

# 基于多尺度生成对抗网络的遮挡行人重识别方法\*

杨婉香<sup>1</sup>, 严 严<sup>1</sup>, 陈 思<sup>2</sup>, 张小康<sup>1</sup>, 王菡子<sup>1</sup>



<sup>1</sup>(福建省智慧城市感知与计算重点实验室(厦门大学 信息学院), 福建 厦门 361005)

<sup>2</sup>(厦门理工学院 计算机与信息工程学院, 福建 厦门 361024)

通讯作者: 严严, E-mail: yanyan@xmu.edu.cn

**摘 要:** 行人重识别是指在多个非重叠摄像头拍摄的场景下, 给定一幅查询行人图像, 从大规模行人图像库中检索出具有相同身份的行人图像, 是一类特殊的图像检索任务。随着深度学习的不断发展, 行人重识别方法的性能得到了显著提升。但是行人重识别在实际应用中经常遭遇遮挡问题(例如, 背景遮挡、行人互相遮挡等)。由于遮挡图像不仅丢失了部分目标信息, 而且引入了额外的干扰, 使得现有方法往往难以学习到鲁棒的特征表示, 从而导致识别性能严重下降。最近, 生成对抗网络在各类计算机视觉任务上展现出强大的图像生成能力。受到生成对抗网络的启发, 提出了一种基于多尺度生成对抗网络的遮挡行人重识别方法。首先, 利用成对的遮挡图像和非遮挡图像训练一个多尺度生成器和一个判别器。多尺度生成器能够对随机遮挡区域进行去遮挡操作, 生成高质量的重构图; 而判别器能够区分输入图像是真实图像还是生成图像。其次, 利用训练好的多尺度生成器, 生成去除随机遮挡的训练图像, 添加到原始训练图像集, 用于增加训练样本的多样性。最后, 基于此扩充训练图像集, 训练分类识别模型, 有效地提高模型在测试图像集上的泛化性。在多个有挑战性的行人重识别数据集上的实验结果, 验证了所提出方法的有效性。

**关键词:** 行人重识别; 生成对抗网络; 遮挡; 数据增强

**中图法分类号:** TP391

中文引用格式: 杨婉香, 严严, 陈思, 张小康, 王菡子. 基于多尺度生成对抗网络的遮挡行人重识别方法. 软件学报, 2020, 31(7): 1943–1958. <http://www.jos.org.cn/1000-9825/5932.htm>

英文引用格式: Yang WX, Yan Y, Chen S, Zhang XK, Wang HZ. Multi-scale generative adversarial network for person re-identification under occlusion. Ruan Jian Xue Bao/Journal of Software, 2020, 31(7): 1943–1958 (in Chinese). <http://www.jos.org.cn/1000-9825/5932.htm>

## Multi-scale Generative Adversarial Network for Person Re-identification under Occlusion

YANG Wan-Xiang<sup>1</sup>, YAN Yan<sup>1</sup>, CHEN Si<sup>2</sup>, ZHANG Xiao-Kang<sup>1</sup>, WANG Han-Zi<sup>1</sup>

<sup>1</sup>(Fujian Key Laboratory of Sensing and Computing for Smart City (School of Informatics, Xiamen University), Xiamen 361005, China)

<sup>2</sup>(School of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, China)

**Abstract:** Person re-identification (ReID) refers to the task of retrieving a given probe pedestrian image from a large-scale gallery collected by multiple non-overlapping cameras, which belongs to a specific task of image retrieval. With the development of deep learning, the performance of person ReID has been significantly improved. However, in practical applications, person ReID usually suffers from the problem of occlusion (such as background occlusion, pedestrian occlusion). The occluded image not only loses partial target information, but also introduces additional interference, which makes the deep neural network difficult to learn robust feature representations and seriously degrades the performance of person ReID. Recently, generative adversarial network (GAN) has shown the powerful image generation ability on various computer vision tasks. Inspired by GAN, a person ReID method is proposed under occlusion based on

\* 基金项目: 国家自然科学基金(61571379, U1605252, 61872307); 福建省自然科学基金(2017J01127, 2018J01576)

Foundation item: National Natural Science Foundation of China (61571379, U1605252, 61872307); Natural Science Foundation of Fujian Province of China (2017J01127, 2018J01576)

