



激光与光电子学进展
Laser & Optoelectronics Progress
ISSN 1006-4125, CN 31-1690/TN

《激光与光电子学进展》网络首发论文

题目: 基于多尺度注意力机制的多分支行人重识别算法
作者: 李聪, 蒋敏, 孔军
收稿日期: 2020-01-06
网络首发日期: 2020-06-02
引用格式: 李聪, 蒋敏, 孔军. 基于多尺度注意力机制的多分支行人重识别算法[J/OL]. 激光与光电子学进展.
<http://kns.cnki.net/kcms/detail/31.1690.TN.20200601.1753.161.html>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

基于多尺度注意力机制的多分支行人重识别算法

李聪, 蒋敏*, 孔军

江南大学物联网工程学院 江苏省模式识别与计算智能工程实验室, 江苏, 无锡 214122

摘 要 行人重识别任务往往存在姿态变化, 遮挡以及行人图像类间相似度高问题, 传统基于深度学习的方法更多地关注行人图像中最显著的特征, 对于次显著细节信息关注不足。针对该问题, 本文提出了一种基于多尺度注意力机制的多分支网络来统筹图像显著信息及次显著信息。首先, 本文提出将多尺度特征融合方法 (MSFF) 与注意力机制相结合, 设计了一个多尺度注意力模块 (MSA), 使得网络可以根据输入信息自适应地调节感受野大小, 实现对于不同尺度信息的充分利用。其次, 建立一个多分支网络, 实现对于全局特征和多元局部特征的协调统一, 并利用多尺度注意力模块, 分别实现对于全局显著信息, 及次显著局部细节信息的加权强化, 得到更具判别性的特征用于最终的识别。实验结果表明, 本文所设计的网络在多个数据集上都取得了较好的表现。

关键词 深度学习; 行人重识别; 注意力机制; 多尺度特征; 局部特征

中图分类号 TP391.4 **文献标志码** A

Multi-branch Person Re-identification Based on Multi-scale Attention

Li Cong, Jiang Min*, Kong Jun

Jiangsu Provincial Engineering Laboratory of Pattern Recognition and Computational Intelligence, School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China

Abstract Person re-identification often have problems such as pose changes, occlusions, and high similarity between different person images. Traditional deep learning-based methods pay more attention to the most significant features in person images, but pay insufficient attention to details. To solve these issues, a multi-branch network based on multi-scale attention mechanism is proposed to coordinate significant information and sub-significant information. Firstly, the proposed algorithm proposes to combine the Multi-scale Feature Fusion module (MSFF) with the attention mechanism to get a Multi-scale Attention module (MSA). This module enables the network to adaptively adjust the size of the receptive field according to the input information, so as to make full use of information of different scales. At the same time, a multi-branch network is established to realize the coordination of global features and multiple local features. Using the Multi-scale Attention module (MSA), weighted enhancement of global information and local detail information can be achieved separately, and a more discriminative feature is obtained for final recognition. The experiment results show that the proposed method has achieved good performance on multiple datasets.

Key words deep learning; person re-identification; attention; multi-scale feature; local feature

OCIS codes 100.4996; 100.5010; 150.0155; 200.4560

基金项目: 国家自然科学基金 (61362030, 61201429)、中国博士后科学基金 (2015M581720, 2016M600360)、科技援疆专项计划 (2017E0279)、江苏博士后科学基金 (1601216C)

*E-mail: minjiang@jiangnan.edu.cn;

1 引言

行人重识别旨在通过匹配行人图像,从部署在不同位置的互不交叉的监控摄像头来寻找特定行人,对于该问题的不断深入研究对公共安全和刑侦以及智能安防等有着非常重要的现实意义^[1]。提取准确而有效的特征是实现行人重识别任务的关键,文献[2]提出了一种基于视角信息嵌入的模型,结合行人图像视角朝向特点,对于行人重识别网络进行了优化。文献[3]提出了一种卷积神经网络判别特征学习的模型,通过加强对样本监督信息的利用和提升对样本间距离关系的约束,提升网络提取判别性特征的能力。

随着视角的变化,图像中行人的面部、手、脚等身体部位会变得不稳定,卷积神经网络(CNN)往往会把焦点放在身体的主要部位,而其他具有描述性的身体部位则容易被忽视^[4]。局部特征可以关注更多的细节,应对遮挡,姿态变化等问题有更好的表现。为了学习更加精细且有判别性的特征,许多工作提出将特征进行切分并分别计算损失的方法,如局部卷积基准网络(PCB)^[5],多粒度网络(MGN)^[6],多尺度卷积特征融合网络^[7]等。但是简单的水平切分往往存在一定的限制,一些完整的特征(比如背包等)可能被强制切分,反而造成了有效信息的丢失。文献[8]提出一种结构化特征丢弃方法(DropBlock),对特征图中的邻近矩形区域进行随机丢弃,实现对于细节信息的关注。文献[4]提出批量特征丢弃方法(BDB),对于一个批量的特征图,随机遮挡同样的矩形区域,强迫网络在保留的区域中去学一些细节的局部特征。然而,这些方法往往难以对于保留区域中的不同局部特征的有效性进行准确地评判,进而实现更加细致地加权强化。

另外,为了处理行人重识别任务中边界框检测不完善以及身体部分错位等问题,许多工作提出利用注意机制来聚焦特定区域捕获特征。文献[9]提出一种注意力组成网络(AACN),利用注意力模块来提取行人图片的姿态信息和局部信息以去除背景的干扰。文献[10]提出了一种双重注意力匹配方法(DuATM)来学习上下文特征序列并同时实现细致地序列比较。文献[11]提出了一个协调注意力网络(HA-CNN),通过融合了基于像素的软注意力与基于区域的硬注意力方法,致力于优化未对齐的行人图像。但这些注意力方法普遍只能提取固定尺度的特征,不能使得网络根据输入信息自适应地调节感受野大小来实现更加合理的权重分配。

