



# 智能系统学报

CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

## 基于多粒度匹配的行人搜索算法

杨玉婷, 苗夺谦

引用本文:

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.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202105038>

## 您可能感兴趣的其他文章

### 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>

### 多层卷积特征的真实场景下行人检测研究

Research on pedestrian detection based on multi-layer convolution feature in real scene

智能系统学报. 2019, 14(2): 306–315 <https://dx.doi.org/10.11992/tis.201710019>

### 基于Faster R-CNN的人体行为检测研究

Research on human behavior detection based on Faster R-CNN

智能系统学报. 2018, 13(6): 967–973 <https://dx.doi.org/10.11992/tis.201801025>

### 粒化的Mean Shift行人跟踪算法

Granular mean shift pedestrian tracking algorithm

智能系统学报. 2016, 11(4): 433–441 <https://dx.doi.org/10.11992/tis.201605033>

### 基于车内外视觉信息的行人碰撞预警方法

Pedestrian collision warning system based on looking-in and looking-out visual information analysis

智能系统学报. 2019, 14(4): 752–760 <https://dx.doi.org/10.11992/tis.201801016>

### 一种特征字典映射的图像盲评价方法研究

Blind quality evaluation with image features codebook mapping

智能系统学报. 2018, 13(6): 989–993 <https://dx.doi.org/10.11992/tis.201805027>

微信公众平台



关注微信公众号，获取更多资讯信息

DOI: 10.11992/tis.202105038

网络出版地址: <https://kns.cnki.net/kcms/detail/23.1538.tp.20211116.1547.002.html>

## 基于多粒度匹配的行人搜索算法

杨玉婷<sup>1</sup>, 苗夺谦<sup>2</sup>

(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:** 行人搜索旨在从一系列未经裁剪的图像中对行人进行定位与识别, 融合了行人检测和行人重识别两个子任务。现有的方法设计了基于 Faster R-CNN 的端到端框架来解决此任务, 但是行人检测和重识别两个子任务之间存在特征优化目标粒度不一致问题。为了解决这一问题, 提出一种双全局池化结构, 使用全局平均池化提取检测分支的共性特征, 使用基于注意力机制的全局  $K$  最大池化提取 re-ID 分支的特性特征, 为两个子任务提取符合各自粒度特性的特征。同时由于 re-ID 子任务的细粒度特性, 还提出一种改善粒度匹配的画廊边界框加权算法, 把查询人和画廊边界框的分辨率差异纳入相似度计算。实验证明融入多粒度的方法有效地提高了单阶段算法在 CHUK-SYSU 和 PRW 数据集上的性能。

**关键词:** 行人搜索; 行人检测; 行人重识别; 多粒度; 特征融合; 深度学习; 鲁棒性; 计算机视觉

**中图分类号:** TP389.1   **文献标志码:** A   **文章编号:** 1673-4785(2022)02-0420-07

中文引用格式: 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.

英文引用格式: YANG Yuting, MIAO Duoqian. Person search algorithm based on multi-granularity matching[J]. CAAI transactions on intelligent systems, 2022, 17(2): 420–426.

## Person search algorithm based on multi-granularity matching

YANG Yuting<sup>1</sup>, MIAO Duoqian<sup>2</sup>

(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:** Person search aims to locate and recognize a specified person from a series of uncropped images, which combines Pedestrian Detection and Person Re-identification (re-ID). Existing methods based on Faster R-CNN have been widely used to solve the two subtasks jointly. However, the optimization goals of the two subtasks are inconsistent. To alleviate this issue, we propose a dual global pooling structure, which applies Global Average Pooling to extract common features in detection branch and applies Global K-Max Pooling to extract discriminative features in re-ID branch. In this way, our method successfully extracts features that conform to the granularity characteristics of the two subtasks. In addition, to relieve the granularity mismatch problem, we propose a multi-granularity gallery boxes re-weighting algorithm, which incorporates granularity difference into similarity measurement. Extensive experiments show that our method greatly improves the performance of the end-to-end framework on two widely used person search datasets, CUHK-SYSU and PRW.