本文由“多媒体内容的多维度相似性计算与搜索”专题特约编辑蒋树强研究员、刘青山教授、孙立峰教授、李波教授推荐。

收稿时间: 2019-04-24; 修改时间: 2019-07-11; 采用时间: 2019-09-17; jos 在线出版时间: 2020-01-13

CNKI 网络优先出版: 2020-01-14 11:25:07, <http://kns.cnki.net/kcms/detail/11.2560.TP.20200114.1125.015.html>

multi-scale GAN. Firstly, the paired occluded images and unoccluded images are used to train a multi-scale generator and a discriminator. The multi-scale generator can restore the lost information for randomly occluded areas and generate high-quality reconstructed images; while the discriminator can distinguish whether the input image is a real image or a generated image. Then, the trained multi-scale generator is used to generate the de-occluded images. Adding these de-occluded images to the original training image set can increase the diversity of training samples. Finally, a classification network is trained based on the augmented training image set, which effectively improves the generalization capability of the trained model on the testing image set. Experimental results on several challenging person ReID datasets demonstrate the effectiveness of the proposed method.

**Key words:** person re-identification; generative adversarial network; occlusion; data augmentation

行人重识别(person re-identification)也称为行人再识别,是指在多个非重叠摄像头拍摄的场景下,给定一幅查询行人图像,从大规模行人图像库中检索出具有相同身份的行人图像,是一类特殊的图像检索任务<sup>[1-3]</sup>.行人重识别广泛应用于视频监控、智能安保等领域.随着深度学习的不断发展,基于深度学习的行人重识别方法显著提高了识别的性能.但是,由于跨摄像机视角的行人图像受姿态变化、光照变化和遮挡等因素的影响,使得行人重识别任务仍然面临着巨大的挑战.其中,遮挡是实际场景中影响行人重识别性能的一个重要因素.

遮挡的行人图像不仅丢失了部分区域的目标信息,而且引入了额外的干扰.同时,由于遮挡的复杂性,比如遮挡类型多样(背包遮挡、背景遮挡、行人相互遮挡等)、遮挡区域不固定等,所以,直接基于行人图像往往难以学习到遮挡鲁棒的行人特征表示.尽管目前有些方法<sup>[4-6]</sup>首先借助行人身体关键点检测的方法,来检测非遮挡的身体区域,然后对非遮挡的身体区域提取特征进行匹配.然而,这需要准确的身体关键点标注信息.一旦关键点标注不准确,往往导致识别错误.并且,有时候遮挡的身体区域可能包含关键的判别信息,而非遮挡的身体区域具有相似的外观.因此,遮挡场景下的行人重识别仍然面临诸多困难.

最近,生成对抗网络(generative adversarial network,简称GAN)<sup>[7]</sup>在图像生成<sup>[8-10]</sup>、风格迁移<sup>[11,12]</sup>等领域展现出强大的图像生成能力.为此,本文提出了一种基于多尺度GAN的遮挡行人重识别方法.借助多尺度GAN,我们可以生成去遮挡的行人图像,从而有效地减少了遮挡在行人特征表示学习过程中的干扰.

具体来说,在多尺度GAN训练阶段,我们首先将无遮挡图像和遮挡图像输入到多尺度GAN(包含一个多尺度生成器和一个判别器)中.在该网络中,多尺度生成器的任务是对输入的遮挡图像尽可能地恢复遮挡区域的信息,并能够生成高质量的虚假图像,可以欺骗判别器;而判别器的任务是,判别输入图像是真实的原始图像还是虚假的生成图像.多尺度生成器和判别器相互对抗并不断调整各自的参数,最终的目的是使判别器无法判断多尺度生成器的输出结果是否真实.其次,我们将随机添加遮挡后的训练图像输入到训练好的多尺度生成器中,生成类似的但不完全一样的去除遮挡的图像(并保持原始身份标签),添加到原始训练图像集中,从而扩充了行人的图像数目.最后,基于此扩充训练图像集,训练一个分类识别模型.在识别场景的测试阶段,对于包含真实遮挡的查询图像,我们首先定位遮挡区域,然后输入到多尺度生成器中,得到去除遮挡的查询图像,最后进行识别.受益于遮挡行人图像中遮挡干扰的去除和有用信息的恢复,包含真实遮挡的查询图像的识别性能明显提高.

本文的主要贡献包括以下3点.

