

一种针对复杂场景的行人再识别优化方法*

王楠 吴健 卢杏坚 钟震宇 黄凯

(广东省科学院智能制造研究所/广东省现代控制技术重点实验室, 广东 广州 510070)

摘要: 针对室外场景中, 因摄像头角度不统一、室外光照多变等因素导致行人难识别的问题, 提出一种针对复杂场景的行人再识别优化方法。通过在 ResNet 的浅层嵌入实例-批归一化层和空间变换网络, 引入 Mish 激活函数, 构建改进的 ResNet-50 特征提取骨干网络, 建立面向复杂场景的行人再识别模型。经在 MSMT17 数据集上验证, rank-1 和 mAP 分别达到 79.8%、58.5%, 说明了该算法的有效性。

关键词: 复杂场景; 行人再识别; 实例-批归一化; 空间变换网络; Mish 激活函数

中图分类号: S225

文献标识码: A

文章编号: 1674-2605(2021)06-0005-06

DOI: 10.3969/j.issn.1674-2605.2021.06.005

0 引言

公共安全是国家安全的重要组成部分, 安防监控系统的规模化普及, 构建了良好的社会治安防控体系。安防监控系统中大量的监控设备带来海量视频数据, 传统基于人工的视频分析方式耗时费力, 已无法满足高效、快速处理的实际需求。随着计算机视觉技术快速发展, 其在众多领域大放异彩^[1-3], 尤其在公共安全的识别、定位、检测等方面, 取得了令人瞩目的成绩。其中人脸识别、指纹识别为最典型的应用, 但多适用于支付、银行、门禁等近距离约束场景, 在跨域远距离监控系统中难以应用。

行人再识别技术是近年来提出的更具挑战性的计算机视觉任务, 旨在对处于不同位置且不同监控摄像头里的指定人员进行身份信息检索。行人再识别技术以整个行人的特征作为相似性度量的依据, 且行人特征不受面部遮挡、低头或背对摄像头的影响, 识别准确率更高, 可应用于行人查找、追踪等领域, 能显著提升破案侦查、行人追踪等效率, 对维护公共安全和社会稳定具有重要意义。

近年来, 许多专家学者对行人再识别技术开展相关研究。FARENZNA 等采用将人体分为多个小部块

的方式进行行人再识别研究^[4]。PEDAGADI 等利用局部线性判定分析方法进行行人再识别^[5]。鲁臣等采用串联多个特征层的方式, 将最后深层特征和浅层特征利用特征层融合的机制提取行人图像特征^[6]。潘少明等采用图卷积神经网络的行人再识别方法, 将一个域的数据信息迁移到目标域的数据中, 提升了模型跨域能力^[7]。陈莹等基于深度学习框架, 提出一种困难样本的三元组损失的行人再识别网络, 以身份和属性标签的方法获得更多行人判别^[8]。

尽管以上专家学者利用行人再识别技术在公共安全领域做了很多探索并取得一定成效, 但当前方法存在 2 个问题限制了其在实际场景的应用。1) 光照干扰。在实际场景下, 跨域远距离监控系统通常由多个摄像头获取监控图像, 不同地点的摄像头可能处于不同的光照亮度下, 导致获取的图像具有不同的光照强度, 给检测模型带来一定误差, 影响检测精度。2) 角度多变。为监控特定区域, 同一地点的摄像头可能有不同的安装角度, 导致图像中的人物存在视觉畸变现象和图像特征差异性大的问题, 使跨域远距离监控系统因误判而丢失目标。

针对上述问题, 本文提出一种针对复杂场景的行人再识别方法。首先, 采用改进的 Mosaic 数据增强

方法生成背景复杂的行人图像,增加网络对复杂场景的鲁棒性;然后,设计基于实例-批归一化(instance-batch normalization, IBN)^[9]的行人图像矫正网络和基于空间变换网络(spatial transformer network, STN)^[10]的行人特征空间变换网络,提升网络的外观变化适应性和空间不变性;接着,采用 Mish 激活函数优化网络激活函数;最后,构建改进的 ResNet 行人再识别模型进行训练。

