计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2022,31(8):88−98 [doi: 10.15888/j.cnki.csa.008647]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

面向多人多生物属性的跨视角步态追踪系统①

黄彬源, 罗咏东, 谢家辉, 李志文, 周成菊, 潘家辉

(华南师范大学 软件学院, 佛山 528225)
通信作者: 周成菊, E-mail: cjzhou@scnu.edu.cn

摘　要: 步态识别是一项新兴的生物识别技术, 可以被广泛地应用在刑事安防, 疫情传播链追踪等领域, 该项技术的本质在于通过人的人体体型和行走姿态来识别人的身份, 年龄, 性别等多种生物属性. 相比其他生物识别技术, 步态识别具有远距离, 全视角, 无感知, 防伪装等显著优势. 基于此, 本文设计了一款面向多人多生物属性的跨视角步态追踪系统, 该系统充分考虑了现实应用场景中存在的多人, 跨视角, 服饰变化等协变量对于步态识别准确率的影响, 并通过更加鲁棒的算法设计从复杂的环境中提取行人的步态信息从而对其身份, 年龄, 性别等生物属性进行准确的分析. 实验结果表明, 在跨视角和多种行走状态的情况下, 本系统中基于深度学习的步态识别算法模型的准确率可以达到 88.0%, 在多视角的情况下, 性别分类准确率可以达到 94.8%, 年龄估计的平均年龄误差约为 7.92 岁, 标准差约为 8.11, 实验结果均优于近年来相关领域的算法, 达到相对领先的水平. 同时系统开发成本低, 面向落地应用场景, 并支持实时性步态检测.

关键词: 步态识别; 多属性识别; 跨视角; 实时检测; 深度学习; 目标检测; 多目标跟踪; 语义分割

引用格式: 黄彬源,罗咏东,谢家辉,李志文,周成菊,潘家辉.面向多人多生物属性的跨视角步态追踪系统.计算机系统应用,2022,31(8):88–98. http://www.c-s-a.org.cn/1003-3254/8647.html

Cross-view Gait Tracking System Oriented to Multiple People and Multiple Biological Attributes

HUANG Bin-Yuan, LUO Yong-Dong, XIE Jia-Hui, LI Zhi-Wen, ZHOU Cheng-Ju, PAN Jia-Hui

(School of Software, South China Normal University, Foshan 528225, China)

Abstract: Gait recognition is an emerging biometric technology, which can be widely used in criminal security, epidemic transmission chain tracking, etc. The essence of this technology is to identify people's identity, age, gender and other biological attributes through their human body shape and walking posture. Compared with other biometric technologies, gait recognition has significant advantages such as long distance, full view, no perception, and anti-counterfeiting. In this study, we design a cross-view gait tracking system for multiple people and multiple biological attributes. The system fully considers the impact of covariates (such as multiple people, cross view and clothing change) on gait recognition accuracy in real application scenarios. It extracts the gait information of pedestrians from complex environments to accurately analyze their biological attributes such as identity, age, and gender through a more robust algorithm design. The experimental results show that the accuracy of the deep learning-based gait recognition algorithm model in this system can reach 88.0% in the case of cross view and multiple walking states and 94.8% in the case of multiple views for gender classification. The average age error of age estimation is about 7.92 years with a standard deviation of about 8.11. These results are better than those of recent algorithms in related fields and reach a relatively leading level. At a low development cost, the system is oriented to application scenarios and supports real-time gait detection.

Key words: gait recognition; multi-attribute recognition; cross view; real-time detection; deep learning; object detection; multi-object tracking; semantic segmentation

① 基金项目: 国家自然科学基金面上项目 (62076103); 广东省自然科学基金面上项目 (2019A1515011375); 广东省科技创新人才专项珠江科技新星专题项目 (201710010038)
收稿时间: 2021-11-12; 修改时间: 2021-12-13; 采用时间: 2021-12-28; csa 在线出版时间: 2022-06-01

1 引言

1.1 研究背景

生物特征识别技术正受到学术界和工业界的广泛认可, 因为它可以通过人类的行为特征进行属性识别. 生物特征识别领域中包含了许多的技术, 其中不乏虹膜识别、人脸识别等广为人知的生物识别技术. 然而, 这些方法包含以下缺陷: 对伪装后的特征辨别效果差, 只能在短距离内进行识别, 同时在大多数情景下都需要受试者的主动配合[1].

以人脸识别在实际中的应用缺陷为例: 犯罪分子通常会以戴口罩或者是易容的方式再次出现, 此时由于人脸特征被隐藏, 通过人脸识别的方式极难识别; 另外是在疫情常态化的当下, 戴口罩成为常规的出行方式, 这同时也加剧了人脸识别的难度; 再则是人脸识别的距离有限, 这对于实际应用中监控摄像头的素质也提出了更高的要求, 因此也增加了硬件成本[1].

而步态识别 (gait recognition) 技术则成为解决上述问题的关键. 步态识别旨在通过分析人的行走模式进而对其进行属性识别, 每个人特有的生理结构决定了其独有的步态, 因此人体的步态具有唯一性和稳定性, 这也成为了步态识别可行性的基础[2]. 步态识别可以在 2K 摄像头 (公安主流摄像头) 下达到最远 50 m 的识别距离, 同时不需要受试者的主动配合, 并且可以从人体的全身捕捉特征, 它不依赖于人体的某一部分, 因此受到的约束就少. 基于上述优势, 步态识别也被称为当下安防领域中极具应用前景的生物识别技术.

1.2 国内外研究现状

目前, 步态识别技术的实现主要分为两大类: 一类是传统的基于机器学习的步态识别算法, 另一类则是基于深度学习的步态识别算法. 基于机器学习的步态识别研究重点之一是解决视角变化的问题[3–11]. 其中, 部分方法通过学习更高的专业知识来提取视角恒定的步态特征, 另一部分方法则通过构建视图转换模型 (view transform model, VTM) 来规范化不同的视角. Kusakunniran 等人[3] 提出了基于视角恒定特征的步态识别框架将不同的视角归一化. 此外, Kusakunniran 等人[4] 还利用截断奇异值分解技术 (truncated singular value decomposition, TSVD) 构建了视角转换模型, 该技术可以将图库样本和探针样本的不同视角转换为同一视角. 上述传统算法虽然可以达到比较高的实验精度, 但在实际应用中却难以克服各种复杂的协变量的影响 (如行人服饰变化, 视角发生较大改变), 缺乏一定的鲁棒性和普适性. 而基于深度学习的步态识别算法虽然没有明确地对视角的变化进行建模, 但依然可以实现良好的跨视角步态识别性能. 现有的基于深度学习的步态识别方法大致可以分为两类: 第 1 种是基于模板图的方式, 该方法将所有的步态轮廓压缩成一个步态信息的模板. Wu 等人[12] 首先介绍了基于 CNN 的从步态能量图像 (GEI) 中捕捉步态模式深度特征的方法. Shiraga 等人[13] 使用 2D CNN 从 GEI 中提取步态特征. 尽管上述的方法尽可能期望使用模板表征丰富的步态时序信息, 但是不可避免的散失了时序信息和细粒度的空间信息[14], 因此并不适合在实际的系统进行应用. 第 2 类则是基于剪影图序列的方法, 利用卷积直接编码来自原始步态轮廓序列的时空表征从而可以获得更为丰富而全面的步态时序信息, Zhang 等人[14] 提出基于 LSTM 的步态识别算法以捕获更长时间范围的时间信息. 为了提高步态识别的灵活性, Chao 等人[15] 提出 GaitSet 算法将步态视为一个集合而不是一个序列, 从而获得更为丰富的步态样本数量并取得优异的性能, Huang 等人[16] 提出了一种基于信息加权模块和局部特征流调节模块进行步态特征学习. Sepas-Moghaddam 等人[17] 提出使用双向递归神经网络的学习关系序列中提取的部分特征. Ding 等人[18] 提出了一种顺序卷积神经网络 SRN 从新颖的角度学习时空特征.

