methods in Table 2 expect for D-person, where a more complex model including global and local information are used. After using re-ranking technique [59], MVP loss achieves the state-of-the-art performance. For instance, the mAP and CMC top-1 accuracy of our method reach 90.9% and 93.3%. Similar empirical results can be found on the *CUHK03* dataset. The CMC top-1 accuracy of MVP loss reaches 91.8% and 93.7% for detected and labeled images. Using re-ranking, the top-1 and top-5 accuracy is 96.4% and 99.4% for *CUHK03* detected images, which is 7.0% and 1.2% higher than the D-person. These two indexes for *CUHK03* labeled images are 97.7% and 99.8% compared with 91.5% and 99.0% using D-person. On more challenging *Duke-MTMC* dataset, MVP loss still outperforms other methods significantly. We reported the optimal mAP and CMC top-1 accuracy with 83.9% and 86.3%, respectively.

For a batch of samples, we chose 6 images and their corresponding hard positive and negative pairs selected by the MVP matching. The visualization of results is shown in Figure 5. 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.

5. Conclusion

In this paper, we proposed a novel maximum-value perfect (MVP) matching strategy for mining easily confused hard samples in person re-ID tasks. The learned MVP matching could deal with unbalanced samples according to the adaptive edge weights in the bipartite graph and emphasize the positive and negative sample pairs automatically. Then we developed a batch-wise loss objective based on the MVP matching pairs and incorporated it into an end-to-end deep metric learning network for recognition. We evaluated the performance of our method with application to person re-ID on five benchmark datasets. The empirical results verified that the proposed method could achieve the state-of-the-art recognition performance with a faster convergence rate. Future work will involve facilitating such a trend and applying this MVP matching to more widespread applications, such as scene reconstruction, 3D facial recognition, and point cloud based object segmentation.

References

[1] Ejaz Ahmed, Michael Jones, and Tim K Marks. An improved deep learning architecture for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3908–3916, 2015. 1

[2] Armen S Asratian, Tristan MJ Denley, and Roland Häggkvist. *Bipartite graphs and their applications*, volume 131. Cambridge university press, 1998. 2, 3

[3] Song Bai, Xiang Bai, and Qi Tian. Scalable person re-identification on supervised smoothed manifold. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2530–2539, 2017. 7

[4] Xiang Bai, Mingkun Yang, Tengteng Huang, Zhiyong Dou, Rui Yu, and Yongchao Xu. Deep-person: Learning discriminative deep features for person re-identification. *arXiv preprint arXiv:1711.10658*, 2017. 2, 7

[5] Apurva Bedagkar-Gala and Shishir K Shah. A survey of approaches and trends in person re-identification. *Image and Vision Computing*, 32(4):270–286, 2014. 1

[6] Sean Bell and Kavita Bala. Learning visual similarity for product design with convolutional neural networks. *ACM Transactions on Graphics (TOG)*, 34(4):98, 2015. 1

[7] Dimitri P Bertsekas. The auction algorithm: A distributed relaxation method for the assignment problem. *Annals of operations research*, 14(1):105–123, 1988. 4

[8] John Adrian Bondy, Uppaluri Siva Ramachandra Murty, et al. *Graph theory with applications*, volume 290. Citeseer, 1976. 3

[9] Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Säckinger, and Roopak Shah. Signature verification using a" siamese" time delay neural network. In *Advances in neural information processing systems*, pages 737–744, 1994. 2

[10] Rainer E Burkard, Mauro Dell'Amico, and Silvano Martello. *Assignment problems*. Springer, 2009. 3, 4

[11] 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, 3

[12] 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. 2, 3

[13] 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, 2, 3

[14] Guodong Ding, Salman Khan, Zhenmin Tang, and Fatih Porikli. Let features decide for themselves: Feature mask network for person re-identification. *arXiv preprint arXiv:1711.07155*, 2017. 7

[15] Michael L Fredman and Robert Endre Tarjan. Fibonacci heaps and their uses in improved network optimization algorithms. *Journal of the ACM (JACM)*, 34(3):596–615, 1987. 3

[16] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015. 5

[17] James Hafner, Harpreet S. Sawhney, William Equitz, Myron Flickner, and Wayne Niblack. Efficient color histogram indexing for quadratic form distance functions. *IEEE transactions on pattern analysis and machine intelligence*, 17(7):729–736, 1995. 2

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. 6

[19] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017. 1, 2, 3

[20] Harold W Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955. 2, 3, 4

[21] Bo Li, Yijuan Lu, Afzal Godil, Tobias Schreck, Masaki Aono, Henry Johan, Jose M Saavedra, and Shoki Tashiro. *SHREC'13 track: large scale sketch-based 3D shape retrieval*. 2013. 1

[22] Dangwei Li, Xiaotang Chen, Zhang Zhang, and Kaiqi Huang. Learning deep context-aware features over body and latent parts for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 384–393, 2017. 1, 3, 7

[23] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 152–159, 2014. 2, 3, 5

[24] Wei Li, Xiatian Zhu, and Shaogang Gong. Person re-identification by deep joint learning of multi-loss classification. *arXiv preprint arXiv:1705.04724*, 2017. 1, 7

[25] Wei Li, Xiatian Zhu, and Shaogang Gong. Harmonious attention network for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2285–2294, 2018. 1

[26] Shengcai Liao, Yang Hu, Xiangyu Zhu, and Stan Z Li. Person re-identification by local maximal occurrence representation and metric learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2197–2206, 2015. 1, 2, 7

[27] Jie Lin, Olivier Morere, Vijay Chandrasekhar, Antoine Veillard, and Hanlin Goh. Deephash: Getting regularization, depth and fine-tuning right. *arXiv preprint arXiv:1501.04711*, 2015. 5