1 数据集建立

1.1 MSMT17 数据集^[11]

MSMT17 数据集包含 15 个摄像头(室外 12 个、室内 3 个);覆盖一个月内不同的天气;每天采样 3 h,涵盖早、中、晚 3 个时段;共采集 4 101 个行人,其中训练集包括 1 041 个行人,测试集包括 3 060 个行人。该数据集具有如下特点:行人和摄像头数量多;场景和背景更加复杂;光照变化更为复杂;较接近真实场景。因此,本文选择 MSMT17 数据集作为训练集和测试集。

1.2 改进的 Mosaic 数据增强方法

为提高数据的鲁棒性和多样性,采用改进的 Mosaic 数据增强方法对行人再识别数据集进行背景复杂化。随机选取 4 幅行人图像,分别对这些图像作如下 4 种操作:

- 1) 亮度增强。考虑到室外场景的光照变化,对图像进行整体的亮度提升。
- 2) 随机旋转。设置随机旋转角度为 $10^{\circ}\sim 60^{\circ}$,随机翻转图像。
- 3) 水平翻转。通过水平翻转图像的方式改变图像的特征分布。
- 4) 缩放。减小原始图像尺寸,使数据多样化。

将处理后的 4 幅图像进行拼接,形成 Mosaic 增强后的图像。图像增强前后对比图如图 1 所示。



图 1 图像增强前后对比图

2 基于改进的 ResNet^[12]行人再识别模型

针对室外复杂场景下光照变化导致行人衣着颜色改变、行人姿态多样等特点,设计基于 IBN 的行人图像矫正网络和基于 STN 的行人特征空间变换网络,引入 Mish 激活函数,重构基于 IBN 和 STN 改进的 ResNet-50 特征提取骨干网络,增强网络对行人外观变化的抗干扰能力,提高网络在跨场景下的鲁棒性。基于改进的 ResNet 行人再识别模型结构如图 2 所示。

在训练阶段,利用改进的 Mosaic 数据增强方法对训练集进行数据增强;以训练集为输入,STN 自适应调整行人异常姿态;STN 提取的特征输入改进的 ResNet-50 特征提取骨干网络,该网络由 5 个模块组成,均嵌入 IBN 模块,以提升网络因光照变化而引发的外观变化的适应性;用 Mish 激活函数代替残差模块中的 ReLU 激活函数,解决负输入失效问题。

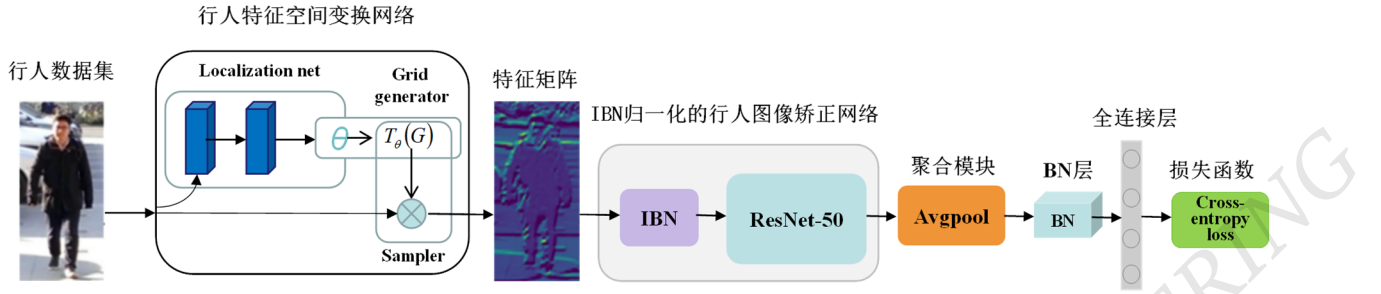


图2 基于改进的 ResNet 行人再识别模型结构

2.1 基于 STN 的行人特征空间变换网络

行人再识别模型因行人姿态和视角不同可能出现行人特征无法匹配的现象，导致其难以在室外真实场景下部署。图 3(a)为同一行人在同一摄像头下表现出双腿打开/交叉、打电话/低头、手部弯曲等姿态；图 3(b)为一个行人由远到近的视角变化。因 STN 具有平移不变性和旋转不变性的特点，故本文利用 STN 自适应（旋转、平移、缩放等）调整行人姿态异常，并基于 ResNet-50 网络，构建基于 STN 的行人特征空间变换网络。



