DOI:10.19651/j.cnki.emt.2519990

# 基于结构感知与动态融合机制的步态识别研究

翟奥北[1] 奚 昊[1] 范金豪[1] 胡传平[2]
(1.郑州大学电气与信息工程学院 郑州 450001; 2.郑州大学网络空间安全学院 郑州 450052)

**摘 要**: 针对基于轮廓的步态识别方法因过度依赖全局表征,而在服装变化等强外观干扰下性能骤降的瓶颈,提出了一种融合结构感知与动态注意力的识别模型。该模型旨在将识别范式从匹配易变轮廓升维至理解内在运动规律。为此,构建了一个双路并行框架:首先,通过一条结构感知路径对人体关键局部区域进行精确建模;其次,引入频率解耦的动态注意力机制以自适应增强最具区分度的特征通道,应对步态相位变化;最后,利用深度语义融合模块将局部结构信息与全局表征进行多尺度协同,生成兼具稳定性与判别力的最终特征。实验结果表明,该模型在 CASIA-B 数据集上的平均准确率达到 89.9%,尤其在服装变化场景下较基线模型提升了 11.0%;同时在大规模 OU-MVLP 数据集上的 Rank-1 准确率达到 89.5%。研究证实,该模型通过协同局部结构感知与全局特征增强,有效提升了步态识别在复杂外观干扰下的稳健性与识别精度。

**关键词**: 步态识别;结构感知;动态注意力;特征融合
**中图分类号**: TP391.41;TN919.81 **文献标识码**: A **国家标准学科分类代码**: 520.2040

# Structure-aware and dynamically fused network for gait recognition

Zhai Aobei[1] Xi Hao[1] Fan Jinhao[1] Hu Chuanping[2]
(1. School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou 450001, China;
2. School of Cyberspace Security, Zhengzhou University, Zhengzhou 450052, China)

**Abstract**: To address the performance bottleneck of contour-based gait recognition, where over-reliance on global representations leads to sharp degradation under strong appearance interferences like clothing changes, this paper proposes a model integrating structure awareness and dynamic attention. The model aims to elevate the recognition paradigm from matching variable contours to understanding intrinsic motion patterns. To achieve this, this study construct a dual-path parallel framework: First, a structure-aware path precisely models key local body regions; second, a frequency-decoupled dynamic attention mechanism is introduced to adaptively enhance the most discriminative feature channels against gait phase variations; finally, a deep semantic fusion module synergizes local structural information with global representations at multiple scales to generate a final feature with both stability and discriminative power. Experimental results show that the model achieves an average accuracy of 89.9% on the CASIA-B dataset, with 11.0% improvement over the baseline under the changing-clothes condition, and a Rank-1 accuracy of 89.5% on the large-scale OU-MVLP dataset. This study confirms that by synergizing local structure perception and global feature enhancement, the proposed model effectively improves the robustness and accuracy of gait recognition under complex appearance interferences.

**Keywords**: gait recognition;structural serception;dynamic attention;feature fusio

## 0 引 言

步态识别作为一种远距离、非接触式的生物特征识别技术,因其在公共安全、智能监控等领域的巨大应用潜力而备受关注。与指纹[1]、人脸[2]等技术相比,步态识别无需目标主动配合[3],为复杂场景下的身份认证提供了独特优势。随着深度学习,特别是卷积神经网络(convolutional neural network, CNN)的应用[4],步态识别的研究范式发生了根本性转变。

自 Wu 等[5]首次将 CNN 引入该任务以来,深度学习模型迅速超越了以步态能量图(gait energy image, GEI)为代表的传统手工特征方法。其中,由 Chao 等[6]提出的

收稿日期:2025-10-10

GaitSet 模型是一个里程碑式的突破,它创新地将步态序列视为无序集合进行处理,极大地提升了模型对视角、速度变化的鲁棒性,奠定了当前主流方法的基础。受此启发,后续研究沿着两条主要技术路线不断演进。

第 1 条路线是在 GaitSet 框架下深化 2D 轮廓特征的表示,如 GaitPart[7] 探索了细粒度的人体部位划分。近期的工作则通过更精细的机制提升性能,例如,赵坚等[8-9]致力于融合判别性的全局与局部特征,而 Li 等[10]则设计了包含多维注意力机制的多分支网络。然而,这类方法本质上仍依赖于 2D 轮廓,在面对服装、背包等强外观干扰时性能瓶颈依然存在。

第 2 条路线则探索超越 2D 轮廓的新范式,以从根本上解决外观依赖问题。例如,吴越等[11]利用 SMPL 三维人体模型将步态分解为内在的体型与姿态,关迪元等[12]则采用 3D 点云序列作为输入,这些方法理论上对外观变化具有更强的鲁棒性。然而,它们通常需要复杂的预处理或特定的传感器,如韩亚丽等[13]的可穿戴外骨骼或刘彦伟等[14]的触觉感知设备,这限制了其在主流 2D 监控场景中的广泛适用性。

综上所述,现有研究面临一个核心权衡:主流 2D 方法虽部署便捷,但面对复杂现实场景时仍显脆弱;而前沿 3D 方法虽准确率高,却存在较高的应用门槛。因此,如何在计算开销可控的 2D 框架内有效克服外观干扰,成为一个亟待解决的关键问题。当前主流方法普遍缺乏精细的结构感知能力,难以在强干扰下聚焦于稳定的人体结构;其静态网络设计也导致特征表达缺乏对步态相位的自适应性[15-16];同时,初级的特征融合机制也忽略了不同特征间的语义差异[17],限制了识别性能。

为系统性地解决上述挑战,本文回归 2D 轮廓框架,提出一种融合结构感知与动态注意力的步态识别新模型,旨在不增加额外模态依赖的前提下,深度挖掘轮廓信息中的稳健运动模式。该模型通过并行的结构感知路径显式建模人体关键局部区域,以抵抗外观变化;并引入动态注意力机制自适应地增强关键特征通道;最后,通过深度语义融合模块实现局部细节与全局表征的高效协同。本文旨在通过这一系列设计,显著提升步态识别在复杂现实场景下的稳健性与精度。

