Deep Top-rank Counter Metric for Person Re-identification

Chen Chen*, Hao Dou*, Xiyuan Hu†, Silong Peng*‡
*Institute of Automation, Chinese Academy of Sciences, Beijing, China
†School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China
‡Beijing VISystem Co. Ltd., Beijing, China
*{chen.chen, douhao2015, silong.peng}@ia.ac.cn, †xiyuan.hu@foxmail.com

***Abstract*—In the research field of person re-identification, deep metric learning that guides the efficient and effective embedding learning serves as one of the most fundamental tasks. Recent efforts of the loss function based deep metric learning methods mainly focus on the top rank accuracy optimization by minimizing the distance difference between the correctly matching sample pair and wrongly matched sample pair. However, it is more straightforward to count the occurrences of correct top-rank candidates and maximize the counting results for better top rank accuracy. In this paper, we propose a generalized logistic function based metric with effective practicalness in deep learning, namely the"*deep top-rank counter metric*", to approximately optimize the counted occurrences of the correct top-rank matches. The properties that qualify the proposed metric as a well-suited deep re-identification metric have been discussed and a progressive hard sample mining strategy is also introduced for effective training and performance boosting. The extensive experiments show that the proposed top-rank counter metric outperforms other loss function based deep metrics and achieves the state-of-the-art accuracies.**

***Index Terms*—person re-identification, metric learning, top-rank counter, deep learning**

I. Introduction

Person re-identification (ReID), which aims at matching pedestrians across non-overlapping camera views, has drawn quite much attention in computer vision community in recent years . Due to the variations in viewpoint, pose, illumination, background and occlusion, the appearances of the same person observed in different camera views are often ambiguous to be re-identified while some other persons might be even more similar to the query. Representative feature extraction and discriminative metric learning algorithms have been proposed to tackle this challenge, especially the rising deep learning based approaches have contributed significantly.

Deep metric learning, as a fundamental problem for deep learning based tasks including person ReID, has successfully progressed, and a number of solid milestones appeared naturally following the incremental guiding force towards learning more discriminative metric. Classification loss was first introduced into person ReID to minimize the "intra-person" variance while maximize the "inter-person" variance . Since classification models discriminate limitedly within the "seen" categories while ReID requires to recognize the "unseen" persons in test phase, contrastive loss [1], triplet loss [2], quadruplet loss [3] and their variants were then proposed in accordance with the requirement by focusing on the relative order correction among local samples instead. They attempt to shorten the distance of images with the same identities (positive pairs) while enlarge the distance of images with different identities (negative pairs). For the purpose of learning the reasonable global order efficiently, triplet loss with batch hard mining (namely hard triplet loss) [4] and their variants [5]–[10] make further improvements with hard sample mining or proper weighting strategy.

Most of loss function based deep metric learning methods have been designed to optimize the distance difference between the positive and negative sample pairs with or without a margin setting, as an indirect path for the purpose of pushing the positive sample towards the nearest place to the anchor, i.e., the top-rank candidate in the end. Since top-rank accuracy serves as one out of two main evaluation criteria of person ReID methods together with mAP accuracy, and proper optimization of top-rank accuracy usually keep consistency with the mAP accuracy improvement, more correctly matched top-rank candidates mainly reflects the ultimate goal of ReID task. Therefore, a loss that directly optimize the top-rank accuracy deals straight with the essential of the ReID problem and reflects the trend.

Inspired by the traditional ranking theories and applications in image retrieval and other related tasks [11], we believe that a good choice of deep metric for person ReID, which originates directly and translates faithfully from the top-rank goal, should maximize the appearances of true top-rank matches, i.e. minimize the counting result of non top-rank appearances as penalties. In order to conform with the rising effective tool of deep learning and fit the strong need of practicalness of deep metric, we propose a straightforward and effective deep ReID metric, namely "*deep top-rank counter metric*". The proposed method first addresses the top-rank constraint and then introduces the top-rank counter. An ideal top-rank counter which translates the exact goal for top-rank accuracy is depicted and then a smooth version for better applicability in deep neural network is proposed. The ideal top-rank counter employs the Heaviside step function to count the non top-rank appearances as loss, representing that loss counts 1 when the truly matched sample is ranked other than

This work was supported by the National Science Foundation of China under Grant NSFC 61906194 and the Liaoning Collaboration Innovation Center For CSLE. The corresponding author is Chen Chen.

the first place while no loss otherwise, instead of calculating the sum of the distance differences between the positive sample pair and negative counterpart like the conventional losses do. In the practical point of view, a generalized logistic function is adopted as the approximated top-rank counter and alleviate the non-differentiable dilemma in deep learning in the meanwhile. Different from previous deep metrics, the proposed deep top-rank counter metric can directly optimize the re-identification accuracy, effectively approaching the optimal top-rank goal by pushing the correct candidates towards the query person till ranked as the first place, and push even further far from the negative samples for better generalization due to its decent property. In addition, the proposed deep top-rank counter metric enables a progressive hard sample mining strategy for training the deep network more effectively and further boosting the ReID performance.

The main contributions of this paper are: 1) we propose a generalized logistic function based deep ReID metric to maximizes the top-rank accuracy with the smoothed and relaxed counter mechanism for better applicability in the deep metric learning context; 2) we address the top-rank constraint, discuss the decent properties of the proposed deep top-rank counter metric, and analyze how these properties qualify the proposed metric as an effective deep metric well suited for person ReID task; 3) a progressive hard sample mining strategy introduced by the deep top-rank counter metric property is introduced and implemented for effective training and performance improvement.