在应用领域方面, 国内的银河水滴科技发布了全球首个步态识别互联系统“水滴慧眼”, 该系统可以实现远距离、多视角的步态识别, 且适用于大范围人群密度测算, 在安防领域、智能交通、智能家居和医疗康养方面有较为广泛的应用. 上述系统功能较为完备且适用范围广, 但仍然存在一定的局限性. 一方面, 上述系统的研发与采购成本均较高; 另一方面, 上述系统更多关注于受试者的步态身份信息, 一旦身份信息失效, 系统则无法提供其他的参考信息. 因此, 本文开发了一款低成本、支持实时多信息跨视角检测的智能安防系统, 在保证安防系统的基础身份识别之外, 拓展了步态年龄与性别信息辅助筛选, 配合路段追踪的功能, 可以很好地满足安防常规需求, 具有重大的现实意义.

2 系统设计

2.1 系统结构介绍

本系统的全名为面向多人多生物属性的跨视角步

态追踪系统, 意指本系统的算法设计面向现实应用场景, 致力于解决现实场景中可能出现的多行人, 跨视角, 多种行走状态等应用难点问题. 系统通过合理而高效的算法设计对上述的协变量进行有效处理并得到了行人的多种生物属性.

本系统宏观上可以分为 4 大模块: 行人检测和追踪, 行人分割, 算法模型训练, 系统实现, 如图 1 所示. 首先, 对于输入的视频序列, 系统首先通过行人检测模块将行人从若干种复杂的街头事物中分离开来; 其次, 通过行人追踪模块从视频中提取出某个特定行人在视频中的完整步态序列. 接着, 系统通过行人分割模块进行前景和背景的分离, 得到行人的二值化步态序列图. 最后, 系统通过特征提取算法模型对提取到的二值化步态序列图进行不同任务的训练并由此分析出行人的身份, 年龄, 性别等多种生物属性. 在系统实现中, 我们提供了一个 Windows 系统下的客户端, 同时客户端已打包好算法模型运行时所需的环境依赖, 因此可以对用户输入的视频进行实时的属性分析.

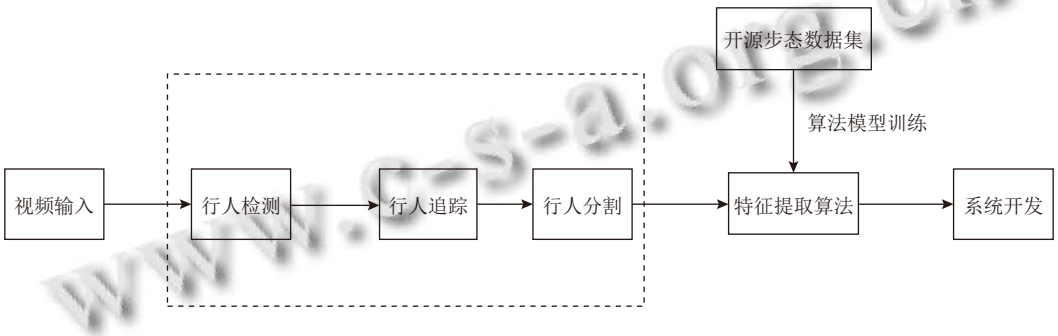

图 1 系统结构图

2.2 行人检测

在行人检测阶段, 我们使用 YOLOv4 算法[19] 对输入的视频序列进行检测. YOLOv4 算法[19] 由 CSPDarknet53 网络、SPP-Net (spatial pyramid pooling networks) 网络、PANet (path aggregation network) 网络以及 YOLOv3 检测头组成, 其中 CSPDarknet53 网络解决了其他大型卷积神经网络框架的梯度信息重复问题, 即将梯度变化集成到特征图中, 有效减少模型参数量和运算量, 在保持推理速度的同时缩小了模型尺寸; SPP-Net 网络主要对任意尺寸的特征图直接进行固定尺寸的池化, 以获得固定数量的行人特征; PANet 网络进行参数聚合以适用于不同水平的行人检测, 最后由 YOLOv3 检测头实现对大中小 3 类目标的检测. YOLOv4 算法对行人特征进行有效提取、集成和映射, 实现了系统运算速度和行人检测精度的完美平衡.

如图 2 所示, 我们将行人视频的帧序列划分成横纵网格, 如果某个行人的中心落在这个网格中, 则这个网格就负责预测边界框的位置信息、置信度以及类别信息. 计算出行人的 ID 置信度分数后, 系统根据预设的阈值过滤分数不佳的边界框, 对保留的边界框进行非最大值抑制算法处理, 得到每个行人的边界框.

2.3 行人追踪

因为系统应用于多人的复杂场景, 需要对视频中出现的每一个行人目标进行追踪. 我们采用多目标追踪中比较成熟的 Deep-Sort[20]. Deep-Sort 算法[20] 主要有 4 个步骤: 数据输入、卡尔曼滤波、匈牙利匹配和输出, 如图 3 所示.

图 2 YOLOv4 目标检测示意图

卡尔曼滤波跟踪根据目标检测算法得到的前一帧某一个目标的目标信息, 并对其进行跟踪和预测, 得到该目标在后一帧的具体信息. 经过卡尔曼滤波器跟踪模块处理后, 视频中每一帧图像中包含的信息不仅是经目标检测算法得到的目标信息, 而且包括跟踪算法得到的目标跟踪信息. 然后使用匈牙利匹配对两种信息进行匹配, 得出最终的检测跟踪结果, 从而避免跟踪目标被多次检测, 降低算法的性能. 同时算法在匹配中还引入了级联匹配, 让更常见的目标分配的优先级更高, 更能应用于复杂的场景.

