基于双流卷积神经网络和生成式对抗网络的行人重识别算法*

林 通,陈 新,唐 晓,贺 玲,李 浩

(中国人民解放军空军预警学院,湖北 武汉 430019)

摘 要:近年来,针对行人重识别问题的深度学习技术研究取得了很大的进展。然而,在解决实际数据的特征样本不平衡问题时,效果仍然不理想。为了解决这一问题,设计了一个更有效的模型,该模型很好地解决了目标的不同姿态的干扰以及数据集中的图片数量不足的问题。首先,通过迁移姿态生成对抗网络生成行人不同姿势的图片,解决姿态干扰及图片数量不足的问题。然后利用两种不同的独立卷积神经网络提取图像特征,并将其结合得到综合特征。最后,利用提取的特征完成行人重识别。采用姿势转换方法对数据集进行扩展,有效地克服了由目标不同姿势引起的识别误差,识别错误率降低了6%。实验结果表明,该模型在Market-1501和DukeMTMC-Reid上达到了更好的识别准确度。在DukeMTMC-Reid数据集上测试时,Rank-1准确度增加到92.10%,mAP达到84.60%。

关键词:行人重识别;卷积神经网络;生成式对抗网络;姿势迁移

中图分类号:TP18 **文献标识码**:A **DOI**:10.19358/j.issn.2096-5133.2020.06.002

引用格式:林通,陈新,唐晓,等.基于双流卷积神经网络和生成式对抗网络的行人重识别算法[J].信息技术与网络安全,2020,39(6):7-12.

Dual stream ConvNet-Gan for person re-identification

Lin Tong, Chen Xin, Tang Xiao, He Ling, Li Hao

(People's Liberation Army Force Army Early Warning Academy, Wuhan 430019, China)

Abstract: The performance of current algorithm of person re-identification(ReID) has been made great progress, but there are still a lot of difficults in dealing with different backgrounds and poses. In order to cope with the difficult in actual scenes, this paper proposes a novel model, which can very well solve the problem that the characters have different poses and the number of pictures in the dataset is insufficient and uneven. Firstly the IDs with a relatively small number of images are screened out to expand by transferring pose and solve the problem of imbalanced datasets. Then two different convolutional neural networks are used to extract picture features and they are joined to get different characteristics. In the end, the task with extracted features is completed. This model solves the impact of different poses on the recognition effect very well. At the same time, it uses two independent feature extraction networks to extract the features of the person more comprehensively. Experimental results confirm that this model significantly improves performance and achieves high accuracy in Market-1501 and DukeMTMC-Reid. When tested on the DukeMTMC-Reid dataset, Rank-1 was increased to 92.10%, and mAP was increased to 84.60%.

Key words: person ReID; CNN; generative adversarial network; pose transfer

0 引言

行人重识别(ReID)的目的是利用计算机视觉技术判断图像或者视频序列中是否存在特定行人的技术。由于其在安全和监控方面的重要应用,受到了学术界和工业界的广泛关注。这一任务极具挑战性,因为不同相机拍摄的图像往往包含由背景、视角、人体姿势等变化引起的显著变化。

在过去的几十年中,大多数现有的研究都集中在度量学习和特征学习上,设计尽可能抵消类内变化的算法已经成为了行人重识别的主要目标之一。

*基金项目:国家自然科学基金(61502522)

起初,行人重识别的研究主要是基于全局特征,即基于全局图像获得特征向量。为获得更好的效果,开始引入局部特征,常用的局部特征提取方法包括图像分割、骨架关键点定位和姿态校正。行人重识别任务还面临一个非常大的问题,即很难获得数据。到目前为止,最大的数据集只有几千个目标和上万张图片。随着生成性对抗网络(GAN)技术的发展,越来越多的学者试图利用这种方法来扩展数据集。然而,大多数现有的针对行人重识别的算法需要大量标记数据,这限制了它在实际应用场景中的鲁棒性和可用性,因为手动标记一个大型数据集是昂贵和困难的。最近的一些文献在使用无监督学习来解决此问题,改进人工标注的特征。由于不同数据集之间的图像有明显差异,在提取图片特征问题上效果仍不理想。本文方法的贡献主要有两个方面。一方面,采用均衡采样策略和姿势迁移方法对数据集进行扩充有效地缓解了行人姿势各异造成的干扰。另一方面,利用两种不同的网络结构从图像的不同方面提取特征,使模型能学习到更为丰富全面的信息。实验结果表明,该方法具有较高的精度。例如,在 Market-1501 数据集上,这一模型 Rank-1 准确度达到了 96.0%,mAP 达到了 90.1%,与现有算法相比,性能有较大提升。