II. Related Work

Person ReID algorithms in general consist of the discriminative feature extraction and the effective metric learning as two main steps, with a post-processing step of re-ranking. With the rise of deep learning, the two main steps fuse with and impact on each other much stronger than before. Based on analysis of person images, researchers attempt to extract more specific person features with human-body priors instead of general image features, for instance utilizing useful information from human body structure and alignment between them [12], [13] human pose [14] and context information [15], [16]. Metric learning aims to learn a proper similarity measurement between sample features.

Traditional metric learning approaches for person re-identification usually follow the similar procedure: extracting manually engineered feature embeddings and then learning a mapping with which the training sampl,es have large inter-class (person identity) variations and small intra-class differences. The Pairwise Constrained Component Analysis (PCCA) proposed in [17], a robust Mahalanobis distance metric for Large Margin Nearest Neighbor classification with Rejection (LMNN-R) [18], and the Relative Distance Comparison (RDC) approach [19] are excellent metric learning methods in traditional metric learning field and push the insight exploration for current deep metric learning society.

The widely used and effective ranking optimization methods inspire us to develop deep metric learning methods for top rank optimization. Prosser *et al*. developed an ensemble RankSVM to learn a subspace so that the correctly matched features ensures the high ranking accuracy [20]. The top rank linear function learning [11], which computes only the ranking function score and can hardly be applied on multiple dimensional cases, shows very costly computation complexity on large-scale dataset as in deep learning scenario. Top-push for video based ReID task [21] adopts the top rank idea and learns a latent subspace with a second-order distance metric, which also has difficulty in application in deep learning context.

Deep metric learning methods can be roughly categorized into pairwise loss [1], triplet loss [2], [7], [10] and quadruplet loss [3], [22]. Contrastive loss [1] deals with two samples at a time, and it is designed to pull the positive pairs closer and push the negative pairs further apart, with a threshold filtering the easy negative samples. Triplet loss [2] studies the relation among an anchor, one positive sample and one negative sample, and sets a margin between the distance of positive pairs and that of negative pairs, in order to enlarge the difference between them for the purpose of better generalization capability. Chen *et al*. [10] propose to check the local orders between positive pair and negative pair and penalize the wrong ordering. Quadruplet loss [3] further introduces another negative sample, and utilizes the distance between two negative samples as an absolute bound for distance between positive pairs. Although these losses optimize in the correct direction, there still exists a gap between their training objectives and the ultimate test goal. In addition, they focus limitedly on local relationship correction and ignore the global ranking need of person ReID, while the proposed loss pushes the positive pair towards the first place in the whole ranking list and directly optimizes the appearances of top-rank matches.

Hard sample mining strategies have been well studied and widely used in order to facilitate the loss function proposals better applied to deep metric learning. The batch-formatted hard sample mining [4], triplet loss with soft margin [4], margin sample mining [22] and deep mining with incremental triplet margin [8] have been designed for the purpose of more reasonable relative order between samples. The works in [5], [6] generate the point-to-point triplet loss into a point-to-set metric as a generalized sampling scheme for strengthened robustness. Inspired by these sample mining schemes, we introduce a progressively hard sample mining strategy based on the proposed loss function, which can make effective improvement in training deep network and further increase the re-identification accuracy.

III. Deep top-rank Counter Metric for Person Re-identification

Let $\mathcal{X}$ and $\mathcal{Y}$ be the set of person images and corresponding label set of person identities, respectively. Given a person image sample $a \in \mathcal{X}$ as an anchor, there exists one of its positive sample $p \in \mathcal{P}_a = \{p \in \mathcal{X} | y(p) = y(a), y \in \mathcal{Y}\}$ that shares the same person identity with the anchor a, and one negative sample $n \in \mathcal{N}_a = \{n \in \mathcal{X} | y(n) \neq y(a), y \in \mathcal{Y}\}$ that has a different person identify from the anchor a. The goal of

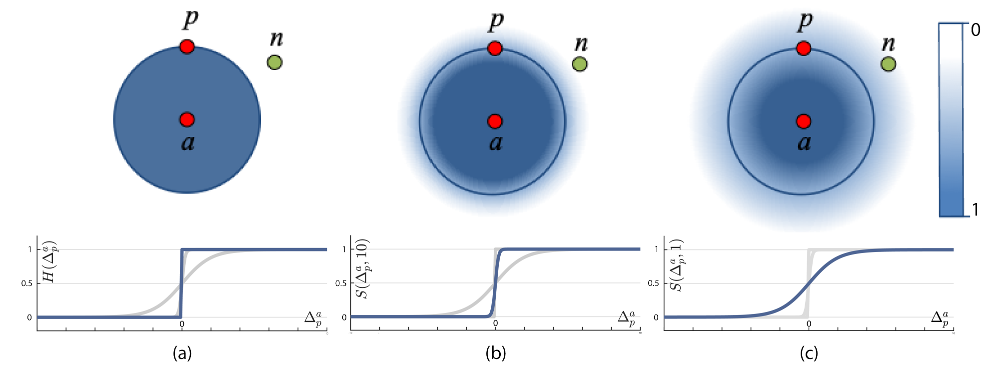

Fig. 1. Illustration of (a) the ideal top-rank counter $H(\Delta_p^a)$ and the smooth top-rank counter representation (b) $S(\Delta_p^a; k = 10)$ and (c) $S(\Delta_p^a; k = 1)$ as approximations. Top row shows the sample location and the loss value distribution, where a denotes anchor sample, p denotes the positive sample, and n denotes the negative sample.

