

DOI: 10.13382/j.jemi.B2306788

融合轮廓增强和注意力机制的改进 GaitSet 步态识别方法*

陈万志¹ 唐浩博¹ 王天元²

(1. 辽宁工程技术大学软件学院 葫芦岛 125105; 2. 国网辽宁省电力有限公司营口供电公司 营口 115005)

摘要:针对传统基于轮廓的步态识别方法受限于输入特征及模型特征提取的能力,从而导致识别准确率不高的问题,提出一种融合轮廓增强和注意力机制的改进 GaitSet 步态识别方法。首先通过预处理获取行人的轮廓图,求得其均值,合成步态 GEI 能量图,将其作为神经网络模型的输入特征,增强了人体外观的表示。其次在提取特征的过程中引入注意力机制,增强模型的特征提取能力,从而提高步态识别的精度。最后在 CASIA-B 和 OU-MVLP 数据集上进行实验,所提方法的平均 Rank-1 准确率分别为 87.7% 和 88.1%。特别是在最复杂的穿大衣行走条件下,相较于 GaitSetv2 算法,准确率提升了 6.7%,表明所提出方法具有更强的准确性。此外,所提方法几乎没有增加额外的参数量、计算复杂度和推理时间,说明其各模块的快速性。

关键词: 步态识别;交叉视角;深度学习;轮廓增强;注意力机制

中图分类号: TP18 **文献标识码:** B **国家标准学科分类代码:** D510.4

Improved GaitSet method for gait recognition via fusion of silhouette enhancement and attention mechanism

Chen Wanzhi¹ Tang Haobo¹ Wang Tianyuan²

(1. College of Software, Liaoning Technical University, Huludao 125105, China; 2. State Grid Yingkou Electric Power Company of Liaoning Electric Power Supply Co., Ltd., Yingkou 115005, China)

Abstract: Aiming at the problem that traditional gait recognition methods based on silhouette are limited by the ability to extract input features and model features, which leads to low recognition accuracy, an improved GaitSet method for gait recognition via fusion of silhouette enhancement and attention mechanism is proposed. Firstly, the outline of the pedestrian is obtained by preprocessing, and its average value is obtained. Then the GEI energy diagram is synthesized, which is used as the input feature of the neural network model to enhance the representation of human appearance. Secondly, the attention mechanism is introduced in the process of feature extraction to enhance the feature extraction ability of the model, so as to improve the accuracy of gait recognition. Finally, experiments are carried out on the CASIA-B and OU-MVLP benchmark data sets, and the average Rank-1 accuracy of the proposed method is 87.7% and 88.1%, respectively. Especially under the most complex walking conditions with overcoat, compared with GaitSetv2 algorithm, the accuracy is improved by 6.7%, indicating that the proposed method has stronger accuracy. Notably, the proposed innovative method adds almost no additional parameter number, computational complexity, and inference time, which proves the rapidity of its individual modules.

Keywords: gait recognition; cross-view; deep learning; silhouette enhancement; attention mechanism

0 引言

步态识别作为一种新兴的生物识别技术,具有广泛的应用前景和经济价值,涉及安全监控、人机交互、医疗

诊断等领域^[1-2]。近年来,随着深度学习技术的不断发展,卷积神经网络(convolutional neural network, CNN)已经被广泛应用于计算机视觉领域,在分类任务中取得了显著的成果。

Shiraga 等^[3]在步态识别领域中第 1 次将卷积神经网络

络应用于该问题,并提出了步态能量图网络(gait energy image net, GEINet),使用步态能量图(gait energy image, GEI)作为输入,最终通过 Softmax 进行分类。随后,Wu 等^[4-5]第 1 次将深度卷积神经网络引入步态识别领域。在此基础上,Wang 等^[6]通过叠加多层自动编码器的方式,把在任意视角拍摄的视频转为固定的视角,提取视角不变的步态特征。

近年来,一些研究者提出了基于人体轮廓的步态识别方法,即直接从步态轮廓序列中提取特征^[7-9]。Chao 等^[7,9]将步态视为由独立帧组成的集合,并采用灵活的方式整合来自同一个人在不同场景下的视频。相较于先前的方法,例如基于模型的方法^[10-11]和基于模板的方法^[12-13],在交叉视角和各种行走条件下的识别准确率上有了较大幅度的提升。Ma 等^[14]基于 GaitSet 模型,提出一种改进的步态识别方法,将人体姿态和人体轮廓进行融合,并引入了人体姿态关键点,从而提高了人体轮廓提取的精度。注意力机制最早由 Hu 等^[15]提出,旨在提高网络对于重要特征的关注度,进而提高网络的性能。随后,Fan 等^[8]提出把人体划分为部件特征的方法,并将注意力机制引入其中,取得了较大幅度的性能提升。

