

结合一阶和二阶空间信息的行人重识别

刘莎¹, 党建武^{1,2*}, 王松^{1,2}, 王阳萍^{2,3}¹兰州交通大学电子与信息工程学院, 甘肃 兰州 730070;²甘肃省人工智能与图形图像处理工程研究中心, 甘肃 兰州 730070;³兰州交通大学计算机科学与技术国家级实验教学示范中心, 甘肃 兰州 730070

摘要 针对行人重识别中行人检测误差引起的空间错位, 基于局部的深度网络模型仅学习相邻局部关系, 导致远距离局部相关性缺失, 因此, 提出了一种结合一阶和二阶空间信息的行人重识别算法。在主干网络上, 学习一阶空间掩模对输入图像的空间权值进行微调, 以减少背景干扰; 通过二阶空间掩模对远距离的依赖关系进行建模, 并将局部特征集成到依赖模型中, 以获取全局特征表示。局部分支引入 DropBlock 对抽取的行人特征进行正则化, 避免了网络模型过于依赖特定部位特征。训练阶段用标签平滑分类损失和引入正样本中心的三元组损失联合优化整个网络。在 Market-1501 和 DukeMTMC-reID 数据集上的实验结果表明, 相比其他主流算法, 本算法的行人重识别精度更高, 且提取的行人特征判别性和鲁棒性更好。

关键词 机器视觉; 行人重识别; 一阶空间掩模; 二阶空间掩模; 中心三元组损失

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/LOP202158.0215005

Person Re-Identification Based on First-Order and Second-Order Spatial Information

Liu Sha¹, Dang Jianwu^{1,2*}, Wang Song^{1,2}, Wang Yangping^{2,3}¹ School of Electronics and Information Engineering, Lanzhou Jiaotong University, Lanzhou, Gansu 730070, China;² Gansu Provincial Engineering Research Center for Artificial Intelligence and Graphic & Image Processing, Lanzhou, Gansu 730070, China;³ National Experimental Teaching Demonstration Center on Computer Science and Technology, Lanzhou Jiaotong University, Lanzhou, Gansu 730070, China

Abstract In order to solve the problem of the spatial dislocation caused by person detection error in person re-identification, the local-based deep neural networks model only learn the adjacent local relationship, resulting in lack of long-distance local correlation. This paper proposes a person re-identification algorithm based on first-order and second-order spatial information. On the backbone network, first-order spatial mask is learned to fine-tune the spatial weight of the input image to reduce the background interference. The second-order spatial mask is used to model the long-distance dependency relationship, and local features are integrated into the dependency model to obtain the global feature representation. In the local branch, DropBlock is introduced to regularize the pedestrian features to avoid the network model relying too much on specific part features. In the training stage, the whole network is optimized by the label-smoothed cross-entropy loss and the triple loss with positive samples' center. Experimental results based on Market-1501 and DukeMTMC-reID data sets show that compared with other mainstream algorithms, the person re-identification accuracy of the algorithm is higher, and the extracted pedestrian features are more discriminative and robust.

收稿日期: 2020-06-05; 修回日期: 2020-07-09; 录用日期: 2020-07-20

基金项目: 甘肃省科技计划项目(18JR3RA104)、甘肃省教育厅科技项目(2017D-08)

* E-mail: dangjw@mail.lzjt.cn

Key words machine vision; person re-identification; first-order spatial mask; second-order spatial mask; center triplet loss

OCIS codes 150.0155; 100.4996

1 引言

行人重识别^[1]是利用计算机视觉技术对无重叠视角相机捕获的行人图像和特定行人图像进行特征提取,并根据距离度量方法判断是否为同一行人。随着人工智能化的推进,行人重识别技术在智能安防、无人超市、人机交互等领域有着广阔的应用前景,但光照、遮挡、分辨率、人体姿势、相机视角、背景和行人服饰等方面的变化较大,使行人重识别极具挑战性。

早期的行人重识别主要研究如何设计具有鲁棒性的手工视觉特征^[2-4]和寻找合适的投影空间以对特征进行相似度量^[4-6]。常见的手工视觉特征结合颜色直方图、纹理直方图对行人表征进行建模,得到的高维行人特征可通过一系列相似性度量约束寻找出一个线性投影空间,以更好地地区分内变化和类间差异。但传统方法设计手工特征时容易忽视一些视觉捕捉不到的信息,且距离度量过程不能很好地利用手工特征的高维信息,往往只能得到次优的重识别结果。因此,人们将深度学习技术应用到行人重识别任务中。不同于传统方法将行人重识别任务分为特征学习和距离度量两部分进行独立优化,深度学习利用多个非线性映射结构提取查询图像中的高级视觉语义,通过目标函数进行端到端的学习优化,将两个环节结合到一起。根据目标函数的优化,深度学习网络可自动从图像数据中学习更具有判别性的特征以适应行人重识别任务,且识别结果较好。Li 等^[7]提出了滤波器对神经网络(FPNN),率先将神经网络用于行人重识别领域,相比传统方法,识别性能得到了一定的提升。但单纯提取行人的全局特征会使深度网络忽略具有判别性的局部细节,在背景复杂、行人检测框不精准、有遮挡的情况下表现不佳。Zhao 等^[8]引入姿态估计模型检测 14 个人体关键点,并根据关键点划分身体部位,按自下而上的顺序进行特征融合,得到了较好的识别结果。毕晓君等^[9]引入 Deepcut 模型获得了 14 通道的关键点置信图,并根据关键点置信图进行粗略姿态和精细姿态建模,以降低视角变化对行人重识别的影响。但该方法需要引入预先训练好的姿态估计模型,会给行人重识别网络引入额外误差。Cheng 等^[10]将查