metric learning is to explicitly or implicitly learn a mapping $f(x;\theta) : \mathcal{X} \mapsto \tilde{\mathcal{X}}$, where θ denotes mapping parameters, so that the features in $\tilde{\mathcal{X}}$ discriminate better from each other than those in $\mathcal{X}$. Let $d(\cdot,\cdot) : \tilde{\mathcal{X}} \times \tilde{\mathcal{X}} \mapsto \mathbb{R}$ be the distance metric between two features in $\tilde{\mathcal{X}}$, and the distance between anchor and positive sample $d\big(f(a;\theta), f(p;\theta)\big)$ is written as d_{ap} for convenience, similarly for d_{an}.

A. Top-rank constraint

The ultimate goal of a person ReID model is to match each of the positive sample p as top-rank candidate, i.e. the first place in ranking list. Given a person image as anchor $a \in \mathcal{X}$ and its positive image $p \in \mathcal{P}_a$, the negative image $n \in \mathcal{N}_a$, j represents random sample image in ranking list other than the anchor. When the top-rank constraint satisfies, the following equation holds:

$$d_{ap} = \min_{j \in \mathcal{X}, j \neq a} d_{aj}. \tag{1}$$

where d_{ap} represents the smallest distance among all possible anchor-related distances.

A lot of effort aiming at top-tank accuracy has been made to learn an effective metric, and most designed loss functions resemble the variations of the above equation. A mild constraint that $d_{ap} - d_{an} < 0$, which focuses only on the correction of local ranking order. The widely used hard triplet loss [4] minimizes $\max_{p \in \mathcal{P}_a} d_{ap} + m - \min_{n \in \mathcal{N}_a} d_{an}$, which introduces hard sample mining and margin to improve the optimization, however, the strong identity-level constraint might cause model collapse. Quadruplet loss [3] further replaces the smallest anchor-negative distance $\min_{n \in \mathcal{N}_a} d_{an}$ with the hardest negative pair distance $\min_{n_1, n_2 \in \mathcal{N}_a} d_{n_1 n_2}$ with the disadvantage of difficult convergence.

We introduce the direct translation of the exact top-rank constraint as

$$d_{ap} < \min_{n \in \mathcal{N}_a} d_{an}, \tag{2}$$

to set the image level top-rank optimization. The proposed top-rank constraint pushes each of the positive sample towards the top-rank position instead of manipulating only the hardest samples, to properly avoid model collapse due to the possible outliers.

B. From top-rank constraint to top-rank counter

When top-rank constraint satisfies, $d_{ap} < \min_{n \in \mathcal{N}_a} d_{an}$ and there exists no distance difference; otherwise the distance difference $d_{ap} - \min_{n \in \mathcal{N}_a} d_{an} > 0$ is obtained. There are two main approaches to measure the corresponding loss. First, directly sum up the distance differences and optimize the averaged loss terms, which has been taken by most metric such as triplet loss. However, the difference-based loss measurement may harm the optimization by the enormously large difference introduced by the extreme hard mining.

Instead, we introduce the case counting system to record the occurrences of unsatisfactory cases, i.e. loss plus one when unsatisfactory case appear once. A natural and ideal counter mechanism in mathematical language leads to the Heaviside step function $H(\cdot)$. Combining with the proposed top-rank constraint, we have the ideal top-rank counter

$$H(a, p, \mathcal{N}_a) = \begin{cases} 1, & d_{ap} - \min_{n \in \mathcal{N}_a} d_{an} \geqslant 0, \\ 0, & otherwise, \end{cases} \tag{3}$$

where $H(a, p, \mathcal{N}_a)$ counts loss of 1 when p is a non top-rank candidate for anchor a in ranking list, and counts no loss when p is the top-rank match.

The Heaviside step function perfectly realizes the non-top-Rank counting functionality, however, it is pitifully not applicable on deep neural network considering the lack of differentiability. For better practicalness in deep learning implementation, an smooth top-rank counter that keeps the counter functionality while improves the differentiability is introduced. We propose to use the generalized logistic function with a parameter k,

$$S(a, p, \mathcal{N}_a; k) = \frac{1}{1 + e^{-k(d_{ap} - \min_{n \in \mathcal{N}_a} d_{an})}}, \tag{4}$$

as the smooth top-rank counter, where k is a designed parameter to control the approximation towards Heaviside step function and compensate for differentiability.

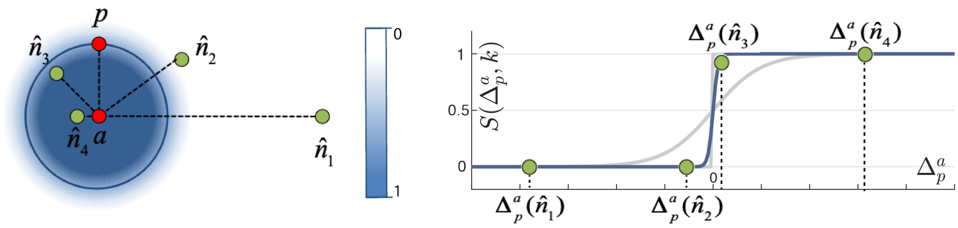

Fig. 2. Illustration of the generalized logistic function based deep top-rank counter metric.

