

半监督单样本深度行人重识别方法^①

单纯, 王敏

(河海大学 计算机与信息学院, 南京 211100)
通讯作者: 单纯, E-mail: 649661204@qq.com



摘要: 本文提出一种采用单样本训练的行人重识别方法, 在迭代的过程中采用一种渐进学习框架, 充分利用有标签数据和无标签数据的特性来优化模型. 本文方法主要分为以下 3 个步骤: (1) 训练卷积神经网络来不断优化模型; (2) 样本评估: 通过本文提出的抽样策略, 使用多个模型共同训练, 共同挑选出较优的伪标签数据; (3) 进行下一轮的训练更新数据. 在训练的过程中, 我们训练数据由有标签数据、伪标签数据、映射标签数据三部分组成, 使用三组数据进行联合学习, 每组数据对应使用相应的损失函数对模型进行优化, 并且随着迭代的进行, 伪标签数据和映射标签数据总是不断更新. 在使用单样本训练条件下, rank-1=65.3, mAP=45.6. 当训练数据的标注率提升至 40% 时, rank-1=83.8, mAP=64.9. 实验结果表明: 本文提出的半监督行人重识别方法可以在使用更少标签数据的情况下, 提供与完全监督学习方法相媲美的结果, 充分体现了本方法的有效性.

关键词: 半监督学习; 单样本; 渐进学习框架; 多模型; 联合学习

引用格式: 单纯, 王敏. 半监督单样本深度行人重识别方法. 计算机系统应用, 2020, 29(1): 256-260. <http://www.c-s-a.org.cn/1003-3254/7270.html>

Semi-Supervised Person Re-Identification with One-Sample

SHAN Chun, WANG Min

(College of Computer and Information, HOHAI University, Nanjing 211100, China)

Abstract: In this study, we propose a one-sample person re-identification method, which adopts a progressive learning framework in the process of iteration in order to making full use of the characteristics of labeled data and unlabeled data to optimize the model. In this framework, we iteratively train convolutional neural network to update the model and utilize multiple-model training together to select the reliable pseudo-label data during label estimation. Then, we update training data for the next round of training. The training data is splitted into three parts: labeled data, pseudo-labeled data, and indexed-labeled data. We set up the corresponding loss function for each set of data and update the CNN model by the joint training on the three parts. In the progress of iteration, the pseudo-labeled data and index-labeled data are constantly updated. Under the one-sample set, rank-1=65.3, mAP=45.6. When the rate of labeled data is increased to 40%, rank-1=83.8, mAP=64.9. The result indicates that the semi-supervised person re-identification method proposed in this study can provide excellent results comparable to the supervised learning method with less labeled data, which fully demonstrates the effectiveness of the method.

Key words: semi-supervised; one-sample; progressive-learning; multi-model; joint-learning

1 引言

行人重识别又称行人再识别^[1,2]. 当人脸识别失效时, 行人重识别就成为了一个非常重要的替代技术. 简

单而言就是给定一个监控行人的图像, 从资料库中检索该行人的图像^[3]. 但是由于不同照相机的参数、位置不同, 行人的姿态变化、光照、遮挡等诸多因素, 行人

① 收稿时间: 2019-06-20; 修改时间: 2019-07-16, 2019-08-06; 采用时间: 2019-08-16; csa 在线出版时间: 2019-12-27

重识别的研究仍面临着许多的挑战。

随着深度学习的发展,卷积神经网络已经被证实 在许多计算机视觉任务中具有十分出众的表现.卷积神经网络模型在行人重识别任务中的潜力也逐渐被发掘出来.相较于传统的行人重识别方法,基于深度卷积神经网络的行人重识别方法,对于个体图片提取的特征具有更好的描述能力以及更强的鲁棒性.当今基于深度学习的行人重识别方法主要可以分为相似度学习以及特征表示两大类,第一类主要是在端到端的深度神经网络中,通过不同的验证损失学习特征,第二类则是将不同 ID 的图片放入到深度神经网络中,通过 ID 损失学习生成一个分类器.目前,绝大部分的行人重识别方法是基于监督式学习方法. Li 等^[4]利用空洞卷积进行多尺度特征提取,并融合局部特征与全局特征. Wojke 等^[5]对传统 Softmax 分类器简单再参数化,优化了余弦相似度,进而实现对特征空间的学习.然而基于监督学习的行人重识别需要大量标注的训练数据,并且这些数据要求来自不同摄像头以及不同的视角.当数据集不断扩大,大量的人工标注数据需要耗费大量的人力和时间,这使得监督式行人重识别方法的实际应用价值大大降低.当摄像头数量大幅增加时,大量人工标注的数据会对产生很多负面影响,从而影响了行人重识别的应用.为了解决上述的困难,研究人员开始逐渐尝试利用未标注的数据,越来越多的半监督式和无监督式行人重识别方法开始出现. Huang 等人^[6]提出了一种将正则化标签赋予生成数据,并将其作为训练样本补充数据的方法. Kipf 等人^[7]提出了通过直接利用神经网络模型进行图形结构编码,在一个有监督目标上,为所有带有标注的节点进行训练. Liu 等人^[8]提出了一种半监督式耦合字典学习方法,主要是在训练阶段,从已标注和未标注的图片中,联合学习两个耦合字典.半监督学习方法的优势十分明显^[9-11],能够充分利用有限的标注数据,将标注数据与未标注数据的优势相结合,从而更好的解决行人重识别任务.

