智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

# 多粒度遮挡特征增强的行人搜索算法

苗春玲, 张红云, 吴卓嘉, 张齐贤, 苗夺谦

引用本文:
苗春玲, 张红云, 吴卓嘉, 等. 多粒度遮挡特征增强的行人搜索算法[J]. 智能系统学报, 2025, 20(1): 230-242.
MIAO Chunling, ZHANG Hongyun, WU Zhuojia, et al. Multi-granularity occlusion feature enhancement algorithm for person search[J]. *CAAI Transactions on Intelligent Systems*, 2025, 20(1): 230-242.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202407031

**您可能感兴趣的其他文章**

基于注意力机制的显著性目标检测方法
Salient object detection method based on the attention mechanism
智能系统学报. 2020, 15(5): 956–963 https://dx.doi.org/10.11992/tis.201903001

面向自动驾驶目标检测的深度多模态融合技术
Deep multi-modal fusion in object detection for autonomous driving
智能系统学报. 2020, 15(4): 758–771 https://dx.doi.org/10.11992/tis.202002010

高斯核函数卷积神经网络跟踪算法
Convolutional neutral network tracking algorithm accelerated by Gaussian kernel function
智能系统学报. 2018, 13(3): 388–394 https://dx.doi.org/10.11992/tis.201612040

深度学习在无人驾驶汽车领域应用的研究进展
Deep learning in driverless vehicles
智能系统学报. 2018, 13(1): 55–69 https://dx.doi.org/10.11992/tis.201609029

基于自编码器的特征迁移算法
Feature transfer algorithm based on an auto-encoder
智能系统学报. 2017, 12(6): 894–898 https://dx.doi.org/10.11992/tis.201706037

行人重识别研究综述
Survey on pedestrian re-identification research
智能系统学报. 2017, 12(6): 770–780 https://dx.doi.org/10.11992/tis.201706084

**DOI**: 10.11992/tis.202407031
**网络出版地址**: https://link.cnki.net/urlid/23.1538.TP.20241203.1646.002

# 多粒度遮挡特征增强的行人搜索算法

苗春玲[1,2], 张红云[1,2], 吴卓嘉[1,2], 张齐贤[1,2], 苗夺谦[1,2]
（1. 同济大学 电子与信息工程学院, 上海 201804; 2. 同济大学 嵌入式系统与服务计算教育部重点实验室, 上海 201804）

**摘　要:** 现有行人搜索方法着重于从有限的标注场景图中学习有效的行人表征, 虽然这些方法取得了一定的效果, 但学习更具有身份辨别力的行人表征通常依赖于大规模的标注数据, 而获取大规模的标注数据是一个资源、劳动密集型的过程。为此, 该文提出了一种场景图多粒度遮挡特征增强算法, 对原始场景图进行多粒度随机遮挡, 扩充训练数据, 并从遮挡后的场景图中生成具有多样化信息的虚拟特征, 最后利用生成的虚拟特征增强真实特征中的行人表征。进一步, 基于生成对抗学习, 该文设计了多粒度特征对齐模块, 用于对齐遮挡图像特征和原始图像特征, 保持两者语义一致性。实验结果表明, 在 CUHK-SYSU 和 PRW 数据集上, 该算法能够显著提升行人搜索任务的搜索精度。

**关键词:** 深度学习; 计算机视觉; 行人搜索; 目标检测; 粒计算; 数据处理; 特征提取; 生成对抗网络; 对齐
**中图分类号:** TP389.1　　**文献标志码:** A　　**文章编号:** 1673−4785(2025)01−0230−13

**中文引用格式: 苗春玲, 张红云, 吴卓嘉, 等. 多粒度遮挡特征增强的行人搜索算法 [J]. 智能系统学报, 2025, 20(1): 230–242.**
**英文引用格式: MIAO Chunling, ZHANG Hongyun, WU Zhuojia, et al. Multi-granularity occlusion feature enhancement algorithm for person search[J]. CAAI transactions on intelligent systems, 2025, 20(1): 230–242.**

# Multi-granularity occlusion feature enhancement algorithm for person search

MIAO Chunling[1,2], ZHANG Hongyun[1,2], WU Zhuojia[1,2], ZHANG Qixian[1,2], MIAO Duoqian[1,2]
(1. College of Electronic and Information Engineering, Tongji University, Shanghai 201804, China; 2. Key Laboratory of Embedded System and Service Computing Ministry of Education, Tongji University, Shanghai 201804, China)

**Abstract**: Existing person search methods focus on efficiently learning pedestrian representations from limited labeled scene images. Although these methods have achieved good results, learning more identity-discriminative pedestrian representations usually relies on large-scale labeled images, while obtaining large-scale labeled data is a resource and labor intensive process. Therefore, we propose a novel multi-granularity occlusion feature enhancement algorithm for person search, which first performs multi-granularity random occlusion on original scene images to expand the training data, and then generates virtual features with diverse information from the occluded scene images. Finally, the generated virtual features are used to enhance the pedestrian representation in the real features. Furthermore, based on generative adversarial learning, a multi-granularity feature alignment module is designed to align the occluded image features and the original image features, and thereby maintain their semantic consistency. Experiments on CUHK-SYSU and PRW datasets show that the proposed algorithm can significantly improve the search accuracy of person search.

**Keywords**: deep learning; computer vision; person search; object detection; granular computing; data processing; feature extraction; generative adversarial networks; alignment

行人搜索任务指在跨摄像头的场景图像中检测和识别特定行人, 可细分为行人检测和行人重识别 (person re-identification, ReID) 两个子任务。

**收稿日期:** 2024−07−25.　**网络出版日期:** 2024−12−04.
**基金项目:** 国家重点研发计划项目（2022YFB3104700）; 国家自然科学基金项目（62376198, 62163016）.
**通信作者:** 苗夺谦. E-mail: dqmiao@tongji.edu.cn.

©《智能系统学报》编辑部版权所有

行人搜索在安防、计算机视觉等领域 (例如搜索与救援、交通管理、城市安全与监控等) 具有广泛的应用前景, 能够为社会的智能化和安全性提供更多的解决方案。

目前, 行人搜索框架主要分为两阶段框架和端到端框架。两阶段框架依次完成行人检测和 ReID 任务, 该框架需要训练两个独立的模型, 包含两个主干网络, 产生较大的时间和资源消耗。为此, 有学者提出了端到端框架[1], 旨在单个模型中同时处理行人检测和行人重识别两个子任务, 此类框架仅需一个主干网络, 通常基于 Faster R-CNN (faster region-based convolutional neural networks)[2]、FCOS (fully convolutional one-stage object detection)[3]、DETR (detection transformer)[4] 等模型构建, 并且在搜索精度上取得了显著提升。

现有的基于端到端框架的改进方法主要通过优化网络结构, 从有限的标注场景图像中挖掘更多的信息[5-9], 学习身份辨别性更强的行人表征, 从而获得更高的行人搜索精度。在当前阶段, 此类方法的表征提取受到了标注数据规模的限制, 而获取大规模的标注数据是资源、劳动密集型的工作。因此如何高效扩充标注数据集, 进一步增强行人表征的可辨别性成为一个关键挑战。一些研究尝试通过图像增强的方式丰富训练数据[10-12], 用于表征学习的训练, 以提升表征的鲁棒性。然而, 此类方法并未有效解决新训练样本中引入的噪声信息导致行人表征的空间分布发生偏移的问题。另有研究者尝试利用真实的无标注场景图像辅助模型训练[13-15], 增强标注图像中的行人表征, 但由于其与标注图像数据在领域上的差异, 且缺少行人边界框标注信息, 该表征增强方法对搜索精度提升有限。

因此, 为了有效增强行人表征的辨别能力, 本文提出了一种新的多粒度遮挡特征增强 (multi-granularity occlusion feature enhancement, MGOFE) 算法。该算法在扩充标注数据集的同时, 通过多粒度特征对齐模块对齐新增样本特征与原始样本特征, 保持两者语义一致性, 并利用多粒度融合特征增强模块对原始样本中的行人表征进行补充和增强。

首先, 通过多粒度区域遮挡 (multi-granularity region occlusion, MGRO) 操作, 随机遮挡图像的不同区域, 获得多个新的遮挡训练样本。其中, 多粒度随机性包括遮挡框大小的随机、遮挡框数量的随机和遮挡框位置的随机。多粒度遮挡虽然能够有效增加训练图像的多样性, 但会导致图像中不同粒度的细节丢失, 从而引起图像特征的分布偏移。

针对上述分布偏移问题, 本算法进一步设计了一个多粒度特征对齐 (multi-granularity feature alignment, MGFA) 模块。MGFA 通过生成对抗网络对齐多粒度遮挡图像和原始图像之间的语义信息。具体而言, 所有粒度遮挡图像的特征均被送入生成器, 重新生成遮挡部分的虚拟特征, 由于遮挡位置的不同, 生成器能够捕获不同区域之间的语义关系, 提升特征重构能力。判别器被用于区分不同粒度遮挡图像中生成的虚拟特征与真实特征, 并引导生成器生成足够“真实”的虚拟特征。通过对齐虚拟特征和真实特征的语义信息, 本算法不仅扩大了训练数据的规模, 同时避免了新增图像特征与原始图像特征之间的分布偏移, 从而充分发挥标注样本扩增的优势。

