

## 基于多尺度和注意力融合学习的行人重识别

王粉花<sup>\*①②③</sup> 赵波<sup>①</sup> 黄超<sup>①</sup> 严由齐<sup>①</sup>

<sup>①</sup>(北京科技大学自动化学院 北京 100083)

<sup>②</sup>(北京科技大学人工智能研究院 北京 100083)

<sup>③</sup>(北京市工业波谱成像工程中心 北京 100083)

**摘要:** 行人重识别的关键依赖于行人特征的提取, 卷积神经网络具有强大的特征提取以及表达能力。针对不同尺度下可以观察到不同的特征, 该文提出一种基于多尺度和注意力网络融合的行人重识别方法(MSAN)。该方法通过对网络不同深度的特征进行采样, 将采样的特征融合后对行人进行预测。不同深度的特征图具有不同的表达能力, 使网络可以学习到行人身上更加细粒度的特征。同时将注意力模块嵌入到残差网络中, 使得网络能更加关注于一些关键信息, 增强网络特征学习能力。所提方法在Market1501, DukeMTMC-reID和MSMT17\_V1数据集上首位准确率分别到了95.3%, 89.8%和82.2%。实验表明, 该方法充分利用了网络不同深度的信息和关注的关键信息, 使模型具有很强的判别能力, 而且所提模型的平均准确率优于大多数先进算法。

**关键词:** 行人重识别; 多尺度; 注意力; 残差网络; 度量学习

中图分类号: TN911.73; TP391

文献标识码: A

文章编号: 1009-5896(2020)12-3045-08

DOI: 10.11999/JEIT190998

## Person Re-identification Based on Multi-scale Network Attention Fusion

WANG Fenhua<sup>①②③</sup> ZHAO Bo<sup>①</sup> HUANG Chao<sup>①</sup> YAN Youqi<sup>①</sup>

<sup>①</sup>(School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China)

<sup>②</sup>(Institute of Artificial Intelligence, University of Science and Technology Beijing, Beijing 100083, China)

<sup>③</sup>(Beijing Engineering Research Center of Industrial Spectrum Imaging, Beijing 100083, China)

**Abstract:** The key to person re-identification depends on the extraction of pedestrian characteristics. Convolutional neural networks have powerful feature extraction and expression capabilities. In view of the fact that different features can be observed at different scales, a pedestrian re-identification method based on Multi-Scale Attention Network(MSAN) fusion is proposed. This method samples the features at different depths of the network and fuses the sampled features to predict pedestrians. Feature maps of different depths have different expressive powers, enabling the network to learn more fine-grained features of pedestrians. At the same time, the attention module is embedded in the residual network, so that the network can pay more attention to some key information and enhance the network feature learning ability. The accuracy of the proposed method on the datasets such as Market1501, DukeMTMC-reID and MSMT17\_V1 reaches 95.3%, 89.8% and 82.2%, respectively. Experiments show that the method makes full use of the information of different depths of the network and the key information of interest, so that the model has strong discriminating ability, and the average accuracy of the proposed model is better than most state-of-the-art algorithms.

**Key words:** Person re-identification; Multiple scale; Attention; Residual network; Metric learning

收稿日期: 2019-12-13; 改回日期: 2020-06-17; 网络出版: 2020-07-20

\*通信作者: 王粉花 wangfenhua@ustb.edu.cn

基金项目: 国家重点研发计划重点专项(2017YFB1400101-01), 北京科技大学中央高校基本科研业务费专项 (FRF-BD-19-002A)

Foundation Items: The Key Projects of National Key R & D Plan (2017YFB1400101-01), Beijing University of Science and Technology Central University Basic Research Business Expenses (FRF-BD-19-002A)

## 1 引言

行人重识别(person re-identification)也称行人再识别,是继人脸识别后的又一研究热点,被广泛认为是一个图像检索的子问题。行人重识别是利用计算机视觉技术判断图像或者视频中是否存在特定行人的技术,即给定一个监控行人图像,检索跨设备下该行人的图像。由于不同摄像头下行人的姿态、远近存在差别,很多情况下无法获取可用的人脸图像,这时候行人重识别技术就发挥作用了。行人姿态的多变性、行人视角差异、不同行人的相似性以及外部因素干扰等多种原因也为行人重识别带来了诸多挑战。传统方法主要基于手工设计特征(颜色空间、纹理空间、局部特征<sup>[1]</sup>和专用特征)和距离度量<sup>[2]</sup>(欧式距离、马氏距离<sup>[3]</sup>和余弦距离)来判断跨设备下是否是同一行人。虽然这些方法可以根据行人的很多属性设计相应的手工特征,但人为设计特征具有一定的局限性。

