

DOI:10.19651/j.cnki.emt.2208924

基于相机风格转换的行人再识别*

邓杰^{1,2} 王旭智^{1,2} 万旺根^{1,2}

(1.上海大学通信与信息工程学院 上海 200072; 2.上海大学智慧城市研究院 上海 200072)

摘要: 不同相机间的风格变化是行人再识别领域的一个重要挑战,为了平滑相机风格差异,丰富行人样本的多样性,本文通过风格转换方法显式学习相机间的不变特征。具体来说,利用循环一致性生成对抗网络为每个行人生成具有其他相机风格的转换图像,并与原始样本一起组成增强数据集进行训练;另外,本文使用注意力机制对特征通道进行重新加权以提取更具判别力的行人外观特征,最后使用多任务损失对再识别网络进行监督训练。实验结果表明,本文方法在公开数据集 Market1501 和 DukeMTMC-reID 上的 mAP 和 top-1 指标分别达到了 86.5%,95.1% 以及 77.1%,87.2%,优于现有算法。相机风格转换作为一种数据增强方法,有效扩充了数据集并降低了人工标注成本,同时提升了在多摄像机场景下的识别准确性。

关键词: 行人再识别;风格转换;循环一致性生成对抗网络;注意力机制

中图分类号: TP391.41;TP332 **文献标识码:** A **国家标准学科分类代码:** 520.60

Camera style transfer for person re-identification

Deng Jie^{1,2} Wang Xuzhi^{1,2} Wan Wanggen^{1,2}

(1. School of Communication and Information Engineering, Shanghai University, Shanghai 200072, China;

2. Institute of Smart City, Shanghai University, Shanghai 200072, China)

Abstract: Style variations among different cameras is an important challenge in the field of person re-identification. To smooth the camera style disparities and enrich the diversity of pedestrian samples, this paper explicitly learns invariant features among cameras through a style transfer approach. Specifically, a cycle consistent adversarial networks (CycleGAN) is used to generate transformed images with other camera styles for each pedestrian, and along with the original samples, form the augmented training set. In addition, this paper uses an attention mechanism to reweight the feature channels to extract more discriminative pedestrian appearance features, and finally, the multi-task loss is used to supervise the training process of the re-identification network. The experimental results show that the mAP and top-1 metrics of the method in this paper achieve 86.5%, 95.1% and 77.1%, 87.2% on the public datasets Market1501 and DukeMTMC-reID, respectively, which are better than the existing algorithms. Camera style transfer as a data augmentation approach effectively expands the dataset and reduces the human labeling cost, while improving the identification accuracy in multi-camera scenarios.

Keywords: person re-identification; style transfer; cycleGAN; attention mechanism

0 引言

行人再识别(person re-identification, re-ID)是利用计算机视觉技术在跨摄像机场景下匹配特定行人的任务,广泛应用于智慧安防、行为分析、内容检索等领域,近年来已成为学术界和工业界的研究热点。给定一个待查询对象,行人再识别的目标是从多个摄像机采集的数据集中检索同

一个人。由于不同摄像机在成像设备、拍照角度、光照条件等方面存在差异,同一行人通常在外观和背景上会发生显著变化;同时,不同摄像机拍摄的行人图像具有不同的风格,这也为识别结果带来了很大的影响。这些因素使得行人再识别仍然是一项充满挑战性的任务。

近年来,基于深度神经网络的行人再识别技术取得了很大的进展,并取得了良好的性能。Zheng 等^[1]提出身份鉴别

收稿日期:2022-01-24

* 基金项目:安徽省自然科学基金(1908085MF178)、安徽省重点研究和开发计划项目(202104b11020031)、中国博士后基金(2020M681264)项目资助