询图像切分为 4 个水平条带,分别提取 4 个局部特征和 1 个全局特征进行融合,用三元组损失优化整个网络。Sun 等^[11]设计了一个精炼部分池化(RPP)模块,重新分配水平条带边缘部分的异常值,以保持水平条带内信息的一致性。为了尽可能地减少水平切分对全局上下文信息连贯性的破坏,徐龙壮等^[12]通过金字塔池化获得多尺度信息,并通过特征融合技术使相邻区域形成语义连贯性,以提供更丰富的上下文信息。

针对大多数局部网络模型只侧重于学习相邻局部之间的相关性,忽略了远距离局部间关系的问题,基于已有研究工作^[13-15],本文提出了一种结合一阶和二阶空间信息的行人重识别网络模型。用卷积分支学习具有判别性的空间信息,一阶空间掩模为全局特征重新分配空间权重,以减少背景噪声对远距离相关性建模的干扰;二阶空间掩模对全局特征进行远距离依赖建模,并集成图像的局部信息。引入 DropBlock 对提取的行人特征进行正则化,避免网络模型过拟合的同时学习更具有判别性的局部特征。将正样本中心引入三元组损失中,使类间特征距离更大的同时保证类内样本特征具有更紧凑的空间分布。在训练阶段,用中心三元组损失联合标签平滑分类损失对整个网络进行优化。在 Market-1501 和 DukeMTMC-reID 数据集上的多次实验结果表明,相比其他算法,本网络模型抽取的特征更具判别性,且能有效提升行人重识别的精度。

2 理论与方法

本网络的框架如图 1 所示,整个网络由全局分支和局部分支组成。用残差网络 ResNet50^[16]作为骨干网络进行特征提取,在第 1、第 2 阶段嵌入一阶空间掩模,调整空间特征权值以降低背景噪声带来的干扰;在第 3 阶段嵌入二阶空间掩模,建立图像局部到局部,人体部位到部位的依赖模型。为了给二阶空间掩模提供足够多的细粒度信息,按照文献^[17]的方法对残差网络进行调整。删除了 ResNet50 第 3 阶段和第 4 阶段的下采样操作并使用扩张卷积^[18],将输出通道数增加一倍,使残差块抽取到的特征拥有更丰富的细节信息,即在输入图像尺寸为 384 pixel×128 pixel×3 的情况下(3 为通道

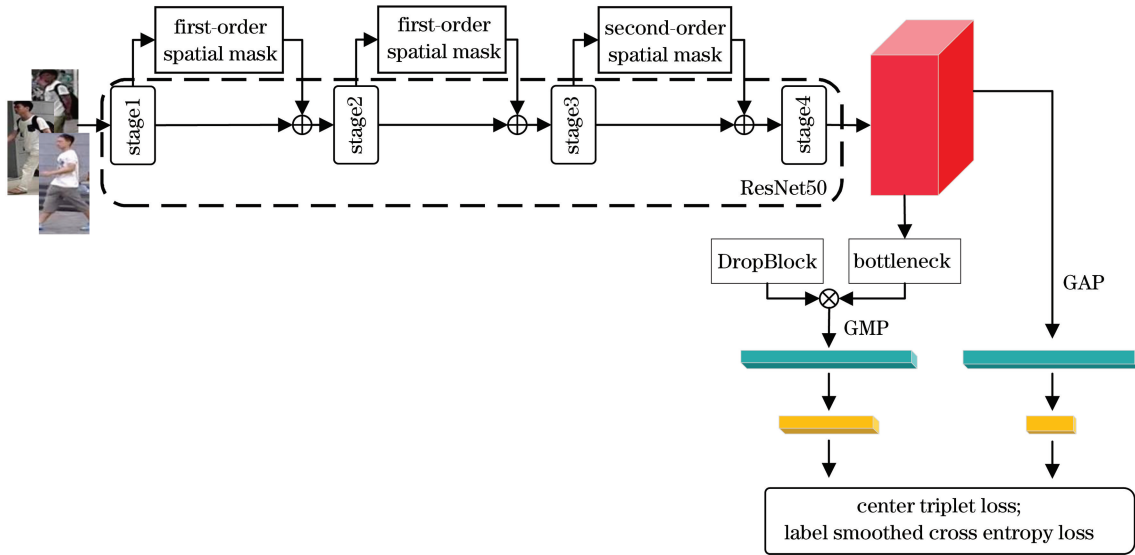


图 1 结合一阶和二阶空间信息的行人重识别网络框架

Fig. 1 Architecture of person re-identification network based on first-order and second-order spatial information

