

# 融合视点机制与姿态估计的行人再识别方法



裴嘉震 徐曾春 胡平

南京工业大学计算机科学与技术学院 南京 211816

(pjz34567@163.com)

**摘要** 行人再识别是视频监控中一项极具挑战性的任务。图像中的遮挡、光照、姿态、视角等因素,会对行人再识别的准确率造成极大影响。为了提高行人再识别的准确率,提出一种融合视点机制与姿态估计的行人再识别方法。首先,采用姿态估计算法 Openpose 定位行人关节点;然后,对行人图像进行视图判别以获得视点信息,并根据视点信息与行人关节点位置进行局部区域推荐,生成行人局部图像;接着,将全局图像与局部图像同时输入 CNN 提取特征;最后,采用特征融合网络将全局与局部的特征融合,以获取更具鲁棒性的特征表示。实验结果表明:提出的方法具有更高的行人再识别准确率,其在 CHUK03 数据集上的 rank1 达到了 71.3%,在 Market1501 和 DukeMTMC-reID 数据集上的 mAP 分别达到了 63.2%与 60.5%。因此,所提方法能够很好地应对行人姿态变化和视角变化等问题。

**关键词** 相机视点;姿态估计;深度学习;行人再识别;特征融合

**中图分类号** TP391

## Person Re-identification Fusing Viewpoint Mechanism and Pose Estimation

PEI Jia-zhen, XU Zeng-chun and HU Ping

College of Computer Science and Technology, Nanjing Tech University, Nanjing 211816, China

**Abstract** Person re-identification is a very challenging task in video surveillance. Person have significant changes in appearance due to occlusion and differences in illumination, posture and perspective, which will ultimately have a great impact on the accuracy of person re-identification. To overcome these difficulties, this paper proposes a method for person re-identification based on viewpoint mechanism and pose estimation. First, the pose estimation algorithm Openpose is used to locate the joint points of person. Then, view discrimination is performed on the image to obtain viewpoint information. Local regions based on viewpoint information and joint point locations is proposed to generate a partial image. Next, the global image and the partial image are input into the CNN simultaneously to extract features. Finally, in order to obtain a more robust feature representation, the feature fusion network is used to fuse the global and local features. Experimental results show that the proposed method has higher person re-identification accuracy. On CHUK03 dataset, the value of rank1 reaches 71.3%, and on Market1501 dataset and DukeMTMC-reID dataset, the mAP reaches 63.2% and 60.5%, respectively. Therefore, the proposed method can well cope with person attitude changes, pose changes and other issues.

**Keywords** Camera viewpoint, Pose estimation, Deep learning, Person re-identification, Feature fusion

## 1 引言

行人再识别(Person Re-identification, Re-ID)起源于多摄像头跟踪,用于判断非重叠视域中拍摄到的不同图像中的行人是否属于同一个人。该技术被广泛应用于各种领域,如视频监控、刑事安防等。近年来,行人再识别技术虽然取得了较多的研究成果,但是依然面临着诸多挑战:1)摄像机拍摄的行人图像的视角多变;2)图像中行人的姿势变化较大。

随着深度学习技术的不断发展和深入,各种基于深度学习的行人再识别方法不断推出。相比于手工设计的传统行人

再识别方法,基于深度学习的行人再识别方法可以更有效地应对行人的复杂变化,取得更优越的性能。同时,随着大量包含注释的行人再识别数据集如 Marke-1501<sup>[1]</sup>, CHUK-03<sup>[2]</sup>, DukeMTMC-reID<sup>[3]</sup>等的推出,基于深度学习的行人再识别方法的准确率不断提升。

克服行人姿势变化与视角多变的常见方法可以分为两种:1)将行人图像均匀分割,提取局部特征;2)提取图像中具有局部不变性的特征作为行人的特征表示。直接将行人图像均匀分割会引入过多的背景噪声,而提取局部不变的特征依赖于人工,稳定性较差,工作量较大。因此,需要一种既能准

到稿日期:2019-05-05 返修日期:2019-07-30 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61672279);国家重点研发计划(2017YFC0805605);江苏省重点研发计划(BE2017617)

This work was supported by the National Natural Science Foundation of China (61672279), National Key R&D Program of China (2017YFC0805605) and Key R&D Program of Jiangsu Province (BE2017617).