## 1 方法与实验

### 1.1 整体框架概述

为提升步态识别在复杂条件下的识别效果,本文设计了一种采用双路并行与渐进式多层级融合的识别网络,其整体架构如图 1 所示。该框架通过一个负责全局建模的主路径和一个为结构感知而专门设计的水平结构感知路径(structural-aware path, SAP)并行提取特征,并利用动态语义融合模块(dynamic semantic fusion module, DSFM)在网络的不同深度进行信息交互,最终生成更具判别力的步态特征。

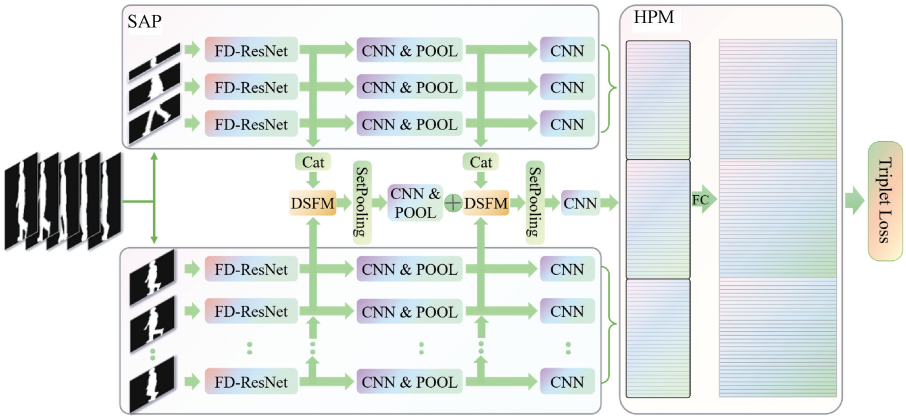



图 1 基于结构感知与动态融合的双路并行步态识别网络框架

Fig. 1 Dual parallel gait recognition network framework based on structural perception and dynamic fusion

具体而言,本文借鉴 GaitSet 的核心思想,将步态序列视为一个无序剪影集合 $X = \{\boldsymbol{x}_j \mid j = 1,\cdots,n\}$,其中 $\boldsymbol{x}_j$ 表示第 $j$ 帧步态剪影图像,$n$ 为序列长度。模型旨在从该集合中学习出具判别性的身份特征 $\boldsymbol{f}_i$。这一过程可形式化表示为:

$$\boldsymbol{f}_i = H(G(F(FD(\boldsymbol{x}_1)),F(FD(\boldsymbol{x}_2)),\cdots,F(FD(\boldsymbol{x}_n)))) \tag{1}$$

其中,$FD$ 是本文提出的浅层基于动态通道压缩比注意力机制的卷积网络(fully dynamic attention residual network, FD-ResNet),$F$ 代表用于提取帧级特征的卷积网络,$G$ 是一个排列不变函数(通过 SetPooling 实现,确保对输入帧顺序的不敏感性),$H$ 则是将集合级特征映射到最终判别空间的模块。

网络包含两条并行的特征提取路径。主路径(图 1 下方)负责处理完整的步态轮廓,旨在捕捉全局运动模式。与之并行的 SAP(图 1 上方)则引入解剖学先验,它将每帧轮廓图显式地分割为 3 个语义区域,并为各区域分配独立的子网络。该设计强制模型学习与特定身体部位相关的、

在服装变化或遮挡下更为稳定的局部特征。两条路径的浅层骨干网络均采用 FD-ResNet，它集成了动态注意力机制，能根据输入自适应地增强关键特征通道。

该框架的另一关键创新在于其多层级的渐进式融合机制。本文设计了 DSFM，取代常规的末端单次融合，在网络的浅层与深层进行两次信息交互。如图 1 所示，浅层融合旨在对齐轮廓、边缘等底层特征，而深层融合则用于整合高度抽象的语义信息，实现全局与局部特征的高效协同。

最后，经过多层级融合后的特征序列，通过 SetPooling 操作聚合时间维度的信息，并送入水平金字塔映射（HPM）模块生成最终的多尺度特征向量。整个网络采用三元组损失（Triplet Loss）进行优化，其基本形式旨在拉近同类样本距离并推远异类样本距离，可表示为：

$$L(a,p,n)=\mathrm{ReLU}(\xi+D(\boldsymbol{f}_a,\boldsymbol{f}_p)-D(\boldsymbol{f}_a,\boldsymbol{f}_n)) \tag{2}$$

其中，$\boldsymbol{f}_a$，$\boldsymbol{f}_p$，$\boldsymbol{f}_n$ 分别是锚点，正样本和负样本的特征表示，$D(\cdot,\cdot)$ 表示特征距离，$\xi$ 表示预设间隔。

### 1.2　结构感知路径

为解决全局特征在服装、背包等外观干扰下失效的问题，本文设计了 SAP，其核心思想源于对人体行走生物力学的观察：在行走过程中，人体不同部位的运动模式与摆动幅度存在显著差异。下肢（腿部）的周期性摆动最具动态性与判别力，是识别身份的关键信息来源；而躯干则相对稳定，主要负责维持身体平衡，头部变化最小，所以影响也是最小。同时，这些肢干并非独立运动，而是以一种高度协调、相互配合的方式共同完成步态周期。这种内在的运动结构稳定性，恰恰是抵抗外部轮廓变化的关键。SAP 通过对人体进行水平区域划分，如图 2 所示。对这些关键部位进行显式建模，旨在通过对人体关键部位的显式建模来提取稳健的局部特征。

