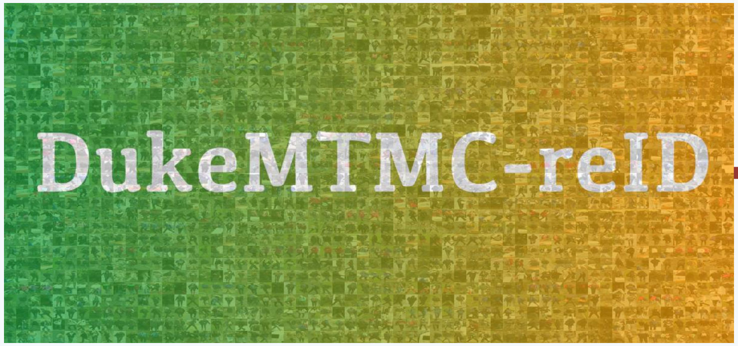
A Bottom-up Clustering Approach to Unsupervised Person Re-identification

Yutian Lin, Xuanyi Dong, Liang Zheng,Yan Yan, Yi Yang

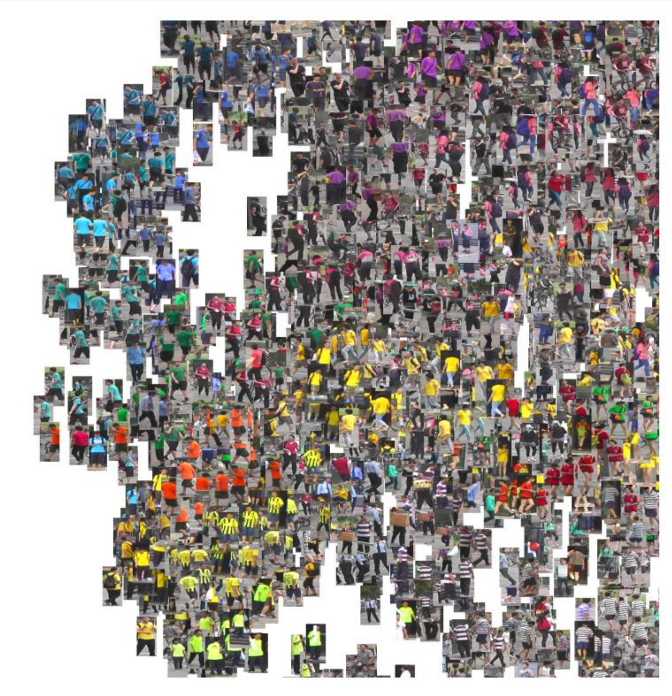
Reference: https://vana77.github.io/vana77.github.io/images/AAAI19.pdf

