

基于深度卷积生成对抗网络和拓展近邻重排序的行人重识别

戴臣超 王洪元 倪彤光 陈首兵

(常州大学信息科学与工程学院 江苏常州 213164)

(17000138@smail.cczu.edu.cn)

Person Re-Identification Based on Deep Convolutional Generative Adversarial Network and Expanded Neighbor Reranking

Dai Chenchao, Wang Hongyuan, Ni Tongguang, and Chen Shoubing

(School of Information Science and Engineering, Changzhou University, Changzhou, Jiangsu 213164)

Abstract Person Re-Identification (Re-ID) focuses on identifying the same person among disjoint camera views. This task is highly challenging, especially when there exists only several images per person in the database. Aiming at the problem of insufficient number of person images in person re-identification dataset, a method that generates extra training data from the original dataset is proposed. There are two challenges in this work, one is how to get more training data from the original training set, and the other is how to deal with these newly generated training data. The deep convolutional generative adversarial network is used to generate extra unlabeled person images and label smoothing regularization is used to process these newly generated unlabeled person images. In order to further improve the accuracy of person re-identification, a new unsupervised reranking framework is proposed. This framework neither requires to recalculate a new sorted list for each image pairs nor requires any human interaction or label information. Experiments on the datasets Market-1501, CUHK03, and DukeMTMC-reID verify the effectiveness of the proposed method.

Key words person re-identification; deep convolutional generative adversarial network; reranking; label smoothing regularization; unsupervised

摘要 行人重识别任务旨在识别不相交摄像头视图下的相同行人。这项任务极具挑战性,尤其是当数据集中每个行人仅仅有几张图片时。针对行人重识别数据集中行人图片数量不足的问题,提出一个从原始数据集中生成额外训练数据的方法。在这项工作之中存在2个挑战:1)如何从原始数据集之中获取更多的训练数据;2)如何处理这些新生成的训练数据。使用深度卷积生成对抗网络来生成额外的无标签行人图片,并采用标签平滑正则化来处理这些新生成的无标签行人图片。为了进一步提升行人重识别准确度,提出了一种新的无监督重排序框架。此框架既不需要为每组图像对重新计算新的排序列表,也不需要任何人工交互或标签信息。在Market-1501, CUHK03和DukeMTMC-reID数据集上的实验验证了所提方法的有效性。

关键词 行人重识别;深度卷积生成对抗网络;重排序;标签平滑正则化;无监督

中图分类号 TP391.41

收稿日期:2019-03-21;修回日期:2019-05-20

基金项目:国家自然科学基金项目(61572085,61502058,61806026);江苏省自然科学基金项目(BK20180956)

This work was supported by the National Natural Science Foundation of China (61572085, 61502058, 61806026) and the Natural Science Foundation of Jiangsu Province of China (BK20180956).

通信作者:王洪元(hywang@cczu.edu.cn)

行人重识别^[1]旨在自动匹配同一个行人在不同摄像机视图下的行人图片,该任务在公共安全方面具有很大的应用潜力.因为光照、遮挡、姿势改变、背景混乱等因素,不同摄像机视图下同一个行人的图片往往有很大不同,如图 1 所示(同一列的图片属于同一个行人),所以该任务具有很大的挑战性,是目前计算机视觉领域的研究热点.近几年在行人重识别领域的进步主要归功于高性能的深度学习算法^[2].和传统方法相比^[3-6],这类算法往往需要大量标注好的行人图片.尽管目前已经发布了一些规模较大的行人重识别数据集,但是在这些数据集中每个行人的图片仍然是有限的.据统计,在 Market1501 数据集^[7]之中平均每个行人只有 17.2 张图片,在 CUHK03 数据集^[8]之中平均每个行人只有 9.6 张

图片,在 DukeMTMC-reID 数据集^[9]之中平均每个行人只有 23.5 张图片.为解决这个问题,有部分研究者试图使用其他场景下标注好的行人图片,例如 Ni 等人^[10]提出一个迁移学习模型试图学习不同场景下行人图片的公共特征.另外一些研究者试图利用无标签行人图片,例如 Xin 等人^[11]提出多视图聚类的方法来为无标签图片分配一个伪标签,然后在训练期间使用具有真实标签和伪标签的数据参与训练,进而提高模型的泛化性能.然而这些方法往往只在特定场合下有用,有时候甚至会导致模型性能下降.本文从另外一个角度出发,使用改进的生成对抗网络,从现有训练集中生成类似行人图片参与训练,从而引入更多的颜色、光照以及姿势变化等信息来正则化模型,提升模型的鲁棒性.



Fig. 1 Pedestrian image pairs of public datasets

图 1 公共数据集行人图像对

另一方面,行人重识别在本质上就是一个图片检索任务,给定一张查询图片,在对应的图库之中检索出与这张查询图片身份相同的行人图片.在这个过程中经常会出现匹配不吻合的现象,导致行人重识别第一匹配率下降.为缓解这个问题,越来越多的研究者在行人重识别任务中加入重排序^[12-15].简单来说,重排序就是重新计算查询图片与图库图片之间的距离,从而使得更多与查询图片身份一致的图库图片排在排序列表更靠前的位置.然而,目前流行的重排序方法往往需要根据每一组图片对的 k 近邻或者 k 互邻重新计算新的排序列表,这使得重排序操作复杂度极高.为了解决这个问题,本文通过引入拓展近邻距离^[16]的概念,提出了一种新

的重排序方法,它不需要为每组图像对重新计算新的排序列表.经实验发现,本文方法可以达到的最高 Rank-1 和 mAP 在 Market-1501 数据集上是 89.70% 和 82.86%,在 CUHK03 数据集上是 87.60% 和 87.46%,在 DukeMTMC-reID 数据集上是 76.44%, 67.59%,并成功用于 2018 全球(南京)人工智能应用大赛多目标跨摄像头跟踪赛题.

