

# 基于视角信息嵌入的行人重识别

毕晓君, 汪灏\*

哈尔滨工程大学信息与通信工程学院, 黑龙江 哈尔滨 150001

**摘要** 提出一种基于视角信息嵌入的行人重识别模型。结合行人图像视角朝向特点,对 PSE (pose-sensitive embedding)网络结构进行了优化。首先将 PSE 特征向量融合部分由特征的融合改成更符合不同视角特征空间性质的三个视角单元特征向量的拼接;其次视角单元从骨架网络更浅层的 blocks-3 进行分离,增加三个视角单元特征空间的差异性;最后利用改进的深度可分离卷积,设计了一个深度可分离模块,对视角单元进一步进行提取特征,防止模型参数过大的同时提高网络非线性能力,从而提高网络的泛化能力。利用 Market1501、Duke-MTMC-reID 和 MARS 数据集对所提的算法进行有效性验证实验,结果表明所提的改进方法取得了更好的识别效果。

**关键词** 机器视觉; 光计算; 行人重识别; 视角信息嵌入; 深度残差卷积神经网络; 深度可分离卷积

中图分类号 TP391

文献标识码 A

doi: 10.3788/AOS201939.0615007

## Person Re-Identification Based on View Information Embedding

Bi Xiaojun, Wang Hao\*

College of Information and Communication Engineering, Harbin Engineering University,  
Harbin, Heilongjiang 150001, China

**Abstract** In this study, we propose a person re-identification model based on view information embedding. In particular, a pose-sensitive embedding (PSE) network structure is optimized based on the perspective towards characteristics of pedestrian images. First, the fusion part of the PSE feature vector is changed from feature fusion into the concatenation of the feature vectors of three view units, which is considerably reasonable for utilizing different view feature spaces. Second, the view units are separated from the shallow blocks-3 of the skeleton network, which improves the difference of the view feature space. Finally, we design a depthwise separable module based on the improved depth separable convolution to extract features of perspective units, preventing the model parameters from being considerably large and improving the network nonlinearity. The results of the validation experiments conducted using the Market1501, Duke-MTMC-reID and MARS datasets demonstrate that the proposed method can achieve a better recognition accuracy when compared with several advanced algorithms.

**Key words** machine vision; optics in computing; person re-identification; perspective information embedding; deep residual convolution neural network; depthwise separable convolution

**OCIS codes** 200.4560; 100.4996; 100.4999; 100.3008

## 1 引 言

行人重识别是用来匹配来自不相交区域摄像头捕捉到的行人图像技术。随着“天网行动”在我国的大力开展,作为智能安防领域主要技术之一的行人重识别技术成为研究热点。受不同摄像头的安放位置、内部硬件差异以及所处区域环境不同等因素的影响,所拍摄的行人图像存在光照、形变、遮挡、姿态差异和视角差异等问题。目前主流的行人重识别的

方法是基于手工特征表示的方法,寻找一个能够充分描述行人图像的描述模型来表示每个特定行人的特征(如纹理特征<sup>[1]</sup>、边缘特征<sup>[2]</sup>、颜色直方图<sup>[3]</sup>和联合特征<sup>[4]</sup>),从而判断不同摄像头中的行人是否为同一个人。尽管采用手工特征表示方法能够考虑到行人的多方面特征,但对于行人重识别这样一个行人图像受多方面因素影响的情况来说,这些手工特征表示方法的能力仍十分有限。

近年来,深度学习理论<sup>[5]</sup>的提出为信号处理尤

收稿日期: 2019-01-11; 修回日期: 2019-02-13; 录用日期: 2019-03-11

\* E-mail: jdzwanghao@hrbeu.edu.cn

其是图像处理带来了全新的进展。特别地,在图像特征提取方面,深度学习方法取得了远超越基于手工特征表示方法的卓越效果。Li 等<sup>[6]</sup>将深度学习方法引入行人重识别中,通过在网络中添加一个补丁匹配层,将两个图像的卷积响应与不同的水平条纹相乘得到最终的特征向量,进而进行匹配,该方法在一定程度上解决了行人匹配错位的问题。Geng 等<sup>[7]</sup>通过改变深度学习训练策略,将分类任务损失函数与验证任务损失函数一同引入行人重识别中,让两种模式互相学习,提高了行人重识别模型的泛化能力。Zheng 等<sup>[8]</sup>将行人图片局部对齐的思想引入模型设计中,利用人体姿态估计模型估计出人体关键部件区域,再分别对每个行人的关键区域进行一对一匹配,最终得到匹配结果。Zhang 等<sup>[9]</sup>基于同样局部对齐的思想,通过在局部匹配过程中引入动态对齐算法,动态计算局部最短距离,从而实现匹配,该方法在有效地减少模型复杂程度的同时保证了匹配准确度。上述基于深度学习的方法虽然取得了优于手工特征表示方法的效果,但忽视了行人的侧面、正面和背面这些视角变化所带来的深层信息,而这些视角信息能更全面地描述行人,对于行人重识别有着重要的影响。2018 年, Saquib 等<sup>[10]</sup>提出一种 PSE (pose-sensitive embedding) 网络,将行人姿态朝向的信息引入网络的设计中,取得了较好的识别效果。但是 PSE 网络没有做到有效地利用姿态朝向的角度信息,导致模型识别能力受限。为了提升行人重识别模型性能进而提高识别率,本文从行人视角信息出发,在 PSE 网络上进行改进,利用视角朝向信息提出基于视角信息嵌入的行人重识别模型,并取得了更好的行人图像识别效果。

