

北 京 化 工 大 学

专 业 学 位

硕士研究生学位论文

题 目 基于度量学习的步态识别方法研究

研 究 生 刘 东

专 业 计 算 机 技 术

指导教师 胡 峻 林 副 教 授

企业导师 罗 瑞 一 高 工

日 期： 二 〇 二 二 年 五 月 二 十 八 日

## 北京化工大学学位论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

作者签名： 刘东 日期： 2022.5.20

## 学位论文使用授权声明

本人同意北京化工大学可以采用影印、缩印或其它复制手段保存、汇编学位论文，允许学位论文及其复制件根据学校规定被查阅、借阅、复制、通过网络传播；学校可以向国家有关部门或其指定机构送交论文原件、复制件及电子文本。

本人 ☒ 同意 / ☐ 不同意 授权学校向《中国博士/优秀硕士学位论文全文数据库》等有关数据库提供学位论文及其电子文本并用于网络检索、查阅和传播。

上述同意与不同意的方框若未勾选，视为本人同意授权。

☐ 论文暂不公开（或保密）注释：本学位论文属于暂不公开（或保密）范围，在\_\_\_\_年解密后适用本授权书。

☒ 非暂不公开（或保密）论文注释：本学位论文不属于暂不公开（或保密）范围，适用本授权书。

作者签名： 刘东 日期： 2022.5.20

导师签名： 胡冬林 日期： 2022.5.20

## 学位论文数据集

中图分类号	TP301.6	学科分类号	520.2040	
论文编号	1001020220499	密 级	公开	
学位授予 单位代码	10010	学位授予 单位名称	北京化工大学	
作者姓名	刘东	学 号	2019210499	
获学位专业 名称	计算机技术	获学位专业 代码	085211	
课题来源	国家自然科学基金	研究方向	计算机视觉	
论文题目	基于度量学习的步态识别方法研究			
关 键 词	步态识别，度量学习，L1 距离度量，特征提取			
论文答辩日期	2022.5.16	论 文 类 型	应用研究	
学位论文评阅及答辩委员会情况				
	姓名	职称	工作单位	学科专长
指导教师	胡峻林	副教授	北京航空航天大学	计算机视觉
评阅人 1	李辉	副教授	北京化工大学	人工智能
评阅人 2	黄仁杰	副教授	西南大学	计算机视觉
评阅人 3				
评阅人 4				
评阅人 5				
答辩委员会主席	祝海江	教授	北京化工大学	计算机视觉
答辩委员 1	赵瑞莲	教授	北京化工大学	软件可靠性
答辩委员 2	赵英	教授	北京化工大学	计算机网络
答辩委员 3	李辉	副教授	北京化工大学	人工智能
答辩委员 4	尤枫	副教授	北京化工大学	Web 用户行为分析
答辩委员 5	江志英	高工	北京化工大学	大数据实时分析

# 基于度量学习的步态识别方法研究

## 摘要

得益于计算机视觉的发展,传统生物识别技术已经取得了广泛的应用,其中最具代表性的就是指纹识别和人脸识别。但是这类方法通常需要待识别者一定的配合,并且识别所需的指纹、面部等生物特征需要在近距离内完成采集。与传统的生物识别方法相比,步态识别以行人行走的姿态作为识别的特征,因此可以实现远距离、无接触的身份识别。近年来,步态识别取得了显著的发展,但是仍然面临着行人着装变化和拍摄视角变化带来的挑战。

针对当前步态识别面临的行人着装变化及拍摄视角变化引起的识别精度下降等问题,本文从度量学习的视角出发,研究步态识别问题。首先,对步态识别中的特征表示方式进行了优化,包括对步态序列的帧图像对齐方式的改进和步态序列帧图像整合方式的选择,并对比了几种特征表示方式的实际表现。其次,本文基于三元组损失提出一种步态识别方法,通过固定锚点样本并同时约束正负样本和锚点样本之间的距离使得样本在特征空间中呈现合理的分布。最后,针对部分步态轮廓图像之间差异过大的问题,本文提出一种基于 L1 距离学习的步态识别方法,寻找一个投影矩阵将原始数据投影到一个特征空间中,使得在该特征空间中,同一类的样本间的 L1 距离小于一个阈值,不同类的样本间的 L1 距离大于另一个阈值,从而提高样本的识别精度。

在广泛使用的 CASIA-B 及 CAISA-C 步态数据集上进行的一系列对比

实验表明,本文提出的特征表示优化方式以及两种步态识别方法具有较好的性能。

**关键词:** 步态识别, 度量学习, L1 距离度量, 特征提取

# GAIT RECOGNITION METHOD BASED ON METRIC LEARNING

## ABSTRACT

Owing to the development of computer vision, traditional biometric technology has been widely used in many real-world applications, the most representative of which are fingerprint recognition and face recognition. However, such methods usually require a certain amount of cooperation from the person to be identified, and biometric features such as fingerprints and faces need to be collected within a short distance. Compared with traditional biometric methods, gait recognition exploits the walking posture of pedestrians as the identification features, so it can realize long-distance, contactless identification. In recent years, gait recognition has made remarkable progress, but it still faces challenges brought about by changes in pedestrians' clothing and view angles.

Aiming at the current gait recognition problems such as the decrease in the recognition accuracy caused by changes in pedestrian clothing and view angles, this thesis studies the problem of gait recognition from the perspective of metric learning. Firstly, the feature representation in gait recognition is optimized, including the improvement of frame image alignment of gait

sequence and the selection of integration of gait sequence frame, and the performance of several feature representations is compared. Secondly, this thesis proposes a gait recognition method based on triplet loss. By fixing anchor samples and constraining the distance between positive and negative samples and anchor samples at the same time, the samples present a reasonable distribution in the feature space. Finally, in view of the problem that the difference between some gait silhouette images is too large, this thesis proposes another gait recognition method based on L1-norm distance, looking for a projection matrix to transform the original data into a subspace in which the L1-norm distance between samples from the same class is less than a threshold, and the L1-norm distance between samples from different classes is larger than another threshold, thereby improving the recognition accuracy.

A series of comparative experiments on the widely used CASIA-B and CAISA-C gait datasets show that the feature representation optimization method and the two gait recognition methods proposed in this thesis have good performance.

**KEY WORDS:** gait recognition, metric learning, L1-norm distance metric, feature extraction

# 目 录

<b>第一章 绪论</b>	<b>1</b>
1.1 研究背景	1
1.2 国内外研究现状	2
1.2.1 步态识别研究现状	2
1.2.2 度量学习研究现状	4
1.3 本文的主要工作	6
1.4 论文的章节安排	6
<b>第二章 理论基础和相关技术</b>	<b>9</b>
2.1 度量学习	9
2.1.1 距离度量	9
2.1.2 损失函数	10
2.2 步态预处理	11
2.2.1 步态轮廓分割	11
2.2.2 轮廓图像归一化	15
2.3 相关数据集及评价指标	17
2.3.1 步态数据集	17
2.3.2 评价指标	18
2.4 本章小结	20
<b>第三章 基于三元组损失的步态识别方法</b>	<b>21</b>
3.1 步态特征提取	21
3.1.1 步态能量图	21
3.1.2 活性能量图	22
3.1.3 特征降维	22
3.2 损失函数及优化	24
3.2.1 目标函数	24



3.2.2 算法实现 .....	24
3.3 实验结果及分析 .....	25
3.3.1 实验设置 .....	26
3.3.2 不同特征对比实验 .....	26
3.3.3 与其他方法的对比实验 .....	29
3.4 本章小结 .....	31
第四章 基于 L1 距离学习的步态识别方法 .....	33
4.1 基于 L1 距离学习的步态识别方法 .....	33
4.1.1 目标函数 .....	34
4.1.2 算法实现 .....	35
4.2 实验结果及分析 .....	36
4.2.1 实验设置 .....	36
4.2.2 不同特征的对比实验 .....	37
4.2.3 与其他方法的对比实验 .....	39
4.2.4 L1 距离度量有效性实验 .....	42
4.3 本章小结 .....	43
第五章 总结与展望 .....	45
5.1 本文总结 .....	45
5.2 未来工作的展望 .....	45
参考文献 .....	47
致谢 .....	53
研究成果及发表的学术论文 .....	55
作者和导师简介 .....	57

# Contents

Chapter 1 Introduction .....	1
1.1 Research background .....	1
1.2 Research status at home and abroad .....	2
1.2.1 Research status of gait recognition .....	2
1.2.2 Research status of metric learning .....	4
1.3 The main work of this paper .....	6
1.4 Chapter organization of this paper .....	6
Chapter 2 Theoretical Basis and Related Technologies .....	9
2.1 Metric learning .....	9
2.1.1 Distance metrics .....	9
2.1.2 Loss functions .....	10
2.2 Gait preprocessing .....	11
2.2.1 Gait silhouette segmentation .....	11
2.2.2 Silhouette image normalization .....	15
2.3 Relevant datasets and evaluation criterion .....	17
2.3.1 Gait datasets .....	17
2.3.2 Evaluation criterion .....	18
2.4 Summary of this chapter .....	20
Chapter 3 Gait Recognition Method Based on Triplet Loss .....	21
3.1 Gait feature extraction .....	21
3.1.1 Gait energy image .....	21
3.1.2 Active energy image .....	22
3.1.3 Feature dimensionality reduction .....	22
3.2 Loss function and optimization .....	24
3.2.1 Objective function .....	24

3.2.2 Algorithm Implementation .....	24
3.3 Experimental results and analysis .....	25
3.3.1 Experiment settings .....	26
3.3.2 Comparative experiments of different features .....	26
3.3.3 Comparative experiments with other methods .....	29
3.4 Summary of this chapter .....	31
 Chapter 4 Gait Recognition Method Based on L1-norm Distance Learning ..	33
4.1 Gait recognition method based on L1-norm distance learning .....	33
4.1.1 Objective function .....	34
4.1.2 Algorithm Implementation .....	35
4.2 Experimental results and analysis .....	36
4.2.1 Experiment settings .....	36
4.2.2 Comparative experiments of different features .....	37
4.2.3 Comparative experiments with other methods .....	39
4.2.4 L1 distance metric effectiveness experiments .....	42
4.3 Summary of this chapter .....	43
 Chapter 5 Summary and future work .....	45
5.1 Summary .....	45
5.2 Future work .....	45
 References .....	47
 Acknowledgment .....	53
 Research results and published academic papers .....	55
 Introduction of advisor and author .....	57

## 第一章 绪论

### 1.1 研究背景

随着现代社会的信息化发展不断进步,对公民的身份进行识别和管理已经成为了现代社会治理中不可或缺的一部分。由于个体所具有的生物特征往往是独一无二且不易发生改变的,因此,利用生物特征进行身份识别逐渐成为一个热门的研究领域。目前,基于生物特征的身份识别已经广泛应用于人们生活的方方面面,从智能手机的解锁,到公司的门禁,再到公共场所的监控安防系统,医学上的亲子鉴定等,这些识别技术都在时时刻刻为我们的生活提供便利。在这些技术中,常用的生物特征主要包括人脸,指纹,虹膜等,利用这些生物特征进行身份识别的技术具有准确率高,识别速度快的优势。但是,传统的生物特征识别方式也有一定的局限性,首先,这类特征需要在近距离内完成采集,当距离过远时,可能无法采集到足够清晰的特征图像,进而无法完成识别。其次,这类方法往往需要待识别者一定的人为配合,如指纹识别需要待识别者进行按压采集,虹膜识别需要待识别者目视摄像头,人脸识别则意味着待识别者需要摘下口罩,这些因素无形中降低了识别的效率。在一些安防监控等应用中,常常需要借助视频片段完成远距离、无感知的身份识别,这时传统的生物识别就无法满足要求。

基于这样的需求,步态识别应运而生。由于每个个体的骨骼,肌肉,身高,体态等方面有所不同,因此行走的姿态也会存在一定的差异,步态识别就是基于这些差异来完成身份识别。与传统的生物识别技术相比,步态识别是从行人行走的姿态中提取特征进行识别,对距离不敏感,因而可以实现远距离、非接触、无感知的身份识别。正是因为这样的优势,步态识别在安防、监控等领域有着非常广阔的应用前景。随着计算机视觉的不断发展,近年来不断有新的步态识别方法提出,但是基于视觉的步态识别仍然面临着一些挑战。首先就是角度的变化,在实际应用中,行人往往不会以某一固定的视角出现在镜头中,因此在识别中需要处理各种不同视角的人体图像。其次,基于图像的步态特征很容易受到来自行人的着装变化或者携带物品如背包等的干扰。最后,一些环境因素如光线,遮挡等也会对实际的图像特征采集带来一些困难。所以,目前步态识别技术仍然处于需要大量探索 and 研究的阶段,距离真正大规模投入实际应用仍有一定的完善空间。

结合步态识别相对于传统生物识别所具有的优势和目前所面临的问题来看,步态识别仍然是一个非常值得研究的领域。

## 1.2 国内外研究现状

本节将主要介绍步态识别和度量学习这两个领域的研究现状，包括这两个领域的研究历史以及国内外研究的最新进展，并分析这两个领域目前仍待解决的问题。

### 1.2.1 步态识别研究现状

步态识别主要借助不同行人行走姿态中存在的差异完成对行人的身份识别。Song 等人<sup>[1]</sup>将步态定义为“使身体发生位移的一系列腿部运动和身体运动的结合”，并最早使用局部相位的概念来定义步态中脚的位置，开发并简化了周期性步态研究的基本定理。步态可以作为一种生物识别方式的依据是，每个生物个体由于骨骼，肌肉，身高，体态等方面都有所差异，因此步态特征也存在着一定的差异，这种差异往往是无法替代的，因而可以通过步态特征唯一地确定某一个个体。步态是一种动态的生物特征<sup>[2]</sup>，相较于静态的密码、指纹以及面部特征等，步态更加复杂并且难以模仿。因为其可以实现远距离、非接触和无感知的身份识别<sup>[3]</sup>，长期以来一直被视作传统生物识别在安防监控领域的替代方案之一。

主流的步态识别方法主要包括基于模型的方法和基于图像的方法。基于模型的方法需要借助一些深度摄像头或者可穿戴的传感器，采集行人行走过程中的运动学数据，建立运动模型并从肢体关节运动的角度、幅度等参数进行分析。这类方法的核心就是运动模型的构建。Wang 等人<sup>[4]</sup>提出一种结合了动态和静态人体生物特征的方法，该方法使用区域匹配和 Procrust 形状分析从每个待识别者的步态序列中提取向量作为人体的静态特征表示。同时，对于每个步态序列，使用每个关节的高斯函数细化的旋转构建运动模型，计算肩部、肘部、臀部、膝盖和脚踝等关节的位置及角度。所提取的静态和动态特征都可以独立使用最近邻方法进行分类识别，在决策层面上，该方法使用不同的规则将两种特征的识别进行融合以提高算法的性能。Lima 等人<sup>[5]</sup>提出一种基于 2D 姿态模型的方法，该方法首先使用姿势估计对人体主要的 18 个关节坐标进行定位，后续通过两个网络模型来处理获得的关节信息，其中 PoseDist 网络从关节信息中提取空间特征和时间特征并以最近邻的方式对特征进行分类，PoseFrame 网络则将关节坐标基于颈部进行归一化并将每一帧的特征向量作为输入。该方法将关节信息作为步态中的关键特征，并在特征上融合了时空信息，对动态的特征进行了很好地表达。

近年来，基于模型的方法有了新的发展，越来越多的研究者将深度摄像头引入运动建模中，Tang 等<sup>[6]</sup>提出一种基于修复的重建三维模型步态识别方法，采用步态局部感兴趣区域（ROI）元素选择方法提取不同视角下的步态轮廓特征，从步态图像序列

中重建三维步态模型。Choi 等人<sup>[7]</sup>提出一种基于 3D 骨架的步态识别方法,该方法基于 Kinect 传感器采集的数据,根据身体的对称性来测量骨架的质量,并基于质量在输入帧与注册帧之间构建了一个质量权重矩阵以降低噪声对识别的影响。通过对每个帧分配不同的权重来增强对不确定的骨架数据识别结果的鲁棒性。Luo 等<sup>[8]</sup>提出一种鲁棒性较好,可以不受服装风格影响的步态识别方法,该方法由三个模块构成:(1) 三维人体姿态、形状和视觉数据估计网络(3D BPSVeNet);(2) 步态语义参数折叠模型;(3) 步态语义特征提取网络。首先,建立 3D BPSVeNet 并从图像序列中提取二维到三维人体姿势和形状语义描述符(2D-3D-BPSDs),然后利用 2D-3D-BPSDs 和识别出的服装信息构造出具有虚拟着装的三维步态模型。将步态模型使用稀疏分布表示(SDR),它将非结构化的原始步态数据转换为步态语义图像的结构化数据,最后由 SoftMax 分类器进行识别。该方法最大的优势在于充分利用参数化的三维步态模型和三维服装模型,很大程度上解决了服装变化和视角变化对于识别的影响,大大提升了鲁棒性。总的来说,基于模型的步态识别方法主要使用深度摄像机或者可穿戴式传感器采集关节运动的角度,幅度等运动参数作为识别的特征,这类方法主要面临着特征采集难度较高,计算量较大等问题。

