http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2025.0833

位置感知与局部预交互的空-地行人重识别方法

毕艺瀚，李冲，王蓉*
(中国人民公安大学 信息网络安全学院，北京 100038)

摘 要：针对现有空-地行人重识别方法中存在的行人特征判别性较差、模型泛化能力不足等问题，提出一种位置感知与局部预交互的空-地行人重识别方法。提出位置偏置自注意力机制，通过将位置信息融入注意力分数，引导模型聚焦输入序列中关键位置的令牌，增强模型对图块之间空间位置关系的感知能力，提升行人特征的判别性与鲁棒性；设计提示-局部预交互模块，通过一次较低成本的预交互，建立提示语义与局部特征之间的细粒度关联，增强提示向量对当前视角图像局部具体细节的感知能力，提升模型的局部细节重建能力；引入标签平滑正则化训练策略，通过将原始One-Hot编码的硬标签转换为软标签，使模型学习更平滑、泛化性更强的特征表示，缓解模型过拟合问题，提升模型的性能与泛化能力。在空-地行人重识别公开数据集LAGPeR上进行大量实验，结果表明，所提方法能够提升重识别性能，具有一定有效性。

关 键 词：空-地行人重识别；Transformer；自注意力机制；局部特征预交互；标签平滑正则化
中图分类号：TP391.4; V279
文献标志码：A **文章编号**：1001-5965(2026)09-3117-08

行人重识别(person re-identification, ReID)旨在利用计算机视觉技术，实现跨视域特定行人的准确匹配与有效追踪[1]，在智能监控、安全防范等领域得到了广泛应用[2]。随着深度学习的发展，现有研究在传统行人重识别领域取得了显著进展，一定程度上解决了光照变化[3]、目标遮挡[4-5]与低图像分辨率差异[6-7]等问题。然而，这些研究主要基于单一平台摄像头网络采集的数据集，通常由固定的地面摄像头或闭路电视系统组成。这一特点使得这些方法在实际场景中的性能高度依赖于地面摄像头的部署密度与覆盖范围。在城市等人流密集的区域，密集的摄像头覆盖有助于实现有效的检索；而在市郊、农田、森林等人流稀疏的区域，稀疏的摄像头部署会显著降低追踪效果，可能导致目标踪迹丢失。

随着无人机(unmanned aerial vehicle, UAV)的快速发展，在无人机上部署摄像头为提升摄像头的覆盖密度提供了一种灵活的解决方案。同时，这种方法还能增强智能监控的态势感知能力[8]。因此，研究在监控网络中同时结合地面摄像头与无人机摄像头的空-地行人重识别(aerial-ground person re-identification, AGPReID)，更符合实际监控场景的需求。

区别于传统行人重识别，空-地行人重识别的核心挑战在于跨摄像头视角变化导致的显著视觉差异，这些视觉差异会阻碍鲁棒性行人特征的提取，降低行人图像匹配的准确性。目前，一些研究尝试解决该问题。Nguyen等[9]第1次从深度学习的视角对空-地行人重识别任务进行研究，提出了一种利用软属性引导模型训练的空-地行人重识别方法，通过利用行人性别、上衣颜色、背包类型等语义属性信息，使模型有效地学习更具身份判别性的行人特征，增强模型对空-地场景中行人外观显著变化的适应能力；同时，构建了空-地行人重识别公开数据集AG-ReID，为空-地行人重识别研究提供了重要的基准。贝俊仁等[10]针对空、地视角差异导致的行人外观显著变化与不对齐问题，提出一种基于隐式解码对齐的空-地行人重识别方法，通过

收稿日期：2025-12-02；录用日期：2025-12-26；网络出版时间：2026-03-03 11:32
网络出版地址：link.cnki.net/urlid/11.2625.V.20260303.0942.001
基金项目：中央高校基本科研业务费专项资金 (2024JKF12)
*通信作者. E-mail：dbdxwangrong@163.com

引用格式：毕艺瀚，李冲，王蓉. 位置感知与局部预交互的空-地行人重识别方法[J]. 北京航空航天大学学报，2026，52(9)：3117-3124.
Bi Y H，Li C，Wang R. Aerial-ground person re-identification method via position-aware and local pre-interaction[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(9)：3117-3124 (in Chinese).

