

中图法分类号: TP391. ; 41 文献标识码: A 文章编号: 1006-8961(2023)05-1477-10

论文引用格式: Hou S H, Fu Y, Li A Q, Liu X, Cao C S and Huang Y Z. 2023. Multifaceted-features enhancement-relevant gait recognition method. Journal of Image and Graphics, 28(05):1477-1486(侯赛辉, 付杨, 李奥奇, 刘旭, 曹春水, 黄永祯. 2023. 多部位特征增强的步态识别算法. 中国图象图形学报, 28(05):1477-1486)[DOI:10.11834/jig.220641]

多部位特征增强的步态识别算法

侯赛辉^{1,2}, 付杨¹, 李奥奇¹, 刘旭², 曹春水², 黄永祯^{1,2*}

1. 北京师范大学人工智能学院, 北京 100875; 2. 银河水滴科技有限公司, 北京 100088

摘要: 目的 基于步态剪影的方法取得了很大的性能提升, 其中通过水平划分骨干网络的输出从而学习多身体部位特征的机制起到了重要作用。然而在这些方法对不同部位的特征都是以相对独立的方式进行提取, 不同部位之间缺乏交互, 有碍于识别准确率的进一步提高。针对这一问题, 本文提出了一个新模块用于增强步态识别中的多部位特征学习。**方法** 本文将“分离-共享”机制引入到步态识别的多部位特征学习过程中。分离机制允许每个部位学习自身独有的特征, 主要通过区域池化和独立权重的全连接层进行实现。共享机制允许不同部位的特征进行交互, 由特征归一化和特征重映射两部分组成。在共享机制中, 特征归一化不包含任何参数, 目的是使不同部位的特征具有相似的统计特性以便进行权值共享; 特征重映射则是通过全连接层或逐项乘积进行实现, 并且在不同部位之间共享权重。**结果** 实验在步态识别领域应用最广泛的数据集 CASIA-B(Institute of Automation, Chinese Academy of Sciences) 和 OUMVLP 上进行, 分别以 GaitSet 和 GaitPart 作为基线方法。实验结果表明, 本文设计的模块能够带来稳定的性能提升。在 CASIA-B 背包条件下, 本文提出的模块相对于 GaitSet 和 GaitPart 分别将 rank-1 的识别准确率提升了 1.62% 和 1.17%。**结论** 本文设计了一个新的模块用于增强步态识别的多部位特征学习过程, 能够在不显著增加计算代价的条件下带来稳定的性能提升。

关键词: 步态识别; 剪影序列; 多部位特征; 分离机制; 共享机制

Multifaceted-features enhancement-relevant gait recognition method

Hou Saihui^{1,2}, Fu Yang¹, Li Aoqi¹, Liu Xu², Cao Chunshui², Huang Yongzhen^{1,2*}

1. School of Artificial Intelligence, Beijing Normal University, Beijing 100875, China;

2. Watix Technology Limited Co., Ltd., Beijing 100088, China

Abstract: Objective Gait recognition can be focused on the identity labels of pedestrians-relevant recognition according to its walking style. Since it can be manipulated without coordination-derived constraints on a long distance scale, more applications potentials are illustrated for such domains like crime prevention, forensic identification, and public security. However, the process of gait recognition is challenged for many factors like camera views, carrying conditions, and different clothes. Current gait recognition tasks can be divided into two categories: model-based and appearance-based. Specifically, the model-based methods can be used to extract the human body structures for gait analysis. Conventional deep learning and graph convolutional network (GCN) based pose estimation is taken to extract gait features from the pose sequences in terms of hand-crafted features to model the walking process in common. First, model-based methods are robust to carrying and clothing theoretically, which is often challenged for human pose-precise low-resolution problems.

收稿日期: 2022-06-21; 修回日期: 2022-11-07; 预印本日期: 2022-11-14

* 通信作者: 黄永祯 huangyongzhen@bnu.edu.cn

基金项目: 中央高校基本科研业务费专项资金资助(2021NTST31)

Supported by: Fundamental Research Funds for the Central Universities (2021NTST31)

