

DOI:10.16356/j.1005-2615.2019.05.012

基于视频的行人再识别

胡彬^{1,2} 杨铖¹ 邵叶秦³ 杨赛⁴

(1.南通大学信息科学技术学院,南通,226019; 2.南通先进通信技术研究院,南通,226019; 3.南通大学交通学院,南通,226019; 4.南通大学电气工程学院,南通,226019)

摘要: 行人再识别是指在无交叉区域的多摄像机视频监控系统中,匹配不同摄像机中的相同行人目标。本文提出了一种基于视频的行人再识别方法,用 HOG3D 来描述一组视频的时空特征,在训练集上用预训练的 DenseNet 来微调模型参数,利用迁移学习得到的模型来提取视频中行人的表观特征,融合两种特征来描述视频序列中的行人。最后将融合的高维特征降维,并用度量学习方法计算行人对之间的距离。本文在 PRID 2011 和 iLIDS-VID 这两个视频数据集上进行了使用,实验结果表明本文的方法取得了较高的累积匹配得分。

关键词: 行人再识别; 时空特征; 迁移学习; 度量学习

中图分类号: TP391 **文献标志码:** A **文章编号:** 1005-2615(2019)05-0669-06

Video-Based Person Re-identification

HU Bin^{1,2}, YANG Cheng¹, SHAO Yeqin³, YANG Sai⁴

(1. School of Information Science and Technology, Nantong University, Nantong, 226019, China; 2. Nantong Research Institute for Advanced Communication Technologies, Nantong, 226019, China; 3. School of Transportation, Nantong University, Nantong, 226019, China; 4. School of Electrical Engineering, Nantong University, Nantong, 226019, China)

Abstract: The task of person re-identification is to match pedestrian images observed from different cameras in a non-overlapping multi-camera surveillance systems. In this article, a video-based person re-identification method is proposed. HOG3D is extracted as temporal and spatial feature and the DenseNet model pre-trained on MSCOCO is adopted to fine-tune the parameters for person re-identification, and the fine-tuned model is used to extract the feature from person image. The two features are combined to describe the person video clip. Finally the metric learning model is applied to measure distance between person pairs. We evaluate our approach by operating in-depth experiments in two video-based benchmarks, and the experimental results on the two benchmark show significant and consistent improvements over the state-of-the-art methods.

Key words: person re-identification; spatial-temporal feature; transfer learning; metric learning

行人再识别指从无重叠的摄像头网络中检索出某个行人,该应用及其相关技术在智能安防领域有着巨大的需求,如疑犯检索,利用这项技术自动从海量视频中检索疑犯,可以节省公安部门大量的人力物力。目前研究以基于静态图像为主,即数据集中一个人最多有两张图片,然后给定行人的一张

图片,从样本集合中检索出该行人^[1-4],这种基于单帧图像的方法存在较大的难度。首先,不同的摄像头存在角度、光线和摄像头参数等多方面因素的不同,导致同一个人可能存在比较大的差异;其次,给定待检索的单帧行人图像没有被遮挡,但是搜索集中的图像可能存在被遮挡的情况,这样会导致匹配

基金项目: 江苏省教育厅自然科学基金(16KJB520037)资助项目;江苏省社会安全图像与视频理解重点实验室创新基金(30916014107)资助项目。

收稿日期: 2018-07-06; **修订日期:** 2018-09-10

通信作者: 胡彬,男,讲师,博士,E-mail:nj-chris@126.com。

引用格式: 胡彬,杨铖,邵叶秦,等. 基于视频的行人再识别[J]. 南京航空航天大学学报,2019,51(5):669-674. HU Bin, YANG Cheng, SHAO Yeqin. Video-Based Person Re-identification[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2019, 51(5): 669-674.