## 2 行人重识别技术

对于一张预查询行人图片,如何在候选库中找到与其匹配的行人图片是行人重识别需要解决的问题。目前主流的基于手工特征表示方法的步骤如下:1)设计一个具有代表性的特征提取器;2)使用该特征提取器分别对预查询行人图片以及候选集中所有的图片进行特征向量提取,并依次利用候选集中的图片特征向量与预查询行人图片的特征向量,计算其向量空间距离;3)将候选集中的图片按特征向量间距离从小到大顺序排列,排在第一位的行人最有可能与预查询行人为同一个人。

上述过程中,关键的步骤在于设计一个具有代表性的特征提取器。目前行人重识别中的特征提取

器主要分为两类,分别是基于手工特征表示的方法和基于深度学习的方法。基于手工特征表示的方法受限于表达能力,在行人重识别的特征提取过程中,很难取得良好的识别效果。而基于深度学习的方法以其优秀的学习机制,能够通过学习得到合适的特征表示,并取得比基于手工特征表达方法更好的效果。目前,在行人重识别中,基于深度学习方法的特征表示有 PSE 网络<sup>[10]</sup>、DPFL (deep pyramidal feature learning)<sup>[11]</sup> 和 QAN (quality aware network)<sup>[12]</sup> 等。本研究算法就是基于 PSE 模型进行改进的。

### 2.1 姿态敏感嵌入网络

最近, Saquib 等<sup>[10]</sup>在计算机视觉顶级会议 CVPR2018 上提出了一种 PSE 网络。该网络具有良好的泛化能力,在行人重识别的多个数据集上取得了最优的结果。PSE 网络的主要思想是将行人姿态与朝向信息引入网络,在骨架网络侧面引入姿态朝向预测模块来辅助骨架网络,最终取得了良好的泛化结果。PSE 网络模型结构如图 1 所示。

PSE 网络模型由骨架、视角预测以及视角单元三个部分组成。整个网络是在 ResNet50 的基础上进行设计,图中 blocks-1、blocks-2、blocks-3 与 blocks-4 对应 ResNet50 中的模块。

PSE 网络的输入是行人图片及其 14 通道的人体姿态热图。layer-0 层是将输入转化为适应 ResNet50 网络输入的形状。骨架部分的各个模块就是 ResNet50 中的对应模块。整个骨架部分用于特征预提取。将视角预测分支从骨架网络的 blocks-1 中的降维步骤后分离出来,用于预测行人图像的朝向,并输出各个朝向对应的概率值。将 ResNet50 中的 blocks-4 的结构复制三份,分别作为代表正面、侧面和背面的视角单元,并分别与视角预测分支的三路 softmax 输出进行加权,三路加权结果从 FC (fully connected layer) 输出并进行特征融合,最终得到一个嵌入向量,用于最终匹配。

### 2.2 PSE 网络模型训练

PSE 网络模型的训练整体分为预训练和训练两个部分。在所有训练之前需将行人重识别数据集分成训练集与测试集,训练集由一定 ID (identity) 数量的行人构成,测试集由 ID 不属于训练集的行人构成。

预训练分为三步。首先对骨架网络部分进行参数初始化,采用在 ImageNet 数据集上预训练的方法进行参数初始化。然后固定骨架网络与视角单元

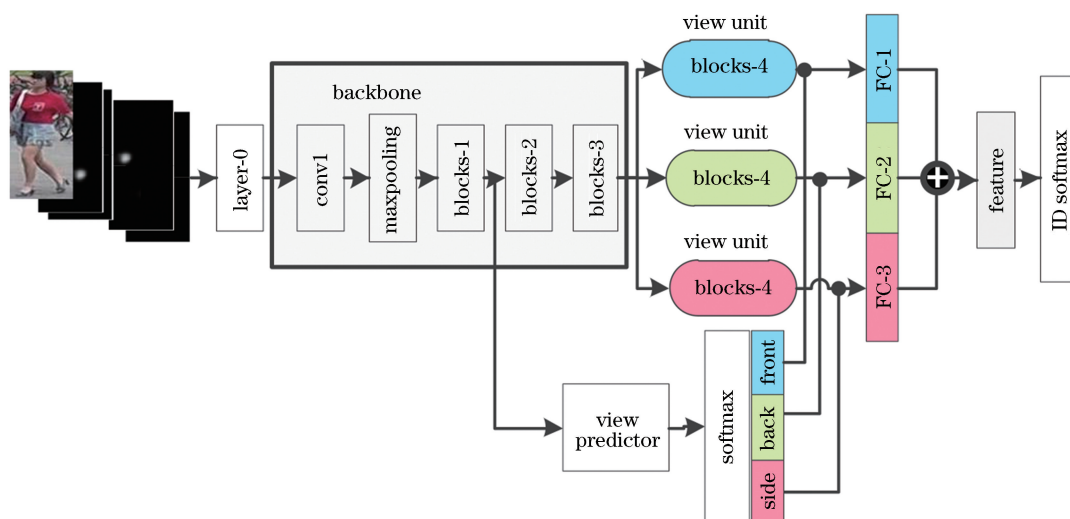


图 1 PSE 网络模型结构

Fig. 1 Structure of PSE network model

部分,使用 RAP(richly annotated pedestrian)数据集对视角预测部分进行预训练。最后固定视角预测部分与骨架部分,使用行人重识别训练集对视角部分进行预训练。