近年来,深度神经网络以及高性能计算硬件迅速发展,尤其是卷积神经网络(Convolutional Neural Network, CNN)的快速发展使得图像处理在各方面有了新的突破。CNN可以自动从原始的图像数据中根据任务需求提取出特征,在行人重识别的应用中取得了显著效果。Sun等人<sup>[4]</sup>提出的PCB方法将行人的局部特征进行水平切块,将网络最终输出的特征图水平分为6块,每一个水平块都进行分类处理。预测时把6个局部特征拼接起来。Luo等人<sup>[5]</sup>为了解决行人姿态不对齐的问题提出了Aligned-ReID方法,首先将特征图水平切块,然后通过计算两张图片局部信息之间的距离,利用最短路径的方式来找到最有效的动态连接。Wang等人<sup>[6]</sup>提出的MGN方法从行人身上的细粒度特征出发,提出了多粒度网络。该网络用全局特征来得到行人粗略的特征,用局部特征来得到行人身上比较

细粒度的特征。陈鸿昶等人<sup>[7]</sup>从行人属性的角度出发,提出一种行人属性分级识别的神经网络模型,根据上一级的识别结果,调整下一级的识别策略,从而提高了小目标属性的识别准确率,进而提高了网络的识别能力。Dai等人<sup>[8]</sup>提出的BFENET方法不让网络太关注于那些显而易见的全局特征,提出了一种批量特征擦除的方法,即对于一个批量的特征图,随机遮挡住同样大小的一块区域,强迫网络去学习剩余区域的一些细节特征。

上述基于深度学习的方法虽然都取得较好的性能,但都是利用网络最后输出的特征进行预测,而没有将网络不同深度的特征利用起来。为此,本文提出一种多尺度和注意力融合的网络结构(Multiple Scale Attention Network, MSAN),利用网络不同深度的特征尺度不同,融合不同尺度特征图的信息,这样可以获得更多中间层的有用信息,从而完成不同的预测任务。同时引入卷积注意力模块(Convolutional Block Attention Module, CBAM)<sup>[9]</sup>,使网络更加关注具有判别力的特征,增强网络对特征的学习能力。对网络的最后一层输出的特征图,采用不同的池化(平均池化和最大池化),然后将池化后的特征进行相关处理,提高网络的学习效率。

## 2 基本原理

### 2.1 算法流程

目标检测中的特征金字塔网络(Feature Pyramid Network, FPN)<sup>[10]</sup>将不同尺度的特征进行融合和不同尺度的特征进行预测两种机制一起使用,即将高层的特征添加到相邻的低层中组合成新的特征,也可以将低层的特征添加到高层,每一层单独进行预测。受FPN网络这种多尺度融合预测的启发,本文基于ResNet50提出了一种多尺度融合预测的网络结构,如图1所示。与ResNet50结构不同的是在Conv4\_x之后没有进行下采样,即Conv4\_x

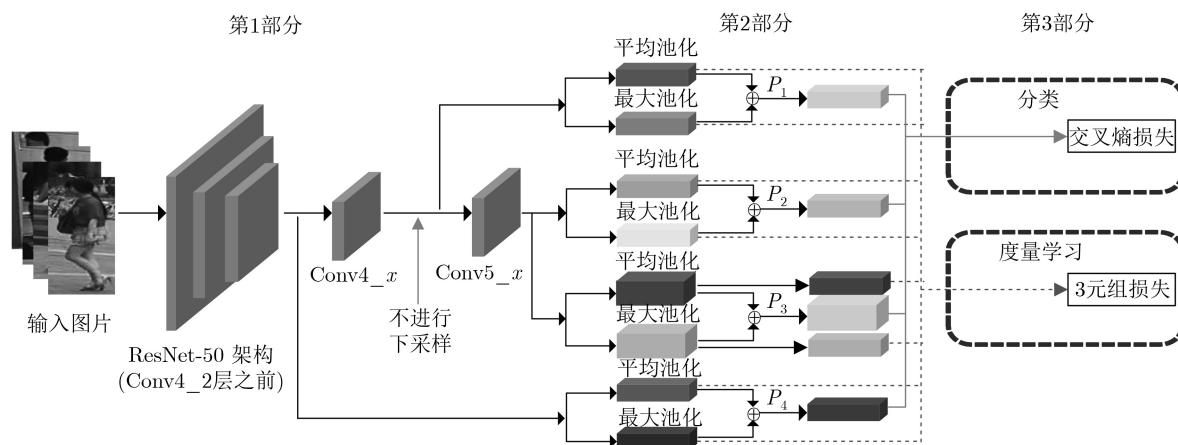


图1 多尺度和注意力融合模型框架图

和Conv5\_x输出的特征图的大小是一样的。如图1所示，整个流程可以分为3个部分，其中第1部分主要负责从模型不同深度提取不同尺度的特征图。第2部分将提取的特征图分别进行平均池化和最大池化，这是为了增加预测所用特征信息，然后将得到的两个池化向量进行叠加处理，最终得到一个 $1 \times 1 \times C$ 的特征向量。为了得到更多特征图信息，对Conv5\_x输出的特征图采用不同的池化方式，分别得到 $1 \times 1 \times C$ 和 $2 \times 2 \times C$ 两种不同的特征向量。第3部分则对处理后的特征向量进行分类和度量学习。