然而,这些基于人体步态轮廓图的方法,受限于输入特征及模型特征提取的能力,并不能很好的表示一个人在行走运动时产生的信息。因此,本文提出一种改进 GaitSet 模型的步态识别方法,其主要创新点主要体现如下 3 个方面:

1) 通过预处理获取行人的不同轮廓图,求得其均值,进而合成步态序列融合 GEI 能量图,共同作为神经网络模型的输入特征,增强了人体外观的表示,有利于提升识别的准确率。

2) 在提取特征的过程中引入基于 GEI 的注意力机制,增强模型的特征提取能力,从而提高模型的性能。

3) 提出的基于轮廓和 GEI 融合的改进 GaitSet 模型步态识别方法,在 CASIA-B^[12]和 OUMVLP^[16]基准步态数据集上的实验表明本模型的有效性与快速性。

1 相关工作

1.1 基于模型的方法

基于模型的步态识别方法是一种使用模型对步态进行建模和识别的技术。该方法将步态视为一种动态系统,并描述其运动规律,以提取步态特征并进行识别。在步态识别领域中,基于模型的方法采用各种模型,如对抗生成网络 GANs^[17]的方法。这些方法^[6,13,18-19]可以利用 GAN 模型去除受试者携带的物品,并生成正常服装与视角下步态的特定侧视图(90°)。如 Yu 等^[19]通过多任务训练优化 GANs 模型,生成周期能量图(PEI 模板)。Xia

等^[20]提出一种基于 GANs 的步态识别算法。该算法的核心在于解决步态识别过程中小物品遮挡问题,其利用对抗生成网络生成上下一致的补全图像,从而有效减小了噪声干扰。Teepe 等^[21]直接从 RGB 图像中估计鲁棒的骨骼姿态,从而使基于模型的步态识别具有更清晰的步态表示。尽管这些方法通过视角变换在一定程度上解决了交叉视角识别的问题,但这种操作可能会导致丢失精细的信息,并且需要耗费高昂的计算成本。

1.2 基于外观的方法

利用受试者行走时的外观信息进行步态识别是一种基于外观特征的方法,通常使用视频或图像序列,如衣物颜色、纹理、轮廓等,对步态进行建模和识别。Han 等^[12]提出一种高效的 GEI 模板,该模板很好地表现出步态周期的形态和速度等特点,较基于模板的方法^[22-23]具有更高的准确率。Murakami 等^[22]和 Chen 等^[24]分别提出 GHI 模板和 CGI 模板,从时间维度上对 GEI 模板进行改进,以更好地保留时序信息。Wu 等^[4-5]提出使用深度卷积神经网络从步态模板中提取特征。胡靖雯等^[25]提出四元组损失,以优化深度神经网络的损失函数。而 Chao 等^[7,9]则提出了一种去除步态序列顺序限制的方法,将步态视为由独立帧组成的集合,并灵活整合来自同一人在不同场景下的视频。最近,Hu 等^[15]提出一种注意力机制网络(squeeze-and-excitation network, SENet),旨在提高网络对重要特征的关注度,进而提高网络的性能。SENet 注意力机制通过对每个通道的特征图进行加权,使得网络更加关注重要的特征。随后,Fan 等^[8]提出把人体划分为部件特征的方法,并将注意力机制引入其中,取得了较大幅度的性能提升。但由于输入特征和模型特征提取能力的限制,这些基于外观的步态识别方法在实际应用中仍存在一定问题。

2 本文方法

2.1 网络框架体系

传统的步态识别方法主要依赖于单一特征表示,这在一定程度上限制了识别性能。本文提出一种人体轮廓与能量图融合的步态识别方法,人体轮廓特征可以有效地捕捉到人体的外形信息,而能量图特征则可以从运动角度描述人体的步态特点。通过将这两种特征融合,可以更好地表征人体步态的空间和运动信息,为后续的识别任务提供更为丰富和可靠的特征表达。该方法基于 GaitSet 模型^[7],通过融合步态 GEI 能量图来增强步态轮廓图的表示能力。同时,引入了 GEI 通道注意力机制来学习不同特征的重要性,从而提高模型的性能,如图 1 所示,其中卷积层由上对角线图案表示,池化层由黑色图案

表示,注意力机制由点图案表示,而将水平池化为条带的操作则由菱形图案表示^[7], f, c, h, w 分别代表帧数、通道数、高度和宽度, p 代表条带数量。具体而言,本文提出方法通过增加步态 GEI 能量图和步态序列作为输入,实现了对不同特征的综合学习和突显重要特征,从而提升了模型的鉴别性。