Intelligent Information Fusion Research Group

一般无监督学习



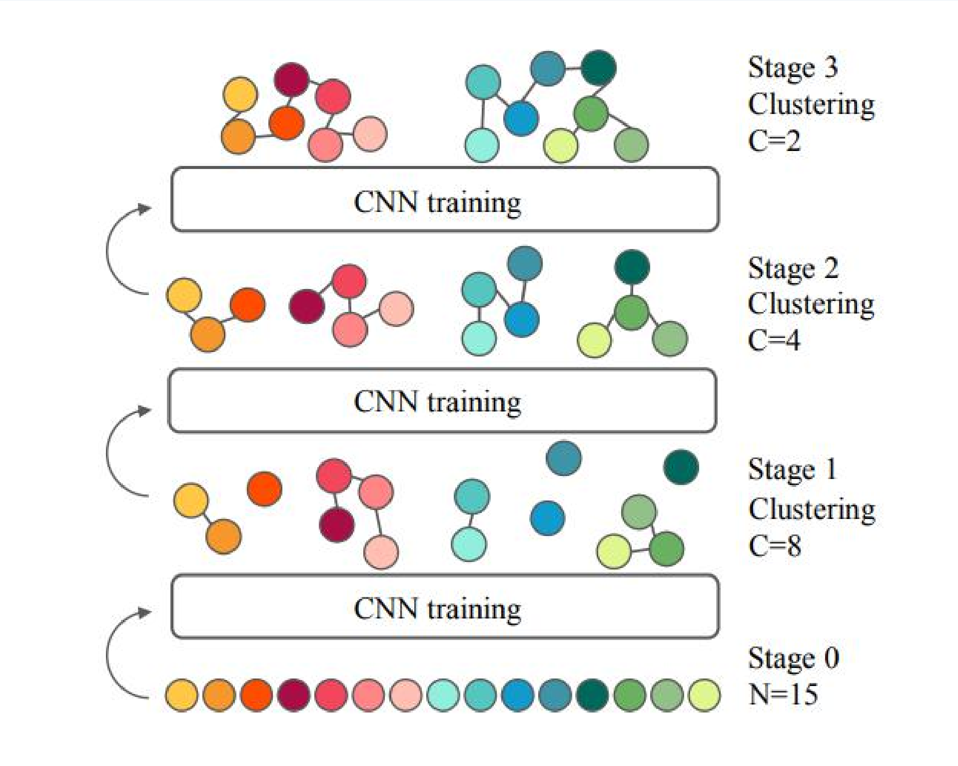
源域有label



目标无label

不是真正的无监督

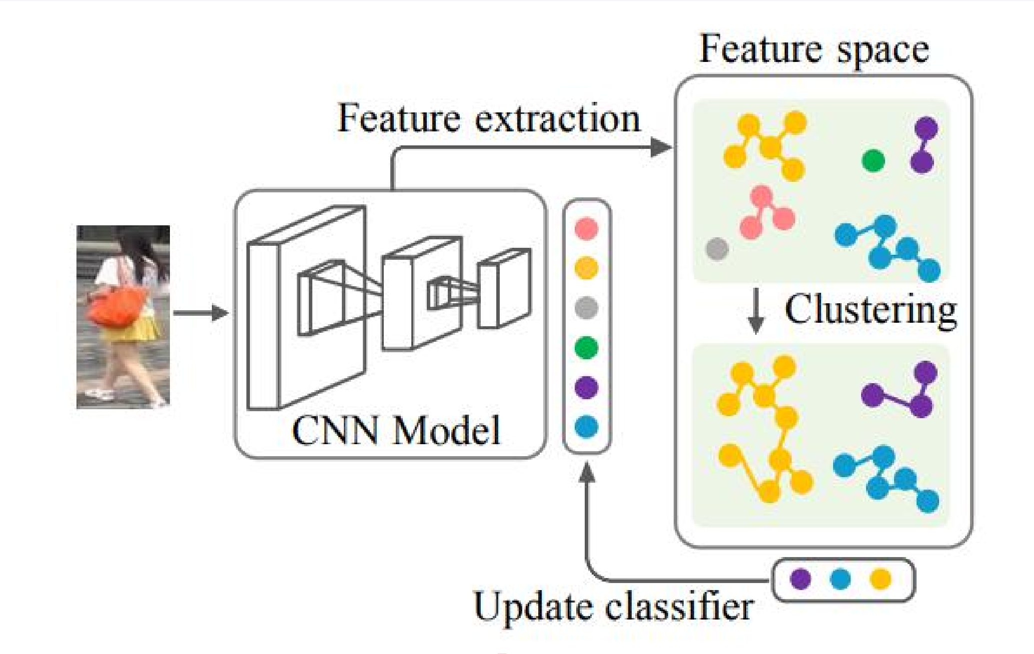
bottom-up clustering (BUC) cluster approach



