

Exploit the Unknown Gradually: One-Shot Video-Based Person Re-Identification by Stepwise Learning

一. 论文概述

作者将基于视频的行人重识别问题视作单样本学习问题, 在文中提出了一种新的方法去利用未标记的片段逐步、稳定地去学习具有辨别性的特征。

开始的时候, 作者直接用每一个标记好身份的片段初始化网络。之后, 在网络的训练过程中, 从未标记的片段数据集中选择一些具有较好相关性的备选片段作为伪标记样本, 然后根据这些备选数据训练 CNN 网络。最后, 完成对行人身份的识别。

相比于前面静态的采样策略, 作者提出了一种更具进步性的采样方法, 可以逐步地扩大备选数据集的大小。

二. 作者方法

模型框架

作者提出了一种逐步学习的方法, 简称为 EUG (Exploit the Unknown Gradually)。其中, CNN 网络用单样本片段进行初始化。EUG 会根据预测的可靠性不断地从未标记的片段数据中选取伪标记片段作为备选数据集, 之后利用这些备选数据和前面标记好的单样本数据重新训练 CNN 网络。EUG 刚开始学习的时候只采用一个很小的子数据集, 这个子数据集仅包括最容易获得的、最可靠的伪标记片段。在后续的学习中, EUG 将逐渐扩大备选数据集, 形成学习更困难和更多样的数据集。

作者将伪标记样本备选数据开发下的单样本重识别问题表示为如下形式：

$$\min_{\theta, w, s_i, \hat{y}_i} \sum_{i=1}^{n_l} \ell(f(w; \phi(\theta; x_i), y_i)) + \sum_{i=n_l+1}^{n_l+n_u} s_i \ell(f(w; \phi(\theta; x_i), \hat{y}_i))$$

其模型框架概观, 如图 2.1 所示。

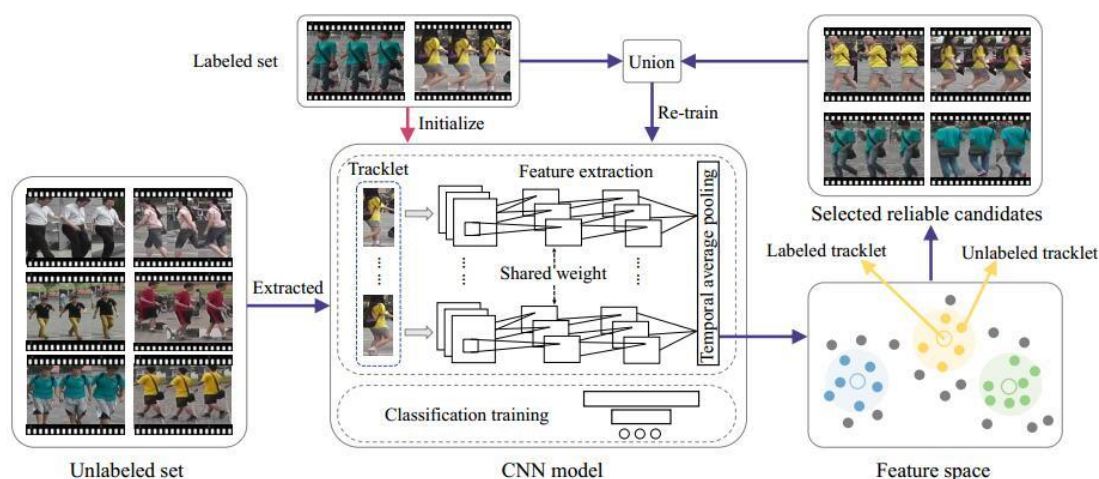


图 2.1 EUG 模型框架的概观

记 \mathcal{L} 表示标记好的数据集, \mathcal{S} 表示选择出来的伪标记备选数据集, $\mathcal{D} = \mathcal{L} \cup \mathcal{S}$ 表示

新的训练数据集。数据集 \mathcal{D} 被用来重新训练模型, 使模型更加稳健。作者在模型中采用带时序平均池的端到端 CNN 网络 (简称为 ETAP-Net) 作为特征嵌入函数来提取识别特征。对于每一个片段, 帧特征在被 CNN 网络提取之后经过时序平均池化形成片段特征。

在标签估计的阶段, 对于每一个未标记的片段, 伪标记片段都按照对应的身份标记的最近邻片段特征空间进行选取。同时采用相异成本 (dissimilarity cost) 来表示两个片段之间的测度距离, 它可以用来衡量这个伪标记的可靠性, 即与真实标记身份的相近程度。

作者将训练的过程视为是一个身份分类的问题, 将标签评估视为是一个对测试片段的特征检索问题。每一次迭代, 首先都选择出一些可靠的伪标记样本扩大备选数据集, 之后再用新生成的备选数据集重新训练网络。

采样方法

作者采用最近邻分类 (简称 NN 分类器) 来进行标签估计。同时, 将未标记数据与它的最近邻标记样本之间的度量距离记作标签估计的可信度。记该距离为相异成本, 表示为:

$$d(\theta; x_i) = \min_{x_l \in \mathcal{L}} \|\phi(\theta; x_i) - \phi(\theta; x_l)\|_2$$