针对这些方法存在的问题,本文提出将多尺度特征融合方法(Multi-scale Feature Fusion, MSFF)与注意力机制相结合,设计了一个多尺度注意力模块(Multi-scale Attention module, MSA),使得网络能够根据输入自适应地调整感受野大小,并实现对于图像中不同尺度目标的加权关注。同时,将MSA模块与批量特征丢弃方法^[2]相结合,设计了一个多尺度注意力特征丢弃模块(Multi-scale Attention-aware Feature DropBlock module, MSA-FD),实现对于局部特征中不同位置的权重调整,使得网络自适应地关注不同尺度的显著信息与次显著信息。另外本文提出一种图像模糊切分策略(Fuzzy partition),与MSA-FD模块相互补充来获取更

加丰富的局部细节信息，最终得到一个基于多尺度注意力机制的多分支网络来统筹全局特征和局部特征，优化行人重识别任务中受遮挡的，未对齐的行人图像。

2 基本原理

2.1 网络框架

本文以 Resnet-50 为基础框架建立了一个多分支网络，分别从多角度提取更具判别性的特征完成行人重识别任务。其中第一个分支是基于多尺度注意力机制的全局分支，该分支在没有区域划分的情况下学习全局特征表示。同时设计了一个多尺度注意力模块（MSA），使得网络能够根据输入信息自适应地调节感受野大小，实现对于全局特征的加权强化。第二个分支和第三个分支分别从两个角度来联合提取多元局部特征，获取更加丰富的细节信息。其中，为了保留更多的细节信息，二、三分支取消了基础网络中的下采样操作。第二个分支是基于多尺度注意力机制的特征丢弃分支，该分支将 MSA 模块与批量特征丢弃方法^[4]相结合，设计了一个多尺度注意力特征丢弃模块（MSA-FD），更好地实现对于保留区域中次显著信息的关注。第三个为图像模糊分块分支，在传统图像均匀切分的基础上进行了微调，采用了一种模糊分块策略，使得上中下三个切块中互相包含一部分信息，减少了暴力切分带来的信息损失的问题。通过全局特征和局部特征的统一结合，网络能够提取更加丰富的信息和细节去表征输入图像的完整情况。

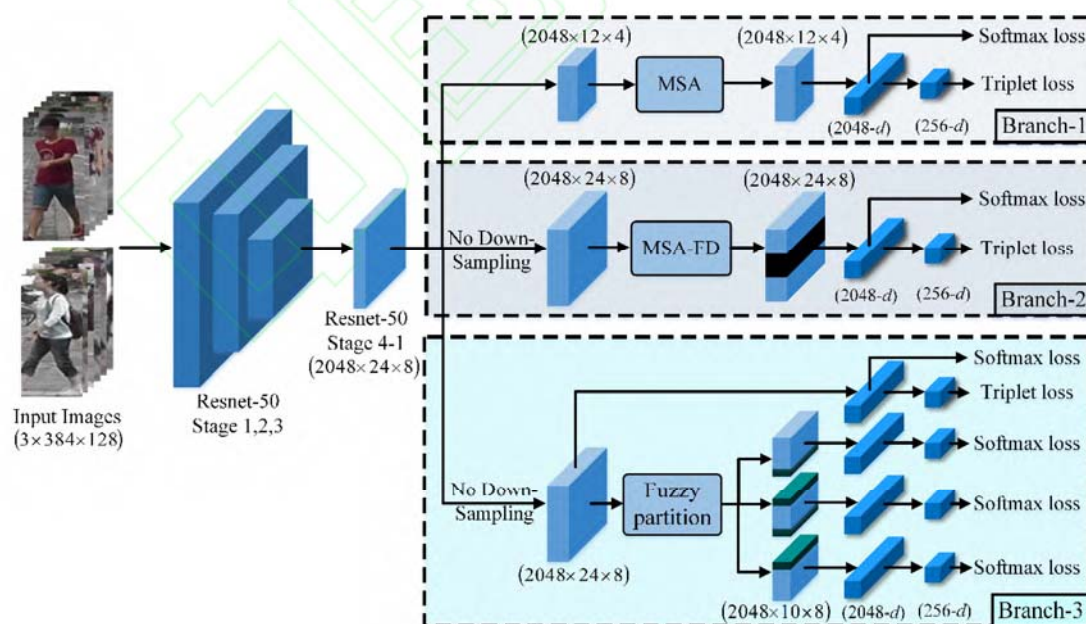


图 1 基于多尺度注意力机制的多分支行人重识别网络框架图

Fig.1 Multi-branch person re-identification framework based on Multi-scale Attention mechanism

2.2 基于多尺度注意力机制的全局分支

传统注意力模块中，每一层人工神经元感受野被设计为相同大小，忽略了不同尺度特征对于模型贡献^{[12][13]}。针对这一问题，本文将多尺度特征融合方法（MSFF）与双重注意力模块^[14]相结合来实现融合利用多尺度信息，最终得到了一个多尺度注意力模块（MSA），使得网络能够根据输入自适应地调节感受野大小，来实现对于全局特征更加细致准确地加权强化。