通信作者:徐曾春(xzc@njtech.edu.cn)

确推荐局部区域又能克服视角多变的行人再识别方法。

本文提出一种融合视点机制与姿态估计的行人再识别方法。首先,基于姿态估计算法确定行人关节点位置,将 15 个关节点划分为 6 个子集;然后,对行人图像进行视图判别,获取图像视点信息,并结合视点信息与关节点位置生成行人局部身体图像;接着,分别从全局与局部图像中学习特征表示,并通过特征融合网络将全局特征图与局部特征图融合,得到融合特征向量;最后,对融合特征向量进行欧几里得距离判别。

## 2 相关工作

传统的行人再识别方法可大致概括为距离度量学习与特征表示两类:1)设计具有高辨识能力的距离度量算法来优化行人图像之间的特征距离,衡量行人图像之间的相似性;2)提取鲁棒性的特征表示,以解决行人的外观变化等问题。

文献[4]提出一种基于子空间的度量学习方法——交叉视图二次判别分析(Cross-view Quadratic Discriminant Analysis, XQDA),用于对空间进行降维与距离度量。文献[5]提出一种加权近似秩分量分析(Weighted Approximate Rank Component Analysis, WARCA)的度量学习算法,该算法将加权近似排序合成损失函数与满足线性正交投影的正则化项组合,以提升再识别的准确率。文献[6]提出一种具有 PSD 约束和非对称样本加权策略的逻辑度量学习算法。

近年来,深度学习通常被用于学习行人的特征表示或距离度量。由于性能卓越,基于深度学习的方法在行人再识别方法中占据主导地位。文献[2,7]以 Siamese 网络为主体网络,以两幅行人图像作为网络输入,通过增加训练样本来提高行人再识别的准确率。文献[8]采用 LSTM 对身体部位的序列性质进行建模,融合视频中的时间特性与深度特征,以提高行人再识别的准确率。文献[9]对不同图像对的判别匹配部分进行改进。文献[10]首次提出三重损失(Triplet Loss),通过选择 3 幅图像构成三元组(Anchor, Negative, Positive)来学习 Triplet loss,该方法可以有效克服行人再识别中类内差异大而类间相似的问题,从而提高了行人再识别的准确率。文献[11]提出一种联合学习框架,将单一图像特征(Single-Image Representation)与交叉图像特征(Cross-Image Representation)相融合,提升了行人再识别的准确率。

与上述方法不同,本文与文献[12-13]的工作类似,使用人体姿态估计算法<sup>[14]</sup>获取行人的局部特征。

## 3 融合视点与姿态估计的行人再识别

本节描述本文方法的总体框架,并主要对特征提取网络和特征融合网络进行介绍。

### 3.1 算法的网络结构

通常,行人姿势的变化以及相机拍摄角度不同会对行人再识别的准确率造成较大的影响。因此,本文算法将行人姿势信息与相机视点信息包含在行人再识别方法中,以提高行人再识别的准确率。本文方法中,两种信息可以单独使用,也可以同时使用并融合到 CNN 结构中。图 1 给出了本文提出的行人再识别网络的整体结构。每轮训练中同时提取全局图像特征与局部图像特征。例如,对于一个样本,首先通过特征提取网络生成局部行人图像;然后将全局图像与局部图像输

入 CNN,两幅图像在所有网络层中皆拥有独立的权重;最后使用特征融合网络将局部特征与全局特征融合。

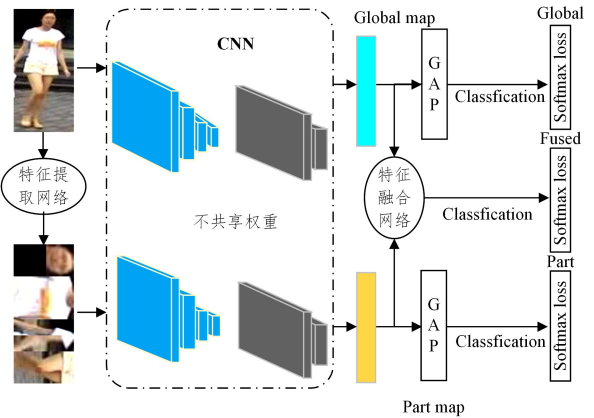


图 1 行人再识别网络的整体框架

Fig. 1 Overall framework of person re-identification

### 3.2 特征提取网络

特征提取网络包含 4 个任务:1)确定行人关节点的位置;2)判别行人图像视点方向;3)根据关节点位置与视点信息进行局部身体区域的推荐;4)将推荐区域重新组合,形成修改后的图像。特征提取网络的结构如图 2 所示。

