

一种多视角全局特征融合的行人再识别方法

江运衡, 赵帅

(北京邮电大学网络与交换技术国家重点实验室, 北京, 100876)

摘要: 行人再识别是通过行人图像数据进行跨时空的行人识别方法, 是用于智慧城市、智慧监控等建设的关键技术之一, 近年来越来越受到研究者的关注。本文提出了一种采用多卷积视角的方式分组挖掘行人全局特征, 再将挖掘到的特征融合用于行人再识别的方法。在开源行人再识别数据集上进行训练和验证, 并且取得了82.1的mAP和92.0%的准确率, 证明了本方法的有效性。

关键词: 计算机应用技术; 计算机视觉; 行人再识别; 全局特征

中图分类号: TP399

A Pedestrian Re-identification Method Based on Multi-view Global Feature Fusion

Jiang Yunheng, Zhao Shuai

(State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, 100876)

Abstract: Pedestrian re-identification is a method for identifying pedestrians across time and space through pedestrian image data. It is one of the key technologies used in construction of smart cities and smart monitoring and has attracted more and more attention from researchers in recent years. This paper proposes a method that uses the multi-convolution perspective to mine the pedestrian's global features in different groups, and the fuses the mined features for pedestrian re-identification. Training and verifying experiments on the open source dataset achieved an mAP of 82.1 and an accuracy rate of 92.0%, which proved the effectiveness of the proposed method.

Key words: computer application technology; computer vision; pedestrian re-identification; global feature

0 引言

随着技术能力的进步, 网络监控技术也迅速发展。现今, 基于建设信息社会、治安维稳、打击犯罪的实际需求, 国家进一步推动智慧城市、天网工程等建设项目, 也因此越来越多的监控摄像机被投入使用, 一个广域的监控网络已经在我国初步完成。近年, 计算机视觉领域的一些基础挑战被快速发展的深度神经网络技术解决, 引入计算机视觉技术进行监控视频的分析处理是建设智慧监控系统的发展趋势。借助计算机视觉技术可以对监控设备采集的视频数据不间断的进行分析和处理, 对可疑人物进行自动识别和报警处理, 例如我国公安机关的“天网工程”在过去几年就通过人脸识别技术对监控视频的分析进行过多次抓捕罪犯的行动, 为打击犯罪做出巨大贡献。

作者简介: 江运衡 (1995 年-), 男, 主要研究方向: 目标检测、行人重识别

通信联系人: 赵帅 (1987 年-), 男, 副教授, 硕导, 主要研究方向: 多模态数据融合、物联网与大数据处理、网络服务与智能. E-mail: zhaoshuaiby@bupt.edu.cn

受限于监控摄像头的位置固定、俯仰角固定、天气带来的光照变化等多种因素,很多场景下监控摄像头可以捕捉行人图像但是无法获得清晰有效的人脸图像,基于实际需要行人再识别相关技术吸引了计算机视觉领域学者的关注。行人再识别(Pedestrians Re-Identification)是指借助计算机视觉技术,在跨摄像头、不同角度、不同光照、不同分辨率的多张行人图像数据中识别出用于查询的行人目标,即给定查询行人图像,系统可从下辖的所有摄像头收集的图像数据中检索到该行人的其他图像数据,整个过程是一个更加细致化的以图搜图的过程,因此行人再识别也被认为是图片检索问题。

行人再识别技术的核心在于如何能够提取有效的、有区分性的高维特征,能够辨别是否是同一个人这个基本问题。本文提出了一种融合了多个视角的全局特征方法,通过不同卷积视角“观察”一张图片由神经网络提取的全局特征图从而生成可供判别的特征,并在开源数据集进行验证。

1 相关工作

行人再识别问题可以被抽象成一个公式语言,如式 1-1.即为给定一张行人查询图片 q 和行人图库数据 $G=\{g_1, g_2, \dots, g_n\}$,行人再识别算法输出:

$$\theta = \arg \min D(q, g_i) \quad g_i \in G \quad (1-1)$$

式 1-1 中 D 表示一种距离度量算法,描述两个行人描述子之间的距离远近,即距离近的两两相似度高,距离远的两两相似度低。综上所述,如何提取行人图像中有效的、有区分性的特征形成该行人的描述子是其核心问题。