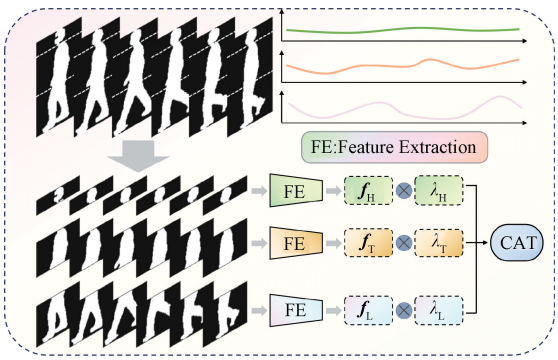



图 2　结构感知部位融合策略

Fig. 2　Structural perception site fusion strategy

SAP 的核心思想是引入解剖学先验。它首先将每帧步态轮廓图依据生物力学原理，水平划分为头部（H）、躯干（T）和下肢（L）3 个语义区域。随后，SAP 为这 3 个区域分配独立的特征提取分支，分别编码得到局部特征向量 $\boldsymbol{f}_H$，$\boldsymbol{f}_T$，$\boldsymbol{f}_L$ 这些局部特征通过一种基于经验权重的加权融合策略，生成最终的结构感知特征 $\boldsymbol{f}_{part}$：

$$\boldsymbol{f}_{part}=\lambda_H\cdot\boldsymbol{f}_H+\lambda_T\cdot\boldsymbol{f}_T+\lambda_L\cdot\boldsymbol{f}_L \tag{3}$$

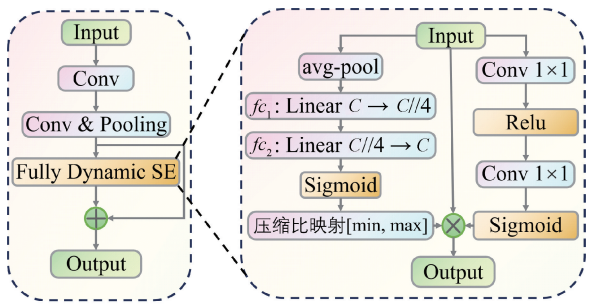
其中，权重 $\lambda_H$，$\lambda_T$，$\lambda_L$ 并非归一化系数，而是反映各部位贡献度的缩放因子。本文通过在 CASIA-B 数据集上的大量实验，确定了一组最优配置 $\lambda_H=0.2$，$\lambda_T=3$，$\lambda_L=5$。这种高对比度的权重分配（贡献度比例约 1∶15∶25）能够有效放大最具判别力的下肢和躯干特征信号，同时保留头部作为稳定参考基准，其效果显著优于简单的特征平均或归一化加权。

在整体网络中，SAP 作为一个并行的关键分支，其输出的结构感知特征 $\boldsymbol{f}_{part}$ 会在网络的不同深度被送入 DSFM 模块，与主路径的全局特征进行多层级融合，从而为最终的步态表示提供丰富的、抗干扰的局部信息。

### 1.3　基于动态通道压缩比注意力机制的改进残差网络

传统注意力机制（如 SE-Net）采用固定的通道压缩比，使其无法自适应地应对由视角、衣物等因素引起的步态特征剧烈变化。为解决这种“注意力僵化”问题，本文提出了 FD-ResNet，它将一个新颖的完全动态通道压缩比注意力模块（fully dynamic squeeze-and-excitation，FD-SE）深度集成到了残差网络中。

FD-ResNet 的结构如图 3(a)所示，它在标准残差块中嵌入了 FD-SE 模块。在整体架构中，FD-ResNet 被用作主路径与 SAP 各分支共享的浅层特征提取骨干，确保在深层处理前，所有特征都经过了动态的筛选与强化。




(a) 改进的动态通道压缩比注意力的特征提取网络结构（FD-ResNet）
(a) Improed dynamic attention residual network (FD-ResNet)

(b) 动态通道压缩比注意力（FD-SE）
(b) Dynamic channel compression ratio attention(FD-SE)

图 3　基于动态通道压缩比注意力的特征提取网络结构

Fig. 3　Structure of the feature extraction network based on dynamic channet compressior ratio attention

FD-SE 模块是一种采用三重乘性门控（triple multiplicative gating）的注意力结构，其核心如图 3(b)所示是通过两个并行分支，对输入特征进行双重动态调制。其工作流程包括：

1）动态通道缩放因子预测：第 1 分支为每个输入样本动态地预测一个最适合的通道缩放因子 $\boldsymbol{r}_c$，代表各通道的“绝对重要性”。计算公式为：

$$\boldsymbol{r}_c = r_{\min} + (r_{\max} - r_{\min}) \cdot \sigma(\boldsymbol{W}_r^{(1)} \cdot \mathrm{ReLU}(\boldsymbol{W}_r^{(2)} \cdot \boldsymbol{z})) \tag{4}$$

其中,$\boldsymbol{W}_r^{(1)} \in \mathbf{R}^{C/4\times C}$ 和 $\boldsymbol{W}_r^{(2)} \in \mathbf{R}^{C\times C/4}$ 为可学习权重矩阵,$\sigma$ 为 Sigmoid 激活函数。

2)注意力图生成:第 2 分支并行地生成一个标准的空间注意力图 $\boldsymbol{s}_c$,用以捕捉特征内的"相对重要性"区域。

$$\boldsymbol{s}_c = \sigma(\mathrm{Conv}_2 \cdot \mathrm{ReLU}(\mathrm{Conv}_1(\boldsymbol{X}_c))) \tag{5}$$

最终,通过三重乘性门控将原始特征 $\boldsymbol{X}_c$、注意力图 $\boldsymbol{s}_c$ 和缩放因子 $\boldsymbol{r}_c$ 进行逐元素相乘,得到最终输出 $\boldsymbol{Y}_c$:

$$\boldsymbol{Y}_c = \boldsymbol{X}_c \odot \boldsymbol{s}_c \odot \boldsymbol{r}_c \tag{6}$$

这种结合了"相对"与"绝对"重要性的双重动态调制机制,赋予了网络更强大、更灵活的特征选择与增强能力,从根本上提升了特征表达的质量。

### 1.4 动态语义融合模块

为解决 GaitSet 等全局模型在衣物、背包等外观干扰下性能下降的问题,本文设计了 DSFM。它作为一个关键桥梁,旨在智能地协同主路径的全局特征与 SAP 路径的局部结构特征。受人类视觉系统"何物"(What)与"何处"(Where)认知策略的启发,DSFM 采用并行的双重注意力机制,对融合后的特征进行通道和空间维度的双重优化,如图 4 所示。