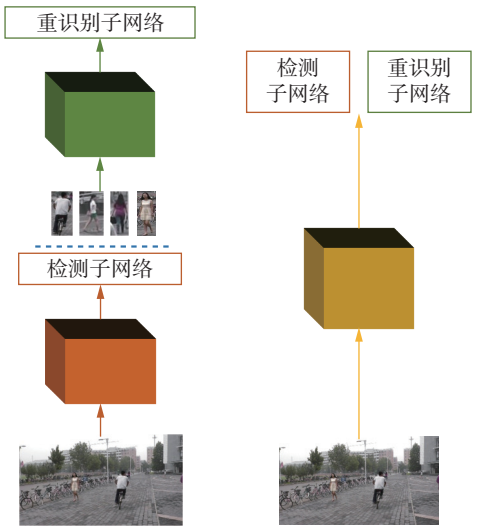
最终, 在特征语义一致的前提下, 通过多粒度融合特征增强 (multi-granularity fusion feature enhancement, MGFFE) 模块, 挖掘新增样本表征中的关键信息, 对重识别行人表征进行增强, 即多粒度虚拟特征中的行人表征进一步用于强化真实特征中的行人表征, 提高其可辨别性。

## 1　相关工作

目前, 行人搜索已经在计算机视觉领域引起了广泛关注。根据子任务的结合方式, 现有行人搜索框架可划分为两阶段框架和端到端框架。如图 1(a) 所示, 两阶段框架是级联行人检测模型[16]和重识别模型。Zheng 等[17] 对各种检测模型和重识别模型组合进行了系统评估, 并为了调整匹配相似度, 提出了一种重加权算法, 以抑制假阳性检测。Chen 等[18] 首次揭示了行人检测与行人重识别之间存在的固有优化冲突, 并提出了 MGTS (mask-guided two-stream) 算法, 通过掩码引导两个并行卷积神经网络, 消除优化冲突问题。Dong 等[19] 提出了 IGPN (instance guided proposal network) 减少区域提议的数量, 从而减轻重识别的负担。

端到端框架联合优化行人检测模型和行人重识别模型, 更简单更高效, 如图 1(b) 所示。Xiao 等[1] 基于 Faster R-CNN 模型设计了首个端到端的行人搜索框架, 并提出了 OIM (online instance matching) 损失函数监督行人表征学习。为了缓解两个子任务的优化目标矛盾性, Chen 等[5] 提出了 NAE (norm-aware embedding) 算法, 通过将行人表征分解为范数和角度, 解耦检测和重识别, 以更好地处理不同子任务的优化目标, 从而提高行

人搜索性能。为了在高质量边界框中提取行人表征，Li 等[20] 提出了 SeqNet (sequential end-to-end network)，将检测和 ReID 作为一个渐进的过程，依次用两个 R-CNN (region-based convolutional neural networks) 头部网络处理。此外，为了更好地利用图像上下文信息，Li 等[20] 还提出了 CBGM (context bipartite graph matching) 后处理方法，将搜索过程建模为一个图匹配问题，以更好地捕获目标行人的身份信息。Jaffe 等[8] 设计了 GFN (gallery filter network) 算法，在检测前根据硬阈值有效剔除不相关场景图像，只对相似度高的图像进行检测和重识别，以减小搜索的图库规模。针对行人边界框内的外观变化和遮挡现象，Zhang 等[21] 提出了 AMPN (attentive multi-granularity perception network)，利用局部区域的区分性特征，提高行人表征的辨别能力。另外，为了解决遮挡和子任务优化冲突的问题，Zhang 等[22] 设计了 ASTD (adaptive shift and task decoupling) 方法，通过尺度感知变换器和任务解耦机制提高行人表征的准确性和鲁棒性。




(a) 两阶段框架　　(b) 端到端框架

**图 1　行人搜索框架对比**

**Fig. 1　Compare of person search frameworks**

除了上述基于 Faster R-CNN 构建的模型，Yan 等[23] 提出了基于全卷积单阶段 (fully-convolutional one-stage, FCOS) 检测器[3] 的 AlignPS (feature-aligned person search network) 模型，该模型自动学习特征点之间的关系，无需预定义的锚框。鉴于 Transformer 在计算机视觉领域中的优异性能，部分研究者将其引入到行人搜索的研究中[24]。Cao 等[25] 设计的 PSTR (person search with transformers) 由用于行人检测的编码器–解码器和用于行人重识别的判别式解码器组成。Yu 等[9] 设计了 COAT (cascade occluded attention transformer) 模型，即级联 R-CNN 样式的变体，该模型利用 Transformer 增强的模型从粗到细地学习姿态、尺寸不变的特征。

上述方法侧重于网络结构的优化，从有限的标注场景图像中充分挖掘行人相关信息，以提高行人表征的辨别力。这类方法取得了良好的效果，然而，目前提升行人搜索性能最有效的方法之一是扩充标注数据集，扩大标注数据的规模，并充分提取行人身份相关信息，以进一步提高行人表征的辨别力。然而大规模标注数据集的获取是成本高昂且耗时的[26]。

因此，部分研究者从图像增强的角度出发，增加训练样本的多样性，让图像数据产生更大信息量，以达到提高模型精度、泛化能力的效果。Zhong 等[10] 用随机像素值遮盖图像中的随机矩形区域，防止模型提取的特征过拟合于特定区域，确保模型关注整幅图像。DeVries 等[11] 使用零值掩码裁剪随机正方形区域。Yun 等[27] 在随机选择的矩形区域填充其他图像的区域像素值。Chen 等[12] 提出的 GridMask 方法对图像进行网格遮盖，优化擦除带来的过度删除问题。然而，上述方法虽然扩充了图像数据，但会引入噪声信息，使得增强图像中提取的行人表征与原始图像中提取的行人表征的空间分布存在差异。

另有研究者采用生成图像的方式丰富图像数据[28]。Wei 等[29] 在尽可能保持前景不变的前提下，通过对图像背景进行转换，生成新的训练样本，但该方法无法进行端到端训练。为此，Zheng 等[30] 设计了图像生成模块和融合特征学习模块联合学习框架，允许模型端到端训练。虽然这些方法生成了新的图像，但生成图像需要耗费额外的资源，且图像级别的生成策略可控性略差，由于场景图包含的信息丰富且范围大，难以保证生成的场景图质量优异。

因而，另有一类方法利用无标注数据辅助模型训练。Li 等[13] 侧重于数据源域与目标域之间的对齐，以及目标域行人检测结果的优化，以提高行人搜索性能。Wang 等[14] 针对标注数据欠缺导致模型泛化性差的问题，使用虚拟数据集丰富数据，并动态训练数据集生成和学习域不变特征。Qi 等[26] 利用非对称领域对抗模块从新增数据中学习迁移信息，增强原始数据集中的特征，解决领域差异问题。Chen 等[15] 通过自监督学习从大量未标注的数据集中学习通用的人类表征，但由于数据量庞大，其训练时间和训练成本较

高。上述方法新增数据集与原始数据集的数据分布差异较大, 模型需要通过大量训练解决其分布偏移问题, 且缺少标注信息, 导致对性能提升有限。

综上所述, 本文在原始数据集场景图的基础上, 利用多粒度思想[31-34]对场景图进行不同粒度的随机区域遮挡, 以扩充标注数据集; 随后, 提出多粒度特征对齐模块将多粒度遮挡图像的虚拟特征与原始图像特征进行对齐, 以保持两者之间的语义关系; 进一步, 构建多粒度融合特征增强模块对生成的虚拟行人表征进行有效融合, 以增强真实特征中的行人表征, 从而提高行人搜索精度。

## 2　多粒度遮挡特征增强算法

### 2.1　问题描述

端到端行人搜索任务的目标是通过一个网络架构检测场景图像中行人的位置, 并识别行人的身份。给定一个原始场景图像$\boldsymbol{i}$, 对图像进行$m$次不同粒度的随机遮挡, 得到$m$张多粒度遮挡图像, 记为$\boldsymbol{M}=\{\boldsymbol{i}_1',\boldsymbol{i}_2',\cdots,\boldsymbol{i}_m'\}$。经过主干网络提取原始场景图和多粒度遮挡场景图的特征, 得到原始场景图的真实特征$\boldsymbol{f}$和多粒度遮挡场景图特征的集合$\boldsymbol{E}=\{\boldsymbol{e}_1',\boldsymbol{e}_2',\cdots,\boldsymbol{e}_m'\}$, 然后, 将集合$\boldsymbol{E}$输入多粒度特征对齐模块的生成器生成虚拟特征图集合$\boldsymbol{F}'=\{\boldsymbol{f}_1',\boldsymbol{f}_2',\cdots,\boldsymbol{f}_m'\}$。在$\boldsymbol{F}'$和$\boldsymbol{f}$的基础上, 通过两个头部网络逐步精确行人的位置并提取行人表征。第 2 个头部网络利用第 1 个头部网络输出的行人边界框, 在特征图上定位行人特征图, 并提取具有辨别力的行人表征$\boldsymbol{P}=\{\boldsymbol{p}_1',\boldsymbol{p}_2',\cdots,\boldsymbol{p}_m',\boldsymbol{p}_t\}$用于重识别。其中, $\boldsymbol{p}_i'$为在第$i$个虚拟特征$\boldsymbol{f}_i'$中提取的虚拟行人表征; $\boldsymbol{p}_t$为在真实特征$\boldsymbol{f}$中提取的真实行人表征。本算法的目标是通过训练数据扩充、特征对齐和特征增强提高行人搜索任务的搜索精度。