深度特征学习自动解码并对齐地面摄像头视角与无人机摄像头视角下行人图像的关键语义信息，无需额外的姿态或部件标注即可实现跨视角的鲁棒性行人特征匹配，有效提升了模型在空-地场景中的特征判别能力。Wang等[11]利用扩散模型的生成能力增强模型对视角变化的鲁棒性，提出一种基于视角感知稳定扩散的空-地行人重识别方法，通过引入视角感知的条件控制机制，并利用文本或视角标签引导稳定扩散模型生成更具身份一致性的跨视角行人图像，弥补真实数据中视角样本的不足，并学习视角不变的特征表示，有效提升了模型在跨视角场景中的判别能力与泛化能力。Mei等[12]针对无人机监控场景中标注数据稀缺的问题，提出了一种基于无监督学习的空-地行人重识别框架，通过引入邻域感知协作学习与渐进图匹配策略，以无监督学习的方式挖掘跨视角行人图像之间的隐含关联；同时，通过设计协作跨模态关联学习模块缓解特征差异、实现模态之间的软对齐，并利用最小行人距离转换策略量化地面摄像头视角与无人机摄像头视角之间的最优行人相似度，在空-地行人重识别公开数据集上取得了较为先进的性能。

目前，空-地行人重识别的研究仍处于起步阶段。上述方法虽然在一定程度上解决了地面摄像头视角与无人机摄像头视角下行人图像的匹配难题，但是仍存在行人特征判别性较差、模型泛化能力不足等问题。

针对上述问题，本文提出一种位置感知与局部预交互的空-地行人重识别方法。该方法使用视觉Transformer[13]（vision Transformer, ViT）作为主干网络，首先，提出位置偏置自注意力机制，提升行人特征的判别性与鲁棒性；其次，设计提示-局部预交互模块，提升模型的局部细节重建能力；最后，引入标签平滑正则化训练策略，缓解模型过拟合问题，提升模型的泛化能力。本文的主要创新点如下：

1）提出位置偏置自注意力机制（positionally biased self-attention mechanism, PBSAM），通过在标准自注意力机制的查询矩阵上添加一个可学习的偏置向量，将位置信息融入注意力分数，引导模型聚焦输入序列中关键位置的令牌，增强模型对图块之间空间位置关系的感知能力，提升行人特征的判别性与鲁棒性。

2）设计提示-局部预交互模块（prompt-local pre-interaction module, PLPIM），通过将提示向量与局部特征进行快速对齐，建立提示语义与局部特征之间的细粒度关联，增强提示向量对当前视角图像局部具体细节的感知能力，提升模型的局部细节重建能力。

3）将标签平滑正则化[14]（label smoothing regularization, LSR）训练策略引入交叉熵损失，通过引入一个平滑因子，将原始One-Hot编码的硬标签转换为软标签，使模型学习更平滑、泛化性更强的特征表示，减少模型对训练数据的过度自信、缓解模型过拟合问题，提升模型的性能与泛化能力。

1 本文方法

如图1所示，本文提出的位置感知与局部预交互的空-地行人重识别方法采用编码器-解码器的Transformer框架。编码器采用视觉解耦的Transformer[15]（view-decoupled Transformer, VDT）网络。相较于传统的ViT网络，本文方法通过引入视角令牌，并在每层对类别令牌进行分层解耦，有效分离类别令牌中的视角相关特征与视角不变特征，同时从输入中提取局部特征。解码器由提示重校准模块[16]（prompt re-calibration module, PRM）、提示-局部预交互模块与局部特征优化模块[16]（local feature refinement module, LFRM）组成。其中，PRM基于当前视角信息自适应生成并重新校准不同视角的提示向量；PLPIM将重新校准的提示向量与当前视角的局部特征进行预交互；LFRM利用PLPIM预交互的提示向量对局部特征进行解码。

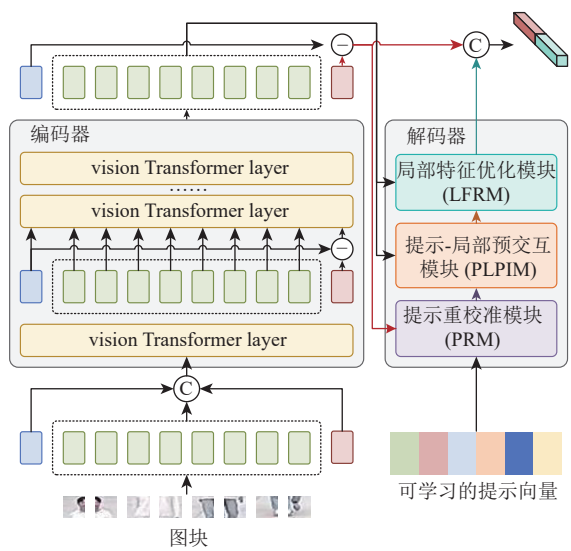

图1 本文方法框架

Fig. 1 Architecture of the proposed method

1.1 位置偏置自注意力机制

为提取更具判别性与鲁棒性的行人特征，本文设计了位置偏置自注意力机制，如图2所示，对Transformer编码器中的自注意力机制进行了关键改进。在标准自注意力机制中，首先，输入经过QKV线性层生成查询矩阵、键矩阵与值矩阵；其次，通过查询矩阵与键矩阵的转置的点积运算，计算注意力分数；然后，利用softmax函数将注意力分数归一化为注意力权重；最后，将注意力权重与值

矩阵相乘, 再经过 Projection 线性层投影得到最后的输出:

$$\mathrm{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{softmax}\left(\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}/\sqrt{d_{\mathrm{k}}}\right)\boldsymbol{V} \tag{1}$$

