

前 言

随着信息技术的飞速发展，人工智能正以前所未有的速度改变着世界。在人工智能的众多应用研究分支中，智能视频分析技术凭借其强大的数据处理与模式识别能力，成为连接物理世界与数字世界的桥梁。特别是在公共安全、人机交互、医疗健康等领域，智能步态视频分析与识别技术以其独特的优势，展现出广阔的应用前景和深远的社会价值。

本书正是在智能视频分析技术迅速崛起的背景下，为了系统梳理和深入探讨该领域的基本原理、关键技术、前沿进展及实际应用而撰写的。本书旨在为相关领域的研究者、工程师、学者及学生提供一个全面、深入、前沿的知识平台，助力其更好地把握智能步态视频分析与识别技术的核心精髓和发展趋势。

步态，作为人类行走过程中的一种自然生物特征，蕴含着丰富的个体信息。步态识别技术可以对步态进行精准捕捉和深入分析，进而实现对个体身份的识别、行为状态的监测和异常步态的预警。这种技术不仅具有非接触性、远距离、难伪装等优点，还能够在复杂多变的场景中保持较高的识别准确率和鲁棒性，因此具有极高的实用价值和研究意义。

本书结构严谨，逻辑清晰，从基础理论到关键技术，再到前沿进展和实际应用，层层递进，深入浅出。首先阐述基于视觉的步态识别的背景和意义，介绍步态识别常用数据集并综述国内外的研究现状，总结和分析当前步态识别的研究和应用中存在的问题和难点。然后针对端到端的步态识别、多样化的随机行走场景和时域步态特征提取中存在的难点问题，分别详细阐述基于语义分割和迁移学习的步态剪影分割、基于卷积神经网络的步态周期检测、基于非周期步态模板的步态识别、基于多步态模板融合及分区思想的步态识别、基于特征差分运动模板的步态识别、基于双向类光流动作表征的步态识别、基于局部时空卷积的步态识别、基于生成对抗网络的步态转换与识别、基于图像修复方法的步态图像遮挡消除等研究内容，从不同角度优化步态识别算法，改善步态识别的适用范围、准确性和实时性等。书中不仅展现了各种算法的原理和实现方法，还通过具体的实验数据和结果分析，让读者更加直观地了解各种算法的性能和优缺点。

本书的完成凝聚了集体的智慧与辛勤付出，是作者研究团队多年研究成果的总结。其中，王科俊统筹全书，丁欣楠为本书的撰写和修改作出主要贡献，苏秉华、张连波和谢海军参与了部分写作和校稿工作，卢桂萍绘制了本书的部分图形，刘亮亮、贾睨烨和吕卓文等对本书的相关成果也作出了重要贡献。在此对所有为

本书撰写和出版作出贡献的人员致以诚挚的感谢。

智能步态视频分析与识别技术是一个涉及多学科交叉的复杂领域，需要不断学习和探索，因此诚挚地邀请广大读者对本书提出宝贵的意见和建议，共同推动该领域的研究和发展。最后希望本书的出版能够有助于智能步态视频分析与识别技术的进步，为相关领域的研究人员提供有益的参考和借鉴。

王科俊

2026 年 2 月

目 录

前言

第 1 章 绪论	1
1.1 步态数据集概述	3
1.2 研究现状概述	6
1.2.1 步态数据预处理方法	7
1.2.2 基于步态模板图像的方法	9
1.2.3 基于一组步态剪影图像的方法	10
1.2.4 基于图像生成转换的方法	11
1.2.5 基于步态分块的方法	13
1.2.6 基于动作的步态识别方法	13
1.3 步态识别面临的挑战	16
第 2 章 基于语义分割和迁移学习的步态剪影分割	20
2.1 主要思路与总体结构	22
2.2 基于注意力机制的语义分割	23
2.2.1 上采样中的注意力模块	23
2.2.2 特征融合中的注意力模块	25
2.2.3 基于 Xception 和残差连接的编码-解码结构	27
2.3 基于行人分割迁移学习的步态识别	29
2.3.1 基于语义分割和迁移学习的端到端步态识别	29
2.3.2 基于语义分割和迁移学习特征融合的端到端步态识别	30
2.4 实验结果与分析	31
2.4.1 基于注意力机制的语义分割网络实验结果与分析	31
2.4.2 基于行人分割迁移学习的步态识别实验结果与分析	36
2.5 本章小结	42
第 3 章 基于卷积神经网络的步态周期检测	44
3.1 卷积神经网络概述	45
3.2 基于分类的步态周期检测	47
3.2.1 周期分类建模	47
3.2.2 基于分类的卷积神经网络步态周期检测	48
3.2.3 采用的卷积神经网络结构	48