Second, the appearance-based methods are oriented to learn gait features in terms of human body structures-potential modeling. The silhouettes are mostly taken as the input, and these methods can be divided into three sub-categories further: template-based, sequence-based, and set-based. Specifically, template-based methods can be used to fuse the silhouettes of a gait circle into a template but the temporal information is sacrificed inevitably. The sequence-based methods can yield the silhouettes of a gait sequence as a video for spatio-temporal features extraction. And, the set-based methods can use the silhouettes of a gait sequence as an insequential set and the permutation invariant is added to the input order. Furthermore, multiple data for gait recognition are categorized into the appearance-based methods, including RGB frames, gray images, and optical flow. Compared to these data modalities and the pose sequences in the model-based methods, the silhouettes are easy to use, which are more suitable for the low-resolution scales. To be noticed, recent silhouettes-based methods for gait recognition can learn multi-part features through slicing the output of the backbone horizontally. However, multi features are extracted solely and the feature-interacted is lacked, which is likely to hinder the recognition accuracy. To resolve this problem, we design a new module to enhance the multifaceted feature learning for gait recognition. **Method** Silhouette-based gait recognition model consists of two parts: backbone-based, and multi-component feature learning. First, we design the backbone in term of the network structures in GaitSet and GaitPart, which can be as two popular methods for silhouette-based gait recognition. For the backbone-relevant, the features are first extracted for each silhouette (regular 2D convolution and max pooling in relevance to spatial dimension), and a set pooling is taken to aggregate the silhouette-level features in a non-sequential set (implemented by max pooling along the temporal dimension). Second, we design a new module for multiple-features learning and try to learn more robust and discriminative features for each motion. The independent-shared mechanism is introduced to learn motion-specific features, which is implemented by regional pooling and fully connected layers are separated. In particular, the interaction can be strengthened across various motions in terms of the coordinated mechanism, which consists of feature normalization and feature remapping. Feature normalization is parameter-free for weight balancing. And, feature remapping is implemented by a fully connected layer or element-wise multiple implecations. **Result** The experiments are carried out on Institute of Automation, Chinese Academy of Sciences (CASIA-B) and OUMVLP, and GaitSet GaitPart are as the baselines. The CASIA-B consists of 124 samples and collects the sequences of regular walking, such as walking with bags, and walking in different clothes for each object. The OUMVLP consists of 10 307 samples, which can collect the sequences of regular walking for each sample. Each sequence for CASIA-B and OUMVLP is recorded by 11 cameras and 14 cameras. GaitSet and GaitPart are commonly-used silhouettes methods as input for gait recognition. To learn the multifaceted features for gait recognition, GaitSet is regarded as an unsequential set and the features are sliced horizontally. To learn more specific features, GaitPart is focused on suppressing the receptive field of convolutional layers and modeling the micro-motion features. To demonstrate its consistency, the identical-view cases-excluded rank-1 accuracy is taken as the main metric for performance comparison. For example, each of rank-1 accuracy for walking with bags on CASIA-B can be optimized by 1.62% and 1.17% based on GaitSet and GaitPart. **Conclusion** A new module is facilitated to enhance the multi-components learning for gait recognition, which is cost-effective and the accuracy is improved in consistency. To be summarized, 1) the lack of interaction to hinder the recognition accuracy is concerned. 2) The independent-shared mechanism is introduced into multifaceted feature learning for gait recognition, and a plug-and-play module is designed to learn more discriminative features for multiple motions. 3) This GaitSet and GaitPart-based method has its potentials for consistent optimization over the baselines under all walking circumstances.

Key words: gait recognition; silhouette sequences; multi-part features; independent mechanism; shared mechanism

0 引言

步态识别的目标是通过被试者的行走方式对其身份进行识别。与指纹、人脸和虹膜等其他生物特征相比,步态具有远距离获取、非受控识别、不易伪

装以及适用范围广等特点,在刑侦破案、防疫流调等领域具有广阔的应用前景。步态的描述方法有多种形式,例如人体测量学数据、空间时间数据、运动学数据和动力学数据等(贲睨烨等,2012),当前最常用的是将人体关键点或剪影作为步态模型的输入,对应的相关研究方法大致可以分为基于模型的方法

和基于外观的方法两类。

基于模型的方法主要通过显式建模人体结构进行身份识别。例如, Liao 等人(2017)从原始视频中估计关键点, 然后结合人体先验知识进行识别; Li 等人(2021)利用 SMPL (skinned multi-person linear) 估计外观和姿态特征进行识别, 整个模型能够以端到端的方式进行训练; Teepe 等人(2021)利用图卷积神经网络从关键点序列提取步态特征; Li 和 Zhao (2022)将关于步态周期的先验知识引入到基于关键点的特征提取过程中。基于模型的方法理论上具有对携带物和着装的不变性, 但是在低分辨率条件下通常较难适用。

基于外观的方法通常不显式建模人体结构, 而是通过分析轮廓变化等信息进行步态识别。基于外观的方法大部分以步态剪影作为输入, 根据对待输入剪影的方式可以进一步划分为基于剪影模板的方法、基于剪影序列的方法、基于剪影集合的方法3个子类。其中, 基于剪影模板的方法将一个周期内的多幅剪影按照一定规则生成模板图像, 如 Han 和 Bhanu (2006) 将一个步态周期内的多幅剪影进行平均生成步态能量图。基于剪影序列的方法将输入剪影视为类似视频的有序序列提取空间和时序特征, 如 Lin 等人(2021)利用3D卷积网络在时间维度上提取局部和全局特征用于步态识别; Fan 等人(2020)通过限制卷积的感受野和编码局部运动信息学习更具有判别性的身体部位特征; Huang 等人(2022)将空间和时序的解耦操作融入到基于3D卷积的框架中, 并提取多尺度特征用于步态识别。基于剪影集合的方法将剪影序列视为无序集合提取步态特征, 如 Chao 等人(2019)首次建立基于剪影集合的步态识别模型, 该模型具有对输入剪影顺序的不变性; Hou 等人(2021)提出了面向集合的残差学习模块, 并融合多层特征增强步态模型的判别性。此外, 在近期的工作中, 王新年等人(2021)提出将无肩姿态能量图、步态参数等姿态特征与步态参数的2维傅里叶变换进行有效融合。徐硕等人(2022)为了在获得高准确率的同时对外观变化具有更好的鲁棒性, 提出了一种双分支网络融合外观特征和姿态特征, 以结合两种不同类型特征的优点。张红颖和包雯静(2022)针对目前基于生成式的识别方法采用特定视角的步态模板转换, 识别率随视角跨度增大而不断下降的问题, 提出融合自注意力机制的生成对抗网

络以进行跨视角步态识别。除了以剪影作为输入以外, 还有一些基于外观的方法以其他类型的数据作为输入, 如光流 (Bashir 等, 2009) 和灰度图 (Castro 等, 2020)。基于外观的方法相对更适用于低分辨率的条件, 因此在学术研究和实际应用中吸引了大量关注。