同时考虑两个重要影响因素：

1. 同一个人的视觉相似度
2. 不同人的视觉差异度

框架：



相似

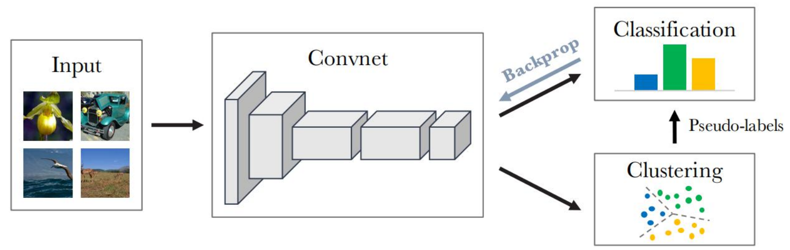
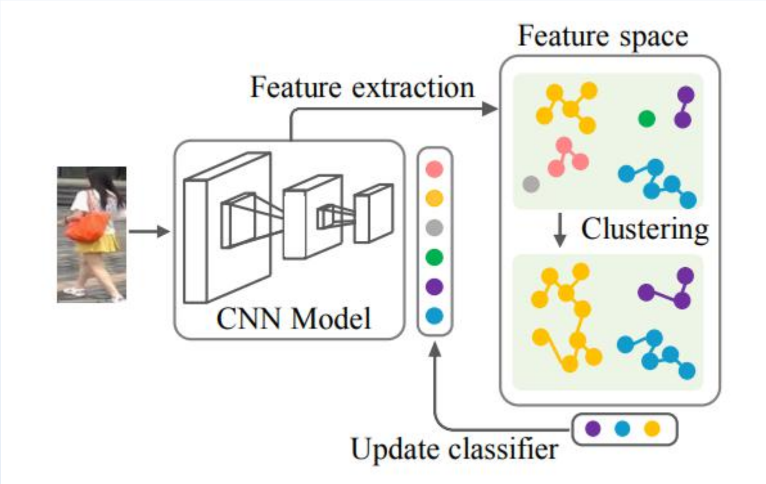


Fig. 1: Illustration of the proposed method: we iteratively cluster deep features and use the cluster assignments as pseudo-labels to learn the parameters of the convnet

有解释为什么可以

Reference: ECCV 2018 - Facebook AI Research - Deep Clustering for Unsupervised Learning of Visual Features

具体分类过程：

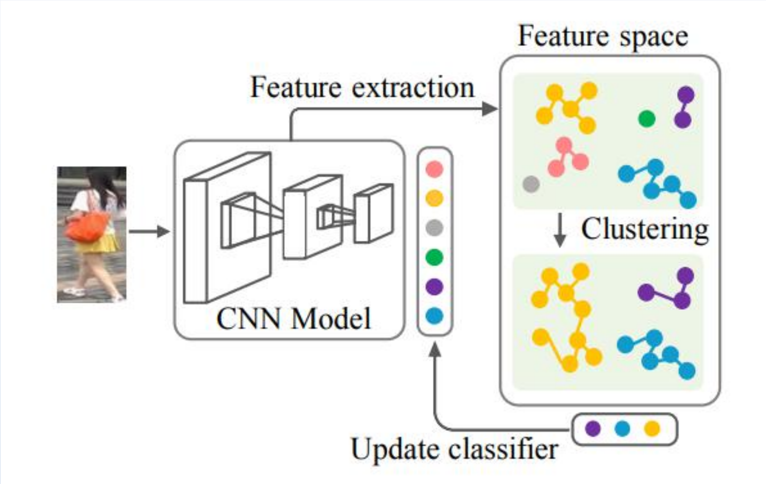


$$p(c|x, \boldsymbol{V}) = \frac{exp(\boldsymbol{V}_c^{\mathrm{T}} \boldsymbol{v}/\tau)}{\sum_{j=1}^{C} exp(\boldsymbol{V}_j^{\mathrm{T}} \boldsymbol{v}/\tau)}, \quad (2)$$

V 为聚簇的Memory Bank
为什么采用 Memory Bank, 而非全连接FC,
OIM loss 论文从实验上说明Bank比FC效果好，同时也稍微说明了好的原因

Reference: CVPR 2017 - Joint Detection and Identification Feature Learning for Person Search