的准确性降低。

由于存在光线、角度和遮挡等问题,从单帧图像中难以提取足够的信息去区分不同的人。实际的监控场景中并不是只有单帧图像,而是有大量的视频资源,视频显然比静态图像更能区分不同的行人。首先,视频图像序列中包含了时空信息,因此可以提取出行人的动作特征;其次从视频中提取的行人外观特征(如颜色直方图)比单帧图像更加稳定,尤其对于遮挡的敏感性更低,因此本文提出了一种基于视频序列的行人再识别方法。由于维梯度方向直方图(Histograms of 3D gradient orientations, HOG3D)^[5]能够较好地描述视频中行人时空特征,因此本文用HOG3D来表示视频序列中行人的时空特征。深度卷积神经网络(Deep convolutional neural networks, DCNN)在图像识别领域取得了显著成果,因此本文通过DCNN提取视频序列中的每一帧图像的特征,然后将视频中的序列图像做平均池化后的特征作为行人的外观特征。该段视频最终由时空特征和外观特征组成,融合的特征通过主成分分析(Principal component analysis, PCA)白化(White PCA, WPCA)进行数据降维,WPCA比PCA更能消除数据的冗余性,最后利用度量学习来度量行人直接的特征距离。

1 相关研究

目前的研究主要集中在两个方面:(1)从静态图像中提取具备区分力的可靠的行人特征以弥补不同摄像头带来的光线、视角和遮挡等问题,(2)学习鲁棒的度量学习算法提高相同样本的匹配率。多数的监控环境都检测不到行人的面部,因此人脸这种生物特征无法使用,颜色和纹理特征是使用最为广泛的特征之一,但是如何设计一个可靠有效的特征需要具备大量的经验。Gray等融合了8种颜色空间的颜色特征和19种纹理特征来描述行人^[6];Farenzena等利用了人体的对称性提出了加权的颜色直方图和结构化的纹理特征来描述行人^[7];局部最大表示(Local maximal occurrence, LOMO)特征^[8]首先采用多尺度Retinex算法对图像进行预处理,然后在色调(H),饱和度(S),明度(V)(Hue, saturation, value, HSV)空间提取颜色直方图,结合尺度不变局部特征(Scale invariant local patterns, SILTP)纹理特征,取得了良好的效果。大多数研究都是融合密集颜色特征和各种纹理特征以描述行人^[8-9]。随着DCNN在图像识别等领域取

得了巨大成功,很多研究者尝试利用DCNN来提取行人特征^[3,10],并取得了较好的结果。但是训练一个成功的网络模型需要海量样本,研究者通常在已有的Imagenet数据集上训练好模型进行参数微调,以获得适合行人再识别的网络模型。总之,研究者利用了多种特征提取方法及融合方法来描述行人,由于融合后的特征维度比较高,多数研究者引入PCA^[2],WPCA等方法以降低特征维度。还有研究者在非线性空间引入Kernel PCA^[11]处理行人特征,识别率较线性空间有较高的提高。

度量学习近几年在解决行人再识别问题中发挥了巨大的作用,大量研究者将度量学习引入到再识别任务中,其中马氏距离是使用最为广泛的方法之一,如式(1)所示, x_i 和 x_j 分别表示两个特征向量, $D_M(x_i, x_j)$ 表示两者之间的距离。度量学习的目的即是要找到投影矩阵 P , $M = P^T P$,使 x_i 和 x_j 属于同类时, D_M 则小,不同类时则 D_M 大。