在上述工作中, 以基于剪影集合的 GaitSet (Chao 等, 2019) 和基于剪影序列的 GaitPart (Fan 等, 2020) 为代表的步态识别方法带来了显著的性能提升, 其中多部位特征学习机制起到了重要作用。这里以 GaitSet 为例对这一机制进行简要说明。为叙述方便, 本文用 F 表示骨干网络提取的步态特征, 其维度表示为 $H \times W \times C$, 其中 H 表示高度, W 表示宽度, C 表示通道维度。如图1所示, GaitSet 的多部位特征学习机制主要包含3个步骤。首先, 沿 H 维度将 F 均匀划分为 N 个部位; 然后, 沿空间维度(宽和高)利用最大池化和平均池化对每个部位的特征进行处理, 得到 N 个1维特征; 最后, 利用不共享权重的全连接层对每个部位的1维特征做进一步映射。上述机制能够使每个部位保留自己的运动特性, 从而能够显著提高步态识别的准确率。但值得注意的是, 在测试阶段通常是将不同部位的特征直接串联在一起作为整个序列的特征表示, 而在训练阶段, 除了共享骨干网络的权重以外, 不同部位的特征提取过程是相对独立的, 彼此之间缺乏交互, 这有碍于识别准确率的进一步提高。

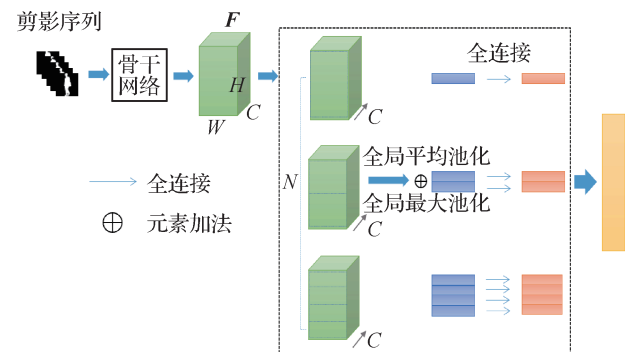


图1 GaitSet中的多部位特征提取

Fig. 1 Multi-part feature learning in GaitSet

针对这一问题, 本文设计了一个即插即用的模块, 通过引入“分离—共享”机制增强步态识别的多部位特征学习过程。分离机制同 GaitSet 中的多部位特征学习类似, 通过不同区域的池化操作和不共

享权重的全连接层允许每个部位学习自身独有的特征,这对于步态识别的过程十分关键。共享机制的引入是为了允许不同部位的特征进行交互,从而更好地相互协作。本文提供了共享机制的两种不同实现方式。共享机制由特征归一化和特征重映射两部分组成,这两个部分相辅相成。特征归一化不带有任何参数,目的是使不同部位的特征具有相似的统计特性(均值、方差等),从而便于进行权重共享;特征重映射可以通过全连接层或逐项乘积实现,在不同部位之间共享权重。本文设计的模块能够很方便地与 GaitSet 或 GaitPart 模型进行结合。在 CASIA-B(Institute of Automation, Chiese Academy of Sciences)和 OUMVLP 数据集上的实验表明,本文设计的模块能够在不显著增加计算代价的条件下带来稳定的性能提升。

综上所述,本文的主要贡献包括:1)分析了现有步态识别方法中广泛采用的多部位特征学习机制存在的不足,指出不同部位特征之间缺乏交互,有碍于

识别准确率的进一步提高;2)将“分离—共享”机制引入到步态识别的多部位特征学习过程中,设计了一个即插即用的多部位特征学习模块;3)在不同数据集和基线方法上的实验结果表明,本文设计的多部位特征学习模块能够在不显著增加计算代价的条件下带来稳定的性能提升。

1 方 法

本文设计的识别流程图如图 2 所示,主要包含骨干网络特征提取和多部位特征提取两部分。其中,多部位特征提取又可以进一步划分为分离机制和共享机制两部分。本文以 GaitSet 为例对每个模块进行简要介绍,其中多部位特征提取中的共享机制是本文的核心工作。值得注意的是,本文设计的多部位特征提取模块具有较强的适用性,也能够以类似的方式应用于 GaitPart 模型。

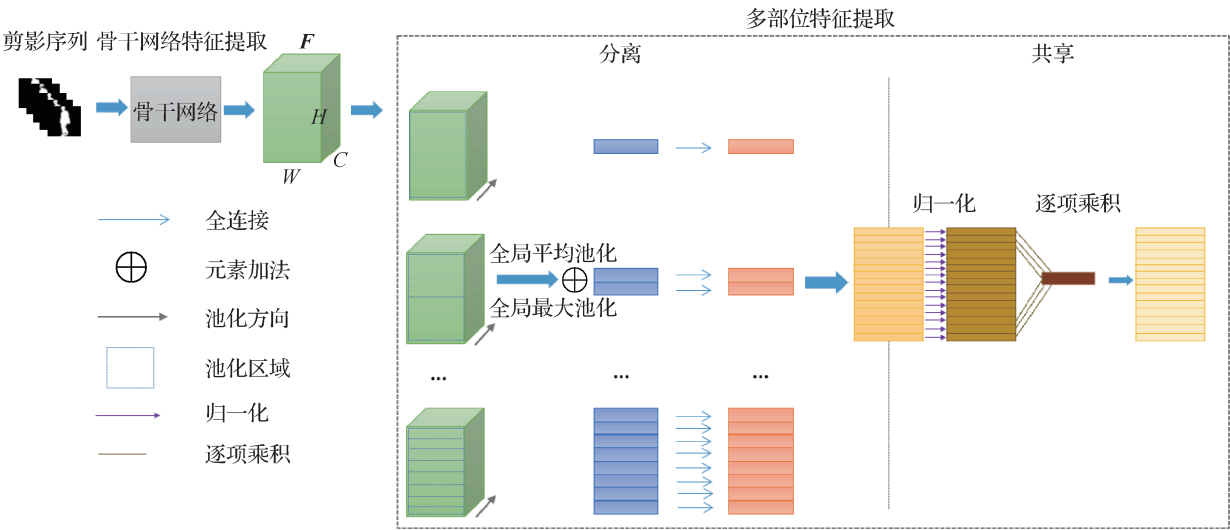


图 2 识别流程图

Fig. 2 The illustration of recognition pipeline

1.1 骨干网络特征提取

骨干网络的设计借鉴了 GaitSet 和 GaitPart 等模型的主要思想,主要包含剪影特征提取和集合池化两部分。剪影特征提取中对每幅剪影单独提取,不同剪影之间相互独立;集合池化是对所有剪影的特征进行融合,获取整个序列的特征表示。