首先对于 MSFF 模块，给定当前输入特征 $X_{bl} \in R^{C \times H \times W}$ ，其中 C ， H ， W 分别代表特征的通道数以及在特征高度和宽度上的元素数目。为了减少模型的参数量，首先利用一个卷积核为 1×1 的卷积操作将特征通道数目调整为原来的 $1/r$ ，得到降维后的特征 $X' \in R^{(C/r) \times H \times W}$ ，（在实验过程中， r 的数值被设置为 8）。将特征 X' 分别送入四个感受野为 R （分别取 3，5，7，9^{[12][13]}）的深度可分离卷积模块（Depthwise Separable Convolutions, DSC）^[15]得到四组不同尺度的特征，从而使最终的特征能够从整个空间尺度范围捕获信息。分别将四组不同尺度的特征传入一个参数共享的压缩激励模块（Shared SE module）^[16]实现特征通道间信息交互，然后将四个尺度的特征进行按元素相加，自适应指导多个尺度特征地动态融合，得到一个包含多尺度信息的融合特征 $X_{ms} \in R^{(C/r) \times H \times W}$ 。将特征 X_{ms} 与通道降维后的输入特征 X' 按元素相加，最终得到多尺度融合特征 $X_o \in R^{(C/r) \times H \times W}$ 进行后续利用。

将第一个分支中经过 Resnet-50 网络第 4 阶段得到的特征 $X_{bl} \in R^{C \times H \times W}$ 分别送入三个 MSFF 模块，得到三个多尺度融合特征 $Q \in R^{(C/r) \times H \times W}$ ， $K \in R^{(C/r) \times H \times W}$ 和 $V \in R^{(C/r) \times H \times W}$ ，利用三者之间的相互关联信息建立注意力模块，可以得到一个更具判别性的全局特征表达。首先，通过对于特征 Q 和 K 进行双线性池化^[17]来完成全局信息的聚合。具体来说，分别对 Q 和 K 进行尺寸调整得到特征 $Q_r \in R^{(C/r) \times HW}$ 和 $K_r \in R^{HW \times (C/r)}$ 。进一步地，对于特征 K_r 通过一个 *softmax* 层得到一个权重特征 $K_s \in R^{HW \times (C/r)}$ 。然后将特征 Q_r 与其进行矩阵相乘操作，得到一个包含二阶统计信息的特征 $G \in R^{(C/r) \times (C/r)}$ 来聚合整个空间的全局信息。

$$G = Q_r \text{softmax}(K_r) \quad (1)$$

从输入的整个空间收集特征的下一步就是根据当前局部特征的需要，对聚合的全局特征进行自适应加权分布。利用这种方式可以实现通过一组紧凑的特征来捕获更复杂的全局关系。同样地，对于特征 $V \in R^{(C/r) \times H \times W}$ ，对其进行尺寸调整得到特征 $V_r \in R^{HW \times (C/r)}$ ，然后附加一个 *softmax* 层得到一个指导全局特征分布的权重向量。权重向量中对应的每个位置都可以接收特定的全局信息，这是对现有局部特征的补充，促进学习更复杂的关系。将该权重向量与聚合的全局特征 G 相乘，进行特征加权分配，最终得到空间层面更加完备的特征 $D \in R^{(C/r) \times H \times W}$ 。

$$D = \text{softmax}(V_r) \times G \quad (2)$$

最后通过一个卷积核为 1×1 的卷积层将特征 D 的通道数目调整到 C ，得到最终的多尺度加权特征 $O \in R^{C \times H \times W}$ 。

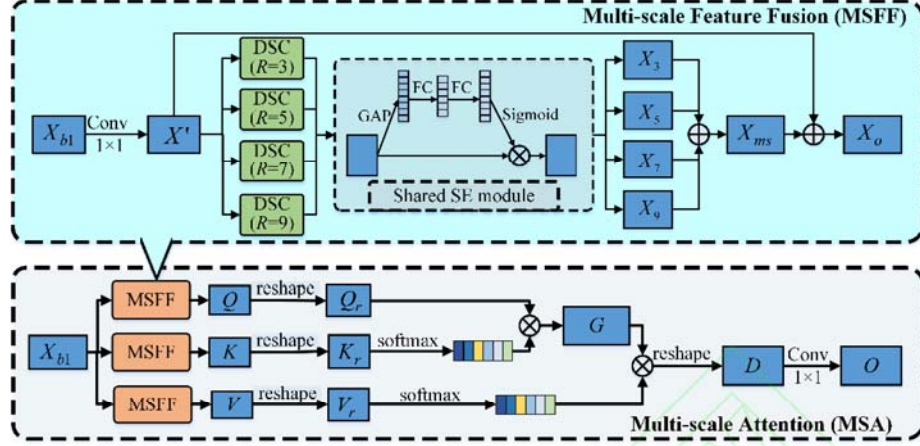


图2 多尺度注意力模块结构图
Fig.2 The structure of Multi-scale Attention module

2.3 基于多尺度注意力机制的特征丢弃分支

联合全局特征和局部特征来进行物体识别被证明是一种有效的策略，批量特征丢弃方法^[4]以批量方式随机丢弃所有输入特征图的相同区域，以增强局部区域更加精细化的特征学习。但该方法难以实现对于保留区域特征的加权关注，因此本文将 MSA 模块与批量特征丢弃方法相结合，设计了一个多尺度注意力机制特征丢弃模块（MSA-FD），实现对于保留局部区域的特征权重调整，增强了模型对于局部区域显著不同尺度特征的关注，减少了无关信息对于模型的影响。