**Keywords:** person search; person detection; person re-identification; multi-granularity; multi-feature fusion; deep learning; robustness; computer vision

行人搜索旨在从跨摄像头下找到与待查询目

标行人类别相同的行人在场景中的位置, 与行人重识别 (re-ID)<sup>[1]</sup> 任务相比, 它包含从画廊 (gallery) 中生成行人边界框和匹配行人两个过程。实际应用获取的图像往往是包含场景信息的完整图像,

收稿日期: 2021-05-26. 网络出版日期: 2021-11-17.

基金项目: 国家自然科学基金项目 (61976158, 61976160, 62076182).

通信作者: 苗夺谦. E-mail: [dqmiao@tongji.edu.cn](mailto:dqmiao@tongji.edu.cn).

没有手动标注的行人位置信息,导致现有的 re-ID 方法无法直接使用。而行人搜索更适合于实际应用,例如视频监控、查找罪犯、跨摄像机人员跟踪以及人机交互等。由于人体姿势、场景照明、行人遮挡、分辨率、背景杂乱等因素的复杂变化,使得行人搜索任务具有挑战性,近年来引起了越来越多的研究关注。现有的行人搜索方法将这项任务分为行人检测和行人重识别两个子任务。在一个端到端的多任务框架中处理两个子任务(单阶段方法),或者通过级联行人检测器和行人 re-ID 特征提取器(两阶段方法),训练两个独立的网络分别处理两个子任务。

目前,现有的单阶段行人搜索方法大多是基于 Faster R-CNN<sup>[2]</sup> 的端到端框架,虽然该多任务框架是高效且易于训练的,但是存在行人检测和 re-ID 分支特征优化目标冲突的问题。检测分支旨在从画廊中获取边界框(bounding boxes),并对边界框进行二分类,只区分行人和背景,而不区分行人的具体类别,其关注的是行人之间的粗粒度共性特征;而 re-ID 分支是将检测出的各个疑似目标行人与查询行人进行匹配的过程,其关注的是不同行人之间的细粒度个性特征。粗粒度共性关注的是全局的特征,细粒度个性关注的是局部的特征,导致行人检测和 re-ID 分支特征优化目标冲突。本文在现有的单阶段模型 OIM<sup>[3]</sup> 中融入注意力机制和多粒度的思想,提出了一种双全局池化结构(dual global pooling),使网络可以灵活捕捉全局和局部的联系,并对两个不同的分支提取出符合自身粒度特性的特征,从而改善了特征之间的共性和特性冲突问题。

同时由于 re-ID 子任务的细粒度特性,我们发现从画廊里检测到的不同边界框粒度差异较大。我们把边界框的分辨率定义为它的粒度,分辨率越高,粒度越细;分辨率越低,粒度越粗。粗粒度的边界框往往更容易与查询人计算出较高的相似度,因为粗粒度的边界框中,行人特征比较模糊,网络无法提取出具有判别性的细粒度特征,从而产生错误的匹配结果。为了缓解粒度不匹配问题,本文提出了一种画廊边界框粒度加权算法(granularity weighted similarity, GWS),无需改变模型的复杂度,将画廊边界框的粒度差异纳入相似度计算,为不同粒度的画廊边界框赋予不同的权重,提高细粒度边界框的权重,降低粗粒度边界框的权重。

本文采用端到端的单阶段的行人搜索框架 OIM 作为基准模型,目标是改进行人检测和 re-ID 分支

特征优化目标冲突问题,以及特征相似度计算时查询行人和画廊边界框粒度不匹配的问题。本文主要工作和贡献如下:

1) 针对检测和 re-ID 分支特征优化目标冲突问题,提出一种基于注意力机制的双全局池化结构,对检测分支使用全局平均池化结构(global average pooling, GAP)提取粗粒度特征,对 re-ID 分支使用全局  $K$  最大池化结构(global  $K$ -max pooling, GKMAP)<sup>[4]</sup> 进行细粒度特征提取。

2) 针对特征相似度计算时查询行人和画廊边界框粒度不匹配的问题,提出一种画廊边界框粒度加权算法 GWS,把画廊边界框的粒度差异纳入相似度计算,使粗粒度的边界框获得较低的权重,细粒度的边界框获得较高的权重。