1 相关理论方法

1.1 局部特征

行人重识别的研究最开始主要是基于全局特征,即基于全局图像获得特征方向。为了得到更好的性能,研究人员逐步研究局部特征。常用的局部特征提取方法包括图像分割、骨架关键点定位和姿态校正。图像分割是一种非常常用的提取局部特征的方法。将图片垂直分成几个部分[1-2]。然后,将分割后的图像块按顺序送入神经网络。最后一个特征融合了所有图像块的局部特征。然而,这种方法的缺点是对图像对齐的要求相对较高。如果两幅图像没有对齐,很可能会造成头部和上身的对比,从而导致模型判断错误。为了解决图像错位问题,一些文献首先利用先验知识对行人进行对齐。这些先验知识主要是经过训练的人体姿势(骨架)模型和骨架关键点(骨架)模型。例如,文献[3]首先利用姿态估计模型对行人的关键点进行估计,然后利用仿射变换对相同的关键点进行对齐。

1.2 基于生成式对抗网络的数据扩充

行人重识别任务面临一个非常严重的问题,那就是很难获得数据。到目前为止,最大的行人重识别数据集只有几千个行人和上万张图片。随着生成式对抗网络(GAN)技术的发展,越来越多的学者试图利用这种方法来扩展数据集。

文献[4]是第一篇以基于生成式对抗网络技术解决行人重识别的论文,发表在 ICCV17 会议上。虽然论文比较简单,但却引出了一系列的佳作。行人重识别的一个问题是不同的相机提供的照片存在差异。这种差异可能来自各种因素,如光线和角度。为了解决这个问题,文献[5]使用生成式对抗网络将图像从一个相机风格迁移到另一个相机。除了相机的差异外,行人重识别面临的另一个问题是不同数据集中存在差异。这很大程度上是由图片拍摄环境造成的。为了克服这种环境变化的影响,文献[6]使用 GAN 将行人从一个数据集的拍摄环境迁移到另一个数据集拍摄环境。行人重识别还要解决的困难之一是行人姿势的不同。为了应对这一问题,文献[7]用生成式对抗网络制作了一系列标准的姿态图,共提取了 8 个姿态,基本涵盖了各个角度。对每个行人生成一组标准的 8 个姿势图片,以解决不同的姿势问题。

2 基于卷积神经网络和生成式对抗网络的行人重识别模型

在这部分,提出了一种新的模型来解决 ReID 问题,如图 1 所示。该模型包含两个部分:基于生成式对抗网络的模型和一个集成 CNN 模型。该模型先将已有的数据集进行扩充,再利用扩充的数据集训练集成模型完成行人重识别任务。

2.1 基于姿势迁移的数据集扩充模型

针对行人重识别任务中,获取数据困难,数据集不均衡,样本图片质量不一的问题,采用生成式对抗网络算法来扩充数据集。使用基于一个比较成功的 GAN 模型[8]的改进方法,该模型可以将图片中人物从一个姿势转移到另一个姿势。$\{P_i^j\}_{i=1\cdots N_j}^{j=1\cdots M}$ 表示输入的图片,j 是人物标签,i 是图片标签,M 是人物的数量,N_j 是第 j 个人的图片数量。模型先检测出输入图片的 18 个关键点,对应人的 18 个关节位置,并将之编码为 S_i^j。P_c 作为姿势条件,P_i 作为待扩充的图片,输入 (P_c, P_i) 及它们的关键点向量 (S_c, S_t),生成器生成一张 t 的图片,其姿势是 c。由鉴别器判断其真实性。

