

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 针对基于深度学习的传统方法对于次显著细节信息关注不足的问题, 提出一种基于多尺度注意力机制的多分支网络来统筹图像的显著信息及次显著信息。首先, 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, 实现了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 图像处理; 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

doi: 10.3788/LOP57.201001

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, Jiangsu 214122, China*

Abstract The traditional method based on deep learning does not focus on sub-significant information. Therefore, a multi-branch network based on multi-scale attention (MSA) mechanism was proposed to coordinate significant and sub-significant information. Firstly, the proposed algorithm combined the multi-scale feature fusion module (MSFF) with the attention mechanism to get an MSA module. 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. Additionally, a multi-branch network was established to realize the coordination of global features and multiple local features. Using the MSA module, 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 performs well on multiple datasets.

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

OCIS codes 100.4996; 100.5010; 150.0155; 200.4560

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

人图像视角朝向的特点, 对于行人重识别网络进行了优化。文献[3]提出了一种基于卷积神经网络(CNN)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, 提升了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

收稿日期: 2020-01-06; 修回日期: 2020-03-02; 录用日期: 2020-03-12

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

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、姿态变化等问题时会有更好的表现。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,从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),通过对特征图中的邻近矩形区域进行随机丢弃,实现了对细节信息的关注。文献[4]提出批量特征丢弃(BDB)方法,该方法对于批量特征图,随机遮挡同样的矩形区域,强迫网络在保留的区域中学习一些细节的局部特征。然而,这些方法往往难以对保留区域中不同局部特征的有效性进行准确评判,故很难实现更加细致的加权强化。

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力机制来聚焦特定区域的捕获特征。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的序列比较。文献[11]提出了一个协调注意力网络(HA-CNN),通过融合基于像素的软注意力与基于区域的硬注意力,优化了未对齐的行人图像。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(MSA)模块,使网络能够根据输入自适应地调整感受野大小,并实现了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(MSA-FD)模块,以实现局部特征中不同位置的权重调整,进而使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,与MSA-FD模块相互补充以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

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MSA模块,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

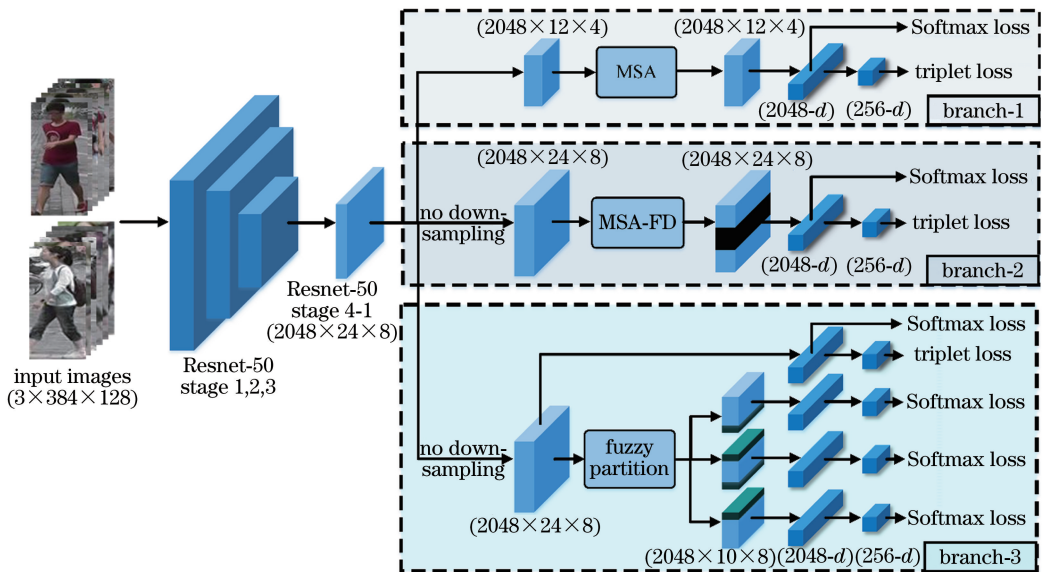


图1 基于多尺度注意力机制的多分支行人重识别网络框架图

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

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,第二、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MSA-FD模块,更好地实现了对保留区域中次显著信息的关注。第三个分支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,采用一种模糊分块策略,使得中下三个切块中的一部分信息相互包含,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来表征输入图像的完整情况。

