DOI: 10.13873/J.1000-9787(2020)10-0119-04

基于场景迁移与区域对齐的行人再识别

白　健[1]，耿树泽[2]，岑世欣[2]

(1. 河北工业大学 人工智能与数据科学学院，天津 300401；
2. 河北工业大学 电子信息工程学院，天津 300400)

摘　要：针对目前现有行人再识别方法训练样本不足的问题，利用语义分割方法将样本图像中的行人区域与背景分离，对背景区域使用生成式对抗网络(GAN)完成图像场景迁移，在保留行人特征的前提下对现有数据集进行扩充。针对数据集中行人区域未对准的情况，提出基于语义分割的滑动窗口行人对准方法，并根据数据集的扩充和对准在残差卷积神经网络结构 ResNet-50 中加入全局特征分支。实验中使用公共数据集 Market-1501 和 DukeMTMC-reID 对上述方法进行测试，在 Rank-1 指标上分别取得了 91.4% 和 81.1% 的准确率。

关键词：行人再识别；深度学习；生成式对抗网络；行人对准；滑动窗口

中图分类号：TP391.41　　**文献标识码**：A　　**文章编号**：1000-9787(2020)10-0119-04

Pedestrian re-identification based on scenes transition and region alignment

BAI Jian[1], GENG Shuze[2], CEN Shixin[2]

(1. School of Artificial Intelligence and Data Science, Hebei University of Technology, Tianjin 300401, China;
2. School of Electronic and Information Engineering, Hebei University of Technology, Tianjin 300400, China)

Abstract: To solve the problem of insufficient training samples of existing pedestrian re-identification methods, semantic segmentation is applied to separate the pedestrian area from the background in the sample image. and the generative adversarial network(GAN) is used to generate the different backgrounds for image scene transition. This mechanism not only expand the existing datasets, but also retain the pedestrian characteristics. The sliding window alignment(SWA) method based on semantic segmentation is proposed to solve the problem of pedestrian area misalignment in datasets. At the same time, the global feature branch is added to the residual convolutional neural network structure ResNet-50. In the experiment, this method is tested on Market-1501 and DukeMTMC-reID datasets, and achieve accuracy of 91.4% and 81.1% in rank-1 index.

Keywords: pedestrian re-identification; deep learning; generative adversarial network(GAN); pedestrian alignment; sliding window

0　引　言

在当前行人再识别的研究中，数据集中的样本易受拍摄角度、光照、拍摄环境等客观因素影响，造成数据样本质量较差。因此行人再识别是目前图像识别领域一项具有挑战性的任务。基于特征表达和距离学习的图像处理方法存在着以下问题：基于颜色特征的方法受光照等变化的影响较大；采集到行人图像存在着分辨率较低的情况，导致基于纹理特征的方法很难提取到有效的特征；行人特征和背景在场景改变后行人特征变换过于复杂，使得基于度量学习的方法在行人匹配上遇到了瓶颈。近年来，深度学习模型在计算机视觉任务上有着非常好的表现。

本文提出了一种基于场景迁移与区域对齐的方法来对行人再识别任务进行优化。使用行人特征恢复的生成式对抗网络(pedestrian feature recovered cycle-GAN, PFRCGAN)进行数据集的扩充和行人特征还原与修复，提高目前现有数据集的多样性；利用基于语义分割[1]的滑动窗口对准法(sliding window alignment, SWA)对行人数据集中行人与行人检测区域[2]未对准的情况进行优化；最后部分根据前两部分对数据集的扩充和对准，本文对 ResNet-50 进行改进，提出新的再识别模型，将基于行人标签的 IDE(ID-discriminative embedding)[3]训练策略中的 ResNet-50 结构加入全局特征分支，可以使神经网络可以兼顾局部特征和全局特征。

收稿日期：2019-05-28

1 本文方法

1.1 基于行人特征恢复的对抗网络

为了使行人再识别模型有更好的鲁棒性和场景变化适应能力[4]，本文首先对当前数据集进行优化和扩充，提出一种基于行人特征恢复的生成式对抗网络(pedestrian features recovered GAN, PFRGAN)。以循环对抗生成网络结构为基础实现场景迁移，即 Cycle-GAN[5]。该结构不需要场景间物体区域完全对应，并且该方法使用循环结构，使得生成的场景更加真实、稳定。

