

中图法分类号: 文献标识码: 文章编号: 1006-8961(XXXX)XX-0001-16

论文引用格式: Liu Yanyan, Chen Xin, Liu Wenyu, Yi Daihao. Occluded Gait Restoration and Recognition via Visibility-Aware Spatio-Temporal Diffusion[J/OL]. Journal of Image and Graphics, XXXX:1-16. DOI: 10.11834/jig.250641. (刘艳艳, 陈欣, 刘文瑜, 易代豪. 可见性感知时空扩散的遮挡步态修复与识别[J/OL]. 中国图象图形学报, XXXX:1-16. DOI: 10.11834/jig.250641. ) [DOI:10.11834/jig.250641]

## 可见性感知时空扩散的遮挡步态修复与识别

刘艳艳, 陈欣\*, 刘文瑜, 易代豪

华南农业大学人工智能与低空技术学院, 广州 510642

**摘要:** 目的 现有的步态识别方法在理想环境下表现出色,但现实监控场景中物体遮挡常导致人体轮廓结构缺失以及传统生成式修复方法存在的盲目修复问题。为了解决这一挑战,提出一种可见性感知时空扩散的遮挡步态修复与识别方法(visibility-aware spatio-temporal diffusion recovery network, VAST-DRNet)旨在准确修复步态序列中被遮挡的区域,并提升其在复杂遮挡场景下的识别性能。**方法** 使用基于可见性感知和3D扩散模型的VAST-DRNet来修复步态序列,在修复每一帧步态图像的空间信息的同时保持帧与帧之间的时空连贯性。VAST-DRNet通过引入区域评分估计器(region scoring estimator, RSE)来检测步态图像的局部遮挡区域,并将其生成的可见性评分作为先验知识注入3D扩散修复模块(visibility-aware diffusion module, VADM)为遮挡区域分配更高的关注度,实现定向时空修复。同时,在识别阶段引入基于可见性空间注意力机制(visibility-driven spatial attention mechanism, V-SAM)的信心引导双分支融合机制,利用评分自适应融合原始遮挡特征与修复后的步态特征,从而优化识别网络的判别能力。**结果** 在CASIA-B数据集基础上构建的混合遮挡数据集OccCASIA-B上的实验表明,该方法在静态物体和人群干扰等复杂场景下Rank-1识别率分别达到83.1%和78.1%,相较于基准GaitGL方法提升了11.2%和7.6%。在混合遮挡强度评估中,在30%-50%的中度遮挡下识别率达到76.5%,在遮挡面积超过50%的极重度遮挡条件下,识别率达到50.6%,体现出在大面积信息丢失时较小的性能退化和更强的鲁棒性。**结论** 该方法能够有效感知遮挡区域并实现高质量的时空一致性修复,通过信心引导的特征融合机制,有效解决了生成模型存在的盲目修复问题,显著提升了模型在复杂现实遮挡环境下的鲁棒性与识别性能。

**关键词:** 步态识别;时空扩散模型;可见性感知;步态轮廓修复;三维卷积神经网络

## Occluded Gait Restoration and Recognition via Visibility-Aware Spatio-Temporal Diffusion

Liu Yanyan, Chen Xin\*, Liu Wenyu, Yi Daihao

College of Artificial Intelligence and Low-altitude Technology, South China Agricultural University, Guangzhou 510642, China

**Abstract: Objective** In real-world surveillance scenarios, the types and degrees of occlusion exhibit significant uncertainty, and different occlusion patterns damage the structural integrity and temporal continuity of gait silhouettes to varying degrees. Recognition methods that extract discriminative identity features solely from the visible regions of incomplete sequences inherently suffer from severe information limitations caused by heavy reliance on visible areas. In contrast,

收稿日期: 2025-12-23; 修回日期: 2026-06-12

\*通信作者: 陈欣 chenxin@scau.edu.cn

基金项目: 国家自然科学基金面上项目(62572200); 广东省自然科学基金面上项目(2024A1515011217)

**Supported by:** National Natural Science Foundation of China General Program(62572200); Guangdong Natural Science Foundation General Program(2024A1515011217)

although gait inpainting methods can partially restore damaged silhouettes, existing generative models are mostly restricted to frame-wise 2D inpainting and suffer from blind restoration issues, resulting in unnatural trajectory breaks and temporal jitter in reconstructed sequences. To address these problems, this chapter proposes an occlusion gait inpainting and recognition method based on visibility-aware spatial-temporal diffusion. It aims to guide a 3D diffusion model for targeted inpainting via visibility-aware occlusion perception, preserving spatial-temporal coherence between consecutive frames while restoring the spatial information of each gait frame, and better extracting the structural integrity and texture authenticity of gait silhouettes in the feature space, thereby recovering highly discriminative identity representations under complex occlusion conditions. **Method** A visibility-aware spatial-temporal diffusion network based on visibility perception and a 3D diffusion model is employed to restore gait sequences, maintaining spatial-temporal coherence between frames while repairing the spatial information of each gait frame. First, the occluded gait sequence is fed into a pre-trained and frozen region score estimator to perform fine-grained estimation of the visibility of each body region in each frame, yielding corresponding visibility scores. This score is then injected as a spatial guidance signal into the visibility-aware diffusion inpainting module to drive the network to capture spatial-temporal dependencies and conduct targeted spatial-temporal denoising and structural completion on occluded regions, generating inpainted sequences with complete structures and coherent dynamics. Next, the original occluded sequence and the inpainted sequence are fed into separate 3D convolutional branches for feature extraction, followed by adaptive fusion in an attention fusion module to comprehensively utilize the authentic identity information in the original sequence and the structural completion information in the inpainted sequence. Meanwhile, a visibility spatial attention mechanism is constructed based on visibility scores and applied to the backbone network via residual connections, further enhancing the model's attention to high-confidence human body regions. **Result** In this study, the OccCASIA-B benchmark dataset is generated and extended based on the existing OccSilGait framework, covering three occlusion patterns: static object occlusion, directional occlusion, and dynamic crowd occlusion. Static occluding objects are synthesized using real-world obstacles from the Mapillary-Vistas street view dataset. Directional occlusion simulates body part missing caused by camera view limitations or pedestrian detection box truncation. Crowd occlusion is simulated by dynamic interference from other pedestrians crossing the subject's path. Experiments on OccCASIA-B demonstrate that the proposed VAST-DRNet maintains a Rank-1 recognition accuracy of 50.6% under extremely severe occlusion with occlusion area exceeding 50%, outperforming the GaitGL model by 3.8%. In static object occlusion scenarios, the method achieves a Rank-1 accuracy of 83.1%, leading GaitGL by 11.2%. For dynamic crowd occlusion, the Rank-1 accuracy reaches 78.1%, surpassing GaitGL by 7.6%. In addition, the region score estimator achieves average MAE, RMSE, and coefficient of determination of 0.0305, 0.0543, and 0.8800 respectively on the visibility regression task, indicating its ability to accurately characterize the occlusion degree of different body parts. Quantitative evaluation of inpainting quality confirms the effectiveness of the 3D diffusion model approach. The introduction of RSE guidance improves the edge F1-score to 0.7021 and the structural similarity index (SSIM) to 88.3%, demonstrating that the model successfully recovers sharp and realistic silhouette boundaries and effectively alleviates the common "flickering" artifacts in frame-wise inpainting. **Conclusion** The proposed gait recognition method in this study effectively leverages the visibility-aware 3D diffusion process to achieve high-quality, spatially-temporally consistent targeted inpainting. By bridging occlusion perception and generative inpainting, the confidence-guided fusion mechanism effectively reduces the interference of generative noise. The significant performance improvement on the OccCASIA-B dataset highlights the robustness of the method and its potential for practical deployment in uncontrolled surveillance environments. Future work will explore latent space diffusion to reduce computational costs.

**Key words:** gait recognition; spatio-temporal diffusion model; visibility awareness; gait silhouette restoration; three-dimensional convolutional neural network

论文引用格式: Liu Yanyan, Chen Xin, Liu Wenyu, Yi Daihao. Occluded Gait Restoration and Recognition via Visibility-Aware Spatio-Temporal Dif-

fusion — SCID [J/OL]. Journal of Image and Graphics. DOI:10.11834/jig.250641. (刘艳艳, 陈欣, 刘文瑜, 易代豪. 可见性感知时空扩散的遮挡步态修复

© 中国图象图形学报版权所有

与识别—SCID[J/OL]. 中国图象图形学报. DOI: 10.11834/jig.250641.)