早期学者们对行人再识别的研究聚焦在如何人为设置有效特征,最常被用到的特征有颜色、纹理。Gheissari^[1]等人在 2006 年提出一种分割算法去分割图像前景、背景,并计算局部的 HS 直方图和边缘直方图作为描述图像的特征。Grey 和 Tao^[2]首先将行人图像水平划分,再对对于每个子区域提取 RGB、HS 和 YCbCr8 个颜色通道特征,用 21 种纹理的滤波器提取纹理特征。2010 年 Bazzani^[3]等人提出了切割行人图像前景、背景后,基于对称、非对称原则自适应的选择行人差异化部位计算其加权颜色直方图、最稳定色彩区、复高结构块等特征。除了最常使用颜色、纹理特征外,属性特征被认为是一种更加鲁棒、更加高阶的特征。Layne^[4]等人在 2012 年开始标注颜色、纹理等低阶特征训练一个属性分类器,通过分类器的属性特征加权后成为特征向量与其他图像特征融合使用。Liu^[5]等人以无监督学习的方式抽象出一些具有共同属性特征的行人模型并自适应的分配各个属性特征之间的权重组合成特征向量。2015 年 Su^[6]等人将同一个人在不同摄像头下图像数据的二进制语义特征编码到一个连续的低秩属性空间以增加相似性匹配过程中的区分度。尽管各种手动设置的特征都被研究过,但是这些特征的表征能力不强,并不能很好的解决行人再识别任务。

2014 年基于深度卷积神经网络的方法刷新了图像分类任务,展现了 CNN 在计算机视觉领域的的能力。同年 D. Yi^[7] 等人将输入的行人图片水平划分成三个有重叠的部件,每个部件送入一个两层卷积和全连接组成的网络提取特征融合成特征向量,特征向量之间用余弦距离进行距离度量。Wei Li^[8] 等人设计了一个 FPNN (filter pairing neural network) 结构,以同一个行人的两张不同图像作为输入,经过一个卷积池化层的处理后水平划分多个部件分析经

过特征提取,在反向传播过程中只更新最大响应滤波器传递的梯度。Varior^[9] 在每个卷积层后面插入一个选通函数去捕捉一对图片输入的更细微的特征,这在当时刷新了很多在标准数据集上的性能,但是该算法需要多次的配对查询,需要大量的冗余计算。Liu^[10] 继承了 Varior 的算法的主要思路,增加了注意力模型机制让网络自适应的关注更重要的局部信息,但是并没有解决多次配对查询的冗余计算问题。使用孪生网络做配对查询的策略有一个很大的缺陷,即每次网络只学习两个或三个输入图像的标签信息来给出这对输入的相似性。为了充分利用所有的标签,一些学者开始研究利用分类识别网络去提取图像特征。2016 年 T. Xiao^[11] 等人采用多数据集联合训练了一个分类网络,取全连接层前的输出结合不同数据集域的评分作为特征向量,取得了不错的效果。Wu^[12] 等人尝试将神经网络全连接层提取的特征向量串联手工设置的特征向量,再馈送到另一个全连接层分类器进行分类训练。

2 多视角全局特征融合网络

本节将介绍本文所提出的 MVG (Multi-View Global feature) 行人再识别模型,包括其特征提取网络和全局特征处理方法以及一些训练技巧和训练过程。

2.1 MVG 网络结构

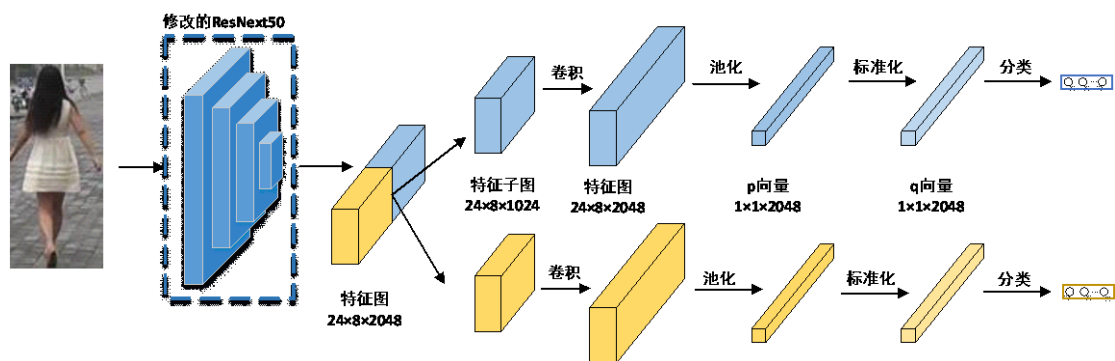


图1 MVG 网络结构

MVG 网络结构如图 1 所示。MVG 选取修改过的 ResNext50^[13]作为网络中的特征提取网络。ResNext50 的结构可以简单看作 6 层,一个输入层、四个中间层和一个输出层。其输入层和四个中间层分别有一次 2 倍的下采样,总共 32 倍下采样。MVG 采用 ResNext50 的输入层和四个中间层作为特征提取网络,并且为了保留更多的特征信息,修改最后一个中间层使它没有 2 倍的下采样从而获得更大尺寸的特征图。