MSA-FD 模块与 MSA 模块结构大致相同，但对于第二个分支中经过 MSFF 模块得到的特征 $K_2 \in R^{(C/r) \times H \times W}$ 和 $V_2 \in R^{(C/r) \times H \times W}$ ，随机选择二者相同位置的连续矩形区域进行丢弃，强迫网络在保留的区域里面去学一些细节的局部特征。其中此处 C, H, W 仅代表其自身含义，不代表具体数值，由于二、三分支中取消了下采样操作，该模块中 H 和 W 的值均为全局分支中 MSA 模块的两倍。在后续注意力权重计算过程中， K_2 和 V_2 两个特征分别起到对于特征进行聚合和分布的作用。在对两个特征利用 *softmax* 层产生权重向量之前，通过特征丢弃的方法随机选择特征图中一个矩形区域进行丢弃。具体来说，首先产生一个二进制掩膜 $m \in R^{1 \times H \times W}$ ，其中随机选择 $(0.3H) \times W$ 的矩形区域元素置为 1，其他位置则置 0。接下来将其 m 乘以 $-\infty$ 并将其与 K_2, V_2 的各通道相加，使得两组特征并被丢弃区域中对应元素均变为 $-\infty$ ，其他位置则保持不变。这样做的目的是在通过 *softmax* 层产生权重向量的时候，利用 $\exp(-\infty) \approx 0$ 的思想，将保留区域和被丢弃区域区分开来。这样，被丢弃区域中所有位置被分配 0 的权重，而保留区域共享和为 1 的概率，保证了重要特征可以得到更多关注，进而实现对于不同局部信息地差别关注。后续的工作与 MSA 模块一致，通过建立 Q_2, K_2 和 V_2 三组

多尺度特征的相互关系，最终得到一个实现区域加权分配的特征，进而提取更加丰富且有判别性的细节信息进行更加准确有效地匹配。

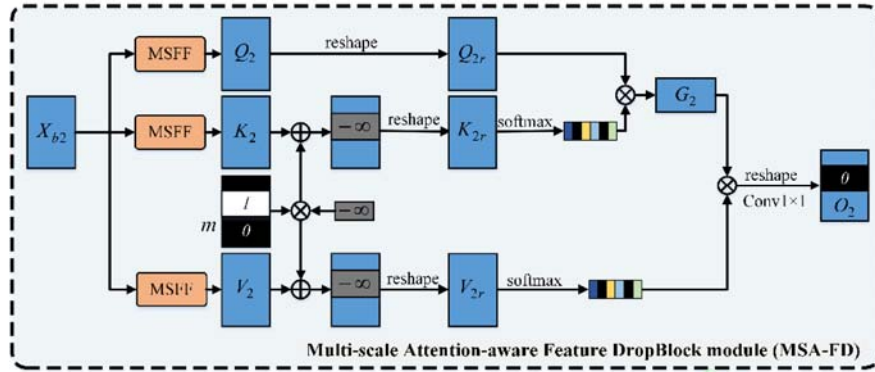


图3 基于多尺度注意力机制的特征丢弃方法结构图

Fig.3 The structure of Multi-scale Attention-aware Feature DropBlock module

2.4 图像模糊分块分支

图像分块策略也是行人重识别任务中获取局部特征的另一有效手段，本文将作为第三分支与第二分支的 MSA-FD 模块联合应用，提取更加细节的有判别力的行人局部特征。本文在传统均匀分块的基础上对于切块策略进行了微调，采用一种模糊分块（Fuzzy partition）的方法。给定网络第三分支传递来的特征 $X_{b3} \in R^{2048 \times 24 \times 8}$ ，如图 4 所示沿着垂直方向将图像分为上中下三部分，每一部分的尺寸分别为 $2048 \times 10 \times 8$ ，上中、中下部分均有一定程度的重合，互相包含切分边缘的部分信息。最后，分别计算每一部分的损失，这样就减少了暴力切分造成的信息丢失的现象。将重叠部分尺寸设定为 2×8 的区域，可以在减少信息损失的同时，又避免了全局显著特征对于细节信息的压制。

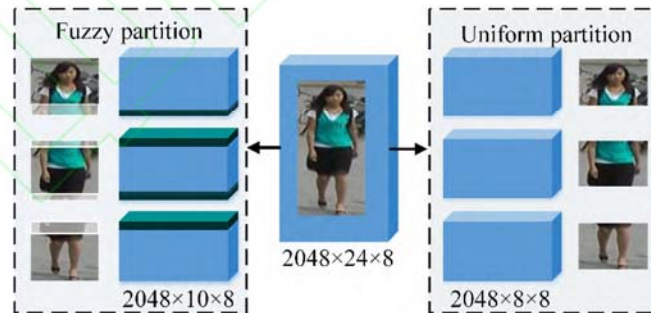


图4 两种分块策略对比图

Fig.4 The comparison of two partition strategies

2.5 训练策略

本文的训练策略沿用了 MGN^[6]方法中三元组损失以及分类损失联合训练的方法，有效地避免了损失权重失衡以及收敛困难的问题。如图 1 所示，对于整体输出的训练，将分类损

失 (softmax) 应用于分支 2 和分支 3 中降维后的 256 维局部特征, 以及三个分支中所有未进行降维的 2048 维全局特征。而对于其他降维的特征则应用三元组损失进行训练。而在测试阶段, 则是将所有的输出特征串接起来, 然后通过计算特征间距离的方式进行排序。

3 实验结果及分析

3.1 数据集及评估协议

3.1.1 数据集介绍

本文在三个主流的行人重识别数据集 Market-1501^[18]、DukeMTMCreID^[19]和 CUHK03^[20]上进行了实验, 并将实验结果与多种最新的行人重识别算法进行了对比。

Market-1501 数据集包含来自 6 个不同摄像机的 1501 个行人的图像。整个数据集分为训练集和测试集两部分, 训练集包含来自 781 个行人的 12936 张图像, 测试集包含来自 750 人的 19732 张图像。3368 张查询图像的行人检测矩形框是人工绘制的, 而图库中的行人检测矩形框则是使用 DPM 检测器^[21]检测得到的。