完成预训练后,进行正式训练。首先根据训练集的 ID 数量,使用 softmax 分类器设置成相同 ID 数量的类别,分类器接在 PSE 网络的全连接层之后,用于分类。一般地,使用卷积神经网络框架对行人图片进行特征表示时,需要使用 softmax 分类函数,由于 PSE 网络模型的本质是一种卷积神经网络框架,这里采用同样的操作。接着使用接有 softmax 分类器的 PSE 网络,在分好的训练集上对行人图片进行分类训练,并采用随机梯度下降算法。最后将训练好的模型参数保存在 PSE 网络模型中,取下 softmax 分类器,得到最终的 PSE 网络。

虽然采用 PSE 网络模型取得了目前最优的效果,但 PSE 网络并没有有效地利用视角朝向信息,它对三路视角单元的特征采用的是特征融合的方式,对于三路视角特征的利用不够充分有效,造成模型的识别能力受限。在此通过实验发现,模型结构上还有更合理的设计方案,下面就这个方案进行说明。

### 3 基于视角信息嵌入的行人重识别模型

在此对 PSE 网络进行了三个方面的改进:1)将最终特征向量的组合方式从特征融合改成特征拼接;2)采用骨架网络更浅层的 blocks-3 作为视角单元分离位置;3)利用改进的深度可分离卷积,设计一个

深度可分离模块,代替 blocks-4 对视角单元进一步进行特征提取。改进后的模型变成三个部分,分别是骨架网络部分、视角预测部分以及视角分离部分。

#### 3.1 基于视角信息嵌入的模型结构

所提出的基于视角信息嵌入的模型结构如图 2 所示。

首先,将视角预测部分输出的三个朝向预测概率与三个视角单元的输出进行加权后,可使网络在进行梯度更新时,针对输入行人图片的朝向着重更新对应朝向的视角单元,如当输入行人图片是正面朝向时,视角预测部分正面概率较大,该大概率会让梯度流向被大权值加权的那一路,即正面朝向的视角单元,因此不同视角单元都能特定学习对应行人朝向的特征,这使得三路分开的视角单元具备分别以正面、侧面和背面方式描述特征的能力。实际上,这种特征表达方式的不同反映在特征层面就是表示特征的特征空间不同。Cao 等<sup>[13]</sup>在人脸识别任务中提出一个假设,认为人脸的侧面与正面分别对应不同的特征空间,并通过实验进行了验证。与人脸类似,不同朝向的行人图像对应的特征空间也是不同的。而在 PSE 网络模型的视角部分,网络对三个复制的分别代表正面、侧面和背面的特征图谱直接进行了加权融合,而三个特征图谱所在的特征空间不同,原 PSE 模型不太适合对来自不同特征空间的特征图谱进行线性加权融合。因此,将特征融合改为特征拼接,即将最终的三路 768 维特征向量首尾连接成 2304 维向量,并将其作为最终的特征向量,这就避免了在不同特征空间进行线性加权融合。

其次,为了更好地利用不同空间的信息,与PSE网络在 blocks-4 分出三份视角单元不同的是,所提出的方法在骨架网络更浅层的 blocks-3 层进行分离,其中 blocks-3\_f 表示 ResNet50 中的 conv4\_1, blocks-3\_b 表示 ResNet50 中的 conv4 的剩余部分,这样的操作可以减少视角单元共享的学习参数,增大分离的三个姿态单元的差异性,从而能够更好地利用不同空间的信息。

最后,在PSE骨架的更浅层分离视角单元,把Blocks-3 原有的一路增加到三路,这大大增加了模

型参数量。为了防止网络过大而产生过拟合,并保证网络的非线性,以充分利用特征信息,在此设计了一种深度可分离模块。深度可分离模块采用深度可分离卷积<sup>[14]</sup>。深度可分离卷积由 depthwise 操作与 pointwise 操作构成,其结构如图 3 所示。depthwise 操作是使用尺寸为  $3 \times 3$ 、数量为输入通道数的卷积核在对应输入通道上进行卷积,pointwise 操作是采用尺寸为  $1 \times 1$ 、数量为输出通道数的卷积核进行一般卷积操作。其中,  $m$  与  $n$  为标量,表示通道数。