3.3	基于回归拟合的步态周期检测	49
3.3.1	步态周期的正弦函数建模	50
3.3.2	基于回归拟合的卷积神经网络步态周期检测	51
3.3.3	采用的卷积神经网络结构	51
3.4	实验结果与分析	52
3.4.1	基于分类的步态周期检测实验结果与分析	53
3.4.2	基于回归拟合的步态周期检测实验结果与分析	54
3.5	本章小结	60
第4章	基于非周期步态模板的步态识别	61
4.1	步态能量图概述	62
4.2	非周期的步态能量图方法	63
4.2.1	非周期的步态能量图方法流程	63
4.2.2	固定帧数的选取	64
4.3	实验结果与分析	67
4.3.1	基于欧几里得距离和最近邻分类器的步态识别实验结果与分析	68
4.3.2	基于 PCA 的步态识别实验结果与分析	69
4.3.3	基于神经网络的步态识别实验结果与分析	71
4.4	本章小结	74
第5章	基于多步态模板融合及分区思想的步态识别	75
5.1	步态模板	76
5.1.1	活动能量图	76
5.1.2	步态熵图像	76
5.2	基于多步态模板融合的步态识别	77
5.2.1	多模板彩色步态能量图构建框架	77
5.2.2	多模板彩色步态能量图	79
5.3	基于分区域的多步态模板融合步态识别方法	80
5.3.1	多模板彩色步态能量图的区域划分	81
5.3.2	基于分区域的多模板彩色步态能量图	82
5.4	实验结果与分析	84
5.4.1	基于多步态模板融合的步态识别实验结果与分析	84
5.4.2	基于分区域的多步态模板融合步态识别实验结果与分析	89
5.5	本章小结	93
第6章	基于特征差分运动模板的步态识别	95
6.1	主要思路与总体结构	96
6.2	基于运动模板的行为信息提取器	97

6.2.1	特征差分运动模板	98
6.2.2	运动模板嵌入方法	101
6.3	多帧聚合器	104
6.4	实验结果与分析	106
6.4.1	实验参数设置	106
6.4.2	行为信息提取器的作用	108
6.4.3	多帧聚合器的作用	109
6.4.4	整体实验	111
6.5	本章小结	112
第 7 章	基于双向类光流动作表征的步态识别	113
7.1	主要思路与总体结构	114
7.2	双向类光流动作表征	117
7.2.1	双边运动导向模块	117
7.2.2	多时空尺度特征扩展	120
7.3	分割误差校正模块	122
7.4	实验结果与分析	124
7.4.1	实验参数设置	124
7.4.2	双边运动导向模块的分析	126
7.4.3	多时空尺度特征扩展的分析	127
7.4.4	分割误差校正模块的分析	128
7.4.5	损失函数的分析	129
7.4.6	整体实验	130
7.5	本章小结	131
第 8 章	基于局部时空卷积的步态识别	132
8.1	主要思路与总体结构	133
8.2	局部时空卷积模块	134
8.2.1	全局双向空间池化	135
8.2.2	局部时空卷积层	137
8.3	局部时空池化方法	140
8.4	实验结果与分析	141
8.4.1	实验参数设置	142
8.4.2	全局双向空间池化的分析	143
8.4.3	局部时空卷积层的分析	144
8.4.4	局部时空池化的分析	145
8.4.5	整体实验	145

8.5 本章小结	147
第9章 基于生成对抗网络的步态转换与识别	148
9.1 步态转换方法的主要思路	148
9.2 步态转换网络的研究与设计	149
9.2.1 构建损失函数	150
9.2.2 生成器网络设计	151
9.2.3 判别器网络设计	155
9.3 步态转换实验过程及结果	156
9.3.1 实验参数设置	156
9.3.2 视角转换实验结果对比	157
9.3.3 形态转换实验结果对比	161
9.3.4 步态转换与识别实验结果对比	162
9.4 本章小结	167
第10章 基于图像修复方法的步态图像遮挡消除	169
10.1 基于相似性度量的步态图像遮挡检验	169
10.2 图像修复	170
10.2.1 传统图像修复方法	170
10.2.2 基于深度学习的图像修复方法	172
10.3 基于门控卷积边缘连接的步态图像遮挡消除模型	177
10.3.1 网络结构	177
10.3.2 注意力机制	180
10.3.3 卷积方式	186
10.4 实验结果与分析	190
10.4.1 遮挡条件下行人步态数据集构建	190
10.4.2 步态图像遮挡检测实验结果与分析	194
10.4.3 基于传统图像修复方法的步态图像遮挡消除实验结果与分析	195
10.4.4 基于深度学习的图像修复方法的步态图像遮挡消除实验结果与分析	198
10.4.5 基于门控卷积边缘连接的步态图像遮挡消除实验结果与分析	200
10.4.6 基于遮挡还原能量图的步态识别实验结果与分析	205
10.5 本章小结	210
参考文献	211

