

基于DCN的无监督域自适应行人重识别方法

杨海伦, 王金聪, 任洪娥, 陶锐

引用本文:

杨海伦, 王金聪, 任洪娥, 等. 基于DCN的无监督域自适应行人重识别方法[J]. 液晶与显示, 2021, 36(11): 1573–1582.

YANG Hai-lun, WANG Jin-cong, REN Hong-e, et al. DCN-Based unsupervised domain adaptive person re-identification method[J]. *Chinese Journal of Liquid Crystals and Displays*, 2021, 36(11): 1573–1582.

在线阅读 View online: <https://doi.org/10.37188/CJLCD.2021-0095>

您可能感兴趣的其他文章

Articles you may be interested in

GAN网络混合编码的行人再识别

Person re-identification of GAN network hybrid coding

液晶与显示. 2021, 36(2): 334–342 <https://doi.org/10.37188/CJLCD.2020-0167>

基于深度卷积生成对抗网络的图像识别算法

Image recognition algorithms based on deep convolution generative adversarial network

液晶与显示. 2020, 35(4): 383–388 <https://doi.org/10.3788/YJYXS20203504.0383>

多粒度特征融合的行人再识别研究

Multi-granularity feature fusion for person re-identification

液晶与显示. 2020, 35(6): 555–563 <https://doi.org/10.3788/YJYXS20203506.0555>

基于条件生成对抗网络的手写数字识别

Handwritten digit recognition based on conditional generative adversarial network

液晶与显示. 2020, 35(12): 1284–1290 <https://doi.org/10.37188/YJYXS20203512.1284>

基于密集连接生成对抗网络的图像颜色迁移

Image color transfer with dense connections generative adversarial networks

液晶与显示. 2020, 35(3): 272–280 <https://doi.org/10.3788/YJYXS20203503.0272>

文章编号:1007-2780(2021)11-1573-06

基于 DCN 的无监督域自适应行人重识别方法

杨海伦¹, 王金聪^{2,3}, 任洪娥^{1,3*}, 陶 锐^{1,4}

(1. 东北林业大学 信息与计算机工程学院, 黑龙江 哈尔滨 150040;

2. 东北林业大学 机电工程学院, 黑龙江 哈尔滨 150040;

3. 黑龙江林业智能装备工程研究中心, 黑龙江 哈尔滨 150040;

4. 呼伦贝尔学院, 内蒙古 呼伦贝尔 021008)

摘要:为解决无监督行人重识别研究中存在的遮挡、域间及相机间风格差异较大等问题,本文提出基于可变形卷积的无监督域自适应模型。针对特征提取过程中的遮挡问题提出基于可变形卷积的 CNN 模型;在预训练阶段提出应用 SPGAN 直接减小域间差异,训练过程中提出使用 CycleGAN 生成不同相机风格图像缓解相机风格差异性;提出多损失协同训练的方法实现 CycleGAN 和复用 CNN 模型的迭代优化进一步提高识别准确率。实验结果表明,本文提出的方法在源域 DukeMTMC-reID/Market-1501 和目标域 Market-1501/DukeMTMC-reID 下进行实验,mAP 和 Rank-1 分别达到了 68.7%、64.1%和 88.2%、78.1%。本文所提出的模型有效缓解了行人被遮挡、域间及相机间风格差异较大等问题,与现有方法比较,有更好的识别效果。

关键词:DCN;生成对抗网络;无监督域自适应;多损失训练

中图分类号:TP391.41 **文献标识码:**A **doi:**10.37188/CJLCD.2021-0095

DCN-Based unsupervised domain adaptive person re-identification method

YANG Hai-lun¹, WANG Jin-cong^{2,3}, REN Hong-e^{1,3*}, TAO Rui¹(1. College of Information and Computer Engineering,
Northeast Forestry University, Harbin 150040, China;2. College of Mechanical and Electrical Engineering,
Northeast Forestry University, Harbin 150040, China;3. Heilongjiang Forestry Intelligent Equipment Engineering Research Center,
Harbin 150040, China;

4. Hulunbuir University, Hulunbuir 0210008, China)

Abstract: In order to solve the problems of occlusion, large differences in styles between domains and

收稿日期:2021-04-12;修订日期:2021-05-17.

基金项目:黑龙江省自然科学基金(No.LH2020F040);中央高校基本科研业务费专项资金(No.2572017PZ10)

Supported by Natural Science Foundation of Heilongjiang Province(No.LH2020F040);Central University
Basic Research Funding for Specialized Project(No.2572017PZ10)

* 通信联系人,E-mail:nefu_rhe@163.com

cameras in the research of unsupervised person re-recognition, this paper proposes an unsupervised domain adaptive model based on deformable convolution. Aiming at the occlusion problem in the feature extraction process, a CNN model based on deformable convolution is proposed. In the pre-training stage, it is proposed to apply SPGAN to directly reduce the difference between domains. During the training process, it is proposed to use CycleGAN to generate images of different camera styles to alleviate the problem of camera style differences. A multi-loss collaborative training method is proposed to realize the iterative optimization of CycleGAN and re-used CNN models to further improve the recognition accuracy. The experimental results show that the method proposed in this paper is tested in the source domain DukeMTMC-reID/Market-1501 and the target domain Market-1501/DukeMTMC-reID, and mAP and Rank-1 reach 68.7%, 64.1% and 88.2%, 78.1%, respectively. The model proposed in this paper effectively alleviates the problems of pedestrians being occluded, and large differences in styles between domains and cameras. Compared with the existing methods, it has a better recognition effect.