$$D_M(x_i, x_j) = (x_i - x_j)^T M (x_i - x_j) \quad (1)$$

研究者利用特征样本对,采用机器学习的方法计算 P ,使样本的类间距变大,类内距变小^[12]。局部自适应决策函数(Learning locally-adaptive decision functions, LADF)方法^[13]结合了距离度量方法和样本对之间的局部自适应约束。简单的度量学习(Keep it simple and straight forward metric learning, KISSME)算法^[9]是一种基于高斯分布模型的简单有效的度量方法,取得了较好的效果。跨视图二次鉴别分析(Cross-view quadratic discriminant analysis, XQDA)^[8]在KISSME基础上通过学习更具区分性的距离函数,进一步提高了效果。还有研究者通过构建三元组样本 (x_i^a, x_i^b, x_i^n) 来设计度量算法^[14],其中 x_i^a 和 x_i^b 是同类, x_i^a 和 x_i^n 为异类,在行人再识别应用中,前者表示两者为同一个人,后者表示两者为不同行人。通过度量学习,让 x_i^a 和 x_i^b 之间的距离尽可能小, x_i^a 和 x_i^n 之间的距离尽可能大,虽然取得了更好的效果,但是构建三元组数据集难度要大很多。

此外,还有研究者利用排序算法如RankingSVM解决行人再识别问题。文献[15]通过boost级联算法,得到一个强RankingSVM排序算法。Zhang等研究者以动态权值参数取代固定参数,得到更有效的排序算法^[16]。综上所述,多数的研究工作都是基于静态图像的,基于视频序列的^[17-19]研究较少。

2 基于视频序列的行人再识别

2.1 特征提取

本文利用 HOG3D 特征^[5]来描述视频序列的时空特征。首先计算三维空间中 SIFT 兴趣点坐标点 (x, y, t) , 然后在该坐标点周围建立立方体, 并将该立方体分割成若干个 cell, 每个 cell 再细分成 8 个 block。对每个 block 里所有点的梯度均值进行统计, 然后以正十二面体作为量化的多面体, 得到每个 block 的特征描述。最后, 对一个 cell 里所有 block 的特征取平均得到每个 cell 的特征描述, 将所有 cell 的特征描述子按次序拼接就得到了 HOG3D 特征。

除了视频的时空特征, 本文还提取了行人的表观特征, 大量研究表明利用 DCNN 提取的图像特征比手工特征更具备优势, 因此本文结合 DCNN 和迁移学习来提取行人的表观特征。由于行人再识别数据集样本数量较少, 难以直接用于训练得到网络参数, 本文利用迁移学习, 将在 MSCOCO 上训练模型参数迁移至行人再识别样本集, 以提高模型分类的准确率。

本文以 DenseNet^[20]作为基础网络模型, 以在 MSCOCO 数据集上训练的参数作为初始参数, 由于 MSCOCO 数据集与行人再识别数据集的差异, 首先需要对网络结构进行细微修改, 然后在行人再识别数据集上进行训练微调参数。DenseNet 缩短了网络前后层之间的连接, 以消除由于 DCNN 层数加深引起的梯度消失问题。对于第 N 层网络, 其输入既包括了原始的输入信息 X_0 , 也包括了经前 $N-1$ 层处理过的 X_0 。当反向传播计算时, X_0 的梯度信息中包含了损失函数对 X_0 的导数部分, 因此有利于梯度传播。借鉴了 GoogleNet^[21], DenseNet 将各层进行了串联, 有

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]) \quad (2)$$

式中 $H_l(\cdot)$ 为 DenseBlock 模块, 其中包含了批量归一化 (Batch normalization, BN)^[22]和线性整流函数和 3×3 卷积这 3 个操作。DenseBlock 模块之间通过 BN, 1×1 卷积和均值池化连接, 这样可以取得更低的错误率, 使用了更少的参数。

2.2 度量算法

针对行人再识别问题, 即给定一张行人图像或者行人视频序列, 要求从样本库中检索出该行人, 对于度量算法而言, 则要求同一个行人之间的距离尽可能小, 而不同行人之间的距离尽可能大。假设二元组 (x_i, x_j) 分别表示同一个人不同图像的特征描述子, 记为正样本, $D(x_i, x_j)$ 表示两个特征之间的距离, (x_i, x_k) 表示来自不同人的特征, 记为负样本,