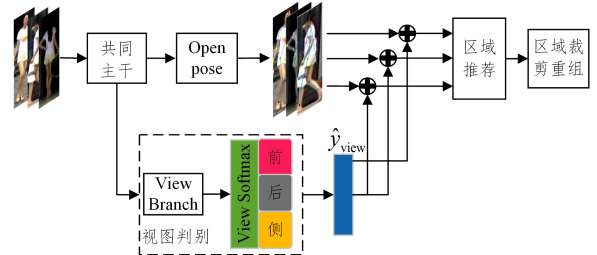


图 2 特征提取网络的结构

Fig. 2 Framework of feature extraction network

#### 3.2.1 相机视点信息

本文使用图像中行人朝向摄像机的方向{前,后,侧}作为相机视点信息。

对于一幅输入图像  $I$ ,设其视图标签为  $Y_v \in \{\text{front, back, side}\}$ 。视图判别器分支从 ResNet-50 结构的第 3 个阶段(即 22 层卷积)后分离,此时的特征图尺寸为  $28 \times 28 \times 256$ 。然后,应用 3 个步长分别为 2,2,5 的连续卷积进一步减小尺寸,得到大小为  $1 \times 1 \times 1024$  的特征向量,再采用 softmax 预测视图,获取每个视图的置信度  $\hat{y}_{\text{view}}$ 。计算公式如下:

$$\mathbf{X}_K = f(I; \theta_K), \mathbf{X}_K \in \mathbb{R}^{1 \times 1 \times 1024} \quad (1)$$

$$\hat{y}_{\text{view}} = f_{\text{view}}(\mathbf{X}_K; \theta_{\text{view}}), \hat{y}_{\text{view}} \in \mathbb{R}^3 \quad (2)$$

其中,  $\mathbf{X}_K$  为图像  $I$  在 25 层卷积之后得到的大小为  $1 \times 1$  的 1024 维特征向量;  $\hat{y}_{\text{view}} = [\hat{y}_{\text{view}}^1, \dots, \hat{y}_{\text{view}}^3]^T$ 。

由于行人再识别公开数据集缺乏视图注释,因此本文在 RAP<sup>[15]</sup> 行人属性数据集上预训练了一个视图判别器。

卷积神经网络的整体时间复杂度遵循下式:

$$\text{TIME} \sim O\left(\sum_{l=1}^D M_l^2 \cdot K_l^2 \cdot C_{l-1} \cdot C_l\right) \quad (3)$$

其中,  $D$  为网络深度(即卷积层数),  $C_l$  为第  $l$  个卷积层的输出通道数,  $M_l$  为第  $l$  个卷积核输出特征图的尺寸,  $K_l$  为第  $l$  个

卷积核的尺寸。分析得出,卷积神经网络的时间复杂度与卷积层数密切相关。

本文方法会增加一定的卷积层,为了降低模型复杂度,在视图判别器与姿态估计网络之间共享低级特征(即早期层)。

视图判别器输出的3个视图置信度被用于辅助推荐行人局部区域,从而避免姿态估计误差导致的部分特征丢失现象。

### 3.2.2 行人关节点检测

目前,大多数行人再识别研究主要通过神经网络提取整幅图像的全局特征,或对图像进行均匀分割以提取局部特征。本文提出的行人再识别方法采用姿态卷积网络获取行人的姿势特征。

由于行人再识别数据集中缺少人体姿势信息的标注,为了获得更好的姿态估计性能,本文采用MPII<sup>[5]</sup>人体姿势数据集上训练的Openpose<sup>[16]</sup>姿态估计模型。

Openpose<sup>[16]</sup>是基于部分亲和域 Part Affinity Fields (PAFs)的自下而上的人体姿态估计算法。其基本原理为:为每个人体关节点预测置信图,每个关节点处都有一个高斯峰值,在得到所有的关节点检测结果后,使用PAFs判断两两关节点是否可以相连,最后进行关节点聚类并连线。

训练过程中,每个关节点的置信度为:

$$S_i^*(m) = \exp\left(-\frac{\|m - x_i\|_2^2}{\sigma^2}\right) \quad (4)$$

其中,  $x_i$  是图像中行人的第  $i$  个身体关节点的真实位置,  $m$  点处的置信度为  $m$  点到真实关节点的高斯距离,  $\sigma$  控制波峰的扩散。