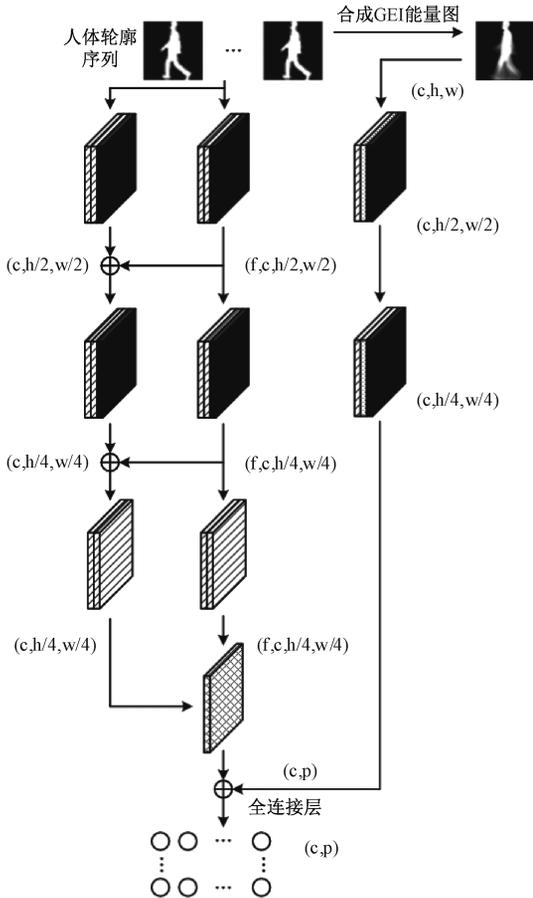


图 1 本文方法的网络框架

Fig. 1 Framework of the proposed method

2.2 人体轮廓图融合 GEI 能量图

步态能量图是一种将步态轮廓序列中的动态信息和静态信息混合的运动模板。它能够计算一个周期内的步态轮廓图在时间序列上的平均值,从而完整地保存边缘、时间、频率等信息。因此,获取行人步态能量图对于识别行人身份具有重要意义。此外,步态能量图还可以用于识别行人的情绪状态,从而更好地了解行人的行为特征。

步态能量图可很好地表达步态的各种特征,在步态识别研究中被广泛使用,如人体运动时的速度和动作姿态等,如图 2 所示。本文步态 GEI 能量图的合成方法是对步态序列进行加权平均,既简单又方便高效。

具体而言,步态 GEI 能量图可表示为:



图 2 步态能量图的合成方式

Fig. 2 Synthetic method of gait energy diagram

$$G(x, y) = \frac{1}{T} \sum_{t=1}^T B_t(x, y) \quad (1)$$

其中, $G(x, y)$ 代表步态 GEI 能量图, $B_t(x, y)$ 代表图像在 t 时刻的像素坐标, T 代表完整的步态周期的帧数。

所采用的步态序列融合步态 GEI 能量图的步态识别方法处理过程可表示为:

$$I = S(X) + G(X) \quad (2)$$

其中,步态轮廓序列 $S(X) = \{x_t | t = 1, 2, \dots, T\}$, X 代表输入序列, T 代表此序列的总帧数, x_t 表示第 t 帧;步态 GEI 能量图 $G(X) = \text{Mean}(S(X))$,即对步态序列求得其平均值,作为 GEI 能量图。

2.3 基于 GEI 的注意力机制

注意力机制^[15]是一种深度学习技术,可以帮助模型更好地理解输入数据,并且可以更好地捕捉输入数据之间的复杂关系。通过计算输入数据中的重要特征,并将这些特征的权重赋予模型,从而使模型能够更好地理解输入数据之间的复杂关系,提高模型的有效性和性能。本文针对 GaitSet 模型提出了一种基于 GEI 能量图的注意力机制,整体结构如图 3 所示,其中上对角线图案代表卷积层,竖线和菱形图案分别代表全局平均池化层和全局最大池化层。该注意力机制模块的输入为 2D 卷积层所提取的特征图,输出为通道重要性选择后的特征图。

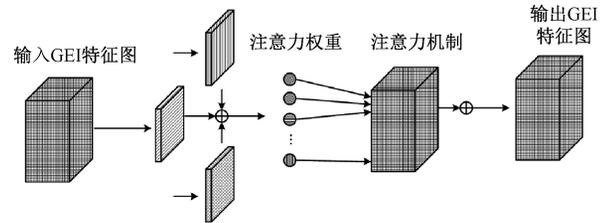


图 3 GEI 注意力机制结构

Fig. 3 GEI attention mechanism structure diagram

首先对输入的特征图进行全局平均池化和全局最大池化,然后将两个池化后的特征进行堆叠,并使用两层全连接层进行深层特征提取,最后求和并经过 sigmoid 函数进行归一化,得到注意力权重。通过将注意力权重与输入的特征图进行像素级相乘,可以得到最终的输出。因此,该通道注意力机制可以有效地提取 GEI 特征图中的重要信息,并忽略不重要的特征图,从而提高模型的性能。其可表示为:

$$Y' = \sigma(\text{Sum}(\text{MLP}([\text{Avg}(Y), \text{Max}(Y)]))) \times Y \quad (3)$$