### 2.2　架构概述

多粒度遮挡特征增强 (MGOFE) 算法的总体架构如图 2 所示。该算法基于端到端框架, 通过多粒度区域遮挡进行数据扩充, 新增训练样本集合$\boldsymbol{M}$, 并由 ConvNeXt[35]主干网络提取多粒度遮挡图像的特征集合$\boldsymbol{E}$。其次, 本算法采用多粒度特征对齐 (MGFA) 模块保持生成的虚拟特征与真实特征语义信息的一致性。其中, 生成器将多粒度遮挡图像的特征作为输入, 生成虚拟特征图集合$\boldsymbol{F}'$, 判别器评估$\boldsymbol{F}'$与真实特征$\boldsymbol{f}$的差异。进一步, MGFA 利用 JS 散度 (Jensen-Shannon divergence, JSD) 拉近真实特征与各个粒度遮挡图像的虚拟特征之间的空间分布。接下来, 利用区域提议网络 (region proposal network, RPN) 生成初步候选框。随后, 兴趣区域对齐操作 (region of interest align, RoI-Align) 根据候选框的位置信息和场景图像特征图, 聚合候选框中的特征, 并将其尺寸规范为$512\times14\times14$。

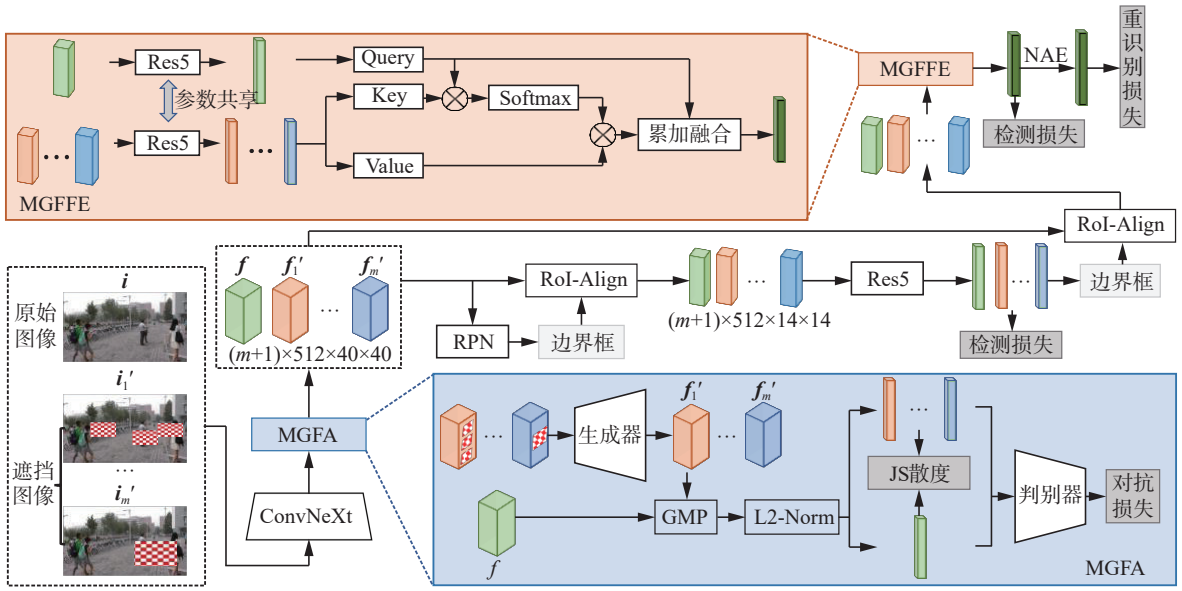



**图 2　MGOFE 的总体架构**

**Fig. 2　Overall architecture of the MGOFE**

为了使提取的行人表征更具代表性, 算法包含两个头部网络, 第 1 个头部网络由 Res5 组成, 对边界框的位置和尺寸进一步微调, 从而引导模型生成更具辨别力的行人表征。其中, Res5 由卷积神经网络和全局最大池化 (global max pooling, GMP) 组成。第 2 个头部网络为多粒度融合特征增强模块 (MGFFE)。该模块利用跨粒度注意力机制, 计算不同粒度的虚拟特征与真实特征中提取

的行人表征之间的多粒度融合权重，以提取多粒度遮挡图像中虚拟行人表征的关键信息，并与真实行人表征进行累加融合，从而增强行人表征的区分能力，提高行人搜索精度。此外，采用规范−感知嵌入 (norm-aware embedding, NAE)[5] 将行人表征分解为范数和角度，分别进行行人检测和行人重识别。

### 2.3　多粒度区域遮挡

多粒度区域遮挡（MGRO）通过随机遮挡原始场景图中不同粒度的区域，得到更多标注图像用于模型训练。该操作的多粒度思想体现在$m$个遮挡图像之间遮挡粒度的差异上。MGRO 根据遮挡框的尺寸动态计算遮挡框的数量，以模拟现实场景中遮挡物尺寸和数量的多样性。图像的遮挡粒度由遮挡框的尺寸和数量共同决定，从而避免遮挡粒度过大导致行人信息完全被遮盖，或遮挡粒度过小以致无法对行人形成有效遮挡的情况。

具体而言，首先，为了保证遮挡框大小的随机性，本文生成$m$个随机数$\rho$，决定$m$个遮挡图像的遮挡框尺寸，$\rho$在均匀分布中采样，记为$\rho \sim U(0,1)$。其次，为了将遮挡图像产生的变化控制在合理的范围内，本文根据遮挡框的大小，确定遮挡框的数量。遮挡框越大，数量越少；遮挡框越小，数量越多。基于此，遮挡框大小和遮挡框数量的计算方式为

$$w_b = \frac{\rho \cdot W}{2}$$

$$h_b = \frac{\rho \cdot H}{2}$$

$$N_b = m - \lfloor \rho \cdot m \rfloor$$

式中：$(w_b, h_b)$为遮挡框大小；$N_b$为遮挡框的数量，范围为$[1, m]$；$W$为场景图的宽；$H$为场景图的高。值得注意的是，遮挡图像内的$N_b$个遮挡框粒度相同，不同遮挡图像间的遮挡框粒度不同，数量不同。

接下来，随机选择遮挡框的位置，即生成$N_b$个随机数$\mu$。本文通过遮挡框的左上角点确定其位置，计算公式为

$$x_{\mathrm{lu}} = \mu \cdot W$$

$$y_{\mathrm{lu}} = \mu \cdot H$$

式中：$(x_{\mathrm{lu}}, y_{\mathrm{lu}})$为左上角点的位置坐标，为了保证遮挡框不超出场景图的边界，$\mu$在均匀分布$U(0, 1-\rho/2)$中采样。

最终，将遮挡框内的像素置为 0，得到$m$幅不同粒度的遮挡图像。通过引入多粒度随机区域遮挡，可以使模型在后续生成特征时，关注图像的不同区域，从而提高生成特征的多样性和鲁棒性。

### 2.4　多粒度特征对齐模块

多粒度特征对齐模块 (MGFA) 作用于所有粒度遮挡图像的特征和原始图像的特征之间，通过生成对抗网络对齐每个粒度的遮挡图像与原始图像的语义信息，减少遮挡噪声带来的特征分布偏移。如图 2 所示，MGFA 由生成对抗网络和 JS 散度组成，其中，生成器为多个卷积层的组合，用于生成多粒度遮挡区域的虚拟特征，判别器为全连接层和激活函数的组合，用于衡量生成的多粒度虚拟特征与真实特征之间的差距。

首先，不同粒度的遮挡场景图特征集合$\boldsymbol{E}$被送入生成器中，以生成虚拟特征，得到不同粒度的虚拟特征集合$\boldsymbol{F}'$。生成器的损失函数为

$$L_{\mathrm{G}} = \frac{1}{mB}\sum_{i=1}^{B}\sum_{k=1}^{m}[\log(1 - D(G(\boldsymbol{e}'_{ik})))]$$