在每次采样时, 根据标签估计的可信度选择表现最好的几个伪标记样本计入备选数据集。其过程如图 2.2 所示。

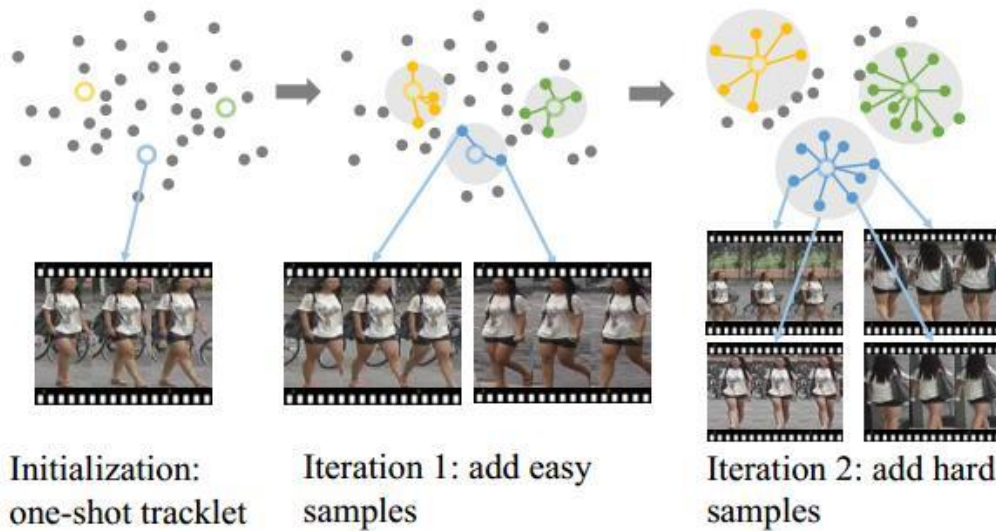


图 2.2 伪标记样本生成过程

通过实验, 作者发现采样的标准和更新的备选数据集的速度将对最终伪标记样本的选择产生很大的影响。如果采用一个缓慢的方式逐渐生成伪标记数据能使模型的识别效果更好。相反, 一个快速扩大备选数据的方式将会使模型的性能明显下降。

所以, 最终论文的作者选择用基于距离衡量的方法作为采样的准则并采用一个很小的更新动量 p 在每一次迭代中扩充备选数据。其具体的逐步生成式算法如图 2.3 所示。

Algorithm 1 Exploit the Unknown Gradually

Input: Labeled data \mathcal{L} , unlabeled data \mathcal{U} , enlarging factor $p \in (0, 1)$, initialized CNN model θ_0 .

Output: The best CNN model θ^* .

- 1: Initialize the selected pseudo-labeled data $\mathcal{S}_0 \leftarrow \emptyset$, sampling size $m_1 \leftarrow p \cdot n_u$, iteration step $t \leftarrow 0$, best validation performance $V^* \leftarrow 0$
- 2: **while** $m_{t+1} \leq |\mathcal{U}|$ **do**
- 3: $t \leftarrow t + 1$
- 4: Update training set: $\mathcal{D}_t \leftarrow \mathcal{L} \cup \mathcal{S}_{t-1}$
- 5: Train the CNN model (θ_t, w_t) based on \mathcal{D}_t .
- 6: Generate the selection indicators s_t via Eq. (5)
- 7: Update \mathcal{S}_t based on s_t via Eq. (3)
- 8: Update the sampling number: $m_{t+1} \leftarrow m_t + p \cdot n_u$
- 9: **end while**
- 10: **for** $i \leftarrow 1$ **to** T **do**
- 11: Evaluate θ_i on the validation set \rightarrow performance V_i
- 12: **if** $V_i > V^*$ **then**
- 13: $V^*, \theta^* \leftarrow V_i, \theta_i$
- 14: **end if**
- 15: **end for**

图 2.3 伪标记样本逐步生成式算法

一. 实验结果

相比于之前最好的方法，文中提出的 EUG 模型在 MARS 和 DukeMTMC 数据集上的表现都表现除了令人满意的结果。其具体实验结果如表格 3-1 所示。

表 3-1 与之前最好方法的比较

Methods	MARS				DukeMTMC-VideoReID			
	rank-1	rank-5	rank-20	mAP	rank-1	rank-5	rank-20	mAP
Baseline (one-shot)	36.16	50.20	61.86	15.45	39.60	56.84	66.95	33.27
DGM+IDE[34]	36.81	54.01	68.51	16.87	42.36	57.92	69.31	33.62
Stepwise[21]	41.21	55.55	66.76	19.65	56.26	70.37	79.20	46.76
EUG ($p = 0.30$)	42.77	56.51	67.17	21.12	63.82	78.64	87.04	54.57
EUG ($p = 0.20$)	48.68	63.38	72.57	26.55	68.95	81.05	89.46	59.50
EUG ($p = 0.15$)	52.32	64.29	73.08	29.56	69.08	81.19	88.88	59.21
EUG ($p = 0.10$)	57.62	69.64	78.08	34.68	70.79	83.61	89.60	61.76
EUG ($p = 0.05$)	62.67	74.94	82.57	42.45	72.79	84.18	91.45	63.23
Baseline (supervised)	80.75	92.07	96.11	67.39	83.62	94.59	97.58	78.34