(1) 我们设计了一个新型的多尺度生成对抗网络,通过一种由粗到细的方式来逐步去除遮挡.我们采用低分辨率行人图像输入去重构高分辨率行人图像,有效地融合不同尺度的图像细节信息,从而提高行人图像去遮挡的性能.特别地,利用提出的多尺度生成对抗网络对于随机遮挡的行人图像可以生成类似但却不完全一样的行人图像,起到数据增强的作用,用来增加训练样本的多样性.而对于真实的遮挡行人图像而言,该网络可以起到去遮挡的作用,从而扩大遮挡行人图像中 useful 信息的区域.

(2) 我们提出了结合内容损失(包括像素损失和特征损失)和对抗损失作为多尺度生成对抗网络的损失函数.对低分辨率生成器采用像素损失,保持底层像素特征的一致性,而对高分辨率生成器采用特征损失,保持高层语义特征的一致性,从而有效地保留了输入图像的身份信息.同时,我们提出的多尺度生成对抗网络是一个预训练网络,可以与现有的行人重识别方法快速且有效地结合在一起,从而有效地提高行人重识别方法的准确率.

(3) 在多个有挑战性的行人重识别数据集中,通过大量的实验设计与分析,验证了我们所提方法的有效性.

本文第1节介绍行人重识别相关工作,第2节介绍本文提出的基于多尺度生成对抗网络的遮挡行人重识别方法,第3节为实验设置与结果分析,最后为本文结论。

## 1 相关工作

传统的行人重识别方法包括特征提取和相似性度量两个步骤,而基于深度学习的方法利用深度神经网络(deep neural network,简称DNN)有效整合上述两个步骤进行端到端的训练,从而有效地提高行人重识别的性能。由于行人图像通常存在没有对齐、遮挡、姿态多变等问题,因此很多学者针对这些问题提出不同的方法以用于学习鲁棒的行人特征表示。例如,设计不同的深度神经网络框架<sup>[13-15]</sup>,用于提取不同身体部件特征;或者设计不同深度度量损失函数<sup>[16,17]</sup>,用于学习判别性的行人特征表示。

针对遮挡场景下的行人重识别问题,现有的方法大致可以分成3类。第1类方法<sup>[18-20]</sup>通过直接添加遮挡样本来到训练集中的方式,来扩充训练集的多样性,从而增强深度神经网络对于遮挡行人图像的识别能力。例如,Zhuo等人<sup>[20]</sup>利用随机遮挡图像和非遮挡图像,并联合身份分类损失和遮挡/非遮挡的二分类损失训练一个深度神经网络。该网络不仅可以识别行人的身份,同时可以区分输入图像是属于遮挡图像数据分布还是非遮挡图像数据分布。这些方法利用随机遮挡样本增加了训练样本的多样性,提高了模型对于遮挡场景下行人重识别算法的性能。但是,由于真实场景下遮挡区域风格的多样性,如何有效添加不同类型的遮挡样本仍然是一个具有挑战性的问题。

上述方法学习到的行人特征表示,不可避免地会包含一些由于遮挡引入的噪音特征。因此,第2类方法<sup>[4,5,21,22]</sup>通过设计合适的特征选择和过滤策略的方式,来减轻遮挡引入的噪音对于行人特征的干扰,从而提高遮挡场景下的行人重识别性能。通常的做法是在特征比对的过程中,降低遮挡区域特征的权重。例如,Zhao等人<sup>[4]</sup>提出SpindleNet方法,利用姿态估计检测身体关键点并划分7个身体部件区域。同时,他们还提出树状结构的特征融合策略,在不同阶段分别融合不同语义层次的部件特征。针对遮挡问题,SpindleNet采用极大值竞争策略(即在不同部件特征融合过程中,按像素层比较特征值,保留极大值的特征),从而达到抑制噪音的目的。类似地,Su等人<sup>[21]</sup>提出PDC(pose-driven deep convolutional)模型,利用全连接层自动学习不同部件特征的权重,并削弱遮挡区域特征的权重。Xu等人<sup>[5]</sup>提出AACN(attention-aware compositional network)方法,利用身体关键点学习各个身体部件的注意力图,从而得到更精确的部件区域。同时,AACN方法利用部件的注意力图计算该部件区域的可视性分数(visibility score),用于衡量部件的遮挡情况。基于该可视性分数,不断调整各个部件特征在特征融合过程中的权重。这些方法需要借助额外的姿态标注信息。当姿态估计不准确时,行人重识别的性能会严重下降。最近,为了有效处理遮挡场景下的行人重识别问题,Yang等人<sup>[22]</sup>提出了利用长短期记忆(long short-term memory,简称LSTM)网络学习行人图像不同部件区域之间的空间关系,从而自动过滤由于遮挡引入的噪音。但是,这类方法通常需要额外的身体标注信息和姿态估计。当身体标注信息和姿态估计不准确时,行人重识别的性能会严重下降。

