DOI: 10.13382/j.jemi.B2407657

融合 3D-CBAM 和跨时间尺度特征分析的步态识别方法*

许振齐 朴 燕 康继元 鞠成伟
(长春理工大学电子信息工程学院 长春 130022)

摘 要:针对传统的步态识别方法忽略了步态特征中的时间信息,提出了一种融合 3D-CBAM 和跨时间尺度特征分析的步态识别框架。研究将注意力模块集成到模型中,自适应地关注输入步态序列关键通道和空间位置,提高模型的步态识别性能。此外,增强的全局和局部特征提取器(EGLFE)中全局特征提取将时间信息和空间信息在一定程度上解耦,在 2D 卷积和 1D 卷积之间添加额外的 LeakyReLU 层,增加了网络的非线性数量,在步态特征提取过程中有助于扩大感受野,从而提升模型对特征的学习能力,实现更好的全局特征提取效果,融合局部特征,弥补局部因分块带来的特征损失。多尺度时间增强模块融合帧级特征和长短期时序特征,增强模型对遮挡的鲁棒性。在 CASIA-B 数据集和 OU-MVLP 数据集上进行训练和测试,在 CASIA-B 数据集上,平均识别准确率为 92.7%,在 NM,BG,CL 上的 rank-1 准确率分别为 98.1%,95.1%,84.9%,实验结果表明,所提方法在正常行走和复杂条件下都表现出很好的性能。

关键词: 步态识别;时空特征;注意力机制;深度学习;多尺度时序特征
中图分类号: TP18;TN911.73 **文献标识码**: A **国家标准学科分类代码**: D510.4

Gait recognition method integrating 3D-CBAM and cross-time scale feature analysis

Xu Zhenqi Piao Yan Kang Jiyuan Ju Chengwei
(School of Electronic and Information Engineering, Changchun University of Science and Technology, Changchun 130022, China)

Abstract: Addressing the limitation of traditional gait recognition methods that neglect temporal information in gait features, we propose a gait recognition framework that integrates 3D-CBAM and cross-temporal scale feature analysis. By incorporating an attention module into the model, it adaptively focuses on critical channels and spatial locations within the input gait sequences, enhancing the model's gait recognition performance. Furthermore, the enhanced global and local feature extractor (EGLFE) decouples temporal and spatial information to a certain extent during global feature extraction. By inserting additional LeakyReLU layers between 2D and 1D convolutions, the number of nonlinearities in the network is increased, which aids in expanding the receptive field during gait feature extraction. This, in turn, boosts the model's ability to learn features, achieving better global feature extraction results. Local features are also integrated to compensate for feature loss due to partitioning. A multi-scale temporal enhancement module fuses frame-level features and short-to-long-term temporal features, enhancing the model's robustness against occlusion. We conducted training and testing on the CASIA-B and OU-MVLP datasets. On the CASIA-B dataset, the average recognition accuracy reached 92.7%, with rank-1 accuracies of 98.1%, 95.1%, and 84.9% for normal (NM), bag (BG), and clothing (CL) conditions, respectively. Experimental results demonstrate that the proposed method exhibits excellent performance under both normal walking and complex conditions.

Keywords: gait recognition; spatiotemporal characteristics; attention mechanisms; deep learning; multi-scale timing features

收稿日期: 2024-07-08 Received Date: 2024-07-08
*基金项目:吉林省科技支撑项目(YDZJ202402041GXJD)资助

0　引　言

步态识别技术是全球领先的生物特征识别方法之一,依据人的体型特征和走路姿态进行身份识别,展现出卓越的稳定性和广泛的适用性。步态识别主要依赖普及度高的视频监控设备,降低了采集成本。能在远距离内有效识别个体,不需要拍摄对象的刻意合作。每个人的步态特征复杂且自然,难以伪造,增强了安全性。同时,步态识别对环境变化的适应性强,不受光线、湿度等条件限制。步态识别技术在多个领域都有广泛的应用前景,如安全监控、医疗康复、智能家居等。然而,在现实场景中,视角变化、不同的穿着条件和遮挡等变化[1-3]会导致步态轮廓发生巨大变化,这对步态识别构成了重大挑战。

现有步态识别技术分为基于模型和基于轮廓两类。基于模型的方法通过估计人体底层结构,结合动力学知识构建三维模型,并参数化人体运动特征进行步态分析。该方法对噪声因素具有鲁棒性,能够很好的描述人体各部位的运动关系和变化,能够克服复杂场景下由于遮挡、着装或者多视角的识别干扰。例如携带和穿着,但遇到分辨率不够清晰的情况特别容易失效,缺乏实用性。基于轮廓的方法直接从视频中学习目标的形状特征,低分辨率条件下也能正常工作,因此准确率更高且实现起来更加方便,但对外观(如姿态、角度、装备)变化[4]较为敏感。基于轮廓的步态识别方法专注于人体的运动[5-6],通常从行人的步态序列中估算并建立相邻帧之间的关系,通过计算视觉之间的相似性来进行步态识别。这比基于模型的方法提供了更大的灵活性,轮廓序列中包含丰富的时间信息和空间信息,充分利用步态序列中包含的时空信息有助于提高步态识别的精度。本研究提出的方法属于基于轮廓的方法,并使用轮廓序列作为输入。