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的大小相同,忽略了不同尺度特征对于模型的贡献^[12-13]。针对这一问题,本研究将MSFF与双重注意力模块^[14]相结合,充分利用不同尺度信息,最终得到了一个MSA,使得网络能够根据输入自适应地调节感受野大小,以实现全局特征更加细致准确的加权强化。

首先对于MSFF模块,给定当前输入特征 $X_{bl} \in \mathbf{R}^{C \times H \times W}$,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 $X' \in \mathbf{R}^{(C/r) \times H \times W}$ (在实验过程中, r 设置为8)。将特征 X' 分别送入4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(DSC)^[15]得到4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4组不同尺度的特征传入一个参数共享的压缩激励模块中^[16]实现特征通道间信息交互,然后将4个尺度的特征按元素相加,自适应指导多个尺度特征的动态融合,得到一个包含多尺度信息的融合特征 $X_{ms} \in \mathbf{R}^{(C/r) \times H \times W}$ 。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 $X_0 \in \mathbf{R}^{(C/r) \times H \times W}$ 已备后续利用。

将第一个分支中经过Resnet-50网络第4阶段得到的特征 $X_{bl} \in \mathbf{R}^{C \times H \times W}$ 分别送入3个MSFF模块,得到3个多尺度融合特征 $Q \in \mathbf{R}^{(C/r) \times H \times W}, K \in \mathbf{R}^{(C/r) \times H \times W}$ 和 $V \in \mathbf{R}^{(C/r) \times H \times W}$,利用三者之间的相互关联信息建立注意力模块,得到一个更具判别性的全局特征表达。首先,对特征 Q 和 K 进行双线性池化^[17],完成全局信息的聚合。具体地来说,分别对 Q 和 K 进行尺寸调整,得到特征 $Q_r \in \mathbf{R}^{(C/r) \times HW}$ 和 $K_r \in \mathbf{R}^{HW \times (C/r)}$ 。进一步地,对于特征 K_r 通过一个Softmax层得到一个权重特征 $K_s \in \mathbf{R}^{HW \times (C/r)}$,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 $G \in \mathbf{R}^{(C/r) \times (C/r)}$,表达式为

$$G = Q_r \times \text{Softmax}(K_r). \quad (1)$$

最后用特征 G 来聚合整个空间的全局信息。

从输入的整个空间收集特征后,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 $V \in \mathbf{R}^{(C/r) \times H \times W}$,对其进行尺寸调整得到特征 $V_r \in \mathbf{R}^{HW \times (C/r)}$,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

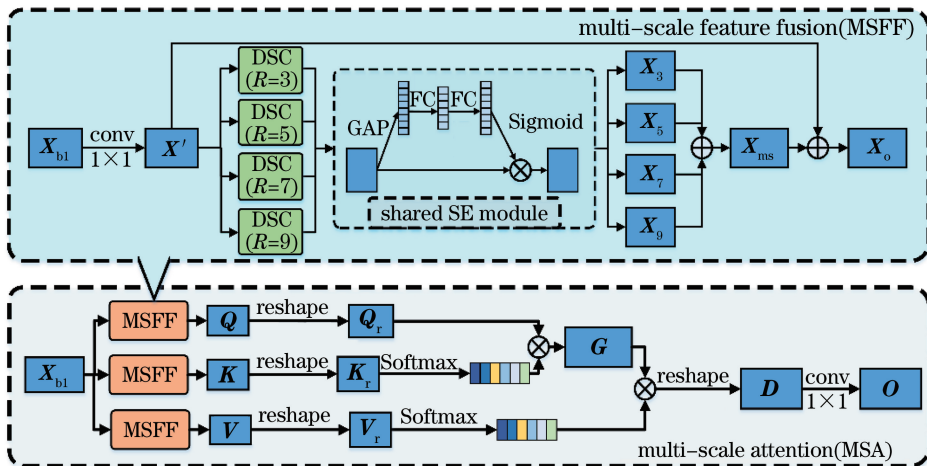


图2 多尺度注意力模块结构图

Fig. 2 Structure of multi-scale attention module

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,有助于学习更复杂的关系。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 $D \in \mathbf{R}^{(C/r) \times H \times W}$,表达式为