### 2.2 ResNet50网络

经典的卷积神经网络显示，网络有足够的深度是模型表现良好的前提，但是网络达到一定深度以后，简单的网络堆叠反而效果变差了。在许多的数据集上都显示出一个普遍的现象：增加网络深度到一定程度时，更深的网络意味着更高的训练误差。误差升高的原因是网络越深，梯度消失的现象就越明显，所以在反向传播的时候，无法有效地把梯度更新到前面的网络层，靠前的网络层参数无法更新，导致训练和测试效果变差。所以针对这个问题He等人<sup>[11]</sup>提出了一种全新的网络，叫深度残差

网络，它允许网络尽可能地加深，如图2所示为ResNet50网络架构图。

ResNet50分别由Conv1, Conv2\_x, Conv3\_x, Conv4\_x, Conv5\_x和池化层6个模块构成。Conv1模块由一个 $7 \times 7$ 的卷积核提取特征，且卷积核的步长为2，使图像的长宽降低为原图像的1/2，然后再经过一个最大池化层，进一步降低图像的分辨率。Conv2\_x, Conv3\_x, Conv4\_x和Conv5\_x模块都使用重复的残差块提取特征，Conv2\_x的网络结构如图3所示，其基本思想是，对输入的特征图，扩展通道至原来的1倍，而长宽都缩减为原先的1/2，每个阶段都会由1个下采样块和2个残差块构成，其中下采样块通过设置最初的卷积步长为2对特征图进行下采样，使得长宽得到缩减。而在残差块中，通过设置卷积相关参数，可以使残差块的输入和输出的特征图尺寸一致，从而进行相加处理，避免深层网络的梯度消失和退化问题。

### 2.3 多尺度模块

所谓多尺度，就是对信号的不同粒度的采样，一般来说粒度更小、更密集的采样可以看到更多的细节，而粒度更大、更稀疏的采样可以看到整体的趋势。卷积神经网络通过逐层抽象的方式来提取目