1 相关工作

1.1 生成对抗网络

生成对抗网络由 2 个子网络组成:生成网络和判别网络.在训练过程中二者相互博弈直到进入

一个均衡和谐的状态.针对普通的生成对抗网络经常会面临训练不稳定、网络难以收敛以及生成图片质量太差等问题,本文采用的深度卷积生成对抗网络^[17] (deep convolutional generative adversarial network, DCGAN)在3个方面作出改进:1)去除掉网络中卷积层后面的全连接层,此时网络变成了全卷积网络,实验发现去掉该全连接层,网络可以收敛的更快;2)在生成网络的输入层之后作批归一化操作,使得训练过程变得更加稳定;3)使用步幅卷积来替换网络中的所有池化层,使得网络可以学习自身的空间下采样,进而生成质量更高的行人图片.

1.2 重排序

最近几年,重排序在行人重识别领域受到了越来越多的关注,Shen 等人^[12]使用 k 近邻生成新的排序列表,并且基于这些排序列表重新计算图片对之间的距离.Ye 等人^[13]将 k 近邻的全局特征与局部特征组合为一个新的查询特征,并且根据这些信息修正初始排名列表.不同于普通的 k 近邻,Zhong 等人^[14]利用 k 互邻来计算杰卡德距离,融合杰卡德距离与原始的欧氏距离获得更有效的排序距离.与基于 k 近邻和 k 互邻的重排序方法不同,本文

引入了拓展近邻距离的概念,通过聚合每一对图片的拓展近邻距离来进行重排序(本文称之为拓展近邻重排序).

由此,本文结合 DCGAN 与拓展近邻重排序进行行人重识别.通过 DCGAN 从原始训练集中生成类似的行人图片参与训练,提高模型的泛化能力,再对初始排序表进行拓展近邻重排序操作,缓解行人匹配错误的现象,提升行人重识别的性能.

2 基于 DCGAN 和拓展近邻重排序的行人重识别

本文提出基于 DCGAN 和拓展近邻重排序的行人重识别方法,整体框架如图 2 所示.首先使用训练好的卷积神经网络分别提取图库图片和查询图片的特征向量,然后基于欧氏度量计算每一张图库图片和查询图片之间的距离生成初始排序表,最后利用拓展近邻重排序对初始排序表进行重排序,生成最终的排序表.本节将介绍卷积神经网络如何在训练期间引入 DCGAN 生成的行人图片,以及拓展近邻重排序的定义与描述.

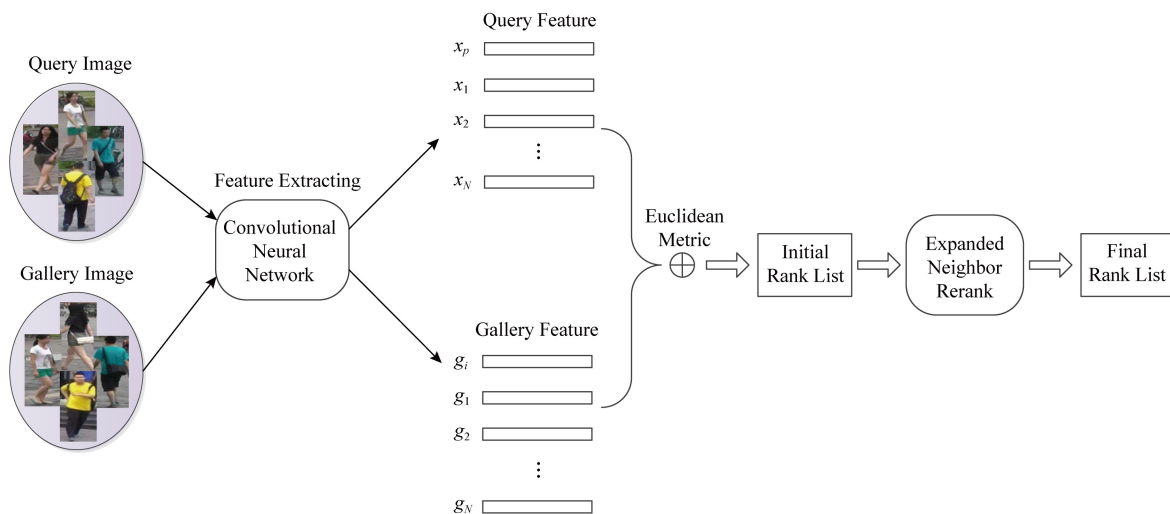


Fig. 2 The whole framework of the method

图 2 本文方法的整体框架

2.1 卷积神经网络

本文采用 ResNet50^[1]作为基础网络结构,并且采用文献[1]中的训练策略.不同于文献[1],本文去掉了最后 1000 维的分类层,并且增加了 2 层全连接层:第 1 个全连接层的输出为 1024 维(称为 FC-1024);第 2 个全连接层的输出是 T 维(称为 FC-T,其中 T 代表训练集中的类别数).经实验证明,添加