这一部分基于 GaitSet 的结构如图 3 所示,为了绘图方便,省略了 GaitSet 原论文中 multilayer global pipeline 这一分支。其中,剪影特征提取部分由多

层卷积和池化操作组成,将剪影序列输入到该网络模块即可得到对应序列中每个剪影的特征。其次是集合池化,该模块将剪影序列视为无序集合,使用统计函数整合所有剪影的特征,得到整个剪影集合的特征。将剪影序列视为一个集合的依据在于每个剪影在步态周期上具有独特的姿态信息,可以通过这种外观信息和人体运动的一般规律,对序列进行时序上的排序,因此可以将剪影序列用一个无序的集合表示。同时考虑到在现实环境中,序列中

剪影的数量是任意的, 所以对于该统计函数的选择, 不仅需要满足集合的无序性, 同时还需满足该集合可以有任意数目的元素。对应满足条件的常

见的统计函数有最大函数、最小函数和中值函数等。在本文的实验中采用最大统计函数, 即最大池化。

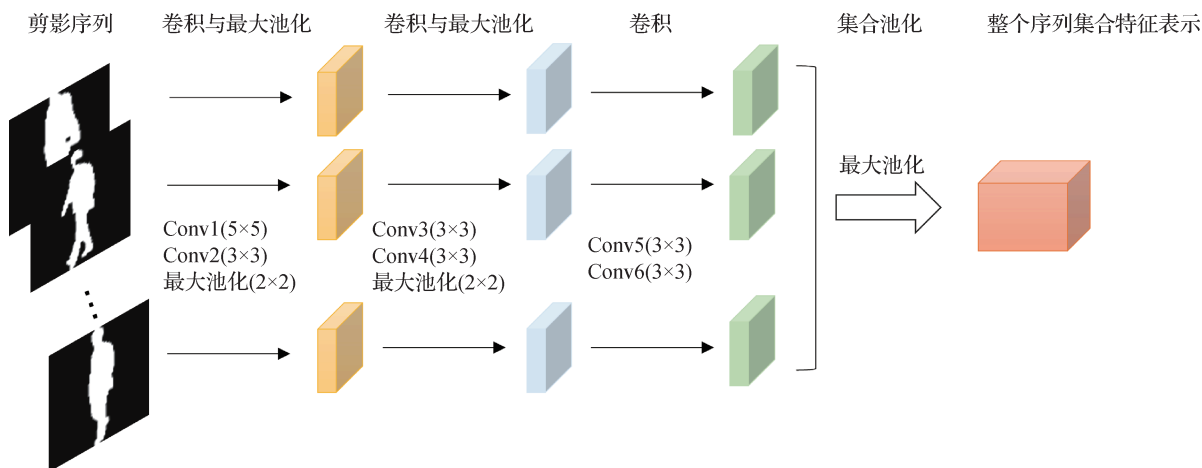


图3 骨干网络结构图

Fig. 3 The illustration of backbone structure

1.2 多部位特征提取

1.2.1 分离机制

分离机制的设计借鉴了GaitSet中的多部位特征学习模块。首先使用多个尺度将骨干网络输出特征图水平划分为多个部分; 其次将得到的多尺度下的多个特征图进行平均和最大池化操作; 最后经过不共享权重的全连接层, 获得具有判别性的多部位特征。这里对于骨干网络输出的特征用 F 进行表示, 使用 S 个尺度进行划分, 一共可以得到多尺度下 N 个部位的特征, 其中 $N = \sum_{s=1}^S 2^{s-1}$ 。然后, 在不同尺度下的空间维度进行池化操作, 对于第 i 个部位特征 f_i , 经过池化操作的特征 $f_i^{\#}$ 可以表示为

$$f_i^{\#} = \text{GAP}(f_i) + \text{GMP}(f_i) \quad (1)$$

式中, GAP 表示全局平均池化(global average pooling), GMP 表示全局最大池化(global max pooling)。最后使用具有独立权重的全连接层对每个部位进行特征映射, 将得到的每个部位的特征用 d 表示, 其维度用 c 表示。

值得一提的是, 使用不共享权重的全连接层的出发点在于不同尺度下的部位拥有不同的感受野, 而在同一感受野下的不同部位对应着不同的空间位置信息。上述过程能够有效保留水平方向上每个身体部位各自的运动特性, 不同部位之间相互独立, 称为分离机制。

1.2.2 共享机制

本文通过引入共享机制, 增加每个部位在特征提取过程中的交互性和协同性, 从而获取更具有判别性的步态特征。本文设计的共享机制由特征归一化和特征重映射两部分组成。

首先, 特征归一化是不带任何参数的操作, 它的目标是使不同部位的特征具有相似的统计特性, 有利于后续进行的特征重映射操作。本文将上步分离机制得到的每个部位特征 d_i ($i = 1, 2, \dots$) 分别通过特征归一化得到对应的部位特征 $d_i^{\#}$, 这一过程可以表示为

$$d_i^{\#} = \frac{d_i - E[d_i]}{\sqrt{\text{Var}[d_i] + \varepsilon}} \quad (2)$$

