



研究与开发

基于多任务学习的行人重识别算法研究

秘蓉新¹, 姚文文¹, 吴兵灏²

(1. 国家计算机网络应急技术处理协调中心, 北京 100029;

2. 北京邮电大学计算机学院, 北京 100876)

摘要: 行人重识别 (person re-identification, re-ID) 在多摄像机之间进行跨镜检索以匹配目标行人图像, 可以在人脸、指纹等生物特征失效的情况下实现行人关联, 已成为智能视频监控系统的核心技术, 对智能安防、智慧城市等领域的产业落地进行了有效赋能。传统的行人重识别算法通常采用表征学习或度量学习方法。基于多任务学习的机器学习模式, 结合表征学习与度量学习方法, 综合利用特征表示和距离度量两方面的优势, 采用分类损失和三元组损失共同训练模型, 使模型在特征提取和相似性度量上都得到充分的训练。实验结果表明, 该方法在行人重识别任务中取得了更好的性能, 验证了鲁棒性和在泛化能力方面的优越性。

关键词: 行人重识别; 智能视频监控; 表征学习; 度量学习; 多任务学习

中图分类号: TN929.53

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2024157

Research on person re-identification algorithm based on multi-task learning

MI Rongxin¹, YAO Wenwen¹, WU Binghao²

1. National Computer Network Emergency Response Technical Team/Coordination Center of China,
Beijing 100029, China

2. School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract: Person re-identification (re-ID) involves the cross-camera retrieval and matching of target pedestrian images, facilitating pedestrian association in scenarios where biometric features such as faces and fingerprints may prove ineffective. It has become a pivotal technology in intelligent video surveillance systems, playing a crucial role in domains like smart security and smart cities. Traditional re-ID algorithms typically employ either representation learning or metric learning methods. A novel approach was proposed which combined representation learning and metric learning methods based on the multi-task learning machine learning paradigm. By capitalizing on the advantages of both feature representation and distance metric, and concurrently training the model using classification loss and triplet loss, comprehensive training for both feature extraction and similarity measurement was ensured. Experimental results validate the effectiveness of the proposed approach, demonstrating superior performance in re-ID tasks

收稿日期: 2024-03-27; 修回日期: 2024-05-10

通信作者: 姚文文, yww@cert.org.cn



and underscoring the robustness and superior generalization capability.

Key words: person re-identification, intelligent video surveillance, representation learning, metric learning, multi-task learning

0 引言

近年来,我国监控摄像机的安装数量持续增长,视频监控系统在智能安防和智慧城市等领域发挥着不可替代的关键作用。面对庞大的监控视频数据,随着人工智能研究热潮的再度兴起和与计算机视觉相关技术的迅速发展,图像与视频中的目标检测、跟踪、分类和识别等技术为监控视频的智能化处理与应用提供了崭新的机遇。

传统的人员识别主要依赖于人脸识别和指纹识别技术。然而,随着应用需求的不断提升和应用场景的不断扩大,传统的人员识别技术因为需要目标人员的配合而受到极大的应用限制。例如,人脸识别技术在摄像机分辨率过低、拍摄视角变化、光照不佳以及帽子或其他配饰遮挡等情况下难以准确提取目标人员的正面脸部特征信息,人脸识别技术难以发挥作用。为了解决这些问题,行人重识别(person re-identification, re-ID)技术应运而生。

行人重识别旨在解决跨摄像机行人的识别与检索问题,其任务目标是判断不同摄像机捕捉到的不同图像中的行人是否为同一个人,行人重识别的任务目标如图1所示。在无法获得目标人员清晰高质量的正面人脸图像的场景中,行人重识别技术能够对目标行人进行跨摄像机连续跟踪,持续确认行人身份,目标人员无接触、无配合、无感觉。re-ID已经成为计算机视觉领域中继人脸识别技术之后又一个重要且热门的研究方向。

传统的行人重识别算法通常采用表征学习或度量学习方法。基于表征学习的方法不直接学习图像之间的相似性,而是通过神经网络根据任务需求自动提取表示特征,常用于分类任务。基于