第3类方法<sup>[23]</sup>利用特定的技术(如生成对抗网络<sup>[7]</sup>)重构出遮挡行人图像中遮挡区域的图像信息,以达到去遮挡的目的,从而有效提高遮挡场景下的行人重识别效果。例如,Hou等人提出STCNet(spatio-temporal completion network)方法<sup>[23]</sup>,通过恢复遮挡区域的表现来处理部分遮挡问题,用于基于视频的行人重识别任务。在STCNet的网络结构中,空间结构生成器根据行人图像的空间结构,利用可见(未遮挡)的身体区域来预测行人缺失(遮挡)的区域。而时间注意力生成器,则利用视频序列中相邻帧的信息来恢复当前帧遮挡区域的表现。

相比于第1类和第2类方法,第3类方法可以直接恢复出遮挡区域的图像信息,而不需要身体标注信息和姿态估计,也无需生成不同类型的遮挡样本。本文提出的方法属于第3类方法。本文提出的方法与STCNet方法处理遮挡行人重识别问题的出发点类似,即直接对遮挡图像进行去遮挡预处理,从而在行人重识别的测试阶段减少遮挡对于性能的影响。然而,在具体实现上,本文提出的方法和STCNet方法截然不同。本文提出多尺度生成对抗网络,基于随机遮挡与非遮挡图像对,通过一种由粗到细的方式来逐步去除遮挡。我们采用低分辨率行人图像输入去重构高分辨率行人图像,从而充分提高行人图像去遮挡的性能。而STCNet方法利用提出空间结构生

成器初步重构出去遮挡的行人图像,然后基于提出的时间注意力生成器优化该重构图像的质量,网络结构相对复杂.因此,相比于 STCNet 方法,本文的优点在于针对基于图像的行人重识别问题,可以更简单、有效地提高行人图像的去遮挡性能,而且不需要借助视频序列相邻帧的信息.

此外,目前已有很多研究者提出了面向各种复杂场景的基于 GAN 的行人重识别方法,代表性的工作有文献 [24-27] 中介绍的相关研究方法.这些方法借助 GAN 强大的图像生成能力,可以生成多样的新的图像数据,从而增加有标签数据,有效提高模型在测试集上的泛化能力.例如,Zhong 等人<sup>[24]</sup>提出摄像机风格自适应方法 (CamStyle),可以作为一种数据增强方法来减少不同摄像机风格之间的差异.具体来说,CamStyle 使用 CycleGAN<sup>[11]</sup>,在保持行人身份不变的情况下,学习不同摄像机视角的图像对之间的迁移关系.从而可以将有标记的行人图像风格迁移到其他摄像机视角,添加到原始训练集中,起到数据增强的作用.Qian 等人<sup>[26]</sup>提出姿态规范化生成对抗网络(PNGAN),可以基于不同姿态图生成相应的姿态迁移行人图像,从而学习姿态不变的行人特征表示.

本文方法与上述基于 GAN 的方法<sup>[24-27]</sup>存在较大差异,包括以下 3 点不同.首先,本文提出的多尺度生成对抗网络在训练阶段是基于非遮挡与随机遮挡图像对用来自动重构部分遮挡区域,并不需要额外的关键点标注信息<sup>[26]</sup>或者摄像机视角信息<sup>[24]</sup>;其次,本文采用多尺度生成器(具有网络结构简单的特点)进行重构,可以充分利用多尺度图像的细节信息,重构出高质量的行人图像.而上述方法<sup>[24-27]</sup>一般采用单尺度生成器对行人图像进行重构.为了生成高质量的行人图像,采用的网络结构相对复杂;最后,本文提出的多尺度生成器可以在识别场景的测试阶段,用于对真实遮挡行人图像进行去遮挡,从而提高遮挡行人重识别性能.而上述方法<sup>[24-27]</sup>一般只是通过生成对抗网络生成新的训练样本增强分类识别模型的泛化性能,而并不能用于提升遮挡场景的识别性能(上述方法无法适用于遮挡场景下的行人重识别问题).