The illustration of both the ideal top-rank counter and its smooth counterpart is shown in Fig.1, where Δ_p^a is used to represent $d_{ap} - \min_{n \in \mathcal{N}_a} d_{an}$ for convenience. The ideal top-rank counter shown in Fig.1(a) directly steps from 0 to 1 at origin in function plot at the bottom row, and leaves two ideally distinguished areas (blue denotes function value 1 and white denotes 0) in illustration of function value distribution with respect to sample relation at the top row; while the logistic function based smooth top-rank counter (shown as (b) and (c) in figure) relaxes the sharp step from 0 to 1 around the origin in function plot and gradually varies the loss value around positive-negative sample boundary. As observed in Fig.1, the ideal and the smooth top-rank counter curves are visually similar, especially when the parameter k gets larger. However, if k is set too large, the stepping region gets too sharp and the gradient quickly gets very small towards two outreaching directions, thus the metric learning system consequently gets more difficult to convergence efficiently. Therefore a proper parameter k could help balance between the consistency with the top-rank constraint and the feasibility of practical implementation.

C. Deep top-rank counter metric

Most deep person ReID metrics regard the distance difference between certain positive sample pair and negative sample pair as a loss term and optimize the sum or average of the distance differences to improve top rank accuracy, however, it is more straightforward and more effective to count the occurrences of the non top-rank matches, regard the counting result as ReID loss and minimize it. Therefore, we further formalize the smooth top-rank counter to develop the deep top-rank counter metric for better deep metric learning for person ReID.

With the generalized logistic function adopted as the smooth top-rank counter to facilitate the practicalness in deep learning, the resulting loss over the whole training set

$$\mathcal{L}(\theta, \mathcal{X}) = \sum_{a \in \mathcal{X}} \sum_{p \in \mathcal{P}_a} S(a, p, \mathcal{N}_a; k), \tag{5}$$

$$= \sum_{a \in \mathcal{X}} \sum_{p \in \mathcal{P}_a} \frac{1}{1 + e^{-k(d_{ap} - \min_{n \in \mathcal{N}_a} d_{an})}}, \tag{6}$$

where θ denotes the deep network parameter set, is defined as the "*deep top-rank counter metric*".

D. Discussion on deep top-rank counter metric

The proposed deep top-rank counter metric not only performs as an adequate loss function that further develops the difference-based loss and approximates of the ideal top-rank counter mechanism, but also qualifies itself for an effective ReID metric due to the following properties and functionalities shown as follows and illustrated in Fig.2:

(a) Straightforward top-rank goal: all the non top-rank candidates (when $\Delta_p^a \geqslant 0$) tend to gain a counter loss approaching 1; while all the top-rank candidates ($\Delta_p^a < 0$) tend to have almost no loss. The proposed deep top-counter metric directly aims at the top-rank accuracy optimization, which keeps high consistency with the ultimate goal of person ReID task;

(b) Implicit Margin Effect: when $\Delta_p^a < 0$, the function curve of top-rank counter $S(\Delta_p^a; k)$ relaxes the ideal counter and approaches zero loss gradually and smoothly. The minimization of the non-zero loss pushes the positive sample that is already top-rank candidate even farther away from the negative samples, which implicitly realizes the similar functionality as the margin setting in hard triplet loss and other margin based loss functions. When the sample distribution varies, the relatively small loss for $\Delta_p^a < 0$ changes its value accordingly and generates the pushing strength with different level. For instance, with the optimization proceeds further from $S(\Delta_p^a(\hat{n}_2); k)$ to $S(\Delta_p^a(\hat{n}_1); k)$, we have $\Delta_p^a(\hat{n}_1)$ pushed further from origin than $\Delta_p^a(\hat{n}_2)$, and the relative relation of a, p, n is better obtained according to the top-rank constraint, which shows more flexible margin effect.

(c) Consistency with mAP accuracy: Rank-1 and mAP accuracies are the evaluation criteria in person re-identification. Proper optimization of Rank-1 accuracy can be beneficial to mAP improvement. The proposed metric adopts the image-level top-rank representation , and the counter loss for each positive sample $p \in \mathcal{P}_a$ given anchor $a \in \mathcal{X}$ is optimized. Therefore, the learned metric will push all positive samples closer towards the query sample independently, which is consistent with the mAP goal.

(d) Robustness and generalization: when Δ_p^a increases towards positive infinity as shown as $\Delta_p^a(\hat{n}_4)$, the corresponding loss infinitely approaches 1 instead of an enormous penalty, which helps to avoid the harmful effect caused by outliers and ensures the robustness; when Δ_p^a decreases toward negative infinity shown as $\Delta_p^a(\hat{n}_1)$, the resulting loss decreases towards a small positive value approaching 0. The relatively small loss pushes the positive pair further discriminative from negative pairs and strengthens the generalization in test phase.