基于图像的方法则是从行走的视频序列中采集人体轮廓作为特征信息,与基于模型的方法相比,这类方法所使用的特征信息更为丰富,而不仅仅局限于特定的关节或者身体部位。基于图像的方法最早从步态序列中提取单一的帧图像作为特征<sup>[9]</sup>,使用主成分分析将帧图像进行降维之后,根据最近邻分类完成图像对之间的匹配识别。然而单一的帧图像往往只包含静态的特征,忽略了步态特征随时间的变化,Chai 等人<sup>[10]</sup>提出一种基于感知曲线的步态识别方法,可以有效保持行人的轮廓变化与时间变化的一致性。首先,在步态序列的每一帧图像中检测到行人的轮廓,然后使用内边界跟踪算法提取行人的二进制轮廓,并计算轮廓的感知形状描述符(Perceptual Shape Descriptor, PSD),将每个步态序列的累积 PSD 生成一条感知曲线作为该样本的步态特征。该方法主要的贡献在于将单一的图像特征与时间特征融合,使得步态特征真正具有了动态性。Han 等人<sup>[11]</sup>提出了应用广泛的步态能量图(Gait Energy Image, GEI)的概念。由于人的行走是一个具有周期性的动作,这一点在文献<sup>[1]</sup>中也有详细的论证,因此,可以从步态序列中提取出单个行走周期内的连续帧,将这些连续帧进行二值化、对齐、裁剪之后,累加并求平均得到一张平均图像,就称为步态能量图。步态能量图很好地融合了时空特征,以较少的数据量保留了较大的信息量,能够在尽可能保留步态特征信息的前提下减少算法的计算量,是目前应用较为广泛的一种步态特征。后来的研究者也对步态能量图提出了一些改进方案,如活性能量图(Active Energy Image, AEI)<sup>[12]</sup>,与步态能量图不同,活性能量图是将一个步态周期内所有相邻的帧图像相减得到差图像,再将所有差图像进行累加求平均,相对于步态能量图,活性能量图可

以提高轮廓图像的质量,并且求差值的方式获得的差图像也可以保留更多的动态特征。基于这类能量图像作为特征,研究者提出了很多步态识别方法。Shiraga 等<sup>[13]</sup>将 GEI 作为一个卷积神经网络 (Convolutional Neural Network, CNN) 的输入,提出 GEINet,它由两个连续的三元组卷积层、池化层、归一层以及两个全连接层组成。Gul 等<sup>[14]</sup>则使用 3D CNN 作为网络结构,对 GEI 作为输入的模型进行了优化策略方面的探索,实现对网络性能的提升。

当然,也有研究者提出直接将步态序列的所有帧图像作为输入的算法。Zhang 等<sup>[15]</sup>引入长短期记忆 (Long Short Term Memory, LSTM) 单元搭建一个时间注意力模型,将步态序列中的帧图像以时间顺序输入,以更好地学习时间步态特征。Chao 等<sup>[16]</sup>认为使用步态序列或者模板图像在一定程度上影响了步态识别的灵活性,因此提出一种端到端的步态识别模型 GaitSet。GaitSet 中,步态序列不需要像步态模板或者步态序列中一样保持时间的顺序,而是被视为一组相互独立的帧。该方法不受帧排列的影响,并且可以自然地整合不同场景下拍摄的不同视频的帧,例如不同的视角、不同的衣服等,在复杂的场景下有较好的性能。Fan 等人<sup>[17]</sup>认为,基于 LSTM 的方法保留了周期性步态不必要的顺序约束,而 GaitSet 虽然打破了这种约束,但是没有明确地模拟时间的变化,因此,他们提出了一种基于局部特征的 GaitPart 模型,该模型应用了最新的协调卷积层 (Focal Convolution Layer) 来增强模型对空间特征的细粒度学习,另一方面还加入了微动捕捉模块 (Micro-motion Capture Module, MCM) 来捕获一定的时间特征,以实现时空特征的融合。

目前步态识别技术在应对单一角度且行走状态不发生改变的情况下,已经取得了理想的效果,而在角度和行走状态发生改变的情况下,步态识别的精度仍然有提升的空间。当然,应对这一挑战也已经有很多积极的尝试,如 TS-GAN<sup>[18]</sup>, GaitGANv2<sup>[19]</sup>, PTSN<sup>[20]</sup>等。总的来说,角度和行走状态的变化仍然是当前步态识别面临的最主要的挑战。

### 1.2.2 度量学习研究现状

度量学习的思想起源于分类问题中的最近邻方法 (Nearest Neighbors, NN)<sup>[21]</sup>,在最近邻中,依据待测样本与注册样本之间的距离为待测样本划分类别,即将待测样本识别为距离其最近的类别。在这个过程中,选择怎样的距离度量就成了一个关键的问题。度量学习的基本任务就是从数据中学习合适的距离度量,以达到更好地区分样本数据的目的。大量的研究已经证明,学习合适的距离度量可以极大程度上提高分类任务的性能<sup>[22-24]</sup>。传统的度量学习在距离度量的选择上经历了从欧氏距离到马氏距离的转变,目前,也有大量基于马氏距离的度量学习研究。

度量学习主要包括无监督和有监督两大类<sup>[25]</sup>, 区别在于无监督度量学习不需要样本数据集的完整监督信息, 而监督度量学习在训练的过程中往往需要借助样本数据的监督信息。无监督度量学习的主要思想是学习一个相较于原始数据更低维的子空间, 并尽量保留原始数据之间的几何关系如距离等。典型的无监督度量学习包括拉普拉斯特征映射<sup>[26]</sup>、多维标度 (Multiple Dimension Scaling, MDS)<sup>[27]</sup>和主成分分析 (Principle Component Analysis, PCA) 等降维算法。监督度量学习则需要通过优化目标函数的方式来学习合适的距离度量, 又可以进一步划分为基于对约束的方法和基于概率框架的方法。Xing 等<sup>[28]</sup>将基于对约束的距离度量学习视为一个凸优化问题, 提出边信息学习 (Learning with Side Information, LSI), 并在 UCI 数据集上提高了算法的集群性能。边际最大化判别分析 (Margin Maximizing Discriminant Analysis, MMDA)<sup>[29]</sup>认为, 理想的特征应该最大化地包含类标签中的信息, 并且特征应该只取决于最佳决策边界, 而与样本数据中非边界部分无关。基于这样的原则, MMDA 将输入的样本数据投影到一组成对正交边距最大化超平面的法线所跨越的子空间上, 这样可以根据投影空间的维度从样本数据中提取尽可能多的特征。相关成分分析 (Relevant Component Analysis, RCA)<sup>[30]</sup>认为将数据沿着相关性最低的维度压缩可以提高算法在分类任务上的性能, 因此该方法使用一个变换矩阵  $W$  对数据进行投影, 该矩阵将较大的权重分配给相关性较高的维度, 将较小的权重分配给相关性较低的维度, 尽可能多地保留原始数据中地相关可变性。

早期的度量学习方法倾向于寻找一个线性映射将样本数据投影到新的子空间中, 这有助于提高学习的性能, 但是线性映射在面对一些非线性特征结构的情况下性能就不那么理想。尽管一些研究者尝试引入核函数的方法<sup>[31-33]</sup>将样本数据嵌入希尔伯特空间来应对非线性问题, 但是核函数可能会带来过拟合的问题。近年来, 随着深度学习的不断发展, 其在应对非线性样本数据时的优秀性能引起了研究者的注意, 传统度量学习也开始逐渐与深度学习结合。Hadshell 等<sup>[34]</sup>提出一种对比损失函数, 首次将深度学习引入度量学习中, 开启了深度度量学习的研究。对比损失通过将样本数据组合成样本对的形式, 使得类内样本尽可能接近, 以达到提高分类精度的目的。在此基础上, Schroff 等<sup>[35]</sup>提出三元组损失, 即将对比损失中的样本对扩展为三元样本组, 包含一个锚点, 一个与锚点同类的样本和一个与锚点异类的样本。其在对比损失的基础上进一步扩大了类内和类间的区别。由于三元组损失采用随机采样的方式生成三元组, 很难收集到足够的难负样本对, 样本组中会存在大量的冗余, 在模型收敛的时候这些样本并不能为学习过程提供梯度。针对这个问题, Ge 等<sup>[36]</sup>提出分层三元组损失, 他们使用一个分层的类级树捕获数据集的内在数据分布, 鼓励模型学习到更多区分性强的特征, 从而达到提高采样效率的目的。针对损失函数还有一些后续的改进, 如  $N$  元组损失<sup>[37]</sup>, 中心损失<sup>[38]</sup>, 代理损失<sup>[39]</sup>, 角损失<sup>[40]</sup>等。也有一些研究将焦点放在网络结构的



设计和采样策略的选择上。Yi 等<sup>[41]</sup>受到孪生神经网络的启发,提出一种用于行人重识别的深度度量学习方法,该方法将两个卷积神经网络通过余弦层连接组成一个孪生卷积神经网络(Siamese Convolutional Neural Network, SCNN),对于给定的两张人物图像,通过三个 SCNN 来判断一对人物图像是否来自同一标签。与传统孪生神经网络不同的是,该方法中的子网络不需要共享相同的权重,因此在应对跨试图的行人重识别任务中表现出优异的性能。Mehralian 等<sup>[42]</sup>将对抗生成网络(Generative Neural Nets, GANs)与度量学习结合,设计了一种基于 GAN 的度量学习框架。该框架中,使用一个神经网络从数据中学习合适的度量,另一个网络则充当第一个网络的监督者,获取第一个网络的输出并给出反馈以协助网络修改权重,通过两个网络交互式学习来提高特征提取的速度。随着深度学习和度量学习的不断融合,近年来,深度度量学习已经广泛应用于人脸识别<sup>[43-45]</sup>,行人重识别<sup>[46-48]</sup>,目标跟踪<sup>[49-50]</sup>等多个领域。

### 1.3 本文的主要工作

本文在对步态识别和度量学习领域的相关文献充分调研的基础上,对步态识别的发展现状作了简要介绍,分析了目前步态识别中仍然面临的挑战。在此基础上,本文做了以下方面的工作:

对步态特征的表示方法进行了探索和优化,在此基础上提出一种基于三元组损失的步态识别方法,通过固定锚点样本,寻找正样本和负样本使得锚点与正样本之间的距离较小,负样本和锚点之间的距离较大来提高识别任务的分类精度。通过实验对比了几种步态特征的优劣,并筛选了性能较好的步态特征表示方式。

针对步态识别中部分轮廓图像之间差异过大的问题,引入 L1 距离度量来提高模型对异常值的鲁棒性,提出一种基于 L1 距离的度量学习方法。寻找一个投影矩阵将原始数据映射到特征空间中,使得在该特征空间中同一类别的样本间的 L1 距离小于一个阈值,同时不同类的样本间的 L1 距离大于另一个阈值,提高步态识别在行走状态及视角变化时的性能表现。在两个公开的步态数据集上分别进行了一系列对比实验,表明所提出的方法能够应对步态识别任务中面临的挑战。

### 1.4 论文的章节安排

第一章主要阐述步态识别技术的研究背景和意义,并介绍了步态识别和度量学习这两个领域的发展历史以及研究现状,在此基础上介绍了本文的工作。

第二章介绍了步态识别和度量学习中的一些理论基础和相关技术。包括度量学习中常用的距离度量和损失函数,步态识别中对轮廓图像的一些预处理方法,最后介绍

了实验所使用的步态数据集和实验结果的评价指标。

第三章提出了一种基于三元组损失的步态识别方法。首先对步态能量图和活性能量图两种步态特征表示方法进行了介绍,并提取了四种不同的步态特征表示进行比较,然后详细介绍了基于三元组损失的步态识别方法的原理及公式推导,在此基础上进行了多组对比实验,分别探究了几种步态特征表示的优劣以及所提出方法与其他度量学习方法的性能比较。

第四章针对步态识别中部分轮廓图像之间差异过大的问题,提出一种基于 L1 距离度量的步态识别方法。首先对方法中涉及的公式进行了详细的推导,并说明所设计的损失函数的具体原理,然后介绍了实验所使用的数据集及实验方式的设计,最后在公开的数据集上进行了一系列对比实验,并结合实验数据论证本文方法能够应对步态识别任务中面临的挑战。

第五章首先对本文所做的步态识别研究领域的主要工作进行了总结,然后分析了当前步态识别技术中存在的问题,指出本文方法仍待提升的方面,最后对步态识别未来大规模应用的研究提出一些可能的方向和建议。



## 第二章 理论基础和相关技术

本章将对度量学习和步态识别这两个领域的基础理论和相关技术作简单的介绍，度量学习方面，主要介绍常用的距离度量和损失函数；步态识别方面，主要介绍步态轮廓的分割以及轮廓图像的归一化处理。第三小节中，还介绍了两个广泛使用的步态数据集和步态识别实验的评价指标，为后续章节的内容做铺垫。

### 2.1 度量学习

#### 2.1.1 距离度量

距离度量学习是分类、聚类、检索等应用中的一个基本问题，其核心思想就是寻找合适的距离度量来确定输入数据之间的相似性或者相异性。在使用最为广泛的分类任务中，度量学习可以通过在样本数据上学习某一种距离度量，使得同类样本之间的距离尽量近，不同类样本之间的距离尽可能远，提升对样本数据的区分能力。

设  $\mathbf{X}$  是一个样本数据集合， $\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k \in \mathbf{X}$  为该数据集合中任意 3 个具有相同维度的向量，则  $\mathbf{X}$  上的一个距离度量  $d: \mathbf{X} \times \mathbf{X} \rightarrow \mathbb{R}$  应同时满足以下四点性质<sup>[22]</sup>：

- (1) 非负性：  $d(\mathbf{x}_i, \mathbf{x}_j) \geq 0$
- (2) 对称性：  $d(\mathbf{x}_i, \mathbf{x}_j) = d(\mathbf{x}_j, \mathbf{x}_i)$
- (3) 一致性：  $d(\mathbf{x}_i, \mathbf{x}_j) = 0 \Leftrightarrow \mathbf{x}_i = \mathbf{x}_j$
- (4) 次可加性：  $d(\mathbf{x}_i, \mathbf{x}_j) + d(\mathbf{x}_j, \mathbf{x}_k) \geq d(\mathbf{x}_i, \mathbf{x}_k)$

L1 距离又称曼哈顿距离，它将两个样本  $\mathbf{x}_i, \mathbf{x}_j$  之间的距离定义为：

$$d(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\|_1 \quad (2-1)$$

欧氏距离在早期度量学习中有着非常广泛的应用，其定义可以写作式 (2-2) 的形式：

$$d(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\|_2 = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j)} \quad (2-2)$$

余弦距离常常用于基于对约束的度量学习中，它的定义如式 (2-3) 所示：

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{1 - \frac{\mathbf{x}_i^T \mathbf{x}_j}{\|\mathbf{x}_i\| \cdot \|\mathbf{x}_j\|}} \quad (2-3)$$

需要注意的是，由于余弦距离并不满足次可加性，因此并不是一种严格的距离度量。但是余弦距离在人脸验证<sup>[51-52]</sup>，文本分类<sup>[53]</sup>等任务中仍然具有重要的作用。

马氏距离 (Mahalanobis distance) 是一种能够很好挖掘样本数据中隐藏的特征和联系的距离度量, 在度量学习中具有非常深远的影响, 它的定义如式 (2-4) 所示:

$$d_M(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T \Sigma^{-1} (\mathbf{x}_i - \mathbf{x}_j)} \quad (2-4)$$

其中,  $\Sigma$  是原始数据中各维度之间的协方差矩阵。从上式可以看出, 马氏距离能够很好地兼顾到数据各个维度方差的影响, 从这个角度看, 马氏距离实际上是欧氏距离的一种扩展。

### 2.1.2 损失函数

近年来, 度量学习逐渐与深度学习相结合, 与传统度量学习相比, 深度度量学习在应对非线性问题上有了新的突破。在深度度量学习中, 损失函数扮演着重要的角色, 它通过提供一种直观的约束, 使得所学习的映射能够更好地在特征空间中对样本进行区分。

对比损失<sup>[34]</sup>基于样本对进行优化, 它将正样本对之间的距离缩小同时将负样本对之间的距离扩大, 设  $D(\mathbf{x}_1, \mathbf{x}_2)$  为一对样本之间的距离, 则对比损失可以描述为式 (2-5) 的形式:

$$L_{Contrastive} = (1-Y) \frac{1}{2} D^2 + Y \frac{1}{2} \{\max(0, m-D)\}^2 \quad (2-5)$$

其中,  $m$  是设置的边界值,  $Y$  是样本对的标签值, 当输入正样本对时  $Y=0$ , 输入负样本对时  $Y=1$ 。从式 (2-5) 中可以直观看出, 对比损失根据一对样本是否来自同一类施加不同的策略, 对于正样本对, 对比损失的值本质就是样本之间的距离, 此时通过最小化对比损失可以缩小在特征空间中同一类样本之间的距离; 而对于负样本对, 只有当负样本对之间的距离小于给定的边界值  $m$  的时候才会触发损失, 此时优化过程的本质是扩大不同类样本在特征空间内的距离。

三元组损失<sup>[35]</sup>将三个输入样本作为一组, 其中包含一个固定的锚点样本  $\mathbf{x}_{anchor}$ , 一个与锚点同类的正样本  $\mathbf{x}_p$ , 一个与锚点异类的负样本  $\mathbf{x}_n$ , 其表达式可以用式 (2-6) 描述:

$$L_{Triple} = \max\{0, D(\mathbf{x}_{anchor}, \mathbf{x}_p) - D(\mathbf{x}_{anchor}, \mathbf{x}_n) + \alpha\} \quad (2-6)$$