$D(x_i, x_k)$ 表示它们之间的距离, 要使任意正样本对之间距离小于所有负样本之间的距离, 可以用如下公式表示

$$D(x_i, x_j) + \rho < \min_{y_i \neq y_j} D(x_i, x_k) \quad (3)$$

式中 ρ 表示松弛变量, 根据式 (2), 得到目标函数如下

$$\min \max \{ D(x_i, x_j) - \min_{y_i \neq y_j} D(x_i, x_k) + \rho, 0 \} \quad (4)$$

但是上述目标函数仅考虑了类间分布, 没有考虑类内分布。因此, 要同时使正样本之间的距离尽可能小, 可以得到如下的目标函数

$$\min \sum_{x_i, x_j, y_i = y_j} D(x_i, x_j) \quad (5)$$

综合考虑类间和类内距离, 得到最终的目标函数如下

$$F(D) = \sum_{x_i, x_j, y_i = y_j} \max \left\{ \begin{array}{l} D(x_i, x_j) - \\ \min_{y_i \neq y_j} D(x_i, x_k) + \rho, 0 \end{array} \right\} + \sum_{x_i, x_j, y_i = y_j} D(x_i, x_j) \quad (6)$$

式中 $D(x_i, x_j)$ 可以使用式 (1) 的马氏距离, 即

$$D(x_i, x_j) = (x_i - x_j)^T M (x_i - x_j) \quad (7)$$

式中 M 是半正定矩阵, 为了求解 M , 定义

$$X_{i,j} = (x_i - x_j)(x_i - x_j)^T \quad (8)$$

$D(x_i, x_j)$ 则可以表示为

$$D(x_i, x_j) = \text{tr}(M X_{i,j}) \quad (9)$$

目标函数 $F(D)$ 则可以改写成关于 M 的函数

$$F(M) = \sum_{x_i, x_j, y_i = y_j} \max \left\{ \begin{array}{l} \text{tr}(M X_{i,j}) - \\ \min_{y_i \neq y_j} \text{tr}(M X_{i,k}) + \rho, 0 \end{array} \right\} + \sum_{x_i, x_j, y_i = y_j} \text{tr}(M X_{i,j}) \quad (10)$$

通过梯度下降法即可求解出 M , 对于给定的行人视频序列特征 x_p , 如何从另一个摄像头中的若干组行人视频序列特征集合 $\{x_g\}$ 中检索出 x_p , 只需要计算

$$D(x_p, x_g) = (x_p - x_g)^T M (x_p - x_g) \quad (11)$$

然后从集合 $\{x_g\}$ 中找出与 x_p 距离最小的。

3 实验分析

3.1 特征提取

给定行人视频序列, 首先按照文献[5]方法提取 HOG3D 特征, 得到 1 200 维特征描述算子, 然后提取 DCNN 特征。

本文选择 PyTorch 作为深度学习平台, 该平台是 Facebook 在 Torch 基础上针对 python 语言发布的一个机器学习工具包。以 DenseNet121 作为基础网络架构, 以在 MSCOCO 数据集上训练的参数作为初始参数。为了防止训练数据量少造成过拟

合,通过数据增强算法来增加训练样本数量。训练时,采用交叉熵代价函数作为损失函数,并使用带momentum项的梯度下降算法作为优化算法,每次训练最多迭代50次,直到找到最优模型,图1给出了一次训练中loss的变化曲线,可以看出,使用预训练的模型可以使损失函数快速下降。训练完成后,去掉最后一层softmax分类层,输入一张样本图片,输出1 024维特征。对于一组行人序列样本,每张图片都得到1 024维特征,采用 1×1 的平均池化处理这些特征,最终得到该序列的特征,维度是1 024。结合1 200维的HOG3D特征,将两种特征级联最终得到2 224维的特征向量。