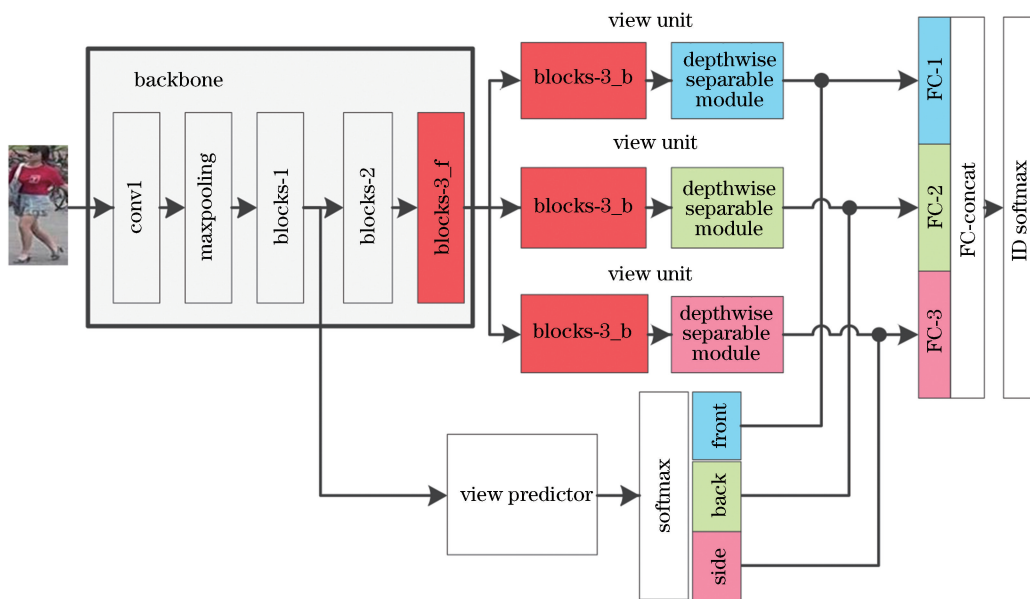


图 2 基于视角信息嵌入模型结构

Fig. 2 Model based perspective information embedding

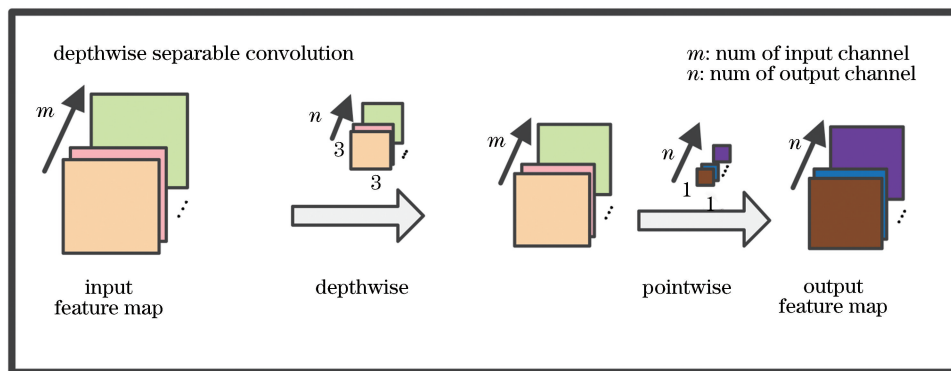


图 3 深度可分离卷积结构

Fig. 3 Depthwise separable convolution

所采用的深度可分离卷积是在原有结构的基础上进行改进。深度网络在提取特征时,往往会提取大量特征,而有些特征并不是识别所需要的特征。一个具备特征选择机制的网络,不仅能够降低网络的冗余,还能更加精准地进行匹配。受文献[15]的

启发,本研究在 depthwise 操作与 pointwise 操作之间引入挤压和激励块。挤压和激励块能对输入的各个通道特征进行加权,通过网络学习,对有利特征进行高加权,对不利特征进行低加权,从而达到特征选择的目的。改进的深度可分离卷积结构如图 4 所

示。深度可分离模块结构如图 5 所示。

深度可分离模块采用两个连续深度可分离卷积堆叠而成,其中每个深度可分离卷积后分别接一个 BN(batch normalization)层与 ReLU 单元。与原始卷积相比,这种结构增加了网络非线性。将堆叠中间层的输出特征图谱与模块输出一同作为输出。文献[16]证明,深度神经网络一般取最后一层