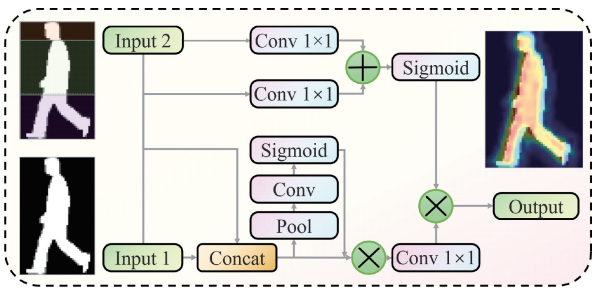



图 4 动态语义融合模块(DSFM)

Fig. 4 Dynamic semantic fusion module(DSFM)

DSFM 的工作流程为:

通道注意力加权(关注"何物"):为判断哪些特征通道最重要,模块首先将全局特征 $\boldsymbol{F}_m$ 与局部特征 $\boldsymbol{F}_s$ 拼接,然后通过全局池化和卷积计算出通道注意力权重 $\boldsymbol{M}_c$。该权重用于增强融合后最具判别力的信息通道。

$$\boldsymbol{M}_c = \sigma(f_{1\times1}(\mathrm{AvgPool}(\mathrm{Concat}(\boldsymbol{F}_m, \boldsymbol{F}_s)))) \tag{7}$$

随后,融合特征 $\boldsymbol{F}_{\mathrm{con}}$ 与通道注意力权重 $\boldsymbol{M}_c$ 逐通道相乘,形成初步加权结果,并再次通过一个 1×1 卷积进行通道维度的整合,输出最终的通道加权特征 $\boldsymbol{F}_c$:

$$\boldsymbol{F}_c = f_{1\times1}(\boldsymbol{M}_c \odot (\mathrm{Concat}(\boldsymbol{F}_m, \boldsymbol{F}_s))) \in \mathbf{R}^{C\times H\times W} \tag{8}$$

空间注意力聚焦(关注"何处"):为定位哪些空间区域最关键,模块并行地将两路输入的特征图相加,并经过 Sigmoid 函数生成空间注意力图 $\boldsymbol{F}_r$。该机制能有效引导模型聚焦于步态轮廓中的腿部、躯干等关键区域,抑制背景噪声。

$$\boldsymbol{F}_r = \sigma(f_{1\times1}(\boldsymbol{F}_m) + f_{1\times1}(\boldsymbol{F}_s)) \tag{9}$$

最终,经过通道注意力增强的特征与空间注意力图进行逐元素相乘,得到最终的输出特征 $\boldsymbol{F}_{\mathrm{out}}$,它既包含了全局的语义信息,又强化了局部的空间显著性。

$$\boldsymbol{F}_{\mathrm{out}} = \boldsymbol{F}_c \odot \boldsymbol{F}_r \tag{10}$$

本质上,DSFM 的双重注意力设计比单一注意力机制能更全面地建模特征,确保模型在复杂条件下仍能稳健地定位并利用关键的人体部位,从而显著提升识别的准确性。

## 2 实 验

### 2.1 数据集与实验设置

为全面评估模型性能,本文选用两大基准数据集:CASIA-B 与 OU-MVLP。这两个数据集分别侧重于行走条件变化和大规模跨视角识别,能系统性地检验本文模型的鲁棒性与泛化能力。

CASIA-B 是检验模型应对行走条件变化的黄金标准。它包含 124 名受试者在 3 种受控情境下的数据:正常行走(NM)、背包行走(BG)和更换外套(CL)。本文遵循其通用的大样本训练(LT)协议:使用前 74 人训练,后 50 人测试。测试时,以 NM 序列为注册集(Gallery),NM、BG、CL 这 3 类序列为探查集(Probe),并排除同视角匹配,以 Rank-1 准确率作为核心指标。

OU-MVLP[18] 是目前规模较大的公开步态数据集,包含 10 307 名受试者,是评估模型大规模跨视角识别能力和可扩展性的关键基准。实验遵循其官方设定,将数据集划分为 5 153 名训练受试者和 5 154 名测试受试者,并采用严格的跨视角匹配来衡量性能。

实验均在 NVIDIA Tesla V100S GPU 上完成,并基于 PyTorch 深度学习框架实现。为确保公平对比,所有模型的训练参数均保持一致。具体设置包括:优化器采用带动量的 SGD,权重衰减系数设为 0.000 5。损失函数采用三元组损失与交叉熵损失的组合,其中三元组损失的间隔(margin)设为 0.2。训练批次大小设为(8,16),即每个批次包含 8 个 ID,每个 ID 包含 16 个样本。初始学习率设为 0.001,并在第 20 000 和 40 000 次迭代时分别衰减为原来的 0.1 倍,总训练迭代次数为 60 000 次。

### 2.2 与主流方法的性能对比

1)在 CASIA-B 数据集上评估

本文首先在 CASIA-B 数据集上将模型与一系列代表性算法进行性能比较。对比方法涵盖了作为本文设计基线的 GaitSet[6],以及其改进版本 GaitSetV2[19]。此外,结果对比还包括了 GaitPart[7](一种采用部位划分思想的性能强大的模型)、GaitGraph[20](基于图结构的模型)、PGOFI[21](基于注意力机制和新颖局部表示的模型)以及文献[22]的工作(一种基于融合轮廓增强的模型)。实验结果如表 1 所示。