## 0 引言

步态识别作为一种基于生物行为特征的身份认证技术,因其非接触感知、远距离采集和难伪装性等优势,在公共安防监控(王茜等,2023)、刑事侦查(李俊翔等,2023)及医疗康复(Thota和Alberts,2013)等领域具有广泛应用前景。然而,尽管现有算法在理想实验环境下已取得较高精度,但在真实场景中,行人常被一些建筑物、车辆等静态障碍物或流动人群等动态物体遮挡(Hasan等,2024)。这种普遍存在的遮挡不仅导致人的步态轮廓在空间上缺失信息,更破坏了时间维度上的运动周期连续性,这也成为制约步态识别实际应用的挑战。

针对遮挡步态识别问题,早期的研究主要采用基于时序预测的重建方法,通过利用未遮挡帧的时序上下文信息来修复步态帧。例如Kumara等人(2021)提出了一种基于关键姿势的修复策略,该方法首先构建通用的关键姿势数据库,通过将遮挡序列与关键姿势进行匹配,利用双向长短期记忆网络(bidirectional long short-term memory, Bi-LSTM)从前向后帧的上下文信息中推断并重建出遮挡帧的潜在向量以达到重建作用。Paul等人(2022)则进一步改进了预测机制,通过去噪训练的方式分别优化前向和后向LSTM分支,最终融合双向预测结果以提升重建的鲁棒性。然而,LSTM等自回归模型在处理长序列时易发生误差累积,且难以捕捉高维图像的精细空间结构。这导致修复结果往往缺乏帧间动态连续性,易产生不自然的抖动,并且生成的轮廓细节模糊,又缺乏真实感。

不同于重建遮挡部位的技术思路,基于特征提取的非修复方法旨在直接从残缺序列的可见区域提取具有判别身份信息特征。高毅和何森(2023)针对密集遮挡场景,通过引入退化水平金字塔结构(DHPP)和随机二值膨胀的数据增广方法,提升网络对特征绝对位置的感知能力与遮挡鲁棒性。Sun等人(2024)设计了一种自适应注意力机制,通过加权的方式强制网络聚焦于未被遮挡的躯干或肢体部位,从而抑制遮挡噪声的干扰。Gupta和Chellappa(2025)提出的MimicGait则从模型无关的知识蒸馏

角度出发,利用可见与受遮挡步态之间的相关性约束来增强遮挡条件下的识别性能,接着提出RG-Gait模型,通过学习残差偏移量来补偿遮挡导致的空间特征损失。此外,Huang等人(2024)从时序建模角度出发,利用混合专家网络(mixture of temporal experts, MoE)对步态周期进行拆解,试图依据步态的周期性规律推断缺失的动作相位。然而,此类方法高度依赖可见区域的信息量,当遭遇大面积遮挡或关键运动部位完全缺失时,仅凭碎片化的局部特征难以构建具有足够区分度的身份表征。

随着深度学习的发展,基于深度生成模型的重建方法利用深层神经网络强大的非线性映射能力,显著提升了遮挡下的修复效果。阳强等人(2024)提出了一种基于先验知识的步态时空序列重建网络,该方法首先利用YOLOv5网络检测每一帧的局部遮挡区域并生成掩码作为先验,随后通过三维卷积神经网络(three-dimensional convolutional neural network, 3DCNN)编码器和Transformer模块在特征空间中同时修复空间缺失与时间断裂,最终生成完整的步态序列。此外,生成对抗网络(generative adversarial networks, GAN)也被广泛应用,如Hasan等人(2024)引入特定的遮挡检测模块来引导生成器对抗训练,以提升修复的视觉逼真度。然而,GAN除了训练不稳定和模式坍塌问题外,其在处理纹理细节时易产生虚假伪影,导致生成内容偏离原始身份特征,反而干扰下游识别。

为了突破对抗生成训练的稳定性瓶颈,近年来,扩散模型在图像生成与修复领域取得了超越GAN的突破性进展(Ho等,2020),该模型通过渐进式去噪生成符合数据真实分布的样本。在步态识别领域,Min等人(2024)提出的ZipGait构建了一种跨模态扩散框架,通过将骨架姿态的结构先验作为条件引导,有效地实现了骨架与轮廓特征在潜在空间的深度对齐与互补。Jin等人(2025)则深入挖掘步态流形结构,设计了几何驱动的扩散模块,将离散的步态帧建模为连续的流场表征,从而显著增强了生成特征的时空一致性。

但将标准扩散模型直接迁移至步态遮挡修复仍面临挑战:1)传统的2D扩散模型缺乏对时间维度的建模,逐帧修复无法保证步态动作的连贯性;2)常规修复过程缺乏对遮挡位置的更加细粒度的显式感知;3)现有的修复方法大多依赖修复后的图像,一旦

生成模型在严重遮挡下产生错误的肢体结构,将造成误判。

针对上述问题,本文提出了一种基于可见性感知时空扩散的遮挡步态修复与识别方法。本文认为不应盲目信任生成模型的修复结果,而要在精确感知遮挡的基础上,通过双分支融合机制,将扩散模型恢复后的时空连贯性步态特征与保留的原始遮挡样本的步态特征进行融合,避免生成伪影的干扰。

本文的主要工作和贡献如下:1)构建了区域评分估计器(RSE),该模块不仅为修复过程提供精确的空间定位引导,更为后续识别阶段提供先验,解决了传统方法缺乏精细化遮挡感知的问题;2)设计了基于3D-UNet的可见性感知3D扩散修复模块(VADM),该模块在RSE生成的空间定位引导下,利用3D卷积捕捉长时序依赖,实现了对遮挡区域准确的时空一致性修复,有效克服了2D模型的时序抖动;3)提出了双分支融合与信心引导机制,通过可见性空间注意力机制(V-SAM)自适应地调节原始特征与修复特征的融合权重,以解决当修复不确定性较高时所产生的伪影干扰,提高严重遮挡情况下的识别能力;4)在CASIA-B数据集上进行了系统性评估,对现实中多种复杂遮挡模式进行了测试,相比其他方法识别率有明显提升。

## 1 本文方法

### 1.1 概况

本文所提出的VAST-DRNet总体架构如图1所示。包含三个核心模块:区域评分估计器、可见性感知3D扩散修复模块以及基于双分支融合与信心引导的步态识别网络。首先,输入的遮挡步态序列被送入预训练并冻结后的RSE模块,对各帧各身体区域的可见性进行细粒度估计,得到对应的可见性评分。随后将该评分作为空间引导信号注入VADM模块中,以驱动网络捕捉时空依赖并对遮挡区域进行定向时空去噪与结构补全,从而生成结构完整且动态连贯的修复序列。

然而,尽管修复序列具备完整的拓扑结构,但生成过程可能引入伪影,而原始遮挡序列虽然信息缺失,却保留了最真实的身份特征。本文通过双分支融合与信心引导机制,将原始遮挡序列和修复序列分别通过两个独立的3D卷积分支提取特征,并由

注意力融合模块进行自适应合并。同时,根据可见性评分构建可见性空间注意力机制,通过残差连接作用于主干网络,进一步增强模型对高置信度人体区域的关注。最终,网络输出高鲁棒性的步态特征,经由时序池化(temporal pooling, TP)聚合为序列特征,并进一步通过水平金字塔池化(horizontal pyramid pooling, HPP)提取多尺度空间特征后,通过联合损失函数进行监督学习,以实现步态识别任务。

### 1.2 区域评分估计器

本模块的主要作用是将输入的遮挡步态序列 $S_{occ}$ 生成逐帧、逐部件的可见性评分向量 $s \in \mathbb{R}^P$ (其中 $P$ 是身体划分的部位总数)。该评分不仅作为VADM修复过程的空间引导信号,还为V-SAM提供了置信度评分先验。RSE模块包括三个主要阶段:多尺度特征提取与融合、部位划分与关系建模以及评分回归。

#### 1.2.1 多尺度特征提取与融合

RSE采用了一个四阶段的级联卷积网络作为骨干,以获取人体深层语义信息和遮挡边界的纹理细节。随着网络层级加深,特征通道数由32逐步扩展至256,以提取高层抽象特征。

网络提取特征过程中,选取第4阶段特征 $F_4$ 作为深层语义特征,并通过空间注意力模块增强对人体有效区域的响应权重。同时,为了弥补下采样产生的边缘细节丢失,提取第2阶段输出的浅层特征 $F_2$ ,经过双线性插值上采样至与 $F_4$ 相同尺寸后进行通道维度的拼接。最终,通过一个 $1 \times 1$ 融合卷积层将总通道数压缩为256,得到语义定位和边缘细节融合下的特征 $F'_{occ}$ ,以实现细粒度部位的评估。

#### 1.2.2 部位划分与Transformer关系建模

如图2所示,将融合特征 $F'_{occ}$ 沿垂直方向均匀地划分为 $P$ 个水平切片 $\{O^{(1)}, O^{(2)}, \dots, O^{(P)}\}$ ,针对每个切片并行执行全局平均池化(global average pooling, GAP)和全局最大池化(global max pooling, GMP),分别捕捉区域内的整体激活强度和显著性特征,拼接后生成维度为512的部位特征向量序列 $V = \{v_1, v_2, \dots, v_P\}$ 。然而,仅单独地评估每个部位的可见性是不够的,因为人体结构具有内在的连通性和依赖性。为此,通过引入Transformer编码器模块对部位特征进行全局上下文关系建模,使得模型能够利用未遮挡部位的信息辅助推断模糊区域的状态,从而提升遮挡评估的鲁棒性。