Key words: DCN; generative adversarial networks; unsupervised domain adaption; multi-loss training

1 引言

行人重识别方法^[1]旨在依据行人的一张查询图像,在大规模图像数据中检索出同一行人最相似的前 n 张图片。基于深度学习理论^[2-4]的行人重识别方法研究主要分为有监督学习的行人重识别^[5]和无监督学习的行人重识别^[6-13]两类。随着行人重识别领域研究的不断深入,数据集规模和数量也在不断扩大,但是其需要耗费大量人力和时间进行数据集的标注。因此,近些年的行人重识别研究逐渐开始转向无监督域自适应(Unsupervised domain adaption, UDA)行人重识别方法。

卷积神经网络^[3](Convolution Neural Networks, CNN)作为深度学习的重要组成部分,它凭借特征提取的巨大优势受到广泛关注,但在卷积操作中也会导致图像中的许多重要信息丢失。可变形卷积网络^[14-15](Deformable Convolutional Networks, DCN)通过加入可变形卷积核模块和可变形兴趣区域(ROI)池化模块,提高了CNN的特征提取能力。此类方法可以在特征提取过程中定位到感兴趣的采样点位置,并通过反向传播进行网络训练。

2018年,Deng等^[6]提出的一种UDA风格转换模型(Similarity Preserving Generative Adversarial Network, SPGAN)和Wei等^[7]提出的行人风格迁移模型(Person Transfer Generative Ad-

versarial Network, PTGAN)实现了风格迁移应用于无监督域自适应问题。CamStyle^[8]通过学习相机的不变性对相机风格进行迁移,在一定程度上解决了过拟合问题。然而,这些方法比较依赖生成图片的质量,没有关注生成对抗网络^[16](Generative Adversarial Network, GAN)与CNN间的迭代优化、影响域适应能力。2019年,Fu等^[9]提出一种基于聚类方法的UDA模型SSG,依据行人的全局和局部特征寻找目标域样本间相似性进行聚类,然而该模型没有从根本上解决域间差异性问题。Song等^[10]提出的UDA模型,在一定程度上提高了域适应能力,但没有过多关注域间、相机间风格差异问题。本文基于Song等的UDA模型提出基于DCN的域自适应模型,其UDA模型的训练过程主要分为3个步骤:首先,在已标注的源域上进行预训练;之后,基于聚类算法执行伪标签的预测,依据伪标签进行特征学习;最后,通过两个步骤的反复迭代进行训练。由于域间、相机间的风格差异性较大,使聚类生成的伪标签变得不准确,影响域自适应能力。因此,缓解域间、相机间风格差异所带来的负面影响是UDA成功的关键。

本文提出将可变形卷积加入到CNN模型中,代替传统的CNN模型,有效缓解了UDA中存在的遮挡问题。预训练阶段提出应用SPGAN直接减小域间差异,训练过程中应用CycleGAN^[6]生成不同相机风格图像解决相机风格差异性问题。为了进一步提高识别准确率,提

出多损失协同训练的方法实现 CycleGAN 和复用 CNN 模型的迭代优化。

2 研究技术与模型架构

2.1 可变形卷积网络结构

本文提出的可变形卷积网络结构主要是将可变形卷积嵌入到 ResNet-50^[17] 网络框架中,替换其中固定卷积层和池化层,更换为可变形卷积层和可变形池化层,提高特征学习能力。二维卷积主要包含两个步骤,首先使用规格固定的卷积核 R 在特征图 x 上进行滑动采样,之后利用采样点乘权重 w 求和。其中 R 决定感受野大小和扩张,如: $R = \{(-1, -1), (-1, 0), \dots, (0, 1), (1, 1)\}$, 传统的卷积过程^[3] 如公式(1)所示:

$$y(P_0) = \sum_{P_n \in R} w(P_n) \cdot x(P_0 + P_n), \quad (1)$$

P_n 是 R 中所列位置的枚举值, P_0 是输出在特征图上的位置。DCN 的卷积过程则在规则网格 R 中增加一个偏移量来提高特征提取的自适应能力,同样位置的 P_0 则需要加一个偏移量 ΔP_n (第 n 个采样点的偏移),采样位置变为不规则可形变的区域。因此,可变形卷积过程^[14] 如公式(2)所示:

$$y(P_0) = \sum_{P_n \in R} w(P_n) \cdot x(P_0 + P_n + \Delta P_n). \quad (2)$$

由于特征图上每个位置的取值需要整数,而公式(2)中 ΔP_n 是非整数,故采取双线性插值的方式处理(寻找坐标距离最近的 4 个像素点,计算该坐标点的值)。

可变形卷积与传统卷积的对比如图 1 所示,传统卷积只需要训练网格点的相应位置点权重,可变形卷积通过增加额外参数来训练卷积网格窗口形变。因此,可变形卷积的采样点选取可以通过学习得到。图 1(a)为传统卷积的规则采样点位置,1(b)为可变形卷积的采样点位置。

