

摘要

行人再识别问题在智慧服务、行为分析、智能安防等方向应用前景广泛，近年来发展迅速，备受瞩目。然而，在现实应用场景中、尤其是将模型部署到全新的无标注数据域时，训练或微调一个行人再识别模型往往要花费大量人工成本来标记数据。这极大地阻碍了现有技术的推广与应用。即使目前已提出许多半监督或无监督行人再识别方法，它们依然标记成本高，或者性能有限。

主动学习是机器学习领域减少标注成本、提升模型学习效率、促进模型与人类互助的一类高效方法。为了缓解行人再识别现实应用时标注繁琐这一问题，本文研究“基于主动学习的少量标注行人再识别方法”，以减少标注代价的同时使行人再识别模型性能令人满意为目的，从样本选择策略的约束建模与改进两个方面递进地选择更有标注价值的样本，使人工标注与模型训练高效协同互助。主要内容有：

第一，为了提升标注样本对的价值，提出了一个多样性约束下的有价值标注样本对选择方法。该方法从全局视角出发，将该任务中样本的成对人工标注形式、样本对选择的策略准则以及模型的分类训练模式这三者结合，兼顾标注样本对的有效性与多样性。具体来说，该方法将样本对的整体选择统一建模成一个优化公式，并与基于聚类——伪标签的基线过程相结合，将标注数量与样本总数解耦，有助于进一步降低标注次数。优化公式由有效性及多样性两种准则组成，联合约束标注样本对的选择。使用两种距离分别衡量聚类簇内与簇间样本对的有效性，确保标注结果对模型训练有良性影响。同时，该方法选择伪标签更多样的标注样本对，计算标注员对每个聚类簇的关注情况作为多样性准则，提高模型对不同行人外观的学习和适应程度。使用贪心法对优化公式进行求解，将对最优值增量最小的样本对逐个选择并人工标注，与迭代训练的基线模型流程匹配。在公开数据集 Market1501 上的实验证明了本方法的有效性。该方法对不同设置下的基线模型都有稳定提升，接近方法上界；同时，相比于其他先进方法，在仅使用 20% 标注成本的前提下，mAP 提升 1.9%。

第二，为了使标注结果能更好地影响样本对选择策略、动态地弥补模型学习的不足，提出了有效性准则优化的标注样本对选择方法。分析发现，模型对标注样本对的学习滞缓时，标注样本簇内伪标签有真实身份混乱的可能。该方法根据样本对的学习结果衡量簇内标签学习的滞缓程度，优化聚类簇内有效性准则，对滞缓程度较大的聚类簇加强约束，增加关注度。同时，使用行人部件计算新的聚类簇间有效性准则，并与基于全局特征的有效性准则交叉使用，使模型对遮挡和未对齐更加鲁棒。在普适性强的基础数据集 Market1501 和更有难度的遮挡数据集 Occluded-Duke 上，相比优化前

mAP 分别提升了 0.6% 与 2.5%，证明了该方法对标注样本对选择策略的有效优化。

第三，为了方便主动学习模式在少标注行人再识别现实场景的运行，降低标注开销，开发了一个面向行人再识别的少量标注与训练辅助工具。该工具包含模型管理、数据管理、任务管理等多块功能，并重点设有人在回路的快速标注与训练任务，辅以完全同步、部分同步两种标注交互模式。工具集成了本文中的样本对选择方法，用户可以在上传无标注数据并运行任务后，以人机交互的形式标注少量标签并训练，从而获得一个适用于新域的性能较好的行人再识别模型。该工具可以很好地推动行人再识别模型在新无标注数据域的快速使用。

综上，本文对于基于主动学习的少量标注行人再识别问题进行研究，从样本对选择策略的建模与优化角度，提出了多样性约束下的有价值样本对的选择方法，并基于学习滞缓程度和行人部件对选择策略的有效性准则进行优化，在有限标注次数的约束下提高了模型性能。本文还就此开发了一个主动学习的行人再识别少量标注与训练辅助工具，促进了该方法对少标注行人再识别问题现实应用场景的帮助。