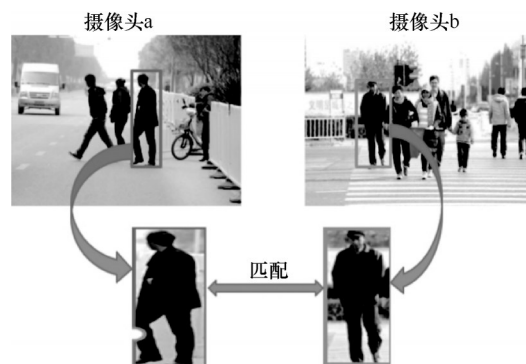


图1 行人重识别的任务目标

度量学习的方法通过神经网络学习任意两张行人图像之间的相似度,从而使得同一行人的不同图像之间具有较高的相似度,不同行人的图像之间具有较低的相似度,常用于聚类任务。通过表征学习方法,模型可以学习到数据特征的有效表示;通过度量学习方法,模型可以学习到适合特定任务的距离度量。结合表征学习和度量学习,可以综合利用特征表示和距离度量两方面的优势,使模型在特征提取和相似性度量上都得到充分的训练。本文采用多任务学习的范式,将表征学习和度量学习方法统一起来,以此达到更好的重识别性能。

本文的主要贡献总结如下。

(1) 基于多任务学习的机器学习模式,结合表征学习和度量学习方法,在同一个端到端模型中将分类任务与聚类任务统一起来,同时使用分类损失和三元组损失对网络进行训练,使模型同时具有较优秀的特征提取和相似性度量能力。

(2) 通过一个参数可学习的全连接层将提取出的全局特征线性变换得到另一个全局特征,这两个特征分别应用于表征学习和度量学习的损失函数计算,消除分类任务与聚类任务之间的学习目标不一致性,使得模型的训练过程更加稳定;

(3) 与其他主流方法的行人重识别模型的对比实验, 说明了本文所提基于多任务学习的行人重识别方法在 Market-1501 数据集和 DukeMTMC-reID 数据集上取得更好的性能, 验证了该方法的鲁棒性和在泛化能力方面的优越性。

1 国内外研究成果

目前国内外行人重识别研究成果可大致分为以下 6 类方法。

(1) 姿态引导方法^[1-3]

该方法利用行人图像中的姿态信息来改善重识别性能, 需要进行额外的人体姿态检测以获得行人在图像中的关键点位置和身体姿势等信息。其优势在于利用了图像中丰富的姿态信息, 使得模型能够更好地理解行人特征, 提高了对不同姿态行人的识别能力。这种方法在处理姿态变化较大的行人图像时尤其有效。

(2) 遮罩引导方法^[4-5]

该方法利用行人图像中的人体部位遮罩信息来改善重识别性能, 需要进行额外的人体部位分割以获得不同部位的遮罩信息, 如头部、上身、下身和鞋子等。其优势在于利用了图像中人体部位的详细信息, 使得模型能够更好地理解人体特征, 提高了对不同人体部位变化的鲁棒性。这种方法在处理不同姿态、发生遮挡情况下的行人图像时非常有效。

(3) 基于纹理的方法^[6-10]

该方法利用图像中行人的服装纹理信息来辅助行人重识别任务, 需要对行人图像进行服装分割以提取纹理特征, 然后进行纹理匹配或者与其他特征相融合。其优势在于利用了行人图像中的服装纹理信息, 这些信息通常能够提供额外的辨识度, 特别是在其他特征不够清晰或者出现部分人体遮挡的情况下。

(4) 基于注意力的方法^[11-13]

该方法利用注意力机制来动态地学习行人图

像中不同区域的重要性, 以此提高模型对行人图像中重要区域的关注程度。其优势在于能够自动地学习图像中重要区域的表示, 并根据任务需求动态地调整注意力, 使模型能够更好地适应不同场景下的行人图像, 提高了重识别的泛化性和鲁棒性。

(5) 基于生成对抗网络 (generative adversarial network, GAN) 的方法^[14-16]

该方法利用生成器网络生成行人图像, 同时使用判别器网络对真实的行人图像与生成的行人图像进行区分。生成器和判别器之间进行对抗性训练, 训练生成器尽量生成逼真的行人图像以欺骗判别器, 训练判别器区分真实和合成的图像。这个过程持续进行直到达到平衡, 生成器可以生成足够逼真的行人图像。这种方法提供了额外的数据增强手段, 通过生成器和判别器的联合对抗性训练, 可以生成更加逼真的行人图像, 帮助模型更好地学习行人的视觉特征, 从而提高重识别性能。

(6) 全局特征方法^[17-20]