式中: $\boldsymbol{Q}$ 为查询矩阵; $\boldsymbol{K}$ 为键矩阵; $\boldsymbol{V}$ 为值矩阵; d_{k} 为键的维度。

如图 2 所示, 位置偏置自注意力机制在标准自注意力机制中引入了位置偏置。如式 (2) 所示, 作为一种高效的位置感知机制, 位置偏置通过在查询矩阵上添加一个可学习的偏置向量, 将位置信息融入注意力分数, 引导模型聚焦输入序列中关键位置的令牌, 增强模型对图块之间空间位置关系的感知能力, 提升行人特征的判别性与鲁棒性。

$$\tilde{\boldsymbol{Q}}=\boldsymbol{Q}+\boldsymbol{P} \tag{2}$$

式中: $\tilde{\boldsymbol{Q}}$为添加位置偏置向量后的查询矩阵; $\boldsymbol{P}$ 为位置偏置。

由式(2)可得引入位置偏置后的自注意力机制:

$$\mathrm{Attention}\left(\tilde{\boldsymbol{Q}},\boldsymbol{K},\boldsymbol{V}\right)=\mathrm{softmax}\left((\boldsymbol{Q}+\boldsymbol{P})\boldsymbol{K}^{\mathrm{T}}/\sqrt{d_{\mathrm{k}}}\right)\boldsymbol{V} \tag{3}$$

将引入位置偏置后自注意力机制的注意力分数拆分为两项:

$$(\boldsymbol{Q}+\boldsymbol{P})\boldsymbol{K}^{\mathrm{T}}/\sqrt{d_{\mathrm{k}}}=\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}/\sqrt{d_{\mathrm{k}}}+\boldsymbol{P}\boldsymbol{K}^{\mathrm{T}}/\sqrt{d_{\mathrm{k}}} \tag{4}$$

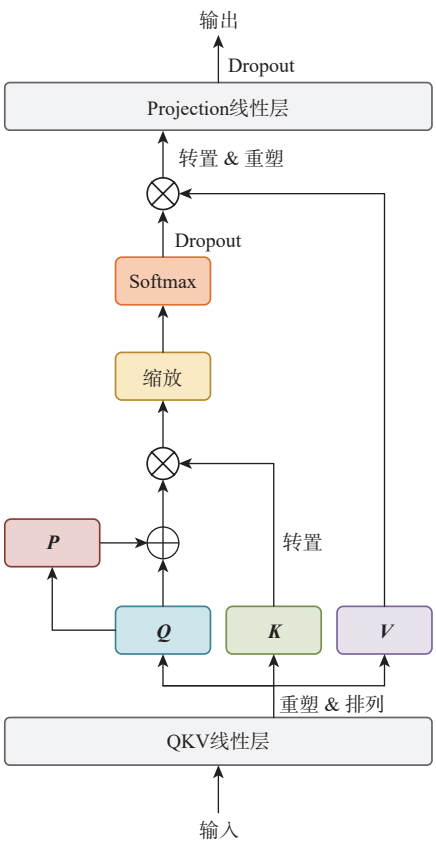

图 2　位置偏置自注意力机制结构

Fig. 2　Architecture of the positionally biased self-attention mechanism

式中: 第 1 项为标准自注意力机制的注意力分数, 即基于内容的寻址; 第 2 项为位置偏置引入的位置相关性分数, 即基于位置的寻址。在第 2 项中, 不考虑缩放因子 $1/\sqrt{d_{\mathrm{k}}}$ 的约束, 通过位置偏置与键矩阵的转置的点积运算, 得到位置相关性分数矩阵:

$$\boldsymbol{S}_{\mathrm{pos}}=\boldsymbol{P}\boldsymbol{K}^{\mathrm{T}} \tag{5}$$

使用 $\boldsymbol{S}_{\mathrm{pos},ij}$ 衡量第 i 个查询向量与第 j 个键向量的位置相关性后, 在模型的反向传播阶段, 损失函数 L 对位置偏置的梯度计算式为

$$\frac{\partial L}{\partial \boldsymbol{P}}=\sum_{i,j}\frac{\partial L}{\partial A_{ij}}\cdot\frac{\partial A_{ij}}{\partial \boldsymbol{S}_{\mathrm{pos},ij}}\cdot\frac{\partial \boldsymbol{S}_{\mathrm{pos},ij}}{\partial \boldsymbol{P}} \tag{6}$$

式中: A_{ij} 为相应的注意力权重。可以看出, 随着位置偏置逐批次更新, 模型动态优化位置相关性分数, 引导注意力权重聚焦行人判别性区域, 从而最小化损失。

1.2　提示-局部预交互模块

为提升模型的局部细节重建能力, 本文设计了提示-局部预交互模块, 如图 3 所示。PLPIM 通过将重新校准的提示向量与当前视角的局部特征进行快速对齐, 使提示向量在输入 LFRM 之前与当前视角的局部特征建立细粒度关联, 使 LFRM 能够专注于更精细的特征融合与解码, 优化解码器的训练动态与计算效率, 提升模型重建局部细节的质量与对视角变化的鲁棒性。