传统的卷积过程主要分为两个步骤:首先利用已定义感受野大小和扩张的网格窗口卷积核 R 在特征图上滑动,之后根据窗口权重相乘求和。

相对于传统卷积过程,可变形卷积过程^[14] 如图 2 所示。使用 3×3 的可变形卷积执行卷积操作,在训练过程中,同时学习用于生成输出特征的

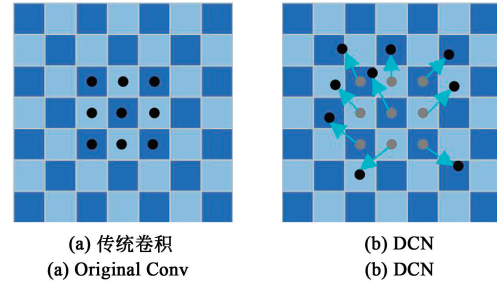


图 1 传统卷积和可变形卷积对比图

Fig.1 Comparison of traditional convolution and DCN

卷积核和偏移量。偏移量特征与输入特征的分辨率相同,通道数需要是采样点个数的 2 倍,即每个位置需要有 x, y 二维的偏移量。可变形卷积单元中增加的偏移量是网络结构的一部分,使用平行的传统卷积单元计算,因此可以通过反向传播实现端到端的训练,偏移量的增加使特征提取过程偏向于感兴趣的区域,即达到采样点位置的自适应变化。

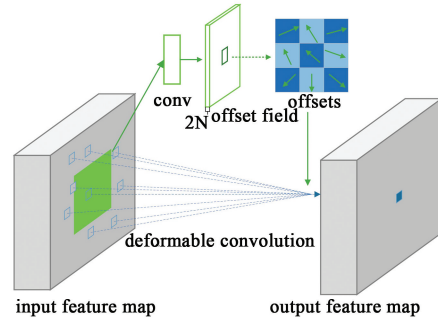


图 2 3×3 可变形卷积

Fig.2 3×3 deformable convolution

可变形池化过程类比于可变形卷积过程,同样采取增加偏移的方法,实际处理过程^[15] 如图 3 所示。给定输入特征图 x , 首先通过一个兴趣区域池化将它划分为 $k \times k$ 个容器, 经过兴趣区域池化后的特征图通过一个全连接层产生归一化偏移, 全连接层的输出为容器个数的 2 倍。

由于需要偏移和兴趣区域尺寸相匹配, 故对偏移做微调, 偏移量可转化为 ΔP_k , 单个容器的输出特征可以表示为公式(3)^[15]:

$$y(k) = \sum_{j=1}^{n_k} x(p_{kj} + \Delta p_k) / n_k, \quad (3)$$

其中: P_{kj} 表示第 k 个容器的 j 个采样点坐标, $x(P_0 + \Delta P_k)$ 通过双线性插值得得, n_k 代表第 k 个

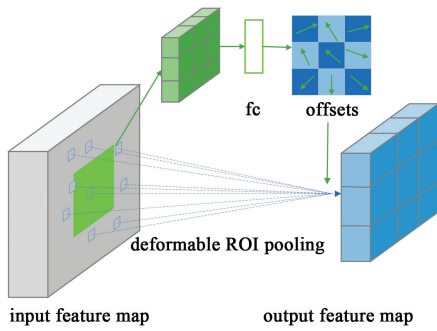


图 3 可变形池化

Fig.3 Deformable ROI pooling

容器的采样点数量, ΔP_k 表示第 k 个容器的偏移量。

本文采取的特征提取网络架构是将可变形卷积结构和可变形池化结构嵌入 ResNet-50 中, 将 ResNet-50 网络中 conv3~conv5 中所有的 3×3 传统卷积替换为可变形卷积, 并于相应位置加入可变形池化层。

2.2 基于源域的有监督预训练

2.2.1 SPGAN 实现域间风格迁移

SPGAN 以 CycleGAN 作为基础框架, 能够解决无监督行人重识别中存在的域间差异问题。首先, 使用两个镜像对称的生成对抗网络搭建环形网络, 实现图像的风格转换; 其次, 加入新的损失函数辅助生成器生成目标域风格的行人图像; 此外, 加入孪生网络及对应的一致性损失函数, 保证转换之后关键身份信息不丢失。

2.2.2 预训练损失函数

使用分类损失函数 L_{cls} 与改进的三元组损失函数 L_{BH} 协同训练, 对预训练模型微调, 总损失函数为:

$$L_{\text{src}} = L_{\text{cls}} + L_{\text{BH}}, \quad (4)$$

分类损失函数^[11]定义为:

$$L_{\text{cls}} = -\frac{1}{N} \sum_{i=1}^N \log p(y_i^a | x_i^a), \quad (5)$$

其中: N 表示批量所包含的图像数量, i 代表图像索引。改进的三元组损失函数^[18]被定义为:

$$L_{\text{BH}} = \sum_i^N [\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + m]_+, \quad (6)$$

其中: x_i^a 、 x_i^p 、 x_i^n 分别为锚样本、正样本及负样本的特征表示, m 为软化的自适应边缘参数。

2.2.3 预训练过程

应用 SPGAN 实现源域风格迁移, 使用基于 DCN 的 CNN 模型在源域上进行预训练。为了提高收敛速度, 梯度变化更加稳定, 在可变形池化层后加入 BN 层; 为了抑制过拟合, 使用 Adam 优化器代替 SGD 优化器。