该方法将整个行人图像视为一个整体, 从中提取出全局特征进行重识别。全局特征能够捕捉到整个行人图像的信息, 因此在一些情况下, 这种方法能够取得很好的效果。前文所提到的表征学习方法和度量学习方法都属于全局特征方法。

2 传统行人重识别算法

2.1 行人重识别问题描述

给定一组图库图像及其对应的身份标签 $\mathcal{G} = \{(x_g, y_g)\} \mid g = 1, 2, \dots, N_g\}$, 可以将行人重识别问题形式化表述为: 在大量候选图像中为查询图像 x_q 选择相关图像的身份标签。具体来说, 候选集 $R(x_q, k)$ 确定了在特征空间中与查询图像 x_q 相似度最高的图像, 即与查询图像 x_q 最有可能是同一行人的图像。候选集 $R(x_q, k)$ 定义如下:

$$R(x_q, k) = \{(x_g, y_g) \mid y_g \in N(x_q, k)\} \quad (1)$$



其中, $N(x_q, k)$ 表示与 x_q 最相似的 k 个样本所对应的标签集合。定义 $E(x)$ 为输入图像 x 映射到特征空间 $\mathbb{R}^d$ 的特征抽取过程, 行人重识别的目标就是寻找一种从行人图像到行人身份的映射。

2.2 基于表征学习的行人重识别算法

基于表征学习的行人重识别方法^[17,21-24]具有两种任务范式: 将行人重识别看作分类任务, 利用行人身份ID等信息作为真值标签, 通过监督学习的方式来训练神经网络; 将行人重识别看作验证任务, 输入两张行人图片, 让神经网络来判断这两张图片中的行人是否为同一个人, 其本质属于二分类任务。

IDE^[17]利用分类损失 (classification loss) 和验证损失 (verification loss) 来训练网络, 每次接收两张行人图像作为输入, 包括分类子网络和验证子网络两部分, 基于表征学习的行人重识别方法IDE模型示意图如图2所示。

分类子网络对图像进行行人身份ID预测, 使用输出的预测身份ID和该图像对应的真值身份ID来计算分类误差损失 (交叉熵损失)。由于网络是对ID标签进行分类, 常称该分类损失为ID损失 (ID loss), 计算式如下。

$$L_{ID} = -\frac{1}{N_b} \sum_j y_j \ln(p(y_j | E(x_j))) \quad (2)$$

其中, N_b 为一个批量 (batch) 中的图片数量, $p(\cdot)$ 表示样本 x_j 属于身份ID标签 y_j 的预测概率。

验证子网络将输入的两张图像的特征融合成一个特征, 根据融合后的这个特征判断这两张图片中的行人是否为同一行人, 其计算式如下:

$$L_{\text{verif}} = -\frac{1}{N_b} \sum_j \sum_{i=1}^2 q_i \ln(p(q_i | f_j \oplus f_{j+1})) \quad (3)$$

其中, q_i 表示是否属于同一行人, 若该图片对属于同一行人则 $q_1=1, q_2=0$, 否则 $q_1=0, q_2=1$, 即二分类标签, $f_x = E(x)$, 表示输入图片 x 经过特征抽取网络后的特征向量, $\oplus$ 表示特征融合, IDE中使用的特征融合方法为按元素做差取平方, 即 $f_1 \oplus f_2 = (f_1 - f_2)^2$ 。

2.3 基于度量学习的行人重识别算法

基于度量学习的行人重识别方法^[19,25-30]通过神经网络学习任意两张行人图像之间的相似度, 其目标为同一行人的不同图像 (正样本对) 之间的相似度尽可能高, 不同行人的不同图像 (负样本对) 之间的相似度尽可能低。

行人重识别任务中最经典也最常使用的度量学习损失为三元组损失 (triplet loss)^[19]。该损失函数接收3张图片作为输入, 这3张图片被称为一个三元组, 记为 (a, p, n) , 其中包含一对正样本 (a, p) 和一对负样本 (a, n) 。三元组损失使得正样本