通过3个阶段的置信网络输出关节点置信图  $P_i \in \mathbb{R}^{(W \times H)}$  ( $i \in (0, \dots, 14)$ ), 其表示某像素在关节点处的高斯响应, 离关节点越近的像素, 其响应值越大。因此, 可以通过非极大值抑制 (Non-Maximum Suppression, NMS) 方法回归出15个关节点  $K_i = [x_i, y_i]$  ( $i = 0, 2, \dots, 14$ ) 的位置。

$$K_i(x, y) = \max_{X \in [1, W], Y \in [1, H]} P_i(X, Y) \quad (5)$$

其中,  $P(X, Y)$  是置信图  $(X, Y)$  处的像素值。

行人的15个关节点的示例如图3所示, 其中, 0表示头, 1表示颈部, 2表示右肩部, 3表示右肘部, 4表示右手腕部, 5表示左肩部, 6表示左肘部, 7表示右手腕部, 8表示右髋部, 9表示右膝, 1表示0右脚踝, 11表示左髋部, 12表示左膝盖, 13表示左脚踝, 14表示中心点。

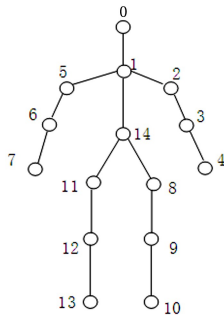


图3 人体关节点示意图

Fig. 3 Human joint point diagram

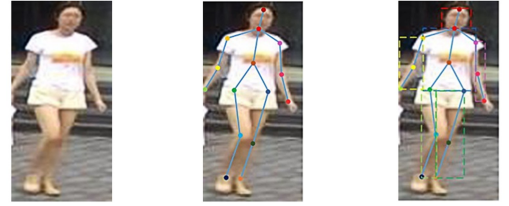
根据具体位置, 将这些关节点划分为6个子集: {头部  $B_1 = \{0, 1\}$ , 躯干  $B_2 = \{1, 2, 5, 8, 11, 14\}$ , 右臂  $B_3 = \{2, 3, 4\}$ , 左臂  $B_4 = \{5, 6, 7\}$ , 右腿  $B_5 = \{8, 9, 10\}$ , 左腿  $B_6 = \{11, 12, 13\}$ }。

若相机视点为{前,后}, 则根据上述6个关节点子集生成6个子区域的边界框。边界框的4个坐标可以根据每个关节点集中身体关节的位置坐标计算得到, 如式(6)所示:

$$Rect_i = \begin{cases} [X-30, X+30, Y_{\min}, Y_{\max}], & \text{if } i = B_1 \\ [\min_{j \in B} (X_j), \max_{j \in B} (X_j), \min_{j \in B} (Y_j), \max_{j \in B} (Y_j)], & (6) \\ \text{else} \end{cases}$$

其中,  $X, Y$  代表每个关节点的横坐标与纵坐标。

相机视点为“前”时, 行人图像如图4所示。

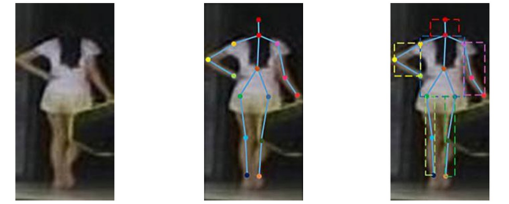


(a) 行人原始图像 (b) 行人关节点位置 (c) 关节点区域推荐

图4 “前”视点的人图像

Fig. 4 Images of person with front view

当相机视点为“后”时, 同样将行人划分为6个子区域, 如图5所示。



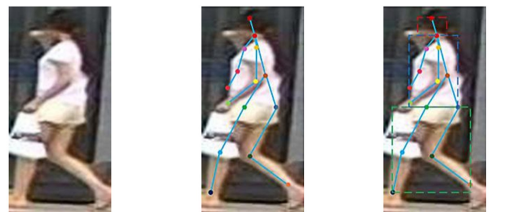
(a) 行人原始图像 (b) 行人关节点位置 (c) 关节点区域推荐

图5 “后”视点的行人图像

Fig. 5 Images of person with back view

相机视点为“侧”时, 将上述6个子集划分为3个宏观子区域{头部  $A_1 = (B_1)$ , 上身  $A_2 = (B_2, B_3, B_4)$ , 下身  $A_3 = (B_5, B_6)$ }, 并为3个子区域生成边框。相机侧面拍照时会出现手臂或腿部被遮挡等情况, 如图6(a)所示, 因此区分相机视点可以有效避免姿态估计错误时划分过多子区域造成的背景噪声干扰问题。边界框的4个坐标可以根据每个关节点集的所有身体关节位置坐标计算得到, 如式(7)所示:

$$Rect_i = \begin{cases} [X-30, X+30, Y_{\min}, Y_{\max}], & \text{if } i = A_1 \\ [\min_{j \in B} (X_j), \max_{j \in B} (X_j), \min_{j \in B} (Y_j), \max_{j \in B} (Y_j)], & (7) \\ \text{else} \end{cases}$$