为了更直观地比较不同模型在应对各种状态干扰时

表 1　不同方法在 CASIA-B 数据集上的性能对比

Table 1　Performance comparison of different methods on the CASIA-B dataset　%

| 状态 | 模型 | 准确率 | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0° | 18° | 36° | 54° | 72° | 90° | 104° | 126° | 144° | 162° | 180° | Mean |
| NM | GaitGraph | 85.3 | 88.5 | 91.0 | 92.5 | 87.2 | 86.5 | 88.4 | 89.2 | 87.9 | 85.9 | 81.9 | 87.7 |
| | 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 |
| | GaitSetV2 | 91.1 | 99.0 | **99.9** | 97.8 | 95.1 | 94.5 | 96.1 | 98.3 | 99.2 | 98.1 | 88.0 | 96.1 |
| | PGOFI | 91.8 | 96.3 | 97.1 | 96.9 | **96.6** | **95.8** | 95.5 | 95.1 | 95.5 | 94.7 | **93.2** | 95.4 |
| | 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 |
| | 文献[22] | 94.0 | 98.4 | 99.0 | 98.1 | 92.9 | 91.5 | 95.5 | 98.6 | **99.4** | **98.6** | 91.3 | 96.1 |
| | 本文 | **94.3** | 97.2 | 99.8 | **98.6** | 95.2 | 95.5 | **96.5** | **98.8** | 99.0 | 98.2 | 92.8 | **96.9** |
| BG | GaitGraph | 75.8 | 76.7 | 75.9 | 76.1 | 71.4 | 73.9 | 78.0 | 74.7 | 75.4 | 75.4 | 69.2 | 74.8 |
| | 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 |
| | GaitSetV2 | 86.7 | 94.2 | 95.7 | 93.4 | 88.9 | 85.5 | 89.0 | 91.7 | 94.5 | 95.9 | 83.3 | 90.8 |
| | PGOFI | 88.3 | 91.3 | 92.5 | 92.1 | 91.8 | 94.5 | 90.8 | 92.2 | 92.7 | 91.1 | 90.2 | 91.6 |
| | GaitPart | 89.1 | 94.8 | **96.7** | **95.1** | 88.3 | **94.9** | 89.0 | **93.5** | 96.1 | 93.8 | 85.8 | 91.5 |
| | 文献[22] | 89.5 | 94.9 | 93.6 | 92.8 | 86.8 | 80.2 | 84.8 | 92.8 | **96.1** | 93.6 | 84.8 | 90.0 |
| | 本文 | **90.3** | **94.9** | 92.6 | 92.1 | **92.5** | 91.0 | **90.9** | 92.4 | 93.5 | 94.0 | **90.5** | **92.2** |
| CL | GaitGraph | 69.6 | 66.1 | 68.8 | 67.2 | 64.5 | 62.0 | 69.5 | 65.6 | 65.7 | 66.1 | 64.3 | 66.3 |
| | 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 |
| | GaitSetV2 | 59.5 | 75.0 | 78.3 | 74.6 | 71.4 | 71.3 | 70.8 | 74.1 | 74.6 | 69.4 | 54.1 | 70.3 |
| | PGOFI | 73.8 | 75.3 | 79.7 | 80.4 | **82.2** | 83.3 | 81.8 | 80.2 | 78.5 | 77.2 | 76.5 | 79.0 |
| | 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 |
| | 文献[22] | 70.8 | 83.7 | 86.3 | 80.9 | 75.3 | 69.4 | 76.4 | 78.6 | 80.9 | 79.1 | 65.1 | 77.0 |
| | 本文 | **75.8** | **86.2** | 86.2 | **84.3** | 80.4 | **79.9** | **82.1** | 80.9 | 81.3 | **81.0** | **77.2** | **81.4** |

的性能下降趋势，本研究将表 1 中的核心结果绘制成了图 5。结合表 1 和图 5 所示，本文提出的模型在 CASIA-B 数据集上取得了 89.9% 的平均识别准确率，显著超越了 GaitSet(84.2%)、GaitPart(88.8%)及 PGOFI(88.7%)等先进方法。5.7% 的性能提升(相较于基线 GaitSet)验证了本文架构性创新的有效性。

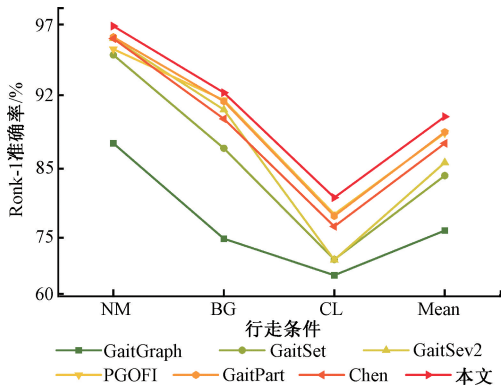



图 5　不同方法在 CASIA-B 数据集上的性能对比

Fig. 5　Performance comparison of different methods on the CASIA-B dataset

在正常行走(NM)和背包行走(BG)场景下，本文的模型分别以 96.9% 和 92.2% 的准确率高于所有对比方法。这表明本文引入的结构感知与动态注意力机制能有效提升步态特征的内在判别力，而不仅仅是应对干扰。

在最具挑战性的服装变化(CL)场景下，性能提升最为显著。GaitSet 等方法性能大幅滑落至 70% 左右，暴露出其处理严重外观变化的瓶颈。GaitPart 通过静态部位划分将准确率提升至 78.7%，初步证明了结构化思想的价值。而本文的模型在此基础上更进一步，达到了 81.4% 的准确率，较 GaitSet 和 GaitPart 分别提升了 11.0% 和 2.7%。

2)在 OU-MVLP 数据集上的跨视角性能评估

为进一步验证模型的泛化能力与在复杂视角变化下的稳健性，本文还在更大规模的跨视角数据集 OU-MVLP 上进行了详尽的实验。OU-MVLP 数据集以其庞大的数据量和多视角覆盖的特点，对模型的泛化能力提出了极高要求。实验结果如表 2 所示，对比方法包括了基线模型 GaitSet、部位划分模型 GaitPart、以及 GaitAMR[23] 和文献[22]方法。

在整体性能上，本文的模型取得了 89.5% 的平均准确率，全面超越了所有对比方法，包括 GaitSet(87.1%)、GaitPart(88.7%)和 GaitAMR(88.3%)。相较于基线 GaitSet，提升了 2.4%，证明了本文架构在处理大规模、多