输出作为特征匹配向量,忽视了处在网络稍浅层的中层特征。而在进行匹配时,不仅深层特征重要,图像的中层特征也很重要,中层特征在识别过程中同样能够提升识别性能。

改进的深度可分离卷积既保证充分利用了每个通道的特征图谱信息,又减少了模型的参数,防止了模型过拟合情况的发生。

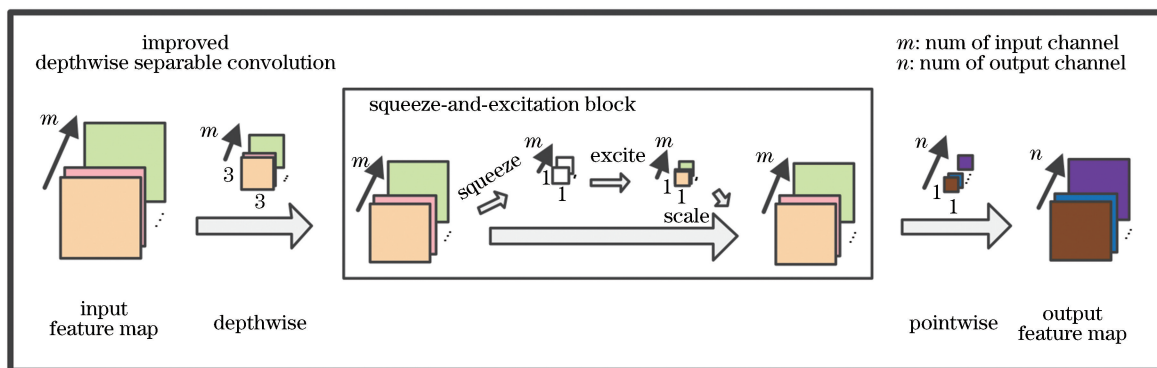


图 4 改进的深度可分离卷积结构

Fig. 4 Improved depthwise separable convolution

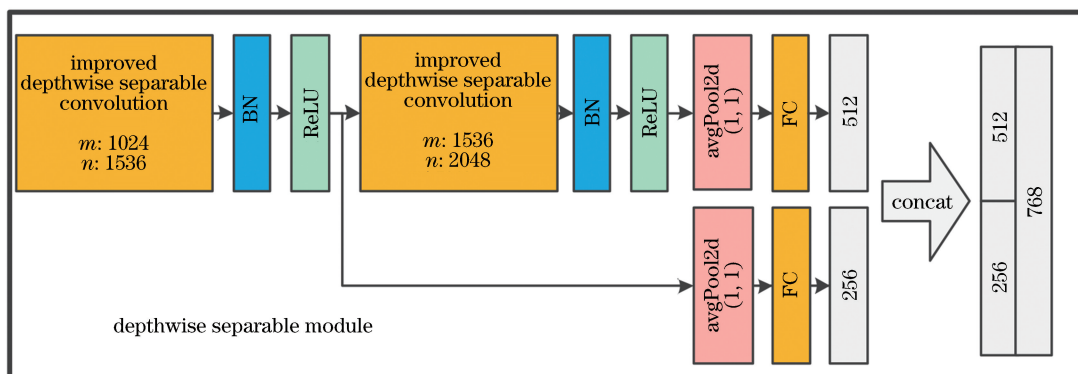


图 5 深度可分离模块结构

Fig. 5 Structure of depthwise separable module

### 3.2 算法实现步骤

为了便于理解,在此给出所提算法的具体实现步骤。

1) 将所提出的模型分为三个部分,分别为 ResNet50 骨架网络、行人预测分支以及视角分离。模型损失函数采用交叉熵函数。

2) 采用 RAP(richly annotated pedestrian)数据集对行人预测分支进行预训练。其中,使用在 ImageNet 上的预训练参数作为骨架网络的参数进行初始化。训练过程中,骨架网络部分的参数固定不训练。

3) 用行人重识别数据集预训练视角分离部分。这里对整个网络进行训练,但是需先固定行人预测分支与骨架网络的参数。进行反向传播时,梯度只

更新视角分离部分,其目的是为视角分离部分的参数提供一个较为理想的初始化状态。

4) 分别设定各个部分的学习率,其中骨架网络、视角预测部分和视角分离部分的识别率分别为 0.01,0.001,0.01,用行人重识别数据集对整个网络进行训练。另外,所提出的算法不需要引入 PSE 网络中的精细粒度的关节信息,仅使用粗粒度的行人朝向信息就能取得较原 PSE 网络更好的效果。

## 4 仿真实验与结果分析

### 4.1 实验环境和参数设置

实验中采用的硬件环境配置为 Intel © Core™ i7 8700K 酷睿六核 3.7 GHz、32 G 内存、64 位处理器,并采用 GeForce GTX 1080 Ti GAMING X

TRIO 11GB GDDR5X 进行运算加速,操作系统为 64 位 Ubuntu 16.04,采用基于 Python3.6 的深度学习框架 Pytorch0.4.0 完成程序编程。基础学习率为 0.01,视角单元部分的学习率为 0.001,迭代代数 130,批量大小为 32。

#### 4.2 数据集描述

为了验证所提出的改进模型在行人重识别任务中的优势,在此选用目前行人重识别主流的公开数据集 Market1501<sup>[17]</sup>、Duke-MTMC-reID<sup>[18]</sup>和 MARS<sup>[19]</sup>进行实验,其中数据集的划分标准与 PSE 网络模型中所采用的标准<sup>[10]</sup>一致。

Market1501 数据集采集自清华大学校园,由来自 6 台摄像机视频上的行人检测器生成的 32668 个边界框组成,包括 1501 个不同 ID 行人。751 人用于训练,750 人用于测试。训练集包含 12936 个图像,候选集设置 19732 个图像,查询集包含 3368 个图像。

Duke-MTMC-reID 数据集是根据 8 台摄像机的数据创建的。包含 1404 个 ID 行人、16522 个训练图像、2228 个查询图像和 17661 个图库图像。由于采用 8 台摄像机捕获了数量较多的图像,Duke-MTMC-reID 数据集成为迄今为止最具挑战性的行人重识别数据集之一。

MARS 数据集是根据 6 台摄像机的数据创建的,由 20478 个视频片段构成,图片总数达到 1191993,包含 1261 个重复出现的行人,625 人用于训练,636 人用于测试。其中训练集由 8298 个视频片段构成,测试集由 12180 个视频片段构成,将图片训练集、测试集分别分割为 509914、681089 张图像。