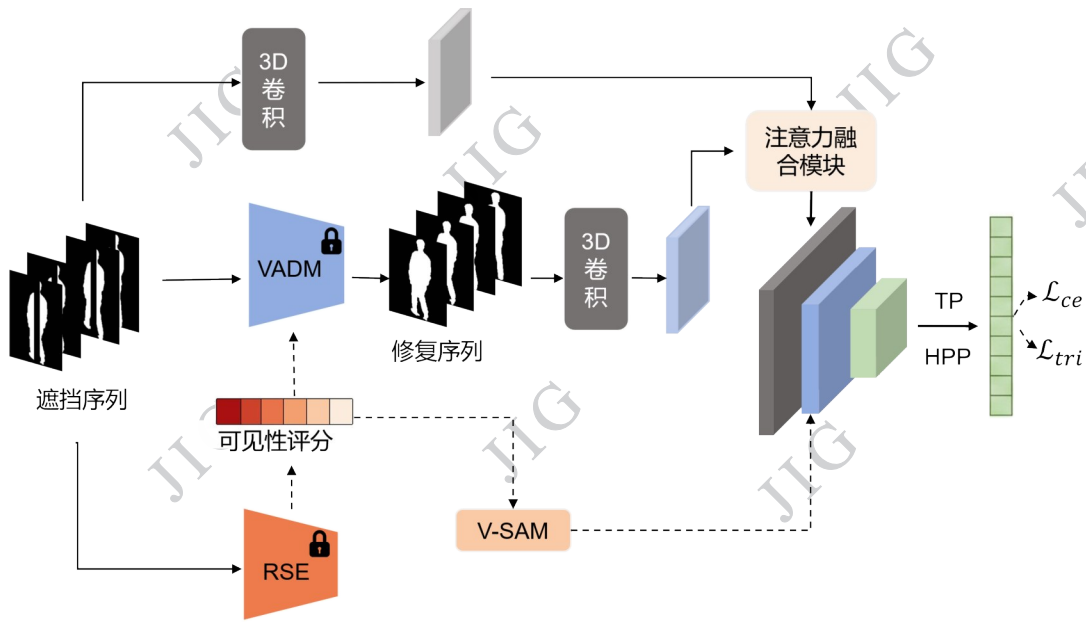


图1 VAST-DRNet 总体架构图

Fig. 1 The overall architecture of VAST-DRNet

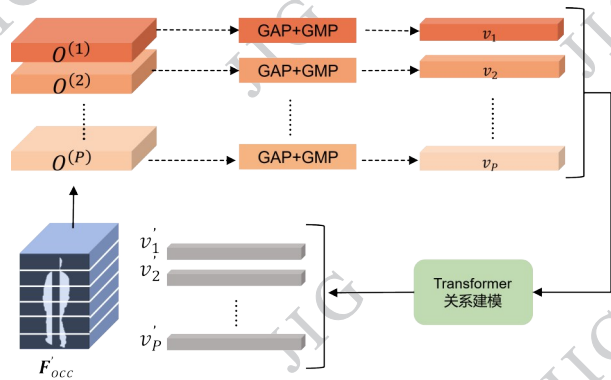


图2 部位划分模块

Fig. 2 Part Segmentation Module

如图3所示,利用自注意力机制显式地捕捉了不同身体部位间的遮挡相关性,具体为

$$Q = VW^Q, K = VW^K, U = VW^V \quad (1)$$

$$A = \text{softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right) \quad (2)$$

$$V' = AU \quad (3)$$

式(1)中,  $W^Q, W^K, W^V$  为可学习的投影矩阵,  $Q, K, U$  分别为查询(Query)矩阵、键(Key)矩阵和值(Value)矩阵。式(2)中,  $A \in \mathbf{R}^{p \times p}$  为部位间关系矩阵,该矩阵量化了部位间的关联程度,  $d_k$  为缩放因子。式(3)中,  $V' = \{v'_1, v'_2, \dots, v'_p\}$  为融入了全局结构上下文的增强特征。

### 1.2.3 评分回归与混合加权监督

最终,将增强特征序列  $V'$  送入  $P$  个并行的回归子网络将 512 维的特征压缩为一维标量,并经 Sigmoid 激活函数映射至  $(0, 1)$  区间,得到各部位的可见性评分,具体为

$$s_p = \sigma(M_p(v'_p)) = \frac{1}{1 + e^{-M_p(v'_p)}} \quad (4)$$

式中,  $M_p$  代表第  $p$  个部位的评估子网络,  $\sigma(\cdot)$  为激活函数。  $s_p$  为第  $p$  个部位的预测可见性评分。

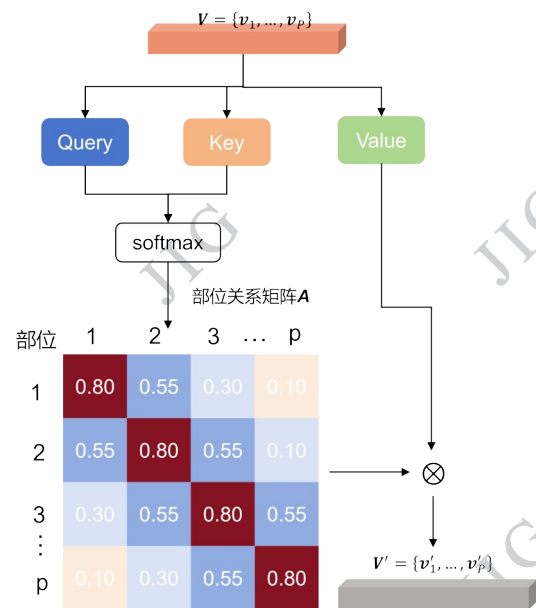


图3 部位自注意力机制

Fig. 3 Part Self-attention mechanism

为了解决遮挡样本分布不均衡以及部位重要性差异问题。通过可见性损失函数  $L_{RSE}$  来训练 RSE 模块,  $L_{RSE}$  由双重加权的平滑  $L1$  损失项  $L_M$  与顺序一致性损失项  $L_C$  构成。  $L_M$  用于衡量预测评分与真实标签的偏差, 并引入了双重加权机制, 具体为

$$L_M = \frac{1}{N \cdot P} \sum_{n=1}^N \sum_{p=1}^P \omega_s \cdot \omega_p \cdot L_{SL1}(s_{n,p}, g_{n,p}) \quad (5)$$

式中,  $L_{SL1}$  为平滑  $L1$  损失函数, 用于衡量预测评分  $s_{n,p}$  与真实可见性标签  $g_{n,p}$  之间的偏差,  $N$  为样本总数,  $P$  为部位总数。  $\omega_s$  为部位权重, 考虑到下肢运动包含了步态识别中最具判别力的动力学信息, 本文采用线性递增策略(从头部至腿部权重由 1.0 增至 2.0), 以强化对贡献度高的下肢区域的预测精度。  $\omega_p$  为样本权重, 针对数据集中严重遮挡样本稀缺导致的类不平衡问题, 本文采用分级动态调整权重策略。依据  $g_{n,p}$  对可见性低于 0.2 的极重度遮挡样本赋予 4.0 的高权重, 对中度和轻微遮挡样本进行阶梯式降权操作, 以此强化模型对遮挡情况严重的样本处理能力。

此外, 为了符合人体结构的连续性先验, 通过约束相邻部位评分的平滑过渡来避免突兀跳变, 本文定义了顺序一致性损失  $L_C$ , 具体为

$$L_C = \sum_{p=1}^{P-1} \max(0, |s_p - s_{p+1}| - \delta) \quad (6)$$

式中,  $s_p$  和  $s_{p+1}$  分别为相邻部位的可见性评分,  $\delta$  为惩罚阈值。最终的总损失函数具体为

$$L_{RSE} = L_M + \lambda L_C \quad (7)$$

式中,  $\lambda$  为平衡超参数。

### 1.3 可见性感知 3D 扩散修复模块

针对传统 2D 修复方法在处理视频序列时存在的帧间抖动与时序不连贯问题, 本文设计了可见性感知 3D 扩散修复模块(VADM)。该模块是针对步态序列的时空特性在去噪扩散概率模型(Ho 等, 2020)的框架进行了改进。其骨干网络采用改进的 3D U-Net 框架(Çiçek 等, 2016), 能够同时在空间维度重建步态轮廓的缺失拓扑, 并在时间维度维持步态运动的动态平滑性。

#### 1.3.1 时空编码器

编码器的主要任务是将高维的含噪视频流映射为紧凑的时空潜在表征。如图 4 所示, 初始输入为  $2 \times 64 \times 64$  的张量, 该张量是由当前时间步的带噪序列  $S_t$  与原始遮挡序列  $S_{occ}$  在通道维度拼接而成, 以此显

式地向网络提供待修复目标与已知观测条件的双重信息。

编码器由四个级联的 3D 卷积阶段构成, 随着网络层数加深, 特征通道数由初始的 32 逐步扩展至 128。在下采样操作中, 本文采用步长为  $1 \times 2 \times 2$  的 3D 卷积替代传统的池化层, 在  $T$  维度恒定的前提下, 将空间分辨率逐步压缩至  $16 \times 16$ 。最终, 编码器输出包含步态序列高级语义信息的  $128 \times 16 \times 16$  的深层时空特征图  $E_{occ}$ 。

#### 1.3.2 评分引导与条件注入

为了解决传统扩散模型盲目去噪的问题, 本文在瓶颈层与解码器的衔接处(即最深层特征)创新地引入了可见性的空间引导信号。首先将离散的时间步  $t$  映射为正弦位置嵌入向量, 同时通过多层感知机(multi-layer perceptron, MLP)将由 RSE 模块生成的可见性评分向量  $s$  映射为同维度的评分嵌入。