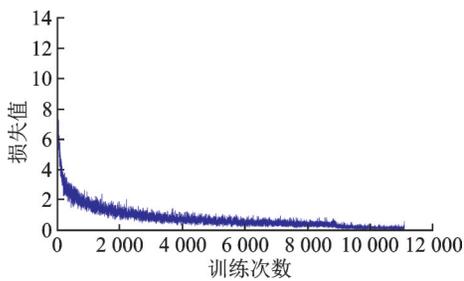


图1 采用预训练模型训练时loss变化曲线

Fig. 1 Loss curve of training with pre-trained model

3.2 评价准则和实验设计

与多数行人再识别算法一致,本文采用累积匹配特性(Cumulative match characteristic, CMC)曲线评价算法性能,即给定查询集和行人图像库,本文主要比较前20的匹配率。将数据集分成训练集和测试集,样本大小统一缩放为256像素 \times 128像素。从训练集提取出每组样本的HOG3D特征和深度特征,通过WPCA降维到110维,根据2.2节所述的度量算法学习出 M 。测试集中,则根据 $D(x_i, x_j) = (x_i - x_j)^T M (x_i - x_j)$ 计算出每对样本直接的距离,按照 D 的大小排序,得到第一匹配率。随机调整训练集和测试集,一共生成10组,取10组匹配率的平均值最为最终结果。

3.3 结果分析

本文在PRID 2011^[23]和iLIDS-VID^[18]这两个视频序列行人再识别数据集上展开了深入的实验。PRID 2011数据集采集自两个固定的室外摄像头,其背景较为干净单一并且遮挡较少。其中一个摄像头中有385个行人,另一个摄像头包含749人,所有人中,有200人在两个摄像头中都出现了。每个视频序列包含的图像帧数为5~675帧,本文挑选了大于27帧的视频一共178人作为本文的实验对象,其中89人为训练集,另89人作为测试集。iLIDS-VID数据集采集自机场大厅,包含300个行人,每个行人有两组来自不同摄像头的视频序列,

即一共有600组视频序列,每组视频序列包含图像帧数从23帧至192帧不等,因此600组视频序列都被用来作为实验对象,训练集和测试集各150人。iLIDS-VID数据集中样本图片的背景相对于PRID 2011更加复杂,也存在遮挡等干扰。

本文与几个基于视频序列的行人再识别方法进行了比较,包括TDL^[17],SDALF^[7],Saliency^[24],RPRF^[14],Sparse re-id^[19],DVDL^[25],DVR^[18]和RNN^[26]。表1和表2给出了两个数据集上的实验结果。

表1 PRID 2011数据集实验结果比较

方法	Rank 1	Rank 5	Rank 10	Rank 20
TDL	56.7	80.0	87.6	93.6
SDALF	5.2	20.7	32.0	47.9
Saliency	25.8	43.6	52.6	62.0
RPRF	19.3	38.4	51.6	68.1
Sparse re-id	35.1	59.4	69.8	79.7
DVDL	40.6	69.7	77.8	85.6
DVR	37.6	63.9	75.3	88.3
RNN	79.4	94.4	99.3	99.3
本文方法	79.6	85.7	92.1	96.3

表2 iLIDS-VID数据集实验结果比较

方法	Rank 1	Rank 5	Rank 10	Rank 20
TDL	56.3	87.6	95.6	98.3
SDALF	6.3	18.8	27.1	37.3
Saliency	10.2	24.8	35.5	52.9
RPRF	14.5	29.8	40.7	58.1
Sparse re-id	24.9	44.5	55.6	66.2
DVDL	25.9	48.2	57.3	68.9
DVR	34.5	56.7	67.5	77.5
RNN ^[26]	55.2	86.5	97.0	97.0
本文方法	62.6	88.5	96.2	98.9

几种方法中,TDL方法结构与本文方法最为接近,该方法同样融合了时空特征和表观特征,其中时空特征提取方法与本文相同,而表观特征则是采用了颜色特征和LBP纹理特征这类手工特征。众多研究证明深度学习提取的特征要明显优于手工特征,因此本文算法在两个数据集上较TDL均有大幅度提高。