这 2 层全连接层可以很有效地提升行人重识别的准确率,并且不会影响模型的收敛速度.与文献[1]采取的策略不同,本文并没有为生成图片赋予一个伪标签,而是分配了一个对现有类别都统一的标签分布.在 2.2 节将对生成图片标签分布的分配进行讨论.结合生成图片的网络训练框架如图 3 所示.首先使用 DCGAN 从现有训练集中生成类似的行人

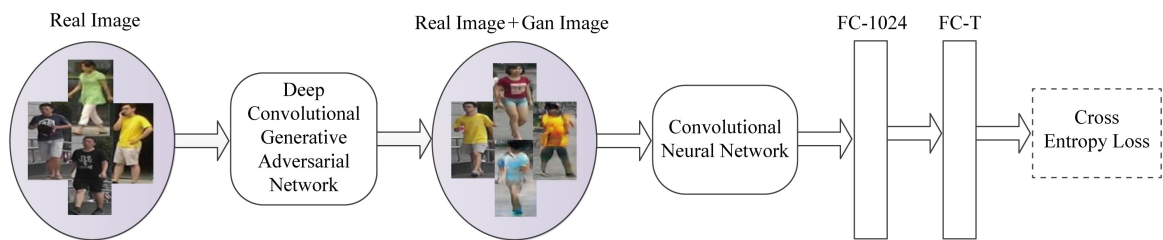


Fig. 3 Network training framework with generated image

图3 结合生成图片的网络训练框架

图片,然后将生成图片与真实图片混合在一起作为卷积神经网络的输入,最后通过最小化交叉熵损失来对网络进行调参使模型更鲁棒。

2.2 标签平滑正则化

标签平滑正则化^[18] (label smoothing regularization, LSR)的思想就是给非真实类别赋予一个较小的值而不是0.这种策略相当于加入一些噪音数据,以防网络过于倾向真实类别.在每一个批次中都有一定数量的真实图片和生成图片,所以本文的损失函数为

$$L_{\text{Loss}} = L_{\text{R}} + L_{\text{G}}, \quad (1)$$

其中, L_{R} 为真实图片的交叉熵损失; L_{G} 为生成图片的交叉熵损失.交叉熵损失定义为

$$L_{\text{Cross}} = - \sum_{n=1}^N \log(p(n))q(n), \quad (2)$$

其中, $n \in \{1, 2, \dots, N\}$, N 是原始训练集中预先定义好的类别总数; $p(n) \in [0, 1]$ 是网络预测输入图片属于第 n 个类别的概率;本文使用softmax函数对 $p(n)$ 进行归一化处理,因此 $\sum_{n=1}^N p(n) = 1$; $q(n)$ 指的是输入图片的真实标签分布.因为训练集中每张行人图片只有一个标签,所以对应的标签分布中必有一项为1而其他项均为0,所以 $q(n)$ 可写为

$$q(n) = \begin{cases} 1, & n = y; \\ 0, & n \neq y; \end{cases} \quad (3)$$

其中, y 代表图片的真实标签.

在文献[18]中,通过标签平滑正则化把非真实类别的分布考虑在内,鼓励网络不要太倾向于真实类别.因此,运用标签平滑正则化策略后,图片的标签分布为

$$q_{\text{LSR}}(n) = \begin{cases} 1 - \epsilon + \frac{\epsilon}{N}, & n = y; \\ \frac{\epsilon}{N}, & n \neq y; \end{cases} \quad (4)$$

其中, $\epsilon \in [0, 1]$.

然而,生成的行人图片不属于任意已知类别,无法为之分配一个准确的标签.不同于以往文献处理无标签图片的策略,本文提出为无标签图片分配一个虚拟的标签分布,设置其标签分布在所有已知类别上都是统一的.因此,对于生成图片,本文改进后的标签分布被定义为

$$q_{\text{LSR}}(n) = \frac{1}{N}, \quad (5)$$

此时,生成图片的标签分布为均匀分布,即默认任意一张生成图片属于任意已知类别的概率是相同的.所以,生成图片的交叉熵损失为

$$L_{\text{G}} = - \frac{1}{N} \sum_{n=1}^N \log p(n). \quad (6)$$

考虑到训练集中行人图片的内容和标签是正确匹配的,本文并没有对训练集中的真实图片进行标签平滑正则化处理,因此真实图片的交叉熵损失为

$$L_{\text{R}} = - \log(p(y)). \quad (7)$$

结合式(6)与式(7),本文的损失函数为

$$L_{\text{Loss}} = - (1 - \beta) \log(p(y)) - \frac{\beta}{N} \sum_{n=1}^N \log(p(n)), \quad (8)$$

其中, β 的取值为0或者1, $\beta=0$ 时,是真实图片的损失;当 $\beta=1$ 时,则是生成图片的损失.虽然本文使用的生成图片质量并不高,无法为之分配一个与图片内容相符的标签,但是本文使用改进后的标签平滑正则化方法可以直接处理在样本空间中位于真实图片附近的生成图片.通过这种方式可以引入更多的颜色、光照、背景、姿势变化等信息来正则化网络模型,进而引导网络去寻找更具有判别力的特征.

2.3 拓展近邻重排序

给定一个行人图库 $G = \{g_i | i = 1, 2, \dots, S\}$ 以及一张查询图片,那么查询图片与图库图片之间的距离可以用欧氏距离来计算:

$$d(p, g_i) = (\mathbf{x}_p - \mathbf{x}_{g_i})^T \times (\mathbf{x}_p - \mathbf{x}_{g_i}), \quad (9)$$

其中, p 指的是查询图片; g_i 指的是图库中的第 i 张

图片; \mathbf{x}_p 指的是查询图片 p 的特征向量; \mathbf{x}_{g_i} 指的是图库图片 g_i 的特征向量.