近年来产生了许多基于轮廓的的步态识别方法,其中一些方法从步态轮廓中提取全局和局部特征。Chao 等[7]提出基于深度集合的步态识别方法 GaitSet,利用 2D 卷积网络直接从原始步态轮廓中提取全局步态特征,然后再采用时间轴上的压缩操作对空间特征进行深度集合。陈万志等[8]利用注意力机制和轮廓增强技术增强 GaitSet 模型特征提取能力,提高步态识别的识别精度。方法 GLN[9]提取不同阶段的轮廓级和集合级特征,然后通过横向连接自上而下地将它们合并。针对轮廓容易受到穿着条件和视点变化的影响,大多数步态识别方法证明从步态序列中提取时间特征是可行的[9-10]。Zhang 等[11]将人体步态分割成不同的局部部分,使用多个独立的二维卷积神经网络(CNN)提取局部特征。Fan 等[12]提出了基于部件的步态识别方法,GaitPart 从时空角度提高步态识别率,利用焦点卷积从单幅图像中提取细粒度的空间信息,通过微动作捕捉模块从每个部件中提取并聚合短时间的步态信息。Huang 等[13]提出一种包含全局和局部分支的双分支步态识别方法 Part3D,设计增强的三维卷积模块分别提取局部和全局的时空步态特征,在基于轮廓的步态识别方法中取得了较好的识别效果。GaitGL 利用局部时间聚合模块(local temporal aggregation,LTA)对局部时间信息进行聚合,同时提出了一种全局和局部特征提取,从步态轮廓中提取全局和局部信息[14]。全局特征主要包含更多的时空步态信息,局部特征更多的关注身体不同部位的时空信息。全局信息和局部信息对步态识别的有效性都有重要影响。因此,充分提取两种步态特征是步态识别的一个关键方面。

然而,大多数现有方法是将特征图在高度上进行水平分割的方法进行分块,出现身体各部位步态特征在水平划分区域中的集中与“块化”,由此导致的分块区域边界特征映射的减弱甚至消失,这将极大地限制步态特征的有效表示。此外,上述方法在时间特征提取上主要依赖短期时间建模,缺乏对步态周期性运动的关注,导致对遮挡的鲁棒性较差。针对现有技术的不足,本研究提出了一种基于融合 3D 卷积注意力机制(3D-convolutional block attention module,3D-CBAM)和跨时间尺度特征分析的步态识别方法,使用 3D-CBAM 来捕获步态序列中重要的通道信息和空间信息,利用增强型全局和局部特征融合模块提取全局特征和局部特征,多尺度时间增强模块将提取的帧级时序特征和多尺度时序特征进行融合,实现帧级时序特征,长短期时序建模,提升步态识别的准确率。

1　网络模型

1.1　系统框架概述

首先概述整个步态识别框架。然后,详细描述了增强卷积模块、多尺度时序特征提取模块,特征映射以及损失函数。最后介绍了训练和测试的策略。

本研究方法的网络框架如图 1 所示。包括特征提取阶段、融合帧级特征提取的多尺度时序特征聚合阶段和特征映射阶段。首先,将步态轮廓序列通过三维卷积提取序列中的浅层特征;提取的浅层特征依次通过注意力模块,局部时间聚合模块,在时间维度上使用局部时间聚合模块来汇总相邻时间点的信息来降低数据维度,同时保留空间细节,从而在时间和空间分析中实现有效权衡。其次,利用增强型全局和局部特征提取器(enhanced global and local feature extractor, EGLFE)综合全局特征和局部特征。在框架中加入最大池化,获取高层次的空间特征。然后,实现帧级时序特征和多尺度时序特征融合。

然后,利用时间池和 GeM (generalized-mean pooling) 池层实现特征映射。最后,选择三元组损失[2,15-16]和交叉熵损失来训练所提出的模型。