性嵌入方法,使用 ImageNet^[2] 预训练模型进行网络微调,将 re-ID 模型训练为图像分类问题。Luo 等^[3]通过对行人再识别领域有效的训练技巧进行总结和实验并将这些技巧结合起来,提出了一个简单而高效的基准网络,在仅使用全局特征的情况下取得了较高的精度。神经网络的训练需要大量的数据,当 CNN 模型相对于训练样本数量过于复杂时会出现过拟合的现象,为了解决这一问题,一些基于生成式对抗网络(generative adversarial networks, GANs)^[4] 的数据增强方法被提出。Zheng 等^[5]使用 DC-GAN^[6]生成无标签样本,并通过异常值标签平滑正则化方法进行均匀的标签分配以提高 CNN 的识别能力,但生成的无标签行人图像不够真实,丢失了大部分行人特征,从而限制了网络性能。PNGAN^[7]提出了一种姿态标准化的生成对抗式再识别框架,通过生成包含不同姿态的标签图像,从原始图像与生成图像中学习互补特征完成再识别任务,但在训练过程中需要额外的姿态估计模型,并且只使用了 8 种标准姿态,没有充分考虑行人样本中姿态的多样性。同样,为应对行人姿态变化带来的挑战,FDGAN^[8]使用人体姿态引导的 GAN 网络学习与身份相关且与姿态无关的行人特征,输入图像中与行人身份无关的信息(如姿态和背景)会被图像编码器从视觉特征中消除,推理过程中不再需要额外的姿态信息,节省了额外的计算成本,但由于生成图像消除了外观信息,导致无法在生成图像中提取到具有鉴别力的细节特征。以上两项工作仅从行人姿态角度出发,通过生成姿态样本库来扩充数据集,一定程度上提高了再识别性能。除了行人姿态,摄像机角度、分辨

率、光照条件等因素导致的相机风格差异也给人再识别任务带来了巨大挑战。为解决该问题,Zhong 等^[9]提出了基于相机风格自适应的 CamStyle 方法,相比于 DC-GAN,该方法使用循环一致性生成对抗网络(cycle consistent adversarial networks, CycleGAN)^[10]生成更为真实的行人图像进行数据增强,但再识别网络提取的行人外观特征不够鲁棒,在算法性能上有较大的提升空间。

为丰富行人再识别数据集并降低人工标注成本,同时为了应对不同相机间的风格差异问题,本文使用 CycleGAN 对行人再识别数据集进行相机风格转换,将原始图像与风格转换图像组合成增强数据集,同时本文提出了一个鲁棒的再识别网络,使用通道注意力 SEBlock^[11](squeeze-and-excitation block)提取更有判别力的全局特征,最后使用三元组变体损失^[12]、中心损失^[13]和标签平滑正则化损失^[14]对网络进行监督训练。

1 相机风格转换

1.1 CycleGAN

给定两个数据集 $\{x_i\}_{i=1}^M$ 和 $\{y_j\}_{j=1}^N$, 分别从两个不同的域 X 和 Y 中得到,其中 $x_i \in X, y_j \in Y$ 。CycleGAN 的目标是学习一个映射函数 $G: X \rightarrow Y$, 使得来自 $G(X)$ 的图像分布在使用对抗损失训练后与 Y 的图像分布难以区分。CycleGAN 包括两个映射函数 $G: X \rightarrow Y$ 和 $F: Y \rightarrow X$, 以及两个对抗判别器 D_X 和 D_Y , 分别用来区分图像是否是其他域转换而来。CycleGAN 的网络结构如图 1 所示。

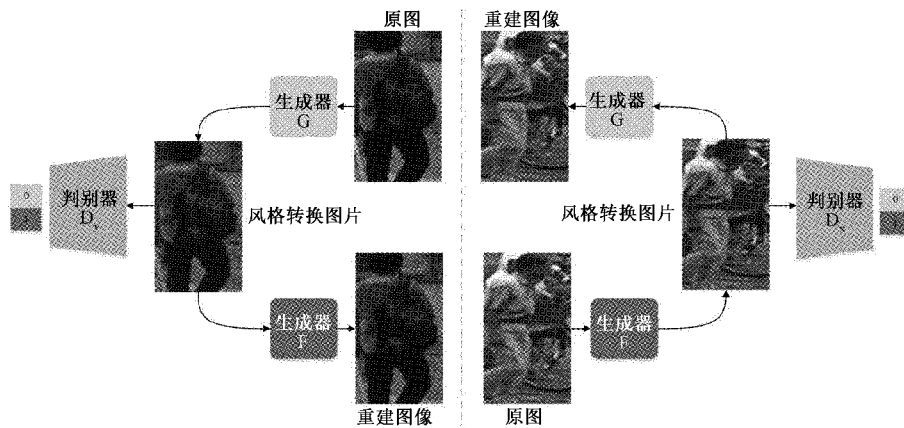


图 1 CycleGAN 网络结构

CycleGAN 使用 GAN 的框架联合训练生成器和判别器,总体损失函数如下式所示:

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{GAN}(D_Y, G, X, Y) + \mathcal{L}_{GAN}(D_X, F, Y, X) + \lambda \mathcal{L}_{cyc}(G, F) \quad (1)$$