3) 实验证明我们的方法极大地提高了单阶段行人搜索算法在 CUHK-SYSU<sup>[3]</sup> 和 PRW<sup>[3]</sup> 上的性能。

## 1 相关工作

行人搜索作为 re-ID 技术的衍生,如图 1(a)所示,历史并不算久。早期行人检测和 re-ID 被当作两个独立的领域进行研究,近年来随着深度学习的不断发展,研究者们提出了融合行人检测和 re-ID 的检索思想,行人搜索的概念应运而生。目前,基于深度学习的行人搜索主要分为单阶段方法和两阶段方法,如图 1(b)所示。



图 1 行人搜索技术示意

Fig. 1 Schematic diagram of person search

### 1.1 端到端的单阶段方法

单阶段方法<sup>[3, 5-8]</sup>是指在一个端到端的网络中实现行人搜索的功能,主流的方法是沿用 Faster R-CNN 框架并加以改进从而实现多任务的功能。最具代表性的是 Xiao 等<sup>[3]</sup>在 2017 年提出的行人搜索框架 OIM。OIM 将基于 ResNet50<sup>[9]</sup> 的 Faster R-CNN 当作骨干网络,通过在网络中增加一个 re-ID 分支以实现行人识别的功能,并提出了一种在线实例匹配损失函数,使网络可以在具有大而稀



疏的分类任务中更好地收敛;为了进一步提高网络减小类内差异的能力, Xiao 等<sup>[5]</sup>提出 IAN, 沿用 Faster R-CNN 框架并通过引入中心损失增强行人特征的判别性;为了充分利用图像的上下文信息, Yan 等<sup>[6]</sup>在 Faster R-CNN 的顶部建立了图模型来学习行人的上下文信息;为了准确提取边界框内的行人特征, Dong 等<sup>[7]</sup>提出 BINet, 在训练阶段,除了画廊图像外,还以裁剪后的行人图片为输入来帮助模型基于人的外观识别身份,此外还设计了两个交互损失来实现两个级别的分支之间的双向交互。为了充分利用查询行人的信息, QEEPS<sup>[8]</sup>提出以查询为导向的方法,在处理画廊图像时将查询图片送入网络,并受 SEblock<sup>[8]</sup>启发,利用 QSSE-Net<sup>[8]</sup>辅助画廊图像进行行人特征的提取。

## 1.2 级联的两阶段方法

两阶段方法<sup>[10-11]</sup>是指先根据检测器从画廊中检测出所有候选人员,再将其输入到 re-ID 特征提取器中进行识别。Chen 等<sup>[10]</sup>首次提出单阶段的方法存在特征优化目标冲突的问题,并提倡使用行人检测器和 re-ID 网络级联的方式实现行人搜索算法,即为检测和 re-ID 提供两个参数独立的模型。Zheng 等<sup>[11]</sup>测试了检测器和识别器的各种组合,提出了用于训练的级联微调策略和用于匹配的置信加权相似度 (confidence weighted similarity, CWS),将分类置信度从检测器传输到 re-ID 网络,有效地提高了识别的效果。一般来说,两阶段的方法在精度上优于单阶段的方法,但是在速度上略逊一筹。实际应用中往往对行人搜索任务的实时性要求较高,因此将两个子任务融合在一个端到端的框架中解决显得更加方便和高效,同时也更符合人类处理问题的方式。