(a) 行人原始图像 (b) 行人关节点位置 (c) 关节点区域推荐

图6 “侧”视点行人图像

Fig. 6 Images of person with sideview

最终, 特征提取网络将裁剪与旋转的身体局部图像重新组合为行人局部特征图像。重新组合后的图像如图7所示。



图7 局部特征的重组图像

Fig. 7 Images of local feature recombination

### 3.3 特征融合网络

本文引入了视点信息与姿势信息,而常规的行人再识别数据集无法用作视图判别器与姿态卷积网络的训练集。若单独采用局部特征作为行人再识别的依据,会导致准确率较低;而简单地将局部特征与全局特征融合容易引入噪声。因此,本文提出一个双线性池化层,将再识别网络从全局特征表示中获取的1024维特征 $g$ 与从局部特征表示中获取的1024维特征 $p$ 先进行双线性变换;然后进行空间全局池化,通过双线性池化操作聚合两种类型的特征图,生成融合特征 $fusion$ ;最后对融合特征进行归一化,得到2048维融合向量 $\widehat{fusion}$ ,特征融合的公式如下:

$$fusion_{(x,y)} = vec(g_{(x,y)} \otimes p_{(x,y)}) \quad (8)$$

$$fusion = pooling_{(x,y)} \{ fusion_{(x,y)} \} = \frac{1}{S} \sum_{x,y} fusion_{(x,y)} \quad (9)$$

$$\widehat{fusion} = \frac{fusion}{\| fusion \|_2} \quad (10)$$

$$\widehat{fusion}_{(x,y)} = vec(\tilde{g}_{(x,y)} \otimes \tilde{p}_{(x,y)}) \quad (11)$$

$$\tilde{g}_{(x,y)} = \frac{g_{(x,y)}}{\sqrt{\| fusion \|_2}} \quad (12)$$

$$\tilde{p}_{(x,y)} = \frac{p_{(x,y)}}{\sqrt{\| fusion \|_2}} \quad (13)$$

联立式(9)一式(11)得:

$$\widehat{fusion} = \frac{1}{S} \sum_{x,y} \widehat{fusion}_{(x,y)} \quad (14)$$

其中:

$$vec(g \otimes p) = [p_1 g]^T [p_2 g]^T \cdots [p_{1024} g]^T \quad (15)$$

上述公式中, $g_{(x,y)}$ 为全局特征图 $G$ 中 $(x,y)$ 处的特征向量, $p_{(x,y)}$ 为局部特征图 $P$ 中 $(x,y)$ 处的特征向量; $S$ 为池化层的步长; $fusion_{(x,y)}$ 为 $(x,y)$ 处的融合特征向量, $\widehat{fusion}$ 为融合向量 $fusion$ 归一化后得到的最终向量。将融合的2048维特征向量作为最终表示,使用欧几里得距离进行行人再识别。特征融合网络的结构如图8所示。

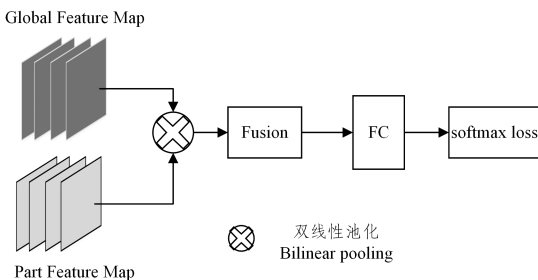


图8 特征融合网络的结构

Fig. 8 Structure of feature fusion network

相似性判别的伪代码如算法1所示。

### 算法1 行人相似性判别(Person similarity discrimination)

Input: Query image I, Test datasets Items, RE-ID model m;

Output: All images similar to the I;

Begin

1. Getting feature vector X from I according to m
2. for j in Items
3. Getting feature vector Y form j according to m;
4. Calculate the Euclidean distance  $d(X, Y)$ ;
5. Sort and output the image j according to the  $d(X, Y)$ ;
6. end for
7. end