其中, $\mathcal{L}_{GAN}(D_Y, G, X, Y)$ 和 $\mathcal{L}_{GAN}(D_X, F, Y, X)$ 是映射函数 G 和 F 以及判别器 D_Y 和 D_X 的对抗损失。 \mathcal{L}_{GAN} 损失函数如下:

$$\mathcal{L}_{GAN}(D_Y, G, X, Y) = \mathbb{E}_{y \sim P_{data}(y)} [\log D_Y(y)] + \mathbb{E}_{x \sim P_{data}(x)} [\log(1 - D_Y(G(x)))] \quad (2)$$

$$\mathcal{L}_{GAN}(D_X, F, Y, X) = \mathbb{E}_{x \sim P_{data}(x)} [\log D_X(x)] + \mathbb{E}_{y \sim P_{data}(y)} [\log(1 - D_X(F(y)))] \quad (3)$$

如图 1 所示,原始输入图像经过生成器 G 生成风格转换图像,再经过生成器 F 生成重建图像。生成重建图像的目的是避免生成器 G 为了欺骗判别器,仅仅只生成具有目标域风格的图像而忽略原图本身的结构与内容,因此需要对原始图像和重建图像进行监督训练,将这部分损失定义为 $\mathcal{L}_{cyc}(G, F)$, 它强制 $F(G(x)) \approx x$ 以及 $G(F(y)) \approx y$, 其中每张图像在循环映射后可以重建。循环一致性

损失函数采用 L1 损失,如下式所示:

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x \sim P_{data}(x)} [\| F(G(x)) - x \|_1] + \mathbb{E}_{y \sim P_{data}(y)} [\| G(F(y)) - y \|_1] \quad (4)$$

式中: λ 作为调节因子为 \mathcal{L}_{GAN} 和 \mathcal{L}_{cyc} 分配权重。

1.2 生成器与判别器网络

CycleGAN 的生成器网络结构如图 2 所示,输入图片经过卷积层之后通道数由 3 维变成 64 维,使用步长为 2 的卷积操作进行 3 次下采样,每次都特征通道数增加 1 倍,直到生成通道数为 512 维的特征图。每次下采样使用由卷积层、实例归一化(instance normalization, IN)、ReLU 激活组成的 CIR 组件,再使用 9 个 ResNet 残差单元提取深度特征,之后经过 3 次转置卷积进行上采样,每次都特征通道数减半,最后通过一个卷积层输出风格转换图像。图 2 中给出了原始图像经过每一模块后生成的特征图的通道数。由于生成结果主要依赖于某个图像的实例,适合对每个特征图做归一化操作,以保证各图像实例间的独立。因此,相比于批归一化(batch normalization, BN),实例归一化更适合图像风格转换任务。

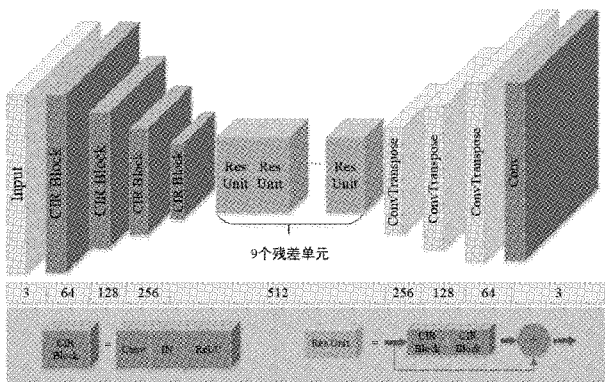


图 2 生成器网络结构

判别器为感受野大小为 70×70 的 PatchGAN 全卷积网络,该网络结构如图 3 所示。判别器网络首先对生成器输出的风格转换图像进行卷积和 Leaky ReLU 激活,通过 3 个由卷积、实例归一化和 Leaky ReLU 组成的 CIR 组件提取图像的深度特征,再通过最后一个卷积层生成一维特征矩阵,该矩阵中的每一个值都代表着风格转换图像中 70×70 像素大小的区域为正样本的概率,将该特征矩阵取平均值后作为判别器的最终输出。