式中：$m$代表遮挡粒度数量；$B$代表训练批量的大小；$\boldsymbol{e}'_{ik}$为第$i$个原始场景图对应的第$k$个粒度的遮挡图像经过主干网络提取的特征；$G(\boldsymbol{e}'_{ik})$为生成器函数，利用遮挡场景图特征$\boldsymbol{e}'_{ik}$生成虚拟特征；判别器函数$D(G(\boldsymbol{e}'_{ik}))$输出的值表示虚拟特征为真实特征的概率。通过最小化生成器损失$L_{\mathrm{G}}$，生成器的目标是生成与真实特征在语义层面相似的虚拟特征。

然后，通过全局最大池化 (GMP) 和 L2 正则化，三维虚拟特征集合$\boldsymbol{F}'$和真实特征$\boldsymbol{f}$被转为一维张量并送入到判别器网络。判别器的损失函数为

$$L_{\mathrm{D}} = \frac{1}{B}\sum_{i=1}^{B}[\log(1 - D(\boldsymbol{f}_i))] + \frac{1}{mB}\sum_{i=1}^{B}\sum_{k=1}^{m}[\log D(G(\boldsymbol{e}'_{ik}))]$$

式中：$m$代表遮挡粒度数量；$B$代表训练批量的大小；$\boldsymbol{f}_i$表示第$i$个真实特征；$\boldsymbol{e}'_{ik}$为第$i$个场景图对应的第$k$个粒度的遮挡图像经过主干网络提取的特征。该损失函数包含两项，第 1 项表示判别器对真实特征的识别能力，用于提高其对真实特征的敏感度；第 2 项表示判别器对生成器生成的虚拟特征的判别能力。通过最小化判别器损失，判别器被优化以尽可能区分真实特征和虚拟特征，从而有效捕获两类特征之间的差异。MGFA 利用生成器和判别器之间的对抗训练，引导生成器在多粒度遮挡图像特征中生成与真实特征空间分布相似的虚拟特征。

为了进一步缩小虚拟特征和真实特征分布之间的差距，MGFA 使用 JS 散度拉近真实特征与各粒度虚拟特征之间的相似度分布。本文将每个粒

度的虚拟特征均与真实特征计算 JS 散度, 并利用所有粒度 JS 散度的均值优化 MGFA 模块训练, 计算公式为

$$L_{\text{dist}}=\frac{1}{m}\sum_{i=1}^{m}D_{\text{JS}}(P(\boldsymbol{f})\|P(\boldsymbol{f}_i'))$$

式中: $P(\boldsymbol{f})$代表真实特征$\boldsymbol{f}$与其他真实特征之间的相似度概率分布, 即将真实特征之间的相似度转换为概率分布; $P(\boldsymbol{f}_i')$表示第$i$个粒度虚拟特征$\boldsymbol{f}_i'$与其他同粒度虚拟特征之间的相似度概率分布; $D_{\text{JS}}$代表 JS 散度, 计算公式为

$$D_{\text{JS}}(P(\boldsymbol{f})\|P(\boldsymbol{f}_i'))=\frac{1}{2}\sum_{f}P(\boldsymbol{f})\log\frac{P(\boldsymbol{f})}{M}+\frac{1}{2}\sum_{f_i'}P(\boldsymbol{f}_i')\log\frac{P(\boldsymbol{f}_i')}{M}$$

式中$M=(P(\boldsymbol{f})+P(\boldsymbol{f}_i'))/2$为混合概率分布。

该模块的损失函数计算公式为

$$L_{\text{mgfa}}=L_{\text{G}}+L_{\text{D}}+L_{\text{dist}}$$

这种联合损失函数有助于对齐生成的虚拟特征与真实特征, 以保持两者的语义一致性。

### 2.5　多粒度融合特征增强模块

为了进一步利用 MGFA 模块生成的遮挡图像的虚拟特征信息, 本文基于注意力机制[36], 设计了多粒度融合特征增强 (MGFFE) 模块。给定真实特征中提取的行人表征和不同粒度遮挡图像虚拟特征中提取的行人表征, MGFFE 通过计算所有粒度虚拟行人表征与真实行人表征之间的融合权重, 动态融合虚拟行人表征中的关键信息, 使真实行人表征可以接收不同粒度虚拟行人表征中的身份信息进行特征增强。

如图 2 所示, 具体来说, MGFFE 通过参数共享的 Res5 网络进一步提取真实特征和不同粒度虚拟特征中的行人表征, 并通过全局最大池化转为一维张量。接下来, 利用跨粒度注意力机制, 实现多粒度特征融合, 以增强真实行人表征。其中, 真实行人表征经过全连接层$\boldsymbol{W}_{\text{Q}}$处理得到 Query, 记为$\boldsymbol{Q}$, 虚拟行人表征分别经过两个不同的全连接层$\boldsymbol{W}_{\text{K}}$和$\boldsymbol{W}_{\text{V}}$处理得到 Key 和 Value, 分别记为 $\boldsymbol{K}$ 和 $\boldsymbol{V}$, 公式化表示为

$$\begin{aligned}\boldsymbol{Q}&=\boldsymbol{p}\boldsymbol{W}_{\text{Q}}\\ \boldsymbol{K}&=\boldsymbol{p}'\boldsymbol{W}_{\text{K}}\\ \boldsymbol{V}&=\boldsymbol{p}'\boldsymbol{W}_{\text{V}}\end{aligned}$$

式中: $\boldsymbol{p}$代表真实行人表征, $\boldsymbol{p}'$代表虚拟行人表征, $\boldsymbol{W}_{\text{Q}}$、$\boldsymbol{W}_{\text{K}}$、$\boldsymbol{W}_{\text{V}}$代表不同全连接层的权重。

$\boldsymbol{Q}$和$\boldsymbol{K}$用于计算真实表征和所有粒度虚拟表征之间的注意力权重。具体而言, $\boldsymbol{Q}$和$\boldsymbol{K}$通过点积计算得到注意力分数, 为了避免点积值随维度$d_k$的增大而逐渐变大, 采用缩放因子$1/\sqrt{d_k}$对$\boldsymbol{Q}\boldsymbol{K}^{\text{T}}$进行缩放。随后, 利用 softmax 函数将缩放后的值归一化为概率分布, 得到多粒度融合的注意力权重$\boldsymbol{W}$。$\boldsymbol{W}$计算公式为

$$\boldsymbol{W}=\text{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\text{T}}}{\sqrt{d_k}}\right)$$

式中$d_k$代表$\boldsymbol{K}$的维度。将此权重与$\boldsymbol{V}$相乘, 得到多粒度的增强信息$\boldsymbol{g}$, 其计算公式为

$$\boldsymbol{g}=\boldsymbol{W}\boldsymbol{V}=[g_1\ g_2\ \cdots\ g_m]$$

式中$g_i$代表不同粒度的增强信息。

将$\boldsymbol{g}$与真实行人表征$\boldsymbol{p}$累加融合, 得到增强后的行人表征$\bar{\boldsymbol{p}}$, 从而$\boldsymbol{p}$可以接收不同粒度虚拟行人表征中的身份相关信息, 实现多粒度融合特征增强, 增强行人表征的计算公式为

$$\bar{\boldsymbol{p}}=\boldsymbol{p}+\sum_{j=1}^{m}(\boldsymbol{g}_j)$$

式中: $j$代表不同的遮挡粒度, $\bar{\boldsymbol{p}}$代表原始场景图的增强行人表征, 其会接收不同粒度虚拟行人表征过滤的信息。

### 2.6　行人搜索损失函数

MGOFE 的检测损失是由 RPN 以及两个头部网络的边界框分类损失和回归损失总和构成[20], 其表示形式为

$$L_{\text{det}}=\sum_{m\in M}(\lambda_m L_{\text{cls}}^m+\gamma_m L_{\text{reg}}^m)$$

式中: $M=\{\text{RPN},\text{RCNN1},\text{RCNN2}\}$; $\lambda_m$和$\gamma_m$为损失的权重。

回归损失计算公式为

$$L_{\text{reg}}=\frac{1}{N_{\text{p}}}\sum_{i=1}^{N_{\text{p}}}L_{\text{loc}}(r_i,\Delta_i)$$

式中: $N_{\text{p}}$为正样本的数量, $r_i$为第 $i$ 个正样本计算的回归值, $\Delta_i$为对应的真实回归值, $L_{\text{loc}}$为Smooth-$L_1$-Loss。