数), 输出尺寸为 $48 \times 16 \times 2048$ 的特征图。

全局分支将第 4 阶段输出的特征图, 经全局平均池化(GAP)层得到 2048 维的特征向量, 并由 1×1 卷积层、批归一化(BN)层和线性修正单元(ReLU)激活函数层组成的特征降维模块, 将特征向量降至 512 维作为全局特征表示, 由标签平滑分类损失和中心三元组损失进行优化。局部分支在第 4 阶段后附加了一个由卷积层和 BN 层、ReLU 层组成的 ResNet 瓶颈块^[16]。为了使网络模型能够学习到更多具有判别性的空间相关性, 减少对某个特定区域的依赖, 避免网络过拟合, 将 DropBlock^[19]应用于瓶颈块产生的特征图, 将其局部特征成块擦除。擦除后的特征图经全局最大池化(GMP)层得到 2048 维的局部特征向量, 用降维模块将其降至 1024 维, 由标签平滑分类损失和中心三元组损失进行优化。测试阶段将全局和局部分支的特征向量级联作为最终特征, 从而计算样本之间的距离。

2.1 一阶空间掩模

在真实监控场景下, 摄像头捕获的行人图像往往具有复杂的背景, 且背景区域可能与行人前景变化相似, 直接利用卷积特征进行二阶空间信息建模可能导致次优结果。因此, 用一阶空间掩模为图像特征重新分配空间权重, 有利于依赖关系获得更多的细节信息, 并抑制其他无用信息, 避免错误建模。

一阶空间掩模的结构如图 2 所示, 由跨通道 L1 范数操作、一个 3×3 的卷积层和一个 Sigmoid 层组成。给定一个残差块的输出特征图 $\mathbf{X} \in \mathbf{R}^{w \times h \times c}$ (w 、 h 、 c 分别为特征图的宽度、高度、通道数), 将特征图

视为沿通道轴的特征向量集合, 每个特征向量集成该空间位置的所有通道信息。沿着通道轴方向计算每个特征向量的 L1 范数, 经过卷积生成空间置信度得分, 并用 Sigmoid 激活函数对卷积结果进行归一化, 得到可对强调或抑制的空间位置进行编码的一阶空间掩模 $M_{FO}(\mathbf{X}) \in \mathbf{R}^{w \times h}$, 可表示为

$$M_{FO}(\mathbf{X}) = \sigma(\mathbf{W} \times \mathbf{X}_{L1-norm}), \quad (1)$$

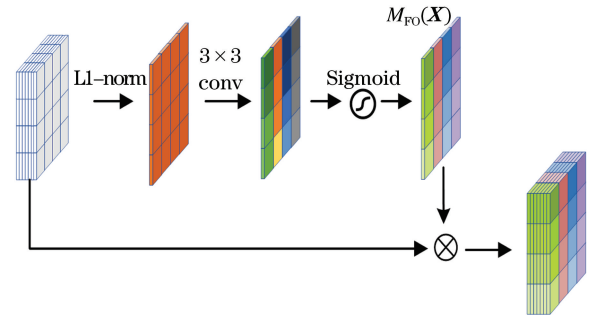


图 2 一阶空间掩模结构

Fig. 2 Structure of the first-order spatial mask

式中, $\mathbf{X}_{L1-norm}$ 为特征图 \mathbf{X} 通道轴上特征向量的 L1 范数, \mathbf{W} 为卷积层参数, σ 为 Sigmoid 函数。Sigmoid 函数将空间置信度得分归一化到 $(0, 1)$ 内, 空间掩模的不断堆叠, 可能会使深层网络的特征值越来越小, 为解决该问题, 用残差学习得到最终的一阶空间特征图 $F_{FOM}(\mathbf{X})$

$$F_{FOM}(\mathbf{X}) = [1 + M_{FO}(\mathbf{X})] \times \mathbf{X}. \quad (2)$$

将残差学习得到的一阶空间特征图 $F_{FOM}(\mathbf{X})$ 作为下一阶段的输入, 以减小变化趋势相似的背景噪声对后续远距离依赖关系建模的干扰。

2.2 二阶空间掩模

目前大多数行人重识别算法是训练深度学习模型对样本图像中的局部和全局、局部和局部之间的信息进行编码,以提取有效的行人特征。为了避免对样本图像进行硬切分引起的空间相关性割裂,二阶空间掩模^[15]将非局部操作与二阶统计量相结合,并利用非局部操作对特征图中位置的相关性进行建模,将卷积操作获取的局部信息集成到远距离依赖建模中,具体如图 3 所示。