其中,  $\alpha$  是预设的边界值。相较于对比损失, 三元组损失更加直接地将正样本对和负样本对同时进行优化, 将类内距离和类间距离之间的相对关系也加入到优化的过程中。

三元组损失的约束作用于锚点和正负样本之间, 结构损失<sup>[54]</sup>则进一步扩大了约束作用的范围, 将样本按批次输入并对批次内所有样本对都进行约束, 具体地说, 设  $P$  为正样本对集合,  $N$  为负样本对集合, 结构损失的表达式如式 (2-7) (2-8) 所示:

$$L_{Structured} = \frac{1}{2|P|} \sum_{(i,j) \in P} \max[0, L(i, j)]^2 \quad (2-7)$$

$$L(i, j) = \log \left( \sum_{(i,k) \in N} \exp\{\alpha - D(\mathbf{x}_i, \mathbf{x}_k)\} + \sum_{(j,l) \in N} \exp\{\alpha - D(\mathbf{x}_j, \mathbf{x}_l)\} \right) + D(\mathbf{x}_i, \mathbf{x}_j) \quad (2-8)$$

结构损失将基于对的距离向量扩展到距离矩阵，通过批处理提升了对样本数据特征的提取能力。图 2-1 显示了三种损失函数的原理和区别。

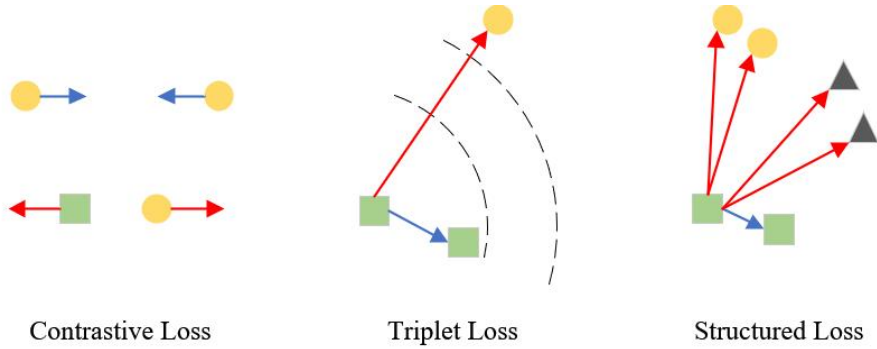


图 2-1 三种损失函数对比

Fig.2-1 Comparison of the three loss functions

## 2.2 步态预处理

步态特征的预处理是步态识别任务中的第一步，由于步态与人脸、指纹等静态生物特征不同，它是一种动态的特征，因此理想的步态特征需要尽可能地保留原始数据中的动态性。根据步态识别方法的不同，步态特征可以分为基于模型的和基于轮廓图像的，本文的研究主要聚焦于基于轮廓图像的步态识别方法，因此本节将主要介绍基于轮廓的步态特征预处理方法。

### 2.2.1 步态轮廓分割

行人的行走是一个动态的过程，因此基于轮廓图像的步态数据往往以视频的形式采集。原始视频中往往包含了一些背景信息，而在步态识别算法中，运动的行人是我们关注的重点，因此在特征提取的过程中需要把不必要的背景信息去除，从原始视频中将人体轮廓提取出来。对视频中的行走目标进行检测是基于轮廓图像的步态识别任务的基础，本节主要介绍目前流行的三种方法，包括背景差分法<sup>[55]</sup>，帧间差分法<sup>[56]</sup>和光流法<sup>[57]</sup>。

### (1) 背景差分法

在步态视频中，我们需要的是人体轮廓，背景则是需要去除的部分，因此背景差分法从这个角度出发，将每一帧中的人体轮廓提取出来。具体实现为，先对数据中不包含行人的背景进行建模，然后将每一帧图像都与背景图像求差值，得到的就是该帧图像中的动态目标。即：

$$M(x, y) = |f_t(x, y) - B(x, y)| \quad (2-9)$$

式中， $B(x, y)$ 为所建立的背景图像， $f_t(x, y)$ 表示当前提取的 $t$ 时刻的帧图像， $M(x, y)$ 表示从当前帧中提取得到的动态目标，将得到的目标与一个预先设置好的阈值进行比较，就得到当前帧图像中的人体轮廓，如式（2-10）所示：

$$S(x, y) = \begin{cases} 1, & M(x, y) \geq \tau \\ 0, & M(x, y) < \tau \end{cases} \quad (2-10)$$

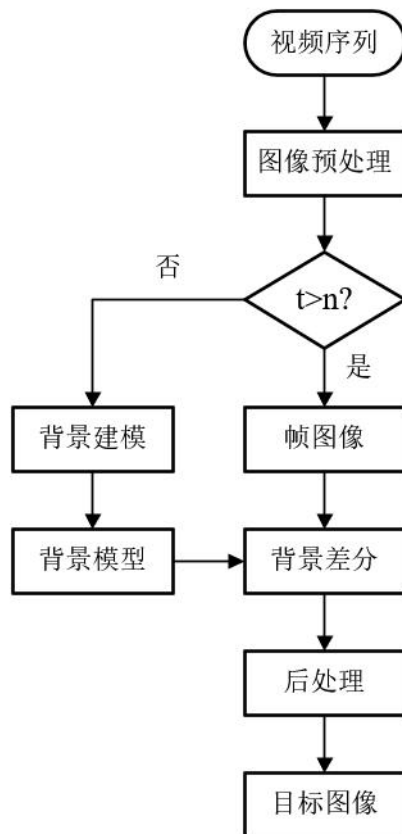


图 2-2 背景差分法的计算流程

Fig.2-2 The calculation flow of the background subtraction algorithm

背景差分法是一种比较简单快捷的目标检测方法，其提取目标的过程如图 2-2 所示，计算过程并不复杂，但是由于该方法中所有轮廓图像都需要通过帧图像与背景图

像相减获得，背景图像的建模质量将直接影响到人体轮廓的采集。在一些真实的识别场景中，背景往往会因为光线，遮挡等因素会发生一些变化，这些变化会直接影响到人体轮廓图像的质量，进而影响识别的精度，因此，在场景比较复杂的情况下，背景差分法很难从图像中提取到高质量的目标。针对传统的背景建模法在背景变化场景下的缺陷，有很多相关的改进，其中最经典的是混合高斯背景建模法<sup>[58]</sup>。

混合高斯建模法将图像内的像素值建模为随图像帧变化而更新的高斯混合模型，并根据模型中每个高斯的持续性和方差来确定某个像素对应的是背景还是前景。具体而言，每个像素在最近的 $t$ 时刻内的灰度值由 $(G_1, G_2, \dots, G_t)$ 表示，则 $G_t$ 的概率密度函数描述为：

$$P(G_t) = \sum_{i=1}^K \omega_{i,t} \cdot \eta(G_t, \mu_{i,t}, \Sigma_{i,t}) \quad (2-11)$$

式中， $K$ 是每个像素高斯分布的数量， $\omega_{i,t}$ 、 $\mu_{i,t}$ 和 $\Sigma_{i,t}$ 分别表示 $t$ 时刻第 $i$ 个高斯分布的权重的估计值，期望值和协方差矩阵，概率密度函数 $\eta$ 可以用式(2-12)描述：

$$\eta(G_t, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left[ -\frac{1}{2} (G_t - \mu)^T \Sigma^{-1} (G_t - \mu) \right] \quad (2-12)$$

$$|G_{t+1} - \mu_{i,t}| < 2.5\sigma_{i,t} \quad (2-13)$$

后续将当前帧的每一个像素与已经建立的 $K$ 个高斯分布按式(2-13)的规则进行匹配，其中 $\sigma_{i,t}$ 表示第 $i$ 个高斯分量的标准差。若满足匹配条件，则认为该像素为背景值并对已有的参数进行更新，更新的具体方式如式(2-14)所示：

$$\begin{cases} \mu_{i,t+1} = (1-\alpha)\mu_{i,t} + \alpha G_{t+1} \\ \sigma_{i,t+1} = (1-\beta)\sigma_{i,t} + \beta (G_{t+1} - \mu_{i,t})^2 \\ \beta = \frac{\alpha}{\omega_{i,t}} \end{cases} \quad (2-14)$$

其中， $\alpha$ 、 $\beta$ 为更新系数。若该像素与 $K$ 个高斯分布均不能匹配，则认为该像素为检测到的运动目标。

混合高斯建模法是对传统背景差分法中背景图像建模的一种优化，它能够应对建模过程中背景发生一定的细微变化导致建模质量下降的问题。当然，混合高斯建模法也有一定的缺陷，如在第一帧场景中即存在运动目标的情况下，该运动目标可能会被识别为背景，进而影响后续的检测和跟踪。

## (2) 帧间差分法

行人的行走是一个动态的过程，体现在图像序列中就是相邻的图像帧之间会存在一定的差异，这些差异就是动态过程的一种表现，帧间差分法的原理就是通过对相邻



帧求差值来获取视频中的人体轮廓。帧间差分法可用式 (2-15) 表示:

$$S(x, y) = \begin{cases} 1, & |f_k(x, y) - f_{k-n}(x, y)| \geq \tau \\ 0, & |f_k(x, y) - f_{k-n}(x, y)| < \tau \end{cases} \quad (2-15)$$

其中,  $f_{k-n}(x, y)$  表示在  $f_k(x, y)$  之前的若干帧图像。

跟背景差分法相比, 由于帧间差分法是在相邻帧之间取差值来提取特征, 因此帧间差分法对背景的变化具有一定的鲁棒性。但是, 当帧图像之间的动作变化较为缓慢的时候, 相邻帧之间的差异较小, 此时帧间差分法所得到的差值图像可能会丢失一部分人体轮廓, 导致得到的轮廓图像不完整。总的来说, 帧间差分法是一种实现简单, 计算较快的目标检测方法, 但是其缺点是对运动目标的速度变化比较敏感, 为了保证分割图像的质量需要根据目标运动的速度对帧差的值进行调整。

### (3) 光流法

光流法认为, 实际场景中的运动可以在图像中通过像素的灰度值变化表现出来。具体来说, 在一个静止的场景中, 由于环境光线的缓慢变化, 背景图像中每一个像素的灰度值也是随时间缓慢变化的, 由此可以将背景建模为一个矢量场。当场景中有运动目标出现时, 由于背景是静止的而目标处在运动中, 目标处的灰度值矢量与背景的矢量会存在一定的差异, 根据这些差异可以在图像中确定运动目标。

我们将图像上某一点  $p(x, y)$  在  $t$  时刻的灰度值定义为  $G(x, y, t)$ , 该点在  $dt$  时间后运动到了另一点  $p'(x+dx, y+dy)$ , 灰度值也变为  $G'(x+dx, y+dy, t+dt)$ , 假设运动前后该目标在图像上的灰度值不发生变化, 则可以得到:

$$G(x, y, t) = G(x+dx, y+dy, t+dt) \quad (2-16)$$

泰勒展开后, 得到式 (2-17):

$$G(x+dx, y+dy, t+dt) = G(x, y, t) + \frac{\partial G}{\partial x} dx + \frac{\partial G}{\partial y} dy + \frac{\partial G}{\partial t} dt + \varepsilon \quad (2-17)$$

其中  $\varepsilon$  为无穷小, 令  $G_x = \frac{\partial G}{\partial x}$ ,  $G_y = \frac{\partial G}{\partial y}$ ,  $G_t = \frac{\partial G}{\partial t}$  分别表示目标像素灰度值随  $x$ ,  $y$ ,

$t$  的变化,  $u = \frac{dx}{dt}$ ,  $v = \frac{dy}{dt}$  表示检测目标在  $x$ ,  $y$  方向上的运动速度, 结合式 (2-16)

和式 (2-17), 进一步整理可得约束方程:

$$\frac{\partial G}{\partial x} dx + \frac{\partial G}{\partial y} dy + \frac{\partial G}{\partial t} dt = G_x u + G_y v + G_t = 0 \quad (2-18)$$

使用光流法检测运动目标时, 需要先对识别的场景图像建立光流场, 再依据给定

的约束条件对图像中的运动目标进行检测。由于光流法是根据检测目标与背景之间的相对运动带来的灰度矢量差异进行检测的,因此对背景的变化并不敏感,但是光流场的建立涉及到对图像中所有像素点进行矢量初始化以及后续运算,因此光流法也面临着计算量较大的问题。

总的来说,三种轮廓分割方法各有优劣,在实际的应用中,可以依据不同的情况对这三种方法进行选择或者组合使用,以获得质量较好的轮廓图像。

### 2.2.2 轮廓图像归一化

在步态识别场景中,图像采集设备往往是固定的,而待识别对象处在行走过程中,其与摄像头的距离、角度是不断变化的,因此采集到的图像中人体轮廓往往是大小不一的,这不利于后续的特征提取。同时,在后续的认识过程中,算法主要关注的是行走的对象,而采集到的图像中包含背景中一些冗余的信息,这些信息也会使算法的性能下降。因此,对分割得到的轮廓图像需要进行对齐和归一化。

首先,对得到的二值轮廓图像进行遍历,自上而下将每一行的像素值累加,首个像素值和不为0的索引即为人体轮廓的头部,最后一个像素值和不为0的索引即为人体轮廓的脚底,同理可以确定人体轮廓的左右边界。为了避免图像中一些噪点的干扰,得到轮廓的范围之后需要统计该范围内的轮廓面积,当面积小于一定的阈值即认为是噪点并将该区域舍弃。

得到轮廓的大致范围,就可以进行轮廓的对齐。在对齐的方式上,主流的方案是基于人体的质心进行对齐。质心坐标可以通过式(2-19)(2-20)的方式计算得到:

$$x_c = \frac{1}{N} \sum_{i=1}^N x_i \cdot i \quad (2-19)$$

$$y_c = \frac{1}{N} \sum_{i=1}^N y_i \cdot i \quad (2-20)$$

其中,  $x_i$ 、 $y_i$  分别表示横纵坐标为  $i$  处的值为1的像素数量,  $N$  表示轮廓区域值为1的总像素数量。确定轮廓的质心之后,按照给定的图像尺寸从质心向四周对图像进行裁剪,操作的流程如图(2-3)所示。

当然,本文中還加入了基于头顶进行对齐的轮廓图像进行对比实验。首先按照前述方法确定人体轮廓范围的大小,并得到轮廓的高度  $h$ 。设缩放后的图像大小设置为  $(x, y)$ , 由于人体轮廓通常高度大于宽度,将轮廓按高度即  $y/h$  缩放至指定大小,最后将头顶置于  $(x/2, 0)$  处进行裁剪。本文在预处理过程中将所有轮廓图像归一化为  $64 \times 48$ , 如图(2-4)所示。

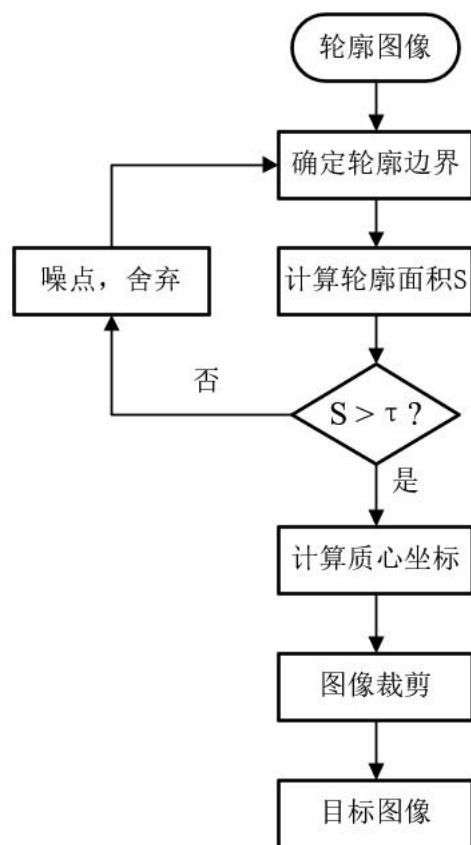


图 2-3 图像归一化流程

Fig.2-3 Image normalization process



原始图像

基于质心对齐

图 2-4 轮廓图像的归一化

Fig.2-4 Normalization of silhouette images

## 2.3 相关数据集及评价指标

自从步态作为一种可识别的生物特征被提出以来,吸引了大量研究者的关注。因为其可以实现远距离、非接触和无感知的身份识别,长期以来一直被视作传统生物识别在安防监控领域的替代方案之一。近年来,研究者提出了很多步态识别方法,也构建了一些步态数据集用于推进步态识别的研究进展。本小节介绍了两个应用比较广泛的步态数据集以及步态识别任务中常用的评价指标。

### 2.3.1 步态数据集

CASIA-C<sup>[59]</sup>是一个在夜间使用热红外摄像机采集的步态数据集,由中国科学院自动化研究所在 2005 年采集并发布。与传统的数据集相比,CASIA-C 更加关注在夜间的步态识别,因此该数据集使用热红外摄像机采集夜间图像,同时,CASIA-C 也提供了轮廓数据版本。在 CASIA-C 数据集中,共包含了 153 个不同的样本,每个样本拥有 4 组正常行走、2 组快步行走、2 组慢步行走和 2 组背包行走共 10 组的步态序列,所有步态序列均在侧面 90°视角下采集。图 2-5 显示了 CASIA-C 数据集中的一些轮廓图像。