如图 4 中, 两个嵌入向量逐元素相加融合成引导嵌入  $\varepsilon_{fused}$  本文将将其广播至与瓶颈层特征  $E_{occ}$  相同的时空尺寸, 并通过残差连接的方式得到引导增强特征  $F_{GEF}$ , 然后注入解码器中, 实现修复引导, 具体为

$$F_{GEF} = \text{ReLU}(E_{occ} + \varepsilon_{fused}) \quad (8)$$

式中, ReLU 为激活函数。

通过引导条件的注入, 修复网络在去噪的最深层能够感知到哪些部位是低可见性的, 需要重点生成。哪些是高可见性的, 只需保持原样, 从而实现靶向性的特征重构。

#### 1.3.3 3D 残差解码器

解码器的任务是将  $F_{GEF}$  逐步恢复至原始分辨率。通过转置 3D 卷积(transposed 3D convolution)进行上采样, 逐级恢复特征图的空间尺寸。为了解决深度网络中梯度消失问题, 并弥补下采样造成的高频细节丢失, 本文在编码器和解码器之间构建了残差跳跃连接。如图 4 所示, 与标准 U-Net 采用的通道拼接不同, 本文采用逐元素相加的方式将编码器的浅层特征融合到解码器的对应层级。使得解码器能够更加专注于学习残差信息, 从而利用深层语义来修正浅层的结构细节。最终, 输出层通过  $1 \times 1$  卷积将特征映射回原始图像空间, 输出预测的清晰步态序列  $\hat{S}_0$ 。

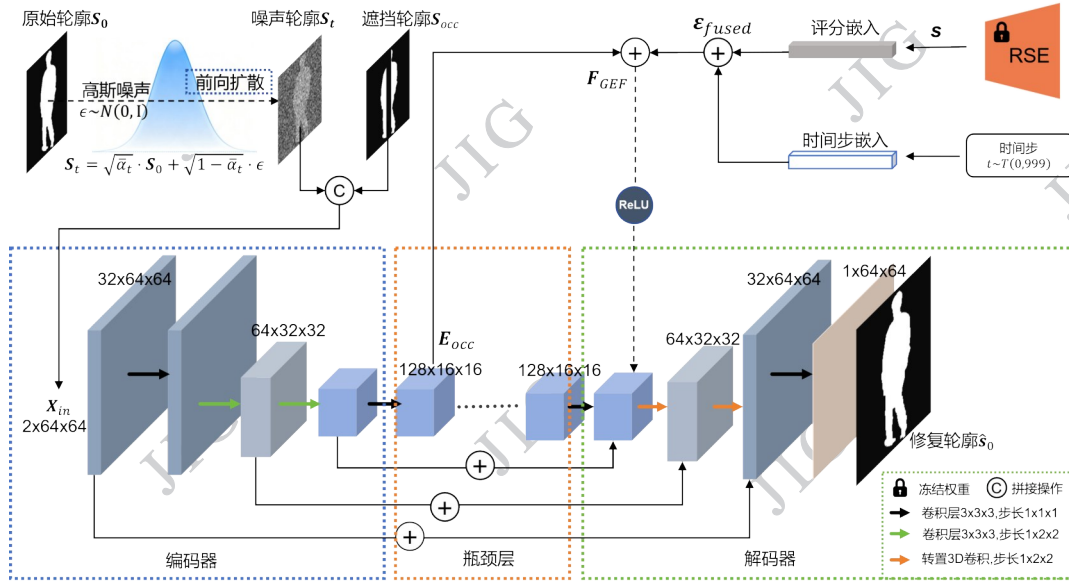


图 4 可见性感知 3D 扩散修复模型

Fig. 4 The architecture of the Visibility-Aware Diffusion Module

### 1.3.4 扩散修复与优化过程

如图 4 所示,前向扩散过程中,将对齐后的原始无遮挡轮廓  $S_0$  作为扩散目标,定义为扩散过程的初始状态。遵循预定义的线性噪声调度策略  $\beta_t$ ,通过马尔可夫链逐步向  $S_0$  添加高斯噪声。得到任意时刻  $t$  的噪声轮廓  $S_t$ ,具体为

$$S_t = \sqrt{\alpha_t} S_0 + \sqrt{1 - \alpha_t} \epsilon, \quad \epsilon \sim \mathcal{N}(0,1) \quad (9)$$

式中,  $\alpha_t = 1 - \beta_t$ ,  $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$  为累积噪声方差系数,  $\epsilon$  为随机噪声。

为了保证生成的步态不仅像素分布正确,而且具备清晰的边缘轮廓,通过扩散重构损失  $L_{diff}$  和边缘保持损失  $L_{edge}$  组成的混合损失函数  $L_{total}$  来驱动网络参数  $\theta$  的优化。 $L_{diff}$  采用均方误差 (MSE) 直接约束模型对原始清晰序列  $S_0$  的预测能力,具体为

$$L_{diff} = \mathbb{E}_{t, S_0} [\|S_0 - f_\theta(\text{Concat}(S_t, S_{occ}, t, s))\|_2^2] \quad (10)$$

式中,  $f_\theta(\cdot)$  表示 VADM 网络的预测输出。

其次,引入边缘保持损失  $L_{edge}$ 。该损失通过计算预测序列与真实序列在拉普拉斯算子 (Mathieu 等, 2015) 滤波下的梯度图差异,显式惩罚边缘区域的模糊伪影,具体为

$$L_{edge} = \|\nabla^2 S_0 - \nabla^2 \hat{S}_0\|_2^2 \quad (11)$$

式中,  $\nabla^2$  表示拉普拉斯边缘提取算子。最终的总损失函数,具体为

$$L_{total} = L_{diff} + \lambda_{edge} L_{edge} \quad (12)$$

式中,  $\lambda_{edge}$  为平衡超参数。在推理阶段,本文采用去

噪扩散隐式模型 (denoising diffusion implicit models, DDIM) 采样策略加速生成过程,从纯高斯噪声中恢复出结构完整、动作连贯的步态序列。

### 1.4 特征识别网络与信心引导融合

将修复后的序列  $\hat{S}_0$  与遮挡序列  $S_{occ}$  并行输入特征提取网络,以提取空间结构完整且包含真实性的判别性步态识别特征。本文选用 GaitGL (Lin 等, 2021) 作为骨干网络,并设计了双分支融合与信心引导机制来优化特征表达。

首先,利用两个权重共享的 3D 卷积分支分别提取  $\hat{S}_0$  和  $S_{occ}$  的浅层特征,并通过注意力融合模块计算像素权重  $W$ ,对分支特征进行加权融合。为了抑制修复伪影对识别精度的干扰,引入 V-SAM。该机制利用 RSE 对  $S_{occ}$  进行可见性评估,将所得评分生成置信度掩码  $M$ ,通过残差门控的方式对特征进行动态校准。

在训练阶段,对于批次大小为  $B \times K$  的输入 (即  $B$  个身份数,每个身份包含  $K$  个样本),采用交叉熵损失  $L_{ce}$  与三元组损失  $L_{tri}$  的联合损失函数进行监督学习,具体为

$$L_{ce} = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{w_i^T z_i}}{\sum_{j=1}^C e^{w_j^T z_i}} \quad (13)$$

式中,  $N = B \times K$  为总样本数,  $C$  为类别总数,  $y_i$  为第  $i$  个样本的真实标签,  $z_i$  为第  $i$  个样本的特征向量,  $W_j$  为全连接层中第  $j$  类的权重向量,  $W_{y_i}$  为对应真实标

签类别的权重向量。

$$L_{tri} = \frac{1}{B \cdot K} \sum_{i=1}^B \sum_{a=1}^K \max(m + d(z_{i,a}, z_{i,p}) - d(z_{i,a}, z_{j,n}), 0) \quad (14)$$

式中,  $z_{i,a}$ 、 $z_{i,p}$ 、 $z_{j,n}$  分别表示锚点、正样本和负样本特征,  $d(\cdot, \cdot)$  为欧氏距离,  $m$  为边界阈值。最终的总识别损失具体为

$$L_{rec} = L_{ce} + L_{tri} \quad (15)$$

## 2 实验与分析

### 2.1 数据集

本文研究的所有实验均在主流公开步态数据集 CASIA-B (Yu 等, 2006) 上进行。该数据集包含 124 名受试者, 涵盖了多视角与多状态的复杂场景。每位受试者在 11 个不同的摄像机视角 (从  $0^\circ$  到  $180^\circ$ , 间隔  $18^\circ$ ) 下采样了 10 组步态序列, 包括 6 组正常行走 (normal, NM)、2 组携包行走 (bag, BG) 和 2 组换装行走 (clothing, CL)。实验遵循广泛采用的大样本训练协议 (Chao 等, 2019), 将前 74 名受试者划分为训练集, 其余 50 名则作为测试集。在测试阶段, NM 条

件的前 4 个序列构成了注册集 (Gallery), 而其余 6 个序列为探针集 (Probe)。CASIA-B 数据集提供的是黑白轮廓图像, 为适配网络输入, 原始轮廓图像经重心对齐和尺寸归一化后, 裁剪成  $64 \times 44$  像素的轮廓序列。

### 2.2 遮挡合成

尽管 CASIA-B 包含了携带包裹和换装的变化, 但仍缺乏针对环境遮挡的系统性评估数据。为此, 本研究参考并改进了 OccSilGait 框架 (Peng 等, 2023), 构建了包含三种模式的遮挡步态数据集 OccCASIA-B。如图 5 展示了各类遮挡模式的合成示例。静态物体遮挡模式旨在模拟行人经过路灯、长椅等现实环境中静止的障碍物场景。

为了提升场景多样性和遮挡难度, 本文在原有基础上, 引入 Mapillary-Vistas 街景数据集中更具普遍性的静态遮挡物, 构建了一个包含长椅、消防栓和植被等 8 种遮挡类型。合成时, 根据行人轮廓的实时高度, 将遮挡物缩放至合理比例 (0.3~0.8), 并随机放置于序列的水平轨迹上, 不仅模拟了外观遮挡, 更重现了现实中因透视关系产生的相对位置变化。