图 3 Deep-Sort 算法效果图

2.4 行人分割

在行人分割阶段, 我们使用 UNet++算法对从视频序列中检测得出的行人序列进行前后景分离, 从而去除衣服条纹以及街道背景等噪声对于识别准确的影响.

UNet++网络由一对完全对称的编码器和解码器构成, 并通过连接的方式, 将编码阶段获得的浅层特征映射同解码阶段获得的深层特征映射结合在一起, 细化图像, 根据得到的特征映射进行预测分割, 最后一层通过卷积做分类. 同时, 将每一层上的特征提取器进行相互的连接, 以达到对特征提取器进行共享的目的. 在训练过程中, UNet++网络可以自行学习得出哪一层的特征信息更为重要, 从而在特征的提取中有更好的表现, 具体表现在图像分割中对边缘的处理更为优异.

由于行人分割所得到的步态剪影序列图将作为特征提取网络的直接输入, 其分割效果以及效率将在很大程度上决定网络是否能够提取得到细粒度的步态特征以及系统能否对输入的视频进行高效的处理, 因此选用合适的算法能帮助系统提高身份及属性识别的准确率. 常见的前后景分离算法有基于机器学习的高斯混合模型算法[21] 以及基于 K 邻近的背景分割算法[22], 以及基于深度学习的 MaskRCNN[23]、CGNet[24]、UNet[25] 和 UNet++[26] 等.

图 4 展示了各种常见的分割算法的结果对比, 其中 MaskRCNN 和 UNet++分割结果更好.

(a) 源视频截图 (b) 使用 MaskRCNN 算法 (c) 使用 CGNet 算法 (d) 使用 UNet 算法 (e) 使用 UNet++算法

图 4 4 种语义分割算法得到的行人步态剪影图

平均交并比 (mIoU) 是指模型对每一类预测的结果和真实值的交集与并集的比值的平均值, 即结果越接近 1 效果越好. 表 1 中所展示的 mIoU 是在同等实验条件下进行测试的结果. 通过比较, 我们发现 UNet++在比 MaskRCNN 更少的参数量的情况下, 达到同等的分割效果. 所以, 结合图 4 语义分割的效果和表 1 算法效率之间的比较, 并考虑实际落地应用需要, 即在保持较好分割效果的同时应保持分割模型的轻量化. 本文选择了轻量化且分割效果较好的 UNet++模型作为我们的分割算法.

表 1 不同分割模型参数量以及平均交并比数据

算法	模型参数量 (Param)	平均交并 (mIoU)
MaskRCNN	63 744 170	0.863
CGNet	6 918 388	0.855
UNet	7 763 041	0.843
UNet++	9 042 175	0.861

2.5 特征提取算法模型

在算法模型方面, 本系统参考 Fan 等人[27] 提出的网络架构, 并在其基础上针对不同识别任务的需要进行模型的优化和改良, 最终在不同的识别任务上达到了相对优异的准确率. 本系统所用于模型训练的特征

提取网络结构如图 5 所示.

Fan 等人[27] 提出的特征提取网络的输入是一系列连续的步态图像 (64 像素×44 像素), 首先, 将包含 t 帧的步态轮廓序列逐帧输入网络. 在网络中, 首先对输入的步态图像进行处理的模块是帧级部分特征提取器 (frame-level part feature extractor, FPFE), 这是一种特殊设计的卷积网络, 用于挖掘行人步态中蕴含的局部细粒度信息, 并输出每帧步态图像f_i的空间特征$F_i, i \in 1,2,\cdots,t$.

FPFE 由 3 个块组成, 每个块由两层焦点卷积层 (FConv) 组成, 目的是提取每帧的部分信息空间特征. FConv (focal convolution layers) 是卷积的一种新应用, 它可以将输入的特征图从上到下分割成若干部分, 然后对每一部分分别进行卷积, 最后水平维度上拼接各个分割出来的子模块成为一个整体模块.

通过 FPFE 模块后的特征图序列记为$SF = \{F_i | i = 1,\cdots,t\}$, 接着, 我们将其输入到水平池化 HP 模块. HP 模块以提取人体不同部分的信息特征为目标, 将特征图F_i水平分割为 n 个部分. 对于F_i的第 j 部分$F_{j,i}$, HP 模块通过全局平均池化和全局最大池化将其向下采样到列向量$P_{j,i}$中, 并将其作为中间结果:

$$P_{j,i} = Avgpool2d(F_{j,i}) + Maxpool2d(F_{j,i})$$

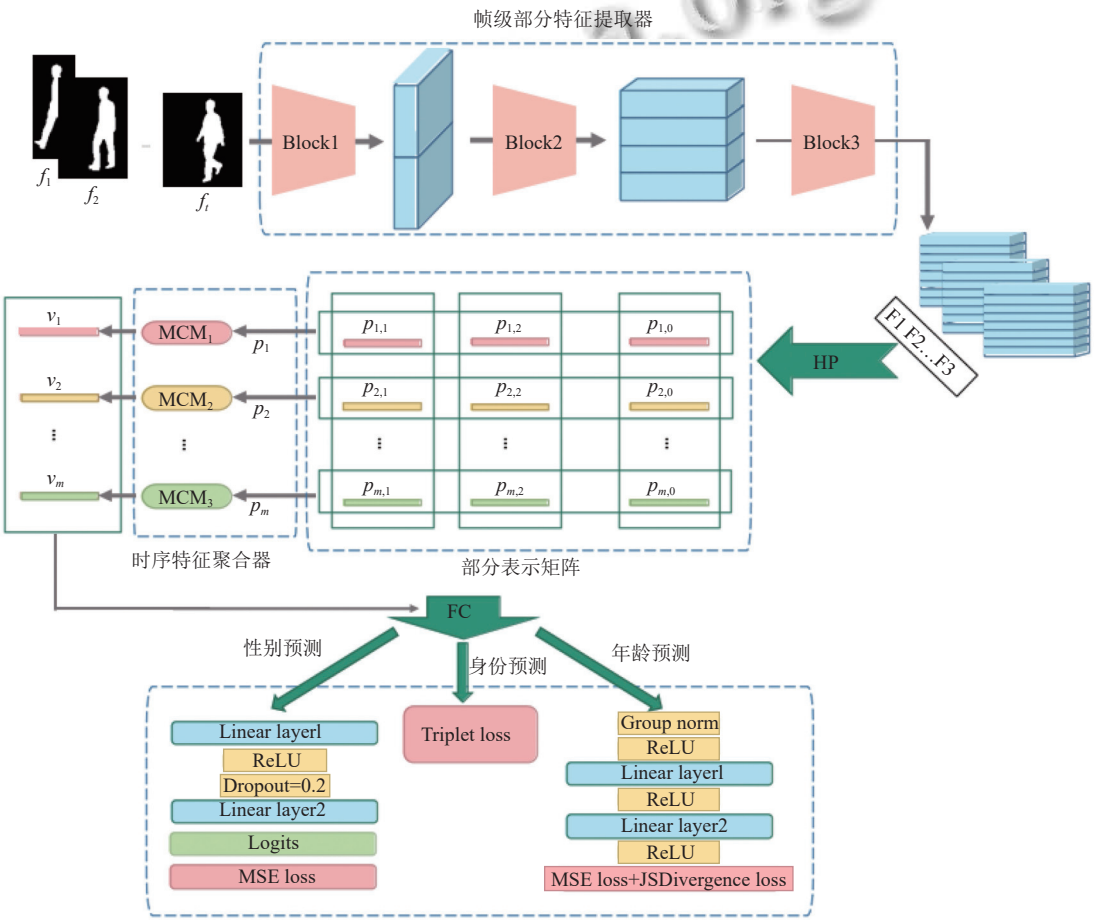

图 5 系统整体算法模型图

经过 HP 模块后, SF 中的每个特征映射将转换为 n 个部分级特征向量, 由这些特征向量得到部分表示矩阵 (part representation matrix, PR-Matrix), 我们将 PR-Matrix 记为:

$$P = (P_{j,i})_{n \times t}$$

PR-Matrix 中相应的向量行可以记为:

$$P_{j,.} = \{P_{j,i} | i = 1,2,\cdots,t\}$$

其中, $P_{j,i}$代表人体第 j 部分, 第 i 个时刻的步态特征. 因此, $P_{j,.}$可以表示人体第 j 个部分的时空运动表示. 我们将$P_{j,.}$通过微动作捕捉模块 (MCM) 聚合到特征向量$v_{j,.}$中, 从而可以提取出第 j 部分的微运动特征, 上述过程用公式表示为:

$$v_{j,.} = MCM(P_{j,.})$$

其中, MCM_j为第 j 个微动作捕捉模块. MCM 的作用在于将经过 HP 模块输出后的帧级部分特征向量映射为微运动特征向量. MCM 包括两个部分: 分解动作模板编辑器 (MTB) 和时间池化 (TP). MCM 模块工作流程图如图 6 所示. 接下来将首先对 MTB 模块进行描述, 然后是 TP 模块.

在 MTB 模块中, 设$Sp = \{P_i|i = 1,2,\cdots,t\}$是 PR-Matrix 的某一行, 代表人体特定部分的步态时空表示. MTB 将尺寸为 2r+1 的一维全局平均池化和一维全局最大池化应用于S_p的每个时刻, 从而得到分解动作特征向量序列S_m, 上述过程用公式表示为:

$$S_m = Avgpool1d(S_p) + Maxpool1d(S_p)$$

为了获得对分解动作更有鉴别性的表示, MTB 引入了通道注意力机制来对每个时刻的特征向量重定权重, 该过程采用了一维卷积核进行权重分配, 重新加权后的微运动分量S_m^{re}的表达式为:

$$S_{\text{logits}} = Conv1dNet(S_p)$$

$$S_m^{\text{re}} = S_m \cdot Sigmoid(S_{\text{logits}})$$

其中, $S_m^{\text{re}} = \{m_i^{\text{re}}|i = 1,2,\cdots,t\}, m_i^{\text{re}}$表示每个时刻的微运动特征表示.

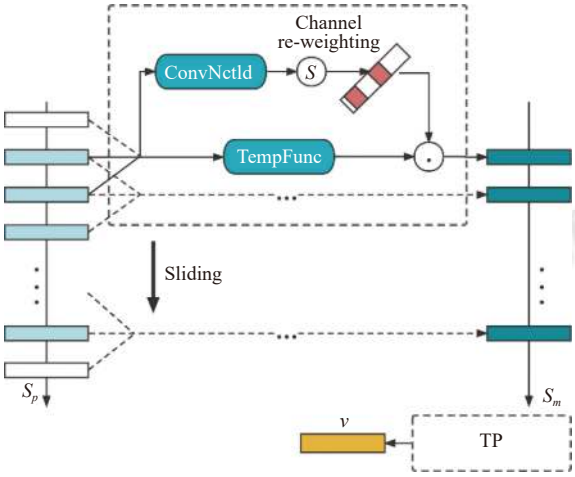

图 6 MCM 模块工作流程图

TP 模块的作用是对 MTB 模块输出的微动作特征向量序列$S_m^{\text{re}} = \{m_i^{\text{re}}|i = 1,2,\cdots,t\}$进行时间聚合, 时间聚合的方式基于最大值池化的序列维度统计函数:

$$TP(S_m^{\text{re}}) = \max(m_1^{\text{re}},\cdots,m_t^{\text{re}})$$

通过 TP 模块之后, 我们会得到 MCM 模块最终输出的特征向量$v_{j,.}$

$$v_{j,.} = TP(S_m^{\text{re}}(t))$$

$v_{j,.}$表征了人体特定部分的微运动特征. 而人体各个部分的完整微运动特征则需要各个部分并行的 MCM 模块进行共同表征, 各个 MCM 模块共同构成了时间聚合器 (temporal feature aggregator, TFA) 模块, TFA 的输出为$v_{\text{final}} = \{v_{j,.}|j = 1,2,\cdots,n\}$.

最后, 我们使用独立的 FC 层将v_{final}映射到度量空间得到向量v_{fc}. 对于身份识别任务, 我们使用三元组损失函数对v_{fc}进行身份识别任务的训练.

对于年龄和性别识别任务, 由于 batch 的大小对年龄的估计结果有较显著的影响, 较大的 batch 所消耗的显存剧增, 而较小的 batch 不利于模型突破局部最优值. 因此我们使用 group normalization 归一化方式, 避免 batch 的大小对数据处理产生显著影响. 在性别分类中, 考虑到二分类问题对于复杂的神经网络而言学习难度较小, 可能会出现在数据集上出现过拟合的现象, 因此我们在线性层特别使用了 dropput 技术. 实验表明, 设定合适概率的 dropput 概率既能给予参数更活跃的搜索空间, 也能使模型在测试时泛化性更强. 两个子任务中对网络的特殊处理有利于增强模型的非线性表达能力.

我们在 GaitPart 网络的末端添加了若干线性层以及 Dropout 层对v_{fc}进行进一步的特征映射, 对于性别任务, 我们使用平均绝对误差作为损失函数; 而对于年龄任务, 我们则联合平均绝对误差和 JS 散度作为损失函数进行训练. 各个部分的实验均采用 Adam 作为训练的优化器以及采用 LeakyReLU/ReLU 作为激活函数.