边界框的分类损失计算公式为

$$L_{\text{cls}}=-\frac{1}{N}\sum_{i=1}^{N}c_i\log(p_i)$$

式中: $N$ 为样本的数量, $p_i$为第 $i$ 个样本的预测分类概率, $c_i$为真实标记。

此外, 本文采用 OIM 损失[1] 监督行人重识别, 重识别损失计算公式为

$$L_{\text{reid}}=\log p_t$$

式中$p_t$是行人表征属于真实身份 $t$ 的概率, 即表征相似度。该损失函数能够推动行人表征向量与目标表征向量相似, 同时拉远与其他身份表征向量的距离。

最终, 完整的行人搜索损失函数表示为

$$L = L_{\text{det}} + L_{\text{reid}}$$

即检测损失和重识别损失的总和。

## 3　实验结果与分析

### 3.1　数据集

CUHK-SYSU[1] 是一个大规模行人搜索数据集，包含手持摄像机街拍和电影中捕获的 18 184 张图像，共 96 143 个行人边界框，具有 8 432 个行人 ID。该数据集分为训练集和测试集，其中训练集包含 11 206 张图像和 5 532 个不同的行人 ID，测试集包括 6 978 张图像和 2 900 个查询人员。值得注意的是，训练集和测试集在图像和行人 ID 上没有重叠。为了评估搜索性能，数据集为每个查询预先定义了不同规模的图库，范围为 [50, 4 000]。如果未指定，图库默认大小为 100。

PRW[17] 数据集由清华大学中 6 台同步监控摄像机拍摄的视频帧组成。该数据集包含 11 816 帧图像，共 43 110 个行人边界框。训练集包含 5 704 帧图像，涵盖 482 个行人 ID，测试集包含 6 112 帧图像和 2 057 个查询人员。查询图库是整个测试集，即图库规模为 6 112。

### 3.2　实验设置与评估指标

实验在 NVIDIA Tesla V00 GPU 上进行，使用 ConvNeXt[35] 作为主干网络。在主干网络提取的特征图上，本文遵循 NPSM[37] 中的锚点设置构建区域提议网络 (RPN)。在 RPN 中，本文将与真实边界框的 IoU≥0.5 的候选框采样为正样本，将 IoU 值在 [0.1, 0.5) 的候选框设置为负样本。

在训练过程中，批量大小设置为 6，采用自适应矩估计 (adaptive moment estimation, Adam) 对模型参数进行优化。在两个数据集上，模型均训练 30 个轮次，同时对梯度进行了修剪，将其范数限制为 10。初始学习率设置为 0.000 1，在第 15 和第 25 个轮次时减小学习率，缩小因子为 10。需要注意的是，本文首先训练多粒度特征对齐模块，待其收敛后，再利用生成的高质量虚拟特征与真实特征共同优化行人搜索任务。

参考先前研究的设置[1,38-39]，本文使用召回率 (recall) 和 0.5IoU 下的平均精度 (average precision, AP) 评估行人检测的性能，使用均值平均精度 (mean average precision, mAP) 和 top-1 评估行人重识别的性能。

### 3.3　定量分析

#### 3.3.1　性能对比

为了验证 MGOFE 算法的有效性，本节将 MGOFE 和现有最先进的行人搜索算法进行比较，结果如表 1 所示。其中，两阶段的对比算法为 MGTS[18]、CLSA (cross-level semantic alignment)[40]、IGPN[19]、RDLR (Re-ID driven localization refinement)[41] 和 TCTS (task-consistent two-stage framework)[42]。单阶段的对比算法包括基于 Faster R-CNN 的 OIM[1]、IAN (individual aggregation network)[43]、NAE+[5]、AGWF (adaptive gradient weighting function)[44]、SeqNet[20]、MHGAM (multi-head global attention module)[45]、SeqNeXt[8] (enhanced SeqNet with a ConvNeXt base)；基于 FCOS 的 AlignPS[23]；基于 Transformer 的 COAT[9]、PSTR[25]、SOLIDER (semantic controllable self-supervised learning)[15]。

**表 1　CUHK-SYSU 和 PRW 数据集上的实验结果对比**
**Table 1　Experimental results comparison on CUHK-SYSU and PRW datasets**　%

| 框架类别 | 对比算法 | CUHK-SYSU | | PRW | |
|---|---|---|---|---|---|
| | | mAP | top-1 | mAP | top-1 |
| 两阶段框架 | MGTS[18] (ECCV2018) | 83.0 | 83.7 | 32.6 | 72.1 |
| | CLSA[40] (ECCV2020) | 87.2 | 88.5 | 38.7 | 65.0 |
| | IGPN[19] (CVPR2020) | 90.3 | 91.4 | 47.2 | 87.0 |
| | RDLR[41] (ICCV2019) | 93.0 | 94.2 | 42.9 | 70.2 |
| | TCTS[42] (CVPR2020) | 93.9 | 95.1 | 46.8 | 87.5 |
| 端到端框架 | OIM[1] (CVPR2017) | 75.5 | 78.7 | 21.3 | 49.4 |
| | IAN[43] (PR2019) | 76.3 | 80.1 | 23.0 | 61.9 |
| | NAE+[5] (CVPR2020) | 92.1 | 92.9 | 44.0 | 81.1 |
| | AGWF[44] (ICCV2021) | 93.3 | 94.2 | 53.3 | 87.7 |
| | AlignPS[23] (CVPR2021) | 94.0 | 94.5 | 46.1 | 82.1 |
| | SeqNet+CBGM[20] (AAAI2021) | 94.8 | 95.7 | 47.6 | 87.6 |
| | COAT[9] (CVPR2022) | 94.2 | 94.7 | 53.3 | 87.4 |
| | MHGAM[45] (IMAVIS2021) | 94.9 | 95.9 | 47.9 | 88.0 |
| | PSTR[25] (CVPR2022) | 95.2 | 96.2 | 56.5 | 89.7 |
| | SeqNeXt[8] (WACV2023) | 96.1 | 96.5 | 57.6 | 89.5 |
| | SOLIDER[15] (CVPR2023) | 95.5 | 96.1 | **59.8** | 86.7 |
| | MGOFE (本文算法) | **96.5** | **96.9** | 59.0 | **90.0** |

注：加粗数字表示最优结果。

在 CUHK-SYSU 数据集上，MGOFE 表现出优异的性能，在两个常用指标 mAP 和 top-1 上，搜索精度优于所有对比算法。具体而言，相比于 SeqNeXt 算法，MGOFE 在 mAP 和 top-1 指标上均提升了 0.4 百分点；相比于 SOLIDER 算法，mAP 和 top-1 指标分别提升 1.0 和 0.8 百分点。性能上的显著提升充分验证了 MGOFE 通过多粒度随机遮挡扩充数据集，进一步利用对齐的遮挡图像特征增强原始图像特征对于搜索精度提升的有效性。

在 PRW 数据集上，虽然 MGOFE 算法在 mAP

指标上略逊于 SOLIDER, 但是 MGOFE 算法在 top-1 指标上超越了 SOLIDER 算法 3.3 百分点, 这是由于 SOLIDER 使用的模型专注于提取人类通用表征, 因此在行人搜索领域中, 行人表征缺乏针对性, top-1 精度会降低。需要注意的是, SOLIDER 算法需要在大规模无标注行人数据集上进行额外的预训练, 增加了模型对于外部数据和计算资源的依赖。

### 3.3.2　消融实验

为了验证多粒度特征对齐模块 (MGFA) 和多粒度融合特征增强模块 (MGFFE) 的有效性, 本文在 CUHK-SYSU 和 PRW 数据集上进行了详细的消融实验。具体而言, 本文将 MGOFE 与以下变体进行比较: 1) 基准算法为以 ConvNeXt 为主干网络的 SeqNet 模型[20], 2) w MGRO 代表在基准算法的基础上, 只增加数据扩充, 3) w/o MGFFE 代表在 MGOFE 中剔除了 MGFFE, 4) w/o MGFA 代表在 MGOFE 中只去除 MGFA。表 2 给出了消融实验对比结果, 其中计算量用浮点运算次数 (floating-point operations, FLOPs) 衡量。

**表 2　消融实验对比结果**

**Table 2　Ablation experiment comparison results**　　%

| 算法 | GPU | 参数量/$10^6$ | 计算量/$10^9$ | CUHK-SYSU | | PRW | |
|---|---|---|---|---|---|---|---|
| | | | | mAP | top-1 | mAP | top-1 |
| 基准算法 | V100(14.1) | 119.0 | 522.4 | 95.8 | 96.2 | 57.9 | 88.5 |
| w MGRO | V100(14.1) | 119.0 | 522.4 | 96.0 | 96.2 | 58.2 | 88.8 |
| w/o MGFFE | V100(14.1) | 120.1 | 527.5 | 96.2 | 96.4 | 58.5 | 89.3 |
| w/o MGFA | V100(14.1) | 119.2 | 522.6 | 96.3 | 96.6 | 58.6 | 89.5 |
| MGOFE | V100(14.1) | 120.3 | 527.7 | **96.5** | **96.9** | **59.0** | **90.0** |

注: 加粗数字表示最优结果。

由表 2 可知, 总体上, 本文提出的 MGOFE 与所有变体相比, 实现了最优性能, 验证了两个子模块的有效性和互补性。具体而言, 仅采用 MGRO 扩充数据集, 带来的性能提升有限, 这是由于 MGRO 未能充分挖掘图像数据中的信息。与 MGOFE 相比, 缺失 MGFFE 模块会导致性能下降明显, 在 PRW 数据集上 mAP 和 top-1 指标分别下降 0.5 和 0.7 百分点, 在 CUHK-SYSU 数据集上 mAP 和 top-1 指标分别下降 0.3 和 0.5 百分点。这表明利用生成的虚拟行人表征对真实行人表征进行增强是非常重要的, 模型能从更多样化的数据中学习身份信息更丰富的行人表征。