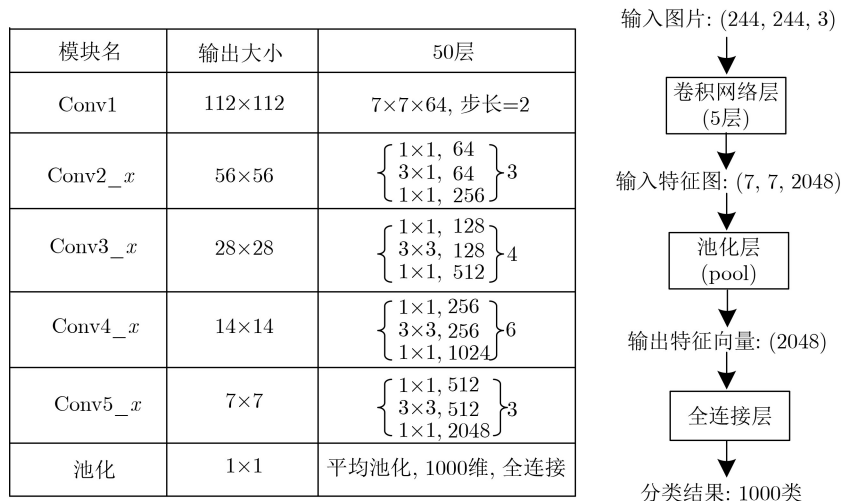


图2 ResNet50网络架构图

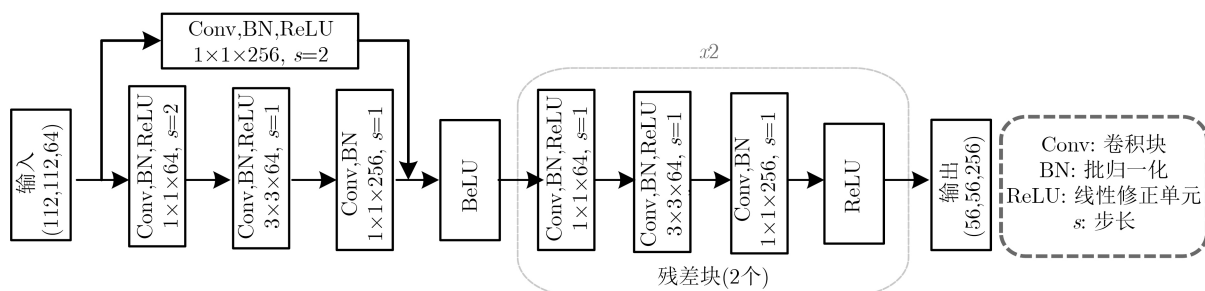


图3 Conv2\_x模块架构图

标的特征，其中最重要的一个概念就是感受野。如果感受野太小，则只能观察到局部的特征，反之感受野太大，则获得更多无效的信息。行人重识别需要根据行人的属性来进行再识别，所以不仅需要感受野小的局部特征来得到行人身上的细粒度特征，还需要感受野大的全局特征得到全局的粗粒度信息。如图4所示，分别从ResNet50网络的Conv4\_1, Conv4\_x和Conv5\_x 3个不同深度对特征进行采样，得到3个特征图，然后利用3个特征图形成不同的网络分支，最后利用这3个分支处理后的信息进行融合预测。

2.4 CBAM模块

CBAM模块包括通道注意力和空间注意力两个子模块，如图5所示。输入特征图首先经过通道注意力，然后通过空间注意力。为了使注意力机制更加高效，通道注意力采用最大池化和平均池化两种方法，得到两个 $1 \times 1 \times C$ 的通道权重的矩阵，这两个权重矩阵会被输入到同一个多层感知机中，如式(1)和式(2)所示

$$M_C(F) = \sigma \{ \text{MLP} [\text{AvgPool}(F)] + \text{MLP} [\text{MaxPool}(F)] \} \quad (1)$$

$$F' = M_C(F) \times F \quad (2)$$

式中， $F$ 表示输入特征， $\text{AvgPool}(F)$ 和 $\text{MaxPool}(F)$ 分别为平均池化和最大池化， $\text{MLP}$ 为多层感知器，

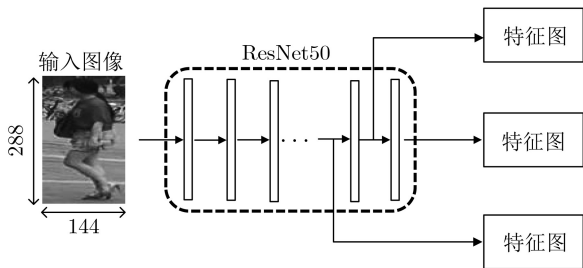


图4 多尺度结构图

$\sigma$ 为Sigmoid激活函数， $M_C(F)$ 为通道注意力模块， $F'$ 表示经过通道注意力机制后得到的特征。

空间注意力模型和通道注意力相似，采用平均池化和最大池化对特征进行处理。不同的是在通道维度上进行压缩采样，生成两个 $W \times H$ 的2维的空间矩阵，然后把它们拼接在一起，输入一个卷积层，进行权重的学习优化，得到空间注意力模块，如式(3)和式(4)所示

$$M_S(F') = \sigma \left( f \left\{ \left[ \text{AvgPool}(F') \right]; \left[ \text{MaxPool}(F') \right] \right\} \right) \quad (3)$$

$$F'' = M_S(F') \times F' \quad (4)$$

式中， $M_S(F')$ 为空间注意力模块， $f$ 为卷积层运算， $\text{AvgPool}(F')$ 和 $\text{MaxPool}(F')$ 分别为平均池化和最大池化， $\sigma$ 为Sigmoid激活函数， $F''$ 表示经过空间注意力机制后得到的特征。

2.5 损失函数

为了增加网络对特征的识别能力，使用交叉熵损失和3元组损失联合训练。在深度Re-ID任务中，这两种损失被广泛使用，交叉熵损失用于分类，3元组损失作为度量学习。对于基础的识别任务，将识别任务当作一个多分类问题。交叉熵损失函数如式(5)所示

$$L_{\text{Softmax}} = - \sum_{i=1}^N \lg \frac{\exp(\mathbf{W}_{y_i}^T \mathbf{f}_i + b_{y_i})}{\sum_{k=1}^C \exp(\mathbf{W}_k^T \mathbf{f}_i + b_k)} \quad (5)$$

式中， $N$ 表示批量大小， $C$ 表示类别数目，即训练集中行人的类别数； $\mathbf{f}_i$ 是第 $i$ 个样本的深度特征，其类别为 $y_i$ ， $\mathbf{W}_k$ 表示 $\mathbf{W}$ 的第 $k$ 行参数， $b$ 为偏置项。