设场景 A 中的图像为$\{a_i\}_A$，场景 B 中的图像为$\{b_j\}_B$，i,j 分别为场景 A 和 B 行人的编号。为了让神经网络更好地适应这些可变因素，本文把场景 A 图像生成为场景 B 中的风格，实现场景迁移。在 Cycle-GAN 中包含两个对称的生成器 G_{AB}，G_{BA}，其中，G_{AB}场景 A 到场景 B 的生成器，G_{BA}为场景 B 到场景 A 的生成器，同时包括两个判别器 D_A，D_B，$E(\cdot)$ 为数学期望，如图 1(a) 所示。

传统 GAN 的对抗目标损失函数为

$$L_{\mathrm{GAN}}(G_{AB},D_B,A,B)=E_{b\sim B}[\log D_B(b)]+E_{a\sim A}[\log[1-D_B(G_{AB}(a))]] \quad (1)$$

Cycle-GAN 相对于传统的 GAN 加入了一致性检验。其思想为把场景 A 中的图片转移到场景 B，在将转移的图片转移到场景 A，计算原始的图片与通过两个生成器返回场景 A 的一致性，如图 1(b)，反向同理，损失函数为

$$\begin{aligned}L_{\mathrm{cyc}}(G_{AB},G_{BA})&=L_{\mathrm{cyc}-1}+L_{\mathrm{cyc}-2}\\&=E_{a\sim A}[\|G_{BA}(G_{AB}(a))-a\|_1]+E_{b\sim B}[\|G_{AB}(G_{BA}(b))-b\|_1]\end{aligned} \quad (2)$$

可得最终损失函数为

$$Loss=L_{\mathrm{GAN}}(D_B,G_{AB},A,B)+L_{\mathrm{GAN}}(D_A,G_{BA},A,B)+\mu L_{\mathrm{cyc}}(G_{AB},G_{BA}) \quad (3)$$

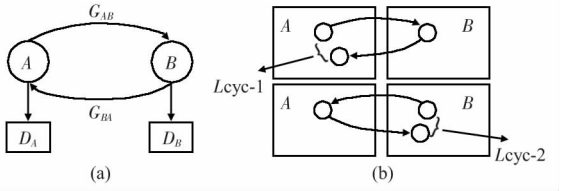

图 1 Cycle-GAN 工作流程

在场景迁移时，Cycle-GAN 不仅仅会生成行人样本的背景部分，也会对行人的外观进行风格转移，使得行人已经存在的具有分辨能力的外观特征发生变化，从而产生人体部位的缺失、行人颜色的变化、以及行人携带背景轮廓等情况。

因此，需要采取一定的手段对行人的特征进行恢复，增加生成图片行人特征的稳定性。本文使用 RefineNet[6] 对 Cycle-GAN 进行优化。RefineNet 可以按行人的身体部分进行分割。首先通过预训练的 ResNet 对图像进行降采样，保留 1/4，1/8，1/16，1/32 尺度的特征图，将其分为 4 个 ResNet 模块，然后分为 4 条路径通过 RefineNet 模块进行融合。将融合后的特征图接入 SoftMax 层，再使用双线性插值输出。对于一张行人样本图片，RefineNet 可以检测行人的头部、大臂、小臂、躯干、大腿、小腿六个部分，并且可以将行人与背景分离。

行人再识别数据集中行人样本的分辨率变化较大，质量较好的图片可以保证身体部位可以被完整地识别出来。然而，数据集中存在一小部分的低分辨率样本，即使增加图片的尺寸，RefineNet 也无法将行人的身体各个部位稳定地分割出来。所以，在使用 RefineNet 结构时，当其识别出行人的身体部位后，将其进行组合，组成人体掩模(mask)。

设定 RefineNet 提取出的掩模为一个只有 0 和 1 组成的矩阵 $\boldsymbol{M}$，其中的点可以表示为 $M_{i,j}$，当该点在图像中的行人区域内时，值为 1，若该点在背景区域，则该点值为 0。则对于原训练集中的图像样本集合$\{\boldsymbol{A}_i\}$ 则有其对应的 Mask 矩阵 $\boldsymbol{M}_{A_i}$，其原图像的行人提取为 $\breve{\boldsymbol{A}}_i$，$\breve{\boldsymbol{A}}_i$ 计算为

$$\breve{\boldsymbol{A}}_i=\boldsymbol{A}_i\odot\boldsymbol{M}_{A_i} \quad (4)$$

式中 $\odot$为矩阵中对应点相乘，设 Cycle-GAN 对应原图像 $\boldsymbol{A}_i$ 转移到其他场景的集合为$\{\boldsymbol{A}_j\}_i$，则复原后的行人样本图像为

$$\tilde{\boldsymbol{A}}_{i,j}=\overline{M}_{Ai}\odot\boldsymbol{A}_j'+\boldsymbol{A}_i\odot\boldsymbol{M}_{A_i}=\overline{\boldsymbol{M}}_{Ai}\odot\boldsymbol{A}_j'+\breve{\boldsymbol{A}}_i \quad (5)$$