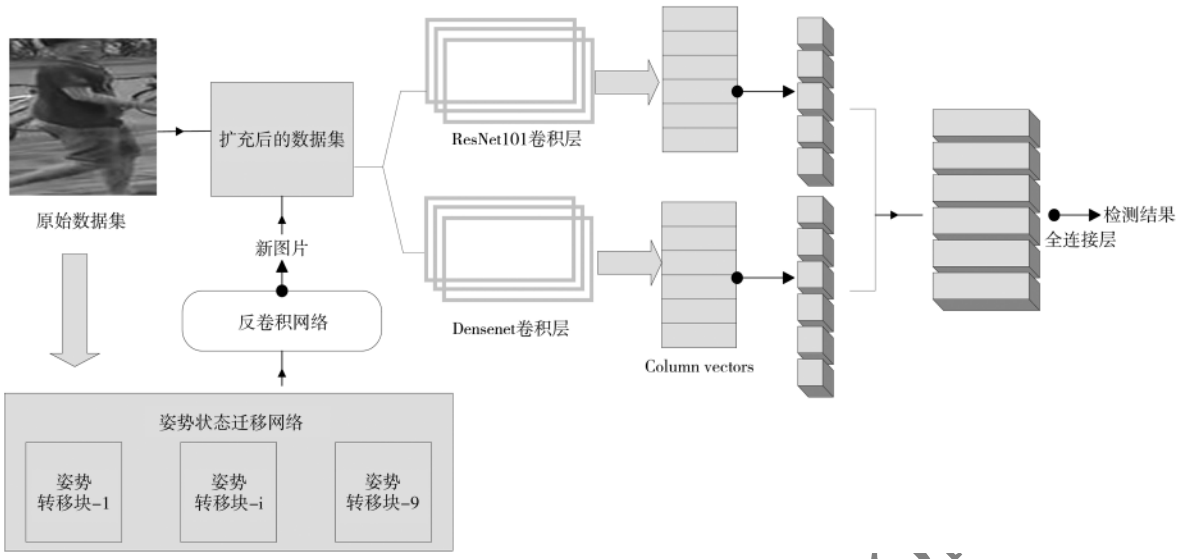

图 1 模型结构

姿势图片 P_c,关键点向量 S_c 经过 2 层降采样卷积层;P_t 也经过两层降采样卷积层,而后输入到姿势-注意转移网络中。姿势-注意转移网络是由几个特别设计的姿势-状态转移块构成的,每个转移块只传输它所关注的某些区域,逐步生成人物图像,允许每个传输块在训练上时行局部传递,因此避免了捕获全局的复杂结构。所有姿势转移块均具有相同的结构,它有两条相互独立的支路构成:图像支路和姿势支路,分别算出生成图片的图像编码和姿势编码。最后将姿势-注意转移网络的输出——图像编码和姿势编码输入到解码器,通过 3 个反卷积网络从图像编码生成图像 P_g,如图 2 所示。

鉴别器也分为两部分:外观鉴别器和姿势鉴别器。外观鉴别器用来判断 P_g 和原图 P_t 是否为同一个人,姿势鉴别器则判断 P_g 的姿势是否与 P_c 相同,两个鉴别器结构相似。先将 P_g 分别与 P_t、P_c 在深度轴上结合,而后分别送入外观鉴别器和姿势鉴别器,得到输出外观相似度 R^A 和姿势相似度 R^S。最终的评分 $R=R^AR^S$。

2.2 行人重识别模型

为了解决行人重识别任务中用于识别的图片人物位置远近不一的问题,提出的行人重识别模型采用多层次的卷积核来提取图片特征,将这些特征拼接,而后进行分类,如图 3 所示。

特征提取部分采用两个算法分别提取特征,将图片分别输入到 ResNet[9]、Densenet[10] 之中,分别提取卷积层输出,得到特征向量 $\boldsymbol{L}_1$、$\boldsymbol{L}_2$ 将之连接,作为特征向量 $\boldsymbol{L}$。

在全连接层,借鉴 PCB 模型的思路,采用 6 个全连接层和 6 个分类器。这里采用交叉熵损失函数。在训练阶段,每个分类器用于预测给定图像的分类(人的身份)。测试阶段,给出两张图片 I_i 和 I_j(i