此外, MGOFE 的性能优于 w/o MGFA, MGFA 模块对 PRW 数据集上的 mAP 和 top-1 指标分别带来了 0.4 和 0.5 百分点的提升, 对 CUHK-SYSU 数据集上的 mAP 和 top-1 指标分别带来了 0.2 和 0.3 百分点的提升。这表明遮挡图像特征和原始图像特征的分布对齐后, 真实行人表征能在虚拟行人表征中得到更加丰富的信息。

表 2 的第 3、4 列分别表示算法的参数量和计算量。从表中分析可得, 相较于基准算法, MGOFE 算法增加了约 $1.3\times10^6$ 的参数量 (增幅约为 1.1%) 和 $5.3\times10^9$ 的 FLOPs (增幅约为 1.0%)。然而, 在有限的参数量和计算量增加的前提下, MGOFE 算法显著提升了行人搜索任务的精度。与基准算法相比, 在 PRW 数据集上, mAP 和 top-1 分别提升了 1.1 和 1.5 百分点; 在 CUHK-SYSU 数据集上, mAP 和 top-1 均提升了 0.7 百分点。同时, 后续实验进一步验证了 MGOFE 算法在鲁棒性和行人表征提取方面的优势。

### 3.3.3　参数实验

为了确定 MGRO 中数据扩充的最佳粒度数量, 本文分别在粒度数为 1、2、3 的情况下进行对比实验, 即 $m=1$, 2, 3。从表 3 中可以看出, $m$ 为 2 时, MGOFE 的精度最高, 这表明粒度数为 2 时, 模型在数据多样性和性能之间取得平衡。更多的粒度设置虽然能够进一步增加数据多样性, 但可能导致训练样本有较多的信息冗余, 不利于模型提取更具辨别力的行人表征。

**表 3　参数 $m$ 定量实验**

**Table 3　Quantitative experiment with the parameter $m$**　　%

| $m$ | CUHK-SYSU | | PRW | |
|---|---|---|---|---|
| | mAP | top-1 | mAP | top-1 |
| 1 | 96.2 | 96.5 | 58.6 | 89.7 |
| 2 | **96.5** | **96.9** | **59.0** | **90.0** |
| 3 | 96.4 | 96.7 | 58.7 | 89.8 |

注: 加粗数字表示最优结果。

针对模型训练的批量大小, 本文设计了参数实验, 如表 4 所示。实验结果表明, 随着批量大小的增加, 模型的训练精度整体呈上升趋势。然而, 当批量大小超过 6 时, 这一趋势较为平缓。此外, 由于批量大小越大显存占用越多, 综合考虑算法性能和显存消耗, 本文选择批量大小为 6。

表 4　批量大小参数实验
Table 4　Batch size parameter experiment　%

| 批量大小 | CUHK-SYSU | | PRW | |
|---|---|---|---|---|
| | mAP | top-1 | mAP | top-1 |
| 3 | 96.3 | 96.5 | 58.8 | 89.7 |
| 4 | 96.4 | 96.7 | 58.7 | 89.9 |
| 5 | 96.5 | 96.8 | 58.9 | 89.8 |
| 6 | 96.5 | **96.9** | **59.0** | 90.0 |
| 7 | 96.5 | 96.8 | 58.9 | **90.1** |
| 8 | **96.6** | 96.8 | **59.0** | 90.0 |

注：加粗数字表示最优结果。

针对 RoI-Align 的池化尺寸，本文设计了相关参数实验。从表 5 中可以看出，随着池化尺寸的增加，搜索精度的增长趋势逐渐变缓。值得注意的是，较大的池化尺寸会显著增加显存消耗，因此本文默认使用14×14的池化尺寸，以在搜索精度与内存消耗之间取得平衡。

表 5　RoI-Align 池化尺寸参数实验
Table 5　RoI-Align pooling size parameter experiment　%

| RoI-Align 池化尺寸 | CUHK-SYSU | | PRW | |
|---|---|---|---|---|
| | mAP | top-1 | mAP | top-1 |
| 7×7 | 96.3 | 96.8 | 58.8 | 89.8 |
| 14×14 | 96.5 | **96.9** | 59.0 | **90.0** |
| 18×18 | **96.6** | **96.9** | **59.1** | **90.0** |

注：加粗数字表示最优结果。

此外，为了确定本文的训练轮次，表 6 给出了模型在训练 24、26、28、30 和 32 个轮次后的精度。结果显示，当模型训练 30 个轮次时已经收敛，再进行训练并未带来精度的提升。

表 6　训练轮次参数实验
Table 6　Training epochs parameter experiment　%

| 训练轮次 | CUHK-SYSU | | PRW | |
|---|---|---|---|---|
| | mAP | top-1 | mAP | top-1 |
| 24 | 96.2 | 96.5 | 58.6 | 89.7 |
| 26 | 96.4 | 96.7 | 58.8 | 89.8 |
| 28 | **96.5** | 96.8 | 58.8 | **90.0** |
| 30 | **96.5** | **96.9** | **59.0** | **90.0** |
| 32 | 96.4 | 96.7 | **59.0** | 89.9 |

注：加粗数字表示最优结果。

### 3.3.4　复杂场景实验

为了验证 MGOFE 提取的行人表征在遮挡和低分辨率等复杂情况下的有效性，本实验选择被遮挡的查询行人和低分辨率查询行人进行测试。该样本源于 CUHK-SYSU 数据集的测试集，Occlusion 子集包含 187 个有遮挡的查询行人，Resolution 子集包含 290 个低分辨率查询行人。结果如表 7 所示，在遮挡的情形下，MGOFE 的 mAP 指标较基准算法提升了 0.9 百分点，在低分辨率的情形下，MGOFE 相较于基准算法性能提升 0.6 百分点。这表明，即使在复杂场景下，MGOFE 仍能提取更具有身份辨别力的行人表征。

表 7　特殊情况下两个算法 mAP 指标对比
Table 7　Comparison of mAP metrics between two algorithms in special scenarios　%

| 算法 | Occlusion | Resolution |
|---|---|---|
| 基准算法 | 91.3 | 91.5 |
| MGOFE | 92.2 | 92.1 |

由于 PRW 数据集来源于监控摄像头拍摄的图像，因此本文设计实验，在跨摄像头场景下构建图库评估算法的性能。跨摄像头场景是指搜索图像的摄像头 ID 与查询行人的摄像机 ID 不一致。跨摄像头的行人图像通常伴随显著的姿态和视角变化，正确匹配需依赖具有较强辨别力的行人表征。从表 8 可以看出，MGOFE 算法在此场景中展现出更优的适用性，能够更全面、更准确地提取行人表征，使得行人表征具有强大的区分能力。

表 8　PRW 数据集跨摄像头场景测试
Table 8　Cross-camera test on the PRW dataset　%

| 算法 | mAP | top-1 |
|---|---|---|
| 基准算法 | 54.8 | 76.9 |
| MGOFE | 55.3 | 77.5 |

为了进一步验证 MGOFE 算法的鲁棒性，本文进行了跨数据集的性能测试。具体而言，本实验采用在 PRW 数据集中训练的模型，测试其在 CUHK-SYSU 数据集上的精度；使用在 CUHK-SYSU 数据集中训练的模型，评估其在 PRW 数据集上的表现。本实验利用 MGOFE 算法、基准算法以及 OIM 算法进行比较，结果如表 9 所示。从表 9 可以看出，尽管在跨数据集的情形下所有算法的精度均有所下降，但本算法仍保持较高的性能，展现出良好的鲁棒性。

表 9　跨数据集性能测试
Table 9　Performance evaluation in a cross-dataset scenario　%

| 算法 | CUHK-SYSU→PRW | | PRW→CUHK-SYSU | |
|---|---|---|---|---|
| | mAP | top-1 | mAP | top-1 |
| OIM | 20.4 | 42.2 | 49.2 | 54.8 |
| 基准算法 | 27.5 | 76.6 | 52.4 | 57.4 |
| MGOFE | **27.7** | **76.8** | **52.6** | **57.7** |

注：加粗数字表示最优结果。

### 3.3.5　大规模图库搜索

本文在不同图库规模下对基准算法和 MGOFE

的性能进行了比较。如图 3 所示, 随着图库规模的增加, 这两种算法的性能均呈单调下降趋势, 这表明在庞大的搜索范围内寻找目标行人的难度增加。然而, 值得注意的是, MGOFE 算法在不同图库规模下, 其 mAP 和 top-1 指标的性能均优于基准算法。这表明 MGOFE 能够提取身份信息更丰富的行人表征, 从而在面对不同规模的图库时保持相对稳定的搜索性能。