PRM 利用全局视角信息与跨视角差分信息对可学习的提示向量进行自适应重新校准, 导致提示向量缺乏对当前视角图像局部具体细节的直接感知能力。在 PLPIM 中, 首先, 将提示向量投影为查询矩阵、位置增强的局部特征投影为键矩阵、原始局

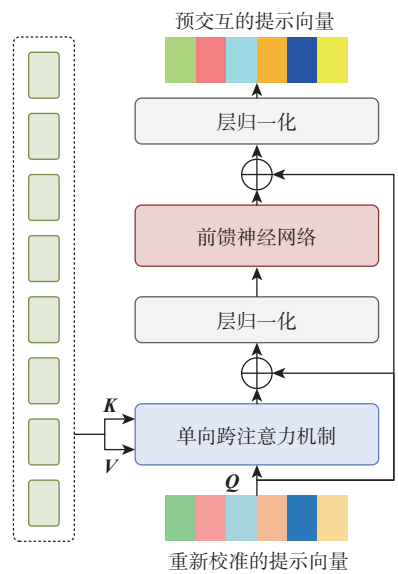

图 3　提示-局部预交互模块结构

Fig. 3　Architecture of the prompt-local pre-interaction module

部特征投影为值矩阵，使用单向跨注意力机制建立提示向量与当前视角局部特征之间的细粒度关联：

$$\mathrm{Attn}(\boldsymbol{P},\boldsymbol{X})=\mathrm{softmax}\left\{(\boldsymbol{P}\boldsymbol{W}_{\boldsymbol{Q}})[(\boldsymbol{X}+\boldsymbol{P}_{\mathrm{E}})\boldsymbol{W}_{\boldsymbol{K}}]^{\mathrm{T}}/\sqrt{d_{\mathrm{k}}}\right\}(\boldsymbol{X}\boldsymbol{W}_{\boldsymbol{V}})\boldsymbol{W}_{\mathrm{O}} \tag{7}$$

式中：$\boldsymbol{P}$为提示向量；$\boldsymbol{X}$为局部特征；$\boldsymbol{P}_{\mathrm{E}}$为位置编码；$\boldsymbol{W}_{\boldsymbol{Q}}$为查询投影矩阵；$\boldsymbol{W}_{\boldsymbol{K}}$为键投影矩阵；$\boldsymbol{W}_{\boldsymbol{V}}$为值投影矩阵；$\boldsymbol{W}_{\mathrm{O}}$为输出投影矩阵。

其次，将跨注意力机制的输出与输入PLPIM的提示向量进行残差连接，并使用层归一化进行处理；然后，利用前馈神经网络（feed-forword neural network, FFN）对归一化后的提示向量进行非线性增强：

$$\mathrm{FFN}(\boldsymbol{x})=\boldsymbol{W}_2\cdot\mathrm{GELU}(\boldsymbol{W}_1\cdot\boldsymbol{x}+\boldsymbol{b}_1)+\boldsymbol{b}_2 \tag{8}$$

式中：$\boldsymbol{x}$为输入FFN的提示向量；$\boldsymbol{W}_1$为FFN中第1个线性层的权重矩阵；$\boldsymbol{b}_1$为FFN中第1个线性层的偏置向量；$\boldsymbol{W}_2$为FFN中第2个线性层的权重矩阵；$\boldsymbol{b}_2$为FFN中第2个线性层的偏置向量。

最后，将FFN的输出与输入PLPIM的提示向量进行残差连接，再经过层归一化处理得到既保留原始语义，又融合当前视角图像局部具体细节的提示向量，如图3所示。

PLPIM通过一次较低成本的预交互，建立提示语义与局部特征之间的关联，增强提示向量对当前视角图像局部具体细节的感知能力，提升模型的局部细节重建能力。

1.3 标签平滑正则化

为缓解模型过拟合问题，提升模型的泛化能力，本文将标签平滑正则化引入交叉熵损失。标签平滑正则化是分类任务中一种广泛使用的训练策略，通过将原始One-Hot编码的硬标签转换为软标签，使模型学习更平滑、泛化性更强的特征表示，减少模型对训练数据的过度自信、缓解模型过拟合问题，提升模型的性能与泛化能力。

1.3.1 交叉熵损失

作为图像分类任务中一种广泛使用的损失函数，交叉熵损失通过度量模型的预测结果与真实数据之间的相似程度评估模型的预测性能。交叉熵损失的计算式为

$$L_{\text{cross-entropy}}=-\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{M}y_{i,k}\ln p_{i,k} \tag{9}$$