重排序就是要重新计算查询图片与图库图片之间的距离,进而生成新的排序表.给定一组图片(包括一张查询图片与一张图库图片).首先寻找 2 张图片对应的拓展近邻集合,然后聚合这 2 个集合中每一对图片之间的距离,并用聚合之后的距离(本文称之为拓展近邻距离)来代替该组图片原本的距离.查询图片的拓展近邻集合 $R(p, K)$ 为

$$\{R(p, M), R(M, N)\} \rightarrow R(p, K), \quad (10)$$

其中, $R(p, M)$ 是与查询图片 p 最相似的 M 张近邻图片; $R(M, N)$ 指的是与 $R(p, M)$ 中每个元素最相似的 N 张近邻图片.同理,图库图片的拓展近邻集合计算方式也是如此.拓展近邻距离可以被定义为

$$d^*(p, g_i) = \frac{1}{2K} \sum_{j=1}^K (d(p_j, g_i) + d(g_{ij}, p)), \quad (11)$$

其中, p_j 指的是查询图片的拓展近邻集合 $R(p, K)$ 中的第 j 个近邻; g_{ij} 指的是图库图片 g_i 的扩展近邻集合 $R(g_i, K)$ 中的第 j 个近邻; $d(\cdot)$ 是图片对之间的距离.

本文采用基于排序列表的方式来计算 $d(\cdot)$,即根据 2 个排序列表前 k 个近邻的位置来计算 2 个排序列表之间的距离.虽然该方法是由 Jarvis 等人^[16]提出的并且已经成功地运用到人脸识别任务^[19]中,但是据我们所了解,该方法是首次运用到行人重识别任务中.基于排序列表的距离计算:

$$d_r(p, g_i) = 1 - \min_{n=1}^N \left(\sum_{n=1}^N \left([k+1 - S_p(n)]_+ \times [k+1 - S_{g_i}(n)]_+ \right) \right), \quad (12)$$

其中, $S_p(n)$ 表示图片 n 在排序列表 L_p 中的位置; $S_{g_i}(n)$ 表示图片 n 在排序列表 L_{g_i} 中的位置; \times 表示矩阵乘法; L_p 与 L_{g_i} 分别表示 p 与 g_i 所对应的排序列表. $[\cdot]_+ = \max(\cdot, 0)$. 因为只考虑排序列表中前 k 张行人图片的情况,所以使用 \max 函数将排序列表中前 k 张以外的图片排除在外. $\min_{n=1}^N(\cdot)$ 是指对 (\cdot) 进行归一化处理.

3 实验

本文在 Market-1501, DukeMTMC-reID 和 CUHK03 数据集^[7-9]上进行实验,并采用累积匹配特征曲线以及平均查准率(mean average precision, mAP)来评价实验性能.累积匹配特征曲线表示查询图片出现在排序后图库列表中的概率.因为在实际

应用中往往只会考虑 Rank-1,即第一次就成功匹配的概率,所以本文主要关注 Rank-1.

3.1 数据集

Market-1501 中的行人图片收集自 6 个摄像机,其中包括来自于 1 501 个行人的 32 668 张标记好的行人图片.该数据集分为训练集,图库集和查询集.训练集由来自 751 个行人的 12 936 张行人图片组成,图库集则是由来自 750 个行人的 19 732 张行人图片组成.查询集中拥有来自 750 个行人的 3 368 张查询图片.

DukeMTMC-reID 是最新发布的大规模行人重识别数据集,其中有 1 812 个行人,一共 1 404 个行人出现在至少 2 个摄像头下,剩余 408 个行人仅出现在一个摄像头之下.它的训练集和测试集各包含 702 个人,训练集包括 16 522 张图片,图库集由 17 661 张图片组成,查询集包括 2 228 张图片.

CUHK03 数据集由来自 1 467 个行人的 14 097 张行人图片组成.每个行人都会被 2 个不同的摄像机拍摄到并且在每个摄像机视图下平均每个行人有 4.8 张图片.该数据集有 2 种类型,一种是由人工标注的,另一种则是由检测器自动标注的.本文使用检测器标注而成的数据集,在训练集中平均每个人有 9.6 张行人图片.