在行人重识别任务中，如图6所示，有时候会遇到距离相对较大的两个相同行人的图片，而两张

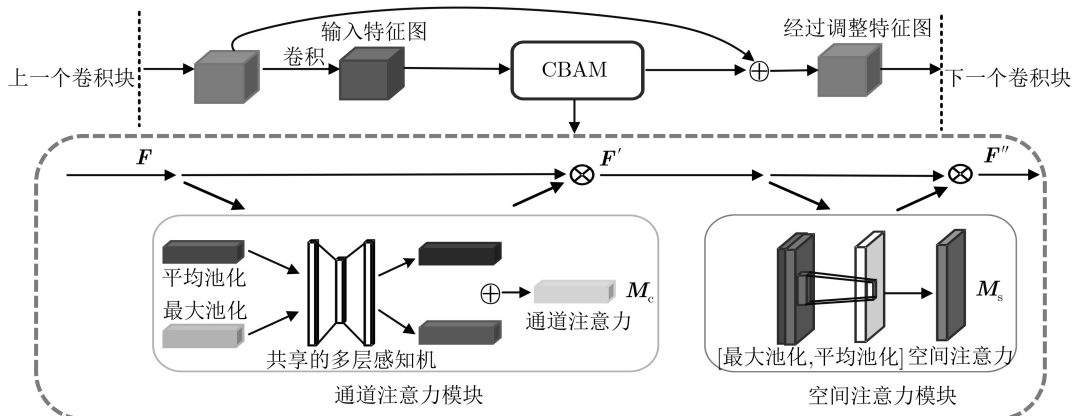


图5 CBAM模块图

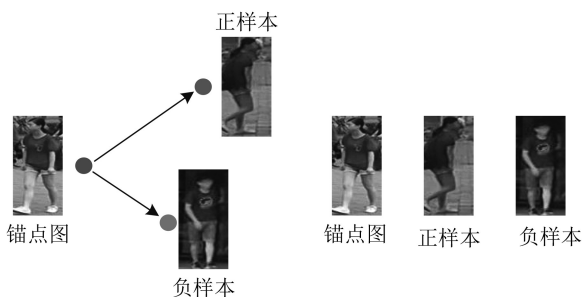


图6 3元组损失

不同行人图片之间的距离却很小,为了挖掘这种难样本,使用3元组损失(triplet loss)<sup>[12]</sup>来优化,选3张图片,分别是锚点图、正样本和负样本,利用这3张图片分别构成正样本对和负样本对,然后拉近相同行人之间的距离,拉远不同行人之间的距离。旨在使减小类内距离,增大类间距离。提高重识别的准确率,3元组损失函数如式(6)所示

$$L_{\text{Triplet}} = - \sum_{i=1}^P \sum_{a=1}^K \left[ \alpha + \max_{p=1 \dots K} \left\| \mathbf{f}_a^{(i)} - \mathbf{f}_p^{(i)} \right\|_2 - \min_{\substack{n=1 \dots K \\ j=1 \dots P \\ j \neq i}} \left\| \mathbf{f}_a^{(i)} - \mathbf{f}_n^{(j)} \right\|_2 \right] \quad (6)$$

式中,  $\mathbf{f}_a^{(i)}$ ,  $\mathbf{f}_p^{(i)}$ ,  $\mathbf{f}_n^{(j)}$ 表示从随机选择的锚点图片、正样本和负样本中提取的特征,这里的正样本是指和随机选择的锚点图片是同一个行人,负样本则相反,  $\alpha$ 表示距离间隔,  $P$ 和  $K$ 表示从批量中选  $P$ 个行人,每个行人抽取  $K$ 张图片。

### 3 实验结果与分析

#### 3.1 实验环境

实验中采用TITAN Xp×2进行运算加速,操作系统为64位Ubuntu16.04,基于python3.6.4的深度学习框架Pytorch1.1.0完成程序编程,基础学习率为0.03,迭代次数为300,批量大小为64。

#### 3.2 实验数据集

为了验证提出模型的有效性,分别在Market-1501<sup>[13]</sup>, DukeMTMC-reID<sup>[14]</sup>和MSMT17\_V1<sup>[15]</sup>3个主流公开数据集上进行有效性实验验证,Market1501数据集包含6个摄像头下1501个行人的32668张图片,其中751个行人的12936张图片作为训练集,另外750个行人的19732张图片作为测试集。DukeMTMC-reID数据集是DukeMTMC数据集的一个子集,用于研究行人重识别,共包含8个摄像头下1812个行人的36411张图片,702个行人的16522张图片,随机选择作为训练集,另外的702个

行人的19889张图片作为测试集。MSMT17\_V1数据集包含4101个行人的126441张图片,训练集包含1041个行人共32621张图片,而测试集包括3060个行人共93820张图片,对于测试集,11659张图片被随机选出来作为查询图片,而其它82161张图片作为图库,这些图片是由15个摄像头在不同场景下拍摄的,涵盖了早上、中午、下午3个时间段,很好地模拟了现实场景下数据的分布。为了能找出效果最佳的模型和调整模型参数,将3个数据集的训练集按9:1的比例分为训练集和验证集。