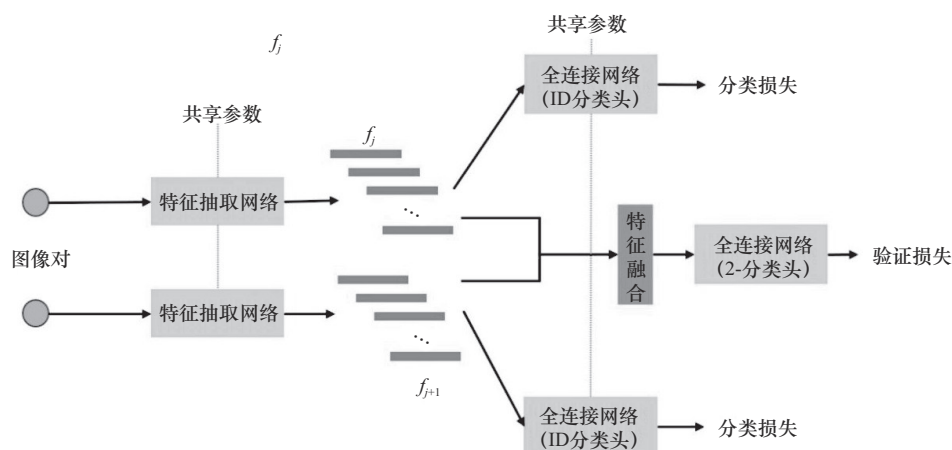


图2 基于表征学习的行人重识别方法IDE模型示意图

对之间的特征距离比负样本对之间的特征距离更接近。三元组损失的计算方法如下。

$$L_{\text{tri}} = \frac{1}{N_b} \sum_j \max [D(f_a, f_p) - D(f_a, f_n) + m, 0] \quad (4)$$

其中, $D(f_1, f_2): \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ 为距离函数, 用于计算两个特征向量在特征空间中的相似度; m 为超参数, 表示距离阈值。通常使用欧几里得距离的平方作为特征向量之间的相似度距离, 即:

$$D(f_1, f_2) = (f_1 - f_2)^T (f_1 - f_2) \quad (5)$$

3 基于多任务学习的行人重识别算法

本文基于多任务学习的机器学习模式, 将表征学习和度量学习两种方法有机地融合到同一个神经网络中, 以同时完成行人重识别中的分类和聚类两种不同任务。在这一模式下, 表征学习方法利用分类损失函数 L_{ID} 来训练神经网络, 使其能够准确地划分不同身份ID的行人图像样本; 而度量学习方法则通过三元组损失函数 L_{tri} 对神经网络进行训练, 以实现相同身份ID的行人图像样本之间的紧密聚类效果。值得注意的是, 本文通过在同一个神经网络中同时使用分类损失函数 L_{ID} 和三元组损失函数 L_{tri} 进行训练, 将两种任务有机地结合在一起。在这一总体模型结构下, 分类损失函数 L_{ID} 主要负责学习分类任务的边界, 而三元组损失函数 L_{tri} 则致力于度量样本之间的相似度, 从而实现了行人重识别任务的全面学习和优化。总损失函数如下。

$$L_{\text{total}} = L_{\text{ID}} + L_{\text{tri}} \quad (6)$$

模型整体架构包含输入层、特征提取层、特

征融合层、多任务学习层和输出层, 基于多任务学习的行人重识别算法模型示意图如图3所示。在输入层中, 模型接收裁剪好的行人图像作为输入, 这些图像来自不同的摄像机或拍摄场景。特征提取层使用在 ImageNet 上预训练的 ResNet50 作为特征提取器, 从输入的行人图像中提取丰富的特征表示。特征融合层使用全局平均池化将所提取的特征进行融合, 以获得全局信息表征。多任务学习层包含一个全连接网络, 使用线性变换前后的两个全局信息表征分别学习分类和聚类两种不同的任务。输出层最终输出行人表征和行人识别的结果。

3.1 特征提取骨干网络

行人重识别, 关键在于训练出一种能够提取高度可靠的行人特征的神经网络。ResNet^[31]是一种典型的深度卷积神经网络模型, 其独特的网络结构设计具有出色的特征提取能力, 在 ImageNet 数据集上展现出卓越的性能。为了能够在模型性能与资源消耗(如显存占用、计算时间等训练成本)之间取得较好的平衡, 本文使用 50 层的 ResNet 系列模型, 即 ResNet50 作为行人重识别算法的特征提取骨干网络。

较高的特征空间分辨率往往意味着提取的特征包含更丰富的语义信息。Sun 等^[32]通过删除 ResNet50 骨干网络中最后一次下采样操作来增加输出特征图的大小。在原始的 ResNet50 模型中, 输入一张大小为 256×128 (单位为像素, 下同) 的图像样本, 骨干网络最终输出的特征图尺寸为 8×4 。本文对 ResNet50 原模型的骨干网络进行了调整, 将最后一次空间下采样操作的步幅由 2 改