式中：M和N分别为类别数量与样本数量；$p_{i,k}$为第i个样本预测为类别k的概率；$y_{i,k}$为第i个样本在第k个类别上的真实标签。

1.3.2 引入标签平滑正则化的交叉熵损失

标准交叉熵损失约束模型对真实类别预测出趋近于1的概率，这种训练目标容易使模型对预测结果过度自信，并使模型学习的特征表示过于尖锐，导致模型的泛化能力不足。

标签平滑正则化通过引入一个平滑因子，对真实类别的概率进行重新分配，即在真实类别上分配略小于1的概率，同时在其余类别上均匀分配少量概率，引导模型在训练过程中关注样本的整体分布，而非仅仅依赖严格的硬标签，使模型在类别之间学习更具鲁棒性的特征表示，缓解模型过拟合问题，提升模型的泛化能力。引入标签平滑正则化的交叉熵损失计算式为

$$L_{\text{cross-entropy}}^{\text{LSR}}=-\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{M}\left[(1-\varepsilon)y_{i,k}+\frac{\varepsilon}{M}\right]\ln p_{i,k} \tag{10}$$

式中：ε为平滑因子。

2 实验与分析

2.1 数据集与评价指标

2.1.1 数据集

在空-地行人重识别公开数据集LAGPeR[16]上对本文方法进行实验验证。LAGPeR数据集由14个地面摄像头与7个无人机摄像头采集得到，包含4 231个行人的63 841张图像。其中，训练集包含2 708个行人的40 770张图像；测试集包含1 523个行人，选择3 046张图像作为待查询图像，其余图像作为查询库图像。同时，构建了一个全面的空-地行人重识别任务设置，包括空对地（aerial to ground, A→G）、地对空（ground to aerial, G→A）与地对全（ground to aerial+ground, G→A＋G）3种实验设置。LAGPeR数据集的基本情况如表1所示。

表1 LAGPeR数据集情况介绍

Table 1 Introduction to LAGPeR dataset

实验设置	子集	视角	摄像头个数	行人个数	图像张数
A→G	Train	A+G	12	2 708	40 770
	Query	A	3	1 523	3 046
	Gallery	G	6	1 523	15 533
G→A	Query	G	6	1 523	3 046
	Gallery	A	3	1 523	7 717
G→A+G	Query	G	6	1 523	3 046
	Gallery	A+G	9	1 523	20 204

2.1.2 评价指标

使用CMC曲线[17]、mAP[18]与mINP[19]作为模型性能的评价指标，对本文方法进行实验验证。其中，CMC曲线通过计算在不同排序的匹配结果中，对待查询的目标行人进行匹配的准确率，较为直观地反映模型的性能；mAP通过计算在所有查询的匹配结果中，每次对待查询的目标行人进行匹配的平均准确率，较为全面地反映模型的性能；mINP通过

计算在所有查询的匹配结果中, 对每个待查询的目标行人进行所有正确匹配需克服的平均负样本难度, 反映模型匹配最困难样本的效率。

2.2　实验环境与参数设置

在 GEFORCE RTX 4090 GPU 上使用 PyTorch 1.4.0 深度学习框架, 对本文方法进行实验验证。模型使用 ViT 作为主干网络, 并使用在 ImageNet[20] 上预训练的参数。模型训练阶段, 输入图像的尺寸设置为 256 像素×128 像素, 总周期数设置为 120, 批量大小设置为 64, 学习率设置为 8×10^{-3}, 平滑因子设置为 0.1。使用 SGD 优化器[21] 对模型参数进行更新, 权重衰减参数设置为 1×10^{-4}。

2.3　实验结果与分析

2.3.1　对比实验

为验证本文方法的先进性, 在 LAGPeR 数据集上进行对比实验, 结果如表 2 所示。表中加粗数据为相应实验设置下的最优结果。可以看出, 在 A→G 实验设置下, 模型的 Rank-1 准确率(在所有测试查询中, 模型返回的相似度最高(排名第 1)的那张恰好是正确目标的概率)达到了 41.99%、mAP 达到了 30.52%、mINP 达到了 12.32%, 均取得了最优结果; 在 G→A 实验设置下, 模型的 Rank-1 准确率达到了 34.93%、mAP 达到了 33.23%、mINP 达到了 24.02%, 均取得了最优结果; 在 G→A+G 实验设置下, 模型的 Rank-1 准确率达到了 22.23%、mAP 达到了 18.17%, 取得了与最优结果接近的次优结果, mINP 达到了 7.24%, 取得了最优结果。综合来看, 本文方法在 LAGPeR 数据集的 3 种实验设置下表现最优, 表明本文方法能够提升空-地行人重识别性能, 具有一定先进性。