#### 3.3 评价指标

本文使用目前行人重识别普遍采用的两种评价指标:平均精确均值(mean Average Precision, mAP)和首位准确率(Rank-1)。

平均精确均值综合考虑了查准率和查全率,能够反映查询图片在图像库中所有正确的图片排在检索列表前面的程度,能更加全面地衡量ReID算法的性能。平均精度均值首先需要计算每一个物体类别的平均精度(AP)。平均精度可以用式(7)表示

$$AP = \frac{\sum P_k}{I_C} \quad (7)$$

式中,  $\sum P_k$ 表示验证集中第  $C$  类的所有精确率的和,  $I_C$ 表示含有第  $C$  个类别目标的图像的数量。平均精度均值为每一个类别的平均精度的均值,如式(8)所示

$$mAP = \frac{\sum_{k=0}^C AP_k}{C} \quad (8)$$

式中,  $\sum_{k=0}^C AP_k$ 表示每一个类别的平均精度,  $C$ 代表总类别数。

Rank- $n$ 表示搜索结果中最靠前(置信度最高)的  $n$ 张图有正确结果的概率,首位准确率(Rank-1)表示搜索结果中置信度最高的图即为正确结果的概率。

#### 3.4 实验仿真与结果分析

##### 3.4.1 多尺度模型有效性验证

多尺度与注意力网络融合模型(MSAN)包含CBAM注意力模块以及从ResNet50网络Conv4\_1, Conv4\_x和Conv5\_x 3个不同深度抽取的特征分支,为了验证多尺度模型的有效性,模型只保留CBAM模块和ResNet50网络Conv5\_x层输出的特征图进行实验。表1给出了单尺度和多尺度模型在3个公开数据集Market1501, DukeMTMC-reID和MSMT17\_V1上的实验结果。其中RK<sup>[16]</sup>表示对需要检测的图库中的图片用  $k$ 阶导数编码的方式进行

表1 多尺度融合模型准确率验证实验结果(%)

方法	Market1501		DukeMTMC-reID		MSMT17_V1	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
SSAN	94.9	87.9	86.1	67.7	81.4	66.3
SSAN(+RK)	95.3	93.7	86.0	75.6	84.6	73.8
MSAN	95.3	87.9	89.8	78.8	82.2	60.6
MSAN (+RK)	95.9	93.9	92.3	89.7	85.0	74.6

重排序, 主要流程是利用获取的图像特征计算出马氏距离和杰卡德距离<sup>[17]</sup>, 然后对这两个距离做一个加权求和, 得到最终的距离。单尺度注意力网络(Single Scale Attention Network, SSAN)表示仅使用ResNet50网络Conv5\_x输出的特征图, 而不使用Conv4\_1和Conv4\_x输出的特征图。MSAN表示分别使用网络不同深度的特征图和注意力融合的多尺度注意力网络。如表1所示, 以不加多尺度特征的SSAN网络作为基础网络, 再此基础上添加多尺度特征后得到MSAN网络, 分别在不同的数据集上进行实验, 可以从表1中实验结果发现MSAN网络在Market1501数据集上首位准确率达到95.3%, 较SSAN网络首位准确率提升0.4%; 在DukeMTMC-reID数据集上首位准确率达到89.8%, 较SSAN网络首位准确率提升3.7%; 在MSMT17\_V1数据集上首位准确率达到82.2%, 较SSAN网络首位准确率提升0.8%。实验结果表明, 添加多尺度特征信息可以提高网络的识别能力。

### 3.4.2 CBAM模块有效性实验验证

为了进一步验证CBAM模块的有效性, 同样地, 只保留多尺度模块(Conv4\_1, Conv4\_x和Conv5\_x输出的特征图)而不包含CBAM模块构成多尺度网络(Multiple Scale Network, MSN), 分别在Market1501, DukeMTMC-reID和MSMT17\_V1 3个主流公开数据集上进行实验。在此基础上, 添加CBAM模块的MSN网络, 分别在3个不同的数据集上进行实验。可以从表2中实验结果发现, 相较于基础网络MSN, 添加CBAM模块的MSN网络, MSAN(MSN+CBAM)网络在Market1501数据集上首位准确率达到95.3%, 较MSN网络提升

0.9%, 平均准确率达到87.9%, 较MSN网络提升1.7%; 在DukeMTMC-reID数据集上首位准确率达到89.8%, 较MSN网络提升2.3%, 平均准确率达到78.8%, 较MSN网络提升1.6%; 在MSMT17\_V1数据集上首位准确率达到82.2%, 较MSN网络提升2.6%, 平均准确率达到60.6%, 较MSN网络提升4.6%。实验结果表明, 添加CBAM模块有助于网络对特征学习, 可以增加网络的识别能力。