本文使用 CycleGAN 为每一个摄像头生成 $L-1$ 个新的风格转换样本, L 为数据集中的相机数量。

1.3 相机风格转换损失函数

如 1.1 节所述,循环一致性损失对原始图像与重建图像进行逐像素约束,以使两者之间更加相似。对于生成器 G 来说,当输入是目标域的图像时,得到的风格转换图像被称为身份映射图像,我们希望生成器 G 对目标域图像产生的影响尽量小,因此需要对目标域图像和身份映射图像进行监督学习,将这部分损失定义为身份映射损失 $\mathcal{L}_{idt}(G, F)$ 。

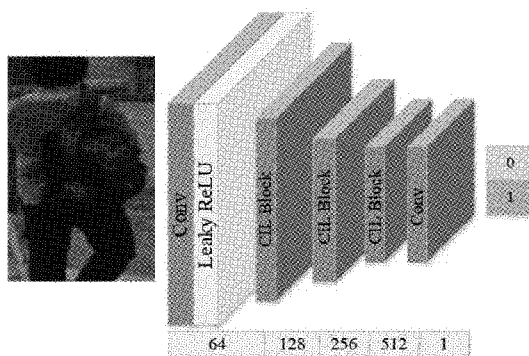


图 3 判别器网络结构

综上,相机风格转换损失整体包含 3 个部分,分别是生成对抗损失 \mathcal{L}_{GAN} , 循环一致性损失 \mathcal{L}_{cyc} 和身份映射损失 \mathcal{L}_{idt} , 其中循环一致性损失和身份映射损失均使用 L1 损失。相机风格转换的总体损失函数如下:

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{GAN}(D_Y, G, X, Y) + \mathcal{L}_{GAN}(D_X, F, Y, X) + \lambda_{cyc} \mathcal{L}_{cyc}(G, F) + \lambda_{idt} \mathcal{L}_{idt}(G, F) \quad (5)$$

其中, λ_{cyc} 取值为 10, λ_{idt} 取值为 0.5。

2 行人再识别

2.1 再识别网络

对于原始图像与风格转换图像组成的增强训练集,本文采用 Luo 等^[3]提出的网络基准与训练技巧,仅使用全局特征训练再识别网络。本文提出的再识别网络结构如图 4 所示,首先原始图像经过风格转换后生成具有其他相机风格的新图像,将两者组合成增强数据集后对再识别网络进行训练。本文使用在 ImageNet 数据集上预训练的 SEResNet50 作为特征提取网络,并加入 SEBlock 对提取到的特征重新分配权重,经过全局平均池化后得到加权全局特征。加权后的全局特征通过两个分支进行联合优化,第一个分支对全局特征使用三元组变体损失与中心损失进行监督训练;另一条分支对全局特征进行批归一化与全连接操作之后使用标签平滑正则化损失进行监督训练。

SEBlock 对卷积特征通道之间的相关性进行建模,通过挤压和激励操作对通道信息进行加权从而获得更好的特征响应。具体来说,SEBlock 对原始特征进行全局平均池化,全连接,ReLU 激活, Sigmoid 等操作后生成特征通道权重 ω , 再与原始特征进行通道相乘得到加权特征。它被定义为特征 F 到 F' 的转换:

$$F' = F * \omega = [\omega_1, \omega_2, \dots, \omega_{2c}] = \sigma(W_2 \delta(W_1 f_{GAP}(F))) \quad (6)$$

其中, σ 表示 Sigmoid 激活函数, δ 表示 ReLU 激活函数。 W_1 和 W_2 是两个全连接层的权重矩阵。 f_{GAP} 是全局平均池化操作, $*$ 表示通道相乘。SEBlock 有选择地学习带有强调性信息的特征,并使用加权向量 ω 对通道特征重新加权来抑制不太有用的特征。

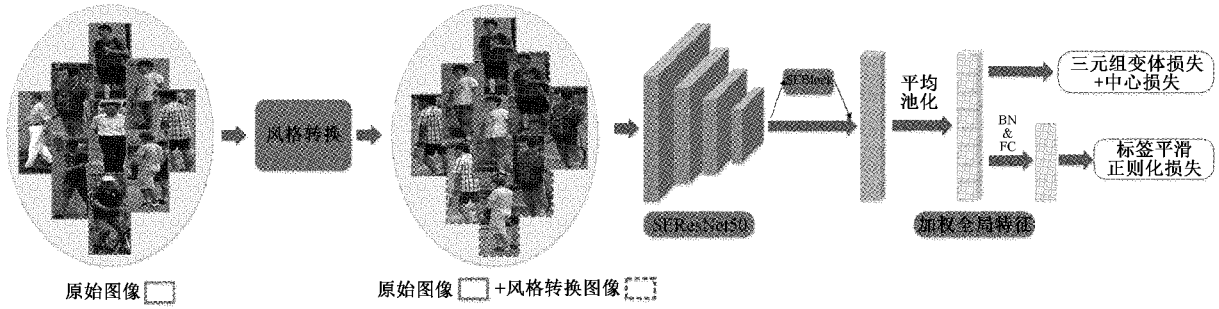


图 4 行人再识别网络

2.2 再识别损失函数

本文使用三元组变体损失、中心损失和标签平滑正则化损失对增强数据集进行监督训练。

三元组变体损失函数表示为:

$$\mathcal{L}_{Tri} = \max(d_p - d_n + \alpha, 0) \quad (7)$$

其中, d_p 是正样本对特征之间的距离, d_n 是负样本对特征之间的距离, α 是边界值,文中设置为 0.3。

三元组损失的目的是拉近正样本特征向量与标签图像特征向量之间的距离,扩大负样本特征向量与标签图像特征向量之间的距离。与传统的三元组损失随机从训练集中挑选出三张图片的方法不同,为了避免网络一直学习简单样本,三元组变体损失使用了批次困难样本挖掘^[12]的方法来提升网络的泛化能力。具体来说,对于 batch 里的每一张图片 a , 计算 a 和样本批次中每一张图片在特征空间中的欧氏距离,挑选出与 a 距离最远的正样本 p 和与 a 距离最近的负样本 n 组成三元组来计算损失。

中心损失可以同时学习各个类别的中心深度特征,并对深度特征与对应的类别中心加以惩罚,表示为:

$$\mathcal{L}_{Center} = \frac{1}{2} \sum_{j=1}^B \|f_{t_j} - c_{y_j}\|^2 \quad (8)$$

其中, y_j 是批次中第 j 张图片的标签, c_{y_j} 表示第 y_j 张标签图像深度特征的中心, B 表示批次大小。该损失可以有效缩小类内差异。

相机风格转换可以有效减少过拟合并实现相机不变特性,但也会给系统带来噪声。同时在使用独热编码计算交叉熵损失时,网络仅关注正确标签位置的损失而忽略错误标签位置的损失,这使得模型过于关注增大预测正确标签的概率而不关注减少预测错误标签的概率,通常带来的结果就是模型的泛化能力较差。因此对于原始图像与风格转换图像,本文均使用标签平滑正则化(label smoothing regularization, LSR)损失进行监督训练,用于平滑标签分布,使网络尽可能地靠近正确分类并远离错误分类。标签平滑正则化损失可以表示为:

$$\mathcal{L}_{LSR} = -(1 - \epsilon) \log p(y) - \frac{\epsilon}{C} \sum_{c=1}^C \log p(c) \quad (9)$$

其中, C 为行人身份数量, $p(y)$ 表示正确标签位置的预测概率, $p(c)$ 表示输入图片中属于标签行人 c 的预测概

率,由 softmax 层进行归一化。 ϵ 为平滑因子,为标签分配权重,取值为 0.1。当 ϵ 为 0 时,标签平滑正则化损失可以退化为交叉熵损失,如下式所示。

$$\mathcal{L}_{Cross} = -\log p(y) \quad (10)$$

本文的总体损失函数由真实图像损失 \mathcal{L}_{real} 与风格转换图像损失 $\mathcal{L}_{transfer}$ 两部分组成,如下式所示:

$$\mathcal{L} = \mathcal{L}_{real} + \alpha \mathcal{L}_{transfer} \quad (11)$$

$$\mathcal{L}_{real} = \mathcal{L}_{Tri} + \beta \mathcal{L}_{Center} + \mathcal{L}_{LSR} \quad (12)$$

$\mathcal{L}_{transfer}$ 同理可得。其中 α 与 β 为调节因子,用于调整各部分损失之间的比例,本文中 α 取值为 0.1, β 取值为 0.0005。

3 实验分析

3.1 数据集及评价指标

本文使用两个常见且具有代表性的行人再识别数据集 Market-1501^[15] 和 DukeMTMC-reID^[15] 进行训练和测试。

Market1501 数据集在大学校园中采集,它包括由 6 个摄像头拍摄到的 1 501 个行人以及 32 668 张标签图像。训练集有 751 人,包含 12 936 张行人图像,测试集有 750 人,包含 19 732 张行人图像,共有 3 368 张查询图像。