The proposed top-rank counter metric can also serve as a basic loss function combining with other metric learning methods or state-of-the-art methods, which may joint utilize

TABLE I
EFFECT OF PARAMETER k WITH VARIOUS SETTINGS ON MARKET-1501 DATASET.

k	1	10	50	100
Rank-1	88.69	**91.48**	87.95	86.49
mAP	76.12	**78.43**	74.91	72.66

TABLE II
EFFECT OF THE PROGRESSIVE HARD SAMPLE MINING STRATEGY ON MARKET-1501 DATASET.

Training phase	Rank-1	Rank-5	Rank-10	mAP
Vanilla training	91.48	96.88	98.16	78.43
Full training	**92.34**	**97.27**	**98.22**	**79.37**

the advantages of the proposed metric and other effective approaches to improve the person ReID accuracy together.

IV. Progressive Hard Sample Mining

Based on the properties of the proposed deep top-rank counter metric, we introduce a progressive hard sample mining strategy to improve the model training and its performance.

The person image set $\mathcal{X}$ consists of two sample sets $\mathcal{X}^+ \triangleq \{a \in \mathcal{X} | \Delta_p^a \geqslant 0\}$ and $\mathcal{X}^- \triangleq \{a \in \mathcal{X} | \Delta_p^a < 0\}$, and the optimization of the deep top-rank counter metric on the sample set $\mathcal{X}^+$ and the whole sample set $\mathcal{X}$ lead to two different training schemes with increasing difficulty level:

- **Vanilla training**: optimize the deep top-rank counter metric on $\mathcal{X}^+$ by penalizing samples with $\Delta_p^a \geqslant 0$ that contribute the non top-rank losses (with $S(\Delta_p^a; k) \in [0.5, 1)$, as shown in Fig.2), in order to push the true matched positive samples towards the top-rank position;
- **Full training**: optimize the deep top-rank counter metric on the whole set $\mathcal{X}$ by penalizing all samples including samples in $\mathcal{X}^+$ with non top-rank losses ($S(\Delta_p^a; k) \in [0.5, 1)$) and also hard samples in $\mathcal{X}^-$ with relatively small losses ($S(\Delta_p^a; k) \in (0, 0.5)$) to push the top-rank matches even farther apart from the negative samples.

In order to alleviate the training difficulty step by step, we propose to use the "*progressive hard sample mining*" strategy. In **Phase 1**, we optimize the deep top-rank counter metric $\mathcal{L}(\theta, \mathcal{X})$ on the sample set $\mathcal{X}^+$ for vanilla training; in **Phase 2**, based on the effective training in Phase 1, we continue to optimize $\mathcal{L}(\theta, \mathcal{X})$ on the whole set $\mathcal{X}$ for full training. The progressive hard sample mining strategy performs successive training phases to first optimize for more top-rank appearances and then to further enlarge the distance between positive-negative pairs for better generalization ability in test phase.

V. Experiments and Results

We conduct extensive experiments from three aspects: 1) performance analysis of the proposed method with different parameter and training strategy settings; 2) comparison between the proposed method and other loss function based methods; 3) comparison with state-of-the-art methods.

A. Dataset and evaluation protocols

We evaluate the proposed method on three large-scale person ReID datasets, Market-1501 [23], DukeMTMC-reID [24] and CUHK03 [25].

- **Market-1501** contains a total of 32,688 images of 1,501 labeled pedestrians captured under 6 camera viewpoints. The pedestrian bounding boxes are detected by Deformable Part Model . The dataset is split into two non-overlapping partitions: 12,936 images (751 identities) for training and 19,732 images (750 identities) for test. In test phase, 3,368 images are chosen as query images. We adopt single-query evaluation mode in our experiments.
- **DukeMTMC-reID** is a subset of Duke-MTMC for ReID task. A total of 36,411 images of 1,812 pedestrians from 8 high-resolution cameras were captured. The dataset is split into the training set of 16,522 images from 702 identities and the test set of 2,228 queries from 702 identities together with 17,661 gallery images. We also employ the single-query setting.
- **CUHK03** contains 14,096 pedestrian images of 1,467 identities, captured by 5 different camera pairs on campus. The dataset provides both DPM-detected and hand-marked pedestrian bounding boxes and we report our results on both sets. Different from the original evaluation protocol, we utilize the more challenging train/test split protocol, i.e. 767 identities for training and the rest 700 for testing, which was proposed in [26].

Following the previous ReID works, we evaluate the cumulative matching characteristics (CMC) at Rank-1 (at least), Rank-5, Rank-10, and mean average precision (mAP) in experiments.

B. Implementation details

For fair comparison, the backbone model, the network design, the data augmentation and the optimization settings adopted in implementation of the proposed top-rank counter metric and other loss functions based metrics are all the same. Only the learning rate schemes utilize the best settings corresponding to each loss design.

Network settings: The backbone model is ResNet-50 pretrained on ImageNet. The network design follows the commonly used ReID baseline with stride $= 2$ in last conv block.

Data augmentation: The training images are resized to the size of 288×144. Standard random crop, horizontal flipping and random erasing (with ratio of 0.5) are applied. In test phase, only horizontal flipping is used. The distance function $d(\cdot, \cdot)$ used for the ranking and mAP computation adopts the Euclidean metric.

Optimization: Our implementation is based on PyTorch platform with NVIDIA Titan X GPU. 32 persons with 4 images per person are chosen to form each of the mini-batch with the fixed size of 128. We use the Adam optimizer with $\epsilon = 10^{-8}, \beta_1 = 0.9$ and $\beta_2 = 0.99$.