表 2 不同方法在 OU-MVLP 数据集上的性能对比

Table 2 Performance comparison of different methods on OU-MVLP dataset %

| 模型 | 准确率 | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0° | 15° | 30° | 45° | 60° | 75° | 90° | 180° | 195° | 210° | 225° | 240° | 255° | 270° | Mean |
| 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 |
| 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 |
| 文献[22] | 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 |
| GaitAMR | **84.1** | 89.0 | 89.7 | 90.0 | 88.7 | 89.2 | 88.2 | 86.9 | 88.4 | 88.7 | 88.9 | 87.2 | 89.1 | 87.7 | 88.3 |
| 本文 | 83.9 | **90.0** | **90.8** | 90.3 | **90.1** | **90.4** | **90.3** | **88.9** | **89.5** | **90.2** | **91.0** | **90.3** | **90.6** | **89.3** | **89.5** |

视角数据时更强的泛化能力。

更重要的是,模型在视角差异巨大的极限条件下表现尤为出色:在正对面的 180°视角,模型准确率达到 88.0%,高于 GaitSet 7.3%,并优于其他所有方法;在 90°和 270°等挑战性视角,模型同样保持了领先的性能。

在几乎所有视角下的一致性领先,强有力地证明了本文提出的结构感知与动态协同策略,不仅能抵抗外观变化,还能学到一种更根本的、视角不变的步态表征。

### 2.3 消融实验分析

1)结构感知路径权重系数

为了验证本文为 SAP 模块设计的非归一化、高对比度权重策略的有效性,并确定最优的权重参数,本文在 CASIA-B 数据集上进行了一组对比实验,针对头部($\lambda_H$)、躯干($\lambda_T$)、下肢($\lambda_L$)不同权重组合的识别准确率对比结果如表 3 所示。

表 3 的结果清晰地揭示了不同权重系数对模型性能的影响。以 $\lambda_H = \lambda_T = \lambda_L = 1$ 为基准,其平均准确率为 85.5%。实验表明,当逐步增加躯干($\lambda_T$)和下肢($\lambda_L$)的权重时,模型性能稳步提升,最高可达 88.3%,这初步验证了放大关键运动部位特征信号的有效性,与下肢最具判别力的生物力学先验相吻合。然而,单纯抑制非关键部位(头部)权重并非最优解,当 $\lambda_H$ 过度降低至 0.1 而未相应提升其他部位权重时,性能反而下降至本次实验的最低点 87.7%。这揭示了最佳性能源于协同优化:最终,通过抑制头部权重($\lambda_H=0.2$)并显著放大躯干与下肢的特征信号($\lambda_T=3$,$\lambda_L=5$),模型性能达到了峰值 89.9%。尤其在最具挑战性的换装(CL)条件下,准确率从基准的 73.9%大幅跃升至 81.4%,强有力地证明了该系数组在抵抗外观干扰方面的优越性。因此,本组详尽的对比实验验证本文所提出的高对比度加权策略的有效性,并确定($\lambda_H=0.2$,$\lambda_T=3$,$\lambda_L=5$)为模型的最终配置。

表 3 不同结构感知路径权重系数实验结果

Table 3 Results of weight coefficients for perception paths with different structures %

| $\lambda_H$ | $\lambda_T$ | $\lambda_L$ | NM | BG | CL | Mean |
|---|---|---|---|---|---|---|
| 1.0 | 1 | 1 | 95.0 | 87.7 | 73.9 | 85.5 |
| 1.0 | 2 | 2 | 95.4 | 88.1 | 72.9 | 85.5 |
| 1.0 | 3 | 3 | 95.8 | 89.3 | 75.1 | 86.7 |
| 1.0 | 3 | 5 | 96.3 | 90.6 | 78.0 | 88.3 |
| 1.0 | 5 | 7 | 96.1 | 90.2 | 77.4 | 87.9 |
| 0.5 | 3 | 3 | 96.0 | 89.1 | 73.4 | 81.2 |
| 0.5 | 3 | 5 | 96.2 | 90.0 | 77.2 | 87.8 |
| 0.3 | 3 | 5 | 96.4 | 90.9 | 78.6 | 88.6 |
| 0.2 | 3 | 3 | 96.7 | 91.6 | 80.2 | 89.5 |
| 0.2 | 3 | 5 | **96.9** | **92.2** | **81.4** | **89.9** |
| 0.1 | 3 | 5 | 95.8 | 89.3 | 78.2 | 87.7 |

2)核心模块贡献度

为了定量地剖析本文所提出的各个创新模块的独立贡献与协同效应,并深入理解性能提升背后的内在机理,本文以 GaitSet 为基线,设计了一系列详尽的消融实验。如表 4 所示,本文逐步引入 SAP、FD-Resnet 以及 DSFM,并对模型的性能演进进行了系统性的评估。

表 4 消融实验结果分析

Table 4 Analysis of ablation experiment results %

| 实验序号 | Baseline | SAP | FD-Resnet | DSFM | NM | BG | CL | Mean |
|---|---|---|---|---|---|---|---|---|
| 1 | √ | | | | 95.0 | 87.2 | 70.4 | 84.2 |
| 2 | √ | | √ | | 95.3 | 88.2 | 72.1 | 85.2 |
| 3 | √ | √ | | | 95.8 | 89.7 | 79.0 | 88.2 |
| 4 | √ | √ | √ | | 96.6 | 90.7 | 80.5 | 89.3 |
| 5 | √ | √ | | √ | 95.8 | 91.1 | 81.0 | 89.3 |
| 6 | √ | √ | √ | √ | **96.9** | **92.2** | **81.4** | **89.9** |

图 6 以折线图的形式直观展示了各模块引入后，模型在 NM、BG、CL 这 3 种场景以及平均性能上的提升趋势，使本文能更清晰地观察到每个组件对最终性能的累积贡献和在不同场景下的表现差异。