图 5 CASIA-B 数据集中的多种模拟遮挡类型示例

Fig. 5 Examples of simulated occlusion types in the CASIA-B dataset

方向性遮挡模式主要模拟因摄像机视角限制或行人检测框截断导致的身体部分缺失。如图 5 所示, 除了上部、下部、左侧、右侧四种常规遮挡方位

外, 本文新增了模拟栏杆遮挡的膝部遮挡和模拟杆状物遮挡的条状遮挡。合成时, 本文根据预设的遮挡比例  $[0.1, 0.5]$ , 在人体轮廓的特定方向移除像素

区域,以制造刚性的局部信息丢失。

人群遮挡模式用于模拟行人间的相互遮挡。如图5中的左穿越和右穿越所示,主要模拟一名干扰行人从左侧或右侧穿过目标对象的过程。合成中,本文将干扰序列视为动态掩码,对其相对移动速度  $v \in [-3, 3]$  (像素/帧)和起始相位进行完全随机化采样。与静态遮挡不同,这种模式会在时序上造成连续且非线性的特征损坏,极大地破坏了步态周期的完整性。

为了使模型能够全面学习并适应复杂多变的遮挡环境,本文在构建 OccCASIA-B 数据集时采用了全类别覆盖的划分策略。即训练集与测试集均涵盖了上述所有定义的静态、方向性及人群遮挡类型。尽管遮挡类型在类别上保持一致,但为了严谨验证模型的泛化能力,本文在合成过程中对测试集的具体遮挡参数进行了严格的独立随机化采样。这一设置确保了2.3节中的测试策略能够真实反映模型在已知遮挡模式下,针对未见过的具体遮挡形式的鲁棒性。

### 2.3 训练细节与评判标准

本文模型基于 PyTorch 框架实现,所有实验均在两块 NVIDIA GeForce RTX 4090 GPU 上进行训练和测试。优化器采用 Adam,初始学习率设置  $1.0 \times 10^{-4}$ ,权重衰减设置为  $5.0 \times 10^{-4}$ 。学习率调度采用余弦退火策略。批处理大小设置为(8,16),即每个批次随机抽取8个身份,每个身份采样16个序列。为了平衡生成质量与推理效率,扩散模型采样步数设置为50步。

为全面评估模型在不同场景下的性能,本文采用两种测试策略:1)分类型测试:在各遮挡类型的子集上分别进行独立测试,用于评估模型在特定遮挡情况下的识别能力;2)混合场景测试:在混合测试集上进行测试,以模拟真实监控场景中遮挡目标与非遮挡目标随机出现的复杂分布。

本文采用步态识别领域标准的 Rank-1、Rank-5、Rank-10 准确率作为步态识别性能指标,所有结果均为排除同视角条件下的平均准确率。为了验证 RSE 模块对遮挡评分回归的准确性,采用平均绝对误差(MAE)和均方根误差(RMSE)用于衡量预测评分与真实标签之间的误差大小,前者反映整体平均偏差,后者对较大误差更为敏感;平均误差(Mean Error)用于反映预测结果是否存在系统性偏高或偏

低;决定系数  $R^2$  则用于衡量回归结果对真实评分变化趋势的拟合程度。

最后,为了评估修复网络对遮挡序列的时空重构质量,采用峰值信噪比(PSNR)和结构相似性指标(SSIM)用于衡量重构结果在像素质量和结构保持方面的表现;边缘 F1 分数(Edge F1)和轮廓交并比(Contour IoU)用于评估轮廓边缘的恢复精度与区域重叠程度;学习感知图像块相似度(LPIPS)则用于衡量重构结果在感知特征空间中的相似性,以反映其视觉自然度。

## 2.4 区域评分估计器性能验证

### 2.4.1 整体与分部位性能分析

为了验证区域评分估计器在细粒度可见性评估任务上的有效性,本文首先基于表1对测试集上的整体与分部位回归性能进行了定量分析。从表中整体指标来看,RSE 模型平均 MAE 为 0.0305,  $R^2$  为 0.8800,而 Mean Error 的均值极为接近 0,这表明模型预测不存在明显的系统性偏差,能够进行客观的可见性评估。

在分部位分析中,虽然下肢区域通常因运动幅度大而难以预测,但得益于本文在1.2.3节中设计的部位加权损失策略,模型对步态识别中贡献度最高的大腿区域表现出色,其  $R^2$  达到了 0.9408,MAE 仅为 0.0302,甚至优于相对静止的躯干区域( $R^2=0.8855$ )和头部区域( $R^2=0.8249$ )。这种针对关键运动部位的精准评估,确保了在复杂遮挡下,模型能优先保留和恢复最具判别力的步态特征,验证了损失函数设计的有效性。

### 2.4.2 不同遮挡程度下的鲁棒性分析

为了检验模型在长尾分布与极端遮挡环境下的稳健性,本文将测试样本按可见性范围划分为极重度(0.0~0.25)、重度(0.25~0.5)、中度(0.5~0.75)及轻度(0.75~0.95)四个等级,具体 MAE 分布如表2所示。其中为了更严谨地评估模型在真实遮挡下的性能,本文在轻度遮挡类别中选择截取可见性为 0.75 至 0.95 的样本,排除了近乎完美的理想轮廓,从而构建了更具挑战性的测试基准。

实验结果表明,RSE 模型在不同遮挡程度下保持了高度的稳定性。在中度遮挡区间,模型取得了最低的预测误差(MAE=0.0513),显示其对最复杂的半遮挡边界具有极强的特征捕捉能力。更为关键的是,在极重度遮挡的极端条件下,尽管样本数量

表1 RSE模块的整体与部位性能评估

Table 1 Overall and Component Performance Assessment of the RSE Module

部位	MAE	RMSE	Mean Error	$R^2$	命中率@0.1/%
头部	0.0220	0.0443	-0.0142	0.8249	95.2
躯干	0.0326	0.0576	-0.0132	0.8855	90.3
大腿	0.0302	0.0511	-0.0160	0.9408	93.3
小腿	0.0373	0.0641	-0.0250	0.8689	89.3
平均	0.0305	0.0543	-0.0171	0.8800	92.0

(328个)远少于其他类别,存在明显的类不平衡问题,模型仍将MAE控制在0.0725的低水平。这证明RSE并未受制于数据分布偏差,在有效信息极度匮乏的情况下,仍能保持与轻度遮挡(MAE=0.0593)相当的判断精度,具备良好的鲁棒性。

表2 RSE模块在不同遮挡程度下的MAE分析

Table 2 MAE Analysis of RSE Module Under Different Occlusion Levels

可见性范围	遮挡程度	样本数量/个	MAE
0.0~0.25	极重遮挡	328	0.0725
0.25~0.5	重度遮挡	772	0.0687
0.5~0.75	中度遮挡	1220	0.0513
0.75~0.95	轻度遮挡	1604	0.0593

#### 2.4.3 遮挡等级可分性与可视化验证

为了进一步量化RSE模块对遮挡程度的判别能力,本文将RSE的预测分数和真实评分按照前述四个可见性区间映射为极重、重度、中度和轻度四类标签,并统计其混淆矩阵,如图6所示。

矩阵整体呈现较明显的对角线分布,说明RSE对不同遮挡等级具有较好的区分能力。其中,中度和轻度遮挡准确率分别达到94.4%和97.1%。对于最具挑战性的极重度遮挡,估计器仍取得了85.7%的正确识别率,这表明在可见信息较为有限的条件下,RSE仍具有较好的判断能力。虽然重度遮挡正确识别率为73.5%存在一定混淆,但其26.5%的错误主要集中在相邻的中度遮挡中,并未出现跨越式极端误判,验证了预测逻辑的连贯性。实验表明RSE不仅能够较准确地回归遮挡评分,而且在离散化后的遮挡等级判别上也具有较好的可分性。即使在重度和极重度遮挡条件下,模型仍能保持较为稳定的分类表现。

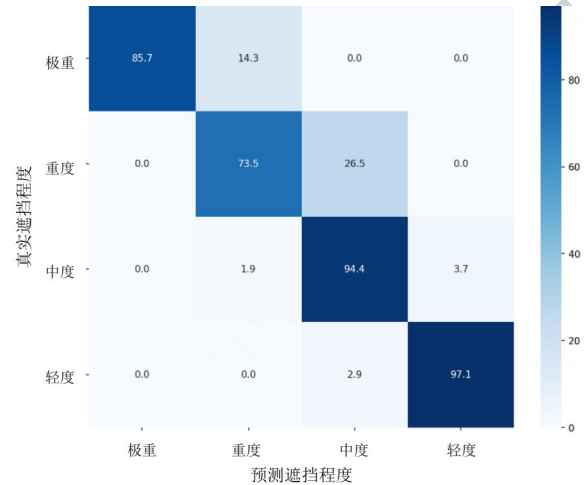


图6 RSE预测遮挡等级的混淆矩阵(%)

Fig. 6 Confusion matrix of RSE predicted occlusion levels (%)