3 实验结果

3.1 系统设计和运行环境

本系统基于“MVC”的设计思想进行设计, 开发出一款集成步态识别算法的 Windows 电脑客户端程序. 程序的交互界面通过 PyQt 进行构建. 本系统成功在 Windows 64 位操作系统上运行与测试. 客户端程序交互界面使用 QtDesigner 完成设计开发, 算法部分使用 Python 语言编译实现, 在恒源云平台上进行训练优化. 硬件层面, 本系统运行和测试的电脑系统为 Windows 10 家庭版, CPU 为 i5-8265U, 内存 8 GB, 显卡为 MX250.

3.2 数据集介绍

3.2.1 CASIA-B 跨视角步态数据集

CASIA-B 是目前最为主流的跨视图步态数据库之一. 它包括 124 个行人样本, 每个行人样本有 10 种行

走状态. 其中, 有 6 组是处于正常步行状态 (NM), 两组是处于携带背包行走状态 (BG), 其余的处于穿着外套步行状态 (CL). 每种行走状态包含 11 个不同角度的步态序列 (视角的取值 0°– 180°, 采样间隔为 18°). 因此, 整个数据集包含有 124 (行人样本) × 10 (行走状态) × 11 (视角) = 13640 个步态序列. 由于本系统面向跨视角的应用识别场景, 因此在本文涉及到的算法模型的训练中, 按照主流的测试协议使用了 CASIA-B 中全部视角的数据进行模型的训练和测试.

3.2.2 OUMVLP 大规模跨视角步态数据集

OU-MVLP 步态数据集是大阪大学发布的, 目前为止世界上最大的并且具有广泛视野的步态数据集. 该数据集中涵盖了 10 307 名受试者, 每个受试者每个角度的一段序列作为标签已知的匹配库 (gallery set) 样本, 另一段序列作为标签未知的待识别 (probe set) 样本. 其中, 男性 5 114 名, 女性 5 193 名, 年龄从 2 到 87 岁不等. 每个受试者将会从 14 个视角进行捕获, 范围在 0°–90°、180°–270°之间, 每 15°为一个分隔. 因此, 整个数据集包含有 10 307 (受试者样本)×14 (视角)×2 (序列数)=288 596 个步态序列.

目前在 OU-MVLP 上主流的测试协议是将 10 307 个受试者样本划分为训练集和验证集, 其中训练集包含前 5 153 个受试者, 测试集包含后 5 154 个受试者. 在训练阶段, 来自所有受试者的 gallery 和 probe 序列的图像将同时用于模型训练; 在测试阶段, 仅使用 gallery 序列独立评估模型表现, 从中随机选择单个帧作为网络输入. 由于本系统面向跨视角的应用识别场景, 并且 OU-MVLP 数据集拥有年龄与性别标记, 因此本文中的步态年龄与性别估计算法将使用 OU-MVLP 步态数据集中的全视角的数据集进行训练.

3.3 算法识别效果

基于步态的身份识别算法模型实验在 CASIA-B 数据集进行训练和测试. 在我们实验的训练阶段, 按照 CASIA-B 上主流的测试协议使用了数据集前 74 个受试者 (ID: 001–074) 的样本数据作为训练集进行训练, 然后将剩余的 50 名受试者 (ID: 075–124) 将作为测试集进行检验. 在训练阶段, 训练集的每个行人输入网络的步态序列长度为 30. 在测试阶段, 我们按照主流协议将测试集拆分为标签已知的匹配库 (gallery set) 和标签未知的待识别 (probe set) 两个部分. 其中标签已知的匹配库包含测试集中所有 ID 行人在 NM 的前 4 组步态序列. 标签已知的匹配库样本通过对测试集行人的步态序列进行身份注册以便后续对比. 标签未知的待识别样本为待查询的集合, 其组成为测试集中所有 ID 行人在其余状态下的序列, 即剩余的 2 组 NM、2 组 CL 和 2 组 BG 的序列集合. 标签未知的待识别样本中的每个个体通过和标签已知的匹配库样本中的所有个体进行逐一对比以评估模型的准确率. 在测试时, 我们对算法模型在每个视角的表现进行独立的评估, 如表 2 中的逐列准确率数据展示的是标签未知的待识别样本中从 0°到 180°的每个查询视角和标签已知的匹配库样本中从 0°到 180°中所有视角 (除去查询视角) 的跨视角识别准确率的平均值. 在测试过程中, 完整的步态序列输入到模型中以提取步态特征. 实验中输入的批大小为 64, 算法模型的训练总共经过 8 000 次迭代, 训练全过程的学习率固定在 1E–4.

本文所采用的算法模型在中科院的 CASIA-B 数据集的最终的训练结果如表 2 所示, 通过比较表明我们的方法优于 Chao 等人[15] 和 Zhang 等人[14] 提出的步态识别算法, 在跨视角的情况且处于多种行走状态下 (NM: 95.9%, BG: 91.2%, CL: 77.0%) 下身份识别的准确率可以达到 88.0%, 说明目前我们的算法模型已经能够捕获到判别性的身份特征并处于相对领先的水平.

对于性别分类与年龄估计问题, 我们采用了和身份识别任务相同的步态特征提取器, 并在此基础上自行设计如图所示的分类器, 将高维空间向量映射至二维输出向量. 其中, 性别分类与年龄估计训练过程如图 7 所示.

在性别分类上, 我们采用性别预测准确率作为我们的评判标准. 图 7 中“train_acc”表示在训练集上的平均分类准确率, “val_acc”表示在验证集上的平均分类准确率. 经过不少于 45 000 次的迭代训练后, 算法基本达到收敛状态, 且在验证集中的最高识别率可达 95.2%, 泛化至测试集后识别率仍可保持在 94.8% 的水平. 该算法在性别分类问题上已获取明晰的判别性表示和鲁棒的决策边界.

在年龄评估上我们采用平均绝对误差 (mean absolute error, *MAE*) 来评估估计年龄的准确性. 假设$\widehat{y_i}$和y_i分别表示第 i 个测试样本的估计年龄和真实年龄, N_S表示测试的样本个数, *MAE* 将被计算为:

$$MAE = \frac{1}{N_S}\sum_{i=1}^{N_S}|\widehat{y_i} - y_i|$$

表 2 算法模型 CASIA-B 数据集的 Rank-1 识别准确率 (不包括相同视角)(%)