Learning rate: The learning rate for the proposed method is fixed to 1.5×10^{-4} first, then exponentially decays to 1.5×10^{-5} and 1.5×10^{-6}, each step lasts 500 epochs, then the learning

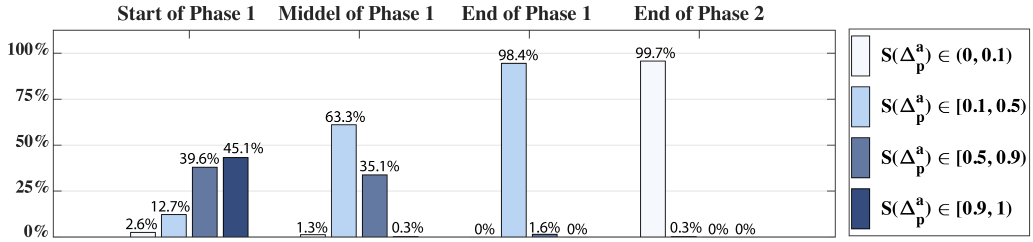

Fig. 3. Illustration of the loss value distribution changing during progressive hard sample mining.

TABLE III
COMPARISON WITH OTHER LOSS FUNCTION BASED METHODS WITH THE SAME BASELINE MODEL.

Loss	Market-1501		DukeMTMC-reID		CUHK03			
					Labeled		Detected	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
Softmax	89.31	73.51	75.54	60.09	58.93	51.95	51.86	46.01
Triplet [2]	74.05	53.70	68.76	51.41	56.43	49.92	52.07	45.80
Quadruplet [3]	87.95	73.57	76.26	61.80	67.36	60.69	63.50	56.95
Hard triplet [4]	90.05	77.56	80.02	66.62	71.07	**65.11**	**68.00**	**62.21**
Deep Top-rank Counter Metric (TRC)	**92.33**	**79.37**	**81.42**	**67.72**	**71.64**	64.38	67.93	60.27

rate is fixed to 1.5×10^{-6} till convergence in both of Phase 1 and Phase 2.

C. Deep top-rank counter metric analysis

Parameter analysis: The parameter k in the proposed deep top-rank counter metric controls the trade-off between the approximation to the ideal counter unit and the usability in deep learning process. We set k with different values and compare the resulting models on Market-1501 dataset using vanilla training as shown in Table.I. As we can see, the proposed deep top-rank counter metric performs well across a large range of k, and produces the best result when $k = 10$. The results keep consistancy with the discussion on k that too samll k deviates the top-rank constraint while too large k leads to sharpe stepping region and difficulty in convergence.

Training strategy evaluation: A progressive hard sample mining strategy is proposed for wisely training and performance boosting. We plot the loss value distributions at four different stages during Phase 1 (vanilla training) and Phase 2 (full training) on Market-1501 dataset in Fig.3 to illustrate the changing trend. It can be observed that 1) during the vanilla training phase, most non top-rank samples have been pushed to top-rank with their loss values changing from $[0.5, 1)$ to $[0.1, 0.5)$; 2) during the full training phase, in addition to optimizing the non top-rank samples, the truly matched top-rank samples with loss in $[0.1, 0.5)$ have been pushed farther away from the negative samples, with their loss values decreasing towards 0 into $(0, 0.1)$. The ReID accuracies at the end of two training phases in Table.II show that the progressive hard sample mining strategy indeed improves the model performance at all ranks, especially from 91.48 to 92.34 at Rank-1 accuracy, by effectively further training on hard samples.

We fix the parameter $k = 10$ and the progressive hard sample mining in the following experiments.

D. Comparison with other loss function based methods

We compare the proposed deep top-rank counter metric with the milestones deep ReID metrics including softmax loss, triplet loss [2], hard triplet loss [4] and quadruplet loss [3] as shown in Table.III. For fair comparison, we use the same pre-trained ResNet-50 backbone model and the same data augmentation, conduct experiments in the same mini-batch configuration and same evaluate performance on three datasets.

Based on the well designed baseline network, softmax loss, as classical identification losses, achieves promising ReID accuracies, while hard triplet loss, among other metric learning methods, produces better performances. The proposed deep top-rank counter metric outperforms all other losses with gains in Rank-1 accuracy up to +2.28%, except for CUHK03 (detected) dataset with only -0.07%gap as the second best. The proposed metric also yields the best Rank-1 accuracy on both Market-1501 and DukemTMC-reID datasets.

E. Comparison with state-of-the-art methods

We compare the proposed metric with state-of-the-art methods on the three datasets in Table.IV. Furthermore, in order to show the effectiveness of deep top-rank counter metric as a basic loss function, we also combine the proposed metric with MGN method [37] by replacing each of the three hard triplet losses in MGN with the proposed deep top-rank counter loss, and the corresponding results are denoted as “deep top-rank counter metric (MGN)” also shown in Table.IV.

Comparison on Market-1501: the proposed deep top-rank counter metric (TRC for short) achieves the performance with 92.33% in Rank-1 accuracy and 79.37% in mAP accuracy.