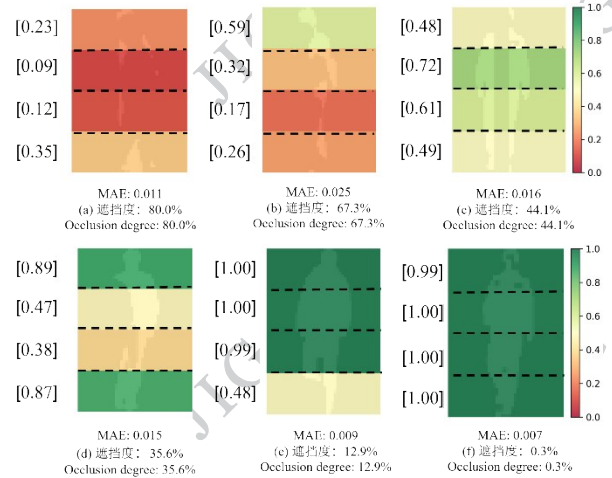


图7 不同遮挡程度下RSE预测结果

Fig. 7 Qualitative results of RSE predictions under different degrees of occlusion

图7进一步从可视化角度展示了RSE对不同遮挡程度样本的评分结果。在遮挡程度80.0%的困难样本图7(a)中,模型准确地为被遮挡最严重的躯干和上腿部分给出了[0.09]和[0.12]的极低置信度

评分,这与热力图的深红色区域完全对应。在遮挡率仅为12.9%的样本图7(e)中,模型能够捕捉到存在于脚部的微小遮挡,给出了[0.48]的局部低分,同时对未受影响的头部和躯干保持了[1.00]的高置信度。而在几乎无遮挡0.3%的样本图7(f)中,所有部位评分均接近[1.00],热力图也相应呈现为代表高可见性的纯绿色。这些结果直观地证实,RSE能够精确感知遮挡的空间位置与严重程度,为VADM模块的定向修复提供了可靠的先验。

## 2.5 主要对比实验结果

### 2.5.1 不同遮挡类型下的主要对比实验分析

为了验证 VAST-DRNet 的有效性,本文遵循 2.3 节定义的测试策略,在 OccCASIA-B 数据集上进行了评估。实验选取了 GaitSet (Chao 等, 2019)、GaitPart (Fan 等, 2020)、GaitGL (Lin 等, 2021)、GaitBase (Fan 等, 2023)、SATOR (Peng 等, 2023)、GaitGS (Xiong 等, 2024)、QAGait (Wang 等, 2024) 及 MimicGait (Gupta 和 Chellappa, 2025) 等步态识别方法作为对比基线。

从表 3 数据可知,在无遮挡条件下,各方法均取得较高的 Rank-1 识别率,其中 GaitGS 取得了 88.4% 的最高平均 Rank-1 识别率,本文方法平均识别率为 87.8%,位列次优;在 CL 条件下,本文方法识别率为

77.4%,为该项最优。这表明本文方法在未受遮挡的常规序列上,基本维持了原有的判别能力。

在静态遮挡和人群遮挡场景下,本文方法的平均 Rank-1 识别率分别为 83.1% 和 78.1%。与对比方法相比,相较于 GaitGL 分别提升了 11.2 和 7.6 个百分点,相较于 SATOR 分别提升了 2.3 和 2.6 个百分点。具体在具有较高难度的 CL 条件中,本文方法在静态遮挡和人群遮挡下的识别率分别为 70.4% 和 62.9%,均高于表格中的其他基线模型。这表明本文方法对服饰变化和复杂遮挡叠加情况下识别性能有一定提升作用。

在方向性遮挡场景下,本文方法的平均 Rank-1 识别率为 80.5%,在 CL 条件中,本文方法的识别率为 66.7% 低于 MimicGait 方法的 68.4%。在应对此类遮挡时本文方法并未达到最优。这可能由于扩散生成范式在逆向去噪重构过程中会带来一定的平滑效应和重采样噪声,简单遮挡场景下的特征锐度略有削弱,从而造成在该子集上微小的性能差距。从整体来看,本文方法的性能优势主要体现在遮挡形式更复杂、干扰更强的场景中,在提升模型对复杂遮挡鲁棒性的同时,在各类遮挡场景中保持了可比的识别精度。

表 3 各遮挡类型纯净子集上的 Rank-1 识别性能对比

Table 3 Comparison of Rank-1 Identification Performance on Clean Subsets of Various Occlusion Types

模型	无遮挡				静态遮挡				方向性遮挡				人群遮挡			
	NM	BG	CL	平均	NM	BG	CL	平均	NM	BG	CL	平均	NM	BG	CL	平均
GaitSet	92.4	84.4	75.6	84.1	85.7	76.5	50.9	71.0	80.6	72.1	43.4	65.4	83.9	71.5	42.6	66.0
GaitPart	91.4	84.3	70.1	81.9	83.2	72.6	55.8	70.5	80.4	72.2	51.8	68.1	81.9	71.9	49.5	67.8
GaitGL	93.5	87.3	75.2	85.3	82.8	74.8	58.0	71.9	82.9	77.9	58.5	73.1	83.5	77.3	50.8	70.5
GaitBase	93.0	82.4	73.4	82.9	80.4	71.9	43.8	65.4	83.8	75.7	46.8	68.8	77.8	69.6	50.3	65.9
SATOR	95.7	89.5	76.4	87.2	90.1	84.4	67.9	80.8	<b>90.8</b>	83.7	67.5	<b>80.7</b>	86.8	78.5	61.1	75.5
GaitGS	<b>96.0</b>	<b>92.1</b>	77.0	<b>88.4</b>	84.5	73.3	47.8	68.5	82.5	76.7	58.6	72.6	82.4	74.2	54.1	70.2
QAGait	95.7	90.2	72.1	86.0	83.4	72.9	51.8	69.4	85.7	77.3	49.6	70.9	80.5	73.2	54.9	69.5
MimicGait	95.4	91.7	76.9	88.0	83.2	78.8	67.5	76.5	84.7	76.6	<b>68.4</b>	76.6	81.2	75.3	61.3	72.6
本文	95.5	90.5	<b>77.4</b>	87.8	<b>90.5</b>	<b>88.5</b>	<b>70.4</b>	<b>83.1</b>	89.3	<b>85.5</b>	66.7	80.5	<b>88.1</b>	<b>83.4</b>	<b>62.9</b>	<b>78.1</b>

注:加粗字体表示各列最优结果。

### 2.5.2 混合遮挡场景下的消融实验

为了验证 VAST-DRNet 模型中各关键组件的有

效性,表 4 展示了在混合遮挡场景下,逐步累加模块后的消融实验结果。基线模型(实验 1)的平均准确

率为83.0%,在此基础上,仅加入V-SAM的实验2平均准确率相比提高了1.6个百分点,说明V-SAM对不同条件下的识别性能均有一定改善作用。相比之下,实验3平均准确率为83.8%,较基线有所提高,但整体低于实验2。其中,实验3在NM和CL条件下分别达到91.5%和74.6%,较基线有所提升,这表明,在缺少显式可见性引导的情况下,单独利用修复模块虽能够提升一定的性能,但作用相对有限。

进一步地,在实验3的基础上加入RSE后的实验4的平均准确率为85.0%,较实验3提高1.2个百分点,同时三种条件下的识别结果均有所提高,这表明说明可见性引导信息的引入有助于进一步发挥修复模块的作用。最终,实验5取得最佳性能,较基线提高3.1个百分点。

消融实验结果表明,V-SAM、RSE和VADM均对模型性能提升具有积极作用。其中,V-SAM能够带来较为直接的性能增益,VADM在引入RSE后表现出更明显的改善,而三者结合后取得了最优的综合识别效果。这说明,本章所提出的各关键组件之间具有较好的互补性,并能够在混合遮挡场景下共同提升模型的识别表现。

表4 混合遮挡场景下VAST-DRNet关键组件的消融实验分析

Table 4 Ablation Study Analysis of VAST-DRNet Key Components under Mixed Occlusion Scenarios

实验序号	RSE	V-SAM	VADM	Rank-1 识别率 /%			
				NM	BG	CL	平均
1	-	-	-	90.1	85.7	73.1	83.0
2	-	√	-	91.8	87.1	75.0	84.6
3	-	-	√	91.5	85.4	74.6	83.8
4	√	-	√	92.4	86.8	75.7	85.0
5	√	√	√	93.6	88.4	76.3	86.1

注:加粗字体表示各列最优结果。“-”代表未采用,“√”代表采用。

### 2.5.3 混合遮挡场景下的鲁棒性分析

为了进一步评估模型在不同遮挡程度下的识别性能,本文根据步态序列的平均遮挡面积比例,将混合遮挡测试集划分为5%~10%、10%~30%、30%~50%和>50%四个区间,实验结果见表5。

随着遮挡比例的增大,各方法的Rank-1准确率

整体呈下降趋势,但本文方法的性能退化幅度相对更小。从5%~10%到>50%区间,Rank-1仅下降了40.3个百分点,而GaitGL、GaitGS等方法分别下降了43.6和49.2个百分点,表明VAST-DRNet在严重遮挡场景下具有更稳定的判别能力。

在遮挡比例为5%~10%的轻微遮挡区间,GaitGS以92.8%的Rank-1准确率取得最优结果,高于本文方法的90.9%。说明在遮挡较轻、轮廓信息相对完整的情况下,本章方法并未表现出明显优势。当遮挡比例为10%~30%时,其准确率均优于所有对比方法。与基线模型GaitGL相比,则分别提升1.6、1.1和0.4个百分点。在30%~50%区间,本章方法的Rank-1和Rank-5分别达到76.5%和89.9%,同样优于对比方法,但其Rank-10略低于最优的GaitGL。该结果表明,随着遮挡程度进一步加深,本章方法仍能够保持较好的识别性能。