DukeMTMC-reID 数据集是由 8 个高分辨率摄像机拍摄的来自 1402 个行人的 36411 张图像组成。训练集由从数据集中随机选取 702 个行人的 16,522 张图像组成。测试集包含剩余的 702 人的 2,228 张查询图像和 17,661 张图库图像。

CUHK03 数据集包含由 5 对摄像机拍摄的 1467 个行人的 14097 张图像组成, 同时分为手动标记行人边界框(CUHK03(Labeled))和 DPM 检测的行人边界框(CUHK03(Detected))两部分。其测试协议分为两种, 本文实验采用了最新的将数据将分为包含 767 个行人的训练集和包含 700 个行人的测试集的分离策略。

3.1.2 评估协议

本文实验中, 为了验证所提出行人重识别模型的有效性, 采用了通用的累积匹配特性曲线(CMC)中的 Rank-1 以及平均精度均值(mAP)两个指标对实验结果进行评估。

CMC-k(也称为 Rank-k 匹配精度)表示正确匹配出现在排名前 k 的检索结果中的概率。Rank-k 即表示按照相似度排序后的前 k 张图像中存在与查询图像属于同一 ID 的准确率。而大多数行人重识别工作中更在意的是, 排序结果第一位的检索结果匹配正确的概率, 即 Rank-1。

另一种广泛使用的度量标准, 即平均精度均值 (mean Average Precision, mAP), 是行人重识别等多标签图像分类任务中常用的评测指标。mAP 指将多分类任务中的平均精度(AP)求和再取平均。具体计算步骤如下:

首先需要计算准确率 (Precision), 其公式如下:

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

其中 TP 为预测正确的正样本数，FP 为预测错误的正样本数，TN 为预测正确的负样本数，FN 为预测错误的负样本数。接下来需要计算平均精度（Average Precision, AP），它是指集合中某一类别所有准确率之和除以含有该类别目标的图像数量：

$$AP = \frac{\sum Presion_c}{images_c} \quad (4)$$

最后是平均精度均值（mAP），由于任务中有不止一个类别，因此需要对所有类别计算平均 AP 值：

$$mAP = \frac{\sum_{k=0}^C AP_k}{C} \quad (5)$$

其中，分子为每一个类别的平均精度， C 为总类别数。

3.2 实验设置

本文实验基于PyTorch深度学习框架进行展开，并使用2个NVIDIA RTX 2080 GPU进行数据并行加速。其中，输入图像的尺寸统一调整为384×128，采用Adam优化器进行参数更新，权重衰减因子设置为0.0005，两个超参数 β_1 ， β_2 被分别设置为0.9和0.999。另外，实验共进行400次迭代，初始学习率被设置为0.0002，并分别在第200次迭代和第300次迭代时，学习率均减小90%。

3.3 实验结果分析

为了验证本文提出模型的有效性，分别在 Market-1501，DukeMTMC-ReID 数据集上对其各模块进行实验验证。其中 baseline 是以 Resnet-50 为基础设计的模型，仅存在一个全局分支。本文通过在 baseline 基础上不断增加特定模块以及从完整模型中去除特定模块两个角度来验证各部分有效性。

3.3.1 多尺度注意力模块对于实验结果的影响

多尺度注意力模块（MSA）通过将多尺度特征融合方法（MSFF）与双重注意力模块^[14]相结合，使得网络能够有效地从整个空间访问多尺度特征，实现对于全局特征的加权强化。为了验证 MSA 模块的有效性，本文在 Market-1501 以及 DukeMTMC-ReID 两个数据集上进行了实验。从表 1 可以看出，将 MSA 模块嵌入 baseline 之后，在 Market-1501 数据集中，Rank-1 和 mAP 分别提升了 1.52%和 1.48%。而在 DukeMTMC-ReID 数据集上，Rank-1 和 mAP 也分别提升了 2.20%和 2.45%。、将 MSA 模块从整体模型中去除之后，Rank-1 和 mAP 在 Market-1501 数据集上分别下降了 0.95%、1.37%；对应地，在 DukeMTMC-ReID 数据集上分别下降了 0.83%、1.74%。实验结果证明本文所提出的 MSA 模块可以起到加权强化全局显著信息的作用。

表 1 多尺度注意力模块的影响

Table1 Effects of Multi-scale Attention module

%

Method	Market-1501		DukeMTMC-ReID	
	Rank-1	mAP	Rank-1	mAP
Baseline	88.63	71.55	82.27	64.57
Baseline(w/ MSA)	90.15	73.03	84.47	67.02
Ours(w/o MSA)	94.42	86.65	89.74	79.18
Ours	95.37	88.02	90.57	80.92

3.3.2 多尺度注意力特征丢弃模块对于实验结果的影响

同样地，本文对于多尺度注意力特征丢弃模块（MSA-FD）的有效性进行了验证。从表 2 中可以看出，将特征丢弃方法得到的局部特征与 baseline 得到的全局特征相结合，模型的表现得到了较大提升。同时相较于未添加多尺度注意力模块的批量特征丢弃方法(BDB)^[4]，本文所提出的 MSA-FD 模块通过调整为丢弃区域中的特征权重分布进一步提高模型表现，在 Market-1501 数据集上进行的实验中，Rank-1 和 mAP 分别提升了 1.08%、0.89%。而在 DukeMTMC-ReID 数据集上，Rank-1 和 mAP 上也分别提升了 1.66%、1.41%。另一方面，将 MSA-FD 模块从整体模型中移除之后，模型的精度在两个数据集上也都有一定程度的下降。证明了本文所提方法的有效性，更加精细准确的局部特征对于模型表现可以起到积极的作用。