表 1 视角预测模块有效性验证实验结果

Table 1 Results of perspective predictor module verification experiment

Method	Market1501		Duke-MTMC-reID		MARS		
	rank-1 / %	mAP / %	rank-1 / %	mAP / %	rank-1 / %	mAP / %	
Proposed	Except perspective	83.6	62.6	74.1	53.7	67.7	50.1
	All	<b>89.9</b>	<b>71.6</b>	<b>79.9</b>	<b>61.7</b>	<b>74.1</b>	<b>57.6</b>

表 1 中,Except perspective 表示去掉视角预测模块的模型,All 表示完整的改进模型。由表 1 可知,在添加了视角预测模块后,模型的性能有了大幅度提升,这验证了视角预测模块的有效性。

为了验证改进深度可分离卷积在所提模型中的有效性,在此进行了改进深度可分离卷积有效性验证实验。实验结果如表 2 所示。

测试集中,查询集由 1980 个视频片段构成,候选集由 9330 个视频片段构成。在此,训练集中所有视频片段的图片均参与训练,对测试集中的查询集与候选集的每个行人视频片段均采用均匀抽样方法进行测试,抽样数目为 15。对该 15 张图片的特征向量进行加权,得到该行人的特征向量。

为防止模型出现过拟合现象,通过对行人图片进行随机水平翻转与随机裁剪实现数据集增强。其中,随机裁剪方法是把输入图片先调整到 288 pixel  $\times$  144 pixel 大小,再使用大小为 256 pixel  $\times$  128 pixel 的框对调整后的图片进行随机裁剪。

#### 4.3 仿真实验与结果分析

实验从模型的有效性性与先进性两个方面进行验证,并与目前最好的几种深度学习方法和主流机器学习方法进行对比验证。两个实验将测试集分成查询集与候选集,本实验采用单查询方式验证结果,单查询方式的查询集由每个 ID 行人在每个摄像头下的一张图片构成,剩下的图片作为候选集。实验评价指标采用行人重识别通用评价指标——累积匹配特征(CMC)的排名 1(rank-1)识别率和平均精度(mAP)。

##### 4.3.1 算法有效性验证实验

为了验证所提模型的有效性,实验选用 Market1501、Duke-MTMC-reID 和 MARS 数据集。

首先,所提出的模型将 PSE 网络中的特征融合改为特征拼接,此时视角预测模块的用途发生改变。因此,为了验证视角预测模块对于改进后的 PSE 网络仍然有效,在此进行了视角预测模块有效性验证实验。实验结果如表 1 所示。

表 2 中,Except SE-Block 表示去除改进深度可分离卷积中的 SE-Block。由表 2 可知,在深度可分离卷积中添加了 SE-Block 之后,模型的性能有了一定的提升,这验证了改进深度可分离卷积的有效性。

为了验证在深度可分离模块中使用中层特征方法的有效性,进行了中层特征方法有效性验证实验,实验结果如表 3 所示。

表 2 改进深度可分离卷积的有效性验证实验结果

Table 2 Results of improved depthwise separable convolution verification experiment

Method	Market1501		Duke-MTMC-reID		MARS		
	rank-1 / %	mAP / %	rank-1 / %	mAP / %	rank-1 / %	mAP / %	
Proposed	Except SE-Block	87.0	67.5	77.8	59.6	71.2	54.2
	All	<b>89.9</b>	<b>71.6</b>	<b>79.9</b>	<b>61.7</b>	<b>74.1</b>	<b>57.6</b>

表 3 中层特征方法的有效性验证实验结果

Table 3 Verification experiment results of mid-level feature method

Method	Market1501		Duke-MTMC-reID		MARS		
	rank-1 / %	mAP / %	rank-1 / %	mAP / %	rank-1 / %	mAP / %	
Proposed	Except Mid-level-feature	87.4	70.9	79.5	57.8	72.1	55.9
	All	<b>89.9</b>	<b>71.6</b>	<b>79.9</b>	<b>61.7</b>	<b>74.1</b>	<b>57.6</b>

表 3 中, Except Mid-level-feature 表示深度可分离模块中不采用中层特征。由表 3 可知, 在添加了 Mid-level-feature 之后, 模型的性能有了一定的提升, 这验证了在深度可分离模块中采用中层特征

方法的有效性。

最后, 通过与 PSE 网络进行比较来验证所提模型的有效性。改进模型有效性验证实验结果如表 4 所示。

表 4 改进模型有效性验证实验结果

Table 4 Results of improved model verification experiment

S/N	Improved method			Market1501		Duke-MTMC-reID		MARS	
	IM1	IM2	IM3	rank-1 / %	mAP / %	rank-1 / %	mAP / %	rank-1 / %	mAP / %
PSE	①	—	—	87.7	69.0	79.8	62.0	72.1	56.9
	②	—	Y	86.8	66.5	77.5	60.5	70.5	54.9
	③	—	Y	82.5	64.2	70.1	54.3	68.4	49.3
	④	—	Y	85.3	65.8	72.5	57.2	70.2	53.8
Ours	⑤	Y	—	87.5	69.0	79.4	61.1	70.9	57.1
	⑥	Y	—	85.9	66.5	75.6	59.9	67.8	54.6
	⑦	Y	Y	86.6	65.6	77.4	59.4	70.1	55.1
	⑧	Y	Y	<b>89.9</b>	<b>71.6</b>	<b>79.9</b>	61.7	<b>74.1</b>	<b>57.6</b>