图 3 二阶空间掩模的示意图

Fig. 3 Schematic diagram of the second-order spatial mask

二阶空间掩模结构如图 4 所示,给定输入特征图 $\mathbf{X} \in \mathbf{R}^{w \times h \times c}$, 将张量整形为 $w h \times c$ 的特征矩阵 \mathbf{G} ; 用 1×1 卷积、BN 层和 Leaky ReLU 层组成一个 $\theta(\mathbf{G})$ 函数; 用一个 1×1 卷积形成 $g(\mathbf{G})$ 函数。 $\theta(\mathbf{G})$ 、 $g(\mathbf{G})$ 函数将输入特征矩阵 \mathbf{G} 的通道 c 减少至 c/r , 其中, r 为降维因子, 然后通过 $\theta(\mathbf{G})$ 计算协方差矩阵 $\mathbf{\Omega}$ 并进行二阶池化, 以获得远距离依赖关系, 可表示为

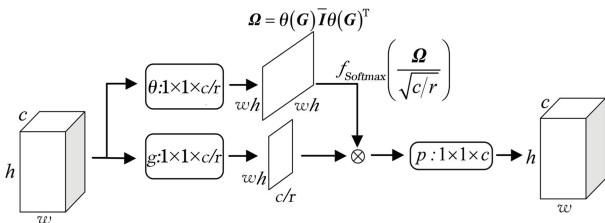


图 4 二阶空间掩模的结构

Fig. 4 Structure of the second-order spatial mask

$$\mathbf{\Omega} = \theta(\mathbf{G}) \bar{\mathbf{I}} \theta(\mathbf{G})^T, \quad (3)$$

式中, $\bar{\mathbf{I}} = \frac{1}{c/r} \left(\mathbf{I} - \frac{1}{c/r} \mathbf{1} \right)$, \mathbf{I} 和 $\mathbf{1}$ 分别为 $n \times n$ 的单位矩阵和所有元素均为 1 的矩阵。用 $1/\sqrt{c/r}$ 作为缩放因子对协方差矩阵进行缩放, 随后用 Softmax 函数对权值进行归一化。通过点积操作, 将局部特

征集成到远距离依赖关系中, 可表示为

$$\mathbf{z} = f_{\text{Softmax}} \left(\frac{\mathbf{\Omega}}{\sqrt{c/r}} \right) g(\mathbf{G}). \quad (4)$$

最后, 用一个 1×1 卷积作为可学习变换 p 将依赖特征矩阵的通道维数从 c/r 还原为 c , 经过适当的重塑, 得到 $w \times h \times c$ 的依赖特征张量

$$M_{\text{SO}}(\mathbf{G}) = p(\mathbf{z}). \quad (5)$$

通过残差学习得到的二阶空间特征可表示为

$$F_{\text{SOM}}(\mathbf{X}) = [\mathbf{1} + M_{\text{SO}}(\mathbf{X})] \times \mathbf{X}. \quad (6)$$

二阶空间特征 $F_{\text{SOM}}(\mathbf{X})$ 作为 ResNet50 下一阶段的输入, 在损失函数的优化下, 学习空间相关性信息, 以鉴别行人身份。

2.3 损失函数

实验用标签平滑正则化的交叉熵损失^[20-21]和改进后的三元组损失优化整个网络模型。交叉熵损失通过样本图像的分类预测值和其真实标签评估网络的性能, 可表示为

$$L_{\text{ide}} = - \sum_{i=1}^N \sum_{k=1}^K y_i^k \log(p_i^k), \quad (7)$$

式中, N 为每批次中包含的样本图像数, K 为每批次中包含的样本类别数, y_i^k 为第 i 张行人图像属于第 k 类的分类真值, p_i^k 为第 i 张行人图像属于第 k 类的预测概率。为了对模型进行正则化以获得更好的泛化能力, 在交叉熵损失函数中引入了标签平滑正则化^[20], 将均匀分布 $\mu(k) = 1/K$ 作为正则项, 将交叉熵损失重新表示为

$$L_{\text{ide}} = - \sum_{i=1}^N \sum_{k=1}^K [(1-\epsilon) y_i^k + \epsilon/K] \log(p_i^k), \quad (8)$$

式中, ϵ 为平滑因子。引入平滑标签正则化的交叉熵损失不仅考虑到了行人样本正确分类的损失, 也考虑了错误分类的损失, 从而提高模型的学习能力, 一定程度上避免模型过拟合。

为了更好地地区分内类间差异, 从输入图像中学习判别性特征, 首先将行人样本分为一系列三元组单元 $\{x_i, x_p, x_n\}$, 其中, x_i 为锚样本特征, (x_i, x_p) 为该组中标签相同 ($y_i = y_p$) 的正样本特征对, (x_i, x_n) 为标签不同 ($y_i \neq y_n$) 的负样本特征对。基本三元组损失函数^[22]可表示为

$$L_{\text{tri}} = \frac{1}{S} \sum_{i=1}^S [d(x_i, x_p) - d(x_i, x_n) + m]_+, \quad (9)$$

