SCIENCE CHINA
Information Sciences

February 2021, Vol. 64 120111:1–120111:2
https://doi.org/10.1007/s11432-019-2943-6

• LETTER •
Special Focus on Deep Learning for Computer Vision

Learning generalizable deep feature using triplet-batch-center loss for person re-identification

Bin HU[†], Jiwei XU[†] & Xinggang WANG[*]

School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China

Received 12 September 2019/Revised 27 March 2020/Accepted 9 May 2020/Published online 27 October 2020

Citation Hu B, Xu J W, Wang X G. Learning generalizable deep feature using triplet-batch-center loss for person re-identification. Sci China Inf Sci, 2021, 64(2): 120111, https://doi.org/10.1007/s11432-019-2943-6

Dear editor,

Metric learning loss functions are important for deep learning-based person re-identification (re-ID). Several loss functions, such as triplet loss, center loss, and sphere loss, have been proposed. However, there are a few discussions on the generalization ability of these loss functions. In this study, we proposed a novel metric-learning loss function, triplet-batch-center loss (TBCL), to learn more generalizable deep features than the previous loss functions. TBCL averages features from the same class within a batch to get a center and requires each sample to be close to its corresponding center and far from the other centers. We carried out extensive experiments on the following datasets: CUHK03, DukeMTMC-re-ID, and Market1501. TBCL leads to state-of-the-art results on these datasets without the image re-ranking post-processing. For instance, our methods achieved 75.9% Rank-1 accuracy and 73.3% mean average precision (mAP) on CUHK03 for the setting of having labelled person using only global features, remarkable better than the previous state-of-the-art person re-ID methods by at least +6.9%/+7.6%.

Motivation. We believe that a suitable metric-learning loss function should be able to minimize intraclass distances and maximize interclass distances. The common shortcoming of the softmax and center loss is that they only increase variation between different identities without considering the variation within the same identity; therefore, their generalization capability is weak. Thus, optimizing the global center point of a class is not the best choice. To train a feature extractor for better generalization ability, we use meta-learning. In meta-learning, specifically few-shot learning, training data are given by a series of episodes, and optimization is carried out within an episode without accessing other training data [1]. As a result, the episode training strategy helps in learning about a generalizable feature extractor. Similar to the episode training, we have a small batch of training images when training a re-ID network. We consider optimizing every training sample to its class center within a batch, every time, so that the generalization performance of the network is improved. Note that our proposed batch-center-based metric learning is totally different from the triplet-center loss in [2] because Ref. [2] optimizes global centers and their distances from training samples, while we focus on learning batch centers.

Moreover, previous triplet-loss-based methods are mainly based on the $L2$ distance metric during the training phase. However, in the test phase cosine similarity can obtain better results than $L2$ distance. Thus, training using $L2$ distance metric seems not to be the best approach. Based on the abovementioned considerations, the proposed TBCL uses cosine similarity rather than Euclidean distance.

Definitions. We fetch $P \times K$ images to build a batch during training. These images are from P different person IDs, and each ID has K samples. We assume that x_{ji} $(1 \leqslant j \leqslant K, 1 \leqslant i \leqslant P)$ represents an extracted feature from a person j and a sample i. The goal of the TBCL is to efficiently and synchronously minimize the intraclass variations and maximize the interclass distances of the deeply-learned features. In a batch, we first average the K samples of the same person so that we can get P center points. Let $(e_{i1}, \ldots, e_{iK})$ represent K samples, and a batch center is defined as follows:

$$c_i = \mathbb{E}_m[e_{im}] = \frac{1}{K}\sum_{m=1}^{K} e_{im}. \tag{1}$$

Then we can obtain $C = \{c_1, c_2, \ldots, c_p\}$. In this study, for simplicity, we use f_i for $f(x_i)$. The final triplet batch center loss is defined as follows:

$$L_{\mathrm{tbc}} = \sum_{i=1}^{P} \max\left(\max_{i \neq j} S(f_i, c_j) + m - S(f_i, c_i), 0\right), \tag{2}$$

where $S(\cdot)$ represents the cosine similarity function defined as follows:

$$S(f_i, c_i) = \frac{f_i c_i}{\|f_i\|\|c_i\|}. \tag{3}$$

[*] Corresponding author (email: xgwang@hust.edu.cn)
[†] Hu B and Xu J W have the same contribution to this work.

Table 1 Performance (%) comparisons by state-of-the-art methods[a)]

Method	Market1501		CUHK03 (labeled)		CUHK03 (detected)		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
MGCAM [3]	84.8	74.3	50.1	50.2	46.7	46.9	–	–
AACN [4]	85.9	66.9	–	–	–	–	76.8	59.3
Pose transfer [5]	87.7	68.9	33.8	30.5	30.1	28.2	68.6	48.4
PSE [6]	87.7	69.0	–	–	30.2	27.3	79.8	62.0
HA-CNN [7]	91.2	75.7	44.4	41.0	41.7	38.6	80.5	63.8
Mancs [8]	93.1	82.3	69.0	63.9	65.5	60.5	84.9	71.8
SphereReID [9]	**94.4**	83.6	66.8*	65.7*	66.1*	63.6*	83.9	68.5
TBCL	94.2	**84.8**	**75.9**	**73.3**	**72.7**	**70.5**	**85.8**	**74.0**

a) Bold numbers denote the best performance. $*$ represents the result of the implementation.