(a) 姿态变化



(b) 视角由远到近

图3 行人姿态和视角变化

基于 STN 的行人特征空间变换网络由定位网络（Localization net）、网络生成器（Grid generator）、采样器（Sampler）3 个基本模块构成，如图 2 所示。

首先，在卷积层和全连接层操作后，将输入图像

发送到定位网络，获得 2×3 矩阵 A_θ ：

$$A_\theta = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \quad (1)$$

然后，在网络生成器中，根据式(2)得到优化后的输入图像：

$$\begin{pmatrix} x_i^s \\ y_i^s \\ 1 \end{pmatrix} = A_\theta \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} = T_\theta(G_i) \quad (2)$$

式中， (x_i^s, y_i^s) 为输入图像坐标； (x_i^t, y_i^t) 为输出图像坐标； $T_\theta(G_i)$ 为输入输出图像之间的映射关系，在采样器中用于对原始像素进行重采样，形成变换后的图像。

2.2 基于 IBN 的行人图像矫正网络

在室外真实场景下，光照随时间变化，不同强度的光照照射在行人身上，会发生不同程度的“变色”现象，如图 4 所示。



图4 室外光照变化引发的“变色”现象

行人外观变化会影响行人再识别模型的准确度。此外，当训练数据集和测试数据集有较大的外观差异

时,模型性能也会显著下降。如数据集中目标光线强烈,训练得到的模型学习了光线强烈的特征,在模型对光线昏暗的数据进行测试时,测试效果较差。为此,本文提出基于 IBN 的行人图像矫正网络,提升模型对图像外观变化的适应性。

嵌入 IBN 的残差模块如图 5 所示。

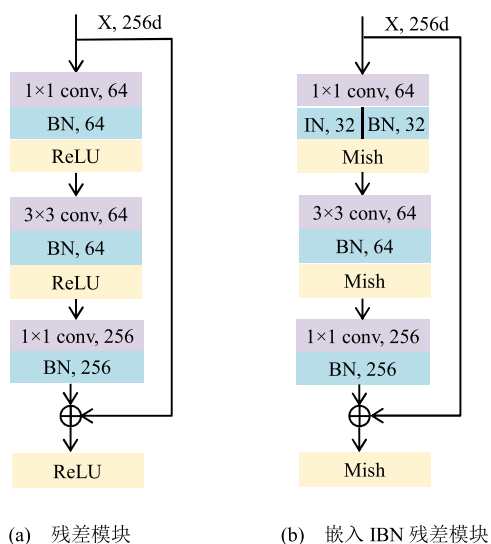


图 5 嵌入 IBN 的残差模块

首先,经过卷积层的特征向量被平分为 2 组;然后,分别经过 IN 层和 BN 层再拼接在一起,以提升模型的学习能力和泛化能力。考虑室外真实场景下,行人再识别数据集中的数据存在跨域问题,将 IBN 层嵌入到 ResNet 网络,用于提升模型在外观变化下的适应能力。

2.3 Mish 激活函数

相比于 Sigmoid、ReLU、Tanh 这 3 种激活函数,Mish 激活函数能在一定程度上解决梯度消失问题;而且相比于 ReLU 激活函数,Mish 激活函数存在一小段负区间,解决了负输入失效问题。

Mish 激活函数表达式为

$$f(x) = x \tanh[\ln(1 + e^x)] \quad (3)$$

Mish 激活函数没有上界有下界,没有上界避免了由于封顶而导致的饱和问题;有下界有助于加快训练

过程,且具有较好的光滑性,可提升泛化能力和优化能力,提高结果质量。因此,本文使用 Mish 激活函数代替残差模块中的 ReLU 激活函数,提高模型的泛化能力。

3 实验验证

3.1 模型训练

实验在 Linux 系统 Visual Studio Code 环境下,基于 pytorch 深度学习框架,采用 python 语言编程完成。实验平台硬件配置:英特尔 i7-7700HQ 处理器、Nvidia GeForce GTX2080、AOC 显示器。本文基于改进的 ResNet 行人再识别模型的超参数设置如表 1 所示。