式中, $E[]$ 表示对每个 d_i 向量各个元素的均值计算, $\text{Var}[]$ 表示方差计算, ε 为一个常量, 避免分母为0, 本文所有实验中 ε 设置为 10^{-5} 。

其次, 特征重映射是带有参数的, 在不同部位之间共享权重, 完成不同部位之间信息的交互和协同。本文采用两种不同的方式实现特征重映射, 分别是共享权重的全连接层(fully-connected layer)和共享权重的逐项乘积(element-wise multiplication)。

1) 共享权重的全连接层。用 w 表示不同部位共享的全连接层的权重(忽略bias向量), 对特征归一化后的每个部位特征 $d_i^{\#}$ 使用 w 进行矩阵乘法, 输出

向量用 $\mathbf{o}_i$ 表示,将不同部位特征串联起来,得到模型的输出向量 $\mathbf{E}_{\text{fully}}$ 。该过程可表示为

$$\mathbf{o}_i = \mathbf{w}^T \times \mathbf{d}_i^{\#} \quad (3)$$

式中, $\times$ 表示矩阵乘法。

$$\mathbf{E}_{\text{fully}} = \Phi(\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_i, \dots) \quad (4)$$

$$\mathbf{d}_i^{\#} \in \mathbf{R}^{c \times 1}$$

$$\mathbf{w} \in \mathbf{R}^{c \times c}$$

$$\mathbf{o}_i \in \mathbf{R}^{c \times 1}$$

$$\mathbf{E}_{\text{fully}} \in \mathbf{R}^{c \times N}$$

式中, Φ 表示串联操作, c 表示每个部位特征的维度, N 表示身体部位的数目。

2) 共享权重的逐项乘积。用 $\mathbf{v}$ 表示一个具有与每个部位特征相同维度的向量。对特征归一化后的每个部位特征 $\mathbf{d}_i^{\#}$, 使用同一个 $\mathbf{v}$ 进行逐项乘积, 输出向量用 $\mathbf{e}_i$ 进行表示, 将不同部位的输出串联起来组成模型输出特征向量 $\mathbf{E}_{\text{element}}$ 。该过程可表示为

$$\mathbf{e}_i = \mathbf{v} \odot \mathbf{d}_i^{\#} \quad (5)$$

$$\mathbf{E}_{\text{element}} = \Phi(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_i, \dots) \quad (6)$$

$$\mathbf{d}_i^{\#} \in \mathbf{R}^{c \times 1}$$

$$\mathbf{v} \in \mathbf{R}^{c \times 1}$$

$$\mathbf{e}_i \in \mathbf{R}^{c \times 1}$$

$$\mathbf{E}_{\text{element}} \in \mathbf{R}^{c \times N}$$

式中, $\odot$ 表示逐项乘积, Φ, c, N 的含义如前所述。值得一提的是, 基于逐项乘积的共享机制实现方式类似于 Layer Normalization 层中的操作。

1.3 模型训练和测试

本文提出的多部位特征学习机制能够很方便地与 GaitSet 或 GaitPart 基线模型结合, 为保证公平的性能对比, 训练和测试阶段的参数跟对应基线模型保持一致。

以 GaitSet 为例, 在训练阶段, 对训练集所有被试者进行均匀采样。每个批次包含的序列样本通常用 $P \times K$ 进行表示, 其中, P 代表被试者数目, K 表示每个被试者选取的剪影序列数目。损失函数采用 Hermans 等人(2017)提出的 Batch All 三元组损失函数。损失函数首先对每个身体部位分别进行计算, 然后在所有身体部位之间进行平均。

在测试阶段, 首先将测试集中所有被试者的序列分为 probe 和 gallery 两部分, 然后计算每个 probe 序列与 gallery 序列的距离, 将距离最近的 gallery 序列的身份标签赋给对应的 probe 序列。其中, 两个序列之间的距离计算方式是首先衡量每个对应身体部

位特征之间的欧氏距离, 然后在不同部位之间进行平均。当 probe 包含多个行走条件的序列时, 通常将 probe 划分为多个子集分别进行测试。

2 实验

2.1 实验设置

2.1.1 数据集

CASIA-B(Yu 等, 2006)是由中国科学院自动化研究所提供的步态数据集, 广泛应用于步态模型评估。该数据集包含 124 个受试者在 3 种行走条件和 11 个角度下的视频序列。其中, 124 个受试者使用标签 001-124 表示。在每个受试者下包含 3 种行走条件, 包括 6 个正常行走(normal walking, NM)、2 个背包行走(walking with bags, BG)和 2 个换衣状态(walking in different clothes, CL), 分别使用 nm-01—nm-06、bg-01 和 bg-02、cl-01 和 cl-02 表示。在每个行走条件下包含 11 种角度(从 $0^\circ \sim 180^\circ$, 间隔为 18°), 使用 000-180 表示。该数据集不仅提供了原始的视频数据, 同时提供了剪影序列数据。本文实验中采用的是剪影数据。实验时, 将数据集划分为 74 个受试者为训练集, 剩余 50 个为测试集。在测试阶段, nm-01—nm-04 作为 gallery, 其余分为 3 组 probe, 分别是 nm-05 和 nm-06、bg-01 和 bg-02、cl-01 和 cl-02。

OUMVLP(Takemura 等, 2018)是日本大阪大学研究者创建的步态数据集, 是目前公开的包含被试者数目最多的步态数据集, 包含了 10 307 个受试者在多个角度的数据, 但仅有正常行走条件下的步态序列。对每个受试者记录了 14 个角度, 即 $0^\circ \sim 90^\circ$ 、 $180^\circ \sim 270^\circ$, 间隔为 15° 。每个角度下包含两段序列(编号为 00, 01)。在本文实验中, 将 5 153 个被试者的数据作为训练集, 将 5 154 个被试者的数据作为测试集。在测试阶段, 将每个被试者编号为 01 的序列作为 gallery, 编号为 00 的序列作为 probe。