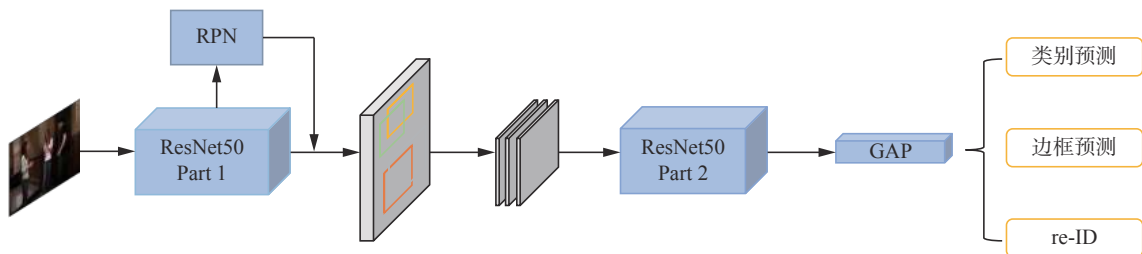


图 2 行人搜索模型 OIM 结构

Fig. 2 Structure of person search model OIM

## 2.2 双全局池化结构

OIM 由检测 (包括分类和回归) 和 re-ID 两个分支组成,并在全连接层之前共享了一个全局平均池化层 (如图 2 所示)。全局池化层通过改变卷积特征图的空间维度,往往可以帮助网络实现更好的分类,然而由于检测和 re-ID 特征的粒度特

## 2 基于注意力机制的多粒度特征建模

在计算机视觉领域,注意力机制主要是为了让模型把注意力放在感兴趣的区域,基本原理很简单:它认为,网络中每层不同特征 (可以是不同通道的,也可以是不同位置的) 的重要性不同,后面的层应该更注重其中重要的信息,抑制不重要的信息。在本节中,首先概述基于 Faster R-CNN 的单阶段行人搜索模型 OIM 的整体结构,然后针对检测和 re-ID 分支优化目标粒度不一致的问题,介绍改进后的模型。

### 2.1 OIM 模型框架

图 2 显示了基于 Faster R-CNN 的单阶段行人搜索方法 OIM 的模型结构,其在 Faster R-CNN 顶部卷积特征的基础上,添加了用于检测、回归和 re-ID 的多任务分支,使模型可以在一个端到端的框架内共同处理行人检测和行人 re-ID 任务。如图 2 所示,给定一张待查询的图像作为输入,首先经过 ResNet50 的第一部分 (res1-res4),将输入图片从原始像素变换到卷积特征图,然后在这些特征图上建立一个 RPN<sup>[2]</sup> 网络 (region proposal network, RPN),从而得到预测候选人的区域提案 (region proposals)。在对这些区域提案进行非极大值抑制<sup>[13]</sup> (non-maximum suppression, NMS) 后,保留 128 个区域提案,并利用 RoI-Align<sup>[14]</sup> 层将这些区域提案转化为  $1024 \times 14 \times 14$  的特征区域。接着将这些池化后的特征区域输入到 ResNet50 的第二部分 (res5) 和一个全局平均池化层 (global average pooling, GAP)。最后经过 2048 维全连接层进行分类和回归,经过全连接层提取 L2 归一化后的 256 维特征,用 OIM 损失<sup>[3]</sup> 对提取的行人特征进行优化。

性不同,使用相同的池化方法无法为每个分支提取出最优的特征。行人检测是粗粒度的任务,关注的是行人这一类别特征之间的整体相似性,需要降低局部判别性特征带来的差异性和对分类结果的影响。如图 3 所示,基于这一特性,本文在检测分支中保留了全局平均池化的结构,便于网

络提取出全局共性信息。而 re-ID 是细粒度的任务,不同类别的行人由于外形、穿着等原因可能看起来很相似,同时同一类别的行人图像可能由于姿态、光照等因素看起来反而不那么相似,因此该任务有着“类间间距小,类内间距大”的特点。对于网络来说,应该更加关注那些能够区分行人的局部特征。本文受到 ELoPE<sup>[4]</sup> 的启发,在网络的 re-ID 分支中融入了注意力机制,如图 3 所示,

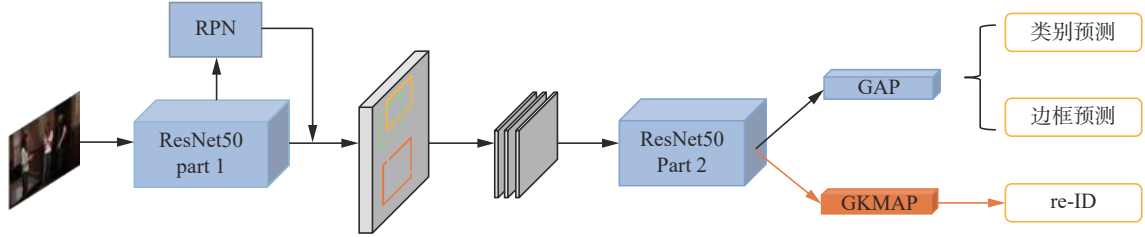


图 3 改进后的 OIM 模型结构

Fig. 3 Structure of improved OIM model

全局  $K$  最大池化定义如下: 给定输入图像  $x$ ,  $y \in \mathbf{R}^{C \times H \times W}$  是最后一个卷积层的输出, 其中  $C$  是通道数, 每一个通道内的特征图的大小为  $H \times W$ 。

