Appendix

1. More Experimental Results

1.1. More Experiments for Parameter Analysis

Analysis of the Maximum Ranking Position η for Positive Sample. The maximum ranking position η is a tunable hyper-parameter in the ranking-based triplet loss (RTL), as shown in Eq. (4) in the main paper, which defines the range $(0, \eta]$ for selecting positive samples and the range $(\eta, 2\eta]$ for negative samples. We conduct several experiments to evaluate the sensitivity of our method to η when transferring from Duke [3] to Market-1501 [2], as shown in Table 1. It shows that when η is equal or larger than 20, we can obtain nearly same and competitive results. And we set $\eta = 20$ in all experiments except this part. The performance drops quickly when η is extremely small. We believe that it is due to the unbalanced identities, *e.g.*, the minimal and maximal numbers of images are 2 and 72 respectively in the training set of Market-1501, which results in a large probability that the selected positive and negative samples are from the same (ground-truth) identity.

D→M				
η	Rank-1	Rank-5	Rank-10	mAP
5	71.85	83.22	87.00	44.37
10	73.78	84.09	87.62	47.64
15	77.43	86.70	89.99	51.77
20	**78.38**	**88.63**	**92.01**	**54.62**
25	78.27	88.63	91.63	55.27
30	78.15	88.93	91.95	54.46
35	78.59	88.48	91.83	55.10

Table 1 – The influence of maximum ranking position η for triplet selection of RTL in our PAST framework on D→M setting.

1.2. Analysis of the Initialization of the Classifier

As described in Section 3.3, we use the feature-based weight initialization for the newly added classifier in the promoting stage. For better displaying the merit of our method, we apply random initialization instead of our initial method for comparison. As shown in Figure 1 , we can observe that in the early iteration, the random initialization is harmful for the optimization of the model. When the process is close to the end, there is no further improvement with random initialization. We also quantitatively represent that our proposed method outperforms random initialization by 1.24% and 1.8% on Rank-1 and mAP, shown in the Tabel. 2. It is noticing that our proposed initial method is good for the convergence of the model training as well as the generalization of the model.

1.3. More Qualitative Analyses

Qualitative Analysis of the Feature Representation. To demonstrate the results intuitively, we visualize the feature embeddings calculated by our PAST framework in 2-D using t-SNE [1]. Three representative classes are displayed by showing the corresponding images in the bottom, *i.e.*, *true positive samples*, *false positive samples* and *false negative samples*. As illustrated in Figure 2, images belonging to the same identity are almost well gathered together, while those from different classes usually stay apart from each other. It implies that our PAST framework can improve the capability of model generalization which is beneficial for learning discriminative feature representation on the target-domain dataset.

Qualitative Analysis of the Triplet Selection. In Figure 3, we visualize the triplet samples generated in the conservative stage for CTL and RTL, respectively. We summarize the main advantages of the proposed PAST method in the following.

1. The proposed PAST algorithm can significantly improve the quality of the clustering assignments during training. As shown in the first row of the iterations from 1 to 4, the images assigned to the same class by the proposed method tend to be more and more similar. On the other hand, the quality of the pseudo labels assigned to each images is steadily improved during training. It means that our PAST framework is beneficial for learning discriminating feature presentation and can assign more reliable pseudo labels to target images. The accurate pseudo labels can be used to promoting stage to improve the model generalization further.

D→M				
Method	Rank-1	Rank-5	Rank-10	mAP
Random	77.14	87.20	90.59	52.82
Ours	**78.38**	**88.63**	**92.01**	**54.62**

Table 2 – The comparison of our feature-based weight initialization and random initialization.

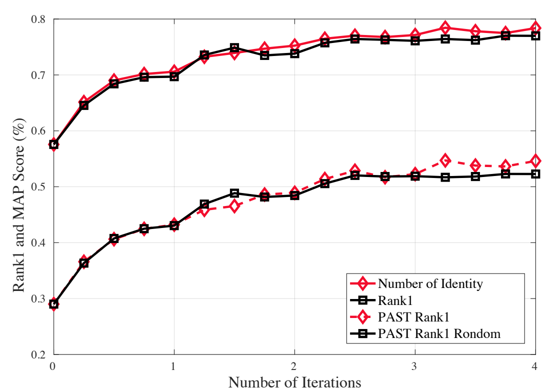

Figure 1 – The performance comparison of random initialization and our feature-based weight initialization on the classifier in the promoting stage.