## 4 实验

### 4.1 数据集

本文采用广泛使用的CHUK03<sup>[2]</sup>, Market-1501<sup>[1]</sup>, DukeMTMC-reID<sup>[3]</sup>数据集进行实验。

### 4.2 实验细节

使用累积匹配曲线(Cumulative Matching Characteristics, CMC)的rank1, rank5, rank10, rank20作为所有算法在3个数据集中的统一评价标准。对于Market-1501和DukeMTMC-reID数据集,额外使用平均精度(Mean Average Precision, mAP)作为评价标准。设定初始学习率为0.001,并在每 $2 \times 10^3$ 次迭代后将其逐渐降低。本文的主体网络为Resnet-50。实验的硬件环境为Intel i7-8700k CPU, 双GTX 1080ti, 64GB内存。所有图像在进入网络之前调整为 $256 \times 512$ 。

### 4.3 实验结果

#### 4.3.1 视图判别器的性能与算法复杂度分析

首先,为了验证视图判别器的准确率,分别对CHUK03, Market1501, DukeMTMC-reID数据集中的3000张图像进行视图标注,其中包含1300张“侧”视点图像。将这3000张图像作为测试集。本文由于只将视图判别器预测的结果应用于区域推荐,而算法为“前”“后”视点图像推荐同样的6个区域,因此认定只有当视图判别器将“前”“后”视点图像误判为“侧”视点或“侧”视点图像被误判时为错检。本文以准确率(Accuracy)作为视图判别器的评价指标。

$$Accuracy = \frac{TP}{P+N} \quad (16)$$

其中,TP为正确检测出视点的图像数量, $P+N$ 为测试集总数。视图判别器在3个数据集上的实验结果如表1所列。

表1 视图判别器在3个数据集上的实验结果

Table 1 Results of view discriminator on three datasets

数据集	正确检测	错检	Accuracy/%
CHUK03	837	163	83.7
Market1501	851	149	85.1
DukeMTMC-reID	853	147	85.3

实验结果表明,本文中的视图判别器在“侧”视点图像较多时,仍可以达到80%以上的视图预测准确率。

本文方法引入了姿势特征并进行了视图判别,会在一定程度上增加算法复杂度。为了检验本文方法是否有助于行人再识别,将其与其他3种方法(仅使用姿势特征的方法、仅使

用全局特征的方法、融合全局特征与姿势特征的方法)在3个公开数据集上进行对比,并以时间开销(即遍历一次测试集的平均时间)作为算法时间复杂度的评价指标,实验结果如表2所列。分析可知,本文方法虽然增加了时间开销,但获得了最优的准确率。在3个数据集中,与融合全局特征与姿势特征

的方法相比,本文方法的  $mAP$  与  $rank1$  的提升率均高于时间开销的增长率,如在 Market-1501 数据集中,  $mAP$  的提升率为 8.6%,  $rank1$  的提升率为 9.8%, 而时间开销的增长率为 7%。因此,本文方法虽然增加了一定的算法复杂度,但其准确率得到了显著提升。

表2 本文方法与对比方法在3个数据集上的实验结果

Table 2 Results of proposed algorithm and each individual algorithm on three datasets

方法	CHUK03		Market1501			DukeMTMC-reID		
	$rank1/\%$	$time/s$	$mAP/\%$	$rank1/\%$	$time/s$	$mAP/\%$	$rank1/\%$	$time/s$
姿势特征	50.2	66.5	42.4	58.8	1461.1	40.7	55.4	1304.6
全局特征	59.9	63.8	54.7	69.9	1381.2	52.1	65.6	1216.3
全局+姿势特征	66.4	79.8	58.2	76.8	1753.1	55.2	71.3	1576.8
本文方法(全局+视点+姿势)	71.3	84.4	63.2	84.3	1875.9	60.5	78.4	1702.7

#### 4.3.2 公开数据集上的结果对比

在 CHUK03 数据集上将本文方法与近几年性能最优的行人再识别方法进行对比,结果如图9所示。对比方法包括:距离度量学习方法(Local Maximal Occurrence+Cross-view Quadratic Discriminant Analysis<sup>[4]</sup> (LOMO+XQDA), Learning Discriminative Null Space<sup>[17]</sup> (LDNS)), 特征提取方法(Gaussian of Gaussian<sup>[18]</sup> (GOG))和深度学习方法(Gated Shape CNN<sup>[8]</sup> (Gated S-CNN), Pose Invariant Embedding<sup>[12]</sup> (PIE))。实验结果表明,本文方法获得了优于所有距离度量学习方法的性能;同时,引入人体姿势信息的 PIE 方法也获得了优于其他方法的性能。本文方法在引入视点机制后,  $rank1$  达到了 71.3%, 比 PIE 方法的  $rank1$  提高了 4.2%。