### 3.4.3 本算法与主流算法的比较

表3给出了在Market1501, DukeMTMC-reID和MSMT17\_V1 3个主流公开数据集上, 算法MSAN与当前最新的几种主流算法性能的比较结果, 其中包括PCB, Aligned-ReID, MGN, BFENET, IANet<sup>[18]</sup>, DGNet<sup>[19]</sup>和OSNet<sup>[20]</sup>等先进的算法。所提算法与这些先进算法都是以ResNet50网络为架构的。

由表3可以看出, 相对于基础网络, 在进行改进后(添加了多尺度和CBAM模块), MSAN网络在Market1501数据集上首位准确率达到95.3%, 平均准确率达到87.9%; 在DukeMTMC-reID数据集上首位准确率达到89.8%, 平均准确率达到78.8%; 在MSMT17\_V1数据集上, 首位准确率达到82.2%, 平均准确率达到60.6%。MSAN算法与Aligned-ReID, MGN, BFENET, IANet等算法相比在3个数据集上识别效果有显著提升, 并且在DukeMTMC-reID和MSMT17\_V1数据集上首位准确率和平均准确率达到最高。实验结果表明, 添加多尺度和注意力的单一改进策略较本文完整算法, 首位准确率和平均准确率有所下降, 将两种方法进行结合, 可以得到更好的实验结果。

表2 CBAM模块准确率验证实验结果(%)

方法	Market1501		DukeMTMC-reID		MSMT17_V1	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
MSN	94.4	86.2	87.5	77.2	79.6	56.0
MSN (+CBAM)	95.3	87.9	89.8	78.8	82.2	60.6
MSN(+RK)	95.3	93.1	90.9	89.2	83.2	72.0
MSN(+CBAM+RK)	95.9	93.9	92.3	89.7	85.0	74.6

表3 所提MSAN算法与其他先进算法的准确率对比(%)

方法	Market1501		DukeMTMC-reID		MSMT17_V1	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
SVDNet <sup>[21]</sup>	82.3	62.1	76.7	56.8	—	—
DPFL <sup>[22]</sup>	88.6	72.6	79.2	60.0	—	—
SVDNet+Era <sup>[23]</sup>	87.1	71.3	79.3	62.4	—	—
TriNET+Era <sup>[23]</sup>	83.9	68.7	73.0	56.6	—	—
DaRe <sup>[24]</sup>	89.0	76.0	80.2	64.5	—	—
GP-reid <sup>[25]</sup>	92.2	81.2	85.2	72.8	—	—
PCB <sup>[4]</sup>	92.3	77.4	81.9	65.3	68.2	40.4
Aligned-ReID <sup>[5]</sup>	92.6	82.3	—	—	—	—
PCB+RPP <sup>[4]</sup>	93.8	81.6	83.3	69.2	—	—
MGN <sup>[6]</sup>	95.7	86.9	88.7	78.4	—	—
BFENET <sup>[8]</sup>	94.2	84.3	86.8	72.1	—	—
IANet <sup>[18]</sup>	94.4	83.1	87.1	73.4	75.5	46.8
DGNet <sup>[19]</sup>	94.8	86.0	86.6	74.8	77.2	52.3
OSNet <sup>[20]</sup>	94.8	84.9	88.6	73.5	78.7	52.9
MSAN	95.3	87.9	89.8	78.8	82.2	60.6
MSAN(+RK)	95.9	93.9	92.3	89.7	85.0	74.6

## 4 结论

本文提出了一种多尺度和注意力融合学习的行人重识别算法。首先, 算法通过采样不同深度的特征, 并利用得到的不同尺度的特征进行融合预测。其次, 在残差网络中加入了注意力模块, 使网络能学习到行人的关键特征。最后, 在网络的最后一层输出的特征图, 分别采用最大池化和平均池化来获取更多的特征信息。损失函数利用分类损失和3元组损失联合约束, 充分利用样本的信息。通过这些方面的改进使得网络可以学习到判别力较强的特征, 提高了行人重识别的精度。并在公开主流数据集上对模型进行了验证, 实验证明, 该模型具有很强的判别力。