其中,输入 Y 是 2D 卷积层的输出。首先对 Y 进行全局平均池化和全局最大池化,然后将这两个池化后的特征堆叠起来,使用多层感知机 (multilayer perceptron, MLP) 进行深度特征提取,MLP 具体包括两层 FC 全连接层,使用 ReLU 作为激活函数。将 MLP 的输出与 Y 进行求和,再通过 sigmoid 函数进行归一化,得到注意力权重。最后,将注意力权重和输入 Y 进行像素级相乘,得到最终的输出特征 Y' 。因此,该通道注意力机制可以有效地提取特征图中的重要信息,并忽略不重要的特征图,从而提高模型的性能。

3 实验与分析

采用 CASIA-B^[12] 和 OUMVLP^[16] 基准步态数据集作为实验数据对所提出方法进行验证和性能评估,实验硬件环境为 i9-12900K CPU + RTX 3080Ti GPU (显存 12 GB),软件环境为 64 位 Windows11 专业版操作系统,采用 Pytorch 深度学习框架和 Python 语言编程。

3.1 数据集和训练设置

1) CASIA-B 数据集

CASIA-B 中包含 124 名受试者,从 0~180° 拍摄了 11 个角度,并且在每个角度下拍摄了 3 种不同的行走状态,分别为普通状态 (NM)、背着包 (BG) 和穿着外套 (CL)。CASIA-B 步态数据集并没有明确规定训练集和测试集。实验采用常见的大样本训练 (LT)、中样本训练 (MT) 和小样本训练 (ST) 3 种设置,具体划分如表 1 所示。

表 1 训练设置

Table 1 Settings in Training

数据集	ID	步态序列
训练集	LT: 001-074	
	MT: 001-062	NM#1-4
	ST: 001-024	
验证集 (Gallery)	LT: 075-124	
	MT: 063-124	NM#1-4
	ST: 025-124	
测试集 (Probe)	LT: 075-124	NM#5-6
	MT: 063-124	BG#1-2
	ST: 025-124	CL#1-2

在测试阶段,每个视角下的 NM 条件的前 4 个序列 (NM#1-4) 被存储在 Gallery 集中,其余 6 个序列被分成 3 个 Probe 子集,分别为 NM 子集 (NM#5-6)、背着包 (BG) 子集 (BG#1-2) 和穿着外套 (CL) 子集 (CL#1-2)。

2) OUMVLP 数据集

OUMVLP 数据集是另一个著名的步态识别基准数

据集,也是最大的公开数据集之一。该数据集提供了 10 307 个独立的个体主体,每个主体以 14 个不同角度 (0°, 15°, ..., 90°, 180°, 195°, ..., 270°) 进行观测。每个角度下都包含两个序列,分别标记为 #00 和 #01。数据集总共使用了 5 153 个主体进行训练,剩余的 5 154 个主体用于测试。在测试过程中,序列索引为 #00 和 #01 的分别被用作 Probe 集和 Gallery 集。

3) 训练和测试

训练:在训练阶段,从训练集中抽取一批大小为 $p \times k$ 的样本,其中 p 代表个体的数量, k 是该批中每个个体拥有的训练样本的数量。需要注意,尽管实验已经表明,当输入是从任何序列收集的一组剪切片段时,模型的表现良好,但是在训练阶段,样本仅由从单个序列采样的剪切片段组成。

测试:给定一个查询集 Q ,目标是检索所有集相同身份的画廊集 G 。首先,将 Q 输入模型,产生步态特征。然后,生成特定的特征向量 F_Q 。类似地,对每个 G 应用相同的操作来得到 F_G 。最后,将 F_Q 与每个 F_G 进行欧氏距离的比较,计算出 Rank-1 的识别精度,即排名第 1 的正确受试者的百分比。

3.2 模型参数

为了提高训练效率,将输入的批次 $p \times k$ 设置为 $p = 8, k = 8$,每个批次的帧数 $f = 30$,即每个批次处理 8 个样本,每个样本包含 30 帧数据。在预处理过程中,将步态轮廓图对齐为 64×44 ,并进行归一化后输入到网络中进行训练^[16]。模型主干的参数设置如表 2 所示。

表 2 模型主干结构的参数设置

Table 2 Parameter setting of model backbone structure

卷积层	通道数	核大小	步长
Conv	32	5	2
Conv	32	3	1
Conv	64	3	1
Conv	64	3	1
Conv	128	3	1
Conv	128	3	1

训练的参数和设置采用 Adam 作为优化器^[26],学习率设置为 1×10^{-4} ;训练和测试采样方式均为采用每个视频随机采样 30 帧。通过梯度下降和反向传播法,使用 batch all+ triplet loss 损失函数^[27]对所设计的网络模型进行优化。在 LT、MT 和 ST 的训练设置下分别迭代 120 000、120 000 和 100 000 轮。