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

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

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已被证明是一种有效的策略,批量特征丢弃方法^[4]以批量的方式随机丢弃所有输入特征图的不同区域,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。本研究将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),实现了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 $K_2 \in \mathbf{R}^{(C/r) \times H \times W}$ 和 $V_2 \in \mathbf{R}^{(C/r) \times H \times W}$,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

留区域里面学习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,由于第二、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 两个特征分别起特征聚合和特征分布的作用。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的一个矩形区域进行丢弃。具体来说,首先产生一个二进制掩模 $m \in \mathbf{R}^{1 \times H \times W}$,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 和 V_2 的各通道相加,使得两组特征被丢弃区域中的对应元素均变为 $-\infty$,其他位置则保持不变。这样做的目的是通过Softmax层产生权重向量时,利用 $\exp(-\infty) \approx 0$ 的思想,将保留区域和被丢弃区域区分开来,即被丢弃区域中所有位置被分配0的权重,而保留区域中所有位置共享和为1的概率,该方式不仅可以保证重要特征得到更多关注,还可实现不同局部信息的差别关注。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。

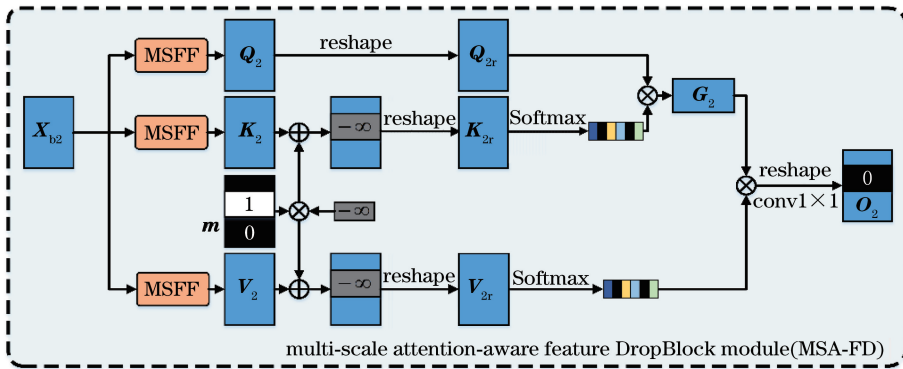


图3 基于多尺度注意力机制的特征丢弃方法结构图

Fig. 3 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,本研究将其作为第三分支与第二分支的MSA-FD模块联合应用,用于提取更加细节的有判别力的行人局部特征。本研究在传统均匀分块的基础上对切块策略进行了微调,采用一种模糊分块的方法,给定网络第三分支传递来的特征 $X_{b3} \in \mathbf{R}^{2048 \times 24 \times 8}$,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,避免全局显著特征对于细节信息的弱化。

2.5 训练策略

本研究的训练策略沿用了MGN^[6]方法中三元组

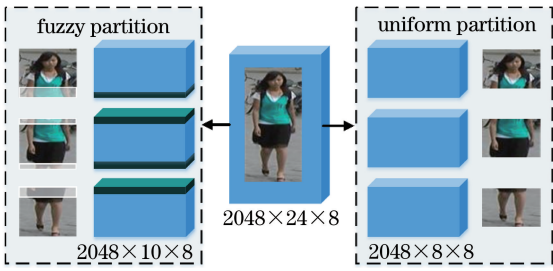


图4 两种分块策略对比图

Fig. 4 Comparison of two partition strategies

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3个分支中所有未进行降维的2048维全局特征。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

本研究在3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个行人的3368张查询图像和16364张图库图像。查询图像中的行人检测矩形框是人工绘制的,而图库图像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的702个行人的16522张图像组成。测试集包含剩余的702个行人的2228张查询图像和17661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和DPM检测的行人边界框两部分。其测试协议分为两种。本实验采用最新的将数据分为包含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

本实验中,为了验证所提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身份信息(ID)的准确率。而大多数行人重识别工作中更在意的是,排名第一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为mAP,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是指将多分类任务中的平均精度(AP)求和再取平均后的结果。具体计算步骤如下:

1) 计算准确率。准确率的计算表达式为

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}}, \quad (3)$$

式中: N_{TP} 为预测正确的正样本数, N_{FP} 为预测错误的正样本数。

2) 计算AP。AP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,表达式为

$$P_A = \frac{\sum P_c}{N_c}, \quad (4)$$

式中: N_c 为第 c 个类别目标的图像数量; $c = 1, 2, \dots, C'$,其中 C' 为总类别数。

3) 计算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,表达式为

$$P_{mAP} = \frac{\sum_{k=0}^{C'} P_{A,k}}{C'}. \quad (5)$$

3.2 实验设置

本实验基于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 pixel×128 pixel,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 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

为了验证本文模型的有效性,分别在Market-1501和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。本研究通过

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的有效性。

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,实现了对全局特征的加权强化。为了验证 MSA 模块的有效性,本研究在 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 数据集上,Rank-1 和 mAP 分别提升了 1.52% 和 1.48%;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%。实验结果证明本研究所提出的 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

Table 1 Effects of multi-scale attention module

Method	unit: %			
	Market-1501		DukeMTMC-ReID	
	Rank-1	mAP	Rank-1	mAP
Baseline	88.63	71.55	82.27	64.57
Baseline(with MSA)	90.15	73.03	84.47	67.02
Ours(without 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,本研究对于多尺度注意力特征丢弃(MSA-FD)模块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由 baseline 得到的全局特征相结合,模型的表现得到了较大的提升。同时相较于未添加多尺度注意力模块的批量特征丢弃(BDB)方法^[4],本研究提出的 MSA-FD 模块通过调整丢弃区域中的特征权重分布进一步提高了模型表现。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%;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不同程度的下降。证明了本文方法的有效性,同时说明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

Table 2 Effects of multi-scale attention-aware feature

Method	DropBlock module unit: %			
	Market-1501		DukeMTMC-ReID	
	Rank-1	mAP	Rank-1	mAP
Baseline	88.63	71.55	82.27	64.57
Baseline(with BDB)	92.68	84.53	86.94	75.33
Baseline(with MSA-FD)	93.76	85.42	88.60	76.74
Ours(without 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,本文分别研究均匀分块(UP)策略以及模糊切块(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

Table 3 Effects of different partition strategies

Partition strategy	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

本文所设计的网络是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,既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。表 4 给出联合多元局部特征策略对于实验结果的具体影响,结果证明本文方法确实可以更好地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	unit: %			
	Market-1501		DukeMTMC-ReID	
	Rank-1	mAP	Rank-1	mAP
Baseline	88.63	71.55	82.27	64.57
Baseline+	93.76	85.42	88.60	76.74
MSA-FD branch				
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,见表 5~7。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。还有一部分工作是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,本文方法的 Rank-1 和 mAP 分别达到了 95.37%、88.02%,表现出了良好的性能。另外采用 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% 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,从表 7 中结果可以看出,在该数据集上,本文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,提取到更具判别力的行人特征用于识别。另外,对于局部特征的提取,本研究联合了基于注意力机制的特征擦除方法与图像分块策略,得到了包含更多细节信息的局部特征。本文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。