具体分类过程：

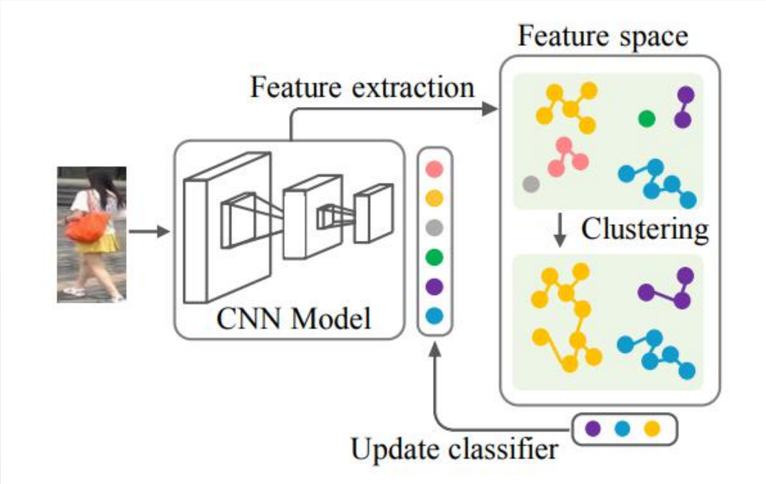


$$p(c|x, \boldsymbol{V}) = \frac{exp(\boldsymbol{V}_c^{\mathrm{T}} \boldsymbol{v}/\tau)}{\sum_{j=1}^{C} exp(\boldsymbol{V}_j^{\mathrm{T}} \boldsymbol{v}/\tau)}, \quad (2)$$

$$\mathcal{L} = -\log(p(\hat{y}_i | x_i, \boldsymbol{V})).$$

$$\boldsymbol{V}_{\hat{y}_i} \leftarrow \frac{1}{2}(\boldsymbol{V}_{\hat{y}_i} + \boldsymbol{v}_i)$$