给定一个  $c \in \{1, 2, \dots, C\}$ , 降序排列后的向量  $V_c$  定义如下:

$$V_c = \text{sort}\{y_{c,h,w} | h \in \{1, 2, \dots, H\}, w \in \{1, 2, \dots, W\}\} \quad (1)$$

基于给定  $K$  值的全局  $K$  最大池化的定义如下:

$$\text{GKMAP} = \frac{1}{K} \sum_{k=1}^K V_{c,k} \quad (2)$$

如果  $K=H \times W$ , 式 (2) 则可以表示标准的全局平均池化。

全局平均池化的定义如下:

$$\text{GAP} = \frac{1}{H \times W} \sum_{k=1}^{H \times W} V_{c,k} \quad (3)$$

本文中, 通过引入双全局池化结构, 改进后的 OIM 模型可以为检测分支提取出全局的粗粒度共性特征, 为 re-ID 分支提取出局部的细粒度个性特征, 从而能够灵活捕捉全局和局部的联系, 专注于提取符合两个子任务粒度特性的特征。

### 3 边界框多粒度加权的相似度计算

OIM 模型在计算查询人和画廊图像的相似度时, 将画廊中检测到的不同边界框权重视作相等, 这就产生了查询行人和画廊边界框粒度不匹配的问题。为了改善粒度不匹配问题, 本文提出一种画廊边界框粒度加权相似度计算 (granularity weighted similarity, GWS)。

GWS 算法定义如下: 给定一个查询人边界框  $q$  和一个画廊边界框  $g$ ,  $q$  和  $g$  的相似度计算定义如下:

用全局  $K$  最大池化结构 (global  $K$ -max pooling, GKMAP) 代替了原来的全局平均池化结构。GKMAP 通过提取特征图中  $K$  个最重要的局部位置, 使得用于 re-ID 的特征更具判别性。全局平均池化是对最后一个卷积层输出的每一个特征图的所有值进行平均运算, 而全局  $K$  最大池化是对最后一个卷积层输出的每一个特征图先选定  $K$  个最大值, 然后对每个特征图中的  $K$  个最大值进行平均运算。

$$\text{sim}(q, g) = \frac{\sum_{i=1}^n x_{q_i} x_{g_i}}{|X_q| \cdot |X_g|} \quad (4)$$

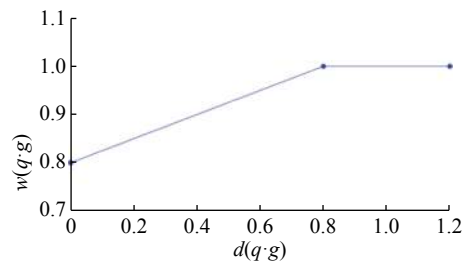
其中  $X_q = \{x_{q_1}, x_{q_2}, \dots, x_{q_n}\}$  和  $X_g = \{x_{g_1}, x_{g_2}, \dots, x_{g_n}\}$  分别表示  $q$  和  $g$  的特征。

我们用  $\text{area}(b) = b_w \cdot b_h$  定义边界框的面积, 并定义函数  $d(q, g)$  来衡量  $q$  和  $g$  之间的粒度差异。

$$d(q, g) = \begin{cases} \frac{\text{area}(g)}{\text{area}(q)}, & \text{area}(g) < \text{area}(q) \\ 1, & \text{其他} \end{cases} \quad (5)$$

式中:  $d(q, g)$  的值越接近 1, 说明  $q$  和  $g$  之间的粒度差异越小,  $d(q, g)$  的值越小, 说明  $g$  相对于  $q$  的粒度越粗。