RNN方法则是构建一个端到端的深度循环神经网络,实现了视频中行人时空特征提取和度量学习,取得了非常好的效果。本文方法在PRID 2011数据集上近Rank 1与其接近,Rank 5和Rank 20都落后于RNN。但是在iLIDS-VID数据集上,由于样本数量有大幅提高,卷积神经网络在图像识别的

优势得到了体现,取得了较RNN较好的识别率。

比较的几种方法中,本文方法和RNN方法都采用了深度学习,其他方法都是基于传统的手工特征提取方法,因此本文的方法和RNN的结果较其他方法有明显的优势。在iLIDS-VID数据集上,由于样本数量的提高,带来的准确率提升则更为明显,这也证实了由神经网络提取的特征比手工特征更具有区分性,缺点则是需要较多的训练样本。

4 结 论

本文提出了一种基于视频的行人再识别方法,针对一段行人视频,提取HOG3D特征表示其时空特征。在训练集上用预训练的DenseNet来微调模型参数,对于视频中的每一帧图像,利用学习的模型来提取表观特征,将平均池化处理后的特征作为视频中行人的外观特征。该段视频最终由时空特征和外观特征组成,度量学习同时要求:(1)任意正样本对之间距离小于所有负样本之间的距离;(2)正样本之间的距离最小化。在PRID 2011和iLIDS-VID这两个视频数据集上取得了较好的实验结果,说明了本文方法的有效性,也证实了由深度卷积神经网络提取的特征比手工特征更具有区分性,因此接下来的工作考虑将HOG3D特征也采用深度学习自动提取,如采用CNN-LSTM融合网络自动提取时空特征。

参考文献:

- [1] BEDAGKAR-GALA A, SHAH S K. A survey of approaches and trends in person re-identification [J]. *Image & Vision Computing*, 2014, 32(4): 270-286.
- [2] SUN C, WANG D, LU H. Person re-identification via distance metric learning with latent variables [J]. *IEEE Transactions on Image Processing*, 2017, 26(1): 23-34.
- [3] SU C, ZHANG S, XING J, et al. Multi-type attributes driven multi-camera person re-identification [J]. *Pattern Recognition*, 2018, 75: 77-89.
- [4] MATSUKAWA T, OKABE T, SUZUKI E, et al. Hierarchical Gaussian descriptor for person re-identification [C]//*IEEE Conference on Computer Vision & Pattern Recognition*. Washington, DC: IEEE Computer Society Press, 2016: 1363-1372.
- [5] KLÄSER A, MARSZALEK M, SCHMID C. A spatio-temporal descriptor based on 3D-Gradients [C]//*British Machine Vision Conference*. Cambridge, UK: British Machine Vision Association, 2008: 1-10.
- [6] GRAY D, TAO H. Viewpoint invariant pedestrian recognition with an ensemble of localized features [C]//*European Conference on Computer Vision*. Berlin: Springer, 2008: 262-275.
- [7] FARENZENA M, BAZZANI L, PERINA A, et al. Person re-identification by symmetry-driven accumulation of local features [C]// *IEEE Conference on Computer Vision & Pattern Recognition*. Washington, DC: IEEE Computer Society Press, 2010: 2360-2367.
- [8] LIAO S, HU Y, ZHU X, et al. Person re-identification by local maximal occurrence representation and metric learning [C]//*IEEE Conference on Computer Vision & Pattern Recognition*. Washington, DC: IEEE Computer Society Press, 2015: 2197-2206.
- [9] KOESTINGER M, HIRZER M, WOHLHART P, et al. Large scale metric learning from equivalence constraints [C]//*IEEE Conference on Computer Vision & Pattern Recognition*. Washington, DC: IEEE Computer Society Press, 2012: 2288-2295.
- [10] LI W, ZHAO R, XIAO T, et al. DeepReID: Deep filter pairing neural network for person re-identification [C]//*IEEE Conference on Computer Vision & Pattern Recognition*. Washington, DC: IEEE Computer Society Press, 2014: 152-159.
- [11] XIONG F, GOU M, CAMPS O, et al. Person re-identification using kernel-based metric learning methods [C]//*European Conference on Computer Vision*. Berlin: Springer, 2014: 1-16.
- [12] PEDAGADI S, ORWELL J, VELASTIN S, et al. Local Fisher discriminant analysis for pedestrian re-identification [C]//*IEEE Conference on Computer Vision & Pattern Recognition*. Washington, DC: IEEE Computer Society Press, 2013: 3318-3325.
- [13] LI Z, CHANG S, LIANG F, et al. Learning locally-adaptive decision functions for person verification [C]//*IEEE Conference on Computer Vision & Pattern Recognition*. Washington, DC: IEEE Computer Society Press, 2013: 3610-3617.
- [14] CHENG D, GONG Y, ZHOU S, et al. Person re-identification by multi-channel parts-based CNN with improved triplet loss function [C]//*IEEE Conference on Computer Vision & Pattern Recognition*. Washington, DC: IEEE Computer Society Press, 2016: 1335-1344.
- [15] ENGEL C, BAUMGARTNER P, HOLZMANN M, et al. Person re-identification by support vector ranking [C]//*British Machine Vision Conference*. Cambridge, UK: British Machine Vision Association, 2010: 1-11.