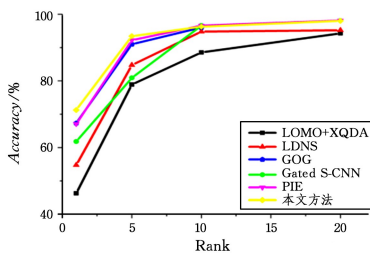


图9 CHUK03数据集上不同方法的实验结果对比

Fig. 9 Comparison of different algorithms on CHUK03 dataset

将本文方法与度量学习方法(Bag of Words+Keep it Simple and Straightforward Metric (BoW+KISSME)<sup>[1]</sup>, WARCA<sup>[5]</sup>, LDNS<sup>[17]</sup>)和深度学习方法(Gate S-CNN<sup>[8]</sup>, Long Short-Term Memory Siamese Convolutional Neural Network (LSTM S-CNN)<sup>[9]</sup>, PIE<sup>[12]</sup>)在 Market-1501 数据集上进行对比,结果如表3所列。可以看出,本文方法的  $mAP$  达到 63.2%,  $rank1$  达到 84.3%, 相比于最佳的比较方法 PIE,  $mAP$  和  $rank1$  分别提高了 7.3% 和 5.0%。

在极具挑战性的行人再识别数据集 DukeMTMC-reID 上,将本文方法与典型的度量学习方法(BoW+KISSME<sup>[1]</sup>, LOMO+XQDA<sup>[4]</sup>)和深度学习方法(Attribute-Complementary Recurrent Network (ACRN)<sup>[19]</sup>, Baseline+Lable Smooth Regularization for Outliers (Basel+LSRO)<sup>[20]</sup>, Singular Value Decomposition Network (SVDNet)<sup>[21]</sup>, PIE)进行对比,结果如表4所列。实验结果表明,本文方法在4个指标上都取得了最优越的性能,与 DukeMTMC-reID 数据集上效果较好的行

人再识别方法 SVDNet 相比,  $mAP$  获得了 3.7% 的提升,  $rank1$  获得了 1.7% 的提升。

表3 本文方法与对比方法在 Market-1501 数据集上的实验结果

Table 3 Results of proposed algorithm and each individual algorithm on Market-1501 dataset

方法	(单位: %)				
	$mAP$	$rank1$	$rank5$	$rank10$	$rank20$
BoW+KISSME <sup>[1]</sup>	20.8	44.4	69.9	72.2	79.0
WARCA <sup>[5]</sup>	—	45.2	68.1	76.0	84.0
LDNS <sup>[17]</sup>	29.9	55.4	—	—	—
LSTM S-CNN <sup>[9]</sup>	35.3	61.6	—	—	—
Gated S-CNN <sup>[8]</sup>	39.6	65.9	—	—	—
PIE <sup>[12]</sup>	55.9	79.3	90.8	94.4	96.7
本文方法	63.2	84.3	90.9	95.5	95.9

表4 本文方法与对比方法在 DukeMTMC-reID 数据集上的实验结果

Table 4 Results of proposed algorithm and each individual algorithm on DukeMTMC-reID

方法	(单位: %)				
	$mAP$	$rank1$	$rank5$	$rank10$	$rank20$
BoW+KISSME <sup>[1]</sup>	12.2	25.1	—	—	—
LOMO+XQDA <sup>[4]</sup>	17.0	30.8	—	—	—
ACRN <sup>[19]</sup>	52.0	72.6	84.8	88.9	—
Basel+LSRO <sup>[20]</sup>	47.1	67.7	—	—	—
SVDNet <sup>[21]</sup>	56.8	76.7	86.4	89.9	—
本文方法	60.5	78.4	87.7	90.4	92.2