检测框  $g$  的权重变化函数  $w(q, g)$  如图 4 所示。

图 4  $w(q, g)$ 示意图Fig. 4 Structure of  $w(q, g)$ 

定义阈值  $k_1$  和  $k_2$ ,  $k_1$  表示开始降低权重时横坐标  $d(q, g)$  的阈值,  $k_2$  表示  $d(q, g)$  为 0 时纵坐标  $w(q, g)$  的值, 若  $d(q, g) \geq k_1$ , 则  $w(q, g)$  设置为 1, 保持不变; 若  $d(q, g) < k_1$ , 则按照线性衰减因子降低检测框  $g$  权重。

粒度加权相似度计算 GWS 的定义如下:

$$\text{GWS}(q, g) = \text{sim}(q, g) \cdot w(q, g) \quad (6)$$

GWS 算法是受到 CWS 算法<sup>[11]</sup> 的启发, 当画

廊边界框的质量较低(如包含大范围背景或者边界框分类错误)时, re-ID 的准确率不可避免地会受到影响, CWS 通过将检测置信度纳入相似度计算来解决这个问题, 使得那些高质量的边界框比低质量的边界框拥有更高的权重, 从而计算出更符合实际的相似度。通过实验证明, GWS 可以在 CWS 的基础上进一步提高 OIM 的精度。

置信度加权相似度计算 CWS 的定义如下:

$$CWS(q, g) = \text{sim}(q, g) \cdot \text{det}(g) \quad (7)$$

$$\text{det}(g) = \frac{\exp(W_y X)}{\sum_{c=1}^C \exp(W_c X)} \quad (8)$$

其中  $C=2$  和  $y=\{0, 1\}$  代表前景和背景边界框两种类别。 $\text{det}(g)$  表示网络的检测分支为边界框  $g$  预测的置信度。

综上, 本文提出的 GWS 可以和 CWS 融合成一种新的画廊边界框加权算法 GWS+, 定义如下:

$$GWS(q, g)^+ = \text{sim}(q, g) \cdot w(q, g) \cdot \text{det}(g) \quad (9)$$

## 4 实验结果

### 4.1 数据集

**CUHK-SYSU:** CUHK-SYSU 是一个大型行人搜索数据集, 由摄像机拍摄的街道行人照片和从电影中收集的照片两部分组成。它包含 1818 张画廊图像, 8432 个带标签的行人和 96143 个带标注的边界框。每个被标记的行人都分配有一个类别标签, 并且属于同一个类别标签的行人至少以不同的角度出现在两张不同的画廊图像中, 未标记类别的行人被标记为未知人员。训练集包含 11 206 个画廊图像和 5 532 个行人类别, 而测试集包含 6 978 个画廊图像和 2 900 个查询人。在测试集中, 对于每个查询人, 其画廊图像的数目在 50~4 000, 本文实验默认将画廊图像的数目设置为 100。

**PRW:** PRW 数据集由从大学校园中拍摄的视频中提取的 11 816 个视频帧组成。它包含 932 个带标签的行人和 34 304 个带标注的边界框。与 CUHK-SYSU 类似, 标注分为带标签的行人类别和未标记的行人。训练集包含 5 704 张图像和 482 个行人类别, 测试集包含 2 057 名查询人员, 每个人都将在具有 6 112 张图像的画廊中进行搜索。因此, 画廊大小明显大于 CUHK-SYSU 的默认设置。

### 4.2 评价指标

本文使用 mAP (mean average precision)<sup>[3]</sup> 和 CMC (cumulative matching characteristics)<sup>[3]</sup> 作为衡量行人搜索性能的标准指标。由于行人搜索包

含了检测边界框的过程, 因此仅当排名候选框与真值边界框 (ground truth) 的 IoU (Intersection over Union) 大于 0.5 时才被认为是正确的, 这是与 re-ID 方法的主要区别。