2.3 基于 DCN 的无监督域自适应模型

基于 DCN 的无监督域自适应模型如图 4 所示。首先, 使用预训练好且基于 DCN 的 CNN 模型进行特征提取, 通过距离度量获得源域和目标域之间的距离矩阵 $\mathbf{M}$, 使用 DBSCAN^[10] 算法进

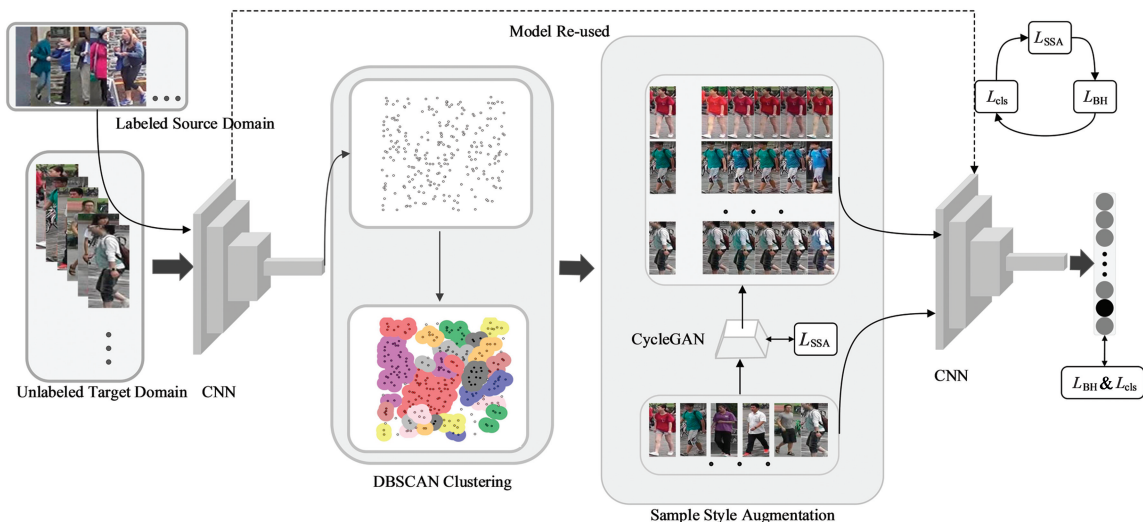


图 4 基于 DCN 的域适应模型框架图

Fig.4 Framework diagram of domain adaptation model based on DCN

行聚类以赋予无标签数据集标签数据。之后,基于目标域样本应用 CycleGAN,生成不同相机风格的图像数据扩充目标域样本。将生成样本和目标域样本一同输入到复用的 CNN 模型(即共享权重),依据生成的伪标签,通过多损失协同训练方式实现 CycleGAN 和复用 CNN 模型的迭代优化。

2.3.1 DBSCAN 聚类方法生成伪标签

DBSCAN 聚类方法是一种基于密度的聚类算法,此方法由特征样本分布紧密程度决定。将紧密相连的样本划分为同一簇,将不同簇划分为不同类别,从而得到不同的聚类结果。

在本框架中,首先利用在源域预训练好的 CNN 模型提取无标签目标数据集的特征矩阵,通过距离度量获得源域数据集和目标数据集之间的距离矩阵 $\mathbf{M}$,然后依据 DBSCAN 聚类方法对其聚类,依据聚类结果赋予无标签的目标域生成伪标签。

$$D = \text{DBSCAN}(\mathbf{M}, \tau, n), \quad (7)$$

其中: τ 为半径,与距离矩阵 $\mathbf{M}$ 共同描述样本分布紧密程度; n 为密度阈值。

2.3.2 样本风格增强

受源域与目标域的域间差异问题影响,基于 DBSCAN 聚类算法生成的伪标签存在部分噪声。除此以外,目标域训练样本的数量限制也导致每个聚类中样本多样性较低,这就导致模型域自适应能力低、学习能力差。针对这个问题,本文提出了数据样本风格增强(Sample Style Augmentation, SSA)的方法,应用 GAN 扩充目标域解决样本不足的问题。使用 GAN 生成的新样本需要满足条件:保留原始身份 ID 的同时,从现有的行人生成新的行人数据,克服相机风格差异。

循环生成对抗网络^[6](Cycle-consistent Adversarial Networks, CycleGAN)的网络结构包括两个生成器 $G(X \rightarrow Y), F(Y \rightarrow X)$, 两个判别器 D_X, D_Y , 旨在实现域 X 到 Y 的普适映射,而不是单一的 x 到 y 的映射,能够满足不同相机风格转换的要求。因此,本文使用 CycleGAN 产生不同相机风格的新行人数据扩充目标域样本数据。例如,数据集中含有 N 个相机,首先进行 CycleGAN 模型训练,该模型能够实现不同相机图

像之间的转换,基于伪标签生成 $N-1$ 张不同相机风格的行人数据图像。通过数据增强,目标域每个聚类样本数量变为原来的 $N-1$ 倍。

2.3.3 多损失协同训练

如图 5 所示, CycleGAN 样本风格增强模型包括两个生成器 $G(X \rightarrow Y), F(Y \rightarrow X)$, 两个判别器 D_X, D_Y 。