在实际应用中,为大量跨镜头的数据标注需要花费大量的精力.在本文中,为每个行人 ID 仅提供一个样本,这样的设置在解决实际问题时更具有意义.为了能够充分利用有限的标注数据,训练过程中,将训练集划分为 3 个部分:有标签数据、伪标签数据、映射标签数据.特别地,选调伪标注数据的规模是渐进式不断扩大,利用两个同时训练的卷积神经网络对于伪标签

进行新一轮筛选,再依据不相似代价从这些伪标注数据筛选出 N^k 个最优项,并加入到下一轮的 ID 分类训练中去.通过这样的方式,可以进一步提升伪标签精度,从而能够更好地优化模型.下面将具体介绍本文所提出的行人重识别方法.

2 半监督式单样本深度行人重识别

图 1 为本文提出方法的总体框架,通过多种数据的联合学习、多模型联合训练和不断扩大伪标签数据规模的渐进式学习方式来逐步优化行人重识别模型,下面将具体介绍该方法.

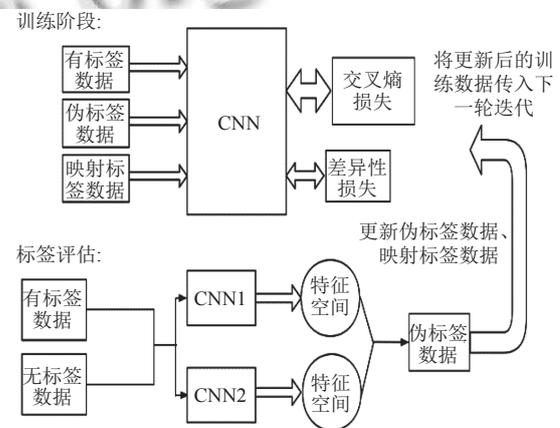


图 1 总体框架

2.1 联合学习方法

首先来介绍模型的更新步骤,在第 k 次迭代时,使用有标签数据 $L = \{(x_1, y_1), \dots, (x_{n_l}, y_{n_l})\}$ 、从伪标签数据 P^k 和映射标签数据 I^k 共同训练模型.特别地,映射标签数据是没有被赋予伪标签的无标注数据,将对应的数据索引作为标签,从而生成映射标签数据:

$$U = \{x_{n_l+1}, x_{n_l+2}, \dots, x_{n_l+u}\} = P^k + I^k$$

利用标注数据和选调出的伪标签数据进行 ID 分类学习.但对于映射标签数据而言,被赋予了对应索引值的映射标签,显然将映射标签数据加入到 ID 分类训练中会损害模型.为了充分利用这些映射标签数据,我们采纳文献^[12]提出的差异性损失,利用索引标注数据来优化卷积神经网络模型:

$$l_e(\mathbf{V}; \tilde{\phi}(\theta; x_i)) = -\log \frac{\exp\left(\frac{v_i^T \tilde{\phi}(\theta; x_i)}{\tau}\right)}{\sum_{j=1}^{|M|} \exp\left(\frac{v_j^T \tilde{\phi}(\theta; x_i)}{\tau}\right)}$$