## 参考文献

- [1] FARENZENA M, BAZZANI L, PERINA A, *et al.* Person re-identification by symmetry-driven accumulation of local features[C]. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, USA, 2010: 2360–2367.
- [2] 周智恒, 刘楷怡, 黄俊楚, 等. 一种基于等距度量学习策略的行人重识别改进算法[J]. 电子与信息学报, 2019, 41(2): 477–483. doi: 10.11999/JEIT180336.
- [3] ZHOU Zhiheng, LIU Kaiyi, HUANG Junchu, *et al.* Improved metric learning algorithm for person re-identification based on equidistance[J]. *Journal of Electronics & Information Technology*, 2019, 41(2): 477–483. doi: 10.11999/JEIT180336.
- [4] HIRZER M, ROTH P M, KÖSTINGER M, *et al.* Relaxed pairwise learned metric for person re-identification[C]. The 12th European Conference on Computer Vision, Florence, Italy, 2012: 780–793.
- [5] SUN Yifan, ZHENG Liang, YANG Yi, *et al.* Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)[C]. The 15th European Conference on Computer Vision (ECCV), Munich, Germany, 2018: 480–496.
- [6] LUO Hao, JIANG Wei, ZHANG Xuan, *et al.* AlignedReID++: Dynamically matching local information for person re-identification[J]. *Pattern Recognition*, 2019, 94: 53–61. doi: 10.1016/j.patcog.2019.05.028.
- [7] WANG Guanshuo, YUAN Yufeng, CHEN Xiong, *et al.* Learning discriminative features with multiple granularities for person re-identification[C]. 2018 ACM Multimedia Conference on Multimedia Conference, Seoul, Korea, 2018: 274–282.
- [8] 陈鸿昶, 吴彦丞, 李邵梅, 等. 基于行人属性分级识别的行人再识别[J]. 电子与信息学报, 2019, 41(9): 2239–2246. doi: 10.11999/JEIT180740.
- [9] CHEN Hongchang, WU Yancheng, LI Shaomei, *et al.* Person re-identification based on attribute hierarchy recognition[J]. *Journal of Electronics & Information Technology*, 2019, 41(9): 2239–2246. doi: 10.11999/JEIT180740.
- [10] DAI Zuozhuo, CHEN Mingqiang, GU Xiaodong, *et al.*

- Batch DropBlock network for person re-identification and beyond[C]. 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 2019: 3691–3701.
- [9] WOO S, PARK J, LEE J Y, *et al.* Cbam: Convolutional block attention module[C]. The 15th European Conference on Computer Vision (ECCV), Munich, Germany, 2018: 3–19.
- [10] LIN T Y, DOLLÁR P, GIRSHICK R, *et al.* Feature pyramid networks for object detection[C]. 2017 IEEE conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 2117–2125.
- [11] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al.* Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770–778.
- [12] HERMANS A, BEYER L, and LEIBE B. In defense of the triplet loss for person re-identification[EB/OL]. <https://arxiv.org/abs/1703.07737>, 2017.
- [13] ZHENG Liang, SHEN Liyue, TIAN Lu, *et al.* Scalable person re-identification: A benchmark[C]. 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 1116–1124.
- [14] RISTANI E, SOLERA F, ZOU R, *et al.* Performance measures and a data set for multi-target, multi-camera tracking[C]. 2016 European Conference on Computer Vision, Amsterdam, The Netherlands, 2016: 17–35.
- [15] 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 Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 79–88.
- [16] ZHONG Zhun, ZHENG Liang, CAO Donglin, *et al.* Re-ranking person re-identification with k-reciprocal encoding[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 1318–1327.
- [17] SALLEH S S, AZIZ N A A, MOHAMAD D, *et al.* Combining mahalanobis and jaccard distance to overcome similarity measurement constriction on geometrical shapes[J]. *International Journal of Computer Science Issues*, 2012, 9(4): 124–132.
- [18] 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, Long Beach, USA, 2019: 2138–2147.
- [19] HOU Ruibing, MA Bingpeng, CHANG Hong, *et al.* Interaction-and-aggregation network for person re-identification[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 9317–9326.
- [20] ZHOU Kaiyang, YANG Yongxin, CAVALLARO A, *et al.* Omni-Scale feature learning for person re-identification[C]. 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 2019: 3702–3712.
- [21] SUN Yifan, ZHENG Liang, DENG Weijian, *et al.* SVDNet for pedestrian retrieval[C]. 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 3800–3808.
- [22] CHEN Yanbei, ZHU Xiatian, and GONG Shaogang. Person re-identification by deep learning multi-scale representations[C]. 2017 IEEE International Conference on Computer Vision Workshops, Venice, Italy, 2017: 2590–2600.
- [23] ZHONG Zhun, ZHENG Liang, KANG Guoliang, *et al.* Random erasing data augmentation[EB/OL]. <https://arxiv.org/abs/1708.04896>, 2017.
- [24] WANG Yan, WANG Lequn, YOU Yurong, *et al.* Resource aware person re-identification across multiple resolutions[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 8042–8051.
- [25] ALMAZAN J, GAJIC B, MURRAY N, *et al.* Re-ID done right: towards good practices for person re-identification[EB/OL]. <https://arxiv.org/abs/1801.05339>, 2018.
- 王粉花: 女, 1971年生, 副教授, 硕士生导师, 研究方向为模式识别和智能信息处理.
- 赵波: 男, 1994年生, 硕士生, 研究方向为计算机视觉.
- 黄超: 男, 1993年生, 硕士生, 研究方向为计算机视觉.
- 严由齐: 男, 1997年生, 硕士生, 研究方向为计算机视觉.

责任编辑: 余蓉