通过上述实验结果可知,本文方法充分考虑了行人姿势信息与相机视点信息,并且引入特征融合网络,将局部特征与全局特征进行融合,最终提高了行人再识别方法的性能。

**结束语** 本文提出了一种融合视点机制与姿态估计的行人再识别方法,首先使用姿态估计网络估计人体关节点,并对图像进行视图判别;然后根据图像视点划分人体关节区域作为局部特征;最后使用特征融合网络将全局特征表示与局部特征表示融合,得到更具鲁棒性的融合特征表示。实验结果表明本文方法在三大主流数据集上均取得了较好的效果。但是,所提方法仍有很多问题需要完善,如将其他人体姿势数据集中训练的模型迁移到该方法中使用时会导致部分关节点定位错误,从而导致准确率下降,这也是后续工作需要突破的难点和重点。

## 参考文献

[1] ZHENG L, SHEN L, TIAN L, et al. Scalable Person Re-identifi-

- cation: A Benchmark[C]//2015 IEEE International Conference on Computer Vision (ICCV). IEEE Computer Society, 2015: 1116-1124.
- [2] 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). IEEE Computer Society, 2014: 152-159.
- [3] RISTANI E, SOLERA F, ZOU R, et al. Performance measures and a data set for multi-target, multi-camera Tracking[C]//European Conference on Computer Vision workshop on Benchmarking Multi-Target Tracking, 2016: 17-35.
- [4] LIAO S, HU Y, ZHU N X, et al. Person re-identification by Local Maximal Occurrence representation and metric learning [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2015: 2197-2206.
- [5] JOSE C, FLEURET F. Scalable metric learning via weighted approximate rank component analysis[C]//European Conference on Computer Vision. Cham: Springer, 2016: 875-890.
- [6] LIAO S, LI S Z. Efficient PSD Constrained Asymmetric Metric Learning for Person Re-identification[C]//2015 IEEE International Conference on Computer Vision (ICCV). IEEE Computer Society, 2015: 3685-3693.
- [7] YI D, LEI Z, LIAO S, et al. Deep Metric Learning for Person Re-identification[C]//2014 22nd International Conference on Pattern Recognition (ICPR). IEEE Computer Society, 2014: 34-39.
- [8] VARIOR R R, SHUAI B, LU J W, et al. A siamese long short-term memory architecture for human re-identification[C]//European Conference on Computer Vision. Cham: Springer, 2016: 135-153.
- [9] VARIOR R R, HALOI M, WANG G. Gated siamese convolutional neural network architecture for human re-identification [C]//European Conference on Computer Vision. Cham: Springer, 2016: 791-808.
- [10] DING S, LIN L, WANG G, et al. Deep feature learning with relative distance comparison for person re-identification[J]. Pattern Recognition, 2015, 48(10): 2993-3003.
- [11] WANG F, ZUO W, LIN L, et al. Joint learning of single-image and cross-image representations for person re-identification [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2016: 1288-1296.
- [12] ZHENG L, HUANG Y, LU H, et al. Pose Invariant Embedding for Deep Person Re-identification[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2017: 1456-1465.
- [13] ZHAO H, TIAN M, SUN S, et al. Spindle Net: Person Re-identification with Human Body Region Guided Feature Decomposition and Fusion[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2017: 1077-1085.
- [14] WANG H, LIU Z F, FANG B F, et al. Human pose estimation based on constellation tree structure appearance model[J]. Computer Science, 2014, 41(3): 76-79.
- [15] LI D, ZHANG Z, CHEN X, et al. A Richly Annotated Dataset for Pedestrian Attribute Recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2016.
- [16] CAO Z, SIMON T, WEI S E, et al. Realtime multi-person 2D pose estimation using part affinity fields[C]//IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2017: 1302-1310.
- [17] ZHANG L, XIANG T, GONG S. Learning a Discriminative Null Space for Person Re-identification[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016: 1239-1248.
- [18] MATSUKAWA T, OKABE T, SUZUKI E, et al. Hierarchical Gaussian Descriptor for Person Re-identification [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016: 1363-1372.
- [19] SCHUMANN A, STIEFELHAGEN R. Person Re-identification by Deep Learning Attribute-Complementary Information[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE Computer Society, 2017: 1435-1443.
- [20] ZHENG Z, ZHENG L, YANG Y. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro [C]//IEEE International Conference on Computer Vision. IEEE Computer Society, 2017: 3774-3782.
- [21] SUN Y, ZHENG L, DENG W, et al. SVDNet for Pedestrian Retrieval[C]//IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 3820-3828.



**PEI Jia-zhen**, born in 1995, master. His main research interests include person re-identification, computer vision and deep learning.



**XU Zeng-chun**, born in 1973, engineer. Her main research interests include computer science and technology, wireless sensor networks and intelligence computation.