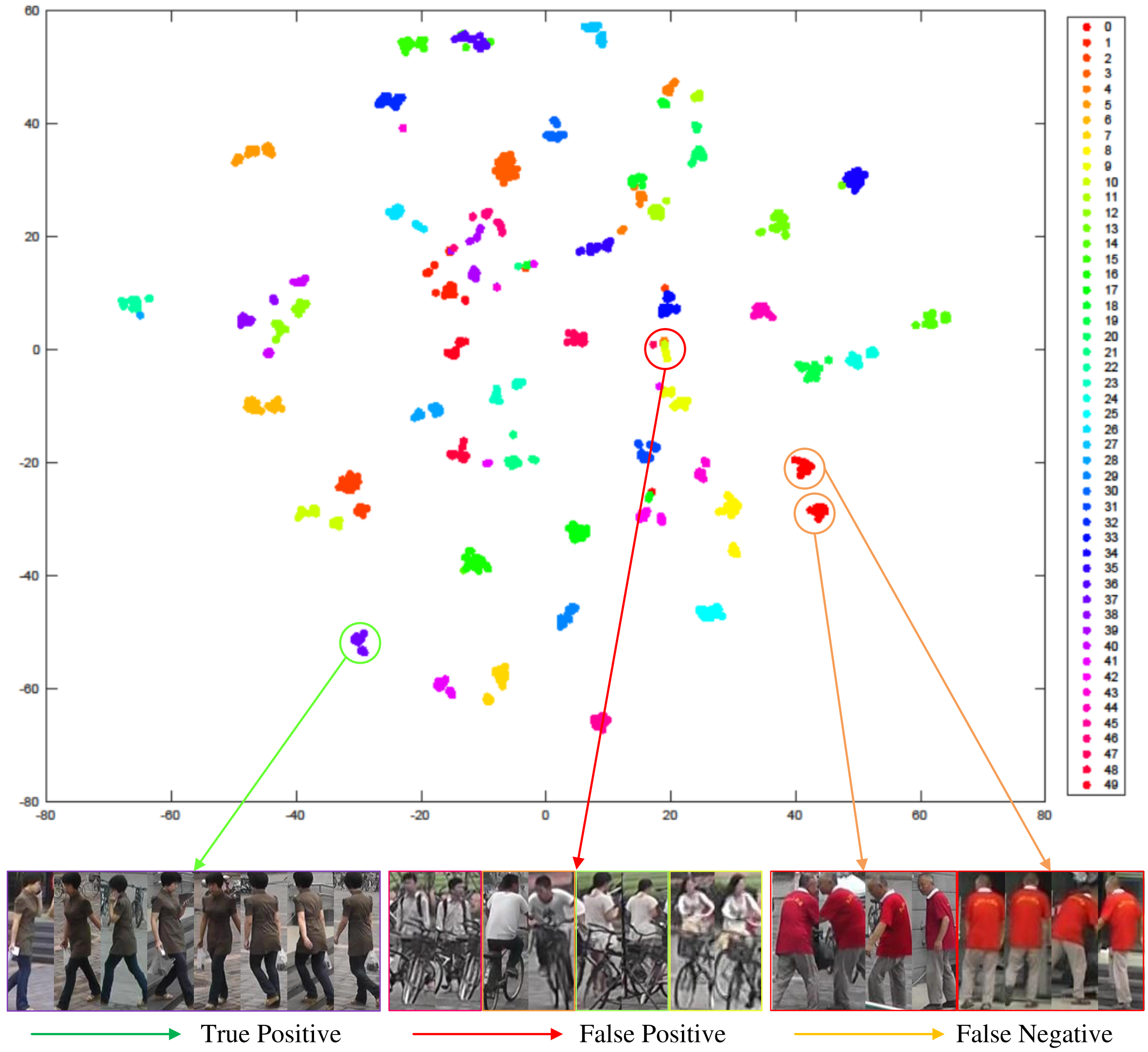

Figure 2 – Qualitative analysis of the feature representation by using t-SNE [1] visualization on a subset of Market-1501 [2] training data. According to the clustering result, we choose the Top-50 identities which contain Top-50 the largest number of images. Points with the same color have the same (ground-truth) identity. The **green circle** means images from the same identity are gathered together, and the cluster is extremely reliable. Images in **orange circle** are both from same identity, yet they are clustered to two different classes. We can see that due to the camera style, images from the two classes have different appearances. In the **red circle**, although our algorithm may gather images from different (ground-truth) identities into the same cluster, these images usually share very similar appearances and are hard to distinguish with each other. For instances, every image in the red circle contains one person with white clothes and a black bicycle.

2. RTL is useful for remedying the variance caused by CTL. Refer to Figure 3 again, we can observe that the third cluster in iteration 2 is noisy and the selected triplets from CTL are not faithful. However, RTL can select correct positive sample even the cluster is dirty. We believe that the reason is that RTL just depends on the similarity ranking matrix and the top η similar images are used for generating positive samples, which is more reliable when the features representation is not so discriminative.

3. RTL helps to further optimize the network, especially in the later iteration. From Figure 3, we can also see that different clusters in one mini-batch may look different due to unique color of clothes, which results in extremely simple negative samples and slows down the optimization when training on CTL. Whereas, considering the triplets generated from the RTL, negative images are extremely similar to the anchors, which is even hard to be well recognized by human beings. For example, at the second column in iteration 4, all images look like one person, although images from the first two rows are same person, while those from the third row belong to another person.