[28] Chunxiao Liu, Chen Change Loy, Shaogang Gong, and Guijin Wang. Pop: Person re-identification post-rank optimisation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 441–448, 2013. 6

[29] David G Lowe et al. Object recognition from local scale-invariant features. In *iccv*, volume 99, pages 1150–1157, 1999. 2

[30] Ketan Mulmuley, Umesh V Vazirani, and Vijay V Vazirani. Matching is as easy as matrix inversion. *Combinatorica*, 7(1):105–113, 1987. 3

[31] James Munkres. Algorithms for the assignment and transportation problems. *Journal of the society for industrial and applied mathematics*, 5(1):32–38, 1957. 3

[32] Bryan A Norman, Wipawee Tharmmaphornphilas, Kim Lascola Needy, Bopaya Bidanda, and Rona Colosimo Warner. Worker assignment in cellular manufacturing considering technical and human skills. *International Journal of Production Research*, 40(6):1479–1492, 2002. 3

[33] 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. 1, 2, 3, 6

[34] James B Orlin. A polynomial time primal network simplex algorithm for minimum cost flows. *Mathematical Programming*, 78(2):109–129, 1997. 4

[35] Qi Qian, Rong Jin, Shenghuo Zhu, and Yuanqing Lin. Fine-grained visual categorization via multi-stage metric learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3716–3724, 2015. 1

[36] Xuelin Qian, Yanwei Fu, Yu-Gang Jiang, Tao Xiang, and Xiangyang Xue. Multi-scale deep learning architectures for person re-identification. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5399–5408, 2017. 7

[37] Colin R Reeves. Modern heuristic techniques for combinatorial problems. advanced topics in computer science. *Modern Heuristic Techniques for Combinatorial Problems: Advanced Topics in Computer Science*, 1995. 3

[38] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In *European Conference on Computer Vision*, pages 17–35. Springer, 2016. 2, 5

[39] M Saquib Sarfraz, Arne Schumann, Andreas Eberle, and Rainer Stiefelhagen. A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 420–429, 2018. 3

[40] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 815–823, 2015. 1, 2, 3, 6

[41] Kihyuk Sohn. Improved deep metric learning with multi-class n-pair loss objective. In *Advances in Neural Information Processing Systems*, pages 1857–1865, 2016. 1, 2

[42] Chi Su, Jianing Li, Shiliang Zhang, Junliang Xing, Wen Gao, and Qi Tian. Pose-driven deep convolutional model for person re-identification. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3960–3969, 2017. 7

[43] Yifan Sun, Liang Zheng, Weijian Deng, and Shengjin Wang. Svdnet for pedestrian retrieval. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3800–3808, 2017. 7

[44] Fang Wang, Le Kang, and Yi Li. Sketch-based 3d shape retrieval using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1875–1883, 2015. 1

[45] Jiang Wang, Yang Song, Thomas Leung, Chuck Rosenberg, Jingbin Wang, James Philbin, Bo Chen, and Ying Wu. Learning fine-grained image similarity with deep ranking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1386–1393, 2014. 1

[46] Taiqing Wang, Shaogang Gong, Xiatian Zhu, and Shengjin Wang. Person re-identification by video ranking. In *European Conference on Computer Vision*, pages 688–703. Springer, 2014. 1

[47] Yicheng Wang, Zhenzhong Chen, Feng Wu, and Gang Wang. Person re-identification with cascaded pairwise convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1470–1478, 2018. 3

[48] Mike Wasikowski and Xue-wen Chen. Combating the small sample class imbalance problem using feature selection. *IEEE Transactions on knowledge and data engineering*, 22(10):1388–1400, 2010. 3

[49] Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian. Person transfer gan to bridge domain gap for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 79–88, 2018. 2, 5

[50] Chao-Yuan Wu, R Manmatha, Alexander J Smola, and Philipp Krahenbuhl. Sampling matters in deep embedding learning. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2840–2848, 2017. 3

[51] Qiqi Xiao, Hao Luo, and Chi Zhang. Margin sample mining loss: A deep learning based method for person re-identification. *arXiv preprint arXiv:1710.00478*, 2017. 3, 6

[52] 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. 1, 2, 3, 6

[53] Shiyang Yan, Jun Xu, Yuai Liu, and Lin Xu. Hornet: A hierarchical offshoot recurrent network for improving person re-id via image captioning. In *arXiv preprint arXiv:1908.04915*, 2019. 1

[54] Liming Zhao, Xi Li, Yueting Zhuang, and Jingdong Wang. Deeply-learned part-aligned representations for person re-identification. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3219–3228, 2017. 7

[55] Liang Zheng, Zhi Bie, Yifan Sun, Jingdong Wang, Chi Su, Shengjin Wang, and Qi Tian. Mars: A video benchmark for large-scale person re-identification. In *European Conference on Computer Vision*, pages 868–884. Springer, 2016. 1, 6

[56] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1116–1124, 2015. 2, 5, 6, 7

[57] Zhedong Zheng, Liang Zheng, and Yi Yang. A discriminatively learned cnn embedding for person reidentification.

ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 14(1):13, 2018. 3, 7

[58] Zhedong Zheng, Liang Zheng, and Yi Yang. Pedestrian alignment network for large-scale person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, 2018. 7

[59] Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li. Re-ranking person re-identification with k-reciprocal encoding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1318–1327, 2017. 8

[60] Haibin Zhu, MengChu Zhou, and Rob Alkins. Group role assignment via a kuhn–munkres algorithm-based solution. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 42(3):739–750, 2012. 3, 4