在表 4 中, Improved method(IM)被分成了三列。第一列的 IM1 表示是否采用拼接特征进行改进。第二列的 IM2 表示视角单元是否在更浅层的 blocks\_3 进行分离。第三列的 IM3 表示是否采用深度可分离模块进行改进。在 IM 这三列中, “Y”表示采用该改进方式, “—”表示不采用该改进方式。①~⑧表示模型序号, 其中①表示 PSE 网络, ②~⑧表示所提的改进模型。

由表 4 可知, 模型⑧在三个行人重识别数据集上都取得了优于 PSE 网络的效果, 验证了所提模型的有效性。下面对实验结果进行分析。

对比⑥和⑧模型可知, 采用浅层分离改进的模型在三个数据集上的识别率比不采用时的识别率

高。说明当模型同时采用拼接特征改进与深度可分离模块改进时, 浅层分离改进方法有效。

对比③和④模型可知, 采用深度可分离模块改进的模型在三个数据集上的识别率比不采用时的识别率高。说明当模型采用浅层可分离改进时, 深度可分离模块的改进方法有效。同时对⑦和⑧模型可知, 采用深度可分离模块改进的模型在三个数据集上的识别率比不采用时的高。说明当模型同时采用拼接特征与浅层可分离进行改进时, 深度可分离模块改进方法有效。

对比④和⑧模型可知, 采用拼接特征改进的模型在三个数据集上的识别率比不采用时的识别率高。说明当模型同时采用浅层可分离改进与深度可

分离模块进行改进时,拼接特征改进的模型有效。

综上所述,当模型同时采用拼接特征、浅层分离以及深度可分离模块改进方法时,所提算法对模型的改进有效,并能够提升模型性能。这验证了所提算法的有效性。

#### 4.3.2 算法运行速度分析实验

对所提算法与 PSE 网络的运行速度进行对比。任取一张查询集中的行人图片,对比所提网络与 PSE 网络完成查询的时间。查询集选用 Market1501 的查询集,候选集选用 Market1501 的候选集。算法运行速度对比实验结果如表 5 所示。

表 5 算法运行速度对比实验结果

Table 5 Results of algorithm running speed comparison experiment

Method	Time /s	
	Total match	Per match (19720)
PSE	141.57	0.0072
Proposed	288.96	0.0147

表 5 中, Total match 表示查询图片与候选集所有图片进行匹配的时间, Per match 表示查询图片

与一张候选图片进行匹配的时间。由表 5 可知,所提算法的运行速度比 PSE 网络的慢,但是相差不大。由于 Market1501 的候选集图片高达 19720 张,一次查询需要匹配 19720 次。而实际场景中,通常行人是逐渐出现在监视区域,候选图片是逐渐增加的,实际场景中更加看重单对匹配的时间(即模型只需计算新增加的匹配而无需重复计算原有候选库中的匹配),对于所提网络,单对匹配时间仅有 0.0147 s,虽然较 PSE 网络略慢,也完全满足现实需要。

#### 4.3.3 算法先进性验证实验

对目前在三个数据集上最主流的算法 DPFL、QMA 和 PSE 进行对比,结果如表 6 所示。

在表 6 中,“—”表示对应方法未在数据集上实验。由表 6 可知,在三个数据集上所提算法的识别效果较原 PSE 网络均有较好的提升。尤其在 Market1501 数据集上,所提方法的 rank1 超过了目前最好的 DPFL 算法,同时在 MARS 数据集上的 rank1 与 mAP 最高。

表 6 算法结果对比

Table 6 Comparison of algorithm results

Method	Market1501		Duke-MTMC-reID		MARS	
	rank-1 /%	mAP /%	rank-1 /%	mAP /%	rank-1 /%	mAP /%
P2S(point to set)	70.7	44.3	—	—	—	—
Spindle	76.9	—	—	—	—	—
Consistent aware	80.9	55.6	—	—	—	—
GAN(generative adversarial networks)	78.1	56.2	67.7	47.1	—	—
Latent parts	80.3	57.5	—	—	71.8	56.1
ResNet+OIM(online instance matching)	82.1	—	68.1	—	—	—
ACRN(attribute-complementary re-ID net)	83.6	62.6	72.6	52.0	—	—
SVD(singular value decomposition)	82.3	62.1	76.7	56.8	—	—
Part aligned	81.0	63.4	—	—	—	—
PDC(pose-driven deep convolutional model)	84.1	63.4	—	—	—	—
JLML(jointly learning multi-loss)	85.1	65.5	—	—	—	—
DPFL	88.6	72.6	79.2	60.6	—	—
Forest	—	—	—	—	70.6	50.7
DGM(dynamic graph matching)+IDE	—	—	—	—	65.2	46.8
QMA	—	—	—	—	73.7	51.7
ResNet baseline	82.6	59.8	71.5	50.3	64.5	49.5
PSE	87.7	69.0	79.8	62.0	72.1	56.9
Proposed algorithm	<b>89.9</b>	71.6	<b>79.9</b>	61.7	<b>74.1</b>	<b>57.6</b>