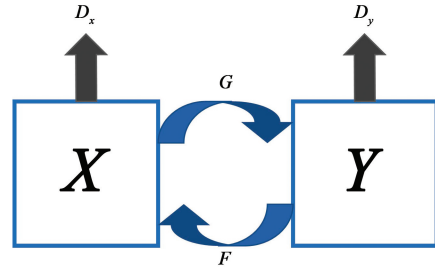


图 5 CycleGAN 网络结构

Fig.5 CycleGAN network structure

该部分选取的损失函数^[6] L_{SSA} 为:

$$L_{SSA} = L_{GAN}(G, D_Y, X, Y) +$$

$$L_{GAN}(F, D_X, Y, X) + L_{cyc}(G, F), \quad (8)$$

利用对抗损失 L_{GAN} 使生成图像在分布上更加接近于目标图像,其中 L_{GAN} 被定义为:

$$L_{GAN}(G, D_Y, X, Y) = E_{y \sim p_{data}(y)} [\log(D_Y(Y))] +$$

$$E_{x \sim p_{data}(x)} [\log(1 - D_Y(G(x)))] \quad (9)$$

使用循环一致性损失 L_{cyc} 防止学习到的映射 G 和 F 产生矛盾, L_{cyc} 可以表示为:

$$L_{cyc}(G, F) = E_{x \sim p_{data}(x)} [\|F(G(x)) - x\|_1] +$$

$$E_{y \sim p_{data}(y)} [\|G(F(y)) - y\|_1] \quad (10)$$

为了进一步提高模型准确率,实现 CycleGAN 模型和复用 CNN 的迭代优化,本文提出改进的三元组损失函数 L_{BH} 、分类损失函数 L_{cls} 和 CycleGAN 损失函数 L_{SSA} 协同训练的多损失协同训练(Multi-loss Training, ML)方法,总体模型损失函数为:

$$L_{loss} = L_{cls} + L_{BH} + L_{SSA}, \quad (11)$$

L_{cls}, L_{BH} 如公式(5)、(6)所示。

3 实验与结果分析

本文在源域 DukeMTMC-reID/Market-1501 和目标域 Market-1501/DukeMTMC-reID 下对本文所提出的模型进行消融实验、对比实验,来验

证本文算法的有效性。

Algorithm 1 模型训练过程

Input: 源域数据集 S , 目标域数据集 T

Output: CNN 模型 F

```

1: CNN 模型  $F$  在源域进行预训练
2: for 每次进行聚类迭代 do:
3:   特征提取:  $F(T), F(S)$ ; 求距离矩阵  $\mathbf{M}$ 
4:   聚类:  $D = \text{DBSCAN}(\mathbf{M}, \tau, N)$ 
5:   for 从  $T$  中取出一个 batch 的样本 do:
6:     样本风格增强, 数据扩充
7:     多损失协同训练, 优化
8:   end for
9: end for
10: return CNN 模型  $F$ 

```

3.1 数据集设置和评估方法

本文采用两个无监督行人重识别领域常用的公开数据集 Market-1501 和 DukeMTMC-reID, 使用累积匹配特性 (Cumulative match characteristic curve, CMC) 中的 Rank-1、Rank-5、Rank-10 和平均精度 (mean average precision, mAP) 作为评价标准。

对于准确率 mAP 的描述如下: 假设包含 N 张图像数据, q_i 表示第 i 张检索图像, 待检测图片集记作 Q ; g_i 表示图片库的第 i 张图片, 图片库数据集记为 G ; q_i 对应行人编号在 G 出现的次数记为 K_{q_i} 。对每张待检索图片的查询过程为: 对待检索图片进行特征提取后, 与 G 中每张图片进行特征比对, 按照距离对图片库升序排列, 排列后的图片库记为 G_{q_i} , q_i 命中图片所组成的数据集

记为 G'_{q_i} , 其中每张图片在 G_{q_i} 的排位记为 r_j , 在 G'_{q_i} 中的排位为 r'_j , 对待检索数据集图像重复这个过程, 得到的准确率计算公式为:

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{K'_{q_i}} (r'_j / r_j) \quad (12)$$

对于 n 位命中率 Rank- n 的描述如下: q_i 对应的 G_{q_i} 在前 n 位包含同一行人, 则 q_i 满足 n 位命中条件, 待检索图像集中所有满足 n 位命中的图片数量记为 N' , 则 Rank- n 可表示为

$$\text{Rank-}n = N' / N \quad (13)$$

Market-1501 数据集共有 6 个相机、1 501 名行人, 共包含了 32 688 张图像数据。其中训练集涵盖 751 人, 共包含 12 936 张图像数据; 测试集涵盖 750 人, 共包含 19 732 张图像数据。DukeMTMC-reID 共有 8 个相机、1 404 个行人, 共包含了 36 411 张图像数据。其中训练集涵盖 702 人, 共包含 16 522 张图像数据; 测试集涵盖 702 人, 共包含 19 889 张图像数据。

3.2 实验环境及参数配置

本实验采用两个 GeForce GTX 2080 Ti GPU, 操作系统为 Ubuntu16.04, 基于 python3.6 和 pytorch1.7.1 完成程序编写。基于 DCN 的域自适应模型的迭代次数是 30 次, 每次迭代 70 个 epoch, 批量大小设置为 128。选用 Adam 优化器, 设置学习率为 6×10^{-5} , 动量设定为 0.9。