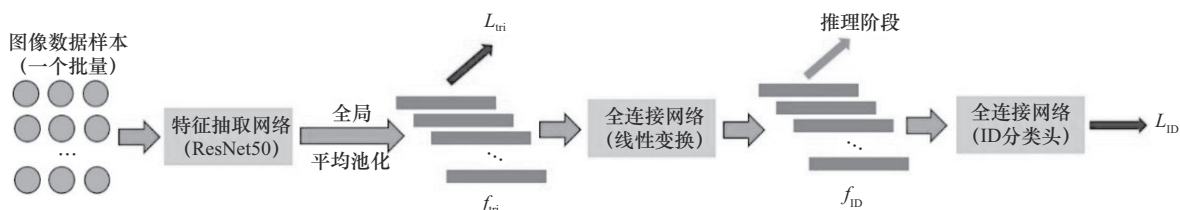


图3 基于多任务学习的行人重识别算法模型示意图



为1, 这样最终输出的特征图尺寸变为 16×8 。相较于原始模型, 修改后的ResNet50骨干网络输出的特征图尺寸增加4倍, 其中蕴含更丰富的语义信息, 使得神经网络能够更加关注行人身份信息的微妙和深层次特征。同时, 由于只需要ResNet50的特征提取部分, 因此将原模型中用于对ImageNet数据集图像进行分类的全连接层删除, 并通过全局平均池化操作输出行人图像样本的全局特征。

3.2 消除多任务间的不一致性

多任务学习是深度学习领域中一种强大的技术, 它允许模型同时学习多个不同的任务, 从而能够更全面地理解数据并提升性能。然而, 这种方法成功的关键在于如何有效地协调和管理各个任务之间的学习目标。当各任务之间的优化目标不一致时, 模型的训练过程可能会变得复杂, 甚至导致学习到的特征表示出现矛盾。

度量学习和表征学习就是两种具有不同优化目标的学习方法。度量学习主要关注样本之间的相似性或距离, 而表征学习则更侧重于从数据中提取有意义的特征表示, 这种差异可能导致它们在学习过程中追求不同的目标。例如, 分类损失 L_{ID} 并没有直接让网络学习两张图片之间的相似性, 它旨在学习能够区分不同类别标签的样本之间的分界面, 而三元组损失 L_{tri} 则侧重于学习样本之间的相对关系, 具有聚类效果, 更多地关注样本之间的相似度。

由于这些目标上的差异, 在多任务学习中使用同一个全局特征来计算这两种损失, 可能会导致模型在训练过程中产生冲突, 无法同时满足两种损失的要求。因此, 为了确保模型在训练阶段的学习目标一致, 本文提出了一种策略: 使用两个不同的全局特征来分别计算分类损失 L_{ID} 和三元组损失 L_{tri} 。这样做的好处是, 每个损失都可以针对其特定的优化目标进行独立计算和优化, 避免了目标之间的冲突。通过这种方式, 可以确保

模型在训练过程中能够同时兼顾分类和度量学习的目标, 学习到更加一致和有效的特征表示。这不仅有助于提高模型的性能, 还能使其在处理多任务问题时更加灵活和准确。

具体而言, 利用修改后的ResNet50骨干网络来提取特征, 提取出的特征 f_{tri} 被用作相似性度量的基础, 这一特征专门用于计算三元组损失 L_{tri} , 它能够帮助模型学习样本之间的相对关系, 进而实现聚类效果。接下来, 为了计算分类损失 L_{ID} , 并没有直接使用 f_{tri} , 而是将其输入一个特殊设计的全连接神经网络 (fully connected neural network, FC) 中。其特点是输入维度与输出维度相等, 称为多任务学习层。通过这一层, 得到了另一个特征 f_{ID} , 它专门用于身份ID分类、计算分类损失 L_{ID} 。这种方式允许针对不同学习任务进行更精细化的特征调整, 从而更好地满足多任务学习的网络训练模式要求。

f_{tri} 和 f_{ID} 之间的全连接神经网络实际上是一个参数可学习的线性变换。在训练过程中, 模型能够自动学习从 f_{tri} 到 f_{ID} 的映射关系。这种设计不仅增加了模型的灵活性, 还使得模型能够更好地协调分类损失 L_{ID} 和三元组损失 L_{tri} 之间的优化目标, 有效地缓解了多任务学习中不同损失函数之间的学习目标不一致性, 从而提高了模型的性能和稳定性。