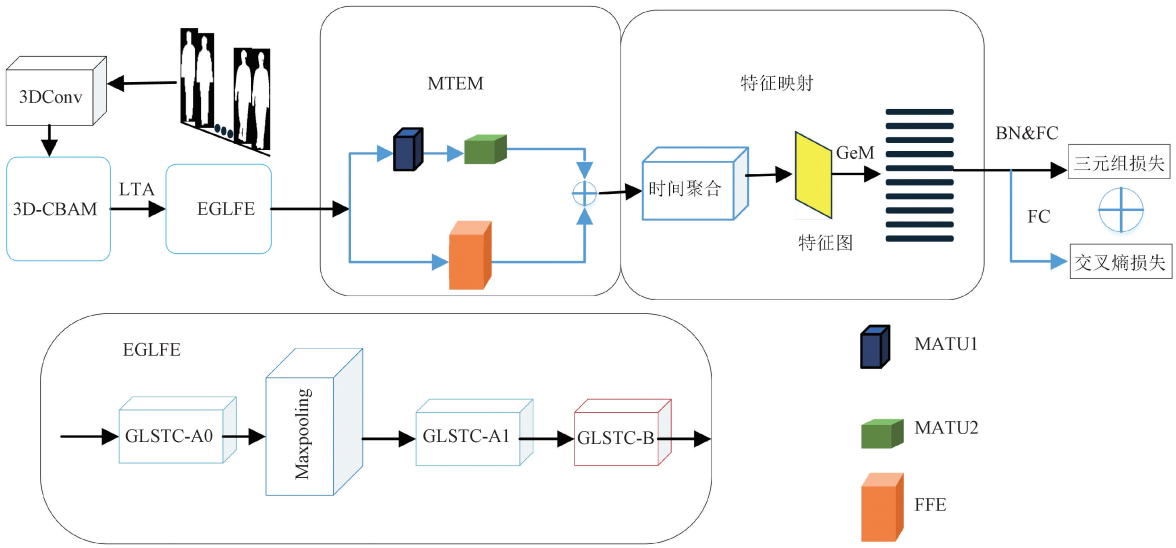

图 1 本文方法的网络框架

Fig. 1 Framework of the proposed method

1.2 基于步态轮廓序列的 3D-CBAM

CBAM 注意力机制是一种轻量级的、有效的注意模块,可以直接应用于卷积神经网络。对于卷积神经网络生成的特征图,CBAM 计算注意权值的两个维度:通道和空间,然后将注意图和特征图对应的元素相乘进行自适应特征细化使模型能够聚焦人体步态的关键信息,提升模型性能。我们提出了针对步态序列的 3D 卷积注意模块,相对 2D 注意卷积模块,同时考虑时间、深度、宽度 3 个维度上的信息,因此能够全面捕捉数据中的特征。3D-CBAM 通过将通道注意力和空间注意力级联,得到最终的注意力步态特征图 F''_{3D},通道注意结构如图 2 所示,首先对输入步态轮廓序列进行三维全局平均池化,并使用多层感知机(multilayer perceptron,MLP)由两个卷积核为 1 的 3D 卷积作为全连接层和一个线性修正激活函数来学习输入数据的特征表示,然后通过 Sigmoid 激活得到注意力权重 F_{C3D}。与二维卷积注意模块不同的是,本文对输入步态轮廓序列只进行三维全局平均池化,以更加完整的保留输入数据中的细节信息,减少信息丢失。将步态轮廓序列特征图与注意力权重进行像素级相乘,得到最终输出。该通道注意模块侧重对步态轮廓序列提取起决定性作用的特征通道。

通道注意模块如图 2 所示,将输入特征图 F_{3D} 处理成 F'_{3D},输出F'_{3D} 可由式(1)和(2)得到。

$$F_{C3D}=\sigma(MLP(AvgPool3D(F_{3D})))=\sigma(\boldsymbol{W}_1(\boldsymbol{W}_0(F^C_{avg}))) \tag{1}$$

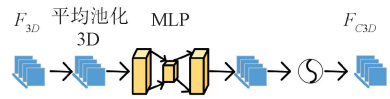

图 2 通道注意模块

Fig. 2 Channel attention module

$$F'_{3D}=F_{3D}\otimes F_{C3D} \tag{2}$$

式中:σ 表示 sigmoid 函数;$AvgPool3D(\cdot)$ 表示全局平均池化;MLP 表示多层感知器;两个输入的权重 $\boldsymbol{W}_1$ 和 $\boldsymbol{W}_0$ 是共享的;符号⊗表示元素乘法。

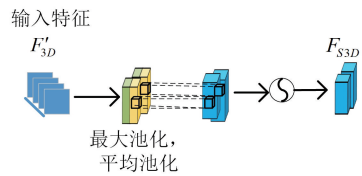

图 3 空间注意模块

Fig. 3 Spatial attention module

空间注意模块如图 3 所示,对输入特征 F'_{3D} 进行处理,空间注意模块更加关注步态轮廓序列中重要像素的位置。最终特征表示 F''_{3D} 的计算过程为:

$$F_{S3D}=\sigma(f^{1\times7\times7}([AvgPool3D(F'_{3D}),MaxPool3D(F'_{3D})])) \tag{3}$$

$$F''_{3D}=F'_{3D}\otimes F_{S3D} \tag{4}$$

式中:σ 表示 sigmoid 函数;$f^{1\times7\times7}$ 表示滤波器大小为 1×

7×7 的三维卷积层；*AvgPool3D*(·) 和 *MaxPool3D*(·) 表示全局平均池化和全局最大池化操作。

1.3　增强型全局和局部特征提取器(EGLFE)

EGLFE 的核心组成部分为 GLSTC(global spatial-temporal convolutional)层。GLSTC 层融合了全局特征提取器和局部特征提取器。增强型全局和局部特征提取器结构为 GLSTC-A0—MaxPooling—GLSTC-A1—GLSTC-B，输入为 $X_{global} \in R^{c1\times t\times h\times w}$，将全局特征映射划分为 n 个局部特征映射 $X_{local}=\{X_{local}^{i} \mid i=1,\cdots,n\}$，其中 n 是分区的数目，$X_{local}^{i} \in R^{c1\times t\times \frac{h}{n}\times w}$ 对应第 i 局部步态部分。研究使用三维卷积分别提取全局和局部步态特征，所有局部特征映射共享相同的卷积权重。将全局时空特征映射和局部特征映射结合起来有两种方法，通过元素相加(GLSTC-Ai，其中 i 值为 0、1 或连接(GLSTC-B)，分别表示为：

$$Y_{GLSTC-B}=\begin{pmatrix}Y_{global}\\ Y_{local}\end{pmatrix} \in R^{c2\times t\times 2h\times w} \tag{5}$$

$$Y_{GLSTC-Ai}=Y_{global}+Y_{local} \in R^{c2\times t\times h\times w} \tag{6}$$

其中，

$$Y_{global}=f_{3\times1\times1}(f_{1\times3\times3}(X_{global})) \in R^{c2\times t\times h\times w} \tag{7}$$

$$Y_{local}=\begin{Bmatrix}f'_{3\times3\times3}(Y_{local}^{1})\\ f'_{3\times3\times3}(Y_{local}^{2})\\ \cdots\\ f'_{3\times3\times3}(Y_{local}^{n})\end{Bmatrix} \in R^{c2\times t\times h\times w} \tag{8}$$