Considering the Eq. (3), the denominator contains the operation of finding the absolute value, which is cumbersome to calculate and is not conducive to gradient back propagation. To solve this issue, we normalize the vectors. f_i^* and c_i^* denote the normalized feature vectors, i.e., $f_i^* = \frac{f_i}{\|f_i\|}$, $c_i^* = \frac{c_i}{\|c_i\|}$. The cosine similarity is calculated as follows:

$$S(f_i, c_i) = \frac{f_i c_i}{\|f_i\|\|c_i\|} = f_i^* c_i^*. \tag{4}$$

Notably, since the position of the centers in each batch is changing, which is completely different from [2], our model's generalization capability is relatively strong. To better understand TBCL, we can study the gradient optimization of TBCL with feature embeddings. We assume $F[\cdot]$ outputs 1 if $L_i > 0$ and outputs 0 otherwise. $q_i = \mathrm{argmin}_{i \neq j} S(f_i, c_j)$ is an integer number which stands for the hardest batch center of the i-th sample, and L_i represents the i-th sample of the TBCL.

$$L_i = \max\left(\max_{i \neq j} S(f_i, c_j) + m - S(f_i, c_i), 0\right). \tag{5}$$

Then, its derivation can be calculated as follows:

$$\begin{aligned} \frac{\partial L_{\mathrm{tbc}}}{\partial f_i} &= \left(\frac{\partial S(f_i, c_j)}{\partial f_i} - \frac{\partial S(f_i, c_i)}{\partial f_i}\right) F[L_i > 0], \\ &= (c_j - c_i) F[L_i > 0]. \end{aligned} \tag{6}$$

Experiments. In the experiments, the re-ID network uses ResNet-50 as the backbone; it outputs a feature vector of 1024-dim, which is connected to two loss functions, i.e., the proposed TBCL and softmax loss. We evaluated our proposed TBCL against previous existing methods on Market-1501, CUHK03, and DukeMTMC-re-ID. For more details of our experiments, please refer to the source code released on the website[1)].

As shown in Table 1, our model achieved the second-best results on Market-1501 and the best results on CUHK03 and DukeMTMC-re-ID. It outperformed the second-best method of Mancs [8] by +0.9%/+2.2%, in Rank-1/mAP, on DukeMTMC-re-ID. It is noted that our backbone is the same as that of [9] (network-D), and we use the same hyperparameters while training. However, our final results have a great improvement.

Conclusion. We highlight the main contributions of our work as follows. (1) We proposed a novel batch-level metric-learning loss function, which benefits from the design of triplet loss, center loss, sphere loss, and episode training. (2) We designed the proposed TBCL to obtain more generalizable deep learning features. It is useful for a wide range of open-set re-ID/retrieval tasks. (3) Without bells and whistles, the TBCL with ResNet-50 has achieved state-of-the-art results.

Acknowledgements This work was supported by Zhejiang Lab (Grant No. 2019NB0AB02) and National Natural Science Foundation of China (NSFC) (Grant Nos. 61876212, 61733007, 6157220).

References

1 Yu Y L, Ji Z, Guo J C, et al. Zero-shot learning via latent space encoding. IEEE Trans Cybern, 2019, 49: 3755–3766

2 He X W, Zhou Y, Zhou Z C, et al. Triplet-center loss for multi-view 3d object retrieval. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2018. 1945–1954

3 Song C F, Huang Y, Ouyang W L, et al. Mask-guided contrastive attention model for person re-identification. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2018. 1179–1188

4 Xu J, Zhao R, Zhu F, et al. Attention-aware compositional network for person re-identification. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2018. 2119–2128

5 Liu J X, Ni B B, Yan Y C, et al. Pose transferrable person re-identification. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2018. 4099–4108

6 Sarfraz M S, Schumann A, Eberle A, et al. A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2018. 420–429

7 Li W, Zhu X T, Gong S G. Harmonious attention network for person re-identification. In: Proceedings of IEEE conference on Computer Vision and Pattern Recognition, 2018. 2285–2294

8 Wang C, Zhang Q, Huang C, et al. Mancs: a multi-task attentional network with curriculum sampling for person re-identification. In: Proceedings of European Conference on Computer Vision (ECCV), 2018. 365–381

9 Fan X, Jiang W, Luo H, et al. SphereReID: deep hypersphere manifold embedding for person re-identification. J Vis Commun Image Represent, 2019, 60: 51–58

1) https://github.com/hustvl/tbcl.