在>50%区间中,各方法性能均明显下降。此时,本章方法的Rank-1准确率仍达到50.6%,为所有方法中最高。但在Rank-5和Rank-10指标上,GaitGL分别达到76.4%和79.6%。这说明,在极重遮挡条件下,不同方法在检索结果分布上表现出一定差异。本章方法在首选匹配结果上仍具有较强优势,而GaitGL在更宽松的检索指标下表现更优。这一现象在一定程度上表明,判别式方法与修复增强方法在极端信息缺失条件下对残余特征利用方式存在差异:前者更容易保留若干相似候选样本,而后者则更强调对目标身份特征的集中恢复,因此在Top-1指标上更具优势,但在更高阶检索指标上未必始终占优。

总体来看,随着遮挡比例增大,各方法的识别性能均呈下降趋势,但本章方法在中高遮挡区间表现出更明显的优势。

## 2.6 修复质量分析

### 2.6.1 定量评估与分析

为了验证VAST-DRNet在步态序列时空重构方面的有效性,本节从定量指标表6和定性视觉效果图8和图9,对比了本文方法与基准2D扩散模型及无RSE引导下3D扩散模型的性能差异。表6展示了不同模型在OccASIA-B数据集上的整体修复性能。

相比于基础的VADM 2D模型,VADM 3D模型在PSNR和SSIM指标上分别提升了2.38dB和

表 5 混合遮挡场景下不同遮挡程度下的识别准确率对比

Table 5 Comparison of Identification Accuracy at Different Occlusion Levels under Mixed Occlusion Scenarios

模型	/%											
	5%~10%			10%~30%			30%~50%			>50%		
	Rank-1	Rank-5	Rank-10	Rank-1	Rank-5	Rank-10	Rank-1	Rank-5	Rank-10	Rank-1	Rank-5	Rank-10
GaitSet	83.7	89.6	92.8	78.2	84.6	88.8	70.5	81.8	86.3	44.9	63.0	73.6
GaitPart	84.6	93.7	95.5	82.3	94.5	96.9	73.0	87.3	92.4	42.1	61.6	71.8
GaitGL	90.4	95.5	95.7	87.7	95.9	98.0	74.2	88.4	<b>95.9</b>	46.8	<b>76.4</b>	<b>79.6</b>
GaitBase	89.6	92.3	95.0	84.0	90.4	93.2	71.5	86.9	91.1	44.4	64.8	78.7
SATOR	91.1	<b>97.1</b>	<b>98.1</b>	88.5	96.8	98.3	74.9	89.1	94.2	47.5	61.8	74.5
GaitGS	<b>92.8</b>	95.5	96.0	89.1	95.9	97.1	72.2	87.4	92.6	43.6	66.4	77.7
QA Gait	89.7	94.7	95.5	81.0	91.2	94.6	71.9	88.2	93.3	46.2	65.4	76.4
MimicGait	89.6	95.3	97.6	85.1	94.2	95.6	75.4	89.7	94.4	46.7	66.1	75.5
本文	90.9	96.9	97.9	<b>89.3</b>	<b>97.0</b>	<b>98.4</b>	<b>76.5</b>	<b>89.9</b>	94.7	<b>50.6</b>	63.0	76.0

注:加粗字体表示各列最优结果。

3.8%,同时 Contour IoU 由 0.8762 提升至 0.8926, LPIPS 由 0.0386 降至 0.0350。这表明,3D 卷积的引入能够有效利用相邻帧的互补信息填补单帧视角下的信息缺失,从而改善整体重构质量和轮廓一致性。但 3D 模型的 Edge F1 为 0.6779,低于 2D 模型的 0.6855。这表明,虽然 3D 时空建模有助于提升整体结构恢复效果,但在缺少显式可见性引导时,模型对局部边缘细节的恢复仍存在一定不足。

表 6 整体修复性能对比

Table 6 Summary Repair Performance

评估指标	VADM		
	2D	3D	3D+RSE(本文)
PSNR/dB	26.18	28.56	<b>31.35</b>
SSIM/%	81.8	85.6	<b>88.3</b>
Edge F1	0.6855	0.6779	<b>0.7021</b>
Contour IoU	0.8762	0.8926	<b>0.9025</b>
LPIPS	0.0386	0.0350	<b>0.0324</b>

注:加粗字体表示各行最优结果。

相比之下,本文的 VADM 方法在引入 RSE 提供的可见性评分作为空间引导后,在所有指标上均取得了最优性能。其中,PSNR、SSIM 和 Contour IoU 分别达到 31.35dB、88.3% 和 0.9025,Edge F1 提升至 0.7021,LPIPS 进一步降至 0.0324。这表明,RSE 提

供的部位级可见性先验有效地抑制了 3D 扩散模型的盲目修复,使扩散模型能够聚焦于遮挡区域进行定向修复,不仅能够提升整体重构质量,还能更有效地改善轮廓边缘与感知细节的恢复效果。

## 2.6.2 混合遮挡场景下的修复定性分析

为了直观评估模型在复杂遮挡场景下的鲁棒性,图 8 和图 9 分别展示了垂直条状遮挡和大面积下肢遮挡两类典型困难场景下的可视化结果,其中同时给出了修复结果与对应误差图。

如图 8 所示,在垂直条状遮挡场景中,人体轮廓的横向结构被局部切断,极易导致特征提取时的空间错位。图中可知,VADM 2D 模型在部分帧的修复中存在较明显的结构失真和时序不稳定现象,其误差图显示出贯穿全身的高亮误差带,说明该方法对此类遮挡的恢复能力有限。相比之下,VADM 3D 模型能够较好地补全被遮挡区域,整体轮廓连续性优于 2D 模型,但在人体边缘处仍存在较明显的误差残留,说明其在边缘细节恢复方面仍有不足。本文(3D+RSE)的修复方法通过修复的序列可以看出其修复效果与原始序列十分接近,修复效果显著提升。这表明在 RSE 引导下的扩散模型能够更好地兼顾结构完整性与边缘清晰度。

如图 9 进一步展示了大面积下肢遮挡条件下的修复结果。从图中可知,2D 模型在部分帧中出现肢体修复模糊或未修复现象,误差主要集中在腿部缺

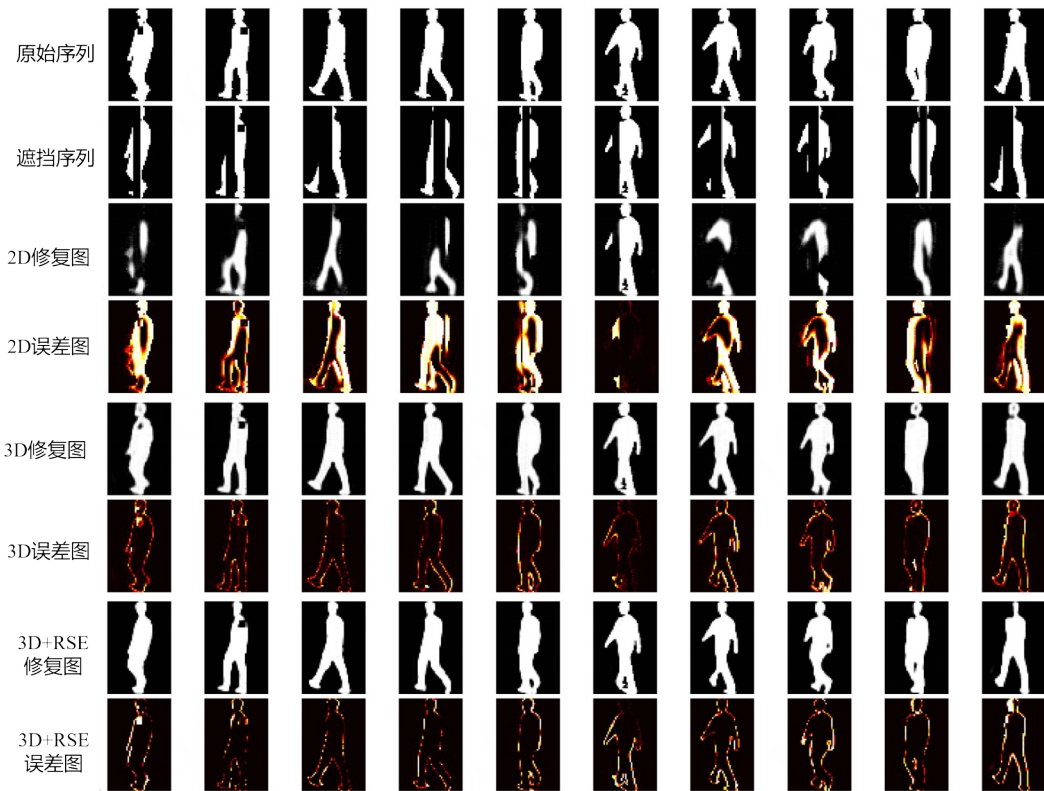


图8 垂直条状遮挡的修复对比

Fig. 8 Comparison of Restoration for Vertical Strip Occlusion

失区域。这进一步表明存在时序一致性缺失问题。3D模型虽然恢复了腿部的大致轨迹,但从误差图的高亮区域的结果可以看到空心红圈人体边缘轮廓,说明其生成结果在轮廓宽度和边界位置上仍存在一定偏差。相比之下,本章方法在下肢连接完整性、轮廓自然性以及跨帧一致性方面表现更好,其误差图几近全黑,尤其在关键边缘区域的误差进一步减小。