式中, $[\cdot]_+$ 为最大函数 $\max(0, \cdot)$, $d(z_1, z_2) = \|z_1 - z_2\|_2^2$ 为 z_1, z_2 特征之间的 L_2 距离, m 为间隔阈值, S 为每批次中的三元组数, p 为第 i 组三元组中的正样本, n 为第 i 组三元组中的负样本。

图 5(a)为标签平滑分类损失和基本三元组损失监督下的样本分布,图 5(b)为标签平滑分类损失和中心三元组损失监督下的样本分布。可以发现,引入正样本中心的三元组损失函数能更好地收紧类内样本分布,同时可使类间样本距离最大化。基本的三元组损失仅考虑了正样本特征对 (z_i, z_p) 和一个负样本特征对 (z_i, z_n) ,忽略了另一个负样本特征对 (z_p, z_n) ,使类内样本分布比较分散,也不能有效增大类间差异。为了有效增强类间距离的约束,引入正样本特征对 (z_i, z_p) 的中心点 c_{ip} 对三元组中的成对关系重新建模,可表示为

$$L_{tri} = \frac{1}{S} \sum_{i=1}^S [d(x_i, c_{ip}) + d(x_p, c_{ip}) - d(x_n, c_{ip}) + m]_+ \quad (10)$$

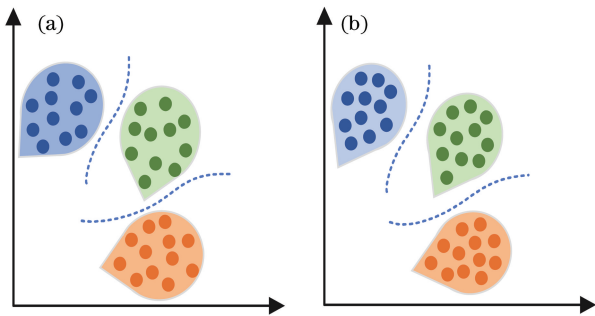


图 5 不同损失函数监督下的二维可视化样本分布。(a)标签平滑分类损失和基本三元组损失;(b)标签平滑分类损失和中心三元组损失

Fig. 5 Two-dimensional visualization of sample distribution supervised by different losses. (a) Label smoothed cross entropy loss and basic triplet loss; (b) label smoothed cross entropy loss and center triplet loss

3 实验分析

3.1 数据集及评价标准

为了验证结合一阶和二阶空间信息网络的有效性,在 Market-1501^[23]和 DukeMTMC-reID^[24]两个大规模的公开行人重识别数据集上进行了测试实验。

Market-1501 数据集在夏季的清华校园内进行采集,行人样本具有较大的姿态和光照变化,包含 6 个室外摄像头(5 个高分辨率摄像头和 1 个低分辨率摄像头)拍摄的 1501 个行人,共 32217 张图像。将 751 个行人的 12936 张图像作为训练集,将 750 个行人的 19732 张图像作为测试集。行人检测框由可变部件模型(DPM)算法和人工进行标注,更

符合真实应用场景。

DukeMTMC-reID 数据集在冬季的杜克大学校园内进行采集,行人样本具有较大的外观变化,包括 8 个室外摄像头拍摄到的 1812 个行人,共 36441 张图像,行人检测框由人工进行标注。由于只有 1404 个行人在超出两个摄像头中出现过,因此,随机采样 702 个行人的所有图像作为训练集,702 个行人的所有图像作为测试集,剩余 408 个人的所有图像作为干扰项加入测试集,可使数据集更具挑战性。

实验使用首位命中率(Rank-1)和平均精度均值(mAP)作为评估行人重识别算法在 Market-1501 和 DukeMTMC-reID 数据集上的性能。Rank-1 指标是将行人重识别作为图像检索问题进行处理,表示在图库图像中第一次检索即命中查询目标的概率。mAP 是平均精度(AP)的均值,可表示为

$$X_{AP} = \frac{\sum_{i=1}^r P(i) \delta(i)}{N} \quad (11)$$

$$X_{mAP} = \frac{\sum_{j=1}^M X_{AP}(j)}{M} \quad (12)$$

式中, r 为查询返回的图库图像总数, N 为返回图像中命中查询目标的图像数量, $P(i)$ 为查询返回图像数 i 的检索准确率,即 i 中命中查询目标样本数量与 i 的比值, M 为查询样本总数, $\delta(i)$ 为指示函数,当返回的第 i 张图库图像命中查询目标时, $\delta(i) = 1$;否则, $\delta(i) = 0$ 。