DukeMTMC-reID 数据集是一个大规模行人再识别数据集。它从 8 个摄像头中采集了 1 404 个行人,共包含 36 411 张标签图像。其中训练集有 702 人,包含 16 522 张行人图像;测试集有 702 人,包含 17 661 张行人图像,共有 2 228 张查询图像。

评价指标:本文使用平均精度均值(mAP)和累计匹配特征(CMC)对行人再识别任务进行评估。平均精确率由精确率-召回率的关系曲线(P-R 曲线)与坐标轴的面积计算得到,对所有类别测试得到各自的平均精确率,取均值得到平均精度均值 $mAP^{[16-17]}$; 累计匹配特征将行人匹配结果按照相关程度的高低进行排序,并格外重视排在第一位(top-1),即相关度最高的结果为正确结果的概率,当数据量较大或者规则需要时,可额外参考后续排名的正确率(top-K)^[18]。

3.2 实验设置

本节实验环境配置: Ubuntu16.04 (LTS)、GTX 1080Ti 12GB、CUDA 10.1、Python 3.7、PyTorch 1.17.0、

torchvision 0.8.1。

训练过程包括相机风格转换与再识别两部分。在相机风格转换阶段,本文为每对摄像头训练一个相机风格转换模型,同时保留行人原始身份用于增强数据集。训练过程中,将输入图片的尺寸调整为 256×256 ,使用 Adam 优化器训练 50 个 epoch, batch size 设置为 1。前 30 个 epoch 将生成器的学习率设置为 0.0002,判别器的学习率设置为 0.0001,之后的 20 个 epoch 学习率线性降低为 0。

再识别模型训练阶段,输入图片的尺寸设置为 256×128 ,每张图片以 0.5 的概率进行镜像翻转和随机擦除^[19],采用 Adam 优化器训练 240 个 epoch,初始学习率设置为 3.5×10^{-5} ,采用学习率预热策略,前 10 个 epoch 线性增长至 3.5×10^{-4} ,分别在第 40 和 70 个 epoch 时下降至 3.5×10^{-5} 和 3.5×10^{-6} 。将训练集设置为原始图像与风格转换图像组合成的混合数据集,训练过程中原始图像与风格转换图像的 batch size 大小分别设置为 64 和 32。

3.3 整体性能对比与分析

表 1 和 2 分别给出了本文所提方法与现有基于 GANs 的方法在两个数据集上的实验结果对比。CamStyle 同样使用 CycleGAN 进行相机风格自适应以增强数据集,相比于该方法,本文提出的再识别网络更具鲁棒性,在 mAP 与 top-1 指标上均有大幅提升。本文方法在 Market1501 数据集上的 mAP 与 top-1 分别为 86.5% 和 95.1%,比 CamStyle 提升了 17.8% 与 7%,在 DukeMTMC-reID 数据集上 mAP 与 top-1 分别为 77.1% 和 87.2%,相比 CamStyle 分别提升了 23.6% 与 11.9%,充分证明了本文提出的再识别网络可以在增强数据集中提取到更为鲁棒的行人特征从而提高网络性能。PN-GAN、FD-GAN 与 GAN 编码方法^[20]从行人姿态角度出发,利用 GANs 生成行人姿态库从而提升再识别性能,与之相比本文方法考虑了光照、色彩、背景等多种因素带来的差异,有效增加了行人样本的多样性,在 mAP 和 top-1 指标上均取得了最优结果。

表 1 Market1501 实验结果对比

方法	mAP	top-1
DC-GAN ^[5]	66.1	84.0
CamStyle ^[9]	68.7	88.1
PN-GAN ^[7]	72.6	89.4
FD-GAN ^[8]	77.7	90.5
GAN 编码 ^[20]	82.2	93.4
本文方法	86.5	95.1

3.4 消融实验与分析

本节以 BOT⁻³ 作为基准网络,进一步验证所提方法中相机风格转换以及 SEBlock 的作用,在表 3 和表 4 中给出了在两个数据集上的消融实验结果,采用 mAP 与 top-1、