局部特征提取器则专注于步态图像中的特定区域，提取出与步态分析密切相关的局部特征。

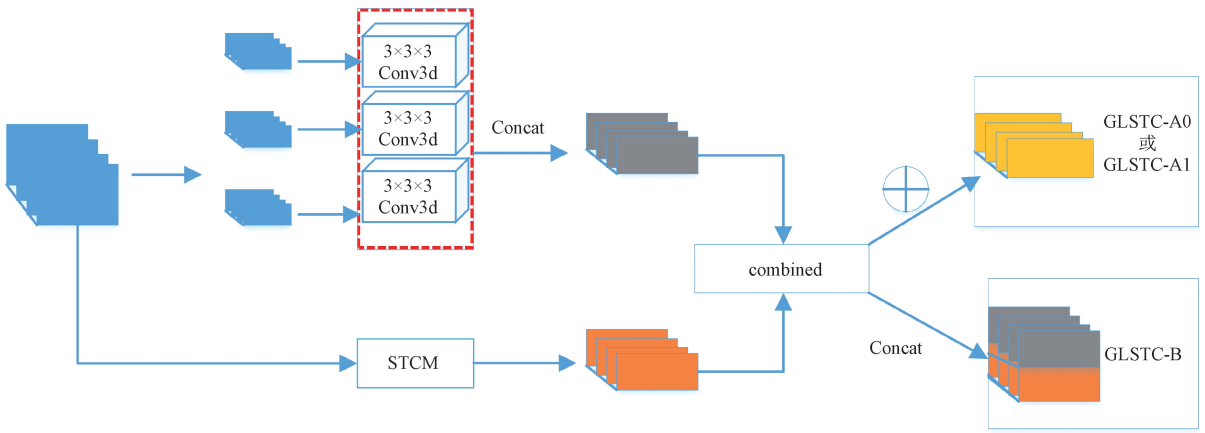

图 4　增强型全局和局部特征提取器

Fig. 4　Enhanced global and local feature extractor

研究将完整的 $N_{i-1}\times t\times D\times D$ 的 3D 卷积通过二维卷积和一维卷积来近似，将空间和时间建模分解为两个独立的步骤。设计了一个时空卷积模块(spatial-temporal convolutional module，STCM)，如图 5 所示，该模块由 $N_{i-1}\times 1\times D\times D$ 的 M_i 二维卷积滤波器和 $M_i\times t\times 1\times 1$ 的 N_i 时域卷积滤波器组成。其中超参数 M_i 确定了中间子空间的维数，其中信号在空间和时间之间进行投影，如式(9)所示。

$$M_i=\left\lfloor\frac{td^2N_{i-1}N_i}{d^2N_{i-1}+tN_i}\right\rfloor \tag{9}$$

其中，使 STCM 模块中的参数数量近似等于实现全三维卷积的参数数量。

STCM 模块在进行全局步态特征提取时没有改变参数的数量，但由于在 2D 卷积和 1D 卷积之间的额外的 LeakyRelu，它使网络的非线性数量增加了 1 倍，在进行步态特征提取时起到扩大感受野的效果。

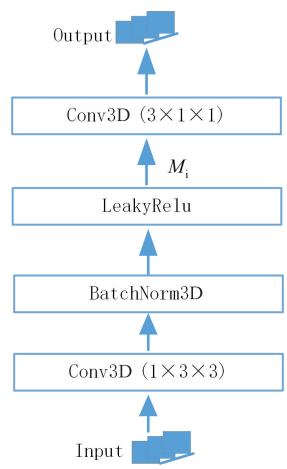

图 5　时空卷积模块

Fig. 5　Spatio-temporal convolution module

1.4 多尺度时间增强模块(MTEM)

在以往的步态识别方法中,大多数方案主要依赖于标准的卷积操作来提取步态序列时间特征。然而,这些方法的局限性在于卷积操作的局部性质,即其有限的感受野,这使得这些方法在捕捉步态序列中的长期时间依赖关系面临挑战。

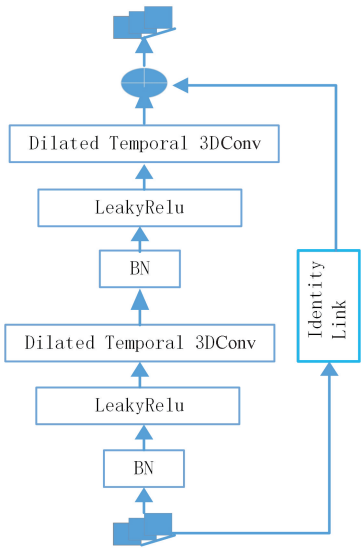

图 6 多尺度时序聚合单元

Fig. 6 Multi-Scale time series aggregation module

考虑到步态序列的特性,相邻帧的采集时间间隔往往非常短暂,通常在 0.04 s 内,这导致相邻帧之间的差异非常细微[17]。因此,仅通过相邻帧之间的时间建模很难学习到具有足够判别性的时间特征。为此,本研究提出了一种多尺度时间增强模块(multi-scale temporal enhancement module,MTEM),该模块由带有残差连接的多尺度扩展卷积块和帧级特征提取模块组成。多尺度时序聚合单元(multi-scale time series aggregation unit,MTAU)提取长短期时序特征,同时帧级特征提取器(frame-level feature extractor,FFE)提取帧级时序特征,在多尺度时间聚合特征的基础上融合帧级特征,捕获步态序列全局信息的同时融合每一帧的细节信息,实现信息互补,使模型获得更好的性能。

如图 6 所示,多尺度时间聚合单元 MTAU 概述,它是扩展的卷积块,由扩展的 3D 卷积层、LeakyRelu 和 BatchNorm 组成,最终多尺度时间增强模块的输出结果为 F_{MTEM},F_{MTEM} 的计算过程为:

$$F_t = MTAU2(MTAU1(Y_{EGLFE})) \tag{10}$$

$$F_f = Conv_{1\times3\times3}(Y_{EGLFE}) \in R^{c2\times t\times h\times w} \tag{11}$$

$$F_{MTEM} = F_t + F_f \tag{12}$$

式中:$MTAU1$ 的空洞率为 2,卷积核大小为 3×1×1;其中 $MTAU2$ 的空洞率为 4,卷积核大小为 3×1×1;FFE 的卷积核大小为 1×3×3;残差连接也在"MTAU" 中使用。

1.5 特征映射

对于时间特征映射,引入了时间 MaxPooling 层来聚合长步态序列的时间信息。假设特征映射 $Z \in R^{C_2\times T_2\times H\times W_2}$ 是特征提取模块的最终输出。时间特征映射表示为:

$$Y = MaxPooling^{1\times T_2\times 1\times 1}(Z) \tag{13}$$

其中,$Y \in R^{C_2\times 1\times H\times W_2}$ 为特征映射的输出,$MaxPooling^{1\times T_2\times 1\times 1}(\cdot)$ 对长度为 T_2 的序列执行 $MaxPooling$ 操作。