3.2 实验设置

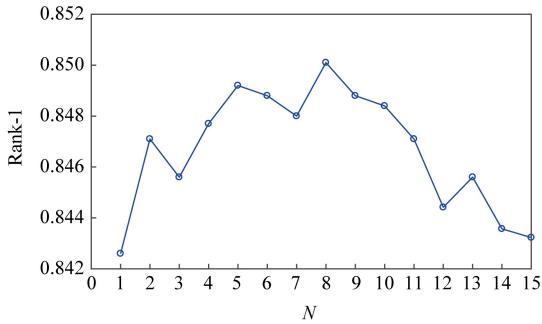
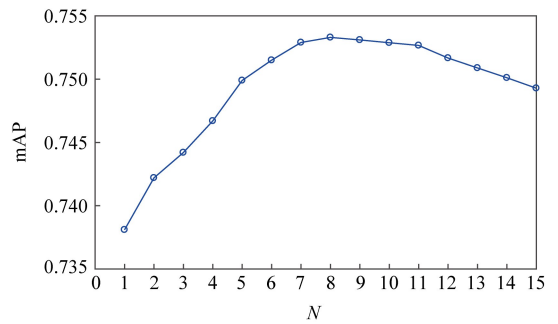
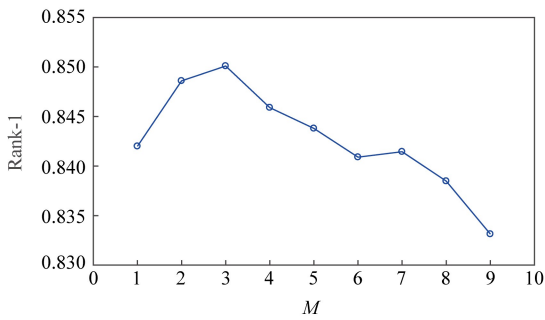
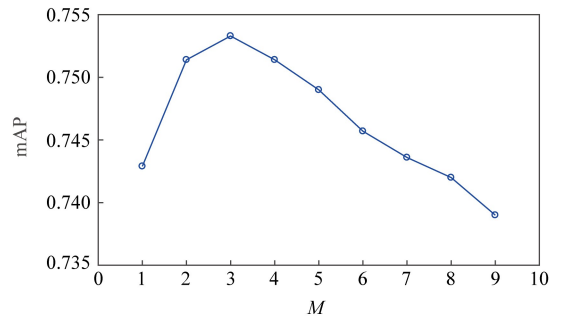
本文在 Market-1501 数据集上采用了 Single Query 和 Multiple Query 设置. Single Query 指查询集中每个行人只有一张图片, Multiple Query 指查询集中每个行人有多张图片.在 CUHK03 数据集上使用了 single-shot 和 multi-shot 设置. single-shot 是指在图库中每个行人只有一张图片, multi-shot 是指在图库中每个行人有多张图片.和本文方法进行比较的有 DNS^[20], Verif.-Identif.^[21], SOMAnet^[22], XQDA^[23]等方法. DNS 方法通过匹配训练数据判别零空间中的行人来克服行人重识别度量学习中的小样本问题. Verif.-Identif. 方法基于验证模型和识别模型可以同时计算验证损失和识别损失,进而可以得到一个更具有判别力的卷积神经网络和相似性度量. SOMAnet 则是基于深度卷积神经网络的框架,它试图通过对人体结构信息建模来缓解行人重识别的类内差异. XQDA 方法首先提取特征,并提出一种度量学习方法来使类内距离变小类间距离变大.

3.3 参数分析

本文在 Market1501 数据集上采用控制变量法对重排序方法中的参数 M 与 N 进行测试,以期

获得最优值.首先固定住一个参数,然后通过调整另一个参数来查看 Rank-1 和 mAP 的变化.从图 4 中可以看出在 Single Query 设置下当 $N=8$ 时,Rank-1 和 mAP 达到最优值.从图 5 中可以看出在 Single

Query 设置下当 $M=3$ 时,Rank-1 和 mAP 达到最优值.经实验测试,在数据集 CUHK03 和 DukeMTMC-reID 上, $M=3$ 与 $N=8$ 时 Rank-1 和 mAP 也可以达到最优.因此,本文取 $M=3$ 和 $N=8$ 为实验参数.

(a) Influence of parameter N on Rank-1(b) Influence of parameter N on mAPFig. 4 Influence of parameter N on Re-ID performance on Market1501 dataset图 4 在 Market1501 数据集上参数 N 对行人重识别性能的影响(a) Influence of parameter M on Rank-1(b) Influence of parameter M on mAPFig. 5 Influence of parameter M on Re-ID performance on Market1501 dataset图 5 在 Market1501 数据集上参数 M 对行人重识别性能的影响

3.4 生成图片的数量对 Re-ID 性能的影响

图 6~8 分别展示了以 CUHK03, Market-1501

和 DukeMTMC-reID 为训练集的生成对抗网络生成的行人图片.随着生成图片数量的增加,行人重识



Fig. 6 Generated images of CUHK03

图 6 CUHK03 数据集的生成图片



Fig. 7 Generated images of Market-1501

图 7 Market-1501 数据集的生成图片

别的表现是否可以获得一个持续性的改进?表 1 展示了不同数量的生成图片参与 Market-1501 数据集训练的实验结果.在 Market-1501 数据集中训练集一共有 12 936 张图片.从表 1 中可以看出:太少的生成图片参与训练时,标签平滑正则化的正则化能力并不充分;太多的生成图片参与训练时,则会引入过多的干扰项.本文采取一个折中处理,即设置参与训练的生成图片与真实图片的比例为 1:1.