### 4.3 实验设置

使用 PyTorch 来实现 OIM 模型, 并在 NVIDIA 2080Ti GPU 上运行实验。采用 ImageNet<sup>[15]</sup> 预先训练的 ResNet50 作为骨干网络, 把前 4 个残差块 (res1-res4) 用作主干网络, 然后用标准的 RPN 生成行人候选边界框, 接下来经过 RoI-Align 层将候选框的尺寸重塑为 14×14, 然后经过 ResNet50 的 res5 残差块。在训练过程中, 采用 SGD 算法, 将动量设置为 0.9, 权重衰减设置为 0.000 1, 批量大小设置为 2。对于 CUHK-SYSU, 学习速率被初始化为 0.001, 在 40 000 迭代之后下降到 0.000 1, 并保持不变, 直到 50 000 迭代。

### 4.4 实验结果与分析

将在 CUHK-SYSU 和 PRW 数据集上进行消融实验, 以探索本文提出的方法对实验结果的贡献, 包括双全局池化结构和 GWS 算法。基于 PyTorch 重新实现了 OIM 模型, 表 1 给出在两个数据集上进行的消融实验结果, 其中 OIM 是我们重新实现的基准模型, OIM+GWS 是基准模型加上 GWS 算法, OIM+CWS 是基准模型加上 CWS 算法, OIM+GWS<sup>+</sup> 是基准模型加上融合了 CWS 的 GWS 算法。2pool 是使用双全局池化结构改进后的模型, 2pool+GWS 是双全局池化结构和 GWS 算法, 2pool+CWS 是双全局池化结构和 CWS 算法, 2pool+GWS<sup>+</sup> 是双全局池化结构加上融合了 CWS 的 GWS 算法。

**双全局池化结构的有效性:** 表 2 给出在 CUHK-SYSU 数据集上单全局池化结构与双全局池化结构的对比实验 (其中  $K=4$ )。其中 OIM\_1GAP 表示使用单全局平均池化的基准模型, OIM\_1GAP 表示使用单全局最大池化的模型, OIM\_1GKMP 表示使用单全局  $K$  最大池化的模型, OIM\_2pool 表示使用双全局池化结构改进后的模型。通过分析可以看出, 无论是共享全局平均池化层, 还是共享全局  $K$  最大池化层, 使用双全局池化层的效果要优于共享全局池化层, 说明对检测分支和 re-ID<sup>[15]</sup> 分支使用不同的池化方法可以优化特征共性-特性冲突问题。同时, 由于本文方法对模型的改动是微小的, 与 OIM 基准模型相比, 本文方法对速度的影响是微小的, 与 OIM 相差无几。

**$K$  的取值:** 在 CHUK-SYSU 数据集上, 在 GKMAP 中为  $K$  选择合适的值的实验探究如图 5 所示。当  $K=1$  时与 GMP 等效,  $K=196$  时与 GAP 等效 (最



后一个卷积层的空间尺寸为  $14 \times 14$ )。如图5所示,观察到  $K=4$  时,结果的效果是最佳的。

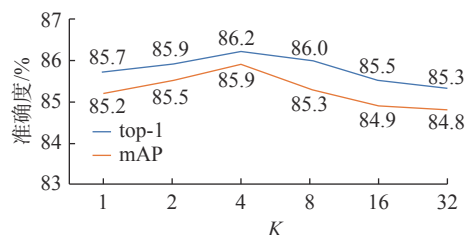


图5  $K$  的取值对比实验

Fig. 5 Different values of  $K$

GWS 的有效性: 图6给出 GWS 的可视化结果示例, GWS 修正了 OIM 的错误结果。我们分别选择了  $k_1=0.8$  和  $k_2=0.8$ 。在表1中, 将 GWS 应用到 OIM 后, 在 CUHK-SYSU 上, mAP 和 top-1 分别提高了 0.5% 和 0.4%; 在 PRW 上, mAP 和 top-1 分别提高了 0.4% 和 0.3%。将 GWS 添加到改进后的 2pool\_OIM 中, 在 CUHK-SYSU 上, mAP 和 top-1 分别提高了 0.4% 和 0.4%; 在 PRW 上, mAP 和 top-1 分别提高了 0.3% 和 0.2%。同时从表1可以看到, 融合了 CWS 的 GWS<sup>+</sup>在 OIM 和改进后的 OIM 仍然保持有效性, 以上结果证明了 GWS 的确可以改善粒度不匹配问题。