Gallery NM#1–4, Probe	Method	0°–180°										Mean
		0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	
NM#5–6	GaitNet[14]	91.2	92.0	90.5	95.6	86.9	92.6	93.5	96.0	90.9	88.8	89.0	91.6
	GaitSet[15]	90.8	97.9	99.4	96.9	93.6	91.7	95.0	97.8	98.9	96.8	85.8	95.0
	Flex-Gait[16]	91.1	97.9	99.6	97.3	94.3	91.9	94.9	98.1	98.8	96.2	86.6	95.1
	PartialRNN[17]	91.1	98.0	99.4	98.2	93.2	91.9	95.2	98.3	98.4	95.7	87.5	95.2
	SCN[18]	89.7	98.5	99.8	97.9	94.4	91.2	94.5	97.1	97.6	97.0	89.4	95.2
	Ours	92.5	97.7	98.8	97.6	94.0	92.4	95.8	98.3	99.3	97.9	90.3	**95.9**
BG#1–2	GaitNet[14]	83.0	87.8	88.3	93.3	82.6	74.8	89.5	91.0	86.1	81.2	85.6	85.7
	GaitSet[15]	83.8	91.2	91.8	88.8	83.3	81.0	84.1	90.0	92.2	94.4	79.0	87.0
	Flex-Gait[16]	84.3	91.2	93.4	91.8	86.1	80.3	84.4	90.9	93.7	90.8	80.1	87.9
	PartialRNN[17]	86.0	93.3	95.1	92.1	88.0	82.3	87.0	94.2	95.9	90.7	82.4	89.7
	SCN[18]	86.7	94.6	96.0	92.5	85.8	80.5	84.9	91.5	96.0	93.1	86.0	89.8
	Ours	87.4	94.6	95.6	92.9	89.1	85.1	89.3	93.8	97.0	94.0	84.0	**91.2**
CL#1-2	GaitNet[14]	42.1	58.2	65.1	70.7	68.0	70.6	65.3	69.4	51.5	50.1	36.6	58.9
	GaitSet[15]	61.4	75.4	80.7	77.3	72.1	70.1	71.5	73.5	73.5	68.4	50.0	70.4
	Flex-Gait[16]	64.7	79.4	84.1	80.4	73.7	72.3	75.0	78.5	77.9	71.2	57.0	74.0
	PartialRNN[17]	65.8	80.7	82.5	81.1	72.7	71.5	74.3	74.6	78.7	75.8	64.4	74.7
	SCN[18]	63.7	79.2	82.3	77.7	69.4	71.5	73.5	77.9	78.4	76.5	62.4	73.9
	Ours	70.9	81.5	84.6	81.4	73.2	71.5	77.0	81.6	81.4	78.8	65.5	**77.0**

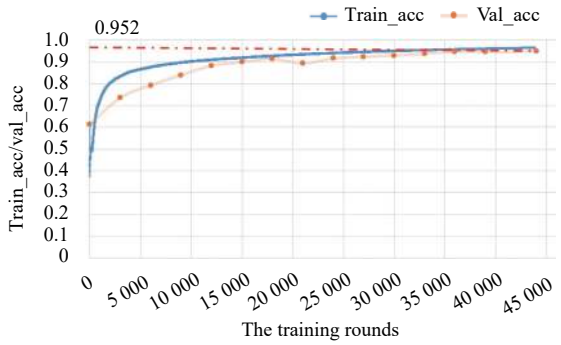

图 7 训练集和验证集上的性别识别率

图 8 中“train_mae”表示在模型在训练集上的平均年龄误差数据, “val_mae”则表示在验证集上的平均年龄误差. 由图可知, 在经过不少于 45 000 次的迭代训练后, 训练集上的平均绝对误差逐渐下降, 并且有不断下降的趋势, 而在验证集上的平均年龄误差并没有随着训练而有明显的下降趋势, 这说明模型已经有效收敛. 算法在验证集中的最低平均年龄误差可达 7.63 岁, 而泛化至测试集后平均年龄误差仍可以保持在 7.92 岁的水平.

为佐证本文在性别分类与年龄估计方面的实验效果, 我们将与 Xu 等人[28] 在 2021 年提出的方法进行对比和分析. 如表 3 所示, 其中, “GaitSet-Based CNN Framework”是利用 PA-GCR (phase-aware gait cycle reconstructor) 重建步态周期、利用 GaitSet 网络提取步态特征的方法, 其跨视角年龄平均分类正确率为 94.3%、平均年龄误差为 8.39 岁. 相比之下, 本文对 14 个视角下行人的性别预测精度达到了 94.8%、年龄估计达到了 7.92 岁, 算法效果更优, 处于当下前沿领域的高水平层级.

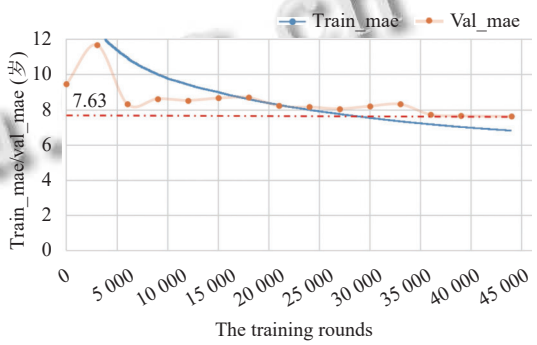

图 8 训练集和验证集上的年龄估计平均年龄误差

表 3 基于步态识别的性别预测结果对比

方法	平均跨视角识别率 (%)	平均年龄误差 (岁)
GaitSet-Based CNN Framework	94.3	8.39
本文方法	**94.8**	**7.92**

3.4 在线系统测试

本系统将在模拟安防情景下进行相应的测试. 在测试中, 我们预先录制了多段模拟的“底库视频”及两

段模拟的“案发现场视频”, 前者模拟嫌疑人可能逃离的路段视频, 后者则模拟案发路段附近所调取的包含目标嫌疑人的视频. 本系统将对“底库视频”中出现的所有行人步态信息在本地数据库中进行注册和录入, 用于与后续“案发现场视频”中的目标嫌疑人员进行比对, 从而确定最终嫌疑人的逃离路段.

图 9 和图 10 是本系统对预先录制好的步态视频进行测试的结果. 图 10 展示的是系统选择嫌疑人 ID 和筛选年龄条件功能. 其中, 图 9 为本系统的主页面, 用于上传底库和案发现场视频, 并通过系统对比找出底库中步态信息与被选择的嫌疑人最为相似的前 3 位行人, 返回相应的步态年龄、性别以及所处路段.