Fig. 8 Generated images of DukeMTMC-reID

图 8 DukeMTMC-reID 数据集的生成图片

Table 1 Impact of the Number of Generated Images on Person Re-identification

表 1 生成图片的数量对 Re-ID 性能的影响

Gan Image	Rank-1	mAP
0	0.799 3	0.587 7
6 000	0.800 5	0.589 1
12 000	0.811 5	0.593 8
18 000	0.810 9	0.588 3
24 000	0.794 8	0.568 3

3.5 标签平滑正则化对 Re-ID 性能的改进

表 2~5 展示了本文方法在 Market-1501, CUHK03 和 DukeMTMC-reID 数据集上的实验结果.在所有表中,LSR 代表在训练期间加入了生成图片,rerank 代表在测试期间使用了拓展近邻重排序操作.从中可以看出:当使用生成图片参与训练时,实验效果明显超过了本文的基准方法 Resnet50;在 Market-1501 数据上 Single Query 与 Multiple Query 情况下,Rank-1 分别提升了 1.22% 和 0.66%,mAP 分别提升了 0.61% 和 1.48% (如表 2 所示);在 CUHK03 数据集 Single-shot 与 multi-shot 情况下,

Rank-1 分别提升了 2.97% 和 0.51%,mAP 分别提升了 2.32% 和 0.70% (如表 3、表 4 所示);在 DukeMTMC-reID 数据上,Rank-1 和 mAP 上分别提升了 1.22% 和 0.64% (如表 5 所示).这使我们有兴趣探索使用现实生活中真实的行人图片参与训练是否也有好的效果.为了验证这一点,本文随机挑选

Table 2 Performance Comparison on the Market-1501 Dataset

表 2 在 Market-1501 数据集上的性能比较

Methods	Single Query		Multiple Query	
	Rank-1	mAP	Rank-1	mAP
Bow ^[7]	0.343 8	0.141 0	0.426 4	0.194 7
MR CNN ^[24]	0.455 8	0.261 1	0.565 9	0.323 6
DNS ^[20]	0.554 3	0.298 7	0.715 6	0.460 3
Gated-SCNN ^[25]	0.658 8	0.395 5	0.760 4	0.484 5
SOMAnet ^[22]	0.738 7	0.478 9	0.812 9	0.569 8
Verif.-Identif ^[21]	0.795 1	0.598 7	0.858 4	0.703 3
ResNet50	0.799 3	0.587 7	0.870 2	0.684 4
DeepTransfer ^[26]	0.837 0	0.655 0	0.896 0	0.738 0
Ours+LSR	0.811 5	0.593 8	0.876 8	0.699 2
Ours+LSR+rerank	0.850 1	0.753 3	0.897 0	0.828 6

Table 3 Performance Comparison on the CUHK03 Dataset Under single-shot Setting

表 3 single-shot 情况下多种方法在 CUHK03 数据集上的比较

Methods	Rank-1	Rank-5	Rank-10	mAP
KISSME ^[27]	0.117 0	0.330 0	0.480 0	
DeepReID ^[8]	0.199 0	0.493 0	0.647 0	
CaffeNet ^[28]	0.358 0	0.653 0	0.779 6	0.426 0
VGG16 ^[29]	0.491 0	0.784 0	0.872 0	0.557 0
SOMAnet ^[22]	0.724 0	0.921 0	0.958 0	
ResNet50	0.705 3	0.915 7	0.959 4	0.752 9
Ours+LSR	0.735 0	0.922 4	0.960 4	0.776 1
Ours+LSR+rerank	0.789 5	0.917 6	0.943 0	0.817 7

Table 4 Performance Comparison on the CUHK03 Dataset Under multi-shot Setting

表 4 multi-shot 情况下 CUHK03 数据集上的性能比较

Methods	Rank-1	Rank-5	Rank-10	mAP
CaffeNet ^[28]	0.433 0	0.635 0	0.768 0	0.372 0
VGG16 ^[29]	0.588 0	0.802 0	0.873 0	0.510 0
Gated-SCNN ^[25]	0.681 0	0.881 0	0.946 0	0.588 0
ResNet50	0.795 7	0.897 4	0.941 1	0.741 5
Ours+LSR	0.800 8	0.908 5	0.945 1	0.748 5
Ours+LSR+rerank	0.876 0	0.910 6	0.938 0	0.874 6

Table 5 Performance Comparison on the DukeMTMC-reID**Dataset****表 5 DukeMTMC-reID 数据集上的性能比较**

Methods	Rank-1	mAP
BOW+KISSME ^[7]	0.2513	0.1217
LOMO+XQDA ^[23]	0.3075	0.1704
OIM ^[30]	0.6810	0.4740
TriNet ^[31]	0.7244	0.5350
ResNet50	0.6925	0.4928
Ours+LSR	0.7047	0.4992
Ours+LSR+rerank	0.7644	0.6759

DukeMTMC-reID 数据集中 12 000 张真实的图片来代替 Market-1501 训练集中的生成图片。

从表 6 可以看出,加入 12 000 张 DukeMTMC-reID 数据集中的真实图片参与训练也有助于模型的正则化,改善了行人重识别的准确度,但是加入 DCGAN 生成的图片参与训练取得的效果更好。

Table 6 Impact of Generated Images and Real Images on Person Re-identification**表 6 生成图片与真实图片对行人重识别性能的影响**

Image	Rank-1	mAP
0	0.7993	0.5877
DukeMTMC-reID-Real-12000	0.8046	0.5840
Market-1501-GAN-12000	0.8115	0.5938

Table 7 Comparison of Various Reranking Methods**表 7 不同重排序方法的比较**

Methods	Market-1501		CUHK03		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
ResNet-50	0.7993	0.5877	0.7053	0.7529	0.6925	0.4928
Ours+LSR	0.8115	0.5938	0.7350	0.7761	0.7047	0.4992
Ours+LSR+SCA	0.8311	0.7401	0.7500	0.8019	0.7387	0.6527
Ours+LSR+k-r	0.8381	0.7454	0.7609	0.8044	0.7451	0.6618
Ours+LSR+rerank	0.8501	0.7533	0.7895	0.8177	0.7644	0.6759