TABLE IV
COMPARISON WITH STATE-OF-THE-ART METHODS. BEST RESULTS ARE MARKED IN BOLD

Method	Market-1501		DukeMTMC-reID		CUHK03			
					Labeled		Detected	
	Rank1	mAP	Rank1	mAP	Rank1	mAP	Rank1	mAP
MGCAM (CVPR18) [27]	83.80	74.30	-	-	50.10	50.20	46.70	46.90
HAP2S (ECCV18) [5]	84.20	69.76	76.08	59.58	-	-	-	-
PSE (CVPR18) [28]	87.70	69.00	79.80	62.00	-	-	30.20	27.30
Pose-Transfer (CVPR18) [14]	87.70	68.90	78.50	56.90	45.10	42.00	41.60	38.70
MLFN (CVPR18) [29]	90.00	74.30	81.00	62.80	54.70	49.20	52.80	47.80
HA-CNN (CVPR18) [30]	91.20	75.70	80.50	63.80	44.40	41.00	41.70	38.60
DuATM (CVPR18) [31]	91.40	76.60	81.80	64.60	-	-	-	-
PCB (ECCV18) [12]	92.30	77.40	81.90	65.30	-	-	61.30	54.20
IA-Net (CVPR-2019) [32]	94.40	83.10	87.10	73.40	-	-	-	-
DG-Net (CVPR-2019) [33]	94.80	86.00	86.60	74.80	-	-	-	-
CAMA (CVPR-2019) [34]	94.70	84.50	85.80	72.90	70.10	66.50	66.60	64.20
SFT (ICCV-2019) [35]	93.40	82.70	86.90	73.20	68.20	62.40	-	-
MHN-6(PCB) (ICCV-2019) [36]	95.10	85.00	89.10	77.20	77.20	72.40	71.70	65.40
MGN (ACM-MM18) [37]	**95.70**	86.90	88.70	78.40	68.00	67.40	66.80	66.00
Deep top-rank Counter Metric (TRC)	92.33	79.37	81.42	67.72	71.64	64.38	67.93	60.27
Deep top-rank Counter Metric (TRC-MGN)	94.63	**87.03**	**89.36**	**78.69**	**77.43**	**74.68**	**74.14**	**71.65**

Compared with the recently proposed metric learning method HAP2S by [5] that develops the point-to-set deep metric with hard sample aware strategy, our method brings +8.13% and +9.41% improvements in Rank-1 and mAP accuracy respectively. The proposed deep top-rank counter metric with MGN (TRC-MGN for short) produces 87.03% in mAP accuracy and outperforms all the other methods with a great improvement (+8.88% gain on average) and achieves the second best Rank-1 accuracy of 94.63% on Market-1501. TRC-MGN shows lower Rank-1 accuracy compared with MGN since our reproduction of MGN remains 94.03% in Rank-1 accuracy, which indicates that TRC-MGN already achieves 0.60% improvement.

Comparison on DukeMTMC-reID: the proposed TRC metric yields promising Rank-1 accuracy of 81.42% and mAP accuracy of 67.72% as a basic metric. Compared with the deep metric HAP2S, TRC metric leads +5.34% gain in Rank-1 and +8.14% in mAP accuracy. The proposed TRC-MGN method achieves the best Rank-1 accuracy and best mAP accuracy among all state-of-the-art methods with large improvement (+5.99% on average in Rank-1 and +10.62% in mAP accuracy). The combined TRC-MGN outperforms the original MGN method with +0.66% gain in Rank-1 and +0.29% gain in mAP accuracy, which shows effective improvement introduced by the proposed deep top-rank counter metric.

Comparison on CUHK03: the challenging training/testing protocol has been implemented, and the proposed TRC metric achieves the second best among state-of-the-art methods, including those ones with sophisticated model designs. TRC-MGN method yields the best performance both in Rank-1 and mAP accuracies under both the labeled setting and detected setting, with +17.70% gain in Rank-1 accuracy and +18.29% gain in mAP on average under the labeled setting, while +20.87% improvement in Rank-1 accuracy and +21.75% improvement in mAP on average under the detected setting. Compared with original MGN showing +9.43% gain in Rank-1 and +7.28% gain in mAP under labeled setting while +7.34% gain in Rank-1 and +5.65% gain in mAP accuracy under detected setting. The large amount of gains indicate the outstanding improvement produced by replacing the hard triplet loss with the proposed deep top-rank counter loss.

Based on the comparison results on three datasets, we can conclude that the deep top-rank counter metric not only improves the top rank accuracy as the model target suggests, but also make the mAP accuracy improved as the property of the deep top-rank counter metric explains. The effectiveness of the proposed metric as its own and also as a basic metric which can be combined with other methods is also properly illustrated and experimentally proved.

VI. CONCLUSION

In this paper, we present a deep top-rank counter metric for person ReID in order to better optimize the top-rank accuracy in the context of deep learning. We discuss the top-rank constraint and employ the generalized logistic function to formalize the proposed deep top-rank counter metric, which relaxes the ideal top-rank counter as well as conquers the non-differentiability dilemma in deep learning. We also analyze the decent properties of the proposed metric with better usability and practicalness for both deep learning and person ReID. A progressive hard sample mining strategy deduced from the metric property has been introduced. The extensive experiments show the superior performance and effectiveness of the proposed deep top-rank counter metric as its own and further as a basic metric that can be combined with other method.

References

[1] Rahul Rama Varior, Mrinal Haloi, and Gang Wang. Gated siamese convolutional neural network architecture for human re-identification. In *European Conference on Computer Vision*, pages 791–808. Springer, 2016.