表 2 多尺度注意力特征丢弃模块的影响

Table2 Effects of Multi-scale Attention-aware Feature DropBlock module

%

Method	Market-1501		DukeMTMC-ReID	
	Rank-1	mAP	Rank-1	mAP
Baseline	88.63	71.55	82.27	64.57
Baseline(w/ BDB)	92.68	84.53	86.94	75.33
Baseline(w/ MSA-FD)	93.76	85.42	88.60	76.74
Ours(w/o MSA-FD)	94.94	86.72	89.59	79.61
Ours	95.37	88.02	90.57	80.92

3.3.3 不同分块策略对于实验结果的影响

为了验证不同的图像分块策略对于模型表现的影响，本文分别将均匀分块（Uniform Partition, UP）策略以及模糊切块（Fuzzy Partition, FP）策略作为单独分支与全局分支相结合进行实验，同时验证了不同重叠尺寸的模糊切块策略对于实验结果的影响。实验结果如表 3 所示，其中 H 代表特征切分后每部分特征的高度。从表中实验结果可以看出，相较于均匀切分策略，模糊切分策略的实验结果有一定的提升。同时，当切分后每部分特征图高度为 10 时，实验结果表现最佳。其中，在 Market-1501 数据集上，模糊切分策略相较于均匀切分策略在 Rank-1 和 mAP 上分别提升了 0.66%和 0.51%。证明模糊分块策略确实可以在一定程度上减少暴力切分带来的信息损失。

表 3 不同分块策略的影响

Table3 Effects of different partition strategies

%

Method	Market-1501	
	Rank-1	mAP
UP branch (H=8)	94.17	86.07
FP branch (H=9)	94.43	86.29
FP branch (H=10)	94.83	86.58
FP branch (H=11)	94.71	86.60
FP branch (H=12)	94.50	86.42

3.3.4 联合多元局部特征策略对于实验结果的影响

本文所设计的网络为了追求更加精细的局部特征,将特征丢弃方法与图像切分策略相结合,即减少了切分策略带来的信息丢失,同时也降低了特征丢弃方法所带来的随机性。这两部分特征与全局特征相结合之后,模型表现出了较好的效果。综上所述,本文所设计的各模块都可以起到提高模型表现的作用,同时各模块之间的组合也可以进一步提升精度,证明本文所提出的方法确实可以更好协调协调全局特征和局部特征之间的关系。

表 4 联合多元局部特征策略对于实验结果的影响

Table 4 Effect of joint multiple local feature strategy

%

Method	Market-1501		DukeMTMC-ReID	
	Rank-1	mAP	Rank-1	mAP
Baseline	88.63	71.55	82.27	64.57
Baseline+MSA-FD branch	93.76	85.42	88.60	76.74
Baseline+FP branch	94.83	86.58	89.50	79.47
All	95.37	88.02	90.57	80.92

3.4 实验结果与主流方法的对比

为了验证本文所提算法的优越性,分别在三个数据集 Market-1501, DukeMTMC-ReID 以及 CUHK03 上与目前多种主流行人重识别算法进行对比。一部分工作将注意力机制引入行人重识别中,如注意力复合网络(AACN)^[9]、基于掩模的对比注意力模型(MGCAM)^[22]、协调注意力网络(HA-CNN)^[11]及多尺度上下文感知网络(MSCAN)^[23]等。还有一部分工作基于局部特征提取进行展开,如局部卷积基准网络(PCB)^[5]、多粒度网络(MGN)^[6]等。对比实验另外涉及一些其他方法如多水平分解网络(MLFN)^[24],行人语义分析网络(SPREID)^[25]等。

从实验结果可以看出,在 Market-1501 数据集上,本文所设计的方法 Rank-1 和 mAP 分别达到了 95.37%、88.02%,表现出了良好的性能。另外在采用 Re-ranking^[26]方法对结果进行优化之后,Rank-1 和 mAP 分别提升至 96.35%以及 94.50%。而在 DukeMTMC-ReID 数据集上,Rank-1 精度为 90.57%,mAP 为 80.92%,使用 Re-ranking 方法后,Rank-1 和 mAP 分别提升至 93.00%、90.74%,超过了当前大多数的方法。同样地,本文也针对 CUHK03 (Labeled) 以及 CUHK03 (Detected)对于算法性能进行了测试,表 7 中结果可以看出,在该数据集上,

本文所提出的方法依旧得到了较好的表现。相较于注意力复合网络(AACN)^[9]等基于注意力机制的工作,本文所提方法通过联合全局和局部注意力方法,同时兼顾了全局显著特征和局部细节特征,可提取更具判别力的行人特征用于识别。另外,对于局部特征地提取,本文联合了基于注意力机制的特征擦出方法与图像分块策略,所得到的局部特征可包含更多细节信息。本文所提出的方法在三个数据集上均取得了较高的精度,超过了当前大多数方法。

表 5 在 Market-1501 数据集上的结果对比

Table 5 Comparison of results on Market-1501

%

Method	Market-1501	
	Rank-1	mAP
MSCAN ^[23]	80.31	57.53
MGCAM ^[22]	83.55	74.25
HA-CNN ^[11]	91.2	75.7
AACN ^[9]	85.90	66.87
SPrEID ^[25]	90.80	76.56
MLFN ^[24]	90.0	74.3
PCB ^[5]	93.8	81.6
MGN ^[6]	95.7	86.9
Ours	95.37	88.02
Ours + Re-ranking	96.35	94.50

表 6 在 DukeMTMC-ReID 数据集上的结果对比

Table 6 Comparison of results on DukeMTMC-ReID

%

Method	DukeMTMC-ReID	
	Rank-1	mAP
JLML ^[27]	73.3	56.4
SVDNet-ResNet50 ^[28]	76.7	56.8
AACN ^[9]	76.84	59.25
SPrEID ^[25]	80.48	63.27
HACNN ^[11]	80.5	63.8
MLFN ^[24]	81.0	62.8
PCB ^[5]	83.3	69.2
MGN ^[6]	88.7	78.4
Ours	90.57	80.92
Ours + Re-ranking	93.00	90.74

