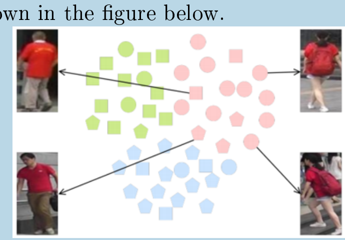
Progressive Unsupervised Domain Adaptation for Image-based Person Re-Identification

Mingliang Yang, Jing Zhao, Da Huang, Ji Wang
{yangmingliang18, zhaojing, wj}@nudt.edu.cn, huangda1109@163.com

Problem

Person Re-Identification (Re-ID) aims at retrieving a specific person from images or videos captured by different cameras from various times and places, which has attracted significant attention in recent years. The most existing approaches have achieved good results in the case of fully supervised training where all the images require manual annotations, consuming large amounts of labor and time.

Unsupervised Domain Adaptation (UDA) has emerged as an effective paradigm for reducing the huge manual annotation cost for Person Re-ID. Many of the recent UDA methods for Re-ID are clustering-based and select all the pseudo-label samples in each iteration for the model training. However, there are many wrong labeled samples that will mislead the model optimization under this circumstance, as shown in the figure below.



Progressive Unsupervised Domain Adaptation

In the case of one-shot video-based Person Re-ID, Wu *et al.* [1] propose the **EUG (Exploit the Unknown Gradually)** method to exploit unlabeled tracklets in a step-wise manner to improve the discriminative capability of the CNN feature representation. They firstly train the model on the labeled tracklets, then gradually but steadily extent the training set by selecting unlabeled samples with higher pseudo-label confidence. Such supervised training and dataset extension stage alternate during the iteration process till all the unlabeled samples are selected, greatly improving the performance of the model.

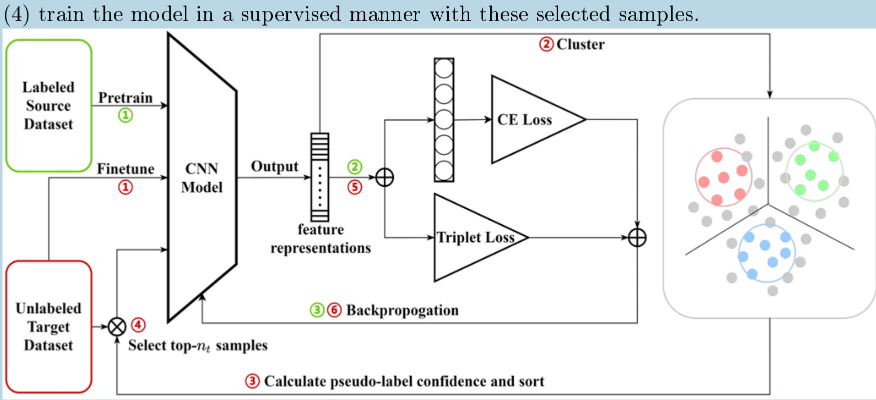
Inspired by EUG, we hope to solve the UDA problem through the progressive method. However, there are some challenges in applying the progressive method to our works. On the one hand, the progressive method is based on a one-shot condition, which is different from the unsupervised situation without any annotation. On the other hand, each tracklet in video-based ReID consists of many slightly different frames, and the model can ignore some noise features by averaging all the frames' features in the early iterations, which is impossible for image based Re-ID with single image. So, it is hard for the model to distinguish which features are important for an identity and which are random noises in the image-based condition.

In this paper, we propose a **Progressive Unsupervised Domain Adaptation (PUDA)** framework for the image-based Person Re-ID to reduce the negative effect of noise pseudo-label samples on the model optimization process by applying the progressive method on this unsupervised task.

The Proposed Framework

Here is the overview of our framework. We first pretrain a CNN model on the labeled source dataset, then finetune the model on the unlabeled target dataset with an iterative approach. The pretraining and finetuning processes are marked with green and red serial numbers respectively.

In each finetuning iteration, we (1) extract the feature representations of all the samples in target dataset; (2) generate pseudo labels for them by clustering algorithms; (3) select more pseudo-label samples with higher confidence as a new training set for current iteration; (4) train the model in a supervised manner with these selected samples.



Experiment Results

We implement our framework based on MMT[2], which is the state-of-the-art method in the unsupervised image-based Person Re-Identification, outperforming MMT by 4.2 points on Market-1501[3] and 1.7 points on DukeMTMC-reID[4], the two large-scale datasets.

Methods	Duke-to-Market				Market-to-Duke			
	mAP	top-1	top-5	top-10	mAP	top-1	top-5	top-10
Pretrained	35.7	65.8	79.8	84.5	35.2	53.1	67.5	72.3
UDAP [24]	53.7	75.8	89.5	93.2	49.0	68.4	80.1	83.5
PCB-PAST [25]	54.6	78.4	-	-	54.3	72.4	-	-
SSG [7]	58.3	80.0	90.0	92.4	53.4	73.0	80.6	83.2
Co-teaching($\hat{C}^t$=500) [27]	71.7	87.8	95.0	96.5	61.7	77.6	88.0	90.7
MMT($\hat{C}^t$=500) [9]	**75.8**	**91.0**	**96.8**	**98.1**	66.6	80.0	89.9	92.6
MMT($\hat{C}^t$=700) [9]	75.0	**91.0**	96.4	97.7	**69.0**	**81.5**	**90.9**	**93.3**
MMT($\hat{C}^t$=900) [9]	73.0	90.8	96.5	98.0	68.0	81.2	90.7	93.1
Ours PUDA($\hat{C}^t$=500, w/o MIS)	73.9	88.0	95.3	97.1	-	-	-	-
Ours PUDA($\hat{C}^t$=700, w/o MIS)	-	-	-	-	39.6	58.0	68.6	72.1
Ours PUDA($\hat{C}^t$=500)	**80.0**	**91.1**	96.6	**97.7**	69.0	81.1	90.4	92.9
Ours PUDA($\hat{C}^t$=700)	76.4	90.8	**96.7**	**97.7**	**70.7**	**82.6**	**91.3**	**93.9**
Ours PUDA($\hat{C}^t$=900)	72.7	90.6	96.3	**97.7**	68.4	81.1	90.4	93.3

Acknowledgements

This work was partially supported by the National Key Research and Development Program of China (2017YFB1301104 and No. 2017YFB1001900), the National Natural Science Foundation of China (No. 91648204 and 61803375), the Youth Talent Lifting Project (No.17-JCJQ-QT-047), and the National Science and Technology Major Project.

References

[1] Yu Wu, Yutian Lin, Xuanyi Dong, Yan Yan, Wanli Ouyang, and Yi Yang. Exploit the unknown gradually: One-shot video-based person re-identification by stepwise learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5177–5186, 2018.

[2] Yixiao Ge, Dapeng Chen, and Hongsheng Li. Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification. In *International Conference on Learning Representations*, 2020.

[3] 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.

[4] 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.