表 2 DukeMTMC-reID 实验结果对比

方法	mAP	top-1
CamStyle ^[9]	53.5	75.3
PN-GAN ^[7]	53.2	73.6
FD-GAN ^[8]	64.5	80.0
GAN 编码 ^[20]	70.5	84.3
本文方法	77.1	87.2

top-5 与 top-10 作为定性指标。与基准网络相比,本文方法在 Market1501 数据集上 mAP 和 top-1 指标均提升了 0.6%,在 DukeMTMC-reID 数据集上分别提升了 0.7% 和 0.8%。其中,在 Market1501 数据集上,使用通过相机风格转换方法增强后的数据集训练基线网络使 top-1 识别准确率达到了 94.9%,高出基准网络 0.4%;在 DukeMTMC-reID 数据集上 top-1 识别准确率达到 87.0%,高出基准网络 0.6%。相机风格转换通过显式学习相机间的不变特征,有效缓解了相机间的风格差异问题,为再识别网络带来良好收益。图 5 中展示了在 Market-1501 数据集中使用 CycleGAN 为不同相机下拍摄的行人生成的风格转换图像,Cam1 表示当前行为 1 号摄像机拍摄的行人标签图像,Cam1->2 表示将其转换为 2 号摄像机拍摄图像的风格,以此类推。

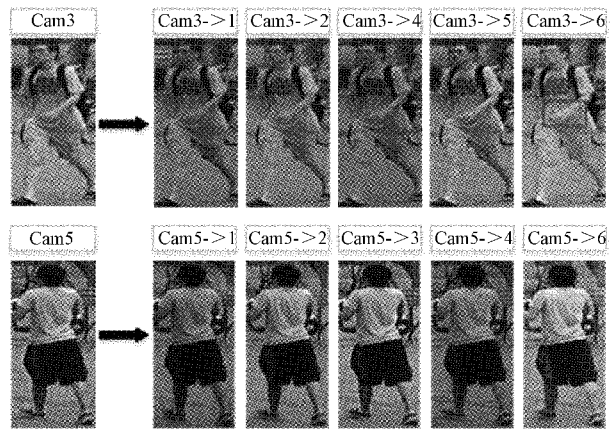


图 5 相机风格转换图像示例

由图 5 可以看出,进行风格转换后的图像保留了原始行人的主要特征,仅在亮度、对比度和色度等方面有所不同,但是在风格转换过程中不可避免地会产生一些噪声,如上图 Cam3->6 框内所示。网络通过 CycleGAN 敏锐地学习到了不同相机间细微的风格变化,为原始数据集提供了更丰富的行人样本,风格转换图像与原始图像具有相似分布并且保留了原始图像的身份标签,因此可以和原始图像一起训练再识别模型。同时,在再识别网络中加入 SEBlock 也为识别结果带来了小幅提升,在两个数据集上的识别准确率均提高了 0.2%,SEBlock 对特征通道重新分配权重有助于网络获取更有判别力的特征从而提升再识别精度。

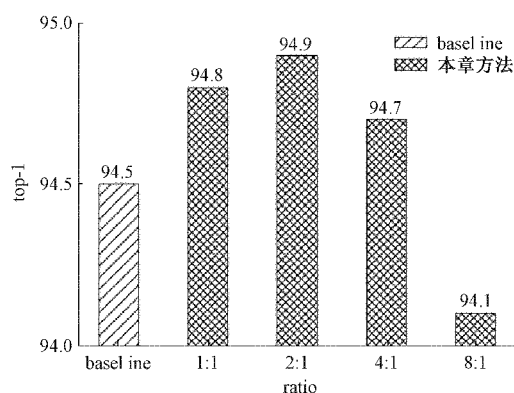
表3 本文方法在 Market1501 数据集上的消融实验结果

baseline ^{-3]}	风格转换	SEBlock	mAP	top-1	top-5	top-10
✓			85.9	94.5	—	—
✓	✓		86.3	94.9	98.1	98.9
✓	✓	✓	86.5	95.1	98.6	99.1

表4 本文方法在 DukeMTMC-reID 数据集上的消融实验结果

baseline ^{-3]}	风格转换	SEBlock	mAP	top-1	top-5	top-10
✓			76.4	86.4	—	—
✓	✓		77.0	87.0	94.3	96.2
✓	✓	✓	77.1	87.2	94.4	96.4

在相机风格转换的实验中,每个 batch 中原始图像与风格转换图像的比例 $M:N$ 是一个重要参数,以 Market1501 数据集为例,图 6 给出了 $M:N$ 不同取值下 top-1 识别准确率的对比实验结果,其中 baseline 表示仅使用原始图像进行训练,相机风格转换过程中 M 取值为 64。可以看出,与 baseline 相比,当 $M:N$ 为 1:1 时 top-1 提高了 0.3%,为 3:1 时提高了 0.2%,为 2:1 时提高了 0.4%达到了最佳性能,而当 $M:N$ 为 8:1 时性能有所下降,这是因为 CycleGAN 无法完美模拟转换过程,在图像生成过程中会产生噪声,而当风格转换图像所占比重较小时,网络无法完全避免这部分噪声带来的影响从而损害了再识别网络的性能。由实验结果可知 $M:N$ 的最佳取值为 2:1。