3.3 CASIA-B 数据集上的步态识别

为了验证本文模型的有效性和优越性,在 CASIA-B 步态数据集上与现有的典型步态识别模型进行了性能比较。选择了先进的基于人体轮廓的步态识别模型进行对比。本文采用了 3.2 节中所述的网络参数设置方案,在

LT、MT 和 ST 的训练模式下进行了对比实验,并统计了 Rank-1 准确率作为评价指标。需要说明的是,本文的实验结果不包括对自身视角下的识别。本文模型在 NM、

BG 和 CL 行走条件下的 Rank-1 识别准确率与其他算法的对比结果如表 3 所示。除了本文的结果外,其他算法的结果均引用自于其所述文献。

表 3 在 CASIA-B 上的跨视角 Rank-1 准确率

Table 3 Cross-view Rank-1 accuracy on CASIA-B

(%)

训练设置	状态	模型	0°	18°	36°	54°	72°	90°	104°	126°	144°	162°	180°	平均
大样本训练 (LT)	NM	PoseGait ^[11]	55.3	69.6	73.9	75.0	68.0	68.2	71.1	72.9	76.1	70.4	55.4	68.7
		CNN-LB ^[4]	82.6	90.3	96.1	94.3	90.1	87.4	89.9	94.0	94.7	91.3	78.5	89.9
		GaitSet ^[7]	90.8	97.9	99.4	96.9	93.6	91.7	95.0	97.8	98.9	96.8	85.8	95.0
		GaitSev2 ^[9]	91.1	99.0	99.9	97.8	95.1	94.5	96.1	98.3	99.2	98.1	88.0	96.1
		GaitGraph ^[21]	85.3	88.5	91.0	92.5	87.2	86.5	88.4	89.2	87.9	85.9	81.9	87.7
		本文模型	94.0	98.4	99.0	98.1	92.9	91.5	95.5	98.6	99.4	98.6	91.3	96.1
	BG	PoseGait ^[11]	35.3	47.2	52.4	46.9	45.5	43.9	46.1	48.1	49.4	43.6	31.1	44.5
		CNN-LB ^[4]	64.2	80.6	82.7	76.9	64.8	63.1	68.0	76.9	82.2	75.4	61.3	72.4
		GaitSet ^[7]	83.8	91.2	91.8	88.8	83.3	81.0	84.1	90.0	92.2	94.4	79.0	87.2
		GaitSev2 ^[9]	86.7	94.2	95.7	93.4	88.9	85.5	89.0	91.7	94.5	95.9	83.3	90.8
		GaitGraph ^[21]	75.8	76.7	75.9	76.1	71.4	73.9	78.0	74.7	75.4	75.4	69.2	74.8
		本文模型	89.5	94.9	93.6	92.8	86.8	80.2	84.8	92.8	96.1	93.6	84.8	90.0
	CL	PoseGait ^[11]	24.3	29.7	41.3	38.8	38.2	38.5	41.6	44.9	42.2	33.4	22.5	36.0
		CNN-LB ^[4]	37.7	57.2	66.6	61.1	55.2	54.6	55.2	59.1	58.9	48.8	39.4	54.0
		GaitSet ^[7]	61.4	75.4	80.7	77.3	72.1	70.1	71.5	73.5	73.5	68.4	50.0	70.4
		GaitSev2 ^[9]	59.5	75.0	78.3	74.6	71.4	71.3	70.8	74.1	74.6	69.4	54.1	70.3
		GaitGraph ^[21]	69.6	66.1	68.8	67.2	64.5	62.0	69.5	65.6	65.7	66.1	64.3	66.3
		本文模型	70.8	83.7	86.3	80.9	75.3	69.4	76.4	78.6	80.9	79.1	65.1	77.0
中样本训练 (MT)	NM	MGAN ^[13]	54.9	65.9	72.1	74.8	71.1	65.7	70.0	75.6	76.2	68.6	53.8	68.1
		GaitSet ^[7]	86.8	95.2	98.0	94.5	91.5	89.1	91.1	95.0	97.4	93.7	80.2	92.0
		GaitSev2 ^[9]	89.7	97.9	98.3	97.4	92.5	90.4	93.4	97.0	98.9	95.9	86.6	94.4
		本文模型	91.2	98.0	98.9	97.9	92.6	90.0	93.7	98.1	98.4	96.0	87.1	94.7
	BG	MGAN ^[13]	48.5	58.5	59.7	58.0	53.7	49.8	54.0	61.3	59.5	55.9	43.1	54.7
		GaitSet ^[7]	79.9	89.8	91.2	86.7	81.6	76.7	81.0	88.2	90.3	88.5	73.0	84.3
		GaitSev2 ^[9]	79.9	89.8	91.2	86.7	81.6	76.7	81.0	88.2	90.3	88.5	73.0	84.3
		本文模型	84.6	89.8	92.7	91.0	85.5	79.1	84.5	92.1	94.4	91.4	78.6	87.6
	CL	MGAN ^[13]	23.1	34.5	36.3	33.3	32.9	32.7	34.2	37.6	33.7	26.7	21.0	31.5
		GaitSet ^[7]	52.0	66.0	72.8	69.3	63.1	61.2	63.5	66.5	67.5	60.0	45.9	62.5
		GaitSev2 ^[9]	52.0	66.0	72.8	69.3	63.1	61.2	63.5	66.5	67.5	60.0	45.9	62.5
		本文模型	61.2	75.8	78.6	76.2	68.8	64.6	68.0	73.8	73.4	70.1	56.3	69.7
小样本训练 (ST)	NM	CNN-LB ^[4]	54.8	—	—	77.8	—	64.9	—	76.1	—	—	—	68.4
		GaitSet ^[7]	64.6	83.3	90.4	86.5	80.2	75.5	80.3	86.0	87.1	81.4	59.6	79.5
		GaitSev2 ^[9]	71.6	87.7	92.6	89.1	82.4	80.3	84.4	89.0	89.8	82.9	66.6	83.3
		本文模型	72.9	85.2	90.8	90.3	82.2	78.5	82.4	89.2	90.8	86.8	70.9	83.6
		GaitSet ^[7]	55.8	70.5	76.9	75.5	69.7	63.4	68.0	75.8	76.2	70.7	52.5	68.6
		GaitSev2 ^[9]	64.1	76.4	81.4	82.4	77.2	71.8	75.4	80.8	81.2	75.7	59.4	75.1
		本文模型	63.4	73.6	78.8	77.0	72.2	65.3	70.3	76.6	80.4	76.7	60.6	72.3
		GaitSet ^[7]	29.4	43.1	49.5	48.7	42.3	40.3	44.9	47.4	43.0	35.7	25.6	40.9
GaitSev2 ^[9]	36.4	49.7	54.6	49.7	48.7	45.2	45.5	48.2	47.2	41.4	30.6	45.2		
本文模型	36.8	49.2	55.7	54.6	51.8	46.6	47.2	49.9	48.2	44.9	32.4	47.0		