## 2 基于多尺度生成对抗网络的遮挡行人重识别

在本节中,首先在第 2.1 节介绍本文提出方法的整体框架和算法流程.其次,在第 2.2 节和第 2.3 节分别介绍提出方法中的两个网络,即去除遮挡的多尺度生成对抗网络和分类识别网络.

### 2.1 整体框架

我们提出的基于多尺度生成对抗网络的遮挡行人重识别方法,包括两个网络:去除遮挡的多尺度生成对抗网络和分类识别网络.我们提出算法的整体框架如图 1 所示.该整体框架包括一个多尺度生成器  $mG$  和一个判别器  $D$  以及一个分类识别网络  $R$ .其中,多尺度生成器的输入是小尺度的遮挡图像,输出是大尺度的去遮挡生成图像;判别器的输入是大尺度的生成图像和原始图像.分类识别网络的训练数据是大尺度的生成图像和原始图像.网络中参数说明见第 2.2 节.算法包括训练过程和测试过程,具体流程如下.

训练过程包括 3 个阶段.

第 1 个阶段,我们事先从原始训练图像集筛选出部分干净的、无遮挡的训练图像.对于选中的每张干净图像,我们人工添加随机遮挡,生成含有随机遮挡的训练图像(具体实现细节见第 3.1 节),从而形成遮挡-非遮挡图像对.基于遮挡-非遮挡图像对,我们训练多尺度生成对抗网络.该网络包括一个由小尺度到大尺度(对应了粗略到细致的行人特征表示)的多尺度生成器  $mG$  和一个判别输入图像是真实原始图像还是重构虚假图像的判别器  $D$ .本文采用 WGAN-GP<sup>[10]</sup>的训练目标函数,交替训练多尺度生成器和判别器,直至收敛.

第 2 个阶段,我们利用训练好的多尺度生成器,对随机遮挡的原始训练图像进行离线遮挡去除.然后将去除遮挡后的训练图像添加到原始训练图像集中,并保留原始的身份标签,从而得到扩充的训练图像集.

第 3 个阶段,基于扩充的训练图像集,训练分类识别网络.我们的分类识别网络是基于改进的 ResNet-50<sup>[28]</sup>网络.具体来说,在原始 ResNet-50 网络的基础上,首先拼接原始 ResNet-50 网络最后 3 个卷积层的输出特征,并添加一个全连接层进行特征降维得到紧凑的特征表示.然后,根据不同的训练数据集,修改最后的全连接层(分类层)的神经元个数.最后,采用 Softmax 分类损失训练该网络.

在测试阶段,为了处理实际场景中的真实遮挡,本文使用手工定位遮挡区域(在实际应用中,可以利用行人

关键点检测的方法来对遮挡的区域进行定位);然后在真实遮挡区域中添加随机遮挡(具体实现细节见第 3.1 节),从而使得随机遮挡图像与训练图像的风格保持一致;接着送入多尺度生成器中,达到去遮挡效果,从而得到去除遮挡后的查询图像.基于训练好的分类识别网络,将分类层之前的全连接层输出作为行人特征表示.对去除遮挡后的查询图像、原始查询图像以及行人数据库中的所有图像,分别提取行人特征、计算查询图像特征(即去除遮挡后的查询图像特征和原始查询图像特征的平均特征)和行人数据库中所有图像特征的相似性(本文基于欧氏距离).最后按照相似性从大到小进行排序,从而完成行人重识别任务.