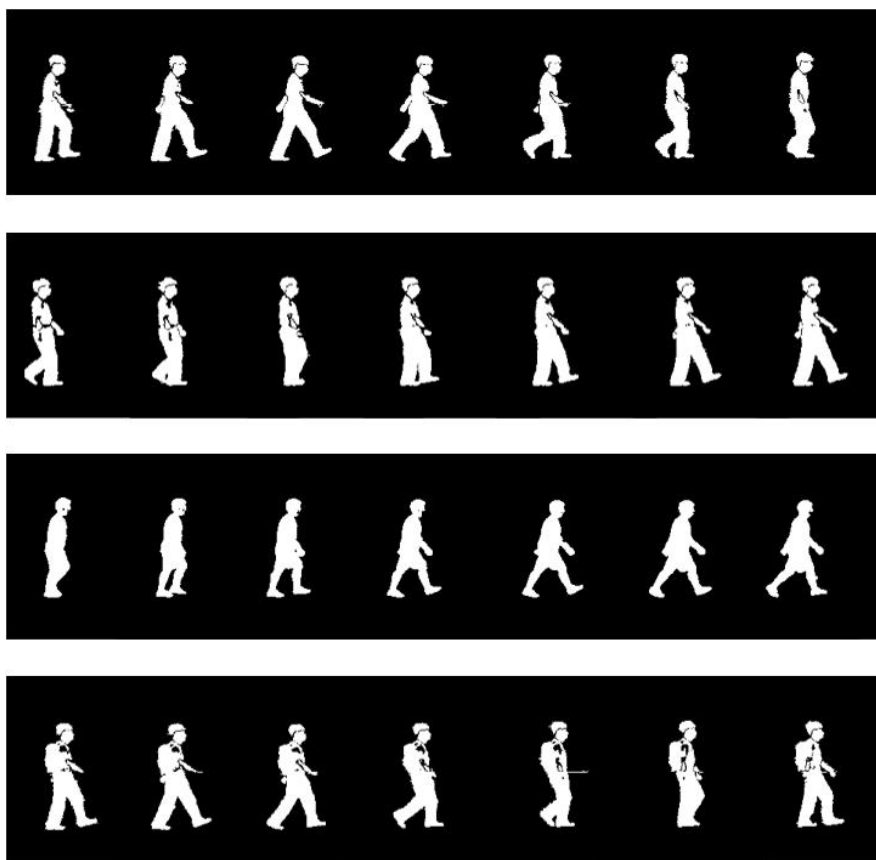


图 2-5 CASIA-C 步态数据集部分轮廓图像示例

**Fig.2-5** Examples of some silhouette images in CASIA-C gait dataset

尽管近年来大量的步态识别算法被提出，目前步态识别仍面临着行走状态以及拍摄视角变化带来的精度下降的挑战。在实际应用中，待识别者往往以各种不同的视角出现在画面中，以往的一些算法在这种情况下可能无法得到理想的结果。相较于过去的一些数据集，CASIA-B<sup>[60]</sup>是一个系统性关注行人的行走状态和视角变化的步态数据集，它同样由自动化研究所构建并发布。CASIA-B 数据集共包含了 124 个不同的样本，每个样本拥有 6 组正常行走、2 组背包行走和 2 组穿着外套行走的步态序列，同时，每一组步态序列又划分了 11 个不同的角度（0°，18°，...，180°），因此，每个样本包含了 110 组步态序列。由于对视角进行了细致地划分，CASIA-B 是一个比较系统且庞大的数据集，也是目前使用最为广泛的步态数据集之一。图 2-6 显示了 CASIA-B 数据集中的一些轮廓图像。

**图 2-6** CASIA-B 步态数据集部分轮廓图像示例**Fig.2-6** Examples of some silhouette images in CASIA-B gait dataset

### 2.3.2 评价指标

为了系统性地评价步态识别算法的性能，方便对不同的算法进行比较和评估，我们需要引入一些评价指标来将算法的性能进行直观地量化。在分类问题中，使用比较广泛的几个评价指标包括准确率（Accuracy）、精度（Precision）、召回率（Recall）、受试者工作特征曲线（Receiver Operating Characteristic Curve，ROC）等。

**表 2-1** 混淆矩阵**Table 2-1** Confusion Matrix

真实类别	预测类别	
	正例	反例
正例	TP	FN
反例	FP	TN

在分类任务中，可以根据一个样本所属的类别以及该样本在算法模型中被预测的类别进行划分，具体如表 2-1 所示。在算法的评价指标中，准确率的定义如式（2-21）所示：

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2-21)$$

由式（2-21）的定义可知，准确率的实际含义是所有预测正确的样本数占总样本数的比例。需要进行区别的是精度和召回率，精度的定义是在模型所有预测为正例的样本中，预测正确的比例，而召回率评估的是模型对正样本的召回能力，具体的定义可用式（2-22）（2-23）描述。

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2-22)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2-23)$$

F1-score 是一个基于精度和召回率产生的指标，其计算公式如式（2-24）所示。

$$F_1 = \frac{2 \times P \times R}{P+R} \quad (2-24)$$

由于在步态识别任务中，往往存在着大量的样本标签，在这样的情况下，一些二分类的指标不能够全面地评估模型的性能，因此需要引入 Macro 度量和 Micro 度量的概念。在步态识别中，设某一样本的类别标签为  $i$ ，则其余  $N-i$  个标签可以视为该样本的反例，在此基础上可以得到该类别对应的  $TP_i$ 、 $TN_i$ 、 $FN_i$ 、 $FP_i$ ，则此时 Macro 度量下精度和召回率可以定义为：

$$\text{macro-P} = \frac{1}{N} \sum_{i=1}^N \frac{TP_i}{TP_i+FP_i} \quad (2-25)$$

$$\text{macro-R} = \frac{1}{N} \sum_{i=1}^N \frac{TP_i}{TP_i+FN_i} \quad (2-26)$$

从式（2-25）（2-26）可以看出，Macro 度量对每一个类别分别求精度和召回率，并最终将所有类别的结果求平均值得到模型在数据集上的精度和召回率。而 Micro 度量下的定义为：

$$\text{micro-P} = \frac{\sum_{i=1}^N TP_i}{\sum_{i=1}^N TP_i + \sum_{i=1}^N FP_i} \quad (2-27)$$

$$\text{micro-R} = \frac{\sum_{i=1}^N \text{TP}_i}{\sum_{i=1}^N \text{TP}_i + \sum_{i=1}^N \text{FN}_i} \quad (2-28)$$

由上式可以看出，Micro 度量充分考虑了所有类别的数量，在一些样本数量不平衡的数据中可以对模型进行更加合理的评估，但是在数据中存在某些样本数量极大的情况下，这些样本类别的结果会影响全局的指标。

## 2.4 本章小结

本章主要对度量学习和步态识别两个领域的一些基础理论和相关技术作了介绍。第一小节对度量学习中涉及到的一些经典的距离度量和损失函数进行了介绍，第二小节从步态特征提取的角度出发，介绍了步态轮廓分割以及图像归一化的相关方法，最后第三小节介绍了两个常用的步态数据集和步态识别算法的评价指标，为后续的实验设计打下基础。

## 第三章 基于三元组损失的步态识别方法

### 3.1 步态特征提取

步态是一种动态的且具有周期性的生物特征，步态数据往往是以视频的形式采集，为了尽可能保留数据中的动态性，基于轮廓图像的步态识别通常将视频中的帧图像处理为步态模板作为特征输入。本小节将主要介绍本章实验中所使用的两种步态模板以及后续的特征降维算法。

#### 3.1.1 步态能量图

人的行走是一个具有周期性的动作，每一个周期内的动作流程都是固定不变的。由于每个个体在骨骼、肌肉、体态以及行走习惯方面的差异，步态特征之间也会存在差异。因此，可以将一系列有顺序的步态帧图像处理为单张的二维步态模板，作为行走过程的步态特征表示。具体地说，将一个步态周期内的帧图像进行归一化之后，将周期内的帧图像相加并求平均值就得到了一张该周期内的步态能量图，计算公式如式(3-1)所示：

$$\mathbf{G}(x, y) = \frac{1}{N} \sum_{i=1}^N \mathbf{B}_i(x, y) \quad (3-1)$$

其中， $N$  表示叠加计算的帧图像的数量。

根据帧图像归一化的对齐方式不同，本文的实验中将步态能量图又分为基于头顶对齐的和基于质心对齐的，如图 3-1 和图 3-2 所示。



图 3-1 基于质心对齐的步态能量图

Fig.3-1 Gait energy image aligned on centroid



图 3-2 基于头顶对齐的步态能量图



**Fig.3-2** Gait energy image aligned on head

### 3.1.2 活性能量图

在步态识别任务中，行人的行走状态变化如携带背包以及着装变化等因素一直是基于轮廓图像的步态识别面临的重要的挑战之一。行走状态的变化会给人体轮廓图像带来一定的影响，这样的影响在静态图像中尤为明显。活性能量图<sup>[12]</sup>（Active Energy Image, AEI）是通过对步态序列中相邻两帧的差值求平均值得到的，由于相邻两帧的差值去除了图像中静态的部分，在一定程度上能够减小行走状态的变化对轮廓带来的影响，同时也能保留步态序列中更多的动态信息。其计算方式可以用式（3-2）（3-3）来描述。

$$A(x, y) = \frac{1}{N} \sum_{i=1}^N D_i(x, y) \quad (3-2)$$

$$D_i(x, y) = \begin{cases} f_i(x, y), & i = 0 \\ \|f_i(x, y) - f_{i-1}(x, y)\|, & i > 0 \end{cases} \quad (3-3)$$

图 3-3 和 3-4 中显示了部分基于头顶对齐和基于质心对齐的活性能量图。

**图 3-3** 基于质心对齐的活性能量图**Fig.3-3** Active energy image aligned on centroid**图 3-4** 基于头顶对齐的活性能量图**Fig.3-4** Active energy image aligned on head

### 3.1.3 特征降维

为了进一步去除数据中的冗余信息和噪声干扰，加快模型的运算速度，从步态序列中提取特征图像之后，还需要对特征数据进行降维。实验在提取到轮廓图像之后，借助主成分分析方法，降低步态特征数据的维度。

PCA 算法的基本思想是将一个较高维度的特征映射到一个相对较低的维度中，映射的同时保留原始数据中一些主要的特征，抛弃一些相对次要的特征，实现对数据的

压缩和去噪。在实际使用中，一般会选择方差最大的方向作为数据的主要特征维度，如图 3-5 所示。

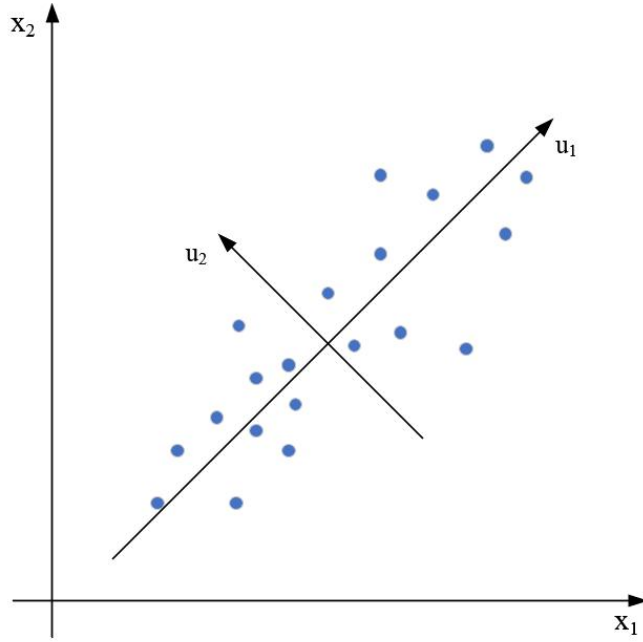


图 3-5 主成分分析方法示意图

Fig.3-5 Schematic diagram of principal component analysis method

设  $\mathbf{X} = \{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^k\}$  是一个  $n$  维的样本集合，包含  $k$  个样本，则该集合可展开为式 (3-4) 的形式：

$$\mathbf{X} = \begin{pmatrix} x_1^1 & x_1^2 & \dots & x_1^k \\ x_2^1 & x_2^2 & \dots & x_2^k \\ \dots & \dots & \dots & \dots \\ x_n^1 & x_n^2 & \dots & x_n^k \end{pmatrix} \quad (3-4)$$

将集合中每一个样本按式 (3-5) 的方式进行标准化：

$$\mathbf{x}^i = \mathbf{x}^i - \frac{1}{k} \sum_{j=1}^k \mathbf{x}^j \quad (3-5)$$

对标准化后的数据中  $n$  维特征两两求协方差，得到大小为  $n \times n$  的协方差矩阵  $\mathbf{C}$ ，该矩阵中的任一元素  $c_{ij}$  表示第  $i$  维特征和第  $j$  维特征的协方差，对角线元素  $c_{ii}$  则表示第  $i$  维特征的方差。将矩阵  $\mathbf{C}$  的特征值按大小降序排列后，将与特征值对应的前  $m$  个特征向量取出，得到：

$$\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m) \quad (3-6)$$

将原数据通过投影矩阵  $\mathbf{W}$  投影，即可得到降至  $m$  维后的数据集  $\mathbf{D}$ ：

$$\mathbf{D} = \mathbf{W}\mathbf{X} \quad (3-7)$$

## 3.2 损失函数及优化

### 3.2.1 目标函数

回顾 2.1 小节可知，三元组损失的描述为：

$$L_{Triple} = \max\{0, d(\mathbf{x}_{anchor}, \mathbf{x}_p) - d(\mathbf{x}_{anchor}, \mathbf{x}_n) + \alpha\} \quad (3-8)$$

其中， $\mathbf{x}_{anchor}$  表示锚点样本， $\mathbf{x}_p$ 、 $\mathbf{x}_n$  分别表示与锚点同类的正样本和与锚点不同类的负样本， $d(\mathbf{x}_i, \mathbf{x}_j)$  表示一对样本之间的距离度量。从上式可以看出，三元组损失的核心思想就是对锚点和负样本的距离与锚点和正样本的距离之差进行约束，当差值小于一个给定的阈值时则施加惩罚，优化的目的是使得负样本和锚点的距离比正样本和锚点的距离大一定的间隔，从而提高算法对不同类样本的判别力。

设一个数据集中包含  $N$  个样本，将其中任意一个样本  $\mathbf{x}_i \in \mathbb{R}^d$  通过线性变换  $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k] \in \mathbb{R}^{d \times s}$ ,  $s \leq d$  投影到特征空间中，其中  $d$  和  $s$  分别是投影前后样本的维度，则在投影的特征空间中，一对样本  $\mathbf{x}_i$  和  $\mathbf{x}_j$  之间的欧式距离可以描述为：

$$d_{\mathbf{W}}(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j\|_2 \quad (3-9)$$

结合式 (3-8) 和 (3-9)，实验所提出的目标函数可以描述为以下形式：

$$\min_{\mathbf{W}} J = \sum_{i=1}^N h(d_{\mathbf{W}}(\mathbf{x}_a, \mathbf{x}_p) - d_{\mathbf{W}}(\mathbf{x}_a, \mathbf{x}_n) + \alpha) \quad (3-10)$$

其中， $\alpha$  为给定的阈值，用于约束锚点样本和正负样本距离的差值， $h(x) = \max(x, 0)$  是铰链损失函数 (Hinge Loss Function)。

### 3.2.2 算法实现

为了求解式 (3-10) 中的目标函数，本章使用梯度下降法，首先将目标函数对投影矩阵  $\mathbf{W}$  求偏导，得到式 (3-11)。

$$\frac{\partial J}{\partial \mathbf{W}} = \sum_{i=1}^N h'(d_{\mathbf{W}}(\mathbf{x}_a, \mathbf{x}_p) - d_{\mathbf{W}}(\mathbf{x}_a, \mathbf{x}_n) + \alpha) \left( \frac{\partial d_{\mathbf{W}}(\mathbf{x}_a, \mathbf{x}_p)}{\partial \mathbf{W}} - \frac{\partial d_{\mathbf{W}}(\mathbf{x}_a, \mathbf{x}_n)}{\partial \mathbf{W}} \right) \quad (3-11)$$

为进一步求解式 (3-11)，需要知道样本间的距离度量对投影矩阵  $\mathbf{W}$  的偏导，有：

$$\frac{\partial d_{\mathbf{W}}(\mathbf{x}_a, \mathbf{x}_p)}{\partial \mathbf{W}} = \frac{(\mathbf{x}_a - \mathbf{x}_p)(\mathbf{x}_a^T \mathbf{W} - \mathbf{x}_p^T \mathbf{W})}{\|\mathbf{W}^T \mathbf{x}_a - \mathbf{W}^T \mathbf{x}_p\|_2} \quad (3-12)$$

同理可得：

$$\frac{\partial d_{\mathbf{W}}(\mathbf{x}_a, \mathbf{x}_n)}{\partial \mathbf{W}} = \frac{(\mathbf{x}_a - \mathbf{x}_n)(\mathbf{x}_a^T \mathbf{W} - \mathbf{x}_n^T \mathbf{W})}{\|\mathbf{W}^T \mathbf{x}_a - \mathbf{W}^T \mathbf{x}_n\|_2} \quad (3-13)$$

将式 (3-12) (3-13) 代入式 (3-11) 即可得到目标函数对投影矩阵的偏导，最后通过式 (3-14) 所示的方法对  $\mathbf{W}$  进行迭代更新，并最终得到最优的投影矩阵。

$$\mathbf{W} = \mathbf{W} - \lambda \frac{\partial J}{\partial \mathbf{W}} \quad (3-14)$$