第1章 绪 论

随着科技的发展和应用场景的扩大，生物特征识别技术正逐渐成为未来安全认证领域的主流技术之一，因为它可以利用每个个体固有的生理或行为特征来进行身份验证或识别，所以具有可复制性低、精度高和应用便捷等优点。与密码和身份证等传统身份认证方式相比，生物特征更难被伪造、盗用或遗忘，故而能够有效地防止身份欺骗和信息泄露，也使得身份认证的过程更加方便和自然^[1,2]。2017年7月，国务院颁布了《新一代人工智能发展规划》，其中明确指出了研发生物特征识别技术的智能安防与警用产品以及利用人工智能来提升公共安全保障能力的发展方向。

智能视频分析技术在公共安全、疫情防控、医疗康复等多个领域展现出巨大的应用潜力。其中，步态识别作为一种新兴的生物特征识别技术，因其非接触、远距离和难伪装的特点，逐渐成为研究热点。步态识别旨在通过人的行走习惯方式进行个体的身份识别，是生物特征识别的一种方式^[3]。电影《碟中谍5：神秘国度》中具象地展现了步态识别未来可能的应用场景。如图1.1所示，由汤姆·克鲁斯饰演的伊森需要进入一家高度安保的数据中心时，被要求在一个狭窄的通道中行走，只有检测并验证为许可人员的步态才能够进入，这也极大程度阻碍了其间谍任务的完成。这一场景生动地展示了步态识别技术在安全领域的应用。在实际应用中，步态识别技术可以通过视频监控或者其他传感器来获取人的步态数据，并通过算法对数据进行处理和分析，从而实现身份验证或识别。



图 1.1 电影中的步态识别系统

基于视觉的步态识别因其固有的优势而具有独特的应用场景^[4,5]。首先,步态识别是一种非侵入式的远距离生物识别技术,这意味着不需要身体接触或使用特定的设备^[6]。与指纹或虹膜等其他生物识别技术相比,步态识别是一种更加用户友好和容易接受的选择。具体而言,人们只需要走路就可以完成身份识别,无须特别配合,也不会影响日常生活^[7]。例如,步态识别技术可以与其他智能设备结合,实现更智能的家居或城市管理,根据家庭成员的步态识别结果,智能家居可以自动调节灯光、温度和音乐等。此外,基于步态的身份识别技术可以为人们提供更好的身份识别体验,同时也可以有效保护个人隐私,对用户来说更便利。由于不需要被识别人员做任何特定的姿势或动作,步态识别在应用中并不突兀,可以用于帮助执法机构在嫌疑人或罪犯不知情的情况下追踪他们的行动,且特别适用于不能对其进行跟随的情况^[8]。另外,步态识别可以使用相对便宜且被广泛布置的标准监控录像设备来实现,成本效益高,因此它可以用于识别某个区域内的人员是否为授权人员,如机场、政府大楼和边防检查等高安全性区域中的个人身份认证,以确保只有授权人员才能进入敏感区域。此外,即使面部等其他识别特征不可见,步态识别也能够通过公共场所密集布置的摄像头来识别潜在的犯罪嫌疑人或恐怖分子,以增强安全性。另外,步态识别有较为宽泛的应用场景,它可以在各种环境下工作。与其他可能受光照、噪声或运动等因素影响的生物识别技术相比,它是一种鲁棒性更强的技术,故可以作为传统生物特征验证方法的一种补充,以提高身份验证的便利性,解决佩戴手套或口罩的情况下指纹或面部识别方法不起作用的问题。最重要的是,步态识别具有动态识别能力,可在人们行走的过程中实现身份识别,与静态技术相比,其更加适用于安全性要求高的场景。总之,基于视频的步态识别技术在公共安全、刑事侦查和智能家居等领域应用前景广阔。随着图像处理和机器学习等基础研究的不断进步,该技术也会越来越成熟并走向实际应用,为人们的生活带来更多的便利和安全保障。