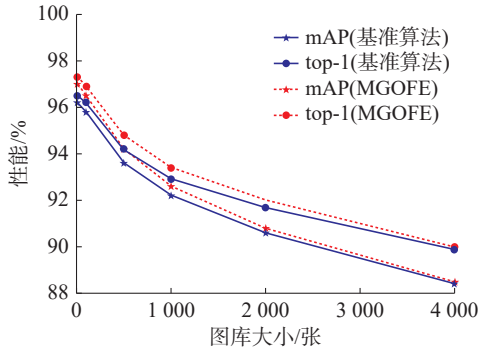



**图 3　基准算法和 MGOFE 在不同图库规模下的性能对比**

**Fig. 3　Performance comparison of baseline algorithm and MGOFE under different gallery sizes**

### 3.4　定性分析

为了直观地展示多粒度特征对齐模块 (MGFA) 的有效性, 本实验使用 t-SNE 可视化原始场景图特征和遮挡场景图特征在对齐前后的分布差异, 结果如图 4 所示。其中, 蓝色点表示原始图像特征, 橙色点表示遮挡图像特征。图 4(a) 表示未对齐的原始图像特征和遮挡图像特征的分布情况, 即 $\boldsymbol{f}$ 和 $\boldsymbol{f}_j'$, 能够看到两种分布存在显著差异。图 4(b) 表示对齐后原始图像特征和遮挡图像特征分布情况, 能够看到两种特征在分布上已不存在偏差。t-SNE 可视化对比直观地展示了 MGFA 能够有效对齐遮挡图像特征和原始图像特征, 有助于多粒度融合特征增强模块通过生成的遮挡图像虚拟特征增强原始图像的行人表征。

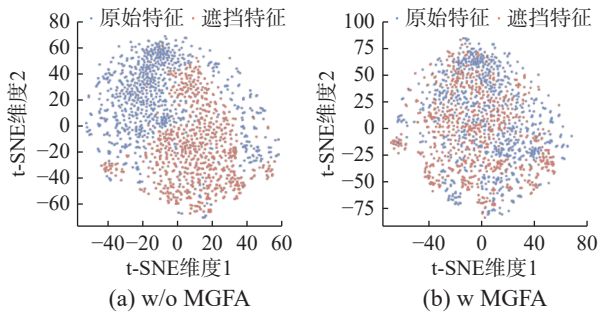



**图 4　在 PRW 数据集上原始特征和遮挡特征的 t-SNE 可视化**

**Fig. 4　t-SNE visualization of original and occluded features on the PRW dataset**

为了更直观地分析 MGOFE 对行人的关注情况, 本实验将基准算法和 MGOFE 算法中行人的特征图进行了可视化。如图 5 所示, MGOFE 算法能更准确地关注到行人区域, 且包含的行人身份信息更丰富, 尤其是当边界框中有干扰的情况下, 效果更为明显。这表明 MGOFE 通过扩充数据集、对齐多粒度图像特征分布和多粒度融合特征增强, 能够提取身份信息更加丰富、更具有辨别力的行人表征。

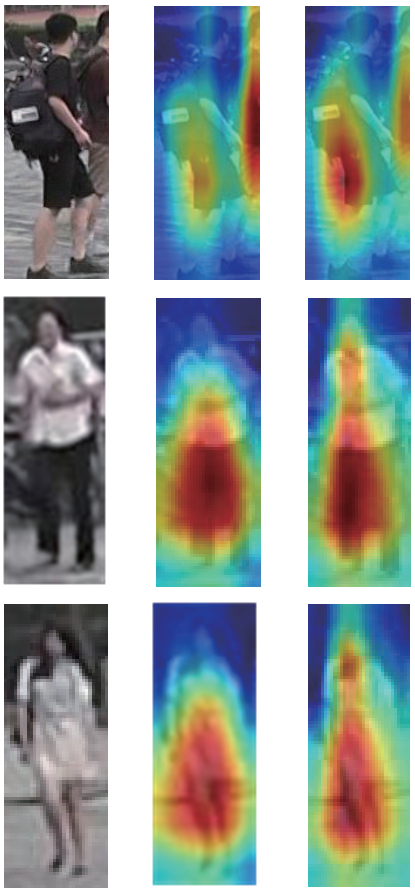



**图 5　特征图可视化**

**Fig. 5　Visualization of feature maps**

## 4　结束语

本文提出了一种新的场景图多粒度遮挡特征增强算法, 它由多粒度区域遮挡、多粒度特征对齐模块和多粒度融合特征增强模块组成。这些部分协同扩充标注数据集, 并对齐多粒度遮挡图像的虚拟特征与真实特征的语义关系, 最后充分利用遮挡样本的行人表征进一步增强真实样本的行人表征。在 CUHK-SYSU 和 PRW 数据集上的实验结果验证了本文算法优越的搜索性能, 可视化实验直观地展示了算法的有效性。本文从数据扩充、特征对齐和特征增强的角度提升搜索精度, 未来的研究将着重通过多视角辅助信息, 进一步提升行人搜索的性能。

## 参考文献:


[1] XIAO Tong, LI Shuang, WANG Bochao. Joint detection and identification feature learning for person search[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2017: 3415−3424.

[2] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(6): 1137−1149.

[3] TIAN Zhi, SHEN Chunhua, CHEN Hao, et al. FCOS: Fully convolutional one-stage object detection[C]//2019 IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society, 2019: 9626−9635.

[4] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]// Proceedings of the 16th European Conference on Computer Vision. Cham: Springer, 2020: 213−229.

[5] CHEN Di, ZHANG Shanshan, YANG Jian, et al. Norm-aware embedding for efficient person search[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2020: 12612−12621.

[6] 张帅宇, 彭力, 戴菲菲. 基于渐进式注意力金字塔的行人重识别方法[J]. 计算机科学, 2023, 50(S1): 452−459.
ZHANG Shuaiyu, PENG Li, DAI Feifei. Person re-identification method based on progressive attention pyramid [J]. Computer science, 2023, 50(S1): 452−459.

[7] 杨静, 张灿龙, 李志欣, 等. 集成空间注意力和姿态估计的遮挡行人再辨识[J]. 计算机研究与发展, 2022, 59(7): 1522−1532.
YANG Jing, ZHANG Canlong, LI Zhixin, et al. Integrated spatial attention and pose estimation for occluded person re-identification[J]. Journal of computer research and development, 2022, 59(7): 1522−1532.

[8] JAFFE L, ZAKHOR A. Gallery filter network for person search[C]//2023 IEEE/CVF Winter Conference on Applications of Computer Vision. Los Alamitos: IEEE Computer Society, 2023: 1684−1693.

[9] YU Rui, DU Dawei, LALONDE R, et al. Cascade transformers for end-to-end person search[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2022: 7257−7266.

[10] ZHONG Zhun, ZHENG Liang, KANG Guoliang, et al. Random erasing data augmentation[C]//Proceedings of the 34th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI: 2020, 13001−13008.

[11] DEVRIES T, TAYLOR G W. Improved regularization of convolutional neural networks with cutout[EB/OL]. (2017−08−15)[2024−07−10]. https://doi.org/10.48550/arXiv.1708.04552.

[12] CHEN Pengguang, LIU Shu, ZHAO Henghuang, et al. GridMask data augmentation[EB/OL]. (2017−01−17)[2024−07−10]. https://doi.org/10.48550/arXiv.2001.04086.

[13] LI Junjie, YAN Yichao, WANG Guanshuo, et al. Domain adaptive person search[C]//Proceedings of the 17th European Conference on Computer Vision. Cham: Springer, 2022: 302−318.

[14] WANG Wenhao, LIAO Shengcai, ZHAO Fang, et al. Domainmix: learning generalizable person re-identifica- tion without human annotations[EB/OL]. (2020−11−24)[2024−07−10]. https://doi.org/10.48550/arXiv.2011.11953.

[15] CHEN Weihua, XU Xianzhe, JIA Jian, et al. Beyond appearance a semantic controllable self-supervised learning framework for human-centric visual tasks[C]// 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2023: 15050−15061.

[16] 程德强, 马尚, 寇旗旗, 等. 基于 YOLOv4 改进特征融合及全局感知的目标检测算法[J]. 智能系统学报, 2024, 19(2): 325−334.
CHENG Deqiang, MA Shang, KOU Qiqi, et al. Target detection algorithm for improving feature fusion and global perception based on YOLOv4[J]. CAAI transactions on intelligent systems, 2024, 19(2): 325−334.

[17] ZHENG Liang, ZHANG Hengheng, SUN Shaoyan, et al. Person re-identification in the wild[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2017: 3346−3355.

[18] CHEN Di, ZHANG Shanshan, OUYANG Wanli, et al. Person search via a mask-guided two-stream CNN model[C]// Proceedings of the 15th European Conference on Computer Vision. Cham: Springer, 2018: 764−781.