- [16] ZHANG Y, LI B, LU H, et al. Sample-specific SVM learning for person re-identification [C]//IEEE Conference on Computer Vision & Pattern Recognition. Washington, DC: IEEE Computer Society Press, 2016: 1278-1287.
- [17] YOU J, WU A, LI X, et al. Top-push video-based person re-identification [C]//IEEE Conference on Computer Vision & Pattern Recognition. Washington, DC: IEEE Computer Society Press, 2016: 1345-1353.
- [18] WANG T, GONG S, ZHU X, et al. Person re-identification by video ranking [C]//European Conference on Computer Vision. Berlin: Springer, 2014: 688-703.
- [19] KARANAM S, LI Y, RADKE R J. Sparse re-id: Block sparsity for person re-identification [C]//IEEE Conference on Computer Vision & Pattern Recognition. Washington, DC: IEEE Computer Society Press, 2015: 33-40.
- [20] HUANG G, LIU Z, WEINBERGER K Q. Densely connected convolutional networks [C]//IEEE Conference on Computer Vision & Pattern Recognition. Washington, DC: IEEE Computer Society Press, 2016: 2261-2269.
- [21] ZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision [C]// IEEE Conference on Computer Vision & Pattern Recognition. Washington, D C: IEEE Computer Society Press, 2016: 2818-2826.
- [22] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [C]//International Conference on Machine Learning. New York: ACM Press, 2015: 448-456.
- [23] HIRZER M, BELEZNAI C, ROTH P M, et al. Person re-identification by descriptive and discriminative classification [J]. Lecture Notes in Computer Science, 2011, 6688(12): 91-102.
- [24] ZHAO R, OUYANG W, WANG X. Unsupervised salience learning for person re-identification [C]//IEEE Conference on Computer Vision & Pattern Recognition. Washington, DC: IEEE Computer Society Press, 2013: 3586-3593.
- [25] KARANAM S, LI Y, RADKE R J. Person re-identification with discriminatively trained viewpoint invariant dictionaries [C]//IEEE Conference on Computer Vision & Pattern Recognition. Washington, DC: IEEE Computer Society Press, 2015: 4516-4524.
- [26] ZHOU Z, HUANG Y, WANG W, et al. See the forest for the trees: Joint spatial and temporal recurrent neural networks for video-based person re-identification [C]//IEEE Conference on Computer Vision & Pattern Recognition. Washington, DC: IEEE Computer Society Press, 2017: 6776-6785.

(编辑:刘彦东)