2.1.2 实现细节

所有模型使用 Pytorch 实现, 在 TITAN-V GPU 上进行训练。

基于数据集中提供的原始剪影序列, 通过使用 Takemura 等人(2018)描述的方法进行对齐操作后作为模型的输入。对于 CASIA-B 数据集, 输入分辨率为 128×88 像素, 每批样本包含 128 个序列, 其中

P 为 8, K 为 16。对于 OUMVLP 数据集, 输入分辨率为 64×44 像素, 每批样本包含 512 个序列, 其中 P 为 32, K 为 16。在训练阶段, 对于每个序列, 随机抽取 30 帧剪影作为模型输入。在测试阶段, 将每个序列中的所有剪影作为模型输入提取步态特征进行比对。

在基于 GaitSet 的实现中, 在分离阶段, 使用了 5 个尺度 (1, 2, 4, 8, 16), 加上 Multilayer Global Pipeline 全局分支, 共得到 62 个部位, 在池化操作后, 特征经过输出通道数为 256 的不共享参数的全连接, 得到该部位的输出。在共享阶段, 将分离机制得到的特征进行归一化处理, 再对每个部分使用输入输出均为 256 的全连接层进行处理或使用维度为 256 的向量进行逐项乘积。对于模型训练, 在 CASIA-B 数据集上采用随机梯度下降法 (stochastic gradient descent, SGD) 优化器, 初始学习率设置为 0.1, 每迭代 10 000 次将学习率降低 10 倍, 在该数

据集上共迭代 40 000 次。在 OU-MVLP 数据集上采用 Adam 优化器, 学习率设置为 10^{-4} , 并在第 150 000 次迭代时, 将学习率降为 10^{-5} , 在该数据集上一共迭代 250 000 次。

基于 GaitPart 的实现方式与 GaitSet 基本类似, 但模型结构和对应的优化过程略有不同。在模型训练过程中, 在 CASIA-B 数据集上采用 Adam 优化器, 学习率设置为 10^{-4} , 并在第 100 000 次迭代时, 将学习率降为 10^{-5} , 在该数据集上共迭代 120 000 次。在 OU-MVLP 数据集上采用 Adam 优化器, 学习率设置为 10^{-4} , 并在 150 000 次迭代时, 将学习率降为 10^{-5} , 在该数据集上一共迭代 250 000 次。

2.2 实验结果对比

表 1—表 3 为在 CASIA-B 数据集上, 分别使用 GaitSet 和 GaitPart 作为基线方法以及两种共享机制的实现方式, 在各个角度下正常行走 (NM)、背包行走 (BG) 和换衣状态 (CL) 的模型准确率。

表 1 不同模型在 CASIA-B 数据集上的正常行走条件下的准确率
Table 1 Accuracy of different models under normal walking conditions on the CASIA-B dataset

模型												/%
	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	平均
GaitSet	93.6	99.7	98.8	97.5	96.1	95.5	96.1	98.7	98.9	97.6	92.0	96.77
GaitSet+Norm+Fully	96.9	100.0	99.6	98.8	98.0	96.5	97.6	98.7	99.9	98.8	94.9	98.16
GaitSet+Norm+Element	97.2	99.6	99.6	98.1	97.9	96.8	98.0	98.9	99.9	99.3	94.5	98.16
GaitPart	96.9	99.5	99.8	98.6	96.4	95.4	96.5	98.7	99.4	98.5	94.2	97.63
GaitPart+Norm+Fully	97.0	99.7	99.9	99.5	97.3	97.0	97.4	99.1	99.9	99.2	95.8	98.35
GaitPart+Norm+Element	97.8	99.8	100.0	99.6	98.0	97.9	98.5	99.5	99.7	99.5	96.1	98.76

注: 加粗字体表示各列最优结果。

表 2 不同模型在 CASIA-B 数据集上的背包行走条件下的准确率
Table 2 Accuracy of different models under carrying bags walking conditions on the CASIA-B dataset

模型												/%
	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	平均
GaitSet	92.3	96.7	96.0	94.85	91.4	90.1	91.4	94.7	96.1	94.75	88.2	93.32
GaitSet+Norm+Fully	93.7	97.3	96.8	94.85	93.3	91.7	92.7	94.9	96.9	96.67	90.3	94.47
GaitSet+Norm+Element	94.8	97.8	96.3	95.05	93.7	92.0	94.2	96.2	97.2	97.07	90.0	94.94
GaitPart	93.1	97.4	97.0	95.46	92.5	89.4	91.9	95.4	96.8	96.06	88.4	93.95
GaitPart+Norm+Fully	93.1	97.9	97.5	95.76	94.1	91.2	94.0	95.5	96.6	97.37	89.8	94.80
GaitPart+Norm+Element	93.4	98.2	97.3	95.86	95.0	92.6	93.8	95.5	98.0	96.97	89.7	95.12

注: 加粗字体表示各列最优结果。

表 3 不同模型在 CASIA-B 数据集上的换衣行走条件下的准确率
Table 3 Accuracy of different models under changing clothes walking conditions on the CASIA-B dataset