其中, V 作为一个查询列表存储了所有的映射标注数据的特征 x_i , 通过这样的设置, 避免了在每次迭代时都要重复从所有的数据中提取特征的复杂运算. τ 控制样本概率分布. 在前向传播过程中, 计算 x_i 与其他所有数据的余弦相似度, 在反向传播过程中, 对 v_i 进行更新:

$$v_i \leftarrow \frac{1}{2}(v_i + \tilde{\phi}(\theta; x_i))$$

并通过 L_2 归一化, 将其转化为单位向量后传入 V 中, 以实现对其第 i 列元素的更新. 通过这种差异性损失, 模型能够学习区分不同输入图片而不是学习分类 ID, 从而使得学习到的表征会获得更多关于输入 ID 的信息. 在此过程中, 尽管这些样本没有任何的标注, 但是模型可以获取更多的样本, 并且这些样本可以通过学习不同行人图片间的差异性来为模型提供微弱的监督信息. 对于有正确标签 ID 的数据集 L , 依照^[10]方法来训练 re-id 模型, 以下是目标方程:

$$\min_{\theta, \omega} \sum_{i=1}^{n_l} l_{CE}(f(\omega; \phi(\theta; x_i)), y_i)$$

其中, $f(\omega; \cdot)$ 表示 ID 分类器, 这个分类器能够将提取出的特征分类到 K 维的置信估计中 (K 表示 ID 数量), l_{CE} 表示预测标签与真实 ID 标签之间的交叉熵损失. 同样的, 可以采用相类似的方式利用伪标签数据集 P 来优化模型:

$$\min_{\theta, \omega} \sum_{i=n_l+1}^{n_l+n_u} s_i l_{CE}(f(\omega; \phi(\theta; x_i)), y_i)$$

对于伪标签数据的使用, 这里的方法与有标签数据的方法基本类似, 不同的是在这里加入一个判别器 $S_i \in \{0, 1\}$, 判别器的值由上一步的变迁估计决定, 判别器将决定所挑选出的伪标注数据能否加入到 ID 分类训练中. 考虑到训练数据有三部分组成, 第一部分是在标注数据集 L 的 ID 分类损失, 第二部分是用于抽选出的伪标签数据 P^k 的 ID 分类损失, 第三部分则是用于映射标注数据集 I^k 的差异性损失. 采用以下的目标方程来进行迭代:

$$\begin{aligned} & \min_{\theta, \omega} c \sum_{i=1}^{n_l} s_i l_{CE}(f(\omega; \phi(\theta; x_i)), y_i) \\ & + c \sum_{i=n_l+1}^{n_l+n_u} s_i^{k-1} l_{CE}(f(\omega; \phi(\theta; x_i)), y_i) \\ & + (1-c) \sum_{i=n_l+1}^{n_l+n_u} (1-s_i^{k-1}) l_e(f(V; \tilde{\phi}(\theta; x_i))) \end{aligned}$$

k 表示第 k 次迭代, c 为调整交叉熵损失与差异性损失所占权重的参数.

2.2 抽样策略与标签估计

在解决单样本类型的任务时, 分类器容易对单样本数据过拟合, 从而进一步影响了分类器对于 ID 分类的鲁棒性. 为了解决过拟合的问题, 着眼于提高选调伪标签精度的方式, 从获取伪标签的数据中挑选出更优的伪标签数据, 将它们加入到 ID 分类的训练中, 因此, 选择一个合适的抽样策略显得十分关键.

本方法采用特征空间上的距离作为衡量伪标签的方式. 对于无标签数据的标签评估, 使用最邻近分类器, 最邻近分类器会在特征空间上为每一个无标签数据查找与其距离最近的有标签数据, 将查找到的有标签数据的标签作为伪标签赋给无标签. 特别地, 在特征空间上的距离会作为标签评估的置信度, 从伪标签数据中进行选调. 为了充分获得并利用这些数据, 在每一轮的迭代中, 从每个模型中选调的伪标签数据 N^k 的规模是不断扩大. 具体的, 通过以下方法为无标签数据进行标签估计:

$$\begin{cases} d(\theta; x_i) = \|\phi(\theta; x_i) - \phi(\theta; x')\| \\ x', y' = \arg \min_{(x_i, y_i) \in L} \|\phi(\theta; x_i) - \phi(\theta; x')\| \\ y_i = y' \end{cases}$$

$d(\theta; x_i)$ 看作是提取特征向量的不相似代价, 用于估计样本标签. 通过上一步, 可以从两个模型中各自挑选出 N^k 个置信度较高的伪标签数据. 但是, 这样的伪标签的数据标签精确度仍然有限, 为此, 将两个模型挑选出的 $2N^k$ 个伪标签数据共同进行进一步的筛选, 将特征索引相同的伪标签数据挑选出来, 从而完成了为标签数据精度的再一次提升:

$$P^t = P^t_{\phi_1} \cap P^t_{\phi_2}$$

2.3 总体迭代策略

通过上面提出抽样方法, 充分保证了所选调的伪标签数据的可靠性. 在最初阶段, 模型训练仅仅使用很小一部分的伪标签数据样本, 随着迭代的进行, 再逐渐扩充伪标签样本的比例. 特别地, 在第一次迭代开始时, 所有的无标签数据均没有伪标签, 模型的优化通过 (1) 在有标签数据上, 进行 ID 分类训练; (2) 对所有的无标签数据, 通过差异性损失来进行无监督训练. 在之后的迭代过程中, 逐渐增加选调伪标签数据的规模, $N_t =$