然而,基于视觉的步态识别在实际应用中面临许多挑战,主要体现在人在行走过程中会不可控地受到外在环境和自身因素的影响^[9,10]。具体而言,由于监控摄像头布置地点的受限性和行人出现的随机性,待识别个体的行走方向、着装和携带物、行走路面、视频分辨率、行走速度和遮挡等的变化导致同一个行人的两段行走视频区别很大^[11,12]。尽管一些外部特征,如服装等,经常被用作跨摄像头的行人重新识别(person re-identification)任务的关键属性^[13],但它不具有唯一性和不变性。对于罪犯等试图故意隐藏身份的人,会出现严重的身份识别错误^[14,15]。因此,如何应对步态识别任务中类内特征的巨大差异一直是研究者致力解决的问题^[16]。步态识别研究最初可以追溯到20世纪80年代^[17],其间很多来自知名研究机构的研究者陆续开展了相关研究,如国外的麻省理工学院、南佛罗里达大学、大阪大学、马里兰大学帕克分校、南安普顿大学和昆士兰科技大学等,国内的中

国科学院自动化研究所、清华大学、复旦大学、山东大学、北京交通大学、哈尔滨工程大学和南方科技大学等。他们针对步态视频中存在的光照、角度、服饰和多人遮挡等干扰,不断地拓展步态识别的研究内容和发展方向^[18-20]。近年来又出现了一批科技公司尝试将步态识别技术落地,真正地将其推广并用于实际生活中,如银河水滴、大华股份和海康威视等。而步态识别系统的布置与应用又对科研领域的识别方法提出了新要求,除了识别准确率外,又期望对识别技术的工作条件、运行耗时、复杂程度和对数据的依赖性等进行了限制。因此,将步态识别真正用于实际的远距离身份判断中仍有许多难点尚未解决。

智能步态视频分析与识别技术是指通过步态特征完成身份验证和识别的技术。因为步态识别实质上需要提取视频中连续存在的行走习惯而完成身份识别,且时域特征不受服装和行走角度等条件变化的影响,所以直接提取时域特征可以得到本质的不受协变量干扰的步态特征。因此,本书围绕步态图像预处理中的步态分割、步态周期检测、步态识别和复杂(多角度和遮挡)条件下的步态识别,全面而系统地阐述智能步态视频分析与识别技术。

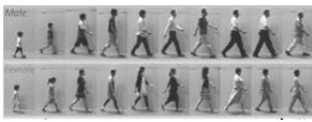
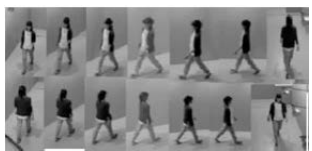
1.1 步态数据集概述

大规模步态数据集对于步态身份识别方法的开发训练、结果评估和性能比较至关重要,尤其是基于深度学习的方法十分依赖大规模高质量的已标注数据。相比于类似大小的面部或指纹数据集,收集大规模步态数据需要更多的时间精力、存储空间和人工成本。为了便于研究,目前国内外已有很多研究机构公开了若干步态数据集以满足多样化和复杂的应用需求,涵盖行走角度、环境条件、着装等与受试者外观变化相关的各种变量参数。包含大规模数据并覆盖各种变量的数据集是深度神经网络能够得到有效训练的必要保障,也是未来步态识别技术发展的重要基础。表 1.1 概述了现有的步态数据集,并展示了相应的步态数据示例。

表 1.1 基于视觉的步态数据集

名称	建立机构	行走条件	样本容量	示例
USF ^[21]	南佛罗里达大学	地面、鞋子、负重、时间和角度(左右两个摄像机前绕椭圆路线行走)	122 人 1870 序列	
HID-UMD ^[22]	马里兰大学帕克分校	角度: 正面(走向、走出)、侧面(向左、向右)	1: 25 人× 4 视角= 100 个序列	

续表

名称	建立机构	行走条件	样本容量	示例
HID-UMD ^[22]	马里兰大学 帕克分校	角度: T 形路径 (正面、侧面)	2: 55 人× 4(2 视角)= 220 个序列	
CMU MoBo ^[23]	卡耐基·梅隆 大学	6 个角度(360°圆周每 60°一个视角), 行走 条件: 抱球、上坡	25 人 600 个序列	
CASIA-A ^[24]	中国科学院 自动化研究所	3 个角度: 侧面的 3 个点	20 人×3 视 角×4 = 240 个序列	
CASIA-B ^[25]	中国科学院 自动化研究所	11 个角度, 行走条件: 正常、携包和穿外套	124 人 13640 个序列	
CASIA-E ^[26]	中国科学院 自动化研究所	室内和室外, 行走条件: 正常、携包和穿外套	1014 人	
OU-ISIR Treadmill ^[27]	大阪大学	25 个角度 (12 个方位角×2 个 倾斜角度+1 个俯视)	168 人	
OU-LP ^[28]	大阪大学	四个角度: 55°、65°、 75°和 85°	4007 人	
OU-LP Bag ^[29]	大阪大学	多自然状态携包	62528 人 177973 个序列	
OU-MVLP ^[30]	大阪大学	14 个角度: 0°~90°, 180°~270°	10307 人 267386 个序列	