具体聚类过程：

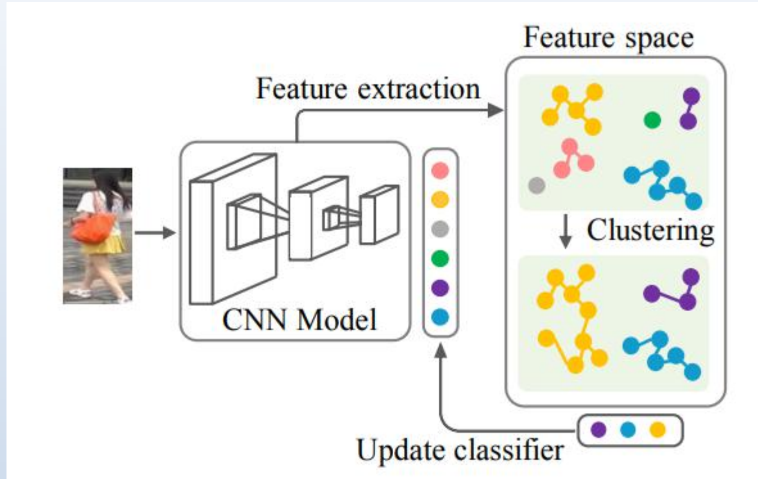


$$D_{distance}(A, B) = \min_{x_a \in A, x_b \in B} d(x_a, x_b),$$

$$d(x_a, x_b) = \|\boldsymbol{v}_a - \boldsymbol{v}_b\|.$$

每次聚类聚top-m个聚簇

具体聚类过程：

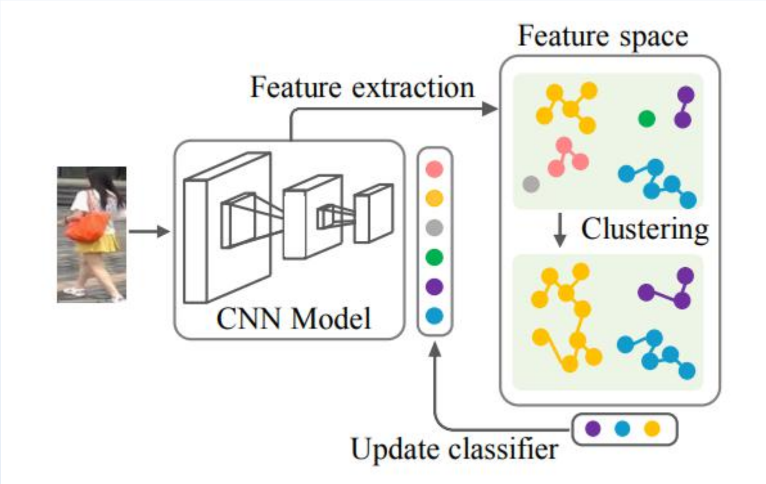


$$D_{distance}(A,B) = \min_{x_a \in A, x_b \in B} d(x_a, x_b),$$

既然是聚类，评估聚簇之间的距离方式就很重要，这里为什么要使用min?

好处：视觉相似的，都可以聚在一起, 特别同摄像头
坏处：容易聚太多

具体聚类过程：



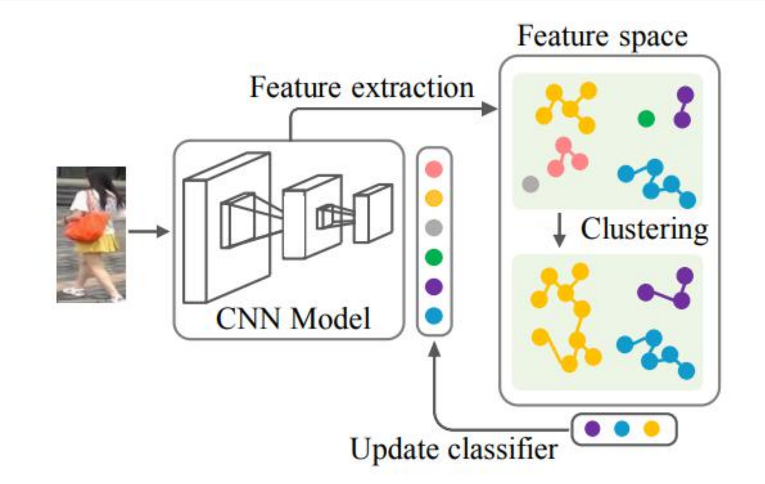
$$D_{distance}(A,B)=\min_{x_a\in A,x_b\in B} d(x_a,x_b),$$

那如果采用 max 或者 centroid 呢？

max 同一个人跨摄像头差异本来就比较大，这样做，容易把这种情况当做不同人

centroid 聚簇中如果有同一个人来自不同摄像头，取平均容易忽略了不同摄像头下行人特征本来的差异性

具体聚类过程：

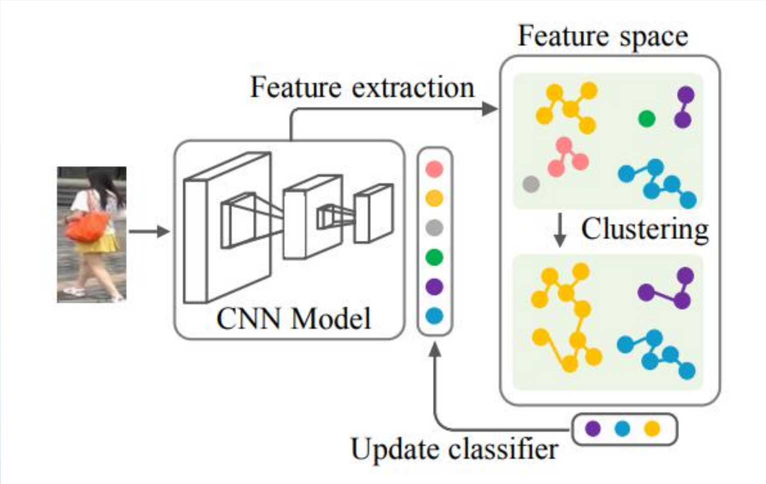


$$D_{distance}(A, B) = \min_{x_a \in A, x_b \in B} d(x_a, x_b),$$

Table 3: The comparison of different merging criteria on Market-1501.