3.3 消融实验

为了验证本文提出模型的有效性, 在两个无监督公开的行人数据集上进行不同域自适应实验。分别以 Market-1501 为源域, 在 DukeMTMC-reID 上进行无监督实验; 以 DukeMTMC-reID 作为源域, 在 Market-1501 上进行无监督实验。本文以 Song 等^[10]提出的自训练框架方法作为 Baseline, 消融实验对比结果如表 1 所示。

表 1 消融实验对比结果

Tab.1 Comparison results of ablation experiments

(%)

算法	DukeMTMC-reID → Market-1501				Market-1501 → DukeMTMC-reID			
	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10	mAP
DirectTransfer	53.3	70.6	76.9	24.8	38.7	55.9	62.5	23.2
Direct Transfer + DCN	54.6	71.0	77.9	25.8	39.9	56.5	64.2	25.3
Baseline ^[10]	76.8	87.2	91.8	56.8	71.8	82.8	83.4	51.4

续 表

(%)

算法	DukeMTMC-reID→Market-1501				Market-1501→DukeMTMC-reID			
	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10	mAP
Baseline + DCN	80.7	90.1	93.6	61.2	74.2	85.6	86.9	55.2
Baseline + DCN + SSA	85.4	93.8	96.0	65.7	76.8	87.4	88.7	59.8
Baseline + DCN + SSA + ML	88.2	94.7	96.8	68.7	78.1	89.3	91.2	64.1

首先,针对基于可变形卷积的 CNN 模型有效性进行实验,验证可变形卷积对特征提取能力的提高。从表 1 中可以看到,在 DukeMTMC-reID/Market-1501 > Market-1501/DukeMTMC-reID 实验中:使用基于 DCN 的 CNN 模型代替 Baseline 中的传统 CNN 模型,mAP 提高了 4.4%、3.8%,Rank-1 提高了 3.9%、2.4%。验证了基于 DCN 的 CNN 模型较传统 CNN 模型在特征提取方面的优越性。为了进一步验证可变形卷积对 CNN 模型特征提取能力的提高,对比 DCN-ResNet50 和 ResNet50 特征提取能力,使用 CAM 实现特征提取过程中第 5 阶段(即 layer4)的可视化热力图,其部分可视化结果如图 6、7 和 8 所示。结合可视化结果和表 1 中的实验结果可知,可变形卷积的增加使得特征提取部位更加偏向于我们感兴趣的区域,即达到采样点位置的

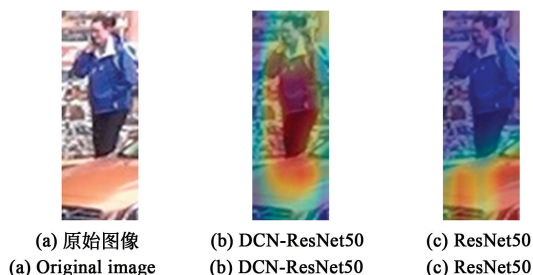


图 6 第 5 阶段输出特征可视化热力图 1

Fig.6 Fifth stage output feature visualization heat map 1

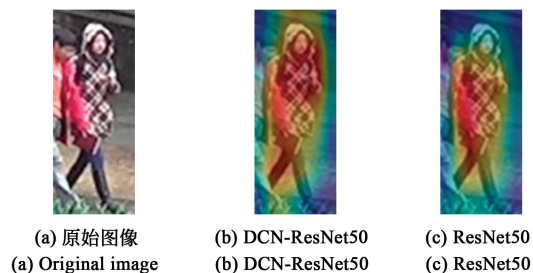


图 7 第 5 阶段输出特征可视化热力图 2

Fig.7 Fifth stage output feature visualization heat map 2

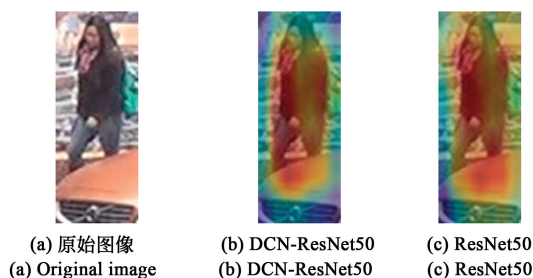


图 8 第 5 阶段输出特征可视化热力图 3

Fig.8 Fifth stage output feature visualization heat map 3

自适应变化,有效缓解遮挡、分辨率低等噪声问题。

由表 1 可知,加入样本风格增强后,使得图像数据得到扩充,提高了基于 DCN 的 CNN 模型学习能力。实验结果表明,使用样本风格增强后,mAP 提高了 4.5%、4.6%,Rank-1 提高了 4.7%、2.6%。证明了使用 CycleGAN 生成不同相机风格下的行人图像,实现了样本数据增强,并且在一定程度上抑制了 DBSCAN 聚类伪标签噪声问题,缓解了不同相机间的风格差异问题。