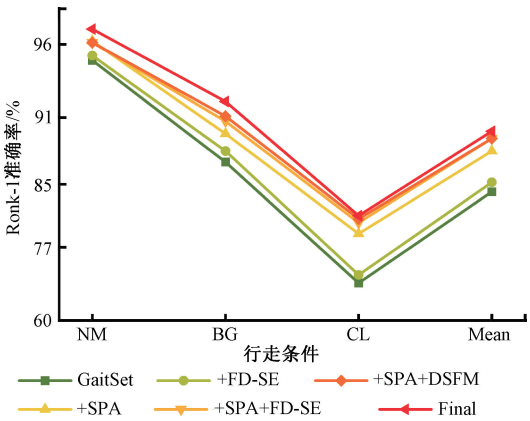



图 6　消融实验结果分析

Fig. 6　Analysis of ablation experiment results

实验清晰地揭示了各组件的累积效应。首先，在 GaitSet 基线平均 84.2% 上，实验对比了两条独立的改进路径：单独引入 FD-Resnet 或单独引入 SAP。实验结果显示，单独引入 FD-Resnet 能将平均准确率提升 1.0%，达到 85.2%，这表明动态注意力机制本身能够通过提纯特征带来一定的性能增益。然而，一个更为显著的提升来自于单独引入 SAP，它带来了 4.0% 的较大提升，使平均准确率达到 88.2%。该提升在最具挑战性的 CL 场景尤为显著，从 70.4% 跃升至 79.0%，这充分证明了相比于单纯的特征提纯，结构化建模在抵抗外观干扰中扮演着更为关键的基础性作用。在此基础上，无论是单独引入 FD-ResNet 还是 DSFM，均能带来约 1.1% 的额外增益，这表明 FD-ResNet 的动态注意力能有效提纯特征，而 DSFM 的双重注意力机制则能高效融合多源信息。最终，当所有模块协同工作时，模型达到了 89.9% 的最优平均性能，这证明了各模块间存在“先提纯，后融合”的良好协同效应，共同构建了一个高效且稳健的识别系统。

### 2.4　步态特征降维对比

为了直观评估模型学习到的步态特征的判别能力和聚类效果，本文采用了 T-SNE（t-distributed stochastic neighbor embedding）降维技术，将高维步态特征投影到二维空间进行可视化分析，如图 7 所示。

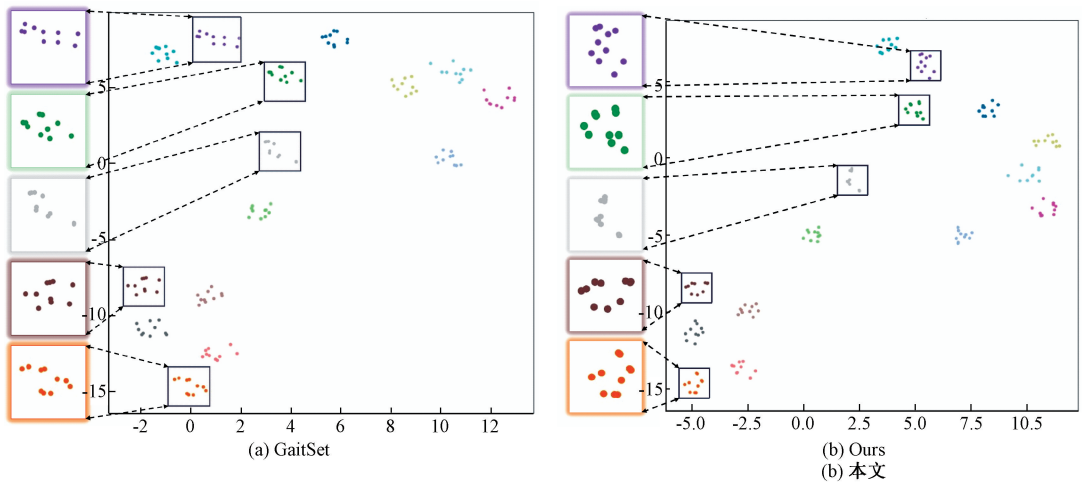



图 7　模型提取的步态特征 T-SNE 降维可视化

Fig. 7　T-SNE dimensionality reduction visualization of gait features extracted by the model

本文从 OU-MVLP 测试集中随机选取 15 名不同个体，每名个体随机抽取 10 个步态序列进行特征提取。图 7(a) 为基线模型 GaitSet 的特征可视化结果，图 7(b) 为本文模型的结果。图中每个点代表一个步态序列，不同颜色对应不同个体身份。

通过对比可以发现，本模型学习到的步态特征在二维空间中形成了更加紧密和边界清晰的簇群。通过局部放大区域的对比，可以清晰地看到：同色点（同一身份的步态特征）的聚类更加集中，类内紧凑性显著增强。这表明本模型能够更有效地辨别不同个体之间的细微步态差异。

## 3　结　　论

针对基于轮廓的步态识别方法在服装变化、背包遮挡等强外观干扰条件下性能显著下降的问题，本研究在保持 2D 轮廓框架部署优势的前提下，提出了一种融合结构感知与动态注意力机制的步态识别模型。该模型通过并行引入结构感知路径，对人体关键部位进行显式建模，并结合动态注意力与多层级语义融合机制，有效强化了稳定运动结构特征的表达能力，缓解了外观变化对识别性能的影响。

在 CASIA-B 和 OU-MVLP 两个主流数据集上的实验

结果表明,所提出的方法在多种行走条件和复杂视角变化下均取得了优于现有主流方法的识别性能,尤其在换装这种高干扰场景中表现出优势。消融实验进一步验证了结构感知建模、动态注意力机制及语义融合策略在提升步态特征判别性方面的有效性。

尽管如此,本文采用的结构划分仍基于固定规则,对不同体型和姿态的自适应能力有限;同时,模型在真实复杂环境下的泛化能力与计算效率仍有进一步优化空间。未来研究可从自适应结构建模、时序动态建模及模型轻量化等方向展开,以推动该方法在实际步态识别应用中的进一步发展。

## 参考文献


[1] YANG W CH, WANG S, HU J K, et al. Security and accuracy of fingerprint-based biometrics: A review[J]. Symmetry, 2019, 11(2): 141.