Criterion	rank-1	rank-5	rank-10	rank-20	mAP
Maximum	62.5	76.8	82.6	87.1	35.0
Centroid	65.8	79.2	83.6	88.4	37.9
Minimum	**66.2**	**79.6**	**84.5**	**88.5**	**38.3**

具体聚类过程：



$$D_{distance}(A,B) = \min_{x_a \in A, x_b \in B} d(x_a, x_b),$$

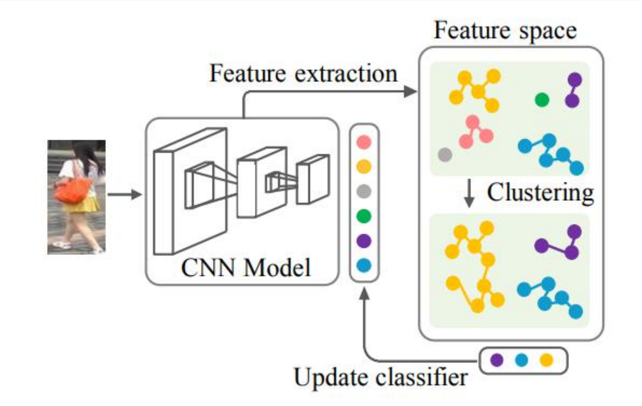
既然是聚类，评估聚簇之间的距离方式就很重要，这里为什么要使用min?

好处：视觉相似的，都可以聚在一起, 特别同摄像头

坏处：容易聚太多

怎么办？？？

具体聚类过程：



$$D_{distance}(A,B) = \min_{x_a \in A, x_b \in B} d(x_a, x_b),$$

加正则

$$D_{diversity}(A,B) = |A| + |B|,$$

|A| = #A

假设每个人的图片数分布还是比较平均的

最终的距离：

$$D(A,B) = D_{distance}(A,B) + \lambda D_{diversity}(A,B)$$

算法

总体算法流程:

Algorithm 1 The Bottom-Up Clustering (BUC) Framework

Require: Unlabeled data $X = \{x_1, x_2, ...x_N\}$;
Merge percent $mp \in (0, 1)$;
CNN model $\phi(\cdot; \boldsymbol{\theta}_0)$.

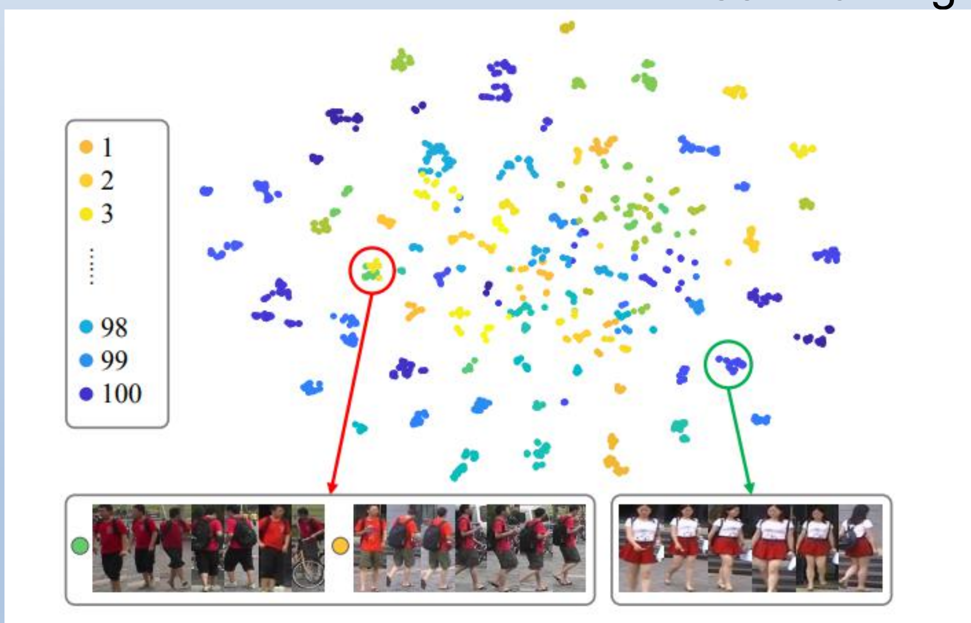
Ensure: Best CNN model $\phi(\cdot; \boldsymbol{\theta}^*)$.