式中， $\lambda$  为学习率。

表 3-1 本章方法的求解过程

Table 3-1 The solution process of the proposed method

算法 1：本章方法的求解过程

输入：训练数据集合及标签信息

    阈值  $\alpha$

    学习率  $\lambda$

    误差值  $\varepsilon$

输出：投影矩阵  $\mathbf{W}$

1. 初始化投影矩阵  $\mathbf{W}$
2. 计算式 (3-10) 的目标函数值  $J$
3. FOR  $i=1,2,\dots,T$
4.     计算式 (3-11) 中的梯度  $\frac{\partial J}{\partial \mathbf{W}}$
5.     通过式 (3-14) 更新投影矩阵  $\mathbf{W}$
6.     通过式 (3-10) 更新目标函数值  $J_i$
7.     IF  $|J_i - J_{i-1}| < \varepsilon$
8.         BREAK
9.     END IF
10. END FOR
11. RETURN  $\mathbf{W}$

### 3.3 实验结果及分析

本章在两个使用比较广泛的步态数据集上分别进行了两组对比实验，第一组实验

对比了几种步态特征表示之间的差异,筛选了性能较好的步态特征表示方式;第二组实验对比了本章提出的基于三元组损失的步态识别方法与几种经典的度量学习方法在步态识别任务中的性能表现,验证所提出方法相较于传统度量学习方法的优势。

### 3.3.1 实验设置

本章的实验分别在 CASIA-B 和 CASIA-C 两个数据集上开展。首先,在两个数据集上分别进行轮廓图像的预处理,将所有图像通过对齐裁剪之后,生成基于质心对齐的步态能量图 (centroid\_GEI)、基于头顶对齐的步态能量图 (head\_GEI)、基于质心对齐的活性能量图 (centroid\_AEI) 和基于头顶对齐的活性能量图 (head\_AEI) 四种特征,所有特征图像的大小统一设置为  $64 \times 32$ 。

其中, CASIA-B 数据集包含 124 个样本,每个样本有 6 个正常行走、2 个背包行走以及 2 个穿着外套行走的步态序列,每个行走状态下又细分了 11 个不同的角度。本章的实验使用前 62 个样本的所有数据作为训练集,后 62 个样本的前 4 个正常行走序列 (NM#1-4) 作为注册集,为了验证模型在行走状态变化的情形下的性能表现,划分了三个验证集,分别为后 62 个样本的 2 个正常行走序列 (NM#5-6)、2 个背包行走序列 (BG#1-2) 以及两个穿外套行走的序列 (CL#1-2),考虑到角度的影响,在实验中对每一个角度都单独测试了识别率。CASIA-C 数据集共包含 153 个样本,实验使用前 76 个样本的所有数据作为训练集,后 77 个样本中,使用每个样本的前 2 个正常行走的序列 (fn00-01) 作为注册集,验证集分为 4 个,分别为 2 个正常行走的序列 (fn02-03)、2 个快步行走的序列 (fq00-01)、2 个慢步行走的序列 (fs00-01) 以及 2 个背包行走的序列 (fb00-01)。实验将所有特征转化为  $1 \times 2048$  维的特征向量并使用 PCA 降至 500 维。

实验在系统为 Ubuntu 18.04 的服务器上运行,首先,实验对比了四种特征的性能表现,分析四种特征在步态识别中的优劣。其次,比较了本章提出的基于三元组损失的方法与几种经典的度量学习方法之间的性能,包括 CSML<sup>[61]</sup>, SILD<sup>[62]</sup>和 ITML<sup>[63]</sup>。

### 3.3.2 不同特征对比实验

本节在 CASIA-B 数据集上对比了 centroid\_GEI、head\_GEI、centroid\_AEI 和 head\_AEI 四种特征的性能表现,表 3-2 显示了四种特征在 CASIA-B 上的平均准确率。

表 3-2 四种特征在 CASIA-B 数据集上的平均准确率(%)  
**Table 3-2** Average accuracy(%) of the four features on CASIA-B dataset

	NM#5-6	BG#1-2	CL#1-2	Mean
centroid_GEI	98.8	58.4	16.6	57.9
head_GEI	98.6	65.9	20.2	61.6
centroid_AEI	98.5	68.7	17.5	61.6
head_AEI	98.2	82.4	24.0	68.2

从表 3-2 可以看出, 基于头顶对齐的特征在正常行走的情况下准确率低于基于质心对齐的特征, 但是在行走状态变化的情况下, 如在 BG#1-2 组中, head\_GEI 特征比 centroid\_GEI 特征的准确率高出 7.5%, head\_AEI 特征比 centroid\_AEI 特征的准确率高出 13.7%, CL#1-2 组中, 基于头顶对齐的特征相比基于质心对齐的特征也都具有明显的优势, 最终, 基于头顶对齐的特征在 CASIA-B 数据集上取得了较高的平均准确率。从图 3-3 和图 3-4 的对比不难看出, 基于头顶对齐的特征图像在上半身的轮廓更加清晰, 可以将行走过程中的动态特征更多地集中到肢体部分, 一定程度上减轻行人携带物品对轮廓图像带来的影响。

为了比较不同特征在应对角度变化时的性能表现, 实验还绘制了四种特征在 BG#1-2 和 CL#1-2 组不同角度下准确率的折线图, 如图 3-6、3-7 所示。

从图中可以看出, 在角度发生变化时, head\_AEI 比其他三种特征取得更稳定的性能表现。这是由于以几何方式求得的质心会随着行人动作的不同而发生变化, 在累加的过程中变化的质心会导致轮廓图像的对齐并不严格, 因此基于质心对齐的特征会包含一些干扰信息, 不能很好地保留原始数据中的步态信息。此外, 相较于 GEI, AEI 所采用的逐帧相减的方式能够去除步态序列中相对静态的部分, 削弱轮廓本身静态差异的同时保留了较多的动态性。因此, 在应对行走状态以及视角变化的场景中, 可以采用基于头顶对齐的特征图像, 以提高模型的性能。本章在后续的对比实验中, 所有方法均采用 head\_AEI 作为特征输入。

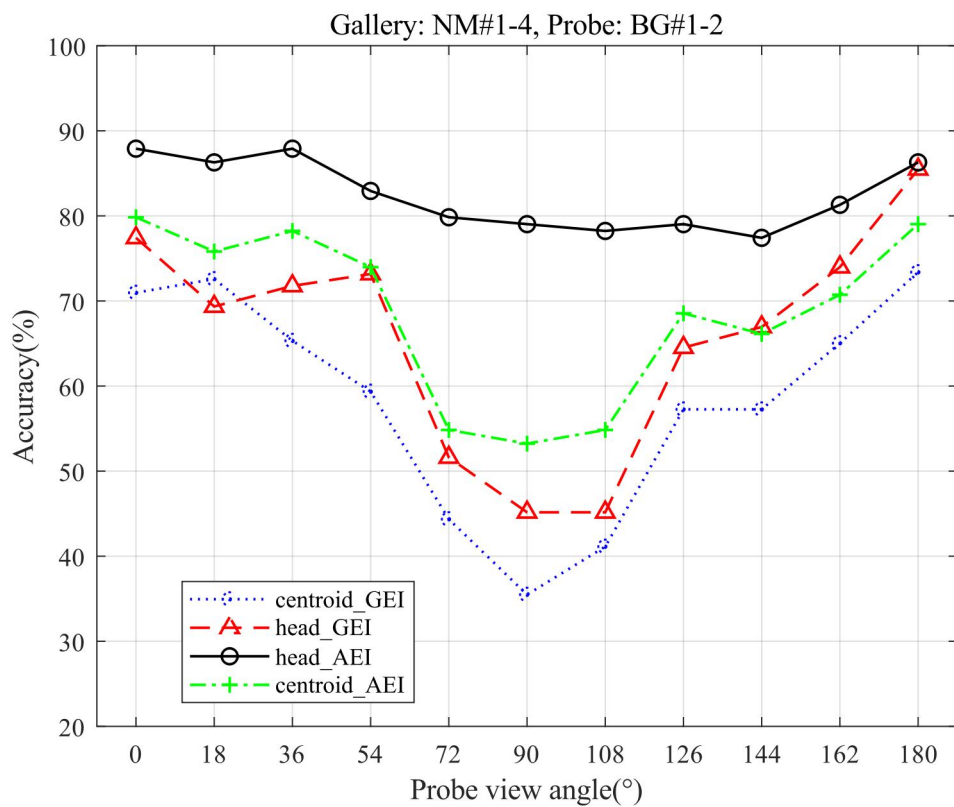


图 3-6 四种特征在 BG#1-2 中各角度下的准确率

Fig.3-6 Accuracy of the four features at various angles in BG#1-2

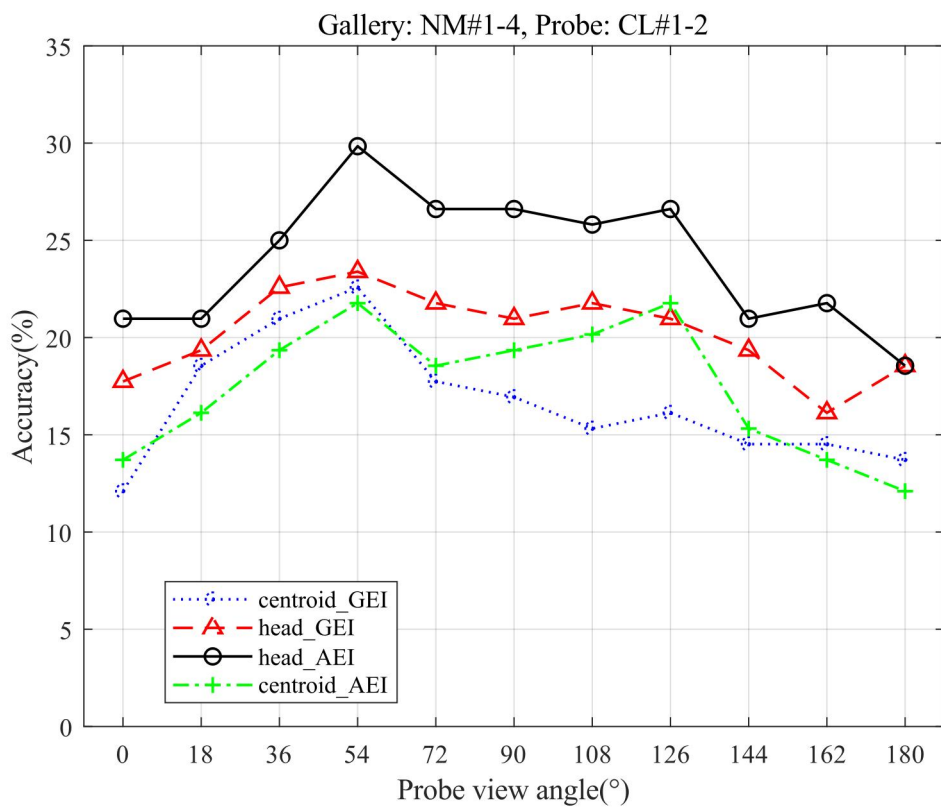


图 3-7 四种特征在 CL#1-2 中各角度下的准确率

**Fig.3-7** Accuracy of the four features at various angles in CL#1-2

## 3.3.3 与其他方法的对比实验

本小节对比了本章所提出的方法和 CSML、SILD、ITML 几种度量学习方法在步态识别任务中的识别性能，首先在 CASIA-B 数据集上进行了对比实验，结果如表 3-3 所示。

**表 3-3** 本章方法和四种算法在 CASIA-B 数据集上的准确率(%)**Table 3-3** Accuracy(%) of proposed method and four algorithms on the CASIA-B dataset

方法	NM#5-6	BG#1-2	CL#1-2
CSML	95.9	52.1	18.2
SILD	97.9	74.9	23.8
L2	97.0	76.4	22.3
ITML	94.4	51.3	9.8
本章方法	98.2	82.4	24.0

表 3-3 中显示了几种算法在 CASIA-B 数据集中每个对照组下所有角度的平均准确率。从表中可以看出，本章方法在 NM#5-6、BG#1-2、CL#1-2 三个对照组中均取得了最高的准确率，分别为 98.2%、82.4%、24.0%。其中，在 BG#1-2 组中，由于行人在携带背包的情形下轮廓变化明显，因此在该对照组中几种算法的准确率差距明显拉大，在此组别中本章方法的准确率相较于其他四种算法具有更明显的优势。而在 CL#1-2 组中，由于轮廓的差异被进一步放大，因此几种算法的准确率都有较大幅度的下降，在此对照组中本章方法同样取得最高的准确率。

**表 3-4** 本章方法和几种方法在 CASIA-C fn02-03 上的实验结果**Table 3-4** Experimental results of proposed method and several methods on CASIA-C fn02-03

方法	CCR(%)	Macro-P(%)	Macro-F1(%)
CSML	92.2	93.5	91.4
ITML	95.5	97.2	95.2
L2	96.1	97.4	95.8
SILD	82.5	86.0	81.6
本章方法	96.8	97.8	96.5



表 3-5 本章方法和几种方法在 CASIA-C fq00-01 上的实验结果

Table 3-5 Experimental results of proposed method and several methods on CASIA-C fq00-01

方法	CCR(%)	Macro-P(%)	Macro-F1(%)
CSML	71.4	71.6	68.5
ITML	81.2	80.5	78.4
L2	83.1	85.2	81.7
SILD	53.6	55.6	50.9
本章方法	89.2	88.4	88.1

表 3-6 本章方法和几种方法在 CASIA-C fs00-01 上的实验结果

Table 3-6 Experimental results of proposed method and several methods on CASIA-C fs00-01

方法	CCR(%)	Macro-P(%)	Macro-F1(%)
CSML	74.7	71.0	70.5
ITML	80.5	78.8	77.4
L2	83.1	84.0	81.7
SILD	61.7	58.1	57.0
本章方法	87.7	90.9	87.3

表 3-7 本章方法和几种方法在 CASIA-C fb00-01 上的实验结果

Table 3-7 Experimental results of proposed method and several methods on CASIA-C fb00-01

方法	CCR(%)	Macro-P(%)	Macro-F1(%)
CSML	53.2	48.5	47.7
ITML	55.2	49.9	49.3
L2	64.9	65.7	61.9
SILD	35.1	30.7	29.8
本章方法	72.7	75.1	70.6

表 3-4、3-5、3-6 和 3-7 分别显示了本章方法和几种算法在 CASIA-C 数据集中 fn02-03、fq00-01、fs00-01 和 fb00-01 的实验结果。从上述表格中可以看出，本章方法四个实验对照组中均取得了最高的精度和 Macro-F1 值。在行走状态发生变化的对照组 fq00-01 和 fs00-01 中，CSML、SILD、ITML 和 L2 四种算法均出现了明显的性能下降，而本章方法性能下降的幅度相对较小，在这两组中分别取得了 88.1%和 87.3% 的 Macro-F1 值。在对照组 fb00-01 中，由于携带背包使得行人的轮廓发生较大的变化，几种方法的性能均有明显的下降，本章方法的 Macro-F1 值为 70.6%，相较于其他四种方法具有明显的优势。综合来看，本章方法在 CASIA-C 数据集上取得了最佳的性能

表现，且在应对行走状态变化的场景下比传统方法具有更好的鲁棒性。

### 3.4 本章小结

本章对步态识别中的特征表示进行了优化改进，并在此基础上提出了一种基于三元组损失的步态识别方法。首先介绍了步态能量图和活性能量图两种步态特征表示，在此基础上根据归一化的方式不同从数据集中提取四种步态特征作为算法的输入。然后基于三元组损失设计了一种算法，该算法通过在数据中寻找锚点样本以及正负样本，使得特征空间中锚点样本和正样本之间相似度尽可能高，锚点样本和负样本之间相似度尽可能低来提高步态识别的准确率。介绍了所使用的目标函数并对目标函数的优化进行了详细的推导。最后在两个步态数据集上进行对比实验，筛选了性能较好的步态特征并通过与几种度量学习算法的对比验证了本章提出的算法的有效性。



## 第四章 基于 L1 距离学习的步态识别方法

第三章提出的基于三元组的步态识别方法，相比于传统度量学习方法，提高了步态识别的精度。但是在角度和行人行走状态发生变化时，轮廓图像之间的差异会进一步放大，进而导致识别的精度下降，针对这一问题，本章在第三章方法的基础上提出一种基于 L1 距离度量的步态识别方法，进一步提高步态识别在跨视角及不同行走状态下的识别精度和稳定性，并通过对比实验验证所提出方法的有效性。