对于空间特征映射,使用 GeM 生成多个水平特征表示,然后将空间信息整合到特征映射中。大多数研究者仅通过平均池化[18]和最大池化的加权和来融合特征,而 $GeM(\cdot)$ 可以直接将这两种不同的操作融合形成一个特征映射,其中 $P = \infty$ 等于 $MaxPooling(\cdot)$,$P = 1$ 等于 $AvgPooling(\cdot)$,然后,使用多个独立的全连接层进一步聚合来自 Y_{GeM} 通道的信息。特征映射可以定义为:

$$S = FC(GEM(Y)) \tag{14}$$

$$GEM(Y) = (AvgPooling^{1\times1\times1\times W_2}((Y)^P))^{\frac{1}{P}} \tag{15}$$

$AvgPooling^{1\times1\times1\times W_2}(\cdot)$ 是核大小为 $(1\times1\times1\times W_2)$ 的平均池化操作。$FC(\cdot)$ 是全连接层,将步态特征映射到更具判别性的嵌入空间中进行最终的步态识别。

1.6 损失函数

研究使用三元组损失[15]和交叉熵损失组成的联合损失函数来训练网络。使用三元组损失来减少类内距离,增加类间距离,另外使用交叉熵损失进行分类,便于在训练过程中对模型进行优化。组合损失函数表示为:

$$L_{conbined} = L_{triplet} + L_{entropy} \tag{16}$$

式中:$L_{triplet}$ 和 $L_{entropy}$ 分别表示三元组损失和交叉熵损失。$L_{triplet}$ 可以表示为:

$$L_{triplet}(a,p,n) = \max\{D_{a,p} - D_{a,n} + margin, 0\} \tag{17}$$

式中:a 表示锚样本;p 表示正样本;n 表示负样本;$D_{a,p}$ 是正样本对之间的距离;$D_{a,n}$ 为负样本对之间的距离;$margin$(余量)是一个调节优化难度的超参数。

2 实验与分析

2.1 数据集和评估方案

本文在 CASIA-B[19]数据集上评估了所提出方法的性能。CASIA-B 数据集是一个广泛使用的跨视角步态识别数据集,由 124 个不同行人的步态信息组成。每个行人有 10 组步态序列,这些序列包含了 3 种不同的行走条件:正常行走(NM,包含 6 组序列,即 NM#01-#06)、携带背包(BG,包含 2 组序列,即 BG#01-#02)以及穿着不同

衣服(CL,通常是大衣或夹克,包含 2 组序列,即 CL#01-#02)。在每组序列中,步态图像是在从 0°~180°的范围内,以 18°为间隔的 11 个不同视角下捕获的。因此,CASIA-B 中有 124(受试者)×10(组)×11(视角)= 13 640 个步态序列。将每个受试者的步态序列分为训练集和测试集。选取 74 个受试者作为训练集,其余 50 个受试者进行测试。

OU-MVLP[20]数据集是最广泛的公共步态数据集之一。该数据集包含 10 307 名年龄在 2~87 岁的受试者的步态视频序列。每个受试者对应两组序列,编号为 seq# 00 和 seq# 01。每个序列包含 14 个不同的采样角度(0°~90°和 180°~270°,采样间隔为 15°)。由于 OU-MVLP 数据集包含更多的主题,因此可以更好地评估模型的泛化潜力。在训练阶段,本文根据文献[7]中的实验设置,使用5 153 名受试者作为训练数据,剩余的5 154 名受试者用于测试。

2.2　实验设置

步态轮廓被归一化为 64×44,将损失函数中的 *margin* 设为 0.2,特征映射中的参数 p 设为 6.5,以 Adam 为优化器,将每个步态序列 T 的长度设为 30。在 CASIA-B 数据集中,参数 P 和 K 分别设置为 4 和 8,P 代表该批次受试者的数量,K 代表每个受试者的样本数量。本文方法的网络框架如图 1 所示,3D-CBAM,GLSTC-A0,GLSTC-A1,GLSTC-B,MATU1,MATU2 的输出通道分别为 32,64, 128,256,256 和 256。实验的学习率 λ 初始化为 10^{-4},总迭代数设置为 80 K。在 70 K 次迭代中,学习率 λ 衰减到 10^{-5}。

对于 OU-MVLP,参数 P 和 K 分别设置为 16 和 4。由于 OU-MVLP 中受试者较多,本研究采用更深层次的网络设置,将“GLSTC-A0”、“GLSTC-A1”、“GLSTC-B”“MTAU1”4 个模块的数量加倍。4 个模块的输出通道分别为 64、128、256 和 256。λ 初始化为 4×10^{-4},总迭代次数设置为 130 K。在 60 K 和 110 K 迭代时,λ 将分别减少到 4×10^{-5} 和 4×10^{-6}。

2.3　实验结果

在 CASIA-B 数据集上,所提方法与 GaitSet[7]、GaitSet 改进[8]、GaitPart[12]、MT3D[10]、GaitGL[14]和 ESNet[13]方法进行了比较,主要结果如表 1 所示。

表 1　不同方法在 CASA-B 数据集的平均 Rank-1 准确率,不包括相同视角的情况

Table 1　Average Rank-1 accuracy of different methods in the CASA-B dataset, excluding the same viewing angle (%)