## 5 结 论

针对目前最好的 PSE 算法进行行人重识别时视角朝向信息利用方面的不足,进行了模型结构的改进。首先改进了特征向量的融合方式,采用特征拼接的方式使得不同视角特征空间的差异能够得到更好的利用。改进了视角单元分离位置,通过在更浅层分离视角单元,增大视角单元间的差异性,更好地构建了不同视角的特征空间。通过引入深度可分离模块,降低了模型的复杂度,防止了模型的过拟合现象。通过这两方面的改进可以得到判别性更强的网络特征,进一步提高了重识别精度。利用 Market1501 和 Duke-MTMC-reid 以及 MARS 数据集这三种主流的数据库对所提算法进行了验证,实验结果表明该算法具有很好的判别性。

实验结果表明,与 PSE 模型相比,所提的改进算法取得了更好的识别效果,同时也提高了模型的泛化能力。对于行人重识别这种存在大量朝向影响的问题,所提算法具有非常广阔的应用前景。

## 参 考 文 献

- [1] Guo P Y, Su A, Zhang H L, *et al.* Online mixture of random Naive Bayes tracker combined texture with shape feature[J]. *Acta Optica Sinica*, 2015, 35(3): 0315002.  
郭鹏宇, 苏昂, 张红良, 等. 结合纹理和形状特征的在线混合随机朴素贝叶斯视觉跟踪器[J]. *光学学报*, 2015, 35(3): 0315002.
- [2] Sun X W, Xu Q S, Cai Y, *et al.* Sea sky line detection based on edge phase encoding in complicated background[J]. *Acta Optica Sinica*, 2017, 37(11): 1110002.  
孙熊伟, 徐青山, 蔡熠, 等. 基于边缘相位编码的复杂背景下海天线检测[J]. *光学学报*, 2017, 37(11): 1110002.
- [3] Gray D, Tao H. Viewpoint invariant pedestrian recognition with an ensemble of localized features [M]//Forsyth D, Torr P, Zisserman A. *Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer, 2008, 5302: 262-275.
- [4] Prosser B J, Zheng W S, Gong S, *et al.* Person re-identification by support vector ranking [C]// *Proceedings of the British Machine Vision Conference*. August 31-September 3, 2010, Aberystwyth, UK. Durham, England, UK: BMVA Press, 2010: 21.
- [5] Chen Y, Fan R S, Wang J X, *et al.* Cloud detection of ZY-3 satellite remote sensing images based on deep learning [J]. *Acta Optica Sinica*, 2018, 38(1): 0128005.  
陈洋, 范荣双, 王竞雪, 等. 基于深度学习的资源三号卫星遥感影像云检测方法[J]. *光学学报*, 2018, 38(1): 0128005.
- [6] 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, 2014: 152-159.
- [7] Geng M Y, Wang Y M, Xiang T, *et al.* Deep transfer learning for person reidentification[EB/OL]. (2016-12-22) [2018-12-20]. <https://arxiv.org/abs/1611.05244>.
- [8] Zheng L, Huang Y J, Lu H C, *et al.* Pose invariant embedding for deep person re-identification [EB/OL]. (2017-01-26) [2018-02-20]. <https://arxiv.org/abs/1701.07732>.
- [9] Zhang X, Luo H, Fan X, *et al.* AlignedReID: surpassing human-level performance in person re-identification[EB/OL]. (2018-01-31) [2018-02-20]. <https://arxiv.org/abs/1711.08184>.
- [10] Saquib Sarfraz M, Schumann A, Eberle A, *et al.* A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 420-429.
- [11] 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.
- [12] Liu Y, Yan J J, Ouyang W L. Quality aware network for set to set recognition [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 4694-4703.
- [13] Cao K D, Rong Y, Li C, *et al.* Pose-robust face recognition via deep residual equivariant mapping [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 5187-5196.
- [14] Howard A G, Zhu M L, Chen B, *et al.* MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17) [2018-12-21]. <https://arxiv.org/abs/1704.04861>.
- [15] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [EB/OL]. (2017-10-25) [2018-12-21].

- <https://arxiv.org/abs/1709.01507>.
- [16] Yu Q, Chang X, Song Y Z, *et al.* The devil is in the middle: exploiting mid-level representations for cross-domain instance matching [EB/OL]. (2018-04-04) [2018-12-21]. <https://arxiv.org/abs/1711.08106>.
- [17] Zheng L, Shen L Y, Tian L, *et al.* Scalable person re-identification: a benchmark [C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015: 1116-1124.
- [18] Ristani E, Solera F, Zou R, *et al.* Performance measures and a data set for multi-target, multi-camera tracking [M]//Hua G, Jégou H. Lecture Notes in Computer Science. Cham: Springer, 2016, 9914: 17-35.
- [19] Zheng L, Bie Z, Sun Y F, *et al.* MARS: a video benchmark for large-scale person re-identification [M]//Leibe B, Matas J, Sebe N, *et al.* Computer Vision-ECCV 2016. Cham: Springer, 2016, 9910: 868-884.