式中 $\tilde{\boldsymbol{A}}_{i,j}$为 $\boldsymbol{A}_i$ 转移到 j 场景下复原后的图像样本 $j=1,2,3,\cdots,N_{\mathrm{scene}}$，$N_{\mathrm{scene}}$为场景个数，$\overline{\boldsymbol{M}}_{A_i}$则为对 $\boldsymbol{M}_{A_i}$中的 0 与 1 取反。

根据行人的身体部位的组合提取行人的掩模(mask)，并提取行人在原数据集掩模位置的图像，并还原到生成的图片中。Cycle-GAN 实现的仅是图像的风格转移，不会改变行人主体在图像中的位置，所以从原图中提取的图像可以按像素位置覆盖在生成的图片上。在一定程度上，保证了背景的多样性，减少了行人特征的改变，使得生成数据集的样本更加可靠。

1.2 滑动窗口对准法

在 1.1 中，本文提取到了行人的 Mask，Mask 为只有 0 和 1 组成的二值矩阵。数据集中有一部分图像 Mask 在样本图像中的面积比例较小，即为数据集中行人与行人检测区域未对准的图片。这类样本会导致神经网络进行特征提取阶段造成一部分特征图无效的情况，不仅产生了噪声，也造成了计算浪费。因此本文提出一种基于语义分割的 SWA 对该情况进行优化。

行人数据集具有很好的上下文特性，数据集中每一张行人图像都具有在竖直方向上的顺序性，即按照从上到下，头部、躯干、腿、脚的顺序。因此，将提取的 Mask 图像进行分条匹配操作，在匹配后保留原始行人比例，使行人的每部

分得到更好的对准。设未对准的图像为 $\boldsymbol{A}=\{a_i\}$,$i=1,2,3,\cdots,N$,i 为行人编号,N 为图像数量,m 和 n 为图像行和列的大小,K 为分块数,k 为图片自上而下的编号,$k=1,2,3,\cdots,K$。S_k 为 Mask 在横条 k 中的面积。新的起始块 k_{head} 与结束块 k_{head} 如下

$$k_{\text{head}}=\min_k\left(S_k>\frac{mn}{\eta}\right),k_{\text{feet}}=\min_k\left(S_k>\frac{mn}{\theta}\right) \tag{6}$$

从而可获得滑动窗口 $\boldsymbol{W}_i$,在新的起始块和结束块的横向进行滑动,S_w 为 Mask 在滑动窗口中的面积,大小为高 $(k_{\text{feet}}-k_{\text{head}}+1)\times m/K$,宽 $(k_{\text{feet}}-k_{\text{head}}+1)\times N/K$。因此 $\boldsymbol{W}_i$ 为滑动窗口截取的区域

$$\boldsymbol{W}_i=\min_{\left(x,\frac{m}{K}(k_{\text{head}}-1)\right)}S_w,x=i\times step,i=0,1,2,\cdots \tag{7}$$

在使用滑动窗口机制时,使用行人在检测限定框中的身高作为第一对准标度,滑动窗口的长宽比与原样本保持一致。并且在对行人样本进行语义分割操作时,一些图像存在着行人头部和足部与背景颜色相近的情况,在头部和足部位置会存在着一定的噪声。为了保留头部和足部的特征,分块后的区域可以较好保留头部和足部的特征信息,这样既可以保证行人不会因为放大造成比例失调,也可以保证行人的特征在裁剪中得以保留。

1.3　双路特征扩展网络

在模型训练部分,本文将行人再识别问题当作一个分类任务,由于生成图和原数据集中的图片都有标签(Label),所以使用身份嵌入(ID-discriminative embedding, IDE)的方式来训练行人再识别的卷积神经网络,最后通过 Soft-Max Loss 进行分类,N 为类别数,C 为数据集种类。在卷积神经网络部分,使用已经预训练的 ResNet-50 框架,训练时将生成图和原数据分为 2 个分支进行训练,并将两个网络训练的参数进行共享,在此称为双路特征扩展网络(double-path feature augmentation net, DFANet)。由于已经对数据集行人样本扩充和对准,有一部分背景面积较大的样本行人区域会被放大,放大的图片所含语义信息较少,会损失图像中的细节。ResNet-50 结构的低层可以对整体图像的细节特征进行保留,但是缺乏对语义特征的表现,所以将高层次特征和低层次特征进行融合,增加低层次特征分支,将低层次中的特征提取出来,使高层次特征与低层次特征进行互补,从而进一步改善图像检索任务,获得较好的效果。