[2] 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.

[3] 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.

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

[5] Rui Yu, Zhiyong Dou, Song Bai, Zhaoxiang Zhang, Yongchao Xu, and Xiang Bai. Hard-aware point-to-set deep metric for person re-identification. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 188–204, 2018.

[6] Sanping Zhou, Jinjun Wang, Jiayun Wang, Yihong Gong, and Nanning Zheng. Point to set similarity based deep feature learning for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3741–3750, 2017.

[7] De Cheng, Yihong Gong, Sanping Zhou, Jinjun Wang, and Nanning Zheng. Person re-identification by multi-channel parts-based cnn with improved triplet loss function. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1335–1344, 2016.

[8] Yingying Zhang, Qiaoyong Zhong, Liang Ma, Di Xie, and Shiliang Pu. Learning incremental triplet margin for person re-identification. *arXiv preprint arXiv:1812.06576*, 2018.

[9] Xing Fan, Wei Jiang, Hao Luo, and Mengjuan Fei. Spherereid: Deep hypersphere manifold embedding for person re-identification. *Journal of Visual Communication and Image Representation*, 2019.

[10] Shi-Zhe Chen, Chun-Chao Guo, and Jian-Huang Lai. Deep ranking for person re-identification via joint representation learning. *IEEE Transactions on Image Processing*, 25(5):2353–2367, 2016.

[11] Nan Li, Rong Jin, and Zhi-Hua Zhou. Top rank optimization in linear time. In *Advances in neural information processing systems*, pages 1502–1510, 2014.

[12] Yifan Sun, Liang Zheng, Yi Yang, Qi Tian, and Shengjin Wang. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 480–496, 2018.

[13] Min Cao, Chen Chen, Xiyuan Hu, and Silong Peng. Region-specific metric learning for person re-identification. In *International Conference on Pattern Recognition*, pages 794–799, 2018.

[14] Jinxian Liu, Bingbing Ni, Yichao Yan, Peng Zhou, Shuo Cheng, and Jianguo Hu. Pose transferrable person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4099–4108, 2018.

[15] Min Cao, Chen Chen, Xiyuan Hu, and Silong Peng. From groups to co-traveler sets: Pair matching based person re-identification framework. In *IEEE International Conference on Computer Vision Workshop*, pages 2573–2582, 2017.

[16] Chen Chen, Min Cao, Xiyuan Hu, and Silong Peng. Key person aided re-identification in partially ordered pedestrian set. In *Proceedings of the British Machine Vision Conference*, 2017.

[17] Alexis Mignon and Frédéric Jurie. Pcca: A new approach for distance learning from sparse pairwise constraints. In *2012 IEEE conference on computer vision and pattern recognition*, pages 2666–2672. IEEE, 2012.

[18] Mert Dikmen, Emre Akbas, Thomas S Huang, and Narendra Ahuja. Pedestrian recognition with a learned metric. In *Asian conference on Computer vision*, pages 501–512. Springer, 2010.

[19] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Reidentification by relative distance comparison. *IEEE transactions on pattern analysis and machine intelligence*, 35(3):653–668, 2012.

[20] Bryan James Prosser, Wei-Shi Zheng, Shaogang Gong, Tao Xiang, and Q Mary. Person re-identification by support vector ranking. In *BMVC*, volume 2, page 6, 2010.

[21] Jinjie You, Ancong Wu, Xiang Li, and Wei-Shi Zheng. Top-push video-based person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1345–1353, 2016.

[22] 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.

[23] 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.

[24] Zhedong Zheng, Liang Zheng, and Yi Yang. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3754–3762, 2017.

[25] 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.

[26] 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.

[27] Chunfeng Song, Yan Huang, Wanli Ouyang, and Liang Wang. Mask-guided contrastive attention model for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1179–1188, 2018.

[28] 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.

[29] Xiaobin Chang, Timothy M Hospedales, and Tao Xiang. Multi-level factorisation net for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2109–2118, 2018.

[30] 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.

[31] Jianlou Si, Honggang Zhang, Chun-Guang Li, Jason Kuen, Xiangfei Kong, Alex C Kot, and Gang Wang. Dual attention matching network for context-aware feature sequence based person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5363–5372, 2018.

[32] Ruibing Hou, Bingpeng Ma, Hong Chang, Xinqian Gu, Shiguang Shan, and Xilin Chen. Interaction-and-aggregation network for person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[33] Zhedong Zheng, Xiaodong Yang, Zhiding Yu, Liang Zheng, Yi Yang, and Jan Kautz. Joint discriminative and generative learning for person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[34] Wenjie Yang, Houjing Huang, Zhang Zhang, Xiaotang Chen, Kaiqi Huang, and Shu Zhang. Towards rich feature discovery with class activation maps augmentation for person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[35] Chuanchen Luo, Yuntao Chen, Naiyan Wang, and Zhaoxiang Zhang. Spectral feature transformation for person re-identification. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.

[36] Binghui Chen, Weihong Deng, and Jiani Hu. Mixed high-order attention network for person re-identification. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.

[37] Guanshuo Wang, Yufeng Yuan, Xiong Chen, Jiwei Li, and Xi Zhou. Learning discriminative features with multiple granularities for person re-identification. In *2018 ACM Multimedia Conference on Multimedia Conference*, pages 274–282. ACM, 2018.