4 总 结

基于深度学习算法的行人重识别方法往往需要大量标记好的训练数据,然而标记大量的行人图片极其耗时.针对这个问题,本文使用改进的生成对抗网络 DCGAN 从现有训练集中生成类似行人图片参与训练,并采用标签平滑正则化方法来同时训练生成图片与真实图片.通过这种方式在训练过程中引入更多的颜色、光照、背景、姿势变化等信息,提升

3.6 拓展近邻重排序对 Re-ID 性能的提高

从表 2~5 中可以看出,拓展近邻重排序可以有效地改善行人重识别的性能:在 Market-1501 数据集 Single Query 情况下,Rank-1 和 mAP 分别提升了 3.86% 和 16%,Multiple Query 情况下,Rank-1 和 mAP 分别提升了 2.02% 和 12.94%;在 CUHK03 数据集 single-short 情况下 Rank-1 和 mAP 分别提升了 5.45% 和 4.16%,multi-short 情况下 Rank-1 和 mAP 分别提升了 7.52% 和 12.61%;在 DukeMTMC-reID 数据集上 Rank-1 和 mAP 分别提升了 5.97% 和 17.67%。

本文还将拓展近邻重排序方法和目前 2 种流行的重排序方法相比较,这 2 种重排序方法分别为稀疏上下文激活重排序^[15](sparse contextual activation, SCA), k 互邻重排序^[14](k -reciprocal, k -r). SCA 通过考虑上下文空间中的原始成对距离,设计一个稀疏上下文激活的特征向量来对图像进行编码. k -r 则是通过 k 互近邻方法降低图像错误匹配情况,最后将欧氏距离和杰卡德距离加权来对排序表进行重排序.表 7 展示了在 Market-1501, CUHK03 和 DukeMTMC-reID 这 3 个数据集上拓展近邻重排序方法和其他 2 种先进重排序方法的实验比较结果.从表 7 中可以看出,本文提出的拓展近邻重排序方法无论是 Rank-1 指标还是 mAP 指标都要比上述 2 种重排序方法效果要好。

模型的鲁棒性.大量实验表明该方法能有效提升行人重识别性能.在此基础上,本文进一步提出了用于行人重识别的新的重排序方法——拓展近邻重排序方法.该方法根据 2 个排序列表前 k 个近邻的位置来计算 2 个排序列表之间的距离,无需重新计算每一对图片的排序列表,通过有说服力的实验,证明该方法比其他重排序方法有更好的性能表现,能更好地提升行人重识别性能.本文所报道的上述方法也在 2018 全球(南京)人工智能应用大赛多目标跨摄像头跟踪赛中得到了成功应用(荣获第 2 名)。

参 考 文 献

- [1] Zheng Liang, Yang Yi, Hauptmann Alexander G. Person re-identification: Past, present and future [EB/OL]. 2016 [2019-05-10]. <https://arxiv.org/pdf/1610.02984.pdf>
- [2] LeCun Y, Bengio Y, Hinton G. Deep learning [J]. *Nature*, 2015, 521(7553): 436-444
- [3] Wang Hongyuan, Ding Zongyuan, Zhang Ji, et al. Person reidentification by semisupervised dictionary rectification learning with retraining module [J]. *Journal of Electronic Imaging*, 2018, 27(4): No.043026
- [4] NI Tongguang, Ding Zongyuan, Chen Fuhua, et al. Relative distance metric learning based on clustering centralization and projection vectors learning for person re-identification [J]. *IEEE Access*, 2018, 6(1): 11405-11411
- [5] Ding Zongyuan, Wang Hongyuan, Chen Fuhua, et al. Pedestrian weight recognition based on distance centralization and projection vector learning [J]. *Journal of Computer Research and Development*, 2017, 54(8): 1785-1794 (in Chinese)
(丁宗元, 王洪元, 陈付华, 等. 基于距离中心化与投影向量学习的行人重识别[J]. *计算机研究与发展*, 2017, 54(8): 1785-1794)
- [6] Liang V E, Lu J, Ge Y. Regularized local metric learning for person re-identification [J]. *Pattern Recognition Letters*, 2015, 68(2015): 288-296
- [7] Zheng Liang, Shen Liyue, Tian Lu, et al. Scalable person re-identification: a benchmark [C] //Proc of IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2015: 1116-1124
- [8] Li Wei, Zhao Rui, Xiao Tong, et al. Deepreid: Deep filter pairing neural network for person re-identification [C] //Proc of the 32nd IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2014: 152-159
- [9] Ristani E, Solera F, Zou R, et al. Performance measures and a data set for multi-target, multi-camera tracking [C] //Proc of European Conf on Computer Vision. Berlin: Springer, 2016: 17-35
- [10] Ni Tongguang, Gu Xiaoqing, Wang Hongyuan, et al. Discriminative deep transfer metric learning for cross-scenario person re-identification [J]. *Journal of Electronic Imaging*, 2018, 27(4): No.043026
- [11] Xin Xiaomeng, Wang Jinjun, Xie Ruji, et al. Semi-supervised person Re-Identification using multi-view clustering [J]. *Pattern Recognition*, 2019, 88(2019): 285-297
- [12] Shen Xiaohui, Lin Zhe, Brandt J, et al. Object retrieval and localization with spatially-constrained similarity measure and k-nn re-ranking [C] //Proc of the 30th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2012: 3013-3020
- [13] Ye Mang, Chen Jun, Leng Qingming, et al. Coupled-view based ranking optimization for person re-identification [C] //Proc of the 21st Int Conf on Multimedia Modeling. Berlin: Springer, 2015: 105-117
- [14] Zhong Zhun, Zheng Liang, Cao Donglin, et al. Re-ranking person re-identification with k-reciprocal encoding [C] //Proc of the 35th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 1318-1327
- [15] Bai Song, Bai Xiang. Sparse contextual activation for efficient visual reranking [J]. *IEEE Transactions on Image Processing*, 2016, 25(3): 1056-1069
- [16] Jarvis R A, Patrick E A. Clustering using a similarity measure based on shared near neighbors [J]. *IEEE Transactions on Computers*, 1973, 100(11): 1025-1034
- [17] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks [EB/OL]. 2016 [2019-05-10]. <https://arxiv.org/abs/1511.06434.pdf>
- [18] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision [C] //Proc of the 34th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 2818-2826
- [19] Schroff F, Treibitz T, Kriegman D, et al. Pose, illumination and expression invariant pairwise face-similarity measure via doppelgänger list comparison [C] //Proc of the 29th IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2011: 2494-2501
- [20] Zhang Li, Xiang Tao, Gong Shaogang. Learning a discriminative null space for person re-identification [C] //Proc of the 34th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 1239-1248
- [21] Zheng Zhedong, Zheng Liang, Yang Yi. A discriminatively learned cnn embedding for person reidentification [J]. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2018, 14(1): 13:1-13:20
- [22] Barbosa I B, Cristani M, Caputo B, et al. Looking beyond appearances: Synthetic training data for deep cnns in re-identification [J]. *Computer Vision and Image Understanding*, 2018, 167: 50-62
- [23] Liao Shengcai, Hu Yang, Zhu Xiangyu, et al. Person re-identification by local maximal occurrence representation and metric learning [C] //Proc of the 33rd IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 2197-2206
- [24] Ustinova E, Ganin Y, Lempitsky V. Multi-region bilinear convolutional neural networks for person re-identification [C] //Proc of the 14th IEEE Int Conf on Advanced Video and Signal Based Surveillance (AVSS). Piscataway, NJ: IEEE, 2017: 1-6
- [25] Varior R R, Haloi M, Wang G. Gated siamese convolutional neural network architecture for human re-identification [C] //Proc of European Conf on Computer Vision. Berlin: Springer, 2016: 791-808
- [26] Geng Mengyue, Wang Yaowei, Xiang Tao. Deep transfer learning for person re-identification [EB/OL]. 2016 [2019-05-10]. <https://arxiv.org/pdf/1611.05244.pdf>