图 2 GAN 模型结构

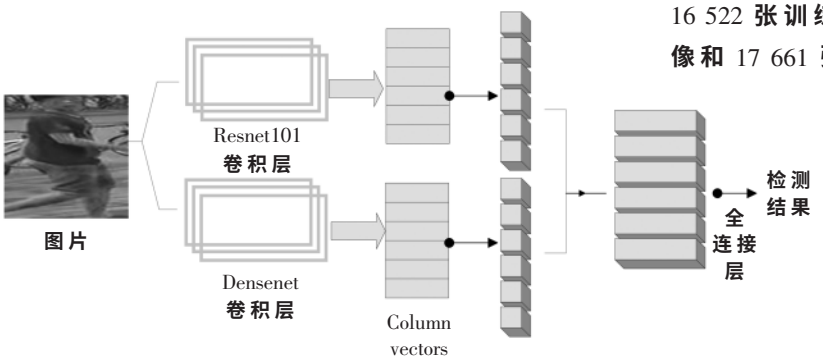

图 3 CNN 模型结构

和 j 图片在数据集中的标签)。

使用特征提取模型提取特征并获得两个特征向量,分别为 L_i 和 L_j。根据余弦距离计算相似度。

$$s(x_i, x_i)=\frac{x_i \cdot x_i}{|x_i||x_j|} \tag{1}$$

3 实验与分析

所有实验是在 Pytorch1.0.1、Python3.6 环境下实现的,硬件设备是 2 块 Titan RTX,显存 24 GB,E5-2680 v3 24 核处理器。

3.1 数据集

当前的 ReID 数据集已大大推动了有关的研究。充足的训练数据使得有可能开发更深的模型并展现出较强的识别能力。尽管目前的算法已在这些数据集上实现了很高的精度,但距离在实际场景中得到了广泛应用还很遥远。因此,有必要分析现有数据集的局限性。

与实际场景中收集的数据相比,当前数据集在四个方面存在局限性:(1)任务种类和摄像机的数量不够大,尤其是与真实监控视频数据相比时。目前,最大的数据集 DukeMTMC-reID 也只有 8 个摄像头,不到 2 000 个人物身份。(2)现有的大多数数据集仅包括单个场景,即室内或室外场景。(3)现有的大多数数据集都是由短时监控视频构建的,没有明显的光照变化。这些局限性使得采用一些方法扩充或增强数据集很有必要。

这里采用 DukeMTMC-reID 数据集,它是 Duke-MTMC 数据集的子集,用于基于图像的行人重识别。有 1 404 个行人出现在两个及以上的摄像头中,408 个行人只出现在 1 个摄像头。选择 702 人作为训练集,其余 702 人被用作测试集。在测试集中,为每个摄像机中的每个行人选取一张查询图像,并将其余图像放入图库中。这样,就可以得到 702 个人的 16 522 张训练图像,其他 702 人的 2 228 张查询图像和 17 661 张图库图像。每张图像都包含其相机 ID 和时间戳。

3.2 模型训练

由于行人重识别数据集由不同人的图像组成,每个人都有不同数量的图像样本。通常情况下,每批次练训练数据是从全部训练集中随机选取的,这样就会使输入的训练数据对于每个 ID 不平等,拥有更多图像样本的 ID 训练的次数就越多,图像样本越少的 ID 训练得越少。这样训练出来的模型会对拥有更多图像样本的 ID 提供更多的信息,但是每个人都应具有同样的重要性,应该受到平等的对待,因此引入了均衡采样策略。对于样本少的 ID,利用 GAN 模型加以扩充。

在训练姿势迁移网络时,采用 Adam 优化器,迭代 120 轮,$\beta_1=0.5$,$\beta_1=0.999$,初始学习率为 2×10^{-4},采用指数衰减。在每一层卷积层和归一化层后使用 Leaky ReLU 做激活函数,负斜率系数设置为 0.2。将生成图片与原数据集合并,得到一个 10 万张图片的训练集。