表 2　在 LAGPeR 数据集上的对比实验结果

Table 2　Comparative experimental results on LAGPeR dataset　%

方法	主干网络	Rank-1			mAP			mINP		
		A→G	G→A	G→A+G	A→G	G→A	G→A+G	A→G	G→A	G→A+G
ViT[13]	ViT	38.67	32.04	18.88	27.25	30.69	15.31			
TransReID[22]	ViT	38.80	33.00	**22.90**	28.80	32.10	**18.80**			
CLIP-ReID[23]	CLIP	24.40	21.30	12.30	17.60	20.80	10.20			
MIP[24]	ViT	39.30	33.90	21.00	29.30	32.60	17.30			
AG-ReID[9]	ViT	40.48	32.96	22.03	28.89	31.91	17.89			
VDT[15]	ViT	40.15	33.55	19.50	28.97	31.98	16.45			
SeCap[16]	ViT	39.89	33.68	21.04	28.67	32.00	16.80	11.16	23.04	6.59
本文方法	ViT	**41.99**	**34.93**	22.23	**30.52**	**33.23**	18.17	**12.32**	**24.02**	**7.24**

2.3.2　消融实验

为验证 3 个模块的有效性, 在 LAGPeR 数据集上进行消融实验, 结果如表 3 所示。可以看出, 使用位置偏置自注意力机制后, 在 A→G 实验设置下, 模型的 Rank-1 准确率提升了 0.39%、mAP 提升了 0.53%、mINP 提升了 0.28%; 在 G→A 实验设置下, 模型的 Rank-1 准确率降低了 0.39%、mAP 提升了 0.12%、mINP 提升了 0.14%; 在 G→A+G 实验设置下, 模型的 Rank-1 准确率提升了 0.53%、mAP 提升了 0.62%、mINP 提升了 0.20%, 表明位置偏置自注意力机制能够提升行人特征的判别性与鲁棒性。加入提示-局部预交互模块后, 在 A→G 实验设置下, 模型的 Rank-1 准确率提升了 1.38%、mAP 提升了 1.05%、mINP 提升了 0.72%; 在 G→A 实验设置下, 模型的 Rank-1 准确率提升了 1.38%、mAP 提升了 0.69%、mINP 提升了 0.31%; 在 G→A+G 实验设置下, 模型的 Rank-1 准确率提升了 0.92%、mAP 提升了 0.49%、mINP 提升了 0.03%, 表明提示-局部预交互模块能够提升模型的局部细节重建能力。引入标签平滑正则化训练策略后, 在 A→G 实验设置下, 模型的 Rank-1 准确率降低了 0.10%、mAP 提升了 0.29%、mINP 降低了 0.08%; 在 G→A 实验设置下, 模型的 Rank-1 准确率提升了 0.56%、mAP 提升了 0.34%、mINP 提升了 0.35%; 在 G→A+G 实验设置下, 模型的 Rank-1 准确率提升了 0.27%、mAP 提升了 0.42%、mINP 提升了 0.43%, 表明标签平滑正则化训练策略能够缓解模型过拟合问题, 提升模型的泛化能力。同时使用位置偏置自注意力机制、加入提示-局部预交互模块、引入标签平滑正则化训练策略后, 在 A→G 实验设置下, 模型的 Rank-1 准确率提升了 2.10%、mAP 提升了 1.85%、mINP 提升了 1.16%, 均取得了最优结果; 在 G→A 实验设置下, 模型的 Rank-1 准确率提升了 1.25%, 取得了与最优结果接近的次优结果, mAP 提升了 1.23%、mINP 提升了 0.98%, 取得了最优结果; 在 G→A+G 实验设置下, 模型的 Rank-1 准确率提升了 1.19%、mAP 提升了 1.37%、mINP 提升了 0.65%, 均取得了最优结果。综合来看, 本文方法在 LAGPeR 数据集的 3 种实验设置下表现最优(见表中加粗数据), 表明 3 个模块协同工作能够提升空-地行人重识别性能,

表 3 在 LAGPeR 数据集上的消融实验结果

Table 3 Results of ablation study on LAGPeR dataset %

PBSAM	PLPIM	LSR	Rank-1			mAP			mINP		
			A→G	G→A	G→A+G	A→G	G→A	G→A+G	A→G	G→A	G→A+G
			39.89	33.68	21.04	28.67	32.00	16.80	11.16	23.04	6.59
√			40.28	33.29	21.57	29.20	32.12	17.42	11.44	23.18	6.79
	√		41.27	**35.06**	21.96	29.72	32.69	17.29	11.88	23.35	6.62
		√	39.79	34.24	21.31	28.96	32.34	17.22	11.08	23.39	7.02
√	√		41.92	34.50	21.50	29.74	33.03	17.40	11.48	23.93	6.80
√	√	√	**41.99**	34.93	**22.23**	**30.52**	**33.23**	**18.17**	**12.32**	**24.02**	**7.24**

具有一定有效性。

2.3.3 可视化分析

本文方法在 LAGPeR 数据集上的 CMC 曲线如图 4 所示。可以看出，同时使用 PBSAM、PLPIM、LSR 训练策略后，本文方法在 LAGPeR 数据集的 3 种实验设置下表现最优。