	Probe	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	平均
NM	GaitSet	90.8	97.9	99.4	96.9	93.6	91.7	95.0	97.8	98.9	96.8	85.8	95.0
	GaitSet 改进	94.0	98.4	99.0	98.1	92.9	91.5	95.5	98.6	99.4	98.6	91.3	96.1
	GaitPart	94.1	98.6	99.3	98.5	94.0	92.3	95.9	98.4	99.2	97.8	90.4	96.2
	MT3D	95.7	98.2	99.0	97.5	95.1	93.9	96.1	98.6	99.2	98.2	92.0	96.7
	GaitGL	96.0	98.3	99.0	97.9	96.9	95.4	97.0	98.9	99.3	98.8	94.0	97.4
	ESNet	95.6	98.6	99.1	97.9	96.7	94.4	96.9	98.7	99.3	98.6	95.1	97.4
	Ours	**96.7**	**99.0**	**99.30**	**98.3**	**97.1**	**95.9**	**98.8**	**99.5**	**99.7**	**99.5**	**95.3**	**98.1**
BG	GaitSet	83.8	91.2	91.8	88.8	83.3	81.0	84.1	90.0	92.2	94.4	79.0	87.2
	GaitSet 改进	89.5	94.9	93.6	92.8	86.8	80.2	84.8	92.8	96.1	93.6	84.8	90.0
	GaitPart	89.1	94.8	96.7	95.1	88.3	84.9	89.0	93.5	96.1	93.8	85.8	91.5
	MT3D	91.0	95.4	97.5	94.2	92.3	86.9	91.2	95.6	97.3	96.4	86.6	93.0
	GaitGL	92.6	96.6	96.8	95.5	93.5	89.3	92.2	96.5	98.2	96.9	91.5	94.5
	ESNet	92.7	95.9	96.3	94.9	93.2	87.7	90.9	96.2	97.3	96.9	91.7	94.0
	Ours	**93.6**	**96.8**	**97.2**	**95.7**	**94.1**	**90.0**	**94.4**	**97.3**	**98.4**	**97.4**	**91.7**	**95.1**
CL	GaitSet	61.4	75.4	80.7	77.3	72.1	70.1	71.5	73.5	73.5	68.4	50.0	70.4
	GaitSet 改进	70.8	83.7	86.3	80.9	75.3	69.4	76.4	78.6	80.9	79.1	65.1	77.0
	GaitPart	70.7	85.5	86.9	83.3	77.1	72.5	76.9	82.2	83.8	80.2	66.5	78.7
	MT3D	76.0	87.6	89.8	85.0	81.2	75.7	81.0	84.5	85.4	82.2	68.1	81.5
	GaitGL	**76.6**	90.0	90.3	87.1	84.5	79.0	84.1	87.0	87.3	84.4	69.5	83.6
	ESNet	75.6	89.2	92.4	90.3	84.3	**80.2**	83.0	86.3	89.0	83.9	69.8	84.0
	Ours	75.7	**89.6**	**92.3**	89.6	**85.9**	79.0	**84.9**	**89.2**	**91.9**	**84.9**	**70.8**	**84.9**

从表 1 可以看出,当外部环境发生变化时,精度明显下降。例如 GaitGL 在 NM,BG 和 CL 3 种情况下的识别准确率分别为 97.4%,94.5%,83.6%。本研究提出的方法在这 3 种情况下的识别准确率分别为 98.1%,95.1%和 84.9%,在 CL 条件下相比 GaitGL 高 1.3%。实验结果表明,该方法在 CL 条件下具有显著优势,表明该模型可以提取更加鲁棒的步态特征,减小遮挡对步态识别的影响,即使在 NM 和 BG 条件下,本研究的结果也要优于 GaitGL,所提出的方法也获得了最好的性能。在真实场景下,人类的步态可以从任意视角和条件下收集,研究步

态识别在各种外部环境因素下的鲁棒性至关重要。进一步计算了 3 种情况下的平均精度,表 2 所示为与目前经典的步态识别方法的对比结果。可以看出本文方法的平均准确率为 92.7,相比 GaitSet、GaitSet 改进、GaitPart、MT3D、GaitGL 和 ESNet 分别高出 8.1、5.0、3.5、1.9、0.9 和 0.9。与其他方法的准确率相比,均有明显提高。主要影响因素包括 1)EGLFE 模块中 GLSTC 层使用 STCM 时空卷积模块提高了模型对全局特征的表达能力,融合局部特征表示,有效弥补因局部分块带来的特征损失;2)利用 3D-CBAM 注意力机制使模型能够更加关注步态序列中重要的通道信息和空间信息,提取更加具有判别性的步态特征,提高模型的精度;3)使用多尺度时间增强模块,使研究模型能够提取更加丰富的时间信息,提高模型对遮挡的鲁棒性。

表 2 CASIA-B 的总体平均 Rank-1 准确度

Table 2 Overall average Rank-1 accuracy of CASIA-B (%)

Method	NM	BG	CL	平均
GaitSet	95.0	87.2	70.4	84.2
GaitSet 改进	96.1	90.0	77.0	87.7
GaitPart	96.2	91.5	78.7	88.8
MT3D	96.7	93.0	81.5	90.4
GaitGL	97.4	94.5	83.6	91.8
ESNet	97.4	94.0	84.0	91.8
本文	**98.1**	**95.1**	**84.9**	**92.7**

表 3 不同方法在 OU-MVLP 数据集的平均 Rank-1 准确率,不包括相同视角的情况

Table 3 Average Rank-1 accuracy of different methods in the OU-MVLP dataset, excluding the same viewing angle (%)