表 1 基于改进的 ResNet 行人再识别模型的超参数设置

参数名称	参数值
Epoch	120
Batch size	64
Learning rate	0.001
Gamma	0.1
size	(128,256)

3.2 评价指标

基于改进的 ResNet 行人再识别模型性能评价指标主要包括 rank- i 和平均精度均值 (mean average precision, mAP), 其中 rank- i 表示排序前 i 位中含有正确行人的概率。

在单目标分类任务中,通常给定查询图像,然后在测试集中通过距离度量函数计算图像相似度。mAP 计算公式为

$$mAP = \frac{\sum_{i=1}^C AP_i}{C} \quad (4)$$

$$AP = \frac{\sum_i^N pre}{N} \quad (5)$$

$$pre = \frac{TP}{TP + FP} \quad (6)$$

式中, TP 为标签为真, 预测为真的数量; FP 为标签为真, 预测为假的数量; N 为该类别图像数量; C 为查询类别的数量; AP 为平均准确率 (average precision, AP); pre 为精确率 (precision)。

3.3 实验结果

3.3.1 复杂光照条件下的实验结果分析

为验证本文方法的实用性, 采用多重粒度网络

(multiple granularity network, MGN)、ResNet-50+CircleLoss 这 2 种常用的行人再识别方法与本文提出的方法进行对比测试。此外, 考虑到 MSMT17 测试集中图像光照的多变性, 首先, 对 MSMT17 测试集中过暗的图像进行亮度增强, 过亮的图像进行亮度消减, 处理流程如图 6 所示; 然后, 采用基于改进的 ResNet 行人再识别模型进行测试, 测试结果如表 2 所示。

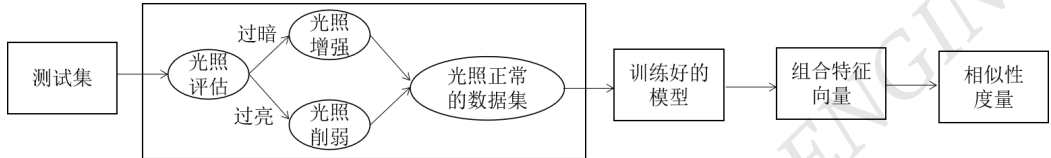


图 6 MSMT17 测试集光照处理流程图

表 2 3 种方法测试结果

方法	MGN	ResNet-50+CircleLoss	本文方法	光照处理后采用本文方法
rank-1/%	76.9	76.3	77	77.8
mAP/%	52.1	50.2	54.4	55.1

由表 2 可知, 经过光照处理后, 本文基于改进的 ResNet 行人再识别模型的 rank-1 和 mAP 分别比未经过光照处理的实验高 0.8% 和 0.7%, 说明本文方法对光照变化复杂的场景具有一定的作用。

3.3.2 复杂光照条件下的综合对比分析

为综合对比本文基于改进的 ResNet 行人再识别模型对光照干扰和角度变化的处理效果, 采用 MGN、ResNet-50+CircleLoss 两种常用的行人再识别方法进行对比分析, 测试结果如表 3 所示。

表 3 2 种方法测试结果

方法	MGN	ResNet-50+CircleLoss	本文方法
rank-1/%	76.9	76.3	79.8
mAP/%	52.1	50.2	58.5

由表 3 可知: 经过光照、异常姿态处理后的数据

集作为基于改进的 ResNet 行人再识别模型的测试数据集, rank-1 分别高于 MGN 和 ResNet-50+CircleLoss 方法 2.9% 和 3.5%; mAP 分别高于 MGN 和 ResNet-50+CircleLoss 方法 6.4% 和 8.3%。