### 4.1 基于 L1 距离学习的步态识别方法

回顾 2.1 节可知，两个样本  $\mathbf{x}_i$ ,  $\mathbf{x}_j$  间的 L1 距离可以定义为：

$$L_1(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\|_1 \quad (4-1)$$

在二维空间中，L1 距离等价于两个点在横纵坐标轴上的差值的绝对值之和，而 L2 距离则表示两个点之间的直线距离，如图 4-1 所示。

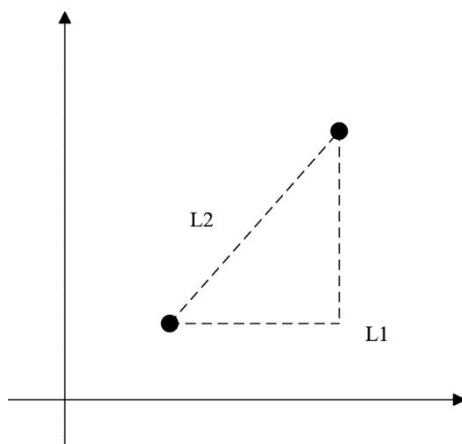


图 4-1 二维空间中的 L1 距离和 L2 距离

Fig.4-1 L1 distance and L2 distance in two-dimensional space

在分类问题中，L2 距离是一种广泛应用的距离度量，然而，大量的研究<sup>[64-68]</sup>已经指出，L2 距离使用的平方运算在很多情况下放大了数据中某些异常值的影响，进而影响分类的结果，使得模型的鲁棒性下降，而 L1 距离是一种对异常值具有良好鲁棒性的距离度量，因此近年来一些研究者将 L1 距离度量引入机器学习中，以提高算法模型的泛化能力及鲁棒性。Yan 等人<sup>[69-70]</sup>将 L1 距离度量引入到双支持向量机（Twin Support Vector Machine, TWSVM）中，通过 L1 距离来最大化类间散度矩阵和类内散度矩阵的比率，显著降低数据中的异常值对模型的影响。Ye 等人<sup>[64]</sup>通过引入一个高

效的迭代框架来解决 L1 范数的最小化最大化问题，并将 L1 范数与线性判别分析（Linear Discriminant Analysis, LDA）相结合，大大提升了传统算法的收敛速度以及对异常值的鲁棒性。

角度和行人着装的变化一直是步态识别任务中面临的重要挑战，这些变化会导致一部分样本数据和同类别的样本之间产生极大的差异，如图 4-2 所示。针对这个问题，本章将 L1 距离度量引入步态识别中，提出一种基于 L1 距离的度量学习方法(L1-Norm Distance Metric Learning, L1ML)。



图 4-2 CASIA-B 数据集中部分正样本对

Fig.4-2 Some positive sample pairs in CASIA-B dataset

#### 4.1.1 目标函数

第三章中基于三元组损失提出了一种步态识别方法，通过采样一个锚点样本，一个正样本和一个负样本，并约束正样本和锚点样本之间的距离尽量小，负样本和锚点样本之间的距离尽量大，使得正负样本之间产生明确的区分。三元组损失对锚点样本和正负样本之间同时进行约束，可以兼顾类内距离的最小化和类间距离的最大化，因此在分类问题中有着广泛的应用。

然而，一些的研究<sup>[71-75]</sup>指出，由于三元组损失中采用的是随机采样的方式，在实际应用中随机采样很难收集到足够的难负样本，大部分样本对迭代过程并不能作出贡献，这不仅会导致模型收敛速度变慢，也会削弱模型对特征的提取能力。一些研究<sup>[76-80]</sup>尝试对三元组损失进行优化，但是难负样本的挖掘依然需要额外的计算。

基于此，本章提出一种基于大边际框架的损失函数，其基本思想是，使样本数据中正样本对的距离小于一个阈值，负样本对之间的距离大于另一个阈值，进一步扩大类内距离和类间距离之间的差距，实现更高的分类精度。

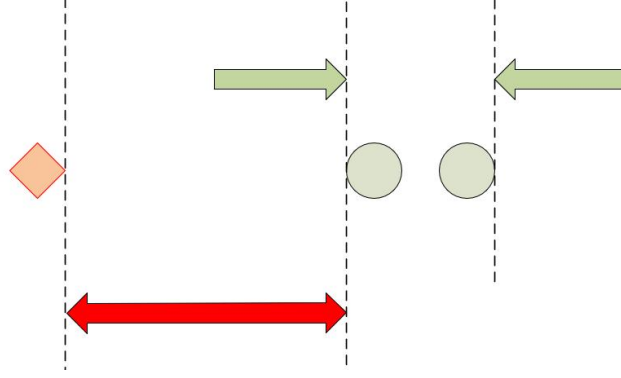


图 4-3 基于大边际框架的损失函数

Fig.4-3 Large margin-based loss function

设一个数据集中包含  $N$  个样本，将其中任意一个样本  $\mathbf{x}_i \in \mathbb{R}^d$  通过线性变换  $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k] \in \mathbb{R}^{d \times s}$ ,  $s \leq d$  投影到特征空间中，其中  $d$  和  $s$  分别是投影前后样本的维度，则在投影的特征空间中，一对样本  $\mathbf{x}_i$  和  $\mathbf{x}_j$  之间的 L1 距离可以描述为：

$$\begin{aligned} d_{\mathbf{w}}(\mathbf{x}_i, \mathbf{x}_j) &= \|\mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j\|_1 \\ &= \sum_{k=1}^s |\mathbf{w}_k^T \mathbf{x}_i - \mathbf{w}_k^T \mathbf{x}_j| \\ &= \sum_{k=1}^s d_{\mathbf{w}_k}(\mathbf{x}_i, \mathbf{x}_j), \end{aligned} \quad (4-2)$$

其中， $\mathbf{w}_k \in \mathbb{R}^{d \times 1}$  是第  $k$  个投影向量。

为了学习到最优的投影矩阵  $\mathbf{W}$ ，本章提出一种基于大边际框架的度量学习方法 L1ML，方法的目标函数可以描述为以下形式：

$$\min_{\mathbf{W}} J = \sum_{l_{ij}=1} h(d_{\mathbf{w}}(\mathbf{x}_i, \mathbf{x}_j) - \tau_p) + \sum_{l_{ij}=-1} h(\tau_n - d_{\mathbf{w}}(\mathbf{x}_i, \mathbf{x}_j)) \quad (4-3)$$

其中， $\tau_p$  和  $\tau_n$  分别是正样本对和负样本对之间的距离阈值，且满足  $0 < \tau_p < \tau_n$ ， $l_{ij} = 1$  表示一对样本为正样本对， $l_{ij} = -1$  表示一对样本为负样本对。

结合图 4-3 和式 (4-3) 可知，损失函数借助铰链损失惩罚距离过远的正样本对和距离过近的负样本对，使得特征空间中样本数据呈现合理的分布，提高算法对步态图像的判别力。

#### 4.1.2 算法实现

L1ML 中，使用梯度下降法求解式 (4-3) 中的目标函数，梯度的计算方式如式 (4-4) 所示：

$$\begin{aligned} \frac{\partial J}{\partial \mathbf{W}} = & \sum_{l_{ij}=1} h'(d_{\mathbf{W}}(\mathbf{x}_i, \mathbf{x}_j) - \tau_p) \frac{\partial d_{\mathbf{W}}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \mathbf{W}} \\ & - \sum_{l_{ij}=-1} h'(\tau_n - d_{\mathbf{W}}(\mathbf{x}_i, \mathbf{x}_j)) \frac{\partial d_{\mathbf{W}}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \mathbf{W}} \end{aligned} \quad (4-4)$$

为求解上式的梯度，需要求得样本间的 L1 距离  $d_{\mathbf{W}}(\mathbf{x}_i, \mathbf{x}_j)$  对投影矩阵  $\mathbf{W}$  的偏导，因此有：

$$\frac{\partial d_{\mathbf{W}_k}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \mathbf{W}} = \text{sgn}(\mathbf{w}_k^T \mathbf{x}_i - \mathbf{w}_k^T \mathbf{x}_j) (\mathbf{x}_i - \mathbf{x}_j), \quad (4-5)$$

其中， $\text{sgn}(x)$  为阶跃函数，有：

$$\text{sgn}(\mathbf{w}_k^T \mathbf{x}_i - \mathbf{w}_k^T \mathbf{x}_j) = \begin{cases} 1, \mathbf{w}_k^T \mathbf{x}_i > \mathbf{w}_k^T \mathbf{x}_j \\ 0, \mathbf{w}_k^T \mathbf{x}_i = \mathbf{w}_k^T \mathbf{x}_j \\ -1, \mathbf{w}_k^T \mathbf{x}_i < \mathbf{w}_k^T \mathbf{x}_j \end{cases} \quad (4-6)$$

整理后可得：

$$\frac{\partial d_{\mathbf{W}}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \mathbf{W}} = (\mathbf{x}_i - \mathbf{x}_j) \mathbf{q}^T, \quad (4-7)$$

其中，

$$\mathbf{q} = (q_1, q_2, \dots, q_s)^T, \mathbf{q}_k = \text{sgn}(\mathbf{w}_k^T \mathbf{x}_i - \mathbf{w}_k^T \mathbf{x}_j). \quad (4-8)$$

最后，使用梯度下降法迭代更新投影矩阵  $\mathbf{W}$ ，并最终求得最优解。 $\mathbf{W}$  的迭代更新方式如式 (4-9) 所示。

$$\mathbf{W} = \mathbf{W} - \lambda \frac{\partial J}{\partial \mathbf{W}} \quad (4-9)$$

其中， $\lambda$  表示学习率。

## 4.2 实验结果及分析

### 4.2.1 实验设置

本章在 CAISIA-B 和 CAISA-C 两个步态数据集中进行实验，数据集的划分如表 4-1、表 4-2 所示。实验的硬件环境与前章中的介绍相同。同时，参与对比实验的其他度量学习方法还有 KISSME<sup>[81]</sup>、GMML<sup>[82]</sup>、CSML、SILD。实验中，将  $\lambda$  设置为  $1 \times 10^{-6}$ ， $\tau_p$  设置为 1， $\tau_n$  设置为 5。

表 4-1 CASIA-B 数据集的划分  
Table 4-1 Division of CAISA-B dataset

Train	Gallery	Probe1	Probe2	Probe3
001-062				
0°-180°	063-124	063-124	063-124	063-124
NM#1-6、	0°-180°	0°-180°	0°-180°	0°-180°
BG#1-2、	NM#1-4	NM#5-6	BG#1-2	CL#1-2
CL#1-2				

表 4-2 CAISA-C 数据集的划分  
Table 4-2 Division of CAISA-C dataset

Train	Gallery	Probe1	Probe2	Probe3	Probe4
001-076					
fn00-04、	077-154	077-154	077-154	077-154	077-154
fq00-01、	fn00-01	fn02-03	fq00-01	fs00-01	fb00-01
fs00-01、					
fb00-01					

#### 4.2.2 不同特征的对比实验

为进一步验证不同特征图像在轮廓变化时的性能差异,本节使用 L1ML 方法对四种特征在 CAISA-B 数据集上进行了对比实验,结果如表 4-3 所示。

表 4-3 四种特征在 CASIA-B 数据集上的平均准确率(%)  
Table 4-3 Average accuracy(%) of the four features on CASIA-B dataset

	NM#5-6	BG#1-2	CL#1-2	Mean
centroid_GEI	98.0	78.1	31.5	69.2
head_GEI	99.3	90.0	34.5	74.6
centroid_AEI	98.8	77.0	31.0	68.9
head_AEI	98.5	90.8	36.8	75.4

从上表可以看出,在正常行走的 NM#5-6 组中,四种特征之间的差异不大,在行走状态发生变化的 BG#1-2 和 CL#1-2 对照组中,基于头顶对齐的特征相对于基于质心对齐的特征具有明显的优势。



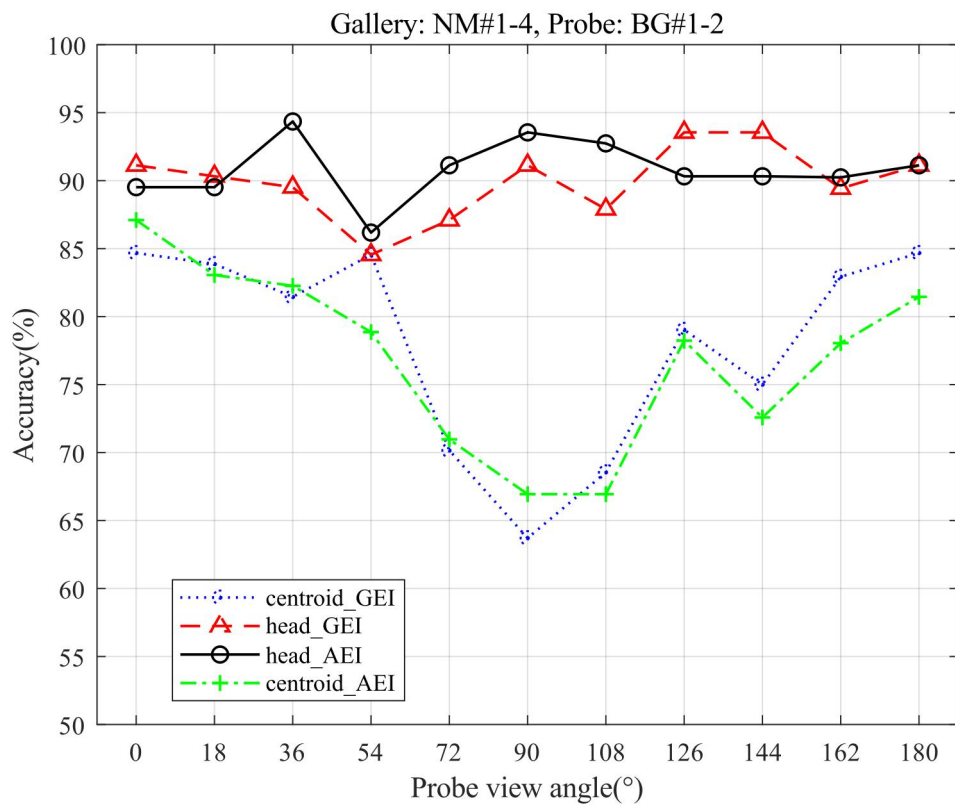


图 4-4 四种特征在 BG#1-2 中各角度下的准确率

Fig.4-4 Accuracy of the four features at various angles in BG#1-2

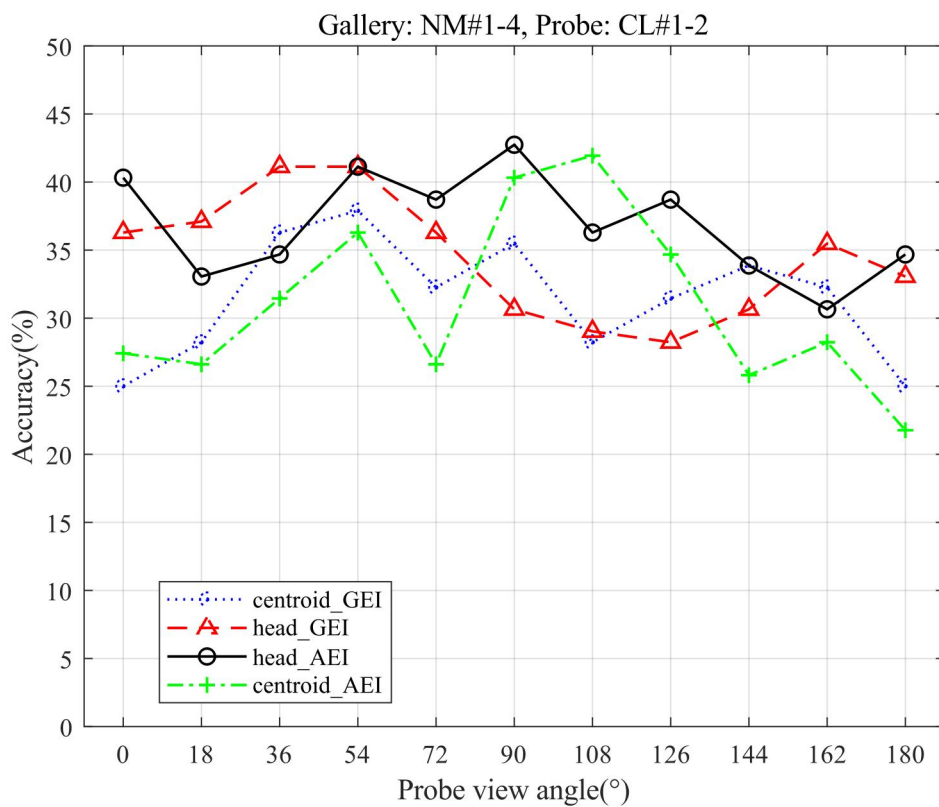


图 4-5 四种特征在 CL#1-2 中各角度下的准确率

**Fig.4-5** Accuracy of the four features at various angles in CL#1-2

图 4-4 和图 4-5 分别显示了四种特征在 BG#1-2 和 CL#1-2 各角度下的性能表现。从图 4-4 可以看出, 基于头顶对齐的特征在角度发生变化时准确率的波动不大, 而基于质心对齐的特征在 90°附近准确率出现了明显的下降。图 4-5 中, 基于头顶对齐的特征整体的准确率较高, 在角度发生变化时准确率的波动也相对较小。上述实验结果进一步验证了 head\_AEI 特征在行走状态及视角变化时的鲁棒性。

### 4.2.3 与其他方法的对比实验

本节首先对 L1ML 和 SILD<sup>[62]</sup>、GMML<sup>[82]</sup>、PoseGait<sup>[83]</sup>、MGAN<sup>[84]</sup>、DV-GEIs<sup>[85]</sup> 几种算法在 CAISIA-B 数据集上各角度下的平均准确率进行了对比, 实验结果如表 4-4 所示。

**表 4-4** L1ML 与其他方法在 CASIA-B 数据集上的平均准确率(%)**Table 4-4** Average accuracy(%) of L1ML and other methods on CASIA-B dataset