续表

名称	建立机构	行走条件	样本容量	示例
OU-LP Age ^[31]	大阪大学	年龄为 2~90 岁	63846 人	
SZU RGB-D ^[32]	深圳大学	角度: 90°和 30°到 60°之间一个角度	99 人×2 视角 ×4 序列= 792 个序列	
ReSGait ^[33]	深圳大学	三种衣服类型, 四种携物状态, 是否使用手机	172 人 870 个序列	
GREW ^[34]	清华大学	开放系统的多个摄像机收集, 无约束条件	26345 人 128671 个序列	

从表 1.1 中可以看出, 早期的步态数据集数据量较小, 包含的行人数量甚至不足 100 人, 如 CMU MoBo、HID-UMD 和 CASIA-A。USF 数据集是在 HID 步态挑战的背景下收集的, 包括 122 名在椭圆路线上行走的受试者的户外步态视频。该数据集涵盖了几种具有挑战性的变量, 包括是否携带公文包、是否在不同的地面上行走、是否穿着不同的鞋子和是否是不同的采集时间, 且数据由左右两个摄像机从 4 个视角采集。SZU RGB-D 则主要考虑了行走方向的问题, 在两个视角上进行数据收集^[32]。CASIA-B 数据集是使用最广泛的步态数据集, 包含来自 124 人的 RGB 图和轮廓图两种形式的多视角步态数据, 覆盖了 11 个不同的行走角度, 范围为 $0^\circ \sim 180^\circ$, 以 18° 为增量。并且, 该数据集考虑了三种不同的行走条件, 即正常行走(NM)、穿外套行走(CL)和携包行走(BG)。除此之外, 中国科学院自动化研究所建立的另一数据集 CASIA-E, 含有的步态变量类型与 CASIA-B 类似^[26], 但样本量增加到 1014 名行人, 包括室外条件, 并记录了所有被采集者的 5 种软生物特征, 包括年龄、性别、身高、体重和国籍。

此外, OU-ISIR Treadmill 是首个在跑步机上进行采集的数据集。之后, 大阪大学又发布了 OU-LP 步态数据集, 这是当时世界上最大的步态数据集, 包括 4007 名行人, 其中 2135 名男性, 1872 名女性。此数据集的优势在于其样本量非常大, 性别比例接近, 且年龄范围从 1 岁到 94 岁不等。在此基础上, 该机构又考虑到携带物、年龄和行走角度的因素分别收集了 OU-LP Bag、OU-LP Age 和 OU-MVLP。它们同样有着大规模的样本数量, 受试者通常都在 10000 名以上, 包含了 100000

段以上的视频序列。

随着数据集规模的增大,自然真实状态场景下的步态采集也逐渐引起了研究者的关注。因为规定状态下步态采集的条件标准是人为设定的,尽管着装、角度和携带物等因素都被全方位地考虑进来,但仍与实际场景有一定的差距,变量多样性不足,所以研究往往会受到在实验室环境中收集的数据集的限制^[34]。并且,受试者在给定环境下行走时可能由于紧张等心理原因无法呈现自然的步态状态。因此,自然真实场景下的数据集建立对步态识别的实际应用意义重大。ReSGait是一种基于真实场景的步态数据集,它是第一个在行人可自由移动且环境参数不受控制的无约束场景下收集的数据集。该数据集由 172 位行人的 870 个视频序列组成,采集记录的时间超过 15 个月。数据集的视频序列除身份信息外,还被标记了性别、服装、携包、行走路线以及是否使用手机的行人属性变量。由于此数据集更接近于现实生活中实际的识别条件,因而为步态识别方法创建了一个新的难度级别。而 GREW 是第一个在自然环境下收集的大规模步态数据集,并为自然条件下的步态识别提供了一个新的基准。GREW 数据集由自然环境下行人视频组成,涵盖了来自数百个摄像机和开放系统中数千小时的数据流。此数据集由 26345 人的 128671 个视频序列组成,且具有丰富的行人属性,十分适用于无约束步态识别方法的训练和评估,为无约束步态识别方法的全面应用和量化分析提供了建设性的解决方案。