考虑到不同文献中的训练设置可能存在差异,为了消除训练设置对结果的影响,本文仅与在相同设置条件下的模型进行了对比。由实验结果可知,所提出方法在不同训练设置下均取得较好的识别效果。在大样本训练下,相对于现有方法,所提出的方法在所有角度下均取得了更高的平均识别率,表明融合人体轮廓与 GEI 能量图

并对 GEI 做注意力机制的方案,具有更强的鲁棒性和准确性。在中、小样本训练下,所提出的方法在大多数角度下均超过了现有方法的识别效果,表明其能够更好地适应样本数量较少的情况。尽管在大样本和小样本的携带背包行走的条件下,准确率略低于 GaitSetv2,但在最复杂的穿大衣行走条件下,准确率大幅领先。例如在大样本

训练条件下,准确率提升了 6.7%。综上所述,所提出方法在不同训练设置下均取得了优异的识别效果,证明其有效性和实用性。

3.4 OUMVLP 数据集上的步态识别

为了验证所提方法的泛化性,在 OU-MVLP 数据集上进行了实验,并与目前先进的 GEINet、MvGGAN、GaitSet 和 GaitSetv2 方法进行对比,以进行公平比较。在 OU-MVLP 数据集中,10 307 个主体被分为训练数据和测试数据,其中 5 153 个主体用于训练,其余的主体用于测试。按照相同的实验设置,测试阶段将序列#00 作为 probe 集,序列#01 作为 gallery 集。实验结果如表 4 所示,除个别角度以外,本文方法都取得了最佳性能。

表 4 在 OUMVLP 上的跨视角 Rank-1 准确率

Table 4 Cross-view Rank-1 accuracy on OUMVLP (%)

Probe	Gallery All 14 Views				
	GEINet	MvGGAN	GaitSet	GaitSetv2	本文模型
0°	11.4	52.6	79.5	81.3	81.3
15°	29.1	62.8	87.9	88.6	90.0
30°	41.5	63.9	89.9	90.2	89.9
45°	45.5	57.5	90.2	90.7	90.5
60°	39.5	55.4	88.1	88.6	89.3
75°	41.8	61.3	88.7	89.1	89.2
90°	38.9	61.9	87.8	88.3	88.4
180°	14.9	54.8	81.7	83.1	83.5
195°	33.1	58.8	86.7	87.7	88.1
210°	43.2	59.3	89.0	89.4	89.3
225°	45.6	58.5	89.3	89.7	89.5
240°	39.4	56.6	87.2	87.8	88.6
255°	40.5	57.5	87.8	88.3	88.5
270°	36.3	56.8	86.2	86.9	86.8
Mean	35.8	58.4	87.1	87.9	88.1

3.5 消融实验

为了验证所提出的人体轮廓图融合 GEI 能量图及注意力机制的有效性,在 CASIA-B 基准数据集中进行消融实验。消融实验的结果为 Rank-1 平均准确率,不包含对自身视角下的识别。