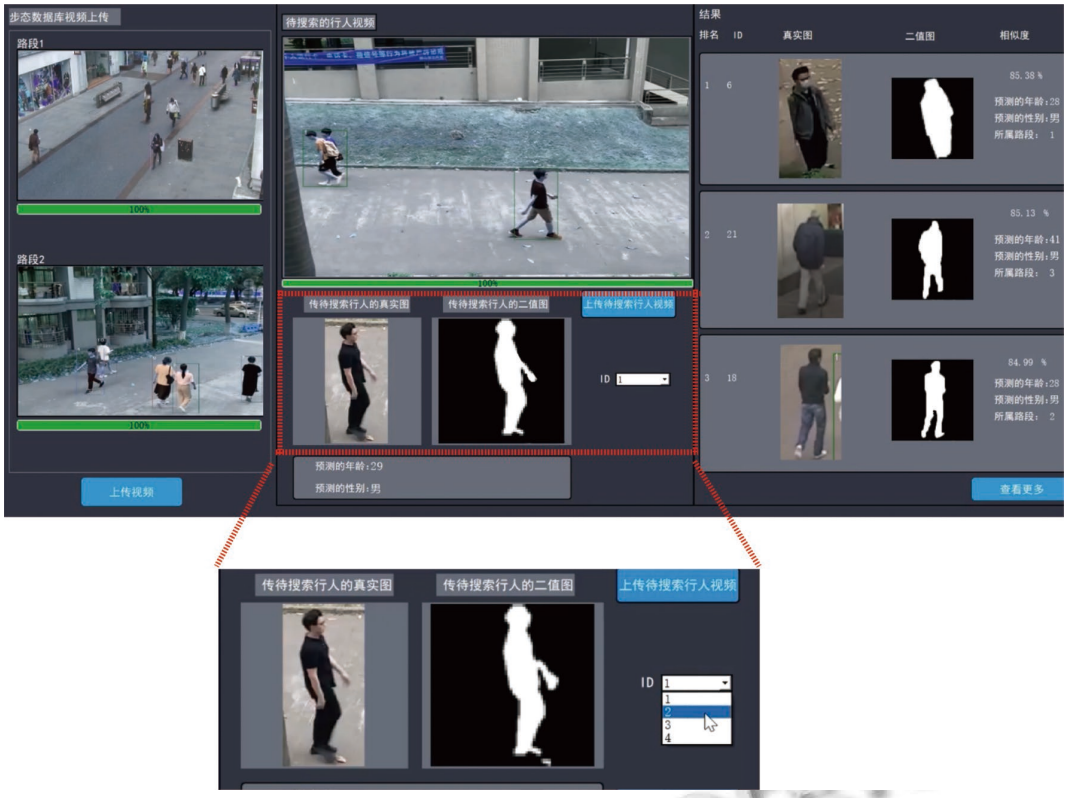

图 9 系统主页面和 ID 选择

在测试中, 我们输入了 7 段测试视频, 视频中共有受试者 229 名, 年龄均为 15–63 岁, 男女分布均等. 测试结果为: 身份匹配 Rank-1 准确率为 81.93%, Rank-5 准确率为 94.16%, 年龄预测平均年龄误差为 7.5 岁, 性别识别准确率为 91.63%. 从测试结果可以看出, 当识别视角发生改变且嫌疑人通过戴口罩进行面部遮挡并更换着装再次出现在已录入底库中的视频中时, 系统也可以准确地对嫌疑人进行识别, 证明了本系统具备较好的鲁棒性, 可以满足实际应用中复杂的场景需求.

在图 10 展示更多检测结果页面中, 系统将展示所有在底库视频中出现的行人步态信息与嫌疑人步态信息的对比结果, 并且具备筛选功能: 通过嫌疑人的年龄以及性别信息对海量的行人信息进行进一步的筛选, 以辅助应对一些由于特殊情况导致的身份识别不准确的情况, 从而加快查找速度.

通过常规的外置摄像头或者安防摄像头, 本系统可以实时捕捉行人的步态信息. 由于算法模型已具有预训练权重并且配备好相应的环境依赖项, 因此当用户上传步态视频后, 系统可以快速地返回相应的结果. 上述测试场景中, 在 1080Ti 的 GPU 算力环境下, 系统处理视频的速度为每帧 0.373 s, 且可在 2 s 内返回对比结果, 因此可以满足实际场景中实时监测的需要, 具有实际的开发意义.

4 总结

本文搭建了一种可以在视频监控条件下实现多行人多生物属性跨视角步态识别的智能安防追踪系统, 该系统适用于安防和寻人等实际应用场景, 可以追踪

不同路段处在不同视角处于不同行走状态的行人步态信息并通过系统内置的算法模型对其身份、年龄和性别等属性进行准确高效的分析. 本系统的算法模型基于深度学习理论及算法, 在 Fan 等人[27] 提出的算法基础上加以优化, 同时在目前最为主流的两个大规模跨视角步态数据集——中科院的 CASIA-B 数据集以及日本大阪大学的 OUMVLP 数据集上进行了训练和测试并取得了处于相对高水平的准确率, 在跨视角的测试前提下, 身份识别在多种行走状态下 (NM, BG, CL) 的平均准确率达到 88.0%, 性别识别准确率达到 94.8%, 年龄估计的平均误差在 7.92 岁, 基本符合实际应用的水准. 与此同时, 系统的开发成本低, 支持实时检测并具备拓展性, 可以根据不同团队的需要进行适应性调整, 因此具有重要的现实意义.