3.2 实验参数设置

实验环境:系统为 Ubuntu16.04,显卡为 NVIDIA Quadro P5000,显存为 16 GB,框架为 PyTorch,用 ImageNet 数据集^[25]上的预训练权重初始化主干网络 ResNet50。为了从每张图像中捕获更详细的信息,将所有输入图像的分辨率调整为 $384 \text{ pixel} \times 128 \text{ pixel}$,并通过水平翻转、归一化和裁剪对图像进行数据增强。设置 DropBlock 的擦除率 γ 为 0.1,擦除块的高度为 5,宽度为 8。标签平滑分类损失中的平滑因子 ϵ 为 0.1,中心三元组损失函数中的间隔阈值 m 为 0.3,批处理大小为 64,每批次中随机抽取 16 个行人,每个行人抽取 4 张图像。学习周期为 400,采用 Adam 优化算法更新参数,初始学习率为 e^{-3} ,在 200 个周期时衰减至 e^{-4} ,在 300 个周期时衰减到 e^{-5} 。

3.3 消融实验

为了验证本算法的有效性,在 Market-1501 和

DukeMTMC-reID 数据集上进行消融实验,结果如表 1 所示。其中,Baseline 为具有全局分支和随机擦除局部分支的基准网络,OM 为一阶空间掩模,SOM 为二阶空间掩模,S 为标签平滑正则化的交叉熵损失,BT 为基础三元组损失,CT 为引入正样本中心的三元组损失。实验 2 在实验 1 的基础上嵌入 SOM 来学习局部显著性特征之间的相关性,在 Market-1501 数据集上的 Rank-1、mAP 分别提高了 0.17、1.01 个百分点,在 DukeMTMC-reID 数据集上的 Rank-1、mAP 分别提高了 1.27、1.25 个百分点,这表明 SOM 能对身体部位相关性进行有效编码,提取具有判别性的行人特征。实验 3 在实验 2

的基础上加入 OM 减少依赖关系错误建模,在 Market-1501 数据集上的 Rank-1、mAP 分别提高了 0.41、0.27 个百分点,在 DukeMTMC-reID 数据集上的 Rank-1、mAP 分别提高了 0.53、0.31 个百分点,这表明 OM 的加入能减少背景对远距离依赖建模的干扰,使网络学习到更具判别性的信息。实验 4 在基础三元组损失上引入正样本中心作为距离度量锚点,相比实验 3 在 Market-1501 数据集上的 Rank-1、mAP 分别提高了 0.25、0.81 个百分点,在 DukeMTMC-reID 数据集上的 Rank-1、mAP 分别提高了 0.29、0.74 个百分点,这表明改进的三元组损失能使类内特征空间分布更紧凑。

表 1 不同模型的实验结果

Table 1 Experimental results of different models

unit: %

NO.	Model	Loss	Market-1501		DukeMTMC-reID	
			Rank-1	mAP	Rank-1	mAP
1	Baseline	S+BT	95.36	87.62	88.28	76.93
2	Baseline+ SOM	S+BT	95.53	88.63	89.55	78.18
3	Baseline+SOM+ OM	S+BT	95.94	88.90	90.08	78.49
4	Baseline+ SOM+ OM	S+CT	96.19	89.71	90.37	79.23

为了进一步探究一阶和二阶空间信息相结合的优越性,给出了部分样本的热力图,如图 6 所示,其中,圆点为参考点,阴影区域为与参考点高度相关的位置。可以发现,图 6(a)在单纯嵌入 SOM 的情况下,对局部特征相关性建模存在一定的误差,用头部

作为参考点时,背景中阶梯与该参考点的相关性较大;用背包作为参考点时,背景中绿植阴影部分与其相关性较大;用背景墙壁作为参考点时,行人、衣物与该参考点的相关性较大。图 6(b)加入 OM 对图像特征重新分配权值后,建模中的误差得到了抑制,并学习到了更详细、有效的局部依赖关系。这表明将 OM 和 SOM 相结合有利于获得依赖关系所需的更多细节信息,避免背景干扰导致的错误建模,提取更具判别性的行人特征。

3.4 实验结果可视化

为了更直观地展现算法的改进效果,将本算法与基准网络进行对比,并将实验结果可视化为图 7、图 8。序号 1~10 分别为算法检索图库返回的相似度排名前十的样本图像,从左到右相似度依次递减,矩形框为检索结果的行人身份与查询目标不一致。可视化示例选取背景与前景变化相似的样本,其中,图 7(a8)~图 7(a10)均为误检,可以发现,本算法降低了这些局部特征不相似的误检样本相似度,并正确检测出了背景变化较大和存在背面特征差异的样本,如图 7(b8)和图 7(b10)。这表明二阶空间掩模对局部显著特征相关性的建模有助于行人重识别精度的提升。相比基准网络,图 8(b6)、图 8(b7)样本在较大的背景变化时查询相似度有明显提升,这表明



图 6 样本热力图。(a)仅使用 SOM 得到的样本热力图;