在 CNN 部分,先进行图像预处理将图像尺寸调整为 384×192×3,然后以 0.5 的概率对图像水平翻转,改变图像的亮度对比度为 1.7 和饱和度 1.5,并使用随机擦除,用均值[0.485,0.456,0.406]和标准差[0.229,0.224,0.225]分别对每个通道的数据进行正则化。

Resnet 采用 Pytorch 预训练的模型 resnet101,Densenet 采用 Pytorch 预训练的模型 densenet121,均使用训练好的参数。将两个模型卷积层的输出设置为 1×1 024,将两个输出连接得到 1×2 048 的向量作为图片特征向量。

用 Softmax 进行分类,在训练阶段,6 个分类器独立预测给定图像的类别(人的身份),从而获得更高的准确性。将 Dropout rate 设置为 0.3,使用 SGD 优化器,并将 momentum 设置为 0.9,权重衰减因子设置为 5×10^{-4},训练时初始学习率设置为 0.1,每训练 40 个 epoch 衰减为原来的 0.1,一共训练 120 轮。

3.3 实验结果与分析

该算法分别在 Market-1501 和 DukeMTMC-ReID 上进行了测试。与现有的算法相比,取得了更好的效果。

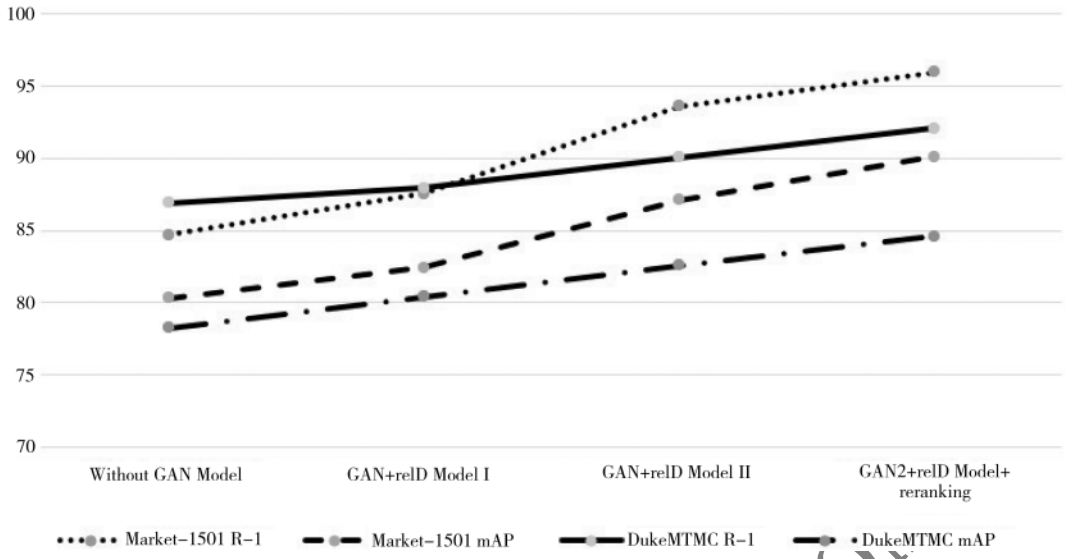

图 4 采用不同的识别策略的性能

从图 4 可以看出,在没有使用 GAN 扩充数据集前,单纯的 ReID Model 在 DukeMTMC-reID 数据集数据集上测试时,Rank-1 为 86.90%, mAP 为 78.25%。在使用了 GAN 模型,随机对一些行人的图片进行扩充(GAN+reID Model Ⅰ)后,Rank-1 和 mAP(mean average precision)分别提高到了 87.95%、80.41%,但是效果提升并不明显。而针对性地挑选出图像相对较少的行人进行 GAN 增强,并采用了一系列数据增强措施,在 ReID Model 部分采用 resnet101 代替 resnet50(GAN+reID Model Ⅱ)后,可以看到模型的效果有明显提升。最后,使用 re-ranking 方法,在 Market-1501 数据集上 Rank-1 的提升到 96.00%, mAP 提升到 90.10%,在 DukeMTMC-reID 数据集上 Rank-1 提升到 92.10%, mAP 提升到 84.60%。