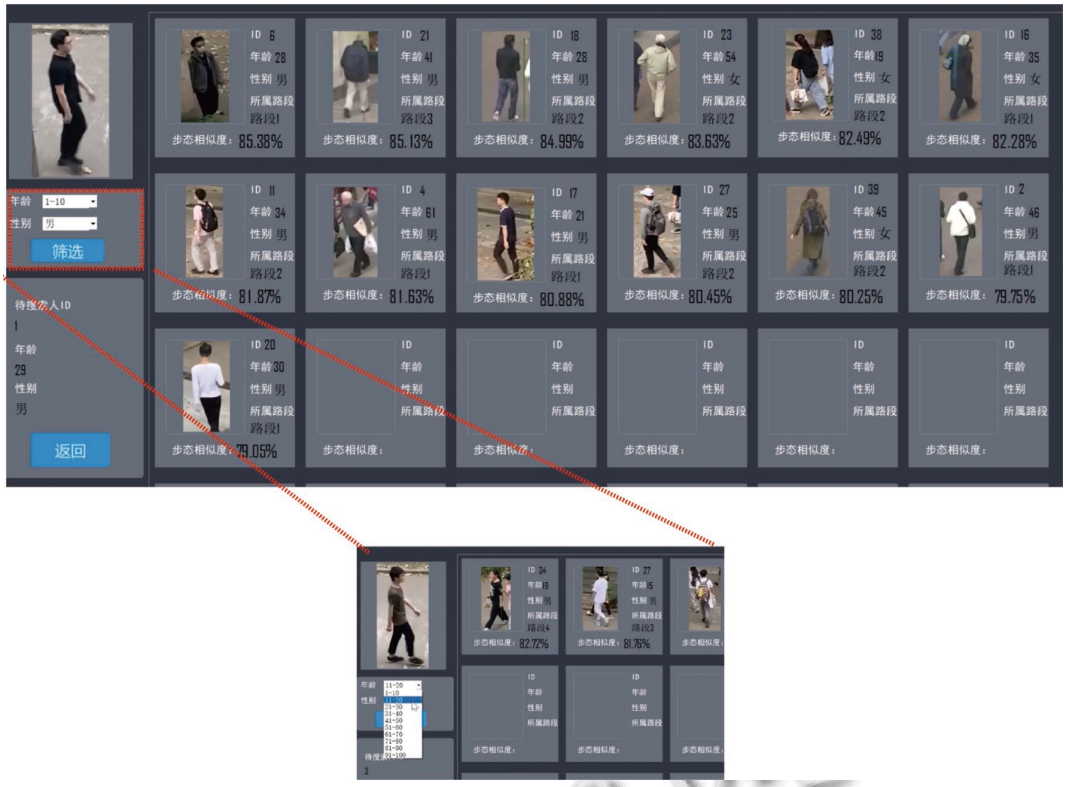

图 10 展示更多检测结果页面和条件筛选

参考文献

1 宁洪伟, 刘胜. 步态识别技术在身份认证中的应用研究. 云南警官学院学报, 2021, (3): 117–121. [doi: 10.3969/j.issn.1672-6057.2021.03.021]

2 张诚. 基于深度学习的步态识别关键技术研究 [硕士学位论文]. 北京: 北京邮电大学, 2016.

3 Kusakunniran W, Wu Q, Zhang J, *et al*. A new view-invariant feature for cross-view gait recognition. IEEE Transactions on Information Forensics and Security, 2013, 8(10): 1642–1653. [doi: 10.1109/TIFS.2013.2252342]

4 Kusakunniran W, Wu Q, Li HD, *et al*. Multiple views gait recognition using view transformation model based on optimized gait energy image. Proceedings of the 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops. Kyoto: IEEE, 2009. 1058–1064.

5 Jean F, Bergevin R, Albu AB. Computing and evaluating view-normalized body part trajectories. Image and Vision Computing, 2009, 27(9): 1272–1284. [doi: 10.1016/j.imavis.2008.11.009]

6 Kusakunniran W, Wu Q, Zhang J, *et al*. Recognizing gaits across views through correlated motion co-clustering. IEEE Transactions on Image Processing, 2014, 23(2): 696–709. [doi: 10.1109/TIP.2013.2294552]

7 Liu NN, Lu JW, Tan YP. Joint subspace learning for view-invariant gait recognition. IEEE Signal Processing Letters, 2011, 18(7): 431–434. [doi: 10.1109/LSP.2011.2157143]

8 Makihara Y, Sagawa R, Mukaigawa Y, *et al*. Gait

recognition using a view transformation model in the frequency domain. Proceedings of the 9th European Conference on Computer Vision. Graz: Springer, 2006. 151–163.

9 Muramatsu D, Shiraishi A, Makihara Y, *et al*. Gait-based person recognition using arbitrary view transformation model. IEEE Transactions on Image Processing, 2015, 24(1): 140–154. [doi: 10.1109/TIP.2014.2371335]

10 Goffredo M, Bouchrika I, Carter JN, *et al*. Self-calibrating view-invariant gait biometrics. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2010, 40(4): 997–1008. [doi: 10.1109/TSMCB.2009.2031091]

11 Han J, Bhanu B, Roy-Chowdhury AK. A study on view insensitive gait recognition. Proceedings of IEEE International Conference on Image Processing 2005. Genova: IEEE, 2005. III–297.

12 Wu ZF, Huang YZ, Wang L, *et al*. A comprehensive study on cross-view gait based human identification with deep CNNs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(2): 209–226. [doi: 10.1109/TPAMI.2016.2545669]

13 Shiraga K, Makihara Y, Muramatsu D, *et al*. GEINet: View-invariant gait recognition using a convolutional neural network. Proceedings of 2016 International Conference on Biometrics. Halmstad: IEEE, 2016. 1–8.

14 Zhang ZY, Tran L, Yin X, *et al*. Gait recognition via disentangled representation learning. Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 4705–4714.

15 Chao HQ, He YW, Zhang JP, *et al*. GaitSet: Regarding gait as a set for cross-view gait recognition. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 8126–8133.

16 Huang GH, Lu Z, Pun CM, *et al*. Flexible gait recognition based on flow regulation of local features between key frames. IEEE Access, 2020, 8: 75381–75392. [doi: 10.1109/ACCESS.2020.2986554]

17 Sepas-Moghaddam A, Etemad A. View-invariant gait recognition with attentive recurrent learning of partial representations. IEEE Transactions on Biometrics, Behavior, and Identity Science, 2021, 3(1): 124–137. [doi: 10.1109/TBIOM.2020.3031470]

18 Ding XN, Wang KJ, Wang CH, *et al*. Sequential convolutional network for behavioral pattern extraction in gait recognition. Neurocomputing, 2021, 463: 411–421. [doi: 10.1016/j.neucom.2021.08.054]

19 Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. arXiv: 2004.10934, 2020.

20 Wojke N, Bewley A, Paulus D. Simple online and realtime tracking with a deep association metric. Proceedings of 2017 IEEE International Conference on Image Processing. Beijing: IEEE, 2017. 3645–3649.

21 Kaewtrakulpong P, Bowden R. An improved adaptive background mixture model for real-time tracking with shadow detection. In: Remagnino P, Jones GA, Paragios N, *et al*. eds. Video-based Surveillance Systems. Boston: Springer, 2002. 135–144.

22 Zivkovic Z, van der Heijden F. Efficient adaptive density estimation per image pixel for the task of background subtraction. Pattern Recognition Letters, 2006, 27(7): 773–780. [doi: 10.1016/j.patrec.2005.11.005]

23 Li MX, Sun YE, Huang H, *et al*. A flexible resource allocation mechanism with performance guarantee in cloud computing. Proceedings of the 2018 4th International Conference on Big Data Computing and Communications. Chicago: IEEE, 2018. 181–188.

24 Wu TY, Tang S, Zhang R, *et al*. CGNet: A light-weight context guided network for semantic segmentation. IEEE Transactions on Image Processing, 2021, 30: 1169–1179. [doi: 10.1109/TIP.2020.3042065]

25 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. Proceedings of the 18th International Conference on Medical Image Computing and Computer-assisted Intervention. Munich: Springer, 2015. 234–241.

26 Zhou ZW, Siddiquee MR, Tajbakhsh N, *et al*. UNet++: A nested U-Net architecture for medical image segmentation. Proceedings of the 4th International Workshop Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Granada: Springer, 2018. 3–11.

27 Fan C, Peng YJ, Cao CS, *et al*. GaitPart: Temporal part-based model for gait recognition. Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 14213–14221.

28 Xu C, Makihara Y, Liao RC, *et al*. Real-time gait-based age estimation and gender classification from a single image. Proceedings of 2021 IEEE Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2021. 3459–3469.

(校对责编: 孙君艳)