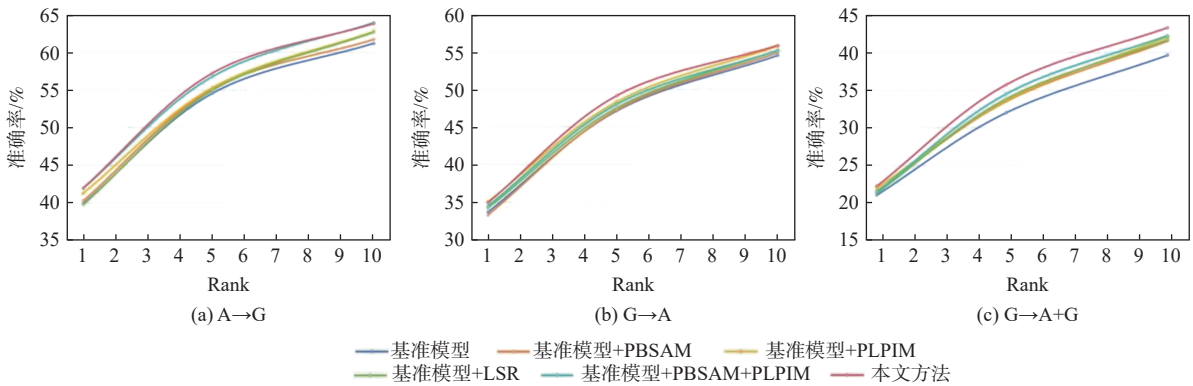

图 4 在 LAGPeR 数据集 3 种实验设置下的 CMC 曲线

Fig. 4 CMC curves under 3 experimental settings of LAGPeR dataset

本文方法在 LAGPeR 数据集上的可视化结果如图 5 所示。其中，Query 表示待查询的目标行人图像，1～5 表示使用本文方法得到的前 5 位匹配结果。绿色边框表示该匹配结果与待查询的目标行人是同一个行人，红色边框表示该匹配结果与待查询的目标行人不是同一个行人。



图 5 在 LAGPeR 数据集上的可视化结果

Fig. 5 Visualization results on LAGPeR dataset

可以看出，对于大部分空-地行人图像（ID 1 至 ID 4），本文方法基本能够匹配所有正确的样本。然而，对于更具挑战性的小部分空-地行人图像（ID 5 与 ID 6），如质量较差或存在大面积物体遮挡的空-地行人图像，本文方法会混淆一些难以判别的样本。这主要是由于图像噪声、物体遮挡等因素导致模型难以从空-地行人图像中提取行人关键特征。

3 结 论

本文提出一种位置感知与局部预交互的空-地行人重识别方法，旨在解决现有空-地行人重识别方法中存在的行人特征判别性较差、模型泛化能力不足等问题，主要结论如下：

1）位置偏置自注意力机制能够通过在注意力分数中融入位置信息，引导模型聚焦输入序列中关键位置的令牌，增强模型对图块之间空间位置关系的感知能力。

2）提示-局部预交互模块能够通过快速对齐提示向量与局部特征，建立提示语义与局部特征之间

的细粒度关联, 增强提示向量对当前视角图像局部具体细节的感知能力, 提升模型的局部细节重建能力。

3) 标签平滑正则化训练策略能够通过将原始 One-Hot 编码的硬标签转换为软标签, 使模型学习更平滑、泛化性更强的特征表示, 缓解模型过拟合问题, 提升模型的性能与泛化能力。

4) 本文方法能够实现较为优异的空-地行人重识别性能: 在 A→G 与 G→A 实验设置下, 模型的 Rank-1 准确率、mAP 与 mINP 均取得了最优结果; 在 G→A+G 实验设置下, 模型的 Rank-1 准确率与 mAP 取得了接近最优结果的次优结果、mINP 取得了最优结果。

参考文献（References）

[1] 孙义博, 张文靖, 王蓉, 等. 基于通道注意力机制的行人重识别方法[J]. 北京航空航天大学学报, 2022, 48(5): 881-889.
Sun Y B, Zhang W J, Wang R, et al. Pedestrian re-identification method based on channel attention mechanism[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(5): 881-889(in Chinese).

[2] 宋晓勇, 孙学宏, 刘丽萍, 等. 基于多细粒度双流网络的行人重识别[J]. 电子测量与仪器学报, 2025, 39(8): 250-257.
Song X Y, Sun X H, Liu L P, et al. Pedestrian re-identification based on a multi-granularity dual-stream network[J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(8): 250-257 (in Chinese).

[3] Huang Y K, Zha Z J, Fu X Y, et al. Illumination-invariant person re-identification[C]//Proceedings of the 27th ACM International Conference on Multimedia. New York: ACM, 2019: 365-373.

[4] Hou R B, Ma B P, Chang H, et al. VRSTC: occlusion-free video person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 7176-7185.

[5] Huang H J, Li D W, Zhang Z, et al. Adversarially occluded samples for person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 5098-5107.

[6] Li X, Zheng W S, Wang X J, et al. Multi-scale learning for low-resolution person re-identification[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2016: 3765-3773.