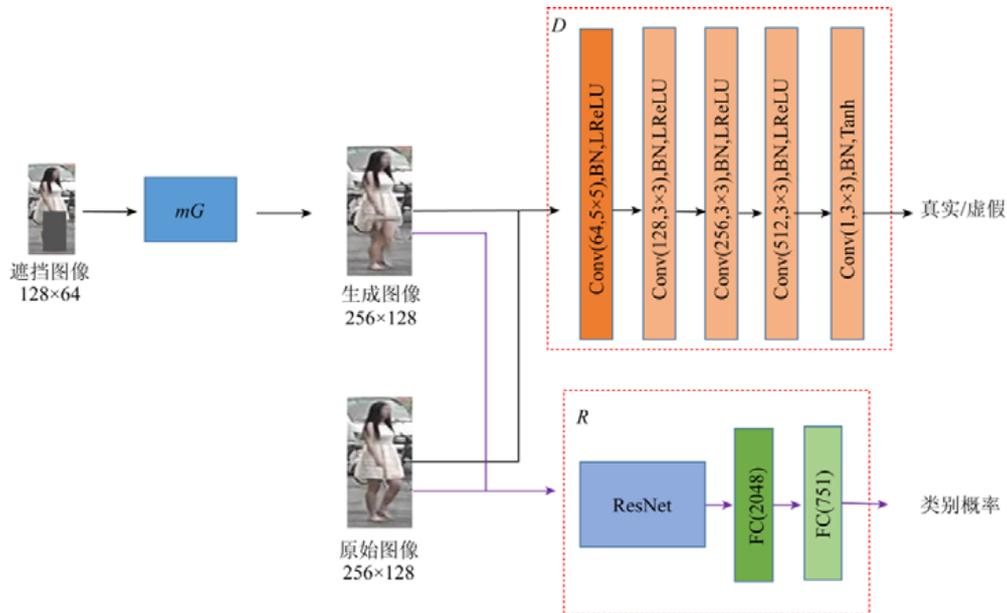


Fig.1 The overall framework of our proposed method

图 1 本文提出方法的整体框架

## 2.2 多尺度生成对抗网络

在本小节中,我们将详细介绍提出的多尺度生成对抗网络(multi-scale GAN).该网络包括一个多尺度生成器  $mG$  和一个判别器  $D$ .下面我们将分别介绍  $mG$  和  $D$  的网络结构以及所使用的对抗损失函数.

现有大部分基于 GAN 的行人重识别方法<sup>[24-27]</sup>在利用 GAN 进行图像变换时,通常对输入图像和输出图像均采用相同的分辨率.为了有效去除遮挡图像,网络结构中需要学习的参数过多,导致计算十分复杂.同时,大部分方法只采用 CNN 得到的特征来重构行人图像,存在一定程度的过拟合.基于上述问题,本文提出了一种多尺度 GAN 网络.该网络设计的基本思想是通过一种由粗到细的方式来逐步去除遮挡.采用低分辨率行人图像输入去重构高分辨率行人图像(有效融合多个尺度的图像细节信息),从而充分提高行人图像去遮挡性能.

我们提出多尺度生成器  $mG$  包括两个生成器  $E$  和  $G$ .第 1 个低分辨率生成器  $E$ ,输入的是小尺度的遮挡图像.该生成器的主要目的是粗略地重构去遮挡的行人图像;第 2 个高分辨率生成器  $G$ ,输入的是大尺度的遮挡图像(由小尺度遮挡图像插值得到).该生成器的主要目的是重构高分辨率的去遮挡行人图像.在  $G$  的网络结构中,我们融合了  $E$  的图像特征(输出重构图之前)来生成高分辨率图像.该做法类似密集连接(dense connection),下一层融合上一层的特征输出,从而可以有效克服 GAN 训练过程中的过拟合问题.本文小尺度采用的图像大小为  $256 \times 128$ ,大尺度采用  $256 \times 128$ .

多尺度生成器  $mG$  网络结构简单,如图 2 所示,包括两个生成器  $E$  和  $G$ . $E$  的输入是小尺度的遮挡图像,输出是小尺度的去除遮挡图像. $G$  的输入是大尺度的遮挡图像,输出是大尺度的去除遮挡图像.具体来说,生成器  $E$  和  $G$  都包含若干个卷积层、残差块(包括两个卷积层和残差连接)以及若干个反卷积层.其中, $E$  包括 3 个卷积层、

3个残差块、2个反卷积层;G包括4个卷积层、9个残差块、3个反卷积层.图2所示的每个卷积层的表示说明如下:Conv( $n,m \times m$ ),IN,ReLU 分别表示卷积核数量是  $n$  个,卷积核大小是  $m \times m$ ,默认设置的填充和步长是 0 和 1;卷积层随后跟着一个实例规范化层(instance normalization,简称 IN)和一个激活层(激活函数采用 ReLU).类似地,反卷积层 DeConv( $n,m \times m$ ),IN,ReLU 中也包含了相同的实例规范化层和激活层.