[2] ADJABI I, OUAHABI A, BENZAOUI A, et al. Past, present and future of face recognition: A review[J]. Electronics, 2020, 9(8): 1188.

[3] 史晓国,云静,张钰莹,等. 步态识别研究综述[J]. 计算机工程与应用, 2025, 61(14): 65-87.
SHI X G, YUN J, ZHANG Y Y, et al. Review of gait recognition research[J]. Computer Engineering and Applications, 2025, 61(14): 65-87.

[4] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6):84-90.

[5] WU Z F, HUANG Y ZH, WANG L, et al. A comprehensive study on cross-view gait based human identification with deep CNNs[J]. IEEE Transaction Pattern Analysis Machine Intelligence, 2017, 39(2): 209-226.

[6] CHAO H Q, HE Y W, ZHANG J P, et al. GaitSet: Regarding gait as a set for cross-view gait recognition[J]. ArXiv preprint arXiv:1811.06186,2019.

[7] FAN CH, PENG Y J, CAO CH SH, et al. GaitPart: Temporal part-based model for gait recognition[C]. IEEE/CVF Conference Computer Vision Pattern Recognition(CVPR), 2020: 14213-14221.

[8] 赵坚, 应娜, 舒勤. 基于全局与局部特征融合的跨视角步态识别算法研究[J]. 杭州电子科技大学学报(自然科学版), 2025, 45(4): 69-77.
ZHAO J, YING N, SHU Q. Research on cross-view gait recognition algorithm based on fusion of global and local features[J]. Journal of Hangzhou Dianzi University(Natural Science), 2025, 45(4): 69-77.

[9] DENG M Q, ZOU Y, ZENG ZH, et al. Cross-view gait recognition based on discriminative global-local feature representation and learning[J]. Engineering Applications of Artificial Intelligence, 2025, 161: 112282.

[10] LI H K, QIU Y D, ZHAO H M, et al. GaitBranch: A multi-branch refinement model combined with frame-channel attention mechanism for gait recognition[J]. Computer Vision and Image Understanding, 2025, 260: 104463.

[11] 吴越,梁铮,高巍,等. 基于 SMPL 模态分解与嵌入融合的多模态步态识别[J]. 浙江大学学报(工学版), 2026,60(1): 52-60.
WU Y, LIANG ZH, GAO W, et al. Multi-modal gait recognition based on SMPL modal decomposition and embedding fusion[J]. Journal of Zhejiang University (Engineering Science), 2026,60(1): 52-60.

[12] 关迪元,张鑫宇,王增烨,等. 融合点云步态模型与深度学习的步态识别算法[J]. 计算机工程与设计, 2025, 46(8): 2312-2319.
GUAN D Y, ZHANG X Y, WANG Z Y, et al. Gait recognition algorithm fusing point cloud gait model and deep learning[J]. Computer Engineering and Design, 2025, 46(8): 2312-2319.

[13] 韩亚丽, 韩子瑒, 金壮壮, 等. 一种主动型踝关节助力外骨骼设计及性能实验[J]. 仪器仪表学报, 2023, 44(11): 109-118.
HAN Y L, HAN Z Y, JIN ZH ZH, et al. Design and performance experiment of an active ankle power-assisted exoskeleton[J]. Chinese Journal of Scientific Instrument, 2023, 44(11): 109-118.

[14] 刘彦伟, 李博文, 胡重阳, 等. 仿壁虎跨尺度粘附结构状态感知方法研究[J]. 仪器仪表学报, 2025, 46(10):318-330.
LIU Y W, LI B W, HU CH Y, et al. Adhesion state sensing method for a gecko-inspired multiscale gripper[J]. Chinese Journal of Scientific Instrument, 2025,46(10): 318-330.

[15] GHAROUN H, MOMENIFAR F, CHEN F, et al. Meta-learning approaches for few-shot learning: A survey of recent advances[J]. ACM Computing Surveys, 2024, 56(12):1-41.

[16] DE BRABANDERE B, JIA X, TUYTELAARS T, et al. Dynamic filter networks[J]. ArXiv preprint arXiv:1605.09673,2016.

[17] 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 Technology, 2024, 47(1): 186-192.

[18] 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 Computer Vision Applications, 2018, 10(1), DOI:10.1186/s41074-018-0039-6.

[19] CHAO H Q, WANG K, HE Y W, et al. GaitSet: Cross-view gait recognition through utilizing gait as a deep set[J]. IEEE Transactions Pattern Analysis Mach ine Intelligence, 2022, 44(7): 3467-3478.

[20] TEEPE T, KHAN A, GLIG J, et al. GaitGraph: Graph convolutional network for skeleton-based gait recognition[J]. ArXiv preprint arXiv:2101.11228,2021.

[21] XU J, LI H, HOU SH J. Attention-based gait recognition network with novel partial representation PGOFI based on prior motion information[J]. Digit Signal Processing, 2023, 133: 103845.

[22] 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. Improved GaitSet method for gait recognition via fusion of silhouette enhancement and attention mechanism[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(1): 203-210.

[23] CHEN J Y, WANG ZH Y, ZHENG C X, et al. GaitAMR: Cross-view gait recognition via aggregated multi-feature representation[J]. Information Sciences, 2023, 636: 118920.

## 作者简介

**翟奥北**,硕士研究生,主要研究方向为模式识别、步态识别。

E-mail:zhaibei713@gs.zzu.edu.cn

**奚昊**,博士研究生,主要研究方向为模式识别、步态识别。

E-mail:hxi@gs.zzu.edu.cn

**范金豪**,博士研究生,主要研究方向为模式识别、场景识别。

E-mail:jhfan@gs.zzu.edu.cn

**胡传平**(通信作者),教授,博士生导师,主要研究方向为公共安全、网络空间安全、模式识别。

E-mail:cphu@zzu.edu.cn