方法	NM#5-6	BG#1-2	CL#1-2
PoseGait	68.7	44.5	35.9
GMML	97.4	78.6	21.5
SILD	97.9	74.9	23.8
MGAN	68.1	54.7	31.5
DV-GEIs	76.4	59.0	39.6
L1ML	98.5	90.8	36.8

从上表可以看出, L1ML 方法在 NM#5-6、BG#1-2、CL#1-2 三组实验中分别取得了 98.5%、90.8%和 36.8%的平均准确率, 对比其他几种算法具有明显的优势。在 BG#1-2 组的实验中, 由于行人背包对轮廓产生的影响, 其他几种方法的平均准确率都出现了明显的下降, 而 L1ML 相对于 NM#5-6 仅出现了小幅度的下降, 可见 L1ML 在背包的情形下依然保持良好的性能。在 CL#1-2 中, 由于穿外套对行人轮廓的影响进一步扩大, 几种算法的平均准确率都出现了明显的下降, L1ML 在该组中的平均准确率仅次于 DV-GEIs 的 39.6%。

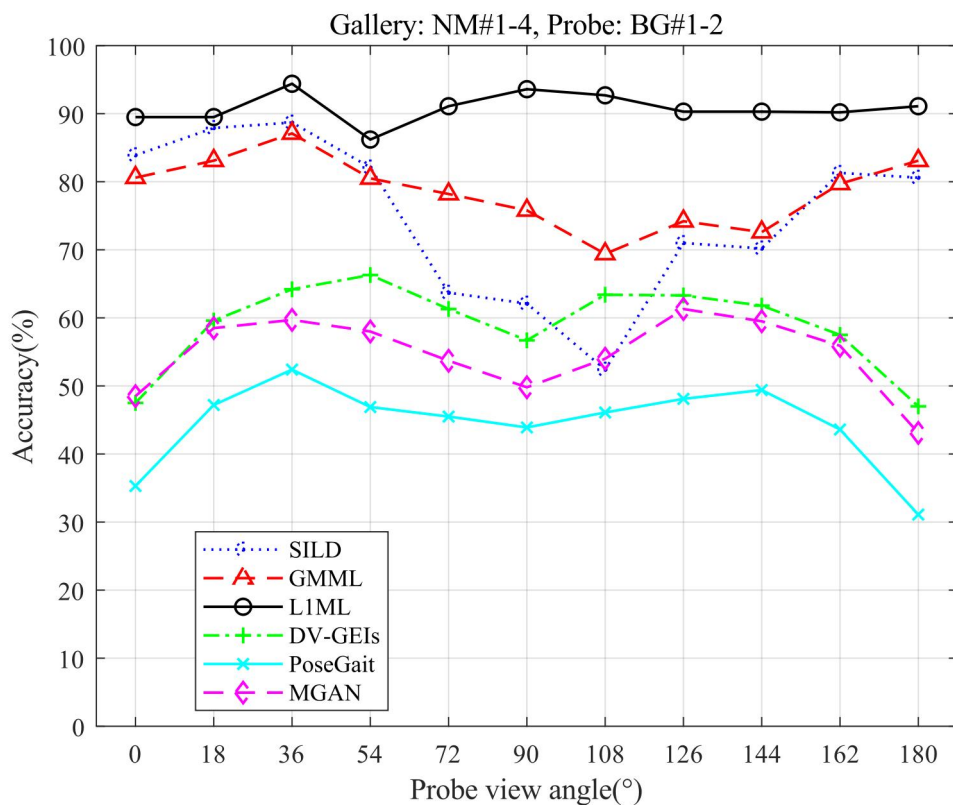


图 4-6 L1ML 和几种方法在 BG#1-2 中各角度下的准确率

Fig.4-6 The accuracy of L1ML and several methods at various angles in BG#1-2

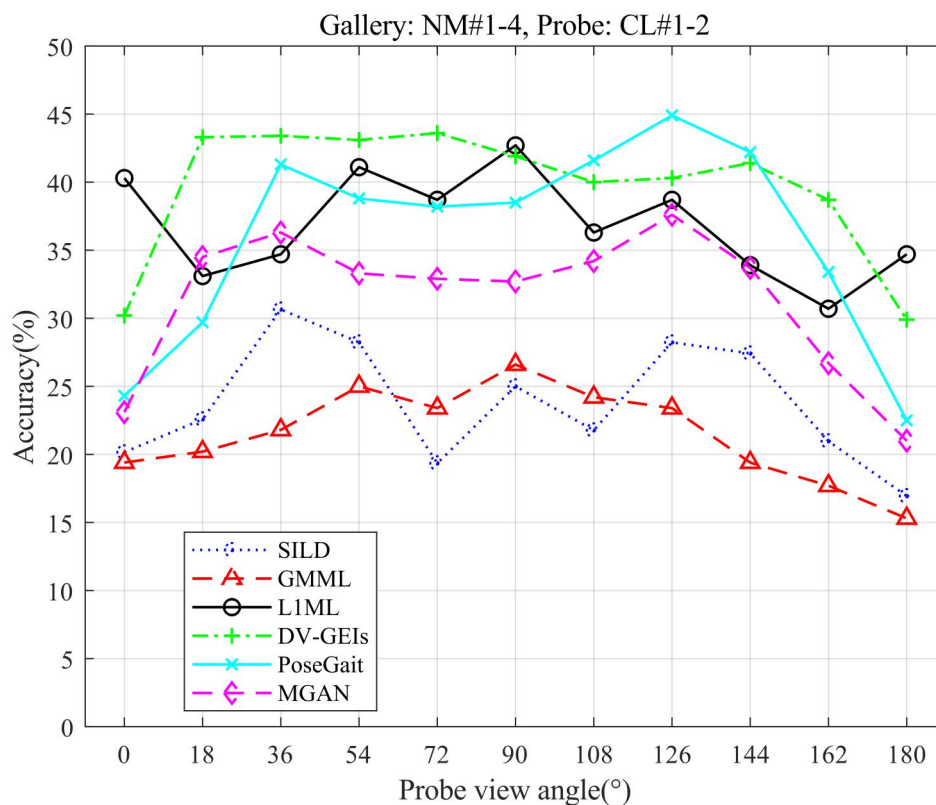


图 4-7 L1ML 和几种方法在 CL#1-2 中各角度下的准确率

**Fig.4-7** The accuracy of L1ML and several methods at various angles in CL#1-2

L1ML 和几种方法在各角度下的识别准确率如图 4-6 和图 4-7 所示。从图 4-6 中可以看出, L1ML 方法的准确率随角度变化的波动不大, 除了在 54°视角下的准确率略低于 90%以外, 在其他角度下的准确率都保持在 90%以上。其他几种方法的准确率在 90°或 108°附近均有明显的下凹, 这是由于在行人背包的情况下, 侧面视角的轮廓差异远远大于正面和后面的差异, 因此在侧面视角下准确率偏低。从图 4-7 可以看出, 所有方法的准确率都出现了较大的波动, 这是由于行人穿着较厚的外套时, 各个视角下的轮廓之间都会存在较大的差异, 此外, 一些比较长的外套还会对行人的腿部产生一定的遮挡, 因此在该组实验中几种算法的准确率表现都不稳定, L1ML 方法的准确率表现略低于 DV-GEIs, 但是相较于其他几种方法 L1ML 仍然具有明显的优势。

表 4-5、4-6、4-7 和 4-8 显示了 L1ML 与几种方法在 CASIA-C 数据集上的表现。其中, 表 4-4、4-5 和 4-6 分别显示了几种方法在行人正常行走、快步行走和慢步行走时的结果, 从表中可以看出, 前两组实验中 L1ML 方法的 Macro-F1 值分别为 99.3% 和 95.2%, 均为几种方法中最高, 第三组中 L1ML 方法的 Macro-F1 值为 92.7%, 略低于 GMMML 方法的 93.0%。表 4-7 显示了几种方法在行人背包行走时的结果, 在该组中, 由于轮廓变化较大, 其他几种方法均出现了明显的性能下降, 而 L1ML 方法的 Macro-F1 值为 88.8%, 仅仅略低于前几组的结果, 比第二的 KISSME 方法高出了 12.4%, 说明 L1ML 方法在轮廓发生明显变化的情况下仍然具有较好的性能。

**表 4-5** L1ML 和几种方法在 CASIA-C fn02-03 上的实验结果**Table 4-5** Experimental results of L1ML and several methods on CASIA-C fn02-03

方法	CCR(%)	Macro-P(%)	Macro-F1(%)
CSML	92.2	93.5	91.4
GMMML	98.7	99.1	98.6
ITML	95.5	97.2	95.2
KISSME	98.7	99.1	98.6
L1ML	99.4	99.6	99.3

**表 4-6** L1ML 和几种方法在 CASIA-C fq00-01 上的实验结果**Table 4-6** Experimental results of L1ML and several methods on CASIA-C fq00-01

方法	CCR(%)	Macro-P(%)	Macro-F1(%)
CSML	71.4	71.6	68.5
GMMML	90.3	92.2	89.5
ITML	81.2	80.5	78.4
KISSME	94.2	97.0	94.2

L1ML	95.5	97.0	95.2
------	------	------	------

表 4-7 L1ML 和几种方法在 CASIA-C fs00-01 上的实验结果

Table 4-7 Experimental results of L1ML and several methods on CASIA-C fs00-01

方法	CCR(%)	Macro-P(%)	Macro-F1(%)
CSML	74.7	71.0	70.5
GMML	93.5	94.9	93.0
ITML	80.5	78.8	77.4
KISSME	92.9	91.1	91.3
L1ML	93.5	93.6	92.7

表 4-8 L1ML 和几种方法在 CASIA-C fb00-01 上的实验结果

Table 4-8 Experimental results of L1ML and several methods on CASIA-C fb00-01

方法	CCR(%)	Macro-P(%)	Macro-F1(%)
CSML	53.3	48.8	47.7
GMML	72.1	71.0	68.7
ITML	55.2	49.9	49.3
KISSME	78.6	80.5	76.4
L1ML	89.6	91.4	88.8

#### 4.2.4 L1 距离度量有效性实验

为了进一步验证 L1 距离度量在应对步态识别中视角和行走状态变化时的有效性,本节实验分别对比了 L1 距离、欧氏距离以及余弦距离度量在相同的损失函数下的性能表现,实验结果如表 4-9 和表 4-10 所示。

表 4-9 三种距离度量在 CASIA-B 数据集上的平均准确率(%)

Table 4-9 Average accuracy(%) of the three distance metrics on CASIA-B dataset

距离度量	NM#5-6	BG#1-2	CL#1-2
L1	98.5	90.8	36.8
L2	97.0	76.4	22.3
Cosine	96.9	80.9	22.1

表 4-10 三种距离度量在 CASIA-C 数据集上的准确率(%)

Table 4-10 Accuracy(%) of the three distance metrics on CASIA-C dataset

距离度量	fn02-03	fq00-01	fs00-01	fb00-01
L1	98.7	94.8	93.5	90.9
L2	96.1	83.1	83.1	65.0
Cosine	94.8	81.8	79.2	64.3

由上述表格可以看出, L1 距离度量在两个数据集上都取得了最高的识别准确率。在行人轮廓图像发生变化的对照组 BG#1-2、CL#1-2 以及 fb00-01 中, L1 距离度量相对于欧氏距离和余弦距离, 均具有非常明显的优势, 这进一步验证了 L1 距离度量在应对视角和行走状态变化时的有效性。

### 4.3 本章小结

针对步态识别中因为视角和行走状态变化导致轮廓图像之间差异过大的问题, 本章提出一种基于 L1 距离学习的步态识别方法 L1ML。为了应对数据中的异常值, L1ML 方法采用 L1 距离作为样本间的距离度量。L1ML 在步态轮廓图像数据中学习一个线性变换, 将原始数据变换到特定的特征空间中, 通过大边际框架设计的损失函数约束正样本对在特征空间的 L1 距离小于一个给定的阈值, 负样本对在特征空间的 L1 距离大于另一个给定的阈值, 使得数据在特征空间呈现合理的分布, 进而提高模型对步态图像的判别力。在两个公开的步态数据集上的实验结果验证了 L1ML 方法应对轮廓变化时的有效性。



## 第五章 总结与展望

### 5.1 本文总结

随着现代社会信息化的不断发展,人们的生活也不断在向电子化、智能化迈进,身份识别已经成为现代社会中不可或缺的一项技术。尽管已有的生物识别技术已经发展得较为成熟,其中最具代表性的就是指纹识别和人脸识别,但是这些已经广泛应用的识别技术依然存在一定的局限性,以人脸识别为例,识别过程往往需要在近距离内完成,并且为了采集到清晰的人脸图像,待识别者需要摘下口罩或其他遮挡物,这在新冠疫情肆虐的当下无疑给人们带来了极大的不便。在一些安防监控类的应用中,也很难要求待识别者做出相应的配合。步态识别是借助行走姿态这一生物特征完成身份的识别,它很好地填补了传统识别方法的缺陷,即可以在远距离无接触的条件下完成识别。基于这样的优势,步态识别已经吸引了很多研究者的注意,但是目前这项技术距离市场化的应用仍然有研究空间,本文的工作在已有研究的基础上开展,主要的内容总结如下:

基于轮廓图像的步态识别中,特征表示的质量直接影响到算法的性能。本文在已有的特征表示的基础上,做出一定的改进,将已有的步态能量图特征和活性能量图特征重新进行归一化,将行走过程中的步态信息更多地集中到肢体部分,使得特征图像尽可能保留更多的步态特征。并在此基础上,提出一种基于三元组损失的度量学习方法,通过固定锚点样本并分别约束正负样本和锚点样本之间的距离提高步态图像的识别精度。在公开的步态数据集上进行的实验表明,优化的步态特征图像在应对视角及行走状态变化时具有较好的鲁棒性。

针对步态识别任务中,部分同类的样本轮廓图像会由于视角及行走状态发生变化而产生较大差异的问题,本文提出一种基于 L1 距离的度量学习方法,使用 L1 距离度量样本间的相似性来增强算法对异常值的鲁棒性。通过投影矩阵将原始样本投影到特征空间中,借助目标函数惩罚距离过远的正样本对和距离过近的负样本对,使得样本在特征空间中呈现合理的分布,进而提高对差异较大的样本的辨别力。实验结果表明,本文所提出的方法在步态数据集上具有较高的识别精度以及较好的鲁棒性。

### 5.2 未来工作的展望

本文在已有的步态识别研究的基础上,对步态特征表示进行了优化并提出了两种



度量学习方法，虽然实验验证了本文工作对步态识别任务性能的提升，但是仍有一定的不足之处和提升空间，具体包含以下几方面：

（1）在行人穿着较厚外套的情况下，轮廓图像之间的差异进一步扩大，此时所提出的方法的识别精度依然不够理想，未来的工作可以针对这一点做进一步的优化，提高算法模型对一些复杂状态下轮廓的识别精度。

（2）基于轮廓图像的步态识别能够在一些步态数据集上取得较好的性能，但是在实际应用中，算法的实时性不强。后续的研究可以聚焦于实时的步态特征提取，实现更高效更快速的步态识别。

（3）借助单一的生物特征进行识别往往具有一些局限性，如何将步态特征和其他生物特征融合，构建适用性更广的身份识别是一个值得研究的方向。

## 参 考 文 献

- [1] Song S M, Waldron K J. An analytical approach for gait study and its applications on wave gaits[J]. The International Journal of Robotics Research, 1987, 6(2): 60-71.
- [2] Huang B, Chen M, Lee K K, et al. Human identification based on gait modeling[J]. International Journal of Information Acquisition, 2007, 4(01): 27-38.
- [3] Kusakunniran W. Review of gait recognition approaches and their challenges on view changes[J]. IET Biometrics, 2020, 9(6): 238-250.
- [4] Wang L, Ning H, Tan T, et al. Fusion of static and dynamic body biometrics for gait recognition[J]. IEEE Transactions on circuits and systems for video technology, 2004, 14(2): 149-158.
- [5] Lima V C, Melo V H C, Schwartz W R. Simple and efficient pose-based gait recognition method for challenging environments[J]. Pattern Analysis and Applications, 2021, 24(2): 497-507.
- [6] Tang J, Luo J, Tjahjadi T, et al. Robust arbitrary-view gait recognition based on 3D partial similarity matching[J]. IEEE Transactions on Image Processing, 2016, 26(1): 7-22.
- [7] Choi S, Kim J, Kim W, et al. Skeleton-based gait recognition via robust frame-level matching[J]. IEEE Transactions on Information Forensics and Security, 2019, 14(10): 2577-2592.
- [8] Luo J, Tjahjadi T. View and Clothing Invariant Gait Recognition via 3D Human Semantic Folding[J]. IEEE Access, 2020, 8: 100365-100383.
- [9] BenAbdelkader C, Cutler R, Davis L. Motion-based recognition of people in eigengait space[C]//Proceedings of Fifth IEEE international conference on automatic face gesture recognition. IEEE, 2002: 267-272.
- [10] Chai Y, Wang Q, Zhao R, et al. A new automatic gait recognition method based on the perceptual curve[C]//TENCON 2005-2005 IEEE Region 10 Conference. IEEE, 2005: 1-5.
- [11] Han J, Bhanu B. Individual Recognition Using Gait Energy Image[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(2): 316-322.
- [12] Zhang E, Zhao Y, Xiong W. Active energy image plus 2DLPP for gait recognition[J]. Signal Processing, 2010, 90(7): 2295-2302.
- [13] Shiraga K, Makihara Y, Muramatsu D, et al. GEINet: View-invariant gait recognition using a convolutional neural network[C]//2016 international conference on biometrics (ICB). IEEE, 2016: 1-8.
- [14] Gul S, Malik M I, Khan G M, et al. Multi-view gait recognition system using spatio-temporal features and deep learning[J]. Expert Systems with Applications, 2021, 179: 115057.