表 7 在 CUHK03 数据集上的结果对比

Table 7 Comparison of results on CUHK03

%

Method	CUHK03(Labeled)		CUHK03(Detected)	
	Rank-1	mAP	Rank-1	mAP
DPFL ^[29]	43.0	40.5	40.7	37.0
MGCAM ^[22]	49.29	49.89	46.29	46.74
HA-CNN ^[11]	44.4	41.0	41.7	38.6
MLFN ^[24]	54.1	49.2	52.8	47.8
SVDNet-ResNet50 ^[28]	40.9	37.8	41.5	37.3
PCB ^[5]	-	-	63.7	57.5
MGN ^[6]	68.0	67.4	66.8	66.0
Ours	77.43	75.84	75.57	73.28
Ours+ Re-ranking	85.30	87.15	83.64	85.17

3.5 计算效率与主流方法的对比

本文所提方法与 Mobilenet_v2^[30]、HA-CNN^[11]、MLFN^[24]等主流方法进行了对比。为保证实验对比的公平性,本文在相同硬件环境下测试了各方法训练运行一个批次(10 张图像)所消耗的时间。从表 8 中可以看出,本文所设计的方法与一些轻量级模型如 Mobilenet_v2 等在实时性方面仍有一定的差距,但其精度表现较好,与 MGN^[6]等方法计算效率相仿。同时需要说明的是,本文所设计的注意力模块旨在增加较少参数的情况下,提高网络学习能力,其增加的参数量相较于整体网络来说是较小的,运行速度相对慢些的主要原因也是来自于及基础网络结构的复杂性。

表 8 在 Market-1501 数据集上的计算速度对比

Table 8 Comparison of computation speed on Market-1501

Method	computation speed	Rank-1 (%)	mAP(%)
	(s/batch)		
Mobilenet_v2 ^[30]	0.138	87.0	68.5
HA-CNN ^[11]	0.237	91.2	75.7
MLFN ^[24]	0.585	90.0	74.3
PCB ^[5]	0.331	93.8	81.6
MGN ^[6]	0.561	95.7	86.9
Ours	0.544	95.37	88.02

4 结 论

本文通过设计一个基于多尺度注意力机制的多分支行人重识别网络,分别实现对于显著信息与次显著信息的强化,更好地协调全局特征和多元局部特征来完成行人重识别任务。与

其他方法相比,本文所提出的方法可以实现更加合理的权重分配,无论从全局特征还是局部特征层面,都可以最大限度对特征进行充分利用,减少无关因素对于模型造成的干扰。此外,对于传统均匀分块方法进行微调,设计了一种模糊图像分块策略,减少了暴力分割造成的信息丢失问题。本文在多个数据集上对于模型表现进行了验证,并提供了关于模型详细的分析,实验结果证明本文所提出的方法具有突出的表现。

参考文献

- [1] Liu K W, Fang P P, Xiong H X, et al. Person Re-Identification Based on Multi-layer Feature[J]. *Laser& Optoelectronics Progress*. 2020,57 (8) :081503.
刘可文,房攀攀,熊红霞等.基于多层级特征的行人重识别[J].激光与光电子学进展, 2020,57 (8) :081503.
- [2] Bi X J, Wang H. Person Re-Identification Based on View Information Embedding[J]. *Acta Optica Sinica*, 2019, 39(6): 615007.
毕晓君,汪灏.基于视角信息嵌入的行人重识别[J]. 光学学报, 2019, 39(6): 615007.
- [3] Chen B, Zha F Y, Li Y Q, et al. Person Re-Identification Based on Convolutional Neural Network Discriminative Feature Learning[J]. *Acta Optica Sinica*, 2018, 38(7): 720001.
陈兵,查宇飞,李运强等.基于卷积神经网络判别特征学习的行人重识别[J]. 光学学报, 2018, 38(7): 720001.
- [4] Dai Z, Chen M, Gu X, et al. Batch DropBlock Network for Person Re-identification and Beyond[C]// 2019 IEEE International Conference on Computer Vision (ICCV), Oct.27-Nov.2, Seoul, Korea. Net York: IEEE, 2019: 3691-3701.
- [5] Sun Y F, Zheng L, Yang Y, et al. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)[M]// Ferrari V, Hebert M, Sminchisescu C, et al. *Lecture notes in computer science*. Cham: Springer, 2018, 11208: 501-518.
- [6] Wang G, Yuan Y, Chen X, et al. Learning discriminative features with multiple granularities for person re-identification[C]//2018 ACM Multimedia Conference on Multimedia Conference. ACM, 2018: 274-282.
- [7] Xu L Z, Peng Li. Person Reidentification Based on Multiscale Convolutional Feature Fusion[J]. *Laser & Optoelectronics Progress*, 2019, 56(14): 141504 .
徐龙壮,彭力.基于多尺度卷积特征融合的行人重识别[J]. 激光与光电子学进展, 2019, 56(14): 141504.
- [8] Ghiasi G, Lin T Y, Le Q V. Dropblock: A regularization method for convolutional networks[C]// *Advances in Neural Information Processing Systems*. 2018: 10727-10737.
- [9] Xu J, Zhao R, Zhu F, et al. Attention-aware compositional network for person re-identification[C] // 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 18-23, 2018, Salt Lake City, UT, USA. Net York: IEEE, 2018: 2119-2128.
- [10] Si J, Zhang H, Li C, et al. Dual attention matching network for context-aware feature sequence based person re-identification[C] // 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 18-23, 2018, Salt Lake City, UT, USA. Net York: IEEE, 2018: 5363-5372.
- [11] Li W, Zhu X T, Gong S G. Harmonious attention network for person re-identification[C] // 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 18-23, 2018, Salt Lake City, UT, USA. Net York: IEEE, 2018: 2285-2294.
- [12] Li X, Wang W, Hu X, et al. Selective kernel networks[C]// 2019 IEEE Conference on Computer

Vision and Pattern Recognition (CVPR), June 16-29, L.A., USA. New York: IEEE, 2019: 510-519.