1.2 研究现状概述

自从在医学上证明步态对于每个个体都是独特的,其便正式作为一种特殊的生物特征进入人们的视野^[35,36]。初期的工作大多考虑利用传感器来获取更易处理的步态信息^[17],而不是以视频数据为主体研究对象,例如,通过加速度计或陀螺仪等来测量步态周期中各个相位的时间、步幅和步速等^[37,38]。但这种方法需要人工干预,鲁棒性不高,且没有发挥步态特征无须用户配合的优势。自 1994 年麻省理工学院的 Niyogi 等^[6]最早提出基于视觉的步态信息的身份认证后,以视频为数据来源的步态识别算法快速发展,并吸引了大量研究者的注意。早期的研究主要通过手动提取步态特征并采用规则匹配的方法进行步态识别,而这些方法往往依赖于人类专家在专业领域的经验作为先验知识,需要耗费大量的人力和时间,且准确度不高^[5]。2000 年,美国国防高级研究计划局(Defense Advanced Research Projects Agency, DARPA)确立开展基于人脸和步态的远距离身份识别(human ID at a distance, HID)重大研究项目^[21],基于视觉的步态识别技术也逐渐成为研究热点。在深度学习的浪潮兴起之前,步态识别研究主要采用了传统的图像处理和模式识别方法,基于人工设计的特征和分类器来实现步态识别,如主成分分析

(principal component analysis, PCA)^[39]、线性判别分析(linear discriminant analysis, LDA)^[40,41]、Gabor 小波^[42]、梯度方向直方图(histogram of oriented gradient, HOG)^[43]、光流法^[44]、隐马尔可夫模型(hidden Markov model, HMM)^[45]、典型相关分析(canonical correlation analysis, CCA)^[46]、视角转换模型(view transformation model, VTM)^[47]、耦合学习^[48]、三维人体建模^[49,50]、 K 近邻^[51]、决策树^[52]和支持向量机(support vector machine, SVM)^[53]等。同时,相关领域研究工作的进展,如背景去除^[54]、人体检测^[55]、骨骼关键点提取^[56]和运动分析^[57]等的发展,也让步态数据的预处理过程更为简单便捷,进一步推进了步态识别方法的优化和完善。继而,随着深度学习的兴起,基于视觉的步态识别研究也迈入了一个新的阶段,其性能也达到了新的基准^[4,16]。因此,本节将着重介绍基于深度学习方法的步态识别相关研究现状。

因为大型步态数据集的支持对于步态识别的研究和评估不可或缺,数据集中数据规模的大小、条件变量的丰富度和采集环境的真实度直接影响了算法的发展方向,深度学习方法是更依赖大量数据来学习和训练。所以,本节首先介绍近年来公开的步态数据集,并介绍步态数据预处理的常用技术,最后阐述相关基于视觉的步态识别方法。

1.2.1 步态数据预处理方法

端到端的步态识别方法通常由几个部分组成,包括行人或运动检测、行人分割和识别算法等^[58]。其简要流程如图 1.2 所示,原始的步态视频序列首先通过行人或运动检测模块,从视频的场景图像序列中检测和提取出行人目标,然后再通过行人分割步骤对已提取的目标进行分割,得到行人目标的二值剪影轮廓图序列。此外,有些算法要求将一段步态视频按照一个步态周期进行截取,获得一段完整的步态剪影序列。最终采用相应的识别算法得到其身份信息。其中,行人分割和周期检测模块通常被视为步态数据的预处理过程,相对于步态识别算法的研究,受到的关注较少。

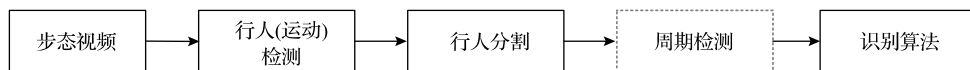


图 1.2 端到端的步态识别流程

1. 行人检测与分割

行人检测旨在从一段图像序列中检测出行人所在的位置,进而分割出行人轮廓。现有的步态数据集往往是在封闭且纯净场景中录制单一人行走图像序列,因此采用前景背景分离即可实现行人检测。而在较复杂的真实场景中,可考虑基

于深度学习的目标检测算法。例如,采用 Faster R-CNN^[59]和 YOLOv3^[60]等实现行人检测与分割;采用 Mask R-CNN^[61]等获取行人轮廓。