图6 $M:N$ 不同取值时的 top-1 对比

4 结 论

本文提出了一种基于相机风格转换的行人再识别方法,该方法使用 CycleGAN 为每个相机学习不同的风格转换图像,并与原始图像共同组成训练集训练再识别网络。同时本文加入 SEBlock 重新分配特征通道权重,使网络获取更有判别力的全局特征,最后使用多任务损失对网络进行监督训练。相机风格转换作为一种数据增强方法,有效

学习了相机间的不变特征,丰富了行人样本并降低了人工标注成本。实验表明,使用所提方法训练的模型能够有效地提高行人再识别的性能。但在风格转换过程中会产生图像伪影并给系统带来噪声,同时随着摄像机数量的增加,需要训练的模型数量也逐渐增多,对于计算资源不足的场景带来了较大挑战。如何获取更高质量的生成图像并降低计算成本是未来的研究方向。

参考文献

- [1] ZHENG L, YANG Y, HAUPTMANN A G. Person re-identification: Past, present and Future[J]. ArXiv Preprint, 2016, ArXiv: 1610. 02984.
- [2] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]. NeurIPS, 2012: 1106-1114.
- [3] LUO H, GU Y Z, LIAO X Y, et al. Bag of tricks and a strong baseline for deep person re-identification[C]. CVPR Workshops, 2019: 1487-1495.
- [4] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C]. NIPS, 2014: 2672-2680.
- [5] ZHENG Z D, ZHENG L, YANG Y. Unlabeled samples generated by gan improve the person re-identification baseline in vitro [C]. ICCV, 2017: 3774-3782.
- [6] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks [C]. ICLR, 2016.
- [7] QIAN X L, FU Y W, XIANG T, et al. Pose-normalized image generation for person re-identification[C]. ECCV, 2018: 661-678.
- [8] GE Y X, LI Z W, ZHAO H Y, et al. FD-GAN: Pose-guided feature distilling GAN for robust person re-identification[C]. NeurIPS, 2018: 1230-1241.
- [9] ZHONG Z, ZHENG L, ZHENG Z D, et al. Camera style adaptation for person re-identification [C]. CVPR, 2018: 5157-5166.
- [10] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]. ICCV, 2017: 2242-2251.
- [11] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[C]. CVPR, 2018: 7132-7141.
- [12] HERMANS A, BEYER L, LEIBE B. In defense of the triplet loss for person re-identification[J]. ArXiv Preprint, 2017, ArXiv: 1703. 07737.
- [13] WEN Y D, ZHANG K P, LI Z F, et al. A discriminative feature learning approach for deep face recognition[C]. ECCV, 2016: 499-515.

- [14] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]. CVPR, 2016: 2818-2826.
- [15] ZHENG L, SHEN L Y, TIAN L, et al. Scalable person re-identification: A benchmark [C]. ICCV, 2015: 1116-1124.
- [16] 李擎, 胡伟阳, 李江昀, 等. 基于深度学习的行人重识别方法综述[J/OL]. 工程科学学报. <https://kns.cnki.net/kcms/detail/10.1297.TF.20211119.1132.003.html>.
- [17] 朱肖磊, 吴训成. 车辆姿态感知注意力增强的车辆重识别[J]. 电子测量技术, 2021, 44(24): 91-97.
- [18] 刘仁春, 孟朝晖. 基于孪生对抗 SGAN 的行人重识别研究[J]. 电子测量技术, 2019, 42(15): 155-160, 166.
- [19] ZHONG Z, ZHENG L, KANG G L, et al. Random erasing data augmentation [C]. AAAI, 2020: 13001-13008.
- [20] 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.

作者简介

邓杰, 硕士, 主要研究方向为计算机视觉、行人检测与重识别。

E-mail: 13262218151@163.com

王旭智, 副教授, 主要研究方向为视频处理, 3D 图像。

E-mail: wangxzw@shu.edu.cn

万旺根, 教授, 博士生导师, 主要研究方向为计算机图形学、信号处理和数据挖掘。

E-mail: wanwg@staff.shu.edu.cn