通过采用两个不同的全局特征 f_{tri} 和 f_{ID} 来分别计算三元组损失 L_{tri} 和分类损失 L_{ID} , 并结合一个参数可学习的线性变换来协调两者之间的关系, 成功构建了一个能够同时处理度量学习和表征学习任务的多任务学习模型。这一模型不仅能够充分利用不同任务之间的互补性, 还能够有效地解决目标不一致性所带来的挑战, 为实际应用提供有力的支持。由于行人重识别任务的最终目标是寻找一种从行人图像到行人身份的映射, 因此在推理阶段, 模型使用网络自动学习到的更适

合身份分类的特征 f_{ID} 进行图像样本之间的相似度计算。

4 实验结果与分析

4.1 实验设置

模型的搭建基于深度学习框架PyTorch。所有输入图像转换为 $3 \times 256 \times 128$ （通道数 $\times$ 高度 $\times$ 宽度）的张量，批量大小（batchsize）设置为32，每个批量中包含4个不同的行人ID，每个行人ID包含8张图像。使用Adam优化器^[33]更新网络中的可学习参数，初始学习率设置为0.000 2， $(\beta_1, \beta_2) = (0.9, 0.999)$ 。式(4)中的距离阈值超参数 m 设置为0.9。在训练阶段，对原始输入图像进行随机擦除^[34]，将随机擦除的概率设置为0.5，随机擦除矩形区域的大小范围设置为0.02~0.4，随机擦除矩形区域的纵横比范围设置为0.3~3.33。在推理阶段不做额外的数据增强处理。

4.2 数据集

本文以行人重识别领域经典数据集Market-1501数据集^[35]和DukeMTMC-reID数据集^[36]作为实验基准数据集。

Market-1501数据集是清华大学在2015年发布的行人重识别数据集。行人图片采集自清华大学校园的6个摄像机（其中有5个高清摄像机和1个低分辨率摄像机），一共标注了1 501个不同身份的行人，通过DPM检测器生成行人检测框，共32 668张图像。训练集中包含12 936张行人图像，对应751个不同身份的行人；测试集中包含19 732张图像，对应750个不同身份的行人。另外，该数据集分别从6个摄像机对应的测试集中为每个行人选取一张图像构成查询图像集，包含3 368张query图片，这些图片的行人检测框由人工裁剪得到，以保证检索目标的正确性。之后，为了模拟大规模检索，加入了约50万幅干扰图像，包含失败的检测框、不在测试集中的行人

等，使该数据集更贴近真实场景。

DukeMTMC-reID数据集是Duke大学在2017年发布的行人重识别数据集。行人图片由8个高清摄像机采集，包括1 812个不同身份的行人，共36 411张图片，其中至少在2个以上摄像机中出现的行人有1 404人，只在1个摄像机中出现的行人有408人。训练集中包含来自出现在多摄像机中的702个行人的16 522张图片，测试集中包含出现在多摄像机中的另外702个行人和只出现在单摄像机中的408个行人的17 661张图片，查询图像集包含2 228张图像。

4.3 评估指标

根据先前的其他行人重识别研究工作，本文使用Rank- N 和精度均值（mean average precision, mAP）这两项评测指标对模型进行性能评估。

Rank- N 指根据查询图像与所有图库图像之间的特征距离进行升序排序后，检索结果的前 N 幅图像中包含正确目标行人图像的概率，概率越大表示重识别算法的效果越好。Rank- N 是一个最直观的指标，常用的有Rank-1、Rank-5、Rank-10和Rank-20，分别代表在第1幅、前5幅、前10幅和前20幅检索图像中包含正确目标行人图像的概率。

mAP对所有类别的精确率-召回率曲线（precision recall curve, PRC）下的面积（area under the curve, AUC）进行平均，综合考虑模型在不同类别上的精确率和召回率，更全面地反映了模型的性能和鲁棒性。

4.4 实验结果

本节将本文所提出的基于多任务学习的行人重识别算法与本文第1节中所介绍的其他类型的行人重识别经典算法进行比较，模型测试在Market-1501数据集和DukeMTMC-reID数据集上进行，以Rank-1和mAP作为性能评估标准，本文模型与其他行人重识别模型对比见表1。