行人检测与分割算法的研究也推动了智能步态视频分析和识别的发展和落地。具体而言,Zhang 等^[62]构建了由 5 种行人部位组成的部位模型,并通过卷积神经网络(convolutional neural network, CNN)度量各个部位可见度,最后综合所有部位的可见得分,得到整体行人的检测结果。该行人检测框架利用一个网络同时进行整体行人检测和遮挡估计,具体而言就是在同一网络中设计并行的两条分支,分别输出两组预测框用以表征整体行人目标以及行人可见部分。该方法显著提高了遮挡及无遮挡检测精度,但是代价也显而易见,即需要数据集额外标注出行人目标的可见部分用于网络训练,故该方法无法在没有可见部分标注的数据集上使用。此外,基于无锚点框的方法逐渐引起研究者的关注^[63]。此类方法并不需要额外设置先验框参数,而是直接利用特征图进行分类预测以及位置回归,推动了目标检测领域的巨大进步。Song 等^[64]提出可以利用直立行人的垂直特征将行人检测划分为 3 个子任务,分别是行人目标上顶点预测、行人目标下顶点预测及行人目标中轴线预测。

Mask R-CNN^[61]是基于 Faster R-CNN 提出的用于图像分割的深度卷积网络,在进行目标检测的同时实现高质量的分割。第一阶段,首先用区域建议网络(region proposal network, RPN)提取出候选目标的边界框,然后将边界框里面的感兴趣区域(regions of interest, RoI)划分为 $m \times m$ 的子区域;第二阶段,与预测类和边界框回归任务并行,增加了为每个边界框输出二分类掩模的分支,可理解为用全卷积神经网络对每个边界框里面的内容进行分割,以像素到像素的方式预测分割掩模。

2. 步态周期检测

不同的行人有不同的步态,但大体行走过程类似。图 1.3 展示正常状态的行人在一个步态周期内的步态序列。一个步态周期的过程可以分为 4 个部分:右双

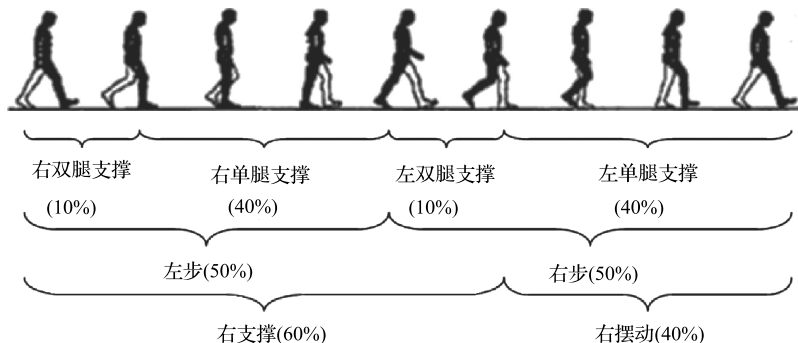


图 1.3 步态周期示意图^[65]

腿支撑,右单腿支撑,左双腿支撑,左单腿支撑。

步态周期检测是实现基于视觉的步态识别的一个重要步骤。与其他生物识别技术不同,步态识别主要依据人体在行走过程中的动作习惯完成身份识别,故不适用于利用单一的步态帧图像进行步态识别。因此,步态识别的输入是视频序列(连续图像)而不是单张的步态轮廓图像。而步态周期检测是指判定输入视频合适长度的过程,一个步态周期通常包含具有完整的步态动作特征的最少的帧数。检测得到的步态周期过短,可能会丢失有效的步态特征,而检测到的步态周期过长,可能会使输入包含冗余数据,占用更多的计算资源。对于比较常用的类能量图像的步态识别方法,需要步态周期检测过程作为前端的预处理步骤,而步态周期检测结果也直接影响了类能量图步态识别结果的准确性。

现有的关于步态周期检测的许多研究都是基于人体行走的几何特征实现的。Collins 等^[66]提出了一种利用人体的宽度和高度进行周期检测的方法,但该方法受人与摄像机之间的距离变化影响较大。Wang 等^[24]考虑到不同视角下轮廓大小的变化,提出了一种基于高宽比的方法,省去了标准化过程。Wang 等^[67]选择腿的平均宽度作为特征来进行周期检测,避免了服饰和背包的影响。此外,身体的面积也可以有效地表示周期特征,Sarkar 等^[21]利用腿的面积作为特征完成周期检测。Ben 等^[68]提出了一种双椭圆拟合方法,将整个轮廓的两个区域拟合为被质心分割成两个椭圆,步态起伏被构造为一个周期函数进行检测。这类基于几何特征的方法简单有效,在近 90°的侧视视角上达到了很高的精度。但对于不同的环境(不同的视角或服饰等),检测精度不理想,没有很好的鲁棒性。