对于原数据集分支,本文采用交叉熵损失函数,对于每一个样本 l_i,i 为行人 ID,计算如下

$$l_i=-\log p(i)\,q(i),q(i)=\begin{cases}1,i=y\\0,\text{others}\end{cases} \tag{8}$$

式中　$p(i)$ 为 SoftMax 对于标签的预测概率,y 为该样本的真实标签,$q(i)$ 为真实分布,通过此方法就可得到原数据分支的损失值。

为防止对于真实数据集的训练造成一定的影响,在生成图部分,本文使用了标签平滑正则化方法(label smoothing regularization, LSR),其交叉熵公式为

$$L_{\text{LSR}}=-(1-\varepsilon)\,p(y)-\frac{\varepsilon}{C}\sum_{i=1}^{C}\log p(i) \tag{9}$$

当预测图像和标签相符时,参数 ε 在真实分布函数中加入了噪声,抑制了正负样本在输出中的差值。降低了 $p(y)$ 到 $p(i)$ 的误差,抑制了在生成集上过拟合的现象。

由于样本中的语义信息是完整的且连续的,如果需要提高图像中某个区域所占的权重,则需要以语义块的方式进行丢弃,因此本文使用了 DropBlock(DB) 结构,可以促进神经网络去学习更加具有鲁棒性的特征,计算为

$$\gamma=\frac{(1-keep\_prob)\,feat\_size^2}{block\_size^2\,(feat\_size-block\_size)^2} \tag{10}$$

DropBlock 有 2 个重要参数: 1) 丢弃单元的大小,即 block_size; 2) 丢弃概率,即 γ,控制着有多少个激活单元会被丢弃。*keep_prob* 为保留激活单元的概率,*feat_size* 为特征图的尺寸。

在训练阶段,提取 DropBlock 上一层的特征图,在此图上生成一个(*feat_size-block_size* + 1) 的正方形采样区域,在这个采样区域上按照关于参数的伯努利分布生成中心点,丢弃 *block_size* 大小的正方形区域,即将这个区域置零。

2　实验与结果评价

2.1　实验数据集

Market-1501 数据集[7] 从 6 个未重叠拍摄场景的摄像头收集了 1 501 名行人的样本 32 668 张,其中,训练样本 12936 张,训练行人类别 751 名,测试样本 19732 张,测试行人类别 750 名,索引样本 3 368 张。在训练集中,平均每名行人有 17.2 张图片,测试集中平均每名行人有 26.2 张图片。Duke-MTMC-reID 数据集[8] 从 8 个未重叠覆盖的摄像头收集了 1812 名行人的样本,其中,有 1404 名行人的图像是从多个角度拍摄的。训练集和测试集包都是 702 人,训练集包含 16522 张图片,测试集包含 17661 张图片,索引图片2228 张。本文将在这两个数据集上对提出的算法进行实验,并与当前主流行人再识别算法进行对比。

2.2　结果评价

表 1 记录了增加 SWA 后在 Market-1501 和 Duke-MTMC-reID 数据集上的表现。在增加 SWA 模块和应用了 DFANet 后,Rank-1 相较于 IDE 模型在 Market-1501 和 Duke-MTMC-reID 数据集上 Rank-1 分别提升了 1.2 % 和 1.3 %。可以证明,对准后的行人样本可以更加贴合行人检测限定框,使得卷积神经网络更加直接地提取出行人有效区域的特征,提高了计算的准确率。

表 1 DFANet 与 SWA 在 Market-1501 和 Duke-MTMC-reID 上的表现 %

方法	Market-1501			Duke-MTMC-reID		
	Rank-1	Rank-5	mAP	Rank-1	Rank-5	mAP
IDE[3]	85.6	94.5	65.1	72.3	84.0	51.8
DFANet + SWA	86.7	94.8	66.1	73.6	85.1	54.7

如表 2,在增加 PFRGAN 生成的补充数据集后,使用 2.3 中的训练策略,Rank-1 指标分别在 Market-1501 和 Duke-MTMC-reID 数据集上提升了 2.2 % 和 2.6 %。经统计,Rank-5 准确率要远大于 Rank-1,在 Market-1501 数据集上相差 6 % 左右,在 Duke-MTMC-reID 数据集上相差 10 % 左右,说明本文算法得到的匹配排序序列已经将较好的结果放入了排序较为靠前的部分,但是在第一匹配率上较为不理想,所以增加了重排序[9] 操作,使得 Market-1501 与 Duke-MTMC-reID 的 Rank-1 分别提升 2.5 % 和 1.3 %,在两个数据集上 Rank-1 分别获得 91.4 % 和 81.1 % 的识别率。