可视化结果表明,相比于2D模型和未引入RSE引导的3D基线模型,本章方法在复杂遮挡场景下能够取得更好的修复效果,不仅提升了人体轮廓的完整性,也在一定程度上改善了边缘细节与时序一致性。

### 结论

针对步态识别中存在遮挡的问题,本文提出了可见性感知时空扩散的遮挡步态修复与识别方法(VAST-DRNet),该模型通过区域评分估计器(RSE)提取可见性评分作为先验知识,利用可见性感知3D扩散修复模块(VADM)在低维时空特征中基于评分引导来修复遮挡内容,在利用3D卷积捕捉时空连贯性的同时实现靶向修复,最终利用可见性空间注意

力机制(V-SAM)引入信心引导机制,对原始特征与修复特征进行双分支融合,保证了识别特征的鲁棒性,从而提高了遮挡条件下的识别率。

在基于CASIA-B扩展的遮挡数据集上的实验结果表明,在各种遮挡模式下,本文方法与其他步态识别算法相比,该算法可以在遮挡程度较大以及复杂动态遮挡场景下,有效改善识别率。

然而,本文方法也有一些局限性。首先,本文采用了扩散概率模型进行迭代去噪修复,相比于单步生成的GAN网络,其推理速度较慢,计算开销较大,难以满足对实时性要求极高的应用场景。其次,修复模块依赖于前端生成的步态轮廓图质量,若在极端光照或背景下导致轮廓完全丢失,区域评分估计器的引导作用将受限。

在未来的工作中,将围绕上述不足进行改进。一方面探索潜在空间扩散等加速采样技术以降低计算成本,另一方面尝试引入RGB图像纹理信息辅助轮廓校正,进一步提高模型在开放场景下的泛化能力。



图9 下肢遮挡的修复对比

Fig. 9 Comparison of Restoration for Lower Limb Occlusion

## 参考文献(References)

- Chao H Q, He Y W, Zhang J P and Feng J F. 2019. GaitSet: Regarding gait as a set for cross-view gait recognition//Proceedings of the AAAI Conference on Artificial Intelligence. Honolulu, USA: AAAI Press: 8126-8133 [DOI: 10.1609/aaai.v33i01.33018126]
- Çiçek Ö, Abdulkadir A, Lienkamp S S, Brox T and Ronneberger O. 2016. 3D U-Net: Learning dense volumetric segmentation from sparse annotation//Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Athens, Greece: Springer: 424-432 [DOI: 10.1007/978-3-319-46723-8\_49]
- Fan C, Peng Y, Cao C, Liu X, Hou S, Chi J, Huang Y, Li Q and He Z. 2020. GaitPart: Temporal part-based model for gait recognition//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 14225-14233 [DOI: 10.1109/CVPR42600.2020.01423]
- Fan C, Liang J, Shen C, Hou S, Huang Y and Yu S. 2023. OpenGait: Revisiting gait recognition toward better practicality//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 9707-9716 [DOI: 10.1109/CVPR52729.2023.00936]
- Gupta A and Chellappa R. 2025. MimicGait: A model agnostic approach for occluded gait recognition using correlational knowledge distillation//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Tucson, USA: IEEE: 4757-4766 [DOI: 10.1109/WACV61041.2025.00466]
- Gupta A, Huang S and Chellappa R. 2025. Mind the Gap: Bridging occlusion in gait recognition via residual gap correction [EB/OL]. [2025-07-15]. <https://arxiv.org/abs/2507.10978> [DOI: 10.48550/arXiv.2507.10978]
- Gao Y and He M. 2023. Gait recognition algorithm in dense occlusion scene. *Optics and Precision Engineering*, 31(2): 263-276 (高毅, 何森. 2023. 密集遮挡条件下的步态识别. *光学精密工程*, 31(2): 263-276) [DOI: 10.37188/OPE.20233102.0263]
- Hasan K, Uddin M Z, Ray A, Hasan M, Alnajjar F and Ahad M A R. 2024. Improving gait recognition through occlusion detection and silhouette sequence reconstruction. *IEEE Access*, 12: 158597-158610 [DOI: 10.1109/ACCESS.2024.3482430]
- Ho J, Jain A and Abbeel P. 2020. Denoising diffusion probabilistic models//Advances in Neural Information Processing Systems. New York: Curran Associates, Inc.: 6840-6851
- Huang P, Peng Y, Hou S, Cao C, Liu X, He Z and Huang Y. 2024. Occluded gait recognition with mixture of experts: an action detection perspective//Proceedings of the European Conference on Computer Vision. Milan, Italy: Springer: 380-397 [DOI: 10.1007/978-3-031-72658-3\_22]

- Jin D, Fan C, Ma J, Zhou J, Chen W and Yu S. 2025. On denoising walking videos for gait recognition//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 12347-12357 [DOI: 10.1109/CVPR52734.2025.01152]
- Kumara S S, Chattopadhyaya P and Wang L P. 2021. BGaitR-Net: Occluded gait sequence reconstruction with temporally constrained model for gait recognition[EB/OL]. [2023-10-01]. <https://arxiv.org/abs/2110.09564> [DOI: 10.48550/arXiv.2110.09564]
- Li J X, Hao G, Wang W and Liu T Z. 2023. Application research of gait recognition technology in public security practical police work. *Police Technology*, (2): 84-86 (李俊翔, 郝刚, 王伟, 刘同中). 2023. 步态识别技术在公安实战警务工作中的应用研究. *警察技术*, (2): 84-86
- Lin B, Zhang S and Yu X. 2021. Gait recognition via effective global-local feature representation and local temporal aggregation//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 14648-14656 [DOI: 10.1109/ICCV48922.2021.01438]
- Min F, Cai Q, Guo S, Yu Y, Fan H and Dong J. 2024. ZipGait: Bridging skeleton and silhouette with diffusion model for advancing gait recognition[EB/OL]. [2024-08-22]. <https://arxiv.org/abs/2408.12111> [DOI: 10.48550/arXiv.2408.12111]
- Paul A, Jain M M, Jain J and Chattopadhyay P. 2022. Gait cycle reconstruction and human identification from occluded sequences [EB/OL]. [2023-10-01]. <https://arxiv.org/abs/2206.13395> [DOI: 10.48550/arXiv.2206.13395]
- Peng Y, Cao C and He Z. 2023. Occluded gait recognition//Proceedings of the International Joint Conference on Neural Networks. Gold Coast, Australia: IEEE: 1-8 [DOI: 10.1109/IJCNN54540.2023.10191651]
- Sun P, Gu K, Wang Y, Yang L and Yao A. 2024. Rethinking visibility in human pose estimation: occluded pose reasoning via transformers//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, USA: IEEE: 5903-5912 [DOI: 10.1109/WACV57701.2024.00580]
- Thota A K and Alberts J L. 2013. Novel use of retro-reflective paint to capture 3D kinematic gait data in non-human primates//Proceedings of the 29th Southern Biomedical Engineering Conference. Miami, USA: IEEE: 113-114 [DOI: 10.1109/SBEC.2013.65]
- Uddin M Z, Muramatsu D, Takemura N, Ahad M A R and Yagi Y. 2019. Spatio-temporal silhouette sequence reconstruction for gait recognition against occlusion. *IPSP Transactions on Computer Vision and Applications*, 11(1): 9 [DOI: 10.1186/s41074-019-0061-3]
- Wang Q, Cai J, Guo B D, Huang Y Z, Cao C S and Wang D. 2023. Research on gait recognition technology for public safety. *Journal of People's Public Security University of China (Science and Technology)*, 29(1): 68-76 (王茜, 蔡竞, 郭柏冬, 黄永祯, 曹春水, 王东). 2023. 面向公共安全的步态识别技术研究. *中国人民公安大学学报(自然科学版)*, 29(1): 68-76
- Wang Z, Hou S, Zhang M, Liu X, Cao C, Huang Y, Li P and Xu S. 2024. QAGait: Revisit gait recognition from a quality perspective//Proceedings of the AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI Press: 5785-5793 [DOI: 10.1609/aaai.v38i6.28391]
- Xiong H, Deng Y, Feng B, Wang X and Liu W. 2024. GaitGS: Temporal feature learning in granularity and span dimension for gait recognition//Proceedings of the 2024 IEEE International Conference on Image Processing. Abu Dhabi, United Arab Emirates: IEEE: 2410-2416 [DOI: 10.1109/ICIP51287.2024.10647578]
- Yang Q, Luo J and Huang Y C. 2024. Gait image spatio-temporal restoration network and its application under occlusion conditions. *Journal of Image and Graphics*, 29(1): 0179-0191 (阳强, 罗坚, 黄宇琛). 2024. 遮挡条件下的步态图像时空修复网络及其应用. *中国图象图形学报*, 29(1): 0179-0191 [DOI: 10.11834/jig.221142]
- Yu S Q, Tan D L and Tan T. 2006. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition//Proceedings of the 18th International Conference on Pattern Recognition. Hong Kong, China: IEEE: 441-444 [DOI: 10.1109/ICPR.2006.67]

### 作者简介

刘艳艳,女,硕士研究生,主要研究方向为步态识别。E-mail: 15779833106@163.com

陈欣,通信作者,女,副教授,主要研究方向为图像处理和步态识别。E-mail: chenxin@scau.edu.cn

刘文瑜,女,硕士研究生,主要研究方向为步态识别。E-mail: 3149033905@qq.com

易代豪,男,硕士研究生,主要研究方向为步态识别与大语言模型。E-mail: 623651791@qq.com