(b)使用 OM 和 SOM 得到的样本热力图

Fig. 6 Sample heat map. (a) Sample heat map obtained using only SOM; (b) sample heat map obtained using OM and SOM



图 7 Market-1501 数据集的可视化结果。(a)基准网络算法;(b)本算法

Fig. 7 Visualization results of Market-1501 data set. (a) Baseline network algorithm; (b) our algorithm



图 8 DukeMTMC-reID 数据集的可视化结果。(a)基准网络算法;(b)本算法

Fig. 8 Visualization results of DukeMTMC-reID data set. (a) Baseline network algorithm; (b) our algorithm

本算法可减小背景的干扰。综上所述,在避免样本局部硬切分的前提下,本算法能减小背景干扰,有效地对样本局部与局部间的相关性进行正确建模,同时引入正样本中心修正梯度反向传播,减小了类内距离。

3.5 与主流算法的对比

表 2 和表 3 是本算法与其他主流行人重识别领域的深度学习算法在 Market-1501 和 DukeMTMC-reID 数据集上的识别结果。这些算法按照提取特

表 2 不同算法在 Market-1501 数据集上的识别结果

Table 2 Recognition results of different algorithms on the Market-1501 data set unit: %

Algorithm	Rank-1	mAP
GAN ^[26]	78.12	56.25
SVDNet ^[27]	82.30	62.10
AACN ^[28]	85.90	66.87
HA-CNN ^[14]	89.00	71.25
GLAD ^[29]	89.90	73.90
PCB+RPP ^[11]	93.80	81.60
MGN ^[30]	95.70	86.90
Pyramid-Net ^[31]	95.70	88.20
Our algorithm	96.19	89.71

征的不同可分为三类:以生成对抗网络(GAN)^[26]、奇异向量分解网络(SVDNet)^[27]为代表的全局特征算法;以注意感知合成网络(AACN)^[28]、和谐注意卷积神经网络(HA-CNN)^[14]、全局-局部对齐描述子(GLAD)^[29]、部分卷积基线网络+部分精炼池化策略(PCB+RPP)^[11]为代表的部位特征算法;以多粒度网络(MGN)^[30]、金字塔网络(Pyramid-Net)^[31]为代表的多尺度特征算法。

表 3 不同算法在 DukeMTMC-reID 数据集上的识别结果

Table 3 Recognition results of different algorithms on the DukeMTMC-reID data set unit: %

Algorithm	Rank-1	mAP
GAN ^[26]	67.68	47.13
SVDNet ^[27]	76.70	56.80
AACN ^[28]	76.84	58.25
HA-CNN ^[14]	78.50	60.25
GLAD ^[29]	80.00	62.20
PCB+RPP ^[11]	83.30	69.20
MGN ^[30]	88.70	78.40
Pyramid-Net ^[31]	89.00	79.00
Our algorithm	90.37	79.23

从表 2 可以发现,在 Market-1501 数据集上,本算法比 GAN 算法的 Rank-1 和 mAP 分别提高了 18.07、33.46 个百分点;比 AACN 算法的 Rank-1 和 mAP 分别提高了 10.29、22.84 个百分点;比 GLAD 算法的 Rank-1 和 mAP 分别提高了 6.29、15.81 个百分点;比 Pyramid-Net 算法的 Rank-1 和 mAP 分别提高了 0.49、1.51 个百分点。从表 3 可以发现,在 DukeMTMC-reID 数据集上,本算法比 SVDNet 算法的 Rank-1 和 mAP 分别提高了 13.67、22.43% 个百分点;比 HA-CNN 算法的 Rank-1 和 mAP 分别提高了 11.87、18.98 个百分点;比 PCB+RPP 算法的 Rank-1 和 mAP 分别提高了 7.07、10.03 个百分点;比 MGN 算法的 Rank-1 和 mAP 分别提高了 1.67、0.83 个百分点。这表明相比目前主流算法,本算法在识别精度上有明显提升。

4 结 论

提出了一种结合一阶和二阶空间信息的行人重识别网络,通过引入二阶空间掩模避免了样本局部硬切分带来的局部相关性缺失。针对背景噪声干扰局部特征远距离建模的问题,在二阶空间依赖建模前的特征提取阶段嵌入一阶空间掩模微调空间权重,抑制错误依赖关系建模。在基础三元组损失中引入正样本中心,更有效地约束类内距离,增大类间距离。在 Market-1501 和 DukeMTMC-reID 数据集上的实验结果表明,相比其他算法,本算法能提取具有鲁棒性的行人特征,有效提升行人重识别的精度。未来工作还需研究如何对行人运动的时空信息进行建模,进一步提高行人重识别的精度。