- [27] Koestinger M, Hirzer M, Wohlhart P, et al. Large scale metric learning from equivalence constraints [C] // Proc of the 30th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2012: 2288-2295
- [28] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [C] // Proc of Advances in Neural Information Processing Systems. San Francisco, CA: Morgan Kaufmann, 2012: 1097-1105
- [29] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [EB/OL]. 2015 [2019-05-10]. <https://arxiv.org/pdf/1409.1556.pdf>
- [30] Xiao Tong, Li Shuang, Wang Baochao, et al. Joint detection and identification feature learning for person search [C] // Proc of the 35th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 3415-3424
- [31] Hermans A, Beyer L, Leibe B. In defense of the triplet loss for person re-identification [EB/OL]. 2017 [2019-05-10]. <https://arxiv.org/pdf/1703.07737.pdf>



Dai Chenchao, born in 1995. Master candidate at Changzhou University. His main research interests include computer vision and person re-identification.



Wang Hongyuan, born in 1960. Received his PhD degree from Nanjing University of Science and Technology in 2004. Professor and master supervisor at Changzhou University. His main research interests include image processing, artificial intelligence and pattern recognition.



Ni Tongguan, born in 1978. Received his PhD degree from Jiangnan University in 2015. Currently associate professor in the School of Information Science and Engineering, Changzhou University, Changzhou, China. His main research interests include pattern recognition, intelligent computation and their application.



Chen Shoubing, born in 1991. Master candidate at Changzhou University. His main research interests include computer vision and person re-identification.

更正声明

由于作者疏忽,刊登于《计算机研究与发展》2017年7期的文章“不完备多粒度决策系统的局部最优粒度选择”遗漏一篇参考文献,特向广大读者致歉.具体更正为:

1) 增加参考文献:

- [37] Gu Shenming, Wan Yahong, Wu Weizhi, et al. On local optimal granularity selections in multi-granular decision systems [J]. Journal of Nanjing University (Natural Sciences), 2016, 52(2): 280-288 (in Chinese)
(顾沈明, 万雅虹, 吴伟志, 等. 多粒度决策系统的局部最优粒度选择[J]. 南京大学学报, 2016, 52(2): 280-288)

2) 增加参考文献[37]引用标注:

- ① 第 1.1 节第一段中“一个信息系统是一个二元组 $(U, C)^{[37]}$ ……”
- ② 第 1.2 节第一段中“决策系统是一个二元组 $S = (U, C \cup \{d\})^{[37]}$ ……”
- ③ 第 1.2 节第 3 段中“隐藏在决策系统 $S = (U, C \cup \{d\})$ 中的规则, 可以用 $t \rightarrow s$ 的形式来表达^[37] ……”
- ④ 第 1.2 节第 4 段中, “规则 $t \rightarrow s$ 的可信度定义为^[37] ……”

作者: 顾沈明、顾金燕、吴伟志、李同军、陈超君
2019年7月