由此可以得出结论, 本文方法对于光照干扰和角度变化下的复杂场景具有一定的改善效果。

4 结论

本文提出基于改进的 ResNet 行人再识别模型, 在 ResNet 的浅层嵌入 IBN 和 STN, 提升网络的外观变化适应性和空间不变性; 采用改进的 Mosaic 数据增强方法模拟复杂背景, 增加数据集的多样性; 同时使用 Mish 激活函数替换 ReLU 激活函数。在 MSMT17 数据集上 rank-1 和 mAP 分别达到 79.8% 和 58.5%, 验证了算法的有效性。

参考文献

- [1] 雷欢, 吴亮生, 焦泽昱, 等. 果园环境中成熟杨梅自动检测方法研究[J]. 自动化与信息工程, 2021, 42(3): 9-14, 26.
- [2] HUANG K, LEI H, JIAO Z, et al. Recycling Waste Classification Using Vision Transformer on Portable Device[J]. Sustainability, 2021, 13(21): 11572.
- [3] 游青山, 冉霞. 基于机器视觉的矿井作业人员行为监测及违章识别系统[J]. 自动化与信息工程, 2021, 42(4): 20-24.

- [4] 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, 2010:2360-2367.
- [5] PEDAGADI S, ORWELL J, VELASTIN S, et al. Local fisher discriminant analysis for pedestrian re-identification [C]//2013 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2013.
- [6] 贾熹滨,鲁臣,Siluyele Ntazana,等.行人再识别中的多尺度特征融合网络[J].北京工业大学学报,2020,46(7):788-794.
- [7] 潘少明,王玉杰,种衍文.基于图卷积神经网络的跨域行人再识别[J].华中科技大学学报(自然科学版),2020,48(9):44-49.
- [8] 陈巧媛,陈莹.基于困难样本三元组损失的多任务行人再识别[J].计算机辅助设计与图形学学报,2019,31(7):1156-1165.
- [9] ULYANOV D, VEDALDI A, LEMPITSKY V. Instance normalization: The missing ingredient for fast stylization[J]. arXiv preprint arXiv:1607.08022, 2016.
- [10] JADERBERG M, SIMONYAN K, ZISSERMAN A. Spatial transformer networks[J]. Advances in neural information processing systems, 2015,28:2017-2025.
- [11] WEI L, ZHANG S, GAO W, et al. Person transfer GAN to bridge domain gap for person re-identification[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2018:79-88.
- [12] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition, 2016:770-778.

A Pedestrian Re-identification Optimization Method for Complex Scenes

Wang Nan Wu Jian Lu Xingjian Zhong Zhenyu Huang Kai

(Institute of Intelligent Manufacturing, Guangdong Academy of Science/
Guangdong Key Laboratory of Modern Control Technology, Guangzhou 510070, China)

Abstract: Aiming at the problem that it is difficult to recognize pedestrians in outdoor scenes due to different camera angles and changeable outdoor lighting, an optimization method for pedestrian re-identification in complex scenes is proposed. By embedding the instance-batch normalization layer and spatial transformer network in the shallow layer of ResNet, introducing the mish activation function, building an improved ResNet-50 feature extraction backbone network, and establishing a pedestrian re-identification model for complex environment. Verified on MSMT17 data set, rank-1 and map reach 79.8% and 58.5% respectively, which shows the effectiveness of the algorithm.

Key words: complex scene; pedestrian re-identification; instance-batch normalization; spatial transformer network; Mish activation function

作者简介:

王楠,女,1989年生,硕士研究生,高级工程师,主要研究方向:机器视觉、深度学习。E-mail: n.wang@giim.ac.cn
黄凯(通信作者),男,1995年生,硕士研究生,主要研究方向:机器视觉、人工智能。E-mail: 2111904371@mail2.gdut.edu.cn

(上接第17页)

作者简介:

杨志明,男,1974年生,本科,工程师,主要研究方向:锂离子电池极片装备制造。
黄天仑(通信作者),男,1991年生,博士,主要研究方向:机器视觉、深度学习。E-mail: tl.huang@giim.ac.cn
谭鹏辉,男,1995年生,博士,主要研究方向:锂离子电池极片制造仿真。
钟震宇,男,1971年生,博士,研究员,主要研究方向:深度学习。
张云,男,1981年生,博士,教授,主要研究方向:锂离子电池制造、智能制造。