将这一模型与最近的方法[11-15]进行比较,在 Market-1501 数据集上就行测试。这些方法主要是基于表示学习、局部特征和 GAN 映射的行人重识别方法。详情见表 1。在每个数据集上,这种方法都有很好的性能,并且与以前的方法相比性能有所提高。

表 1 模型性能比较

方法	Market-1501		DukeMTMC	
	R-1	mAP	R-1	mAP
Svdnet+randomerasing	89.13	83.90	84.02	78.28
HPM	94.20	82.7	86.6	74.3
DG-Net(RK)	95.40	92.49	90.60	88.31
OSNet	94.80	84.90	88.60	73.50
MGN	95.70	86.90	88.70	78.40
Incremental Learning	89.30	71.80	80.00	60.20
Parameter-Free Spation Attention	94.70	91.70	89.00	85.90
本文	96.00	90.10	92.10	84.60

4 结论

本文提出的基于姿势迁移和 CNN 的模型应用在行人重识别中,通过两个方面提高模型性能。首先,使用均衡采样策略和姿态转移方法来扩展数据集,从而解决了数据量少和不平衡的问题。这使得能够训练更复杂的神经网络并提取更多特征。其次,该方法使用两种不同的卷积网络结构提取图像特征,使所获得的信息更加全面,从而进一步提高了准确性。尽管该方法有效,但仍有待改进。更好的策略是考虑使用不同尺寸的卷积核以提取不同级别的特征。在未来的工作中将继续研究更有效的模型。

参考文献

[1] SUN Y, ZHENG L, YANG Y, et al. Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline)[J]. arXiv: 1711.09349, 2017.

[2] VARIOR R R, SHUAI B, LU J, et al. A siamese long short-term memory architecture for human re-identification[J]. arXiv: 1607.08381v1, 2016.

[3] Liang Zheng, Huang Yujia, Lu Huchuan, et al. Pose invariant embedding for deep person re-identification[J]. arXiv: 1701.07732, 2017.

[4] ZHENG Z, ZHENG L, YANG Y. Unlabeled samples generated by gan improve the person re-identification

baseline in vitro[J]. arXiv:1701.07717, 2017.
[5] ZHONG Z, ZHENG L, ZHENG Z, et al. Camera style adaptation for person re-identification[J]. arXiv:1711.10295, 2017.
[6] WEI L, ZHANG S, GAO W, et al. Person transfer GAN to bridge domain gap for person re-identification[J]. arXiv:1711.08565, 2017.
[7] QIAN X, FU Y, WANG W, et al. Pose-normalized image generation for person re-identification[J]. arXiv:1712.02225, 2017.
[8] ZHU Z, HUANG T, SHI B, et al. Progressive pose attention transfer for person image generation[J]. arXiv:1904.03349v3, 2019.
[9] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[J]. arXiv:1512.03385v1, 2015.
[10] HUANG G, LIU Z, LAURENS V D M, et al. Densely connected convolutional networks[J]. arXiv:1608.06993v5, 2016.
[11] INSAFUTDINOV E, PISHCHULIN L, ANDRES B, et al. DeeperCut: a deeper, stronger, and faster multi-person pose estimation model[C]. arXiv:1605.03170, 2016.
[12] KIM T, CHA M, KIM H, et al. Learning todiscover cross-domain relations with generative adversarial networks[J]. arXiv:1703.05192, 2017.
[13] LEDIG C, THEIS L, HUSZAR F, et al. Photo-realistic single image super-resolution using a generative adversarial network[J]. arXiv:1609.04802, 2016.
[14] LI D, CHEN X, ZHANG Z, et al. Learning deep context-aware features over body and latent parts for person re-identification[C]. arXiv:1710.06555, 2017.
[15] 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), 2017.

(收稿日期:2020-04-24)

作者简介:

林通(1998-),男,本科,主要研究方向:人工智能、数据处理。

陈新(1981-),男,博士,副教授,主要研究方向:大数据、人工智能。

唐晓(1983-),女,博士,讲师,主要研究方向:人工智能。

版权声明