图6 GWS 可视化结果示例

Fig. 6 Visualization examples of GWS

表1 本文方法在 CUHK-SYSU 和 PRW 上的实验结果  
Table 1 Result of proposed method on CUHK-SYSU and PRW %

方法	CUHK-SYSU		PRW	
	mAP	top-1	mAP	top-1
OIM	84.7	85.3	34.0	75.9
OIM+GWS	85.2	85.7	34.4	76.6
OIM+CWS	86.9	87.2	35.7	77.2
OIM+GWS <sup>+</sup>	87.3	87.6	36.0	77.5
2pool	85.9	86.2	35.1	76.8
2pool+GWS	86.3	86.6	35.4	77.0
2pool+CWS	87.7	87.7	36.3	77.9
2pool+GWS <sup>+</sup>	88.1	87.9	36.6	78.1

表2 单双全局池化层结构对比实验 (取  $K=4$ )  
Table 2 Experiments of different pooling structures %

方法	mAP	top-1	$\Delta$ mAP	$\Delta$ top-1
OIM_1GAP	84.7	85.3	—	—
OIM_1GMP	85.0	85.4	0.3	0.1
OIM_1GKMP	85.5	85.7	0.8	0.4
OIM_2pool	85.9	86.2	1.2	0.9

## 5 结束语

本文针对单阶段行人搜索模型存在的特征优化目标冲突问题, 在现有的单阶段模型中融入注意力机制和多粒度的思想, 提出了一种双全局池化结构, 使网络可为不同分支提取出符合自身粒度特性的特征。针对查询人和画廊边界框粒度不匹配的问题, 本文提出一种改善粒度匹配的画廊边界框加权算法, 将检测框的分辨率差异纳入相似度计算中从而改善了粒度不匹配的问题。本文方法有效地提高了单阶段算法在 CHUK-SYSU 和 PRW 数据集上的性能。

## 参考文献:

- [1] 刘皓. 基于深度学习的行人再识别问题研究 [D]. 合肥: 合肥工业大学, 2017.  
LIU Hao. Research on problems of person re-identification based on deep learning[D]. Hefei: Hefei University of Technology, 2017.
- [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] XIAO Tong, LI Shuang, WANG Bochao, et al. Joint detection and identification feature learning for person search[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 3376–3385.
- [4] HANSELMANN H, NEY H. ELoPE: fine-grained visual classification with efficient localization, pooling and embedding[C]//Proceedings of 2020 IEEE Winter Conference on Applications of Computer Vision. Snowmass, USA, 2020: 1236–1245.
- [5] XIAO Jimin, XIE Yanchun, TILLO T, et al. IAN: the individual aggregation network for person search[J]. Pattern recognition, 2019, 87: 332–340.
- [6] YAN Yichao, ZHANG Qiang, NI Bingbing, et al. Learning context graph for person search[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and

- Pattern Recognition. Long Beach, USA, 2019: 2153–2162.
- [7] DONG Wenkai, ZHANG Zhaoxiang, SONG Chunfeng, et al. Bi-directional interaction network for person search[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020: 2836–2845.
- [8] MUNJAL B, AMIN S, TOMBARI F, et al. Query-guided end-to-end person search[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 811–820.
- [9] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 770–778.
- [10] 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. Munich, Germany, 2018: 764–781.
- [11] ZHENG Liang, ZHANG Hengheng, SUN Shaoyan, et al. Person re-identification in the wild[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 3346–3355.
- [12] LAN Xu, ZHU Xiatian, GONG Shaogang. Person search by multi-scale matching[C]//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany, 2018: 553–569.
- [13] HOSANG J, BENENSON R, SCHIELE B. Learning non-maximum suppression[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 6469–6477.
- [14] HE Kaiming, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy, 2017: 2980–2988.
- [15] DENG Jia, WEI Dong, RICHARD Socher, et al. ImageNet: A large-scale hierarchical image database[C]//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Florida, USA, 2009: 248–255.

#### 作者简介:



杨玉婷, 硕士研究生, 主要研究方向为深度学习、计算机视觉、行人搜索和粒计算。



苗夺谦, 教授, 主要研究方向为人工智能、机器学习、大数据分析和粒计算。主持、参与国家自然科学基金项目及横向项目多项。发表学术论文 180 余篇。