1) 人体轮廓融合 GEI 的有效性

此部分消融实验结果对比如表 5 所示,本文提出的人体轮廓图融合 GEI 能量图可以显著提高步态识别的准确性。这是因为通过融合这两种信息,可以更准确地捕捉人体的运动特征和姿势信息,使得对人体运动的描述更加准确,从而提高识别的准确性。此外,人体轮廓图还能够提供更多的信息,如人体的姿态、动作等,这些信息能够被用于更加准确地描述人体行为。相比之下,GEI 能量图虽然也能够提供一定的信息,但是其对于背景和人体的分离不够明显,容易受到背景干扰,单独使用时影

响识别的准确性。因此,使用人体轮廓图融合 GEI 的方法相对更为优秀,能够更加准确地描述人体行为,提高步态识别的准确性。

表 5 人体轮廓融合 GEI 的消融实验结果

Table 5 Results of human body silhouette fusion GEI ablative experiment

实验	轮廓图	能量图	平均准确率/%
A	不使用	使用	41.7
B	使用	不使用	86.8
C	使用	使用	87.7

2) 基于 GEI 的注意力机制的有效性

此部分消融实验结果对比如表 6 所示,本文提出的基于 GEI 的注意力机制可以显著提高步态识别的准确性。这是因为通过引入注意力机制,可以更加准确地捕捉人体运动特征并抑制对于识别无用的特征,从而提高识别的准确性。由 B~E 可知,注意力机制的卷积核被设置为 3×3 大小时,获得最佳性能,当卷积核增大时,由于滑动窗口的增大导致信息可能会丢失,因此本文最终选择 3×3 尺寸的卷积核作为注意力机制的最终设定。

表 6 基于 GEI 的注意力机制消融实验结果

Table 6 Results of attention mechanism based on GEI ablation experiments

实验	卷积核	平均准确率/%
A	—	87.3
B	3×3	87.7
C	5×5	87.6
D	7×7	87.6
E	9×9	87.4

3) 计算速度分析

在实际应用中,计算速度尤为重要。由于现有方法并未具体给出计算速度的结果,因此本文只与自己的方法做计算速度分析。表 7 展示了本文所提方法的模型参数量、计算复杂度和推理时间结果。其中,方法 1 为从本文方法里,去除 GEI 分支;方法 2 为从本文方法里,去除 GEI 注意力机制模块。推理时间的计算方式是以 10 个包含 30 帧的视频作为输入,计算进行 10 次重复实验的推理时间,并取平均值。从表格中可以观察到,引入 GEI 分支和 GEI 注意力机制几乎没有增加额外的参数量、计算复杂度和推理时间,验证了所提方法及其各个模块的快速性。

表 7 计算速度分析

Table 7 Computational speed analysis

方法名称	参数量	计算复杂(GFlops)	推理时间/ms
方法 1	2 144 032	3.26	7.02
方法 2	2 149 072	3.27	7.13
本文方法	2 149 085	3.27	7.27

4 结 论

本文提出轮廓增强和注意力机制相融合的改进 GaitSet 模型的步态识别方法,预处理获取行人轮廓图并生成步态 GEI 能量图作为神经网络模型输入特征,同时引入 GEI 注意力机制增强特征提取能力,解决了现有方法的输入特征和模型特征提取限制。在 CASIA-B 基准数据集上,该方法在正常行走、背包行走和穿大衣行走条件下的 Rank-1 准确率均优于现有方法。实验结果表明,该方法具有更强的鲁棒性和准确性,可以更好地适应样本数量较少的情况。此外,引入 GEI 分支和 GEI 注意力机制,几乎没有引入额外参数量、计算复杂度和推理时间,证明所提方法及其各个模块的快速性。因此,所提出方法在实际应用中有一定的借鉴意义。未来工作中将进一步探究如何更好地利用深度学习技术进行步态识别,并将其应用于更广泛的领域,以实现更高效、准确的识别。

参考文献

[1] 陆兴华,蔡韬. 基于 CNN 的安防监控步态特征提取研究[J]. 计算机技术与发展, 2019, 29(11): 123-127.

LU X H, CAI T. Research on gait feature extraction in security monitoring system based on CNN[J]. Computer Technology and Development, 2019, 29(11): 123-127.

[2] 张红颖,田鹏华. 结合残差网络与多级分块结构的步态识别方法[J]. 电子测量与仪器学报, 2022, 36(6): 66-72.

ZHANG H Y, TIAN P H. Gait recognition method combining residual network and multi-level block structure[J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(6): 66-72.

[3] SHIRAGA K, MAKIHARA Y, MURAMATSU D, et al. GEINet: View-invariant gait recognition using a convolutional neural network [C]. 2016 International Conference on Biometrics. Halmstad, Sweden: IEEE, 2016: 1-8.