1.2.2 基于步态模板图像的方法

步态模板图像是一种经典的手工步态视频预处理方法。由于高维的视频数据量较大且存在很多信息冗余,很难被机器直接理解和学习。传统的方法很难直接从序列的剪影中提取特征,而采用深度学习方法处理高维的视频数据也面临很多难点。步态模板图像是为了缓解这个问题而构造出的能够集中整合整个视频中步态信息的单张步态图像,通常是对视频图像序列按照一定的规则叠加而成^[69]。尽管步态模板图像的维数要小得多,但因其简单高效,所以在应用中取得了很大的成功。主要的步态模板图像有步态能量图(gait energy image, GEI)^[70]、步态历史图(motion history image, GHI)^[71]和步态熵图(gait entropy image, GEnI)^[7]等,如图 1.4 所示,它们可以通过一张图像清晰地反映一段序列剪影的外观轮廓信息。

基于深度学习的方法可以在步态模板图像上进行二次特征提取,相对于传统方法能够自动地学习到更具表征性的步态特征,同时对于其他因素的影响具有一定的鲁棒性,这也是早期基于深度学习的步态识别方法的主要研究趋势。

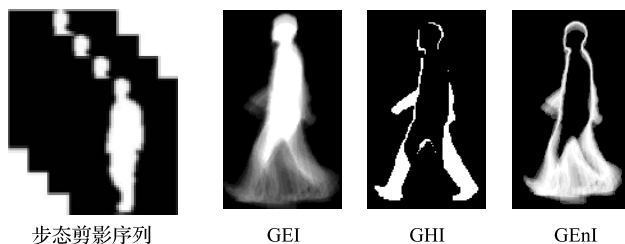


图 1.4 步态模板图像示例

Shiraga 等^[72]将最流行的 GEI 作为输入,专门设计了由两个连续的卷积层、池化层和全连接层构成的 GEI 网络。Hossain 等^[73]在使用多层卷积提取 GEI 模板的同时融合了深度特征,取得了显著改进。Yan 等^[74]提出将 GEI 送入卷积网络,并加入多任务学习的方法,在步态识别的同时预测目标行人的其他状态,结合所有任务的损失函数来进行互学习。Zhang 等^[75]提出将 GEI 送入具有两个共享权重的孪生卷积神经网络以完成身份验证。Wu 等^[76]同样基于 GEI 设计了多种卷积网络结构,针对各种识别条件(交叉视图和交叉行走条件)提供了广泛的实证评估。Xu 等^[77]首先使用卷积网络学习 GEI 模板的属性,然后运用带有动态路由的胶囊网络来保留每个模板中的关系以找到鲁棒性更强的特征,这是因为胶囊网络分别利用胶囊激活值和协议路由算法,有效弥补了卷积网络的缺点^[78]。Takemura 等^[79]讨论了多种基于卷积神经网络的跨视图步态识别的网络架构,权衡由不同身份信息和不同行走角度引起的空间位移之间的差异,并使用带有双输入的孪生网络和三重输入的三元组网络度量样本之间的单差异或多差异,以进行对比匹配。Xu 等^[80]基于 GEI 提出了用于跨视角识别的成对空间变换器网络(pairwise spatial transformer network, PSTN),可以估计一个非刚性变形场以将匹配对中的特征转换到其中间视角中。

1.2.3 基于一组步态剪影图像的方法

步态剪影轮廓可以理解为去掉背景后的人体掩模,但即使在剪影轮廓分割的过程中去掉了颜色和纹理,人体轮廓仍然包含充足的外观信息^[24]。Vinyals 等^[81]以无序的图像集作为输入,研究表明即使输入数据未按照时间顺序排列,也可以保证网络良好的性能,这一想法也被广泛应用于自然语言处理中,继而推广至步态识别任务。

相应地,一组忽视了顺序的步态剪影集相对于步态模板图像能够表现更多的特征,所以很多工作以其为输入开展研究。其中最引人注目的是 GaitSet 结构^[82-84]。如图 1.5 所示,该结构首先使用基础的卷积网络从每个轮廓中独立提取帧级外观特征。进而,采用多种时域池化的方法将帧级特征整合为视频级特征,且设计了多层全局支路,在每个层中分别提取多尺度的视频级步态表征。最后应用水平金