[19] DONG Wenkai, ZHANG Zhaoxiang, SONG Chunfeng, et al. Instance guided proposal network for person search [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2020: 2585−2594.

[20] LI Zhengjia, MIAO Duoqian. Sequential end-to-end network for efficient person search[C]// Proceedings of the 35th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2021: 2011−2019.

[21] ZHANG Qixian, WU Jun, MIAO Duoqian, et al. Attentive multi-granularity perception network for person search [J]. Information sciences, 2024, 681: 121191.

[22] ZHANG Qixian, MIAO Duoqian, ZHANG Qi, et al. Learning adaptive shift and task decoupling for discriminative one-step person search[J]. Knowledge-based systems, 2024, 304: 112483.

[23] YAN Yichao, LI Jinpeng, QIN Jie, et al. Anchor-free person search[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2021: 7686−7695.

[24] 邓家欣. 基于深度神经网络的行人搜索方法研究[D]. 成都: 电子科技大学, 2024.
DENG Jiaxin. Research on person search method based on deep neural network[D]. Chengdu: University of Electronic Science and Technology of China, 2024.

[25] CAO Jiale, PANG Yanwei, ANWER R M, et al. PSTR: end-to-end one-step person search with transformers[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2022: 9448−9457.

[26] 祁磊, 任子豪, 刘俊汐, 等. 虚实结合的行人重识别方法[J/OL]. 计算机研究与发展. [2024−07−10]. http://kns.cnki.net/kcms/detail/11.1777.TP.20240108.1607.014.html.
QI Lei, REN Zihao, LIU Junxi, et al. Person re-identification method based on hybrid real-synthetic data[J/OL]. Journal of computer research and development. [2024−07−10]. http://kns.cnki.net/kcms/detail/11.1777. TP.20240108.1607.014.html.

[27] YUN Sangdoo, HAN Dongyoon, CHUN Sanghyuk, et al. CutMix: regularization strategy to train strong classifiers with localizable features[C]//2019 International Conference on Computer Vision. Los Alamitos: IEEE Computer Society, 2019: 6022−6031.

[28] 戴臣超, 王洪元, 倪彤光, 等. 基于深度卷积生成对抗网络和拓展近邻重排序的行人重识别[J]. 计算机研究与发展, 2019, 56(8): 1632−1641.
DAI Chenchao, WANG Hongyuan, NI Tongguang, et al. Person re-identification based on deep convolutional generative adversarial network and expanded neighbor reranking[J]. Journal of computer research and development, 2019, 56(8): 1632−1641.

[29] WEI Longhui, ZHANG Shiliang, GAO Wen, et al. Person transfer gan to bridge domain gap for person re-identification[C]//2018 IEEE/CVF Conference on Com- puter Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2018: 79−88.

[30] ZHENG Zhedong, YANG Xiaodong, YU Zhiding, et al. Joint discriminative and generative learning for person re-identification[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2019: 2138−2147.

[31] 苗夺谦, 张清华, 钱宇华, 等. 从人类智能到机器实现模型——粒计算理论与方法[J]. 智能系统学报, 2016, 11(6): 743−757.
MIAO Duoqian, ZHANG Qinghua, QIAN Yuhua, et al. From human intelligence to machine implementation model: theories and applications based on granular computing[J]. CAAI transactions on intelligent systems, 2016, 11(6): 743−757.

[32] 郭豆豆, 徐伟华. R-FCCL: 一种面向高维数据的模糊概念认知学习方法[J/OL]. 计算机研究与发展. [2024−03−09]. http://kns.cnki.net/kcms/detail/11.1777.TP.20240307.1525.002.html.
GUO Doudou, XU Weihua. R-FCCL: a novel fuzzy-based concept-cognitive learning approach for high-dimensional data [J/OL]. Journal of computer research and development. [2024−03−09]. http://kns.cnki.net/kcms/detail/11.1777.TP.20240307.1525.002.html.

[33] LI Yanping, MIAO Duoqian, ZHANG Hongyun, et al. Multi-granularity cross transformer network for person re-identification[J]. Pattern recognition, 2024, 150: 110362.

[34] 杨玉婷, 苗夺谦. 基于多粒度匹配的行人搜索算法[J]. 智能系统学报, 2022, 17(2): 420−426.
YANG Yuting, MIAO Duoqian. Person search algorithm based on multi-granularity matching[J]. CAAI transactions on intelligent systems, 2022, 17(2): 420−426.

[35] LIU Zhuang, MAO Hanzi, WU Chaoyuan, et al. A convnet for the 2020s[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2022: 11976−11986.

[36] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: Curran Associates, 2017: 6000−6010.

[37] LIU Hao, FENG Jiashi, JIE Zequn, et al. Neural person search machines[C]//2017 IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society, 2017: 493−501.

[38] MUNJAL B, AMIN S, TOMBARI F, et al. Query-guided end-to-end person search[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2019: 811−820.

[39] 赵才荣, 齐鼎, 窦曙光, 等. 智能视频监控关键技术: 行人再识别研究综述[J]. 中国科学: 信息科学, 2021, 52(12): 1979−2015.
ZHAO Cairong, QI Ding, DOU Shuguang, et al. Key technology for intelligent video surveillance: a review of person re-identification[J]. Scientia sinica informationis,

2021, 52(12): 1979−2015.

[40] LAN Xu, ZHU Xiatian, GONG Shaogang. Person search by multi-scale matching[C]//Proceedings of the 15th European Conference on Computer Vision. Cham: Springer, 2018: 536−552.

[41] HAN Chuchu, YE Jiacheng, ZHONG Yunshan, et al. Re-ID driven localization refinement for person search[C]//2019 IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society, 2019: 9813−9822.

[42] WANG Cheng, MA Bingpeng, CHEN Xilin, et al. TCTS: a task-consistent two-stage framework for person search[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2020: 11949−11958.

[43] XIAO Jimin, XIE Yanchun, TILLO T, et al. IAN: the individual aggregation network for person search[J]. Pattern recognition, 2019, 87: 332−340.

[44] HAN B J, KO K, SIM J Y. End-to-end trainable trident person search network using adaptive gradient propagation[C]//2021 IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society, 2021: 905−913.

[45] LI Yang, XU Huahu, BIAN Minjie, et al. Cross-scale global attention feature pyramid network for person search [J]. Image and vision computing, 2021, 116: 104332.

**作者简介：**

苗春玲，硕士研究生，主要研究方向为行人搜索和深度学习。E-mail：miaochunling@tongji.edu.cn。

张红云，副教授，博士生导师，主要研究方向为粒计算和计算机视觉。E-mail：zhanghongyun@tongji.edu.cn。

苗夺谦，教授，博士生导师，国际粗糙集学会会士，中国人工智能学会会士，嵌入式系统与服务计算教育部重点实验室副主任，上海市计算机学会副理事长，上海市人工智能学会副理事长。主要研究方向为人工智能、机器学习、粒度计算、粗糙集。主持完成国家自然科学基金项目 6 项，主持并参与省部级自然科学基金项目与科技攻关项目 30 余项。获得教育部科技进步一等奖、上海市技术发明一等奖、重庆市自然科学一等奖和中国人工智能学会吴文俊人工智能自然科学二等奖。发表学术论文 180 余篇。E-mail：dqmiao@tongji.edu.cn。

# 第二届中国具身智能大会（CEAI 2025）
# The 2rd China Embodied AI Conference（CEAI 2025）

由中国人工智能学会（CAAI）主办，CAAI 具身智能专委会（筹）、中国科学院计算技术研究所、同济大学和上海交通大学承办的中国具身智能大会（CEAI 2025）定于 2025 年 3 月 28—30 日在北京市举行。本次大会聚焦具身智能领域的最新科研进展和产业应用前沿，以构建广泛覆盖学术界、产业界、政策制定部门以及社会公众的高水平交流与合作平台为目标，推动技术创新、成果转化与产业协同发展。大会立足具身智能技术发展的全局需求，围绕科学研究、技术突破与产业实践的关键议题，致力于促进国内外专家学者深入交流，强化学术界与产业界的互动合作，形成跨领域、多维度的协同创新体系。

CEAI 2025 诚邀国内外顶级专家学者参与，阵容强大，涵盖院士、行业领军人物以及一线科研人员，共同探讨具身智能领域的未来发展。CAAI 名誉理事长、中国工程院李德毅院士，CAAI 理事长、中国工程院戴琼海院士担任大会荣誉主席；中国工程院高文院士，CAAI 监事长、中国工程院蒋昌俊院士，中国工程院于海斌院士共同担任大会主席；中国科学院计算技术研究所蒋树强研究员，上海交通大学卢策吾教授，清华大学刘华平教授，浙江大学杨易教授共同担任程序主席。

**大会具体详情请见通知：**https://www.caai.cn/index.php?s=/home/article/detail/id/4385.html。

**大会秘书处：**CAAI 具身智能专委会（筹）

**联系方式：**caaiembodiedai@163.com, 18503@tongji.edu.cn