将所有的输入行人图片重新缩放到 384×128 的高宽尺度,通过特征提取网络得到通道数为 2048 的 24×8 的特征图,再将特征图在通道维度进行两等分,得到两个通道数为 1024 的 24×8 的特征子图。每个特征子图作为一个分支采用相同策略互相独立进行处理,首先通过卷积将通道扩容至 2048,用平均池化降低尺寸使特征图转化为特征向量 p,再通过一次批标准化形成特征向量 q,最后通过全连接层分类器进行分类。在测试阶段,串联两个分支的 p 向量作为行人的描述子进行距离度量。

模型的核心在于特征图在通道维度的划分和通道扩容。卷积的本质是一个滑动窗口,以卷积所含参数的视角滑动观察图片。特征提取网络输出的特征图是关于行人的一个全局特征,通过卷积对全局特征进行通道扩容的本质是利用卷积参数的视角进行更细致的观察、更深度的挖掘。而对特征图的通道划分是采用分组卷积实现的,这借鉴了 ResNext 设计思路,

能够降低上下层连接、减少参数量和计算量，从一定程度上加速网络收敛、抑制了网络过拟合。

2.2 网络训练

2.2.1 数据集及训练环境

为了验证本文提出模型的有效性，本文选择在行人再识别领域被广为使用的开源数据集 Market1501^[14]进行训练和验证。Market1501 数据是在夏季的清华大学校园内使用 6 个摄像头采集拍摄的，总共采集了 1501 个行人，共计 32668 张图片数据，每个行人的数据至少来自两个摄像头。其训练集和测试集分布情况如表 1 所示，训练集共 751 个行人的 12936 张图片数据，测试集分为查询图片和图集，通过查询对象在图集的正确匹配情况来区分模型的优劣。

表 1 Market1501 数据集

用途	行人数	图片数
训练	751	12936
测试（查询）	750	3368
测试（图集）	750	16364

本文所提模型在 Ubuntu 操作系统下利用 Python 语言和 Pytorch 深度学习框架实现，并且用到了英伟达的 GPU 硬件及相关技术进行加速。各软硬件具体型号或版本如表 2 所示。

表 2 训练环境

处理器	Intel(R) Xeon(R) Gold 5118 CPU
内存	256GB
显卡	Tesla P40
显存	24GB
GPU 库	CUDA10.1、CUDNN7603
编程语言	Python 3.6
深度学习框架	Pytorch 1.4.0

2.2.2 训练策略及训练

模型用 Market1501 的训练集进行训练，训练过程使用了两张显卡，采用 Adam 优化器，每个 batch 送入 64 张图片，总共训练了 100 个 epoch。为了将模型训练到一个更有效、更精确的状态，得到更好的行人再识别效果，本文在训练过程中采取了一系列帮助训练的策略，包括学习率热身、随机遮挡、标签平滑、双损失函数。

1. 学习率热身

常规的训练方法是先用一个较大的学习率让模型快速收敛，再用较小的学习率让模型收敛到一个更好的状态。学习率热身策略则是指先用一个较小的学习率让网络进行热身训练，再进行上述训练过程，这个策略在多个学者的不同实验中被证明对模型效果有增益作用。本文的学习率设置情况如图 2 所示，先以 0.00001 的学习率进行 10 个 epoch 的热身，再依次以 0.0001、0.00001 和 0.000001 的学习率分别训练 30 个 epoch。