方法	0°	15°	30°	45°	60°	75°	90°	180°	195°	210°	225°	240°	255°	270°	平均
GaitSet	79.5	87.9	89.9	90.2	88.1	88.7	87.8	81.7	86.7	89.0	89.3	87.2	87.8	86.2	87.1
GaitSet 改进	81.3	90.0	89.9	90.5	89.3	89.2	88.4	83.5	88.1	89.3	89.5	88.6	88.5	86.8	88.1
GaitPart	82.6	88.9	90.8	91.0	89.7	89.9	89.5	85.2	88.1	90.0	90.1	89.0	89.1	88.2	88.7
GaitGL	84.9	90.2	91.1	91.5	91.1	90.8	90.3	88.5	88.6	90.3	90.4	89.6	89.3	88.5	89.2
ESNet	84.8	89.6	91.0	91.3	90.7	90.4	89.9	88.5	87.5	90.1	90.2	89.4	89.3	88.5	89.4
本文	**85.7**	**90.3**	**91.1**	**91.3**	**91.2**	**90.8**	**90.7**	**89.4**	**88.3**	**90.1**	**90.2**	**89.8**	**89.9**	**89.0**	**89.8**

在 OUMVLP 数据集上,还将本研究的方法与领先的基于视频的方法进行了比较,包括 GaitSet、GaitSet 改进、GaitPart、MT3D、GaitGL 和 ESNet。表 3 为这些方法在 OU-MVLP 数据集上的性能。结果表明,本文的方法优于其他方法,在大多数视图中达到了最高的准确率,平均准确率为 89.8%。结果表明,该方法可以有效地应用于大规模数据集。

2.4 复杂性分析

本研究对模型进行了复杂度分析,结果如表 4 所示。在 CASIA-B 数据集上,使用相同的设备和设置,分别对本研究的模型和 GaitGL 模型进行了实验。与 GaitGL 相比,本研究的模型拥有更多的参数,这表明本研究的模型在特征表示能力上更为出色,但同时也需要更多的计算资源来进行训练和推理。从表 4 可以看出,在 CASIA-B 数据集上,本研究的模型的推理速度为 31.35/it·s^{-1},而 GaitGL 的推理速度为 93.07/it·s^{-1}。尽管本研究的模型的推理速度相较于 GaitGL 稍慢,但它仍然保持了一个可接受的速度,与步态识别领域普遍采用的推理速度标准相吻合。

表 4 CASIA-B 下 GaitGL 和本文方法的复杂度结果

Table 4 Complexity results of GaitGL and our method under the CASIA-B dataset

模型	Parameters Count/(×10^6)	Inference Speed/(it·s^{-1})
GaiGL	3.10	93.07
本文	7.90	31.35

两种模型均在两台 NVIDIA GeForce RTX 3080 显卡上进行了计算性能评估。

2.5 消融实验

本研究步态识别方法包括 EGLFE 和多尺度时序特征提取模块 MTEM,3D 卷积注意模块(3D-CBAM)等关键模块。因此,本研究设计了不同的消融研究以分析每个关键模块的贡献。由于 CASIA-B 包含了更多的行走条件来验证不同情况下的模块,本研究对该数据集进行消融实验。消融实验结果如表 5 所示,选择 GaitGL 作为基线,仅通过使用 EGLFE 模块,平均精度提高 0.2,EGLFE 模块中使用 STCM 模块,与普通的 3DConv 相比,(2+1)D 分解[21]在步态特征提取任务中具有增加非线性与函数复杂性、优化便捷性、更好的时空特征提取能力以及减少参数冗余与计算成本等优势,在进行步态特征提取时起到扩大感受野的效果,同时还能提取细节的空间信息。

表 5 不同条件下的特征提取准确率

Table 5 Accuracy of feature extraction under different conditions (%)

方法	EGLFE	MTEM	3D-CBAM	NM	BG	CL	平均
a	-	-	-	97.4	94.5	83.6	91.8
b	√	-	-	97.5	94.7	83.9	92.0(↑0.2)
c	√	√	-	97.4	94.8	84.8	92.3(↑0.5)
d	√	-	√	97.7	94.7	84.3	92.2(↑0.4)
f	√	√	√	98.1	95.1	84.9	92.7(↑0.9)

MTEM 对时间特征进行多尺度提取,同时提取帧级特征,丰富特征信息,在多尺度时间聚合特征的基础上融合帧级特征提取,提取步态序列中的多尺度时间信息的同时也能够对每帧提取图像的细节特征,有效提高了网络对遮挡的鲁棒性。在方法 b 基础上加入 MTEM,平均精度提高了 0.5%,在 BG 条件和 CL 条件下,性能分别提高了 0.3 和 1.3。方法 d 是在方法 b 上融合 3D-CBAM 模块,相较于 GaitGL 算法,模型能够更加关注步态特征的关键信息,提升模型的性能,在 NM、BG 和 CL 条件下提高了识别准确率。方法 f 同时使用 EGLFE、MTEM 和 3D-CBAM 模块,在正常行走条件下,该方法达到了平均 98.1%(97.4%)的准确率。由于对特征信息的挖掘较深入,模型的鲁棒性相对较强,对于协变量因素的影响也有一定的突破,在穿大衣和背包的条件下,平均识别率分别达到了 95.1%(94.5%)、84.9%(83.6%),本研究的 baseline 模型比 GaitGL 在相同条件下的平均识别率都有一定的提高。

3　结　论

为了充分利用空间特征信息和时序特征信息,提出了融合多尺度时序与增强时空表征的步态识别框架。首先,为了提取更加全面的步态信息,所提方法采用局部特征提取和全局特征提取鲁棒性的步态特征表示。为了提取有效的时空信息,本研究引入了时空卷积模块,先提取步态序列中包含的空间信息,然后进行时序特征提取,在保证参数量一致的情况下,扩大了特征提取的感受野。引入多尺度特征提取模块,实现帧级时序特征和多尺度时序特征融合,提取多尺度的时序特征,减小遮挡对步态识别的影响。最后,引入广义均值池化层以自适应聚合空间的信息,提高了特征映射的性能。实验结果验证了所提方法的有效性。通过在公共数据集上进行对比,发现所提方法在复杂条件下的准确率虽有显著提升,但相比于 NM 条件,复杂条件下的准确率仍有待提高。因此,如何提高复杂条件下步态识别的准确率是未来研究的重点。

参考文献

[1] SHEN CH F, YU SH Q, WANG J L, et al. A comprehensive survey on deep gait recognition: Algorithms, datasets and challenges [J]. EEE Transactions on Biometrics, Behavior, and Identity Science, 2024, DOI: 10.1109/TBIOM.2024.3486345.