$N_t + \alpha \cdot n_u$, $\alpha \in (0, 1)$ 作为迭代过程中控制扩大选调伪标速度的参数, α 参数决定了训练代价与模型精度, 一个较小的参数, 意味着更多的迭代次数以及训练时间, 但是, 这样会获得更加可靠的伪标签数据, 从而获得更加精确的模型. 本方法采用多模型共同训练, 选调的伪标签数据为各自模型中估计不相似代价最小的 N' 个伪标签, 这样可以保障被赋予伪标签的数据具有相对较高的可靠性, 因此, 本方法在迭代的过程中会选择个相对较大的参数 α , 从而平衡了模型精度与训练代价.

算法 1. 模型优化算法

- (1) 初始化两个不同的 CNN 模型, 抽调伪标签数据 $P_0 \leftarrow 0$, 设置抽调样本扩充因数 α ;
- (2) 将两个不同的 CNN 模型同时进行训练, 在迭代进行过程中, 通过 3 组数据优化模型参数;
- (3) 对无标签数据进行伪标签评估;
- (4) 通过两个模型, 选取较优伪标签数据;
- (5) 更新抽样规模;
- (6) 评估模型阶段, 通过在验证集上的表现, 挑选出两个不同的最优模型;
- (7) 更新下次迭代使用的伪标签数据和映射标签数据.

3 实验与分析

3.1 实验数据与评价标准

本文所使用的数据集, 是当前数据量较大的行人重识别数据集: MARS^[13]. 该数据集总共包含了 1261 个行人 ID, 分别在 6 个摄像头下进行收集. 其中训练集总计包含了 625 个行人 ID, 8298 小段行人轨迹, 共 509 914 张图片. 测试集包含 636 个行人 ID, 12 180 小段行人轨迹. MARS 数据集是提供序列信息的大规模行人重识别数据集.

当前对行人重识别识别率的标准主要有以下两种:
(a) rank- n : 给定查询集中的一张照片, 计算它与图库中所有图片的相似度, 从高到低排列, 如果前 n 项中有相同的行人, 则识别正确. 在这样的评价标准中, rank-1 的准确率最为重要, rank-5 与 rank-10 最为辅助参考
(b) mAP: 强调的是查全率与查准率之间的平衡, 也是先对相似度进行排序, 然后从高到低统计从第一项到最后一项相同行人图片间正确识别的比率.

需要说明的是, 同一摄像头连续拍摄的两幅图像中, 同一行人往往是比较相似的, 评价标准中排除了所有这些情况, 只统计一个摄像头中的行人在其他摄像头拍摄的图像中查询到的结果, 一定程度抑制了数据

本身不平衡可能对结果造成的影响.

3.2 实验环境及参数配置

本算法使用深度学习框架 Pytorch 实现, 在配有 1080tiGPU 上进行实验模型训练和测试评估. 对于模型选择, 使用 InceptionV3^[14]和在 Imagenet 上预训练过的 Resnet50^[15]两个模型.

3.3 实验比较

对于不同的选调伪标签扩展因数 α 会影响伪标签数据的可靠性, 因此, 在实验进行的过程中, 使用不同的 α , 以充分了解其对于模型的影响.

从表 1 中可以看到选择一个较小的 α 使得模型性能更优, 由于扩充因数 α 决定了标签更新速度, 因此选择较小的 α 会增加时间代价.

表 1 不同样本扩充因数 α 下模型性能

因数 α	rank-1	mAP
0.3	47.8	23.5
0.2	53.0	31.2
0.1	60.2	37.7
0.05	65.3	45.6

为了验证方法的有效性, 使用 Mars 数据集, 分别与当前基于监督式、半监督式的行人重识别方法进行比较.