表1 本文模型与其他行人重识别模型对比

方法类型	模型	Market-1501		DukeMTMC-reID	
		Rank-1	mAP	Rank-1	mAP
姿态引导	PIE ^[2]	87.7	69.0	79.8	62.0
遮罩引导	MaskReID ^[5]	90.0	75.3	78.8	61.9
基于纹理	AlignedReID ^[6]	90.6	77.7	81.2	67.4
	SCPNet ^[7]	91.2	75.2	80.3	62.6
基于注意力	HA-CNN ^[13]	91.2	75.7	80.5	63.8
基于GAN	Camstyle ^[14]	88.1	68.7	75.3	53.5
	PN-GAN ^[15]	89.4	72.6	73.6	53.2
全局特征	IDE ^[17]	79.5	59.9	65.2	45.0
	SVDNet ^[18]	82.3	62.1	76.7	56.8
	TriNet ^[19]	84.9	69.1	72.4	53.5
	AWTL ^[20]	89.5	75.7	79.8	63.4
	本文模型	91.3	82.9	83.5	73.8

本文所提出的基于多任务学习的行人重识别模型相较于纯粹的表征学习模型 IDE 和纯粹的度量学习模型 TriNet 取得了显著的性能提升。在 Market-1501 数据集上, 本文模型相较于 IDE 的 Rank-1 指标和 mAP 指标分别提升了 11.8% 和 23%, 相较于 TriNet 的 Rank-1 指标和 mAP 指标分别提升了 6.4% 和 13.8%; 在 DukeMTMC-reID 数据集上, 本文模型相较于 IDE 的 Rank-1 指标和 mAP 指标分别提升了 18.3% 和 28.8%, 相较于 TriNet 的 Rank-1 指标和 mAP 指标分别提升了 11.1% 和 20.3%。实验结果进一步验证了本文所提出的基于多任务学习的行人重识别模型在准确性和鲁棒性方面的优越性。

5 结束语

本文基于多任务学习的机器学习模式, 结合表征学习和度量学习方法, 通过一个参数可学习的全连接层消除表征学习和度量学习的任务不一致性, 同时使用分类损失和三元组损失共同训练模型, 综合利用特征表示和距离度量两方面的优势, 使模型同时具备优秀的特征提取和相似性度量能力。实验结果表明, 该方法在行人重识别任

务中取得了更好的性能, 验证了该方法的鲁棒性和在泛化能力方面的优越性。

参考文献:

- [1] WEI L H, ZHANG S L, YAO H T, et al. Glad: Global-local-alignment descriptor for pedestrian retrieval[C]//Proceedings of the 25th ACM international conference on Multimedia. New York: ACM Press, 2017: 420 - 428.
- [2] ZHENG L, HUANG Y J, LU H C, et al. Pose invariant embedding for deep person re-identification[EB]. 2017: 1701.07732.
- [3] SARFRAZ M S, SCHUMANN A, EBERLE A, et al. A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 420-429.
- [4] KALAYEH M M, BASARAN E, GÖKMEN M, et al. Human semantic parsing for person re-identification[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 1062-1071.
- [5] QI L, HUO J, WANG L, et al. MaskReID: a mask based deep ranking neural network for person re-identification[EB]. 2018: 1804.03864.
- [6] ZHANG X, LUO H, FAN X, et al. AlignedReID: surpassing human-level performance in person re-identification[EB]. 2017: 1711.08184.