3.5 计算效率与主流方法的对比

本文方法与 Mobilenet_v2^[30]、HA-CNN^[11]、MLFN^[24]等主流方法进行了对比。为保证实验对

表5 在 Market-1501 数据集上的结果对比

Table 5 Comparison of results on Market-1501

Partition strategy	Market-1501	
	Rank-1	mAP
MSCAN ^[23]	80.31	57.53
MGCAM ^[22]	83.55	74.25
HA-CNN ^[11]	91.20	75.70
AACN ^[9]	85.90	66.87
SPReID ^[25]	90.80	76.56
MLFN ^[24]	90.00	74.30
PCB ^[5]	93.80	81.60
MGN ^[6]	95.70	86.90
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]	73.30	56.40
SVDNet-ResNet50 ^[28]	76.70	56.80
AACN ^[9]	76.84	59.25
SPReID ^[25]	80.48	63.27
HACNN ^[11]	80.50	63.80
MLFN ^[24]	81.00	62.80
PCB ^[5]	83.30	69.20
MGN ^[6]	88.70	78.40
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		CUHK03	
	(Labeled)		(Detected)	
	Rank-1	mAP	Rank-1	mAP
DPFL ^[29]	43.00	40.50	40.70	37.00
MGCAM ^[22]	49.29	49.89	46.29	46.74
HA-CNN ^[11]	44.40	41.00	41.70	38.60
MLFN ^[24]	54.10	49.20	52.80	47.80
SVDNet-ResNet50 ^[28]	40.90	37.80	41.50	37.30
PCB ^[5]	—	—	63.70	57.50
MGN ^[6]	68.00	67.40	66.80	66.00
Ours	77.43	75.84	75.57	73.28
Ours+Re-ranking	85.30	87.15	83.64	85.17

比的公平性,本研究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]等方法的计算效率相仿。同时需要说明的是,本文所设计的注意力模块可以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仍然是基础网络结构相对复杂。

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		
	time per batch /s	Rank-1 /%	mAP /%
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

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,通过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,都可以最大限度地充分利用特征,减少无关因素对于模型造成的干扰。此外,对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,减少了因暴力分割造成的信息丢失的问题。本研究在多个数据集上对于模型表现进行了验证,并进行了详细的分析,实验结果证明本文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. 基于多层次特征的行人重识别[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): 0615007.
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): 0615007.
- [3] Chen B, Zha Y F, Li Y Q, et al. Person re-identification based on convolutional neural network discriminative feature learning [J]. Acta Optica Sinica, 2018, 38(7): 0720001.
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): 0720001.

- [4] Dai Z Z, Chen M Q, Gu X D, et al. Batch DropBlock network for person re-identification and beyond[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 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]//Computer Vision - ECCV 2018. Cham: Springer International Publishing, 2018: 501-518.
- [6] Wang G S, Yuan Y F, Chen X, et al. Learning discriminative features with multiple granularities for person Re-identification [C]//2018 ACM Multimedia Conference on Multimedia Conference - MM '18. New York: ACM Press, 2018: 274-282.
- [7] Xu L Z, Peng L. 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/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 2119-2128.
- [10] Si J L, Zhang H G, Li C G, et al. Dual attention matching network for context-aware feature sequence based person re-identification [C] //2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 5363-5372.
- [11] Li W, Zhu X T, Gong S G. Harmonious attention network for person re-identification[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 2285-2294.
- [12] Li X, Wang W H, Hu X L, et al. Selective kernel networks [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New

- York: IEEE Press, 2019: 510-519.
- [13] Zhou K Y, Yang Y X, Cavallaro A, et al. Omni-scale feature learning for person re-identification[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27- November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 3701-3711.
- [14] Chen Y P, Kalantidis Y, Li J S, et al. A²-nets: double attention networks [EB/OL]. (2018-10-27) [2019-12-26]. <https://arxiv.org/abs/1810.11579>.
- [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 Press, 2017: 1800-1807.
- [16] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 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), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1449-1457.
- [18] Wang H X, Gong S G, Xiang T. Highly efficient regression for scalable person re-identification [EB/OL]. (2016-12-05) [2019-12-26]. <https://arxiv.org/abs/1612.01341>.
- [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 Press, 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, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 152-159.
- [21] Felzenszwalb P F, Girshick R B, McAllester D, et al. Object detection with discriminatively trained part-based models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627-1645.
- [22] Song C F, Huang Y, Wanli O Y, et al. Mask-guided contrastive attention model for person re-identification[C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 1179-1188.
- [23] Li D W, Chen X T, 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 Press, 2017: 7398-7407.
- [24] Chang X B, Hospedales T M, Xiang T. Multi-level factorisation net for person re-identification [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 2109-2118.
- [25] Kalayeh M M, Basaran E, Gökmen M, et al. Human semantic parsing for person re-identification [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 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 Press, 2017: 3652-3661.
- [27] Li W, Zhu X T, Gong S G. Person re-identification by deep joint learning of multi-loss classification[EB/OL]. (2017-05-12) [2019-12-26]. <https://arxiv.org/abs/1705.04724>.
- [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 Press, 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 Press, 2017: 2590-2600.
- [30] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residuals and linear bottlenecks [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4510-4520.