- [15] Zhang Y, Huang Y, Yu S, et al. Cross-view gait recognition by discriminative feature learning[J]. IEEE Transactions on Image Processing, 2019, 29: 1001-1015.
- [16] Chao H, He Y, Zhang J, et al. GaitSet: Regarding gait as a set for cross-view gait recognition [C]//Proceedings of the AAAI conference on artificial intelligence. IEEE, 2019, 33(01): 8126-8133.
- [17] Fan C, Peng Y, Cao C, et al. GaitPart: Temporal part-based model for gait recognition[C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. IEEE, 2020: 14225-14233.
- [18] Wang Y, Song C, Huang Y, et al. Learning view invariant gait features with Two-Stream GAN[J]. Neurocomputing, 2019, 339: 245-254.
- [19] Yu S, Chen H, Garcia Reyes E B, et al. GaitGan: Invariant gait feature extraction using generative adversarial networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. IEEE, 2017: 30-37.
- [20] Liao R, Cao C, Garcia E B, et al. Pose-based temporal-spatial network (PTSN) for gait recognition with carrying and clothing variations[C]//Chinese conference on biometric recognition. Springer, 2017: 474-483.
- [21] 刘冰, 李瑞麟, 封举富. 深度度量学习综述[J]. 智能系统学报, 2019, 14(06): 1064-1072.
- [22] Wang F, Sun J. Survey on distance metric learning and dimensionality reduction in data mining[J]. Data mining and knowledge discovery, 2015, 29(2): 534-564.
- [23] Yang L, Jin R. Distance metric learning: A comprehensive survey[J]. Michigan State University, 2006, 2(2): 4.
- [24] Weinberger K Q, Saul L K. Distance Metric Learning for Large Margin Nearest Neighbor Classification[J]. Journal of Machine Learning Research, 2009, 10: 207-244.
- [25] Lu J, Hu J, Zhou J. Deep metric learning for visual understanding: An overview of recent advances[J]. IEEE Signal Processing Magazine, 2017, 34(6): 76-84.
- [26] Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation[J]. Neural computation, 2003, 15(6): 1373-1396.
- [27] Carroll J D, Arabie P. Multidimensional scaling[J]. Measurement, judgment and decision making, 1998: 179-250.
- [28] Xing E P, Ng A Y, Jordan M I, et al. Distance metric learning with application to clustering with side-information[C]//Proceedings of the 15th International Conference on Neural Information Processing Systems. IEEE, 2002: 521-528.
- [29] Kocsor A, Kovács K, Szepesvári C. Margin maximizing discriminant analysis[C]//European Conference on Machine Learning. Springer, 2004: 227-238.

- [30] Shental N, Hertz T, Weinshall D, et al. Adjustment learning and relevant component analysis[C]//European conference on computer vision. Springer, 2002: 776-790.
- [31] Schölkopf B. Statistical learning and kernel methods[M]//Data Fusion and Perception. Springer, Vienna, 2001: 3-24.
- [32] Cristianini N, Shawe-Taylor J, Elisseeff A, et al. On kernel-target alignment[C]//Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic. IEEE, 2001: 367-373.
- [33] Campbell C. An introduction to kernel methods[J]. Studies in Fuzziness and Soft Computing, 2001, 66: 155-192.
- [34] Hadsell R, Chopra S, LeCun Y. Dimensionality reduction by learning an invariant mapping[C]//2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). IEEE, 2006, 2: 1735-1742.
- [35] Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, 2015: 815-823.
- [36] Ge W. Deep metric learning with hierarchical triplet loss[C]//Proceedings of the European Conference on Computer Vision (ECCV). IEEE, 2018: 269-285.
- [37] Sohn K. Improved deep metric learning with multi-class N-pair loss objective[C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. IEEE, 2016: 1857-1865.
- [38] Wen Y, Zhang K, Li Z, et al. A discriminative feature learning approach for deep face recognition [C]//European conference on computer vision. Springer, 2016: 499-515.
- [39] Movshovitz-Attias Y, Toshev A, Leung T K, et al. No fuss distance metric learning using proxies[C]//Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2017: 360-368.
- [40] Wang J, Zhou F, Wen S, et al. Deep metric learning with angular loss[C]//Proceedings of the IEEE international conference on computer vision. IEEE, 2017: 2593-2601.
- [41] Yi D, Lei Z, Liao S, et al. Deep metric learning for person re-identification[C]//2014 22nd international conference on pattern recognition. IEEE, 2014: 34-39.
- [42] Mehralian S, Teshnehlal M, Nasersharif B. Unrestricted deep metric learning using neural networks interaction[J]. Pattern Analysis and Applications, 2021, 24(4): 1699-1711.
- [43] Feng Y, Wu F, Ji Y, et al. Deep Metric Learning with Triplet-Margin-Center Loss for Sketch Face Recognition[J]. IEICE Transactions on Information and Systems, 2020, 103(11): 2394-2397.
- [44] Yu J, Hu C H, Jing X Y, et al. Deep metric learning with dynamic margin hard sampling loss for

- face verification[J]. Signal, Image and Video Processing, 2020, 14(4): 791-798.
- [45] Zhou X, Jin K, Xu M, et al. Learning deep compact similarity metric for kinship verification from face images[J]. Information Fusion, 2019, 48: 84-94.
- [46] Feng Y, Yuan Y, Lu X. Person reidentification via unsupervised cross-view metric learning[J]. IEEE Transactions on Cybernetics, 2019, 51(4): 1849-1859.
- [47] Han P, Li Q, Ma C, et al. HMMN: Online metric learning for human re-identification via hard sample mining memory network[J]. Engineering Applications of Artificial Intelligence, 2021, 106: 104489.
- [48] Chen X, Xu H, Li Y, et al. Person Re-Identification by Low-Dimensional Features and Metric Learning[J]. Future Internet, 2021, 13(11): 289.
- [49] Zhao X, Xu Z, Zhao B, et al. Object tracking with structured metric learning[J]. IEEE Access, 2019, 7: 161764-161775.
- [50] Yuan D, Kang W, He Z. Robust visual tracking with correlation filters and metric learning[J]. Knowledge-Based Systems, 2020, 195: 105697.
- [51] Nguyen H V, Bai L. Cosine similarity metric learning for face verification[C]//Asian conference on computer vision. Springer, 2010: 709-720.
- [52] Chen J, Guo Z, Hu J. Ring-regularized cosine similarity learning for fine-grained face verification[J]. Pattern Recognition Letters, 2021, 148: 68-74.
- [53] Li B, Han L. Distance weighted cosine similarity measure for text classification[C]// International conference on intelligent data engineering and automated learning. Springer, 2013: 611-618.
- [54] Oh Song H, Xiang Y, Jegelka S, et al. Deep metric learning via lifted structured feature embedding[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, 2016: 4004-4012.
- [55] Kumar S, Yadav J S. Video object extraction and its tracking using background subtraction in complex environments[J]. Perspectives in Science, 2016, 8: 317-322.
- [56] Lipton A J, Fujiyoshi H, Patil R S. Moving target classification and tracking from real-time video[C]//Proceedings fourth IEEE workshop on applications of computer vision. WACV'98 (Cat. No. 98EX201). IEEE, 1998: 8-14.
- [57] 商磊, 张宇, 李平. 基于密集光流的步态识别[J]. 大连理工大学学报, 2016, 56(02): 214-220.
- [58] Stauffer C, Grimson W E L. Adaptive background mixture models for real-time tracking[C]//Proceedings. 1999 IEEE computer society conference on computer vision and pattern recognition (Cat. No PR00149). IEEE, 1999, 2: 246-252.
- [59] Tan D, Huang K, Yu S, et al. Efficient night gait recognition based on template matching[C]//18th International Conference on Pattern Recognition (ICPR'06). IEEE, 2006, 3: 1000-1003.

- [60] Yu S, Tan D, Tan T. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition[C]//18th International Conference on Pattern Recognition (ICPR'06). IEEE, 2006, 4: 441-444.
- [61] Nguyen H V, Bai L. Cosine similarity metric learning for face verification[C]//Asian conference on computer vision. Springer, 2010: 709-720.
- [62] Bessaoudi M, Ouamane A, Belahcene M, et al. Multilinear side-information based discriminant analysis for face and kinship verification in the wild[J]. Neurocomputing, 2019, 329: 267-278.
- [63] Choi J, Min C, Lee B. Mathematical Analysis on Information-Theoretic Metric Learning With Application to Supervised Learning[J]. IEEE Access, 2019, 7: 121998-122005.
- [64] Ye Q, Yang J, Liu F, et al. L1-norm distance linear discriminant analysis based on an effective iterative algorithm[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 28(1): 114-129.
- [65] Lin G, Tang N, Wang H. Locally principal component analysis based on L1-norm maximisation[J]. IET Image Processing, 2015, 9(2): 91-96.
- [66] Nie F, Huang H, Ding C, et al. Robust principal component analysis with non-greedy l1-norm maximization[C]//Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Volume Two. IEEE, 2011: 1433-1438.
- [67] Ye Q, Zhao H, Li Z, et al. L1-Norm distance minimization-based fast robust twin support vector k-plane clustering[J]. IEEE transactions on neural networks and learning systems. IEEE, 2017, 29(9): 4494-4503.
- [68] Li C N, Shao Y H, Deng N Y. Robust L1-norm two-dimensional linear discriminant analysis[J]. Neural Networks, 2015, 65: 92-104.
- [69] Yan H, Ye Q L, Yu D J. Efficient and robust TWSVM classification via a minimum L1-norm distance metric criterion[J]. Machine Learning, 2019, 108(6): 993-1018.
- [70] Yan H, Ye Q, Zhang T, et al. Least squares twin bounded support vector machines based on L1-norm distance metric for classification[J]. Pattern recognition, 2018, 74: 434-447.
- [71] Deng J, Guo J, Xue N, et al. Arcface: Additive angular margin loss for deep face recognition [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. IEEE, 2019: 4690-4699.
- [72] Wang F, Cheng J, Liu W, et al. Additive margin softmax for face verification[J]. IEEE Signal Processing Letters, 2018, 25(7): 926-930.
- [73] Zhang X, Fang Z, Wen Y, et al. Range loss for deep face recognition with long-tailed training data[C]//Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2017: 5409-5418.

- [74] Zhang X, Fang Z, Wen Y, et al. Range loss for deep face recognition with long-tailed training data[C]//Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2017: 5409-5418.
- [75] Liu H, Zhu X, Lei Z, et al. Adaptiveface: Adaptive margin and sampling for face recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2019: 11947-11956.
- [76] Zhao D, Chen C, Li D. Multi-stage attention and center triplet loss for person re-identification[J]. Applied Intelligence, 2022, 52(3): 3077-3089.
- [77] Liu H, Tan X, Zhou X. Parameter sharing exploration and hetero-center triplet loss for visible-thermal person re-identification[J]. IEEE Transactions on Multimedia, 2020, 23: 4414-4425.
- [78] Bhattacharya J, Sharma R K. Ranking-based triplet loss function with intra-class mean and variance for fine-grained classification tasks[J]. Soft Computing, 2020, 24(20): 15519-15528.
- [79] Feng Y, Wu F, Ji Y, et al. Deep Metric Learning with Triplet-Margin-Center Loss for Sketch Face Recognition[J]. IEICE Transactions on Information and Systems, 2020, 103(11): 2394-2397.
- [80] D'Innocente A, Garg N, Zhang Y, et al. Localized Triplet Loss for Fine-Grained Fashion Image Retrieval[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2021: 3910-3915.
- [81] Koestinger M, Hirzer M, Wohlhart P, et al. Large scale metric learning from equivalence constraints[C]//2012 IEEE conference on computer vision and pattern recognition. IEEE, 2012: 2288-2295.
- [82] Zhao P, Wu T, Zhao S, et al. Robust transfer learning based on geometric mean metric learning[J]. Knowledge-Based Systems, 2021, 227: 107227.
- [83] Liao R, Yu S, An W, et al. A model-based gait recognition method with body pose and human prior knowledge[J]. Pattern Recognition, 2020, 98: 107069.
- [84] He Y, Zhang J, Shan H, et al. Multi-task GANs for view-specific feature learning in gait recognition[J]. IEEE Transactions on Information Forensics and Security, 2018, 14(1): 102-113.
- [85] Liao R, An W, Li Z, et al. A novel view synthesis approach based on view space covering for gait recognition[J]. Neurocomputing, 2021, 453: 13-25.

## 致 谢

三年前研究生报道的场景还犹在眼前，转眼已经要毕业了。回顾在北化的这三年研究生生活，我首先想到的一个词就是成长，其次就是孤独。在马上就要毕业，进入人生下一个阶段的这个时刻，我想对所有相处过的人表示感谢。

首先要感谢胡峻林老师，还记得第一次见到胡老师的场景，那个时候胡老师的年轻和活力就给我留下了深刻的印象。胡老师是一位内敛，智慧，富有亲和力的老师，在这三年里，胡老师给过我很多的帮助和指导，在科研受挫的时候也是胡老师一直在鼓励我，并且给了我很多积极的建议。

其次，感谢父母的付出和支持。无论遇到什么样的困难，家人永远都是最可靠的后盾。无论是生活还是学习，背后都有父母默默无闻的付出，感谢父母一如既往地在这个方面的支持和鼓励。

同时，感谢 614 实验室的李瑞瑞老师和同学们。在我的印象中，无论多么晚，614 的灯光永远都在科技大厦六楼亮着。614 是一个学习氛围很浓厚的实验室，每次在实验室学习的时候，周围翻书的声音，键盘敲击的声音，学术讨论的声音无不在激励着我抓紧时间。感谢身边有这样一群勤奋上进的同学，让我时时刻刻都能看到身边的榜样。

最后，感谢自己的不放弃。回想起自己无数次独自往返在宿舍和实验室的路上，找工作时一次次面试的碰壁，做实验时一次次的失败和错误，所有这些孤独、挫折和失败，都没能使我放弃。

每一个阶段都有各自的风景，感谢在北京化工大学这 7 年的时光，这段风景会永远在我的记忆中熠熠闪光。





## 研究成果及发表的学术论文

### 发表及已接受的论文

1. Liu D, Hu J. L1-Norm Distance Metric Learning for Gait Recognition[C]//*2021 13th International Conference on Wireless Communications and Signal Processing*. (EI). IEEE, 2021: 1-4.
2. 刘东, 胡峻林. 基于度量学习的步态识别比较研究[J]. *计算机技术与发展*, 已接收



## 作者和导师简介

### 作者介绍:

刘东, 男, 1996 年出生, 汉族, 江苏人; 硕士研究生, 主要研究方向为度量学习, 着重于步态识别任务的研究。



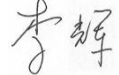
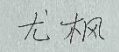

### 导师介绍:

胡峻林, 男, 1986 年出生, 汉族, 甘肃人; 副教授, 博士生导师, 主要研究方向为计算机视觉、模式识别、度量学习、人脸识别等。



# 北京化工大学

## 专业学位硕士研究生学位论文答辩委员会决议书

研究生姓名: <u>刘东</u> 学科名称: <u>计算机技术</u>				
论文题目: <u>基于度量学习的步态识别方法研究</u>				
学校导师姓名: <u>胡峻林</u>		职称: <u>副教授</u>		
企业导师姓名: <u>罗瑞一</u>		职称: <u>高级工程师</u>		
论文答辩日期: <u>2022.5.16</u>		地点: <u>腾讯会议 323-297-218</u>		
<b>论文答辩委员会成员</b>				
姓名	职称	工作单位	本人签名	是否来自企业或工程部门
祝海江	教授	北京化工大学		否
赵瑞莲	教授	北京化工大学		否
赵英	教授	北京化工大学		否
李辉	副教授	北京化工大学		否
尤枫	副教授	北京化工大学		否
江志英	高工	北京化工大学		否

注：此表用于存档，除本人签名务必用钢笔填写外，其余处必须用计算机打印。  
答辩委员会对论文的评语（选题是否来源生产实际且具有明确的生产背景和应用价值、论文工作的技术难度和工作量；是否具备了解决工程实际问题的新思想、新方法；是否创造了经济效益和社会效益；是否具备了综合运用科学理论、研究方法和技术手段解决工程实践问题的能力，论文的不足之处）：

论文以基于度量学习的步态识别方法研究为课题，选题具有一定的理论意义和实际应用价值。针对步态识别中视角及行走状态变化带来的识别精度下降问题，改进了步态特征表示方式，并在此基础上应用了 L1 距离度量的方法。

论文论述清楚，文献综述较全面，写作规范，理论分析逻辑清楚，实验结果详实。论文工作表明作者掌握了本学科领域基础理论和专门知识。该论文达到了专业硕士学位论文的水平。

答辩过程中表述清楚，回答基本问题正确。同意毕业论文通过。

对学位论文水平的总体评价	优秀	良好	一般	较差
			✓	

答辩委员会表决结果：

同意授予专业硕士学位 6 票，不同意授予专业硕士学位 0 票，弃权 0 票。根据投票结果，答辩委员会做出建议授予该同学专业硕士学位的决议。

答辩委员会主席签字：



2022 年 5 月 16 日