表 2 本文方法在数据集 Market-1501 和 Duke-MTMC-reID 上的表现 %

方法	Market-1501		Duke-MTMC-reID	
	Rank-1	mAP	Rank-1	mAP
LOMO + XQDA[10]	43.8	22.2	30.8	17.0
IDE[3]	72.5	46.7	65.2	45.9
Re-rank[9]	77.1	63.6	–	–
SVDNet[11]	82.3	62.1	76.7	56.8
GAN[12]	83.9	66.1	67.7	47.1
DFANet + SWA + PFRGAN	88.9	70.7	76.2	54.7
DFANet + SWA + PFRGAN + Rerank	91.4	84.7	81.1	69.3

3 结束语

本文主要研究基于场景迁移与区域对准的行人再识别方法。较好地解决了行人与检测框不贴合的情况和行人再识别训练样本不足的问题。在 Marcket-1501 数据集和 Duke-MTMC-reID 数据集上 Rank-1 指标分别取得了 91.4 % 和 81.1 % 的准确率。此外,本方法也存在着一定的局限性。在使用基于语义分割的滑动窗口对准法时,由于一些未对准的图像存在着人体部分缺失的情况,该方法并不能对行人样本进行特征补全。该不足将会成为未来研究的重点。

参考文献:

[1] 田萱,王亮,丁琪. 基于深度学习的图像语义分割方法综述[J]. 软件学报,2019,30(2):440 - 468.

[2] 张汇,杜煜,宁淑荣,等. 基于 Faster RCNN 的行人检测方法[J]. 传感器与微系统,2019,38(2):147 - 149,153.

[3] ZHENG L, YANG Y, HAUPTMANN A G. Person re-identification: Past, present and future [J]. arXiv preprint arXiv: 1610.02984,2016.

[4] 蒋桧慧,张荣,李小宝,等. 基于特征融合与改进神经网络的行人再识别[J]. 传感器与微系统,2017,36(8):121 - 125.

[5] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks [C]// Proceedings of the IEEE International Conference on Computer Vision,2017: 2223 - 2232.

[6] LIN G, MILAN A, SHEN C, et al. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017: 1925 - 1934.

[7] 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.

[8] ZHENG Z, ZHENG L, YANG Y. Unlabeled samples generated by gan improve the person re-identification baseline in vitro [C]// Proceedings of the IEEE International Conference on Computer Vision,2017: 3754 - 3762.

[9] ZHONG Z, ZHENG L, CAO D, et al. Re-ranking person re-identification with k-reciprocal encoding [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1318 - 1327.

[10] LIAO S, HU Y, ZHU X, et al. Person re-identification by local maximal occurrence representation and metric learning [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2015: 2197 - 2206.

[11] SUN Y, ZHENG L, DENG W, et al. SVDN for pedestrian retrieval [C]// Proceedings of the IEEE International Conference on Computer Vision,2017: 3800 - 3808.

[12] ZHENG Z, ZHENG L, YANG Y. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro [C]// Proceedings of the IEEE International Conference on Computer Vision,2017: 3754 - 3762.

作者简介:

白 健(1990 -),男,通讯作者,硕士研究生,研究方向为行人再识别,E-mail: baijian1990@nuaa.edu.cn。

耿树泽(1989 -),男,博士研究生,研究方向为行人再识别。

岑世欣(1991 -),男,博士研究生,研究方向为图像处理与深度学习。

(上接第 118 页)

[16] WU D H, LIU Z T, WANG X H, et al. Composite magnetic flux leakage detection method for pipelines using alternating magnetic field excitation [J]. NDT & E International,2017,91: 148 - 155.

[17] PHAM H Q, TRINH Q T, DOAN D T, et al. Importance of magnetizing field on magnetic flux leakage signal of defects [J]. IEEE Transactions on Magnetics,2018(99): 1 - 6.

作者简介:

巩文东(1985 -),男,博士研究生,讲师,主要研究领域为电磁无损检测技术,E-mail: 1078@sdp.edu.cn。

杨 涛(1970 -),男,教授,博士生导师,主要研究领域为复合材料成型技术与装备、机电一体化技术等,E-mail: yangtao626@163.com。

连 超(1983 -),男,副教授,主要研究领域为深海科考设备。

张宏杰(1978 -),男,副教授,主要研究领域为电磁无损检测等。