[7] Wang Y, Wang L Q, You Y R, et al. Resource aware person re-identification across multiple resolutions[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 8042-8051.

[8] Schumann A, Metzler J. Person re-identification across aerial and ground-based cameras by deep feature fusion[J]. Automatic Target Recognition XXVII, 2017, 10202: 102020A.

[9] Nguyen H, Nguyen K, Sridharan S, et al. Aerial-ground person re-ID[C]//Proceedings of the IEEE International Conference on Multimedia and Expo. Piscataway: IEEE Press, 2023: 2585-2590.

[10] 贝俊仁, 张权, 赖剑煌. 基于隐式解码对齐的空地行人重识别方法[J]. 自动化学报, 2025, 51(9): 1988-2000.
Bei J R, Zhang Q, Lai J H. Implicit decoder alignment for aerial-ground person re-identification[J]. Acta Automatica Sinica, 2025, 51(9): 1988-2000(in Chinese).

[11] Wang Y H, Hu X, Wang L X, et al. SD-ReID: view-aware stable diffusion for aerial-ground person re-identification[EB/OL]. (2026-05-14)[2026-05-29]. https://arxiv.org/abs/2504.09549.

[12] Mei L, Cheng Y W, Chen H X, et al. Unsupervised aerial-ground re-identification from pedestrian to group for UAV-based surveillance[J]. Drones, 2025, 9(4): 244.

[13] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[EB/OL]. (2021-06-03)[2025-12-02]. https://arxiv.org/abs/2010.11929.

[14] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 2818-2826.

[15] Zhang Q, Wang L, Patel V M, et al. View-decoupled transformer for person re-identification under aerial-ground camera network[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2024: 22000-22009.

[16] Wang S N, Wang Y L, Wu R Q, et al. SeCap: self-calibrating and adaptive prompts for cross-view person re-identification in aerial-ground networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2025: 22119-22128.

[17] Wang X G, Doretto G, Sebastian T, et al. Shape and appearance context modeling[C]//Proceedings of the IEEE 11th International Conference on Computer Vision. Piscataway: IEEE Press, 2007: 1-8.

[18] Zheng L, Shen L Y, Tian L, et al. Scalable person re-identification: a benchmark[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2016: 1116-1124.

[19] Ye M, Shen J B, Lin G J, et al. Deep learning for person re-identification: a survey and outlook[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(6): 2872-2893.

[20] Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2009: 248-255.

[21] Kingma D P, BA J. Adam: a method for stochastic optimization [EB/OL].(2017-01-30)[2025-12-02].https://arxiv.org/abs/1412.6980.

[22] He S T, Luo H, Wang P C, et al. TransReID: Transformer-based object re-identification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2022: 14993-15002.

[23] Li S Y, Sun L, Li Q L. CLIP-ReID: exploiting vision-language model for image re-identification without concrete text labels[EB/OL]. (2023-01-01)[2025-12-02]. https://arxiv.org/abs/2211.13977.

[24] Wu R Q, Jiao B L, Wang W X, et al. Enhancing visible-infrared person re-identification with modality- and instance-aware visual prompt learning[EB/OL]. (2024-06-18)[2025-12-02]. https://arxiv.org/abs/2406.12316.

Aerial-ground person re-identification method via position-aware and local pre-interaction

BI Yihan，LI Chong，WANG Rong*

(School of Information and Cyber Security，People's Public Security University of China，Beijing 100038，China)

Abstract：In response to the issues of poor discriminative power in pedestrian features and insufficient model generalization in existing aerial-ground person re-identification methods, this paper proposes an aerial-ground person re-identification method via position-aware and local pre-interaction. A positionally biased self-attention mechanism is proposed, which incorporates positional information into attention scores to guide the model in focusing on tokens at key positions within the input sequence. This enhances the model's perception of spatial relationships between image patches and improves the discriminability and robustness of pedestrian features. A prompt-local pre-interaction module is created that uses a single, inexpensive pre-interaction to create fine-grained linkages between prompt semantics and local attributes. This strengthens the prompt vector's ability to perceive specific local details in the current view and enhances the model's capability for fine-grained detail reconstruction. The label smoothing regularization training strategy is introduced, converting original one-hot encoded hard labels into soft labels. This encourages the model to learn smoother and more generalizable feature representations, mitigates overfitting, and improves overall model performance and generalization. The validity of the proposed strategy is confirmed by extensive trials on the public aerial-ground person re-identification dataset LAGPeR, which show that it effectively enhances Re-Identification performance.

Keywords：aerial-ground person re-identification；Transformer；self-attention mechanism；local feature pre-interaction；label smoothing regularization

Received：2025-12-02；**Accepted**：2025-12-26；**Published Online**：2026-03-03 11：32

URL：link.cnki.net/urlid/11.2625.V.20260303.0942.001

Foundation item：The Fundamental Research Funds for the Central Universities (2024JKF12)

* **Corresponding author**. E-mail：dbdxwangrong@163.com