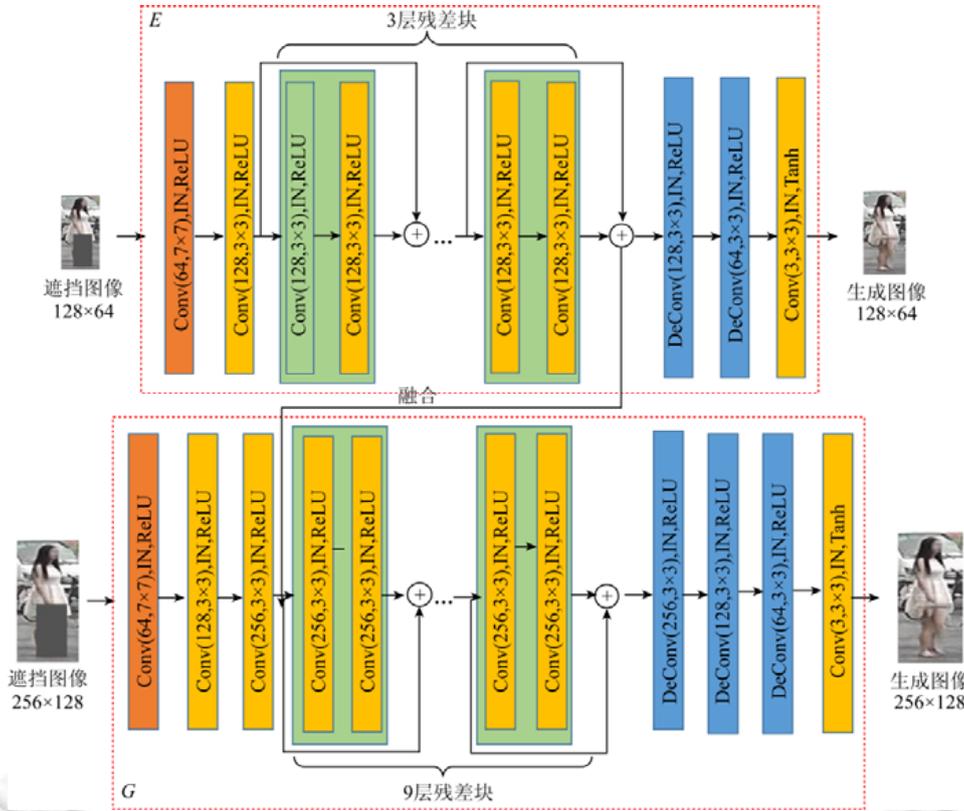


Fig.2 The framework of the multi-scale generator  $mG$

图2 多尺度生成器  $mG$  的网络框架

判别器  $D$  的网络结构如图 1 所示,只包括 5 个卷积层.在图 1 中,BN(batch normalization)表示批量归一化层,Tanh 和 LReLU 分别表示激活函数为 Tanh 和 Leaky ReLU 的激活层.

为了使得多尺度生成器输出的生成图像看起来尽可能逼近真实图像,我们提出结合内容损失(包括像素损失和特征损失)和对抗损失作为整个网络的损失函数.具体来说,我们对生成器  $E$  采用  $l_1$  像素损失,用来计算真实图像和生成图像在像素空间上的差异,保证可以粗略重构出去遮挡的行人图像. $l_1$  像素损失用来保持底层像素特征的一致性.而对生成器  $G$  则使用 VGG-19 特征损失,计算真实图像和生成图像在预训练模型 VGG-19<sup>[29]</sup> 上的卷积特征(本文采用的是第 14 层的输出)之间的差异,尽可能地保持行人图像高层语义特征的一致性.

$l_1$  像素损失表示为

$$L_p = \|X - E(Z)\|_1 \tag{1}$$

其中, $Z$  表示遮挡图像, $X$  表示对应的无遮挡图像; $\|\cdot\|_1$  表示  $l_1$  范数; $E(\cdot)$  表示生成器  $E$  的重构映射变化函数.

VGG 特征损失表示为

$$L_f = \|f(X) - f(G(Z))\| \tag{2}$$

其中, $f(\cdot)$  表示 VGG-19 模型的卷积特征映射函数; $\|\cdot\|$  表示  $l_2$  范数; $G(\cdot)$  表示生成器  $G$  的重构映射变化函数.