- [7] FAN X, LUO H, ZHANG X, et al. SCPNet: spatial-channel parallelism network for joint holistic and partial person re-identification[EB]. 2018: 1810.06996.
- [8] SUN Y F, ZHENG L, YANG Y, et al. Beyond part models: person retrieval with refined part pooling (and A strong convolutional baseline)[C]//European Conference on Computer Vision. Cham: Springer, 2018: 501-518.
- [9] ZHENG F, SUN X, JIANG X Y, et al. A coarse-to-fine pyramidal model for person re-identification via multi-loss dynamic training[EB]. 2018: 1810.12193.
- [10] DAI Z Z, CHEN M Q, ZHU S Y, et al. Batch feature erasing for person re-identification and beyond[EB]. 2018: 1811.07130.
- [11] WANG C, ZHANG Q, HUANG C, et al. Manes: A multi-task attentional network with curriculum sampling for person re-identification[C]//European Conference on Computer Vision. Cham: Springer, 2018: 384-400.
- [12] SI J L, ZHANG H G, LI C G, et al. Dual attention matching network for context-aware feature sequence based person re-identification[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 5363-5372.
- [13] LI W, ZHU X T, GONG S G. Harmonious attention network for person re-identification[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 2285-2294.
- [14] ZHONG Z, ZHENG L, ZHENG Z D, et al. CamStyle: a novel data augmentation method for person re-identification[J]. IEEE Transactions on Image Processing, 2019, 28(3): 1176-1190.
- [15] QIAN X L, FU Y W, XIANG T, et al. Posenormalized image generation for person re-identification[C]//Proceedings of the European Conference on Computer Vision (ECCV). [S.l.: s.n.], 2018: 650 - 667.
- [16] 郑声晟, 殷海兵, 黄晓峰, 等. 基于GAN的无监督域自适应行人重识别[J]. 电信科学, 2021, 37(2): 99-106.
ZHENG S C, YIN H B, HUANG X F, et al. GAN-based unsupervised domain adaptive person re-identification[J]. Telecommunications Science, 2021, 37(2): 99-106.
- [17] ZHENG Z D, ZHENG L, YANG Y. A discriminatively learned CNN embedding for person re-identification[EB]. 2016: 1611.05666.
- [18] SUN Y F, ZHENG L g, DENG W J, et al. Svdnet for pedestrian retrieval[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 3800-3808.
- [19] HERMANS A, BEYER L, LEIBE B, et al. In defense of the triplet loss for person re-identification[EB]. 2017: 1703.07737.
- [20] RISTANI E, TOMASI C. Features for multi-target multi-camera tracking and re-identification[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 6036-6046.
- [21] GENG M Y, WANG Y W, XIANG T, et al. Deep transfer learning for person re-identification[EB]. 2016: 1611.05244.
- [22] LIN Y T, ZHENG L, ZHENG Z D, et al. Improving person re-identification by attribute and identity learning[EB]. 2017: 1703.07220.
- [23] ZHENG L, YANG Y, HAUPTMANN A G, et al. Person re-identification: past, present and future[EB]. 2016: 1610.02984.
- [24] MATSUKAWA T, SUZUKI E. Person re-identification using CNN features learned from combination of attributes[C]//Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR). Piscataway: IEEE Press, 2016: 2428-2433.
- [25] VARIOR R R, HALOI M, WANG G. Gated Siamese convolutional neural network architecture for human re-identification [C]//European Conference on Computer Vision. Cham: Springer, 2016: 791-808.
- [26] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: a unified embedding for face recognition and clustering[C]//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2015: 815-823.
- [27] LIU H, FENG J S, QI M B, et al. End-to-end comparative attention networks for person re-identification[J]. IEEE Transactions on Image Processing, 2017, 26(7): 3492-3506.
- [28] CHENG D, GONG Y H, ZHOU S P, et al. Person re-identification by multi-channel parts-based CNN with improved triplet loss function[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 1335-1344.
- [29] CHEN W H, CHEN X T, ZHANG J G, et al. Beyond triplet loss: a deep quadruplet network for person re-identification [EB]. 2017: 1704.01719.
- [30] XIAO Q Q, LUO H, ZHANG C. Margin sample mining loss: a deep learning based method for person re-identification[EB]. 2017: 1710.00478.



- [31] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 770-778.
- [32] SUN Y F, ZHENG L, YANG Y, et al. Beyond part models: person retrieval with refined part pooling (and A strong convolutional baseline)[C]//European Conference on Computer Vision. Cham: Springer, 2018: 501-518.
- [33] KINGMA D P, BA J. Adam: a method for stochastic optimization[EB]. 2014: 1412.6980.
- [34] ZHONG Z, ZHENG L, KANG G L, et al. Random erasing data augmentation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 13001-13008.
- [35] ZHENG L, SHEN L Y, TIAN L, et al. Scalable person re-identification: a benchmark[C]//Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2015: 1116-1124.
- [36] RISTANI E, SOLERA F, ZOU R, et al. Performance measures and a data set for multi-target, multi-camera tracking[C]//European Conference on Computer Vision. Cham: Springer, 2016: 17-35.

[作者简介]



秘蓉新 (1984-), 女, 国家计算机网络应急技术处理协调中心助理研究员, 主要研究方向为人工智能、网络安全、信息安全。



姚文文 (1984-), 女, 国家计算机网络应急技术处理协调中心助理研究员, 主要研究方向为网络安全战略、信息科技发展态势。



吴兵灏 (2001-), 男, 北京邮电大学计算机学院 (国家示范性软件学院) 硕士生, 主要研究方向为深度学习、计算机视觉与行人重识别。