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

[5] WU Z, HUANG Y, WANG L. Learning representative deep features for image set analysis [J]. IEEE Transactions on Multimedia, 2015, 17 (11): 1960-1968.

[6] WANG Y, SONG C, HUANG Y, et al. Learning view invariant gait features with two-stream gan [J].

Neurocomputing, 2019, 339: 245-254.

[7] CHAO H, HE Y, ZHANG J, et al. GaitSet: Regarding gait as a set for cross-view gait recognition [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 8126-8133.

[8] FAN C, PENG Y, CAO C, et al. GaitPart: Temporal part-based model for gait recognition [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2020: 14213-14221.

[9] CHAO H, WANG K, HE Y, et al. GaitSet: Cross-view gait recognition through utilizing gait as a deep set[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, DOI: 10.1109/TPAMI. 2021. 3057 879.

[10] HONG C, YU J, TAO D, et al. Image-based three-dimensional human pose recovery by multiview locality-sensitive sparse retrieval [J]. IEEE Transactions on Industrial Electronics, 2014, 62(6): 3742-3751.

[11] 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.

[12] HAN J, BHANU B. Individual recognition using gait energy image[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(2): 316-322.

[13] 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, 2019, 14(1): 102-113.

[14] MA Y, WEI C, LONG H. A gait recognition method based on the combination of human body posture and human body contour[J]. Journal of Physics: Conference Series, 2020, 1631(1): 012031.

[15] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2018, DOI: 10.1109/ CVPR. 2018. 00745.

[16] 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): 4.

[17] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks; arXiv: 1406. 2661 [Z/OL]. arXiv, 2014 (2014-06-10). http://arxiv.org/abs/ 1406. 2661.

[18] YU S, CHEN H, REYES E B G, et al. GaitGAN: Invariant gait feature extraction using generative

- adversarial networks [C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu, HI, USA; IEEE, 2017; 532-539.
- [19] YU S, LIAO R, AN W, et al. GaitGANv2: Invariant gait feature extraction using generative adversarial networks[J]. Pattern Recognition, 2019, 87: 179-189.
- [20] XIA L, WANG H, GUO W. Gait recognition based on wasserstein generating adversarial image inpainting network[J]. Journal of Central South University, 2019, 26(10): 2759-2770.
- [21] TEEPE T, KHAN A, GILG J, et al. Gaitgraph: Graph convolutional network for skeleton-based gait recognition[C]. 2021 IEEE International Conference on Image Processing. Anchorage, AK, USA; IEEE, 2021; 2314-2318.
- [22] MURAKAMI M, TAN J K, KIM H, et al. Human motion recognition using directional motion history images [C]. ICCAS 2010, Gyeonggi-do, South Korea, 2010; 1445-1449.
- [23] SHUTLER J D. Statistical gait recognition via velocity moments [C]. IEE Colloquium on Visual Biometrics, 2000; 1-10.
- [24] CHEN W, JUNPING Z, LIANG W, et al. Human identification using temporal information preserving gait template[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11): 2164-2176.
- [25] 胡靖雯, 李晓坤, 陈虹旭, 等. 基于深度学习的步态识别方法[J]. 计算机应用, 2020, 40(S1): 69-73.
- HU J W, LI X K, CHEN H X, et al. Gait recognition method based on deep learning [J]. Journal of Computer Applications, 2020, 40(S1): 69-73.
- [26] KINGMA D P, BA J. Adam: A method for stochastic optimization; arXiv: 1412. 6980 [Z/OL]. arXiv, 2017 (2017 -01-29). <http://arxiv.org/abs/1412.6980>.
- [27] HERMANS A, BEYER L, LEIBE B. In defense of the triplet loss for person re-identification; arXiv: 1703. 07737 [Z/OL]. arXiv, 2017 (2017-11-21). <http://arxiv.org/abs/1703.07737>.

作者简介



陈万志(通信作者), 2015 年于辽宁工程技术大学(中国测绘科学研究院联合培养)获得博士学位, 现为辽宁工程技术大学副教授, 硕士生导师, 主要研究方向为人工智能与智能信息处理、网络与信息安全和工控软件与数据分析。

E-mail: chenwanzhi@lntu.edu.cn

Chen Wanzhi (Corresponding author) received his Ph. D. degree from Liaoning Technical University (China Academy of Surveying and Mapping Science Joint Cultivation) in 2015. Now he is an associate professor and master's degree supervisor in Liaoning Technical University. His main research interests include artificial intelligence and intelligent information processing, network and information security and industrial control software and data analytics.



唐浩博, 2021 年于辽宁工程技术大学获得学士学位, 现为辽宁工程技术大学硕士研究生, 主要研究方向为图像与视觉信息计算。

E-mail: 1779114938@qq.com

Tang Haobo received his B. Sc. degree from Liaoning Technical University in 2021. Now he is a M. Sc. candidate at Liaoning Technical University. His main research interests include network security and intrusion detection.