模型												/%
	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	平均
GaitSet	71.0	83.7	84.3	82.2	76.8	75.1	75.3	79.0	80.9	75.7	66.9	77.36
GaitSet+Norm+Fully	72.6	84.8	88.1	83.6	79.4	79.1	78.9	83.9	84.5	80.0	68.8	80.34
GaitSet+Norm+Element	74.9	87.3	88.1	85.4	83.1	80.8	81.4	84.9	81.6	79.3	68.0	81.35
GaitPart	77.0	86.6	90.3	84.2	80.9	77.5	80.0	84.4	87.1	81.2	72.9	82.01
GaitPart+Norm+Fully	71.3	85.7	89.2	84.5	79.3	75.6	78.2	81.5	84.1	79.5	66.9	79.62
GaitPart+Norm+Element	76.5	86.9	88.4	86.7	83.3	79.6	80.9	83.6	84.4	80.0	70.8	81.92

注:加粗字体表示各列最优结果。

在共享机制的实现中,本文用 Norm 代表特征归一化操作,特征重映射分别使用了共享参数的全连接(fully)和逐项乘积(element)两种方式。从表中可以看出,在正常行走(NM)和背包行走(BG)条件下,“分离—共享”机制能够带来稳定的性能提升,说明共享机制的引入能够有效地协同不同部位的特征学习,从而获取更具有判别性的步态特征。在复杂的换装行走(CL)条件下,在以 GaitSet 作为基线方法时,“分离—共享”机制能够稳定地提升识别准确率;而在以 GaitPart 作为基线方法时,分离共享机制会使识别准确率略微降低,这很有可能是由于过拟合导致的。但是从

3个行走条件下平均识别准确率的角度分析,“分离—共享”机制能够为 GaitSet 和 GaitPart 带来稳定的性能提升。此外,与最近基于剪影序列的方法(Huang 等, 2022)相比,本文方法在换装条件下的性能略低,但是在正常行走和背包的情况下能够实现较高的准确率。本文工作和 Huang 等人(2022)工作从不同的维度对步态识别任务进行了进一步探索。

表 4 为 GaitSet 和 GaitPart 两个基线方法及其改进模型在 OUMVLP 数据集上的表现。同样,使用“分离—共享”机制的模型在多个视角下准确率都有稳定的提升。

表 4 不同模型在 OUMVLP 数据集上的准确率
Table 4 Accuracy of different models on OUMVLP dataset

																/%
模型	0°	15°	30°	45°	60°	75°	90°	180°	195°	210°	225°	240°	255°	270°	平均	
GaitSet	79.24	87.57	89.99	90.11	88.10	88.72	87.68	81.70	86.46	89.05	89.26	87.27	87.67	86.31	87.08	
GaitSet+Norm+Fully	81.95	88.73	90.31	90.56	88.92	89.45	88.58	83.86	87.72	89.38	89.60	88.13	88.31	87.27	88.06	
GaitSet+Norm+Element	81.82	88.85	90.31	90.62	88.93	89.41	88.57	84.00	87.88	89.46	89.68	88.21	88.41	87.25	88.10	
GaitPart	82.63	89.06	90.84	90.96	89.69	89.85	89.38	85.10	87.64	89.92	90.06	89.01	88.93	88.05	88.65	
GaitPart+Norm+Fully	83.80	89.80	90.86	91.01	89.85	89.87	89.37	86.09	88.83	89.96	90.17	88.96	88.98	88.19	88.98	
GaitPart+Norm+Element	83.23	89.59	90.73	90.91	89.56	89.68	89.22	85.67	88.64	90.00	90.07	88.73	88.80	87.94	88.77	

注:加粗字体表示各列最优结果。

2.3 实验结果讨论

本文提出了共享机制的两种实现方式,由于在不同部位之间共享权重,引入的参数数量以及增加的运算量都相对有限。从表 1—表 4 的实验结果可以看出,在大部分情况下,逐项乘积(element)的方法

能够取得比全连接(fully)方法更高的识别准确率。通过比较两种实现方式发现,逐项乘积的方法引入的参数数量仅与单个身体部位特征的维度相等(忽略 bias 向量)。换言之,本文能够在不显著增加计算代价的前提下带来稳定的性能提升。

3 结 论

本文针对目前步态识别领域广泛采用的多身体部位特征学习机制中存在的问题,设计了一种“分离—共享”机制,用于学习更具有判别力的各个身体部位特征。本文方法首先通过分离机制,提取多身体部位独立的运动特性,接着通过共享机制,建立身体多部位特征之间的联系,使多部位特征之间更好地进行信息交互。本文使用GaitSet和GaitPart作为基线模型,分别尝试了两种共享机制的实现方式,并在CASIA-B和OUMVLP数据集上进行训练和测试。实验结果表明,与基线模型相比,本文提出的增强多部位特征学习的策略能够在不显著增加计算代价的条件下,稳定地提高步态模型在多个视角和行走状态下的识别准确率。

目前,步态识别研究中广泛使用的多部位特征学习模块是从行人重识别中借鉴而来,在步态相关研究工作中缺乏对这一模块的原理分析。本文的核心观点是不同身体部位的特征学习之间缺乏交互有碍于步态识别准确率的进一步提高。为了缓解这一问题,本文将“分离—共享”机制引入到多部位特征学习中,并提供了共享机制的两种简单有效的实现方式。针对不同部位之间缺乏交互这一问题的进一步解决方案,以及设计更为复杂的共享机制,都值得在未来工作中进一步探索。

参考文献(References)