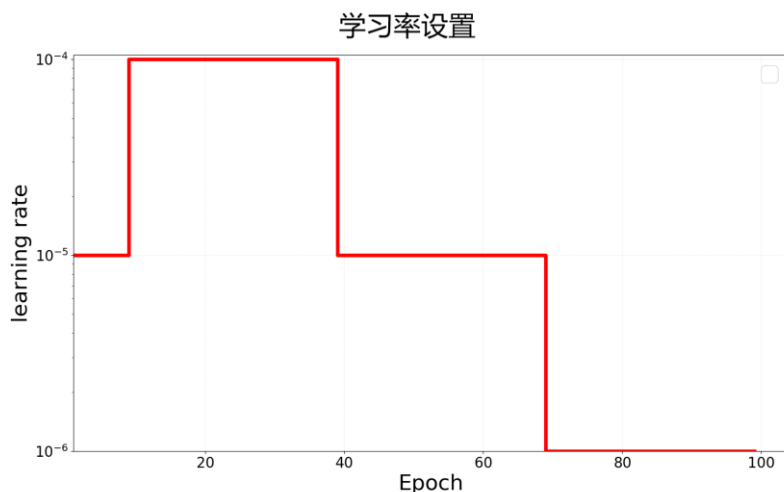


图2 学习率变化

2. 数据增强

本文对每次输入的图片采取了二分之一概率随机水平翻转和随机遮挡的数据增强手段，如图3所示。数据增强对于小型数据集而言是被普遍采用的手段，能够扩大输入数据间的差异性从而增强模型的鲁棒性。



图3 数据增强可视化示例

3. 标签平滑

传统的分类网络使用的标签都是 one-hot 编码形式，即一个长度为分类任务类别数的向量，响应位置为 1，其余位置为 0。标签平滑技术最早在[15]中提到，其核心操作是针对 one-hot 编码进行微小程度的平滑，具体如式 2-1 所示，式中 y 是 one-hot 标签， y' 是平滑后的标签， ε 是平滑系数， K 是类别数。

$$y' = (1 - \varepsilon)y + \frac{\varepsilon}{K} \quad (2-1)$$

标签平滑技术可以抑制模型过度信任 one-hot 编码中的 1 而忽视 0 的问题，尤其是在数据集较小的情况下，从一定程度上防止模型过拟合。

4. 双损失函数

不同于传统的分类任务，本文在训练过程中使用了交叉熵损失函数做分类损失的同时还

增加了 Triplet loss^[16]进行高层特征的度量训练，Triplet loss 会在一个 batch 中挑选最难的正样本和最难的负样本进行一个三元损失度量。双损失函数的设置可以增加模型对类间距离和类内距离的区分能力，从而提升重识别的准确率。

3 实验与分析

155 基于上述训练得到的网络参数，本文在 Market1501 的测试集上进行模型验证，验证结果如表 3 所示。其中 mAP 指标和 Rank-1 指标分别达到了 82.1 和 92.0%，体现了提出的模型的有效性。

表 3 验证结果

指标	mAP	Rank-1	Rank-5	Rank-10
结果	82.1	92.0%	97.7%	98.8%

160 为了进一步探讨模型的各个分支和特征向量对性能的影响，本文对其两个分支的不同特征向量进行了单独的测试，其结果如表 4 所示。

表 4 各部件性能分析

	mAP	Rank-1	Rank-5	Rank-10
上支路 p 向量	81.5	91.5%	97.7%	98.7%
上支路 q 向量	81.6	91.4%	97.6%	98.8%
下支路 p 向量	81.8	91.6%	97.5%	98.7%
下支路 q 向量	81.8	91.6%	97.4%	98.7%
上下支路 q 向量串联	82.1	91.9%	97.7%	98.7%
MVG(上下支路 p 向量串联)	82.1	92.0%	97.7%	98.8%

对比发现单支路的性能表现也足够出色，下支路表现比上支路更强，可能下支路在训练过程中收敛到了一个更佳的状态。p 向量和 q 向量用作描述子的测试结果无显著差异。两支路对应特征向量串联能够带来 mAP 和 Rank-1 的一定程度提升。

165 本文将 MVG 的结果与其他行人再识别方法在 Market1501 上的指标进行了比较，其结果如表 5 所示。可以发现 MVG 超越了多数其他行人再识别方法，而对比性能超出 MVG 的方法发现，MVG 的网络结构相较之下更加简单，参数量更少，更易于实现。

表 5 与其他行人再识别方法对比

	mAP	Rank-1
IDE ^[17]	46.0	72.5%
PAN ^[18]	63.4	82.8%
SVDNet ^[19]	62.1	82.3%
DPFL ^[20]	73.1	88.9%
HA-CNN ^[21]	75.7	91.2%
SCPNet ^[22]	75.2	91.2%
PCB ^[23]	77.3	92.4%
BDB ^[24]	85.0	94.5%
MGN ^[25]	86.9	95.7%
MVG	82.1	92.0%