我们采用 WGAN-GP<sup>[10]</sup>作为对抗损失,它可以表示为

$$L_{adv} = E_{X \sim P_c} [D(X)] - E_{Z \sim P_o} [mG(Z)] + \lambda E_{\hat{X} \sim P_c} \left[ \left( \left\| \nabla_{\hat{X}} D(\hat{X}) \right\|_2 - 1 \right)^2 \right] \quad (3)$$

其中,  $P_c$  和  $P_o$  分别表示非遮挡图像  $X$  和遮挡图像  $Z$  的数据分布;  $mG(\cdot)$  表示多尺度生成器的重构映射变化函数;  $\lambda$  表示权重参数;  $\hat{X} = tX + (1-t)mG(Z)$  ( $0 \leq t \leq 1$ ) 表示的是从  $P_c$  和  $P_o$  采样的数据分布;  $\nabla_{\hat{X}}$  表示  $D$  关于图像  $\hat{X}$  的梯度. 具体解释可以参考 WGAN-GP<sup>[10]</sup>.

因此,我们提出的多尺度生成对抗网络的整体损失函数可以表示为

$$L = L_{adv} + \lambda_1 L_p + \lambda_2 L_f \quad (4)$$

其中,  $\lambda_1$  和  $\lambda_2$  控制  $L_1$  像素损失和特征损失的权重. 根据实验经验, 本文均设置为 10.

基于上述整体损失函数, 我们交替训练多尺度生成器  $mG$  和判别器  $D$ , 从而生成纹理清晰、视觉效果逼真的行人图像.

### 2.3 分类识别网络

我们采用在大规模数据集 ImageNet 上预训练的 ResNet-50<sup>[28]</sup> 模型为基础, 将 ResNet-50 最后 3 个卷积层的输出特征进行拼接, 并添加一个全连接层将特征维度降至 2 048 维, 用来作为我们的分类识别网络. 同时按照训练数据集中的行人身份类别, 修改最后的全连接层(分类层)的神经元个数(比如 Market1501 的训练集包括的行人身份为 751 个, 则分类层相应的神经元个数设置为 751).

如第 2.1 节中所述, 在完成第 1 个阶段生成对抗网络的训练后, 在第 2 个阶段可以得到扩充的训练图像集. 基于该扩充训练图像集, 我们采用 Softmax 分类损失训练该分类识别网络, 如公式(5)所示.

$$L_s = -\sum_{i=1}^B \log(p_i^{y_i}) \quad (5)$$

其中,  $B$  表示小批次训练样本数量.  $p_i^{y_i}$  表示样本  $x_i$  经过 Softmax 层计算的样本属于真实类别  $y_i$  的预测概率.

训练好该分类识别网络之后, 我们将分类层之前的全连接层输出作为最后的鲁棒行人特征表示, 并进行后续的计算(见第 2.1 节).

## 3 实验结果及分析

在本节中, 我们首先说明本文提出方法的实验设置. 其次, 分析本文提出的多尺度生成对抗网络的模型简化测试(ablation study), 即分析采用不同类型的生成器, 对于生成图像视觉效果的影响和对于行人重识别性能的影响. 最后, 与现有的先进方法进行行人重识别性能比较, 包括不同遮挡比例的遮挡行人重识别性能对比.

### 3.1 实验设置

**数据集和评价标准.** 为了验证本文提方法的有效性, 我们在 4 个具有挑战性的行人重识别数据集上进行大量的实验, 包括 Market1501<sup>[30]</sup>、DukeMTMC-reID<sup>[31]</sup>、CUHK03<sup>[32]</sup> 和 Partial REID<sup>[33]</sup> 数据集.

Market1501 数据集包括 1 501 个不同身份的行人图像, 由 6 个不同视角的摄像机拍摄得到, 其中图像数据划分成 751 个身份的行人图像(总共 12 936 张图像)用于训练, 750 个身份的行人图像(包括 19 732 张图像作为数据库图像和 3 368 张图像作为查询图像)用于测试.

DukeMTMC-reID 数据集包括 1 404 个不同身份的行人图像, 由 8 个不同视角的摄像机拍摄得到, 其中图像数据划分成 702 个身份的行人图像(总共 16 522 张图像)用于训练, 702 个身份的行人图像(包括 17 661 张图像作为数据库图像和 2 228 张图像作为查询图像)用于测试.

CUHK03 数据集包括 1 467 个不同身份的行人, 由 6 个不同视角的摄像机拍摄得到, 其中, 1 367 个不同身份的行人图像(总共 13 132 张图像)用于训练