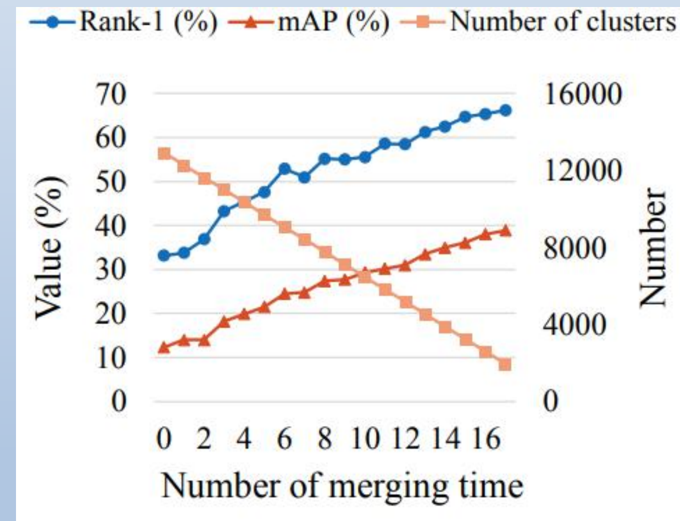
Initialize: Cluster label $Y = \{\hat{y}_i | \, 1 \leq i \leq N\}$
    Number of cluster $C = N$
    Number of merging image $m = \lceil mp * C \rceil$
while $C > m$ **do**
    Train CNN model $\phi(x; \boldsymbol{\theta})$ with X and Y
    Clustering with m:
        $C \leftarrow C - m$
        Update Y with the new cluster labels
    Initialize the lookup table $\boldsymbol{V}$ with new dimensions
    Re-train the CNN model with parameters $\boldsymbol{\theta}$
    Evaluate on the validation set $\rightarrow$ performance P
    if $P > P^*$ **then**
        $P^* = P_t$
        Best model = $\phi(x; \boldsymbol{\theta})$
    end if
end while

实验:

Methods	Venue	Labels	Market-1501				DukeMTMC-reID			
			rank-1	rank-5	rank-10	mAP	rank-1	rank-5	rank-10	mAP
BOW (Zheng et al. 2015)	ICCV15	**None**	35.8	52.4	60.3	14.8	17.1	28.8	34.9	8.3
OIM* (Xiao et al. 2017)	CVPR18	**None**	38.0	58.0	66.3	14.0	24.5	38.8	46.0	11.3
UMDL (Peng et al. 2016)	CVPR16	Transfer	34.5	52.6	59.6	12.4	18.5	31.4	37.6	7.3
PUL (Hehe et al. 2018)	TOMM18	Transfer	44.7	59.1	65.6	20.1	30.4	46.4	50.7	16.4
EUG* (Wu et al. 2018a)	CVPR18	OneEx	49.8	66.4	72.7	22.5	45.2	59.2	63.4	24.5
SPGAN (Deng et al. 2018)	CVPR18	Transfer	58.1	76.0	82.7	26.7	46.9	62.6	68.5	26.4
TJ-AIDL (Wang et al. 2018)	CVPR18	Transfer	58.2	-	-	26.5	44.3	-	-	23.0
BUC without diversity regularizer	AAAI19	**None**	62.9	77.1	82.7	33.8	41.3	55.8	62.5	22.5
BUC	AAAI19	**None**	**66.2**	**79.6**	**84.5**	**38.3**	**47.4**	**62.6**	**68.4**	**27.5**

self-training 有潜力





(b)

总结：

1. 卖点好，self-training，完全无监督

2. 论文流畅，简单易读

3. 实验严谨，对重点需要解读的部分，有实验证明，有原因阐述

4. 图也画的漂亮