参 考 文 献

- [1] Yang F, Xu Y, Yin M X, et al. Review on deep learning-based pedestrian re-identification[J]. Journal of Computer Applications, 2020, 40(5): 1243-1252. 杨锋, 许玉, 尹梦晓, 等. 基于深度学习的行人重识别综述[J]. 计算机应用, 2020, 40(5): 1243-1252.
- [2] Liao S C, Hu Y, Zhu X Y, et al. Person re-identification by local maximal occurrence representation and metric learning[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 2197-2206.
- [3] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(7): 971-987.
- [4] Zhu X B, Che J. Person re-identification algorithm based on feature fusion and subspace learning[J]. Laser & Optoelectronics Progress, 2019, 56(2): 021503. 朱小波, 车进. 基于特征融合与子空间学习的行人重识别算法[J]. 激光与光电子学进展, 2019, 56(2): 021503.
- [5] Zhao R, Ouyang W L, Wang X G. Unsupervised saliency learning for person re-identification[C]//2013 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2013, Portland, OR, USA. New York: IEEE, 2013: 3586-3593.
- [6] Xiong F, Gou M R, Camps O, et al. Person re-identification using kernel-based metric learning methods[M]//Fleet D, Pajdla T, Schiele B, et al. Computer Vision-ECCV 2014. Lecture Notes in Computer Science. Cham: Springer, 2014, 8695: 1-16.
- [7] Li W, Zhao R, Xiao T, et al. DeepReID: deep filter pairing neural network for person re-identification[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE, 2014: 152-159.
- [8] Zhao H Y, Tian M Q, Sun S Y, et al. Spindle net: person re-identification with human body region guided feature decomposition and fusion[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 907-915.
- [9] Bi X J, Wang H. Person re-identification based on view information embedding[J]. Acta Optica Sinica, 2019, 39(6): 0615007. 毕晓君, 汪灏. 基于视角信息嵌入的行人重识别[J]. 光学学报, 2019, 39(6): 0615007.
- [10] 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]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 1335-1344.
- [11] Sun Y F, Zheng L, Yang Y, et al. Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline) [EB/OL]. [2020-05-23]. <https://arxiv.org/abs/1711.09349>.
- [12] Xu L Z, Peng L. Person reidentification based on multiscale convolutional feature fusion[J]. Laser &

- Optoelectronics Progress, 2019, 56(14): 141504.
徐龙壮, 彭力. 基于多尺度卷积特征融合的行人重识别[J]. 激光与光电子学进展, 2019, 56(14): 141504.
- [13] 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.
- [14] Li W, Zhu X T, Gong S G. Harmonious attention network for person re-identification[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 2285-2294.
- [15] Bryan B, Gong Y, Zhang Y Z, et al. Second-order non-local attention networks for person re-identification[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE, 2019: 3759-3768.
- [16] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 770-778.
- [17] Dai Z Z, Chen M Q, Zhu S Y, et al. Batch feature erasing for person re-identification and beyond [EB/OL]. [2020-05-21]. <https://arxiv.org/abs/1811.07130>.
- [18] Yu F, Koltun V, Funkhouser T. Dilated residual networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 636-644.
- [19] Ghiasi G, Lin T Y, Le Q V. Dropblock: a regularization method for convolutional networks [EB/OL]. [2020-05-24]. <https://arxiv.org/pdf/1810.12890>.
- [20] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 2818-2826.
- [21] He T, Zhang Z, Zhang H, et al. Bag of tricks for image classification with convolutional neural networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE, 2019: 558-567.
- [22] Hermans A, Beyer L, Leibe B. In defense of the triplet loss for person-re-identification [EB/OL]. [2020-05-24]. <https://arxiv.org/abs/1703.07737>.
- [23] Zheng L, Shen L Y, Tian L, et al. Scalable person re-identification: a benchmark[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015: 1116-1124.
- [24] Ristani E, Solera F, Zou R, et al. Performance measures and a data set for multi-target, multi-camera tracking [EB/OL]. [2020-05-20]. <https://arxiv.org/abs/1609.01775>.
- [25] Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database [C]//2009 IEEE Conference on Computer Vision and Pattern Recognition, June 20-25, 2009, Miami, FL, USA. New York: IEEE, 2009: 248-255.
- [26] Zheng Z D, Zheng L, Yang Y. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro [C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 3774-3782.
- [27] Sun Y F, Zheng L, Deng W J, et al. SVDNet for pedestrian retrieval [C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 3820-3828.
- [28] Xu J, Zhao R, Zhu F, et al. Attention-aware compositional network for person-re-identification [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 2119-2128.
- [29] Wei L H, Zhang S L, Yao H T, et al. GLAD: global-local-alignment descriptor for scalable person re-identification [J]. IEEE Transactions on Multimedia, 2019, 21(4): 986-999.
- [30] Wang G S, Yuan Y F, Chen X, et al. Learning discriminative features with multiple granularities for person re-identification [EB/OL]. [2020-05-21]. <https://arxiv.org/pdf/1804.01438.pdf>.
- [31] Zheng F, Deng C, Sun X, et al. Pyramidal person re-identification via multi-loss dynamic training [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE, 2019: 8506-8514.