[2] ZHANG Z Y, TRAN L, LIU F, et al. On learning disentangled representations for gait recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(1): 345-360.

[3] 王铭,林贝贝,张顺利. 基于帧级-时空双分支网络的步态识别方法[J]. 北京邮电大学学报,2023,46(3):73-77,96.
WANG M, LIN B B, ZHANG SH L. A gait recognition method based on frame-level and spatio-temporal dual-branch network [J]. Journal of Beijing University of Posts and Telecommunications, 2023, 46(3): 73-77,96.

[4] 魏淇,潘晴,田妮莉. 基于局部特征的时间注意力图卷积步态识别方法[J]. 国外电子测量技术,2023,42(10):19-24.
WEI Q, PAN Q, TIAN N L. A gait recognition method based on temporal attention graph convolution with local features [J]. Foreign Electronic Measurement Technology, 2023, 42(10): 19-24.

[5] AN W ZH, YU SH Q, MAKIHARA Y, et al. Performance evaluation of model-based gait on multi-view very large population database with pose sequences[J]. IEEE Transactions on Biometrics, Behavior, and Identity Science, 2020, 2(4): 421-430.

[6] 邹雪,谭棉,严晓波,等. 基于多尺度特征融合的跨视角步态识别[J]. 电子测量技术,2024,47(1):186-192.
ZOU X, TAN M, YAN X B, et al. Cross-view gait recognition based on multi-scale feature fusion [J]. Electronic Measurement Techniques, 2024, 47(1): 186-192.

[7] 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, 2019, 33(1): 8126-8133.

[8] 陈万志,唐浩博,王天元. 融合轮廓增强和注意力机制的改进 GaitSet 步态识别方法[J]. 电子测量与仪器学报,2024,38(1):203-210.
CHEN W ZH, TANG H B, WANG T Y. An improved GaitSet gait recognition method incorporating contour enhancement and attention mechanism [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(1): 203-210.

[9] HOU S H, CAO CH SH, LIU X, et al. Gait lateral network: Learning discriminative and compact representations for gait recognition[C]. European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 382-398.

[10] LIN B B, ZHANG SH L, BAO F. Gait recognition with multiple-temporal-scale 3D convolutional neural network[C]. Proceedings of the 28th ACM International Conference on Multimedia, 2020: 3054-3062.

[11] ZHANG Y Q, HUANG Y ZH, YU SH Q, et al. Cross-view gait recognition by discriminative feature learning[J]. IEEE Transactions on Image Processing, 2019, 29: 1001-1015.
[12] FAN CH, PENG Y J, CAO CH SH, et al. Gaitpart: Temporal part-based model for gait recognition [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 14225-14233.
[13] HUANG T H, BEN X Y, GONG CH, et al. Enhanced spatial-temporal salience for cross-view gait recognition[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(10): 6967-6980.
[14] LIN B B, ZHANG SH L, YU X. Gait recognition via effective global-local feature representation and local temporal aggregation[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 14648-14656.
[15] HERMANS A, BEYER L, LEIBE B. In defense of the triplet loss for person re-identification [J]. 2017, DOI: 10.48550/arXiv.1703.07737.
[16] SEPAS-MOGHADDAM A, ETEMAD A. Deep gait recognition: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(1): 264-284.
[17] SUN Y, LONG H, FENG X L, et al. GaitASMS: Gait recognition by adaptive structured spatial representation and multi-scale temporal aggregation [J]. Neural Computing and Applications, 2024, 36(13): 7057-7069.
[18] FU Y, WEI Y CH, ZHOU Y Q, et al. Horizontal pyramid matching for person re-identification [C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 8295-8302.
[19] YU SH Q, TAN D L, TAN T N. 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.
[20] TAKEMURA N, MAKIHARA Y, MURAMATSU D, et al. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition [J]. IPSJ Transactions on Computer Vision and Applications, 2018, 10: 1-14.
[21] TRAN D, WANG H, TORRESANI L, et al. A closer look at spatiotemporal convolutions for action recognition [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6450-6459.

作者简介

许振齐，现为长春理工大学硕士研究生，主要研究方向为图像处理、机器视觉。
E-mail：3498036029@qq.com

Xu Zhenqi is now a M. Sc. candidate at Changchun University of Science and Technology. His main research interests include image processing and machine vision.

朴燕（通信作者），1988 年于哈尔滨工业大学获得学士学位，1995 年于中科院长春物理研究所获得硕士学位，2000 年于中国科学院长春光学精密机械与物理研究所获得博士学位，现为长春理工大学教授，主要研究方向机器视觉与成像技术。
E-mail：piaoyan@cust.edu.cn

Piao Yan (Corresponding author) received her B. Sc. degree from Harbin Institute of Technology in 1988, her M. Sc. degree from Changchun Institute of Physics, Chinese Academy of Sciences in 1995 and her Ph. D. degree from Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences in 2000. Now she is a professor at Changchun University of Science and Technology. Her main research interests in machine vision and imaging technology.

康继元，现为长春理工大学硕士研究生，主要研究方向为图像处理、机器视觉。
E-mail：1292455382@qq.com

Kang Jiyuan is now a M. Sc. candidate at Science and Technology. Her main research interests include image processing and machine vision.

鞠成伟，现为长春理工大学硕士研究生，主要研究方向为图像处理、机器视觉。
E-mail：1337624084@qq.com

Ju Chengwei is now a M. Sc. candidate at Changchun University of Science and Technology. His main research interests include image processing and machine vision.