[13] Zhou K, Yang Y, Cavallaro A, et al. Omni-scale feature learning for person re-identification[C]// 2019 IEEE International Conference on Computer Vision (ICCV), Oct.27-Nov.2, Seoul, Korea. New York: IEEE, 2019: 3702-3712.

[14] Chen Y, Kalantidis Y, Li J, et al. A²-Nets: Double Attention Networks[C]//Advances in Neural Information Processing Systems. 2018: 352-361.

[15] Chollet F. Xception: Deep learning with depthwise separable convolutions[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 1251-1258.

[16] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]// 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 7132-7141.

[17] Lin T Y, RoyChowdhury A, Maji S. Bilinear cnn models for fine-grained visual recognition[C]// 2015 IEEE international conference on computer vision (ICCV). 2015: 1449-1457.

[18] Wang H, Gong S, Xiang T. Highly efficient regression for scalable person re-identification[J]. arXiv preprint arXiv:1612.01341. 2016.

[19] Zheng Z D, Zheng L, Yang Y. Unlabeled samples generated by gan improve the person re-identification baseline in vitro[C]// 2017 IEEE International Conference on computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 3774-3782.

[20] Li W, Zhao R, Xiao T, et al. Deepreid: Deep filter pairing neural network for person re-identification[C]//2014 IEEE conference on computer vision and pattern recognition (CVPR). 2014: 152-159.

[21] Felzenszwalb F, Mcallester D, McAllester D, et al. Object detection with discriminatively trained part-based models [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627-1645.

[22] Song C, Huang Y, Ouyang W, et al. Mask-guided contrastive attention model for person re-identification[C] // 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018:1179-1188.

[23] Li D, Chen X, Zhang Z, et al. Learning deep context-aware features over body and latent parts for person re-identification[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 384-393.

[24] Chang X B, Hospedales T M, Xiang T. Multi-level factorisation net for person re-identification[C]// 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018:2109-2118.

[25] Kalayeh M M, Basaran E, Gökmen M, et al. Human semantic parsing for person re-identification[C] // 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 1062-1071.

[26] Zhong Z, Zheng L, Cao D L, et al. Re-ranking person re-identification with k-reciprocal encoding[C]. // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 3652-3661.

[27] Li W, Zhu X, Gong S. Person re-identification by deep joint learning of multi-loss classification[J]. arXiv preprint arXiv:1705.04724. 2017.

[28] Sun Y F, Zheng L, Deng W J, et al. Svdnet for pedestrian retrieval[C]// 2017 IEEE International Conference on computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017:

3820-3828.

[29] Chen Y B, Zhu X T, Gong S G. Person re-identification by deep learning multi-scale representations[C]// 2017 IEEE International Conference on computer Vision Workshops (ICCVW), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 2590-2600.

[30] Sandler M, Howard A, Zhu M. Mobilenetv2: Inverted residuals and linear bottlenecks. [C]// 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 4510-4520.

作者简介:

李聪（1994—），男，硕士研究生，主要从事深度学习，计算机视觉，行人重识别方面的研究。E-mail: cvlicong@163.com

导师简介:

蒋敏（1978—），女，博士，教授，博士生导师，主要从事机器视觉与图像理解，人体行为理解、目标重识别与目标跟踪，人工智能与大数据分析等方面的研究。E-mail: minjiang@jiangnan.edu.cn（通讯联系人）

网络首发:

标题: 基于多尺度注意力机制的多分支行人重识别算法

作者: 李聪, 蒋敏, 孔军

收稿日期: 2020-01-06

录用日期: 2020-03-12

DOI: 10.3788/lop57.201001

引用格式:

李聪, 蒋敏, 孔军. 基于多尺度注意力机制的多分支行人重识别算法[J]. 激光与光电子学进展, 2020, 57(20): 201001.

网络首发文章内容与正式出版的有细微差别, 请以正式出版文件为准!

您感兴趣的其他相关论文:

基于卷积神经网络的棋子定位和识别方法

韩燮 赵融 孙福盛

中北大学大数据学院, 山西 太原 030051

激光与光电子学进展, 2019, 56(8): 081007

基于深度卷积神经网络的道路场景深度估计

袁建中 周武杰 潘婷 顾鹏笠

浙江科技学院信息与电子工程学院, 浙江 杭州 310023

激光与光电子学进展, 2019, 56(8): 081501

基于深度学习的红外与可见光决策级融合跟踪

唐聪 凌永顺 杨华 杨星 同武勤

国防科技大学电子对抗学院, 安徽 合肥 230037

激光与光电子学进展, 2019, 56(7): 071502

基于卷积神经网络与长短期记忆神经网络的多特征融合人体行为识别算法

黄友文 万超伦 冯恒

江西理工大学信息工程学院, 江西 赣州 341000

激光与光电子学进展, 2019, 56(7): 071505

基于深度学习航拍图像检测的梯度聚类算法

解博 朱斌 张宏伟 马旗 张扬

国防科技大学脉冲功率激光技术国家重点实验室, 安徽 合肥 230037

激光与光电子学进展, 2019, 56(6): 061007