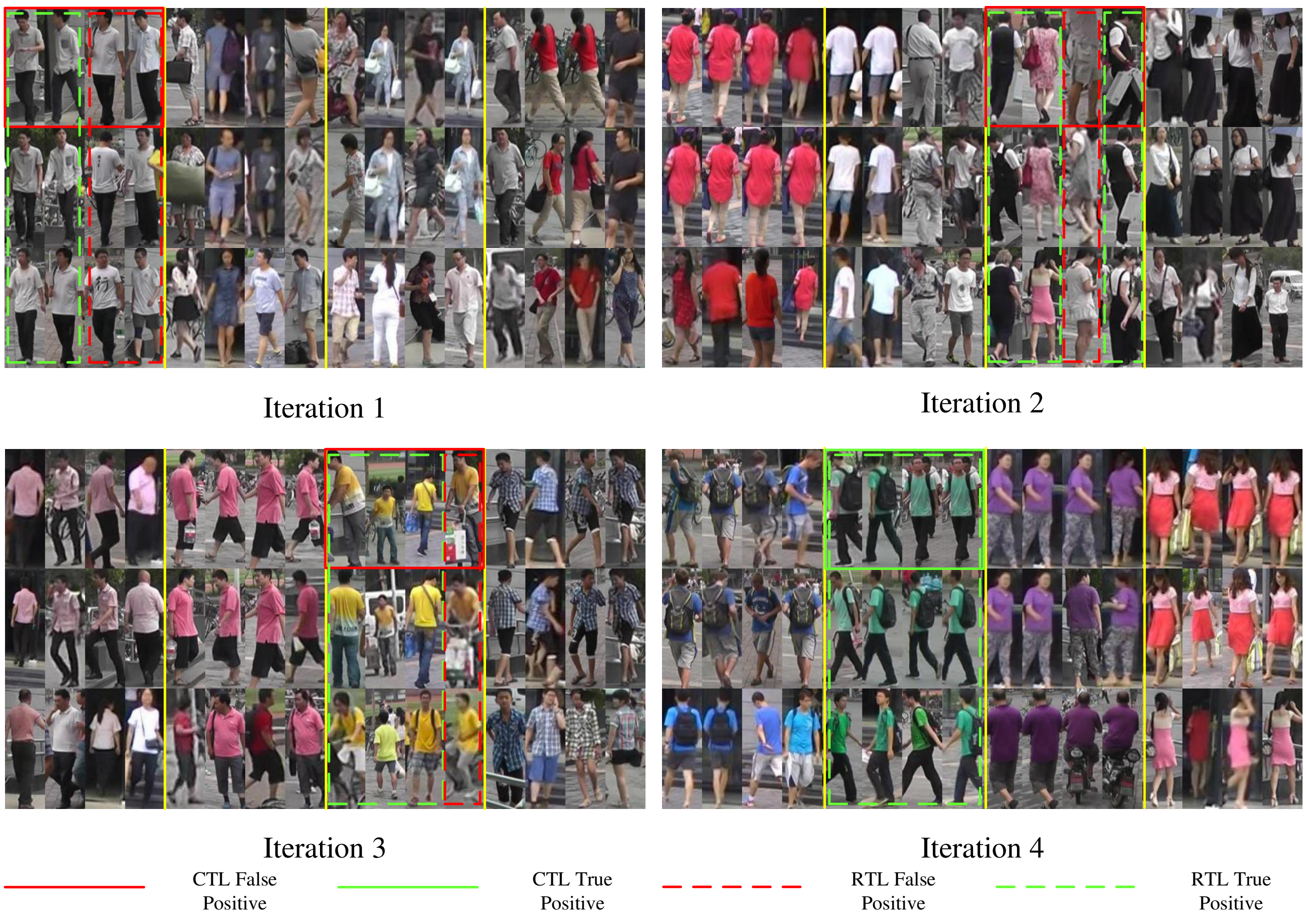

Figure 3 – Quality of the triplet selection over training iterations. Images from different clusters are divided by **yellow line**. The **red line** means generated triplets are not completely correct, while **green line** represents generated triplets are completely correct. The solid line and dashed line are for triplets, which are generated from CTL and RTL respectively. We use Duke [3] as the source domain and Market-1501 [2] as the target domain.

References

[1] Laurens Van Der Maaten. Accelerating t-sne using tree-based algorithms. *J. Mach. Learn. Res.*, 15(1):3221–3245, 2014.

[2] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Proc. IEEE Int. Conf. Comp. Vis.*, pages 1116–1124, 2015.

[3] Zhedong Zheng, Liang Zheng, and Yi Yang. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *Proc. IEEE Int. Conf. Comp. Vis.*, pages 3754–3762, 2017.