4 结论

本文提出了一种多视角全局特征的行人重识别模型,对于特征提取网络提取的行人全局特征,我们用分组卷积进行通道扩容的方式对全局特征进行更细致的挖掘得到最终的描述子。本文所提模型的结构简单,参数量少,并且在 Market1501 数据集上进行训练和验证,取得了 82.1 的 mAP 和 92.0% 的 Rank-1 准确率。

[参考文献] (References)

- [1] Gheissari N, Sebastian T B, Hartley R. Person reidentification using spatiotemporal appearance[C]//2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). IEEE, 2006, 2: 1528-1535.
- [2] Gray D, Tao H. Viewpoint invariant pedestrian recognition with an ensemble of localized features[C]//European conference on computer vision. Springer, Berlin, Heidelberg, 2008: 262-275.
- [3] Bazzani L, Cristani M, Perina A, et al. Multiple-Shot Person Re-identification by HPE Signature[C]// 2010 International Conference on Pattern Recognition. IEEE Computer Society, 2010:1413-1416
- [4] Layne R, Hospedales T M, Gong S, et al. Person re-identification by attributes[C]//Bmvc. 2012, 2(3): 8.
- [5] Liu X, Song M, Zhao Q, et al. Attribute-restricted latent topic model for person re-identification[J]. Pattern recognition, 2012, 45(12): 4204-4213.
- [6] Su C, Yang F, Zhang S, et al. Multi-task learning with low rank attribute embedding for person re-identification[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 3739-3747.
- [7] Yi D, Lei Z, Liao S, et al. Deep metric learning for person re-identification[C]//2014 22nd International Conference on Pattern Recognition. IEEE, 2014: 34-39.
- [8] Li W, Zhao R, Xiao T, et al. Deepreid: Deep filter pairing neural network for person re-identification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 152-159.
- [9] Varior R R, Haloi M, Wang G. Gated siamese convolutional neural network architecture for human re-identification[C]//European conference on computer vision. Springer, Cham, 2016: 791-808.
- [10] Liu H, Feng J, Qi M, et al. End-to-end comparative attention networks for person re-identification[J]. IEEE Transactions on Image Processing, 2017, 26(7): 3492-3506.
- [11] Xiao T, Li H, Ouyang W, et al. Learning deep feature representations with domain guided dropout for person re-identification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1249-1258.
- [12] Wu S, Chen Y C, Li X, et al. An enhanced deep feature representation for person re-identification[C]//2016 IEEE winter conference on applications of computer vision (WACV). IEEE, 2016: 1-8.
- [13] Xie S, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1492-1500.
- [14] Zheng L, Shen L, Tian L, et al. Scalable person re-identification: A benchmark[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1116-1124.
- [15] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2818-2826.
- [16] Hermans A, Beyer L, Leibe B. In defense of the triplet loss for person re-identification[J]. arXiv preprint arXiv:1703.07737, 2017.
- [17] Zheng L, Yang Y, Hauptmann A G. Person re-identification: Past, present and future[J]. arXiv preprint arXiv:1610.02984, 2016.
- [18] Zheng Z, Zheng L, Yang Y. Pedestrian alignment network for large-scale person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 29(10): 3037-3045.
- [19] Sun Y, Zheng L, Deng W, et al. Svdnet for pedestrian retrieval[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 3800-3808.
- [20] Chen Y, Zhu X, Gong S. Person re-identification by deep learning multi-scale representations[C]//Proceedings of the IEEE International Conference on Computer Vision Workshops. 2017: 2590-2600.
- [21] Li W, Zhu X, Gong S. Harmonious attention network for person re-identification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 2285-2294.
- [22] Fan X, Luo H, Zhang X, et al. Scpnet: Spatial-channel parallelism network for joint holistic and partial person re-identification[C]//Asian Conference on Computer Vision. Springer, Cham, 2018: 19-34.
- [23] Sun Y, Zheng L, Yang Y, et al. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 480-496.
- [24] Dai Z, Chen M, Gu X, et al. Batch DropBlock network for person re-identification and beyond[C]//Proceedings of the IEEE International Conference on Computer Vision. 2019: 3691-3701.
- [25] Wang G, Yuan Y, Chen X, et al. Learning discriminative features with multiple granularities for person

re-identification[C]//Proceedings of the 26th ACM international conference on Multimedia. 2018: 274-282.