- Bashir K, Xiang T and Gong S G. 2009. Gait representation using flow fields//Proceedings of 2009 British Machine Vision Conference. London, UK: BMVA Press; #113 [DOI: 10.5244/C.23.113]
- Ben X Y, Xu S and Wang K J. 2012. Review on pedestrian gait feature expression and recognition. Pattern Recognition and Artificial Intelligence, 25(1): 71-81 (黄晔烨, 徐森, 王科俊. 2012. 行人步态的特征表达及识别综述. 模式识别与人工智能, 25(1): 71-81) [DOI: 10.3969/j.issn.1003-6059.2012.01.010]
- Castro F M, Marín-Jiménez M J, Guil N and Pérez de la Blanca N. 2020. Multimodal feature fusion for CNN-based gait recognition: an empirical comparison. Neural Computing and Applications, 32(17): 14173-14193 [DOI: 10.1007/s00521-020-04811-z]
- Chao H Q, He Y W, Zhang J P and Feng J F. 2019. GaitSet: regarding gait as a set for cross-view gait recognition. Proceedings of the AAAI Conference on Artificial Intelligence, 33(1): 8126-8133 [DOI: 10.1609/aaai.v33i01.33018126]
- Fan C, Peng Y J, Cao C S, Liu X, Hou S H, Chi J N, Huang Y Z, Li Q and He Z Q. 2020. GaitPart: temporal part-based model for gait recognition//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 14213-14221 [DOI: 10.1109/CVPR42600.2020.01423]
- Han J and Bhanu B. 2006. Individual recognition using gait energy image. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(2): 316-322 [DOI: 10.1109/TPAMI.2006.38]
- Hermans A, Beyer L and Leibe B. 2017. In defense of the triplet loss for person re-identification [EB/OL]. [2022-06-12]. <https://arxiv.org/pdf/1703.07737.pdf>
- Hou S H, Liu X, Cao C S and Huang Y Z. 2021. Set residual network for silhouette-based gait recognition. IEEE Transactions on Biometrics, Behavior, and Identity Science, 3(3): 384-393 [DOI: 10.1109/TBIOM.2021.3074963]
- Huang T H, Ben X Y, Gong C, Zhang B C, Yan R and Wu Q. 2022. Enhanced spatial-temporal salience for cross-view gait recognition. IEEE Transactions on Circuits and Systems for Video Technology, 32(10): 6967-6980 [DOI: 10.1109/TCSVT.2022.3175959]
- Li N and Zhao X B. 2022. A strong and robust skeleton-based gait recognition method with gait periodicity priors. IEEE Transactions on Multimedia: #3154609 [DOI: 10.1109/TMM.2022.3154609]
- Li X, Makihara Y, Xu C and Yagi Y. 2021. End-to-end model-based gait recognition using synchronized multi-view pose constraint//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision Workshops. Montreal, Canada: IEEE: 4089-4098 [DOI: 10.1109/ICCVW54120.2021.00456]
- Liao R J, Cao C S, Garcia E B, Yu S Q and Huang Y Z. 2017. Pose-based temporal-spatial network (PTSN) for gait recognition with carrying and clothing variations//Proceedings of the 12th Chinese Conference on Biometric Recognition. Shenzhen, China: Springer: 474-483 [DOI: 10.1007/978-3-319-69923-3_51]
- Lin B B, Zhang S L and Yu X. 2021. Gait recognition via effective global-local feature representation and local temporal aggregation//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 14628-14636 [DOI: 10.1109/ICCV48922.2021.01438]
- Takemura N, Makihara Y, Muramatsu D, Echigo T and Yagi Y. 2018. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. IPSJ Transactions on Computer Vision and Applications, 10: #4 [DOI: 10.1186/s41074-018-0039-6]
- Teepe T, Khan A, Gilg J, Herzog F, Hörmann S and Rigoll G. 2021. Gaitgraph: graph convolutional network for skeleton-based gait recognition//Proceedings of 2021 IEEE International Conference on Image Processing. Anchorage, USA: IEEE: 2314-2318 [DOI: 10.1109/ICIP42928.2021.9506717]
- Wang X N, Hu D D, Zhang T and Bai G X. 2021. Gait recognition using

pose features and 2D Fourier transform. Journal of Image and Graphics, 26(4): 796-814 (王新年, 胡丹丹, 张涛, 白桂欣. 2021. 姿态特征结合2维傅里叶变换的步态识别. 中国图象图形学报, 26(4): 796-814) [DOI: 10.11834/jig.200061]

Xu S, Zheng F, Tang J and Bao W X. 2022. Dual branch feature fusion network based gait recognition algorithm. Journal of Image and Graphics, 27(7): 2263-2273 (徐硕, 郑锋, 唐俊, 鲍文霞. 2022. 双分支特征融合网络的步态识别算法. 中国图象图形学报, 27(7): 2263-2273) [DOI: 10.11834/jig.200730]

Yu S Q, Tan D L and Tan T N. 2006. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition//Proceedings of the 18th International Conference on Pattern Recognition. Hong Kong, China; IEEE: 441-444 [DOI: 10.1109/ICPR.2006.67]

Zhang H Y and Bao W J. 2022. The cross-view gait recognition analysis based on generative adversarial networks derived of self-attention mechanism. Journal of image and Graphics, 27(4): 1097-1109 (张红颖, 包雯静. 2022. 融合自注意力机制的生成对抗网络跨

视角步态识别. 中国图象图形学报, 27(4): 1097-1109) [DOI: 10.11834/jig.200482]

作者简介

侯赛辉,男,讲师,主要研究方向为计算机视觉。

E-mail: housaihui@bnu.edu.cn

黄永祯,通信作者,男,副教授,主要研究方向为机器学习、计算机视觉。E-mail: huangyongzhen@bnu.edu.cn

付杨,男,硕士研究生,主要研究方向为计算机视觉。

E-mail: aleeyanger@gmail.com

李奥奇,男,硕士研究生,主要研究方向为计算机视觉。

E-mail: 201811210216@mail.bnu.edu.cn

刘旭,男,博士,算法工程师,主要研究方向为计算机视觉。

E-mail: xu.liu@watix.ai

曹春水,男,博士,算法工程师,主要研究方向为计算机视觉。

E-mail: chunshui.cao@watix.ai