从表 2 中可以看出: 通过本文提出的行人重识别方法, 在单样本设置下, 模型 rank-1=65.3, mAP=45.6 这是一个不逊于基于监督学习方法的效果. 与当前先进的行人重识别方法相比较, 本文提出的方法更少地依赖标签数据, 进一步提升行人重识别的现实应用价值. 当然, 单样本的设置会使得模型优化远不及多样本, 且当模型的样本输入从单样本输入设置为多样本输入时, 总体的识别率会得到显著提升, 比如: 当标注率提升为 40%, rank-1=83.8, mAP=64.9. 实验结果表明本文提出的方法的有效性.

表 2 与当前先进算法的性能比较

方法	样本标注率	rank-1	mAP
DGM+IDE ^[16]	单样本	36.8	16.9
EUG ^[10]	单样本	62.67	42.45
MSCAN ^[15]	100%	71.8	56.0
DCML ^[16]	100%	72.93	56.88
Ours	单样本	65.3	45.6

4 结论

本文提出的方法中, 对于每个 ID 训练仅只提供一

个有正确标签的样本,在这样的设置下,模型的学习会受到限制。为了应对单样本设置下行人重识别任务,设计了一种多个卷积神经网络联合训练的方法。对于每个卷积神经网络,并且采用了渐进式训练框架,使用标签数据、伪标签数据、映射标签数据联合训练的方式。在标签评估阶段,利用两个模型联合选调出较为精确的伪标签数据,对伪标签数据集进行更新。虽然我们的方法呈现出了一个可以与多样本设置相媲美的效果,但是,解决单样本设置下的行人重识别任务还面临诸多挑战,未来仍有很多工作有待去解决。

参考文献

- 1 Zheng WS, Gong SG, Xiang T. Person re-identification by probabilistic relative distance comparison. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence, RI, USA. 2011. 649–656.
- 2 Chen YC, Zheng WS, Lai JH, *et al.* An asymmetric distance model for cross-view feature mapping in person reidentification. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 27(8): 1661–1675. [doi: [10.1109/TCSVT.2016.2515309](https://doi.org/10.1109/TCSVT.2016.2515309)]
- 3 Gong S, Cristani M, Yan S. Person re-identification. Computer Vision and Pattern Recognition, 2014, 42(7): 301–303.
- 4 Li DW, Chen XT, Zhang Z, *et al.* Learning deep context-aware features over body and latent parts for person re-identification. The IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 384–393.
- 5 Wojke N, Bewley A. Deep cosine metric learning for person re-identification. IEEE Winter Conference on Applications of Computer Vision. Lake Tahoe, NV, USA. 2018. 748–756.
- 6 Huang Y, Xu JS, Wu Q, *et al.* Multi-pseudo regularized label for generated data in person re-identification. IEEE Transactions on Image Processing, 2019, 28(3): 1391–1403. [doi: [10.1109/TIP.2018.2874715](https://doi.org/10.1109/TIP.2018.2874715)]
- 7 Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv: 1609.02907, 2017.
- 8 Liu X, Song ML, Tao DC, *et al.* Semi-supervised coupled dictionary learning for person re-identification. IEEE Conference on Computer Vision and Pattern Recognition, 2014: 3550–3557.
- 9 Ding GD, Zhang SS, Khan S, *et al.* Center based pseudo-labeling for semi-supervised person re-identification. IEEE International Conference on Multimedia & Expo Workshops. San Diego, CA, USA. 2018. 1–6.
- 10 Li JW, Ma AJ, Yuen PC. Semi-supervised region metric learning for person re-identification. International Journal of Computer Vision, 2018, 126(8): 855–874. [doi: [10.1007/s11263-018-1075-5](https://doi.org/10.1007/s11263-018-1075-5)]
- 11 Chen F, Chai JH, Ren DH, *et al.* Semi-supervised distance metric learning for person re-identification. 2017 IEEE/ACIS 16th International Conference on Computer and Information Science. Wuhan, China. 2017. 733–738.
- 12 Wu Y, Li YT, Dong XY, *et al.* Exploit the unknown gradually: One-shot video-based person re-identification by stepwise learning. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 5177–5186.
- 13 Zheng L, Bie Z, Sun YF, *et al.* MARS: A video benchmark for large-scale person re-identification. European Conference on Computer Vision. Cham, 2016: 868–884.
- 14 Szegedy C, Vanhoucke V, Ioffe S, *et al.* Rethinking the inception architecture for computer vision. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2818–2826.
- 15 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 770–778.
- 16 Ye M, Ma AJ, Zheng L, *et al.* Dynamic label graph matching for unsupervised video re-identification. 2017 IEEE International Conference on Computer Vision. Venice, Italy. 2017. 5142–5150.