为了验证多损失协同训练方法,使用 CycleGAN 进行样本风格增强,并重用基于 DCN 的 CNN 模型,利用多损失协同训练方法实现 CycleGAN 模型和复用 CNN 模型的迭代优化。实验结果表明,使用多损失协同训练后,mAP 提高了 3.0%、4.3%,Rank-1 提高了 2.8%、1.3%。结果表明,多损失协同训练方法对于模型间的迭代优化训练过程有一定程度的提高,可获得更好的行人重识别效果;验证了 CycleGAN 损失函数(L_{SSA})在模型训练中的作用,能够提高 CycleGAN 生成图像数据的质量。

3.4 本文方法与现有主流方法对比实验

将本文方法与主流无监督行人重识别方法进行比较,表 2 给出以 Market-1501/DukeMTMC-reID 为源域,以 DukeMTMC-reID/Market-1501 为

表 2 算法结果对比
Tab.2 Comparison of algorithm results (%)

算法	DukeMTMC-reID→Market-1501				Market-1501→DukeMTMC-reID			
	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10	mAP
MMFA ^[18]	56.7	75.0	81.8	27.4	45.3	59.8	66.3	24.7
SPGAN + LMP ^[6]	57.7	75.8	82.4	26.7	46.4	62.3	68.0	26.2
CamStyle ^[8]	58.8	78.2	84.3	27.4	48.4	62.5	68.9	25.1
CR-GAN ^[20]	77.7	89.7	92.7	54.0	68.9	80.2	84.7	48.6
PCB-PAST ^[21]	78.4	—	—	54.6	72.4	—	—	54.3
SSG ^[9]	80.0	90.0	92.4	58.3	73.0	80.6	83.2	53.4
ACT ^[12]	80.5	—	—	60.6	72.4	—	—	54.5
ECN++ ^[22]	84.1	92.8	95.4	63.8	74.0	83.7	87.4	54.4
SNR ^[23]	82.8	—	—	61.7	76.3	—	—	58.1
MPLP ^[13]	84.4	92.8	95.0	60.4	72.4	82.9	85.0	51.4
DAAM ^[24]	86.4	—	—	67.8	77.6	—	—	63.9
AD-Cluster ^[11]	86.7	94.4	96.5	68.3	72.6	82.5	85.5	54.1
This paper	88.2	94.7	96.8	68.7	78.1	89.3	91.2	64.1

目标域的比较结果,“—”表示对应方法未在数据集上实现,其中包括:依据图像属性的 MMFA^[19] 无监督行人重识别方法;基于域适应的 7 种无监督行人重识别方法 SPGAN+LMP^[6]、CamStyle^[8]、AD-Cluster^[11]、ACT^[12]、CR-GAN^[20]、PCB-PAST^[21]、ECN++^[22];基于全局与局部特征多重聚类的 SSG^[9]方法;针对无监督行人重识别的多标签分类来预测多类别标签,学习 ID 特征 MPLP^[13]方法;基于风格正则化的 SNR^[23]方法;基于注意力机制的域对齐方法 DAAM^[24]。

本文方法高于所比较的主流竞争方法,在 Market-1501 上测试,Rank-1 准确率为 88.2%,mAP 为 68.7%。相较于目前性能优异的 AD-Cluster、DAAM、MPLP、ACT、ECN、SNR 等均有不同幅度的提升。在 DukeMTMC-reID 上进行测试,本文方法获得的 Rank-1 准确率为 78.1%,mAP 为 64.1%。均高于目前主流方法,且较 AD-Cluster 的 mAP 精度提高了 10%,Rank-1 精度提高了 5.5%,验证了本文方法的有效性。

总体来说,本文在 CNN 模型中嵌入可变形卷积,提高了 CNN 模型对行人的特征提取效果,

缓解了数据集中普遍存在的遮挡、分辨率低等问题;使用样本风格增强和多损失协同训练方法,缓解了不相交相机视图下的类内变化大和类间模糊的问题。本文提出方法实现了无监督行人重识别跨域的域自适应,抑制了跨域所导致的性能大幅下降问题,具备在无标注目标数据集上的泛化能力。

4 结 论

本文针对传统 CNN 模型特征提取过程的遮挡、分辨率低等噪声问题,提出基于可变形卷积的 CNN 模型,使特征提取部位更偏向于感兴趣的区域,即达到采样点位置的自适应变化。针对域间、不同相机风格间差异性较大的问题,在预训练阶段提出应用 SPGAN 直接减小域间风格差异性,训练过程中提出应用 CycleGAN 生成不同相机风格图像数据解决相机风格差异性问题。为了进一步提升无监督域自适应行人重识别准确率,提出多损失函数协同训练的方法实现 GAN 和复用 CNN 模型的迭代优化过程。在当前无监督行人重识别领域的公开数据集设定前提下,对本文提

出的方法进行验证。实验表明,在源域 DukeMTMC-reID/ Market-1501 和目标域 Market-1501/ DukeMTMC-reID 下实验, mAP 和 Rank-1 分别达到了 68.7%、64.1% 和 88.2%、78.1%。实验结果表明,本文提出的基于 DCN 的域自适应模型,

能够有效解决数据集中存在的遮挡、分辨率低等噪声问题;能够降低域间、不同相机风格间的差异性;有效提升无监督域自适应行人重识别的准确率。但仍有进步空间,未来工作可以在现有方法基础上进一步改进以提升特征聚类的准确度。

参 考 文 献:

- [1] YE M, SHEN J B, LIN G J, *et al.* Deep learning for person re-identification: a survey and outlook [J]. *arXiv*: 2001.04193v1.
- [2] 黄乐弘,曹立华,李宁,等.深度学习的空间红外弱小目标状态感知方法[J].中国光学,2020,13(3):527-536.
HUANG L H, CAO L H, LI N, *et al.* Deep learning-based spatial infrared dim and small target state perception method [J]. *Chinese Optics*, 2020, 13(3): 527-536. (in Chinese)
- [3] 李勇,杨德东,毛宁,等.基于卷积神经网络的响应自适应跟踪[J].液晶与显示,2018,33(7):596-605.
LI Y, YANG D D, MAO N, *et al.* Response adaptive tracking based on convolution neural network [J]. *Chinese Journal of Liquid Crystals and Displays*, 2018, 33(7): 596-605. (in Chinese)
- [4] 陈晓冬,艾大航,张佳琛,等.Gabor 滤波融合卷积神经网络的路面裂缝检测方法[J].中国光学,2020,13(6): 1293-1301.
CHEN X D, AI D H, ZHANG J C, *et al.* Gabor filter fusion network for pavement crack detection [J]. *Chinese Optics*, 2020, 13(6): 1293-1301. (in Chinese)
- [5] YADAV A, VISHWAKARMA D K. Person re-identification using deep learning networks: a systematic review [J]. *arXiv preprint arXiv:2012.13318*, 2020.
- [6] DENG W J, ZHENG L, YE Q X, *et al.* Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification [C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018: 994-1003.
- [7] WEI L H, ZHANG S L, GAO W, *et al.* Person transfer GAN to bridge domain gap for person re-identification [C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018: 79-88.
- [8] ZHONG Z, ZHENG L, ZHENG Z D, *et al.* CamStyle: a novel data augmentation method for person re-identification [J]. *IEEE Transactions on Image Processing*, 2019, 28(3): 1176-1190.
- [9] FU Y, WEI Y C, WANG G S, *et al.* Self-similarity grouping: a simple unsupervised cross domain adaptation approach for person re-identification [C]//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul, Korea: IEEE, 2019: 6112-6121.
- [10] SONG L C, WANG C, ZHANG L F, *et al.* Unsupervised domain adaptive re-identification: theory and practice [J]. *Pattern Recognition*, 2020, 102: 107173.
- [11] ZHAI Y P, LU S J, YE Q X, *et al.* Ad-cluster: augmented discriminative clustering for domain adaptive person re-identification [C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020: 9021-9030.
- [12] YANG F X, LI K, ZHONG Z, *et al.* Asymmetric co-teaching for unsupervised cross-domain person re-identification [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(7): 12597-12604.
- [13] WANG D K, ZHANG S L. Unsupervised person re-identification via multi-label classification [C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020: 10981-10990.
- [14] DAI J F, QI H Z, XIONG Y W, *et al.* Deformable convolutional networks [C]//*Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice: IEEE, 2017: 764-773.
- [15] ZHU X Z, HU H, LIN S, *et al.* Deformable ConvNets V2: more deformable, better results [C]//*Proceedings*

- of 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019: 9308-9316.
- [16] 杨琦, 车进, 张良, 等. GAN 网络混合编码的行人再识别[J]. 液晶与显示, 2021, 36(2): 334-342.
YANG Q, CHE J, ZHANG L, *et al.* Person re-identification of GAN network hybrid coding [J]. *Chinese Journal of Liquid Crystals and Displays*, 2021, 36(2): 334-342. (in Chinese)
- [17] HE K M, ZHANG X Y, REN S Q, *et al.* Deep residual learning for image recognition [C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016: 770-778.
- [18] HERMANS A, BEYER L, LEIBE B. In defense of the triplet loss for person re-identification [J]. *arXiv preprint arXiv:1703.07737*, 2017.
- [19] LIN S, LI H L, LI C T, *et al.* Multi-task mid-level feature alignment network for unsupervised cross-dataset person re-identification [J]. *arXiv preprint arXiv:1807.01440*, 2018.
- [20] CHEN Y B, ZHU X T, GONG S G. Instance-guided context rendering for cross-domain person re-identification [C]//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul, Korea: IEEE, 2019: 232-242.
- [21] ZHANG X Y, CAO J W, SHEN C H, *et al.* Self-training with progressive augmentation for unsupervised cross-domain person re-identification [C]//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul, Korea: IEEE, 2019: 8222-8231.
- [22] ZHONG Z, ZHENG L, LUO Z M, *et al.* Learning to adapt invariance in memory for person re-identification [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(8): 2723-2738.
- [23] JIN X, LAN C L, ZENG W J, *et al.* Style normalization and restitution for generalizable person re-identification [C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020: 3143-3152.
- [24] HUANG Y R, PENG P X, JIN Y, *et al.* Domain adaptive attention model for unsupervised cross-domain person re-identification [J]. *arXiv preprint arXiv:1905.10529*, 2019.

作者简介:



杨海轮(1998—),男,辽宁锦州人,硕士研究生,2020年于长春理工大学获得学士学位,主要从事图像识别与智能控制方面的研究。E-mail: yanghailun1998@163.com



任洪娥(1962—),女,吉林白山人,博士,教授,2009年于东北林业大学获得博士学位,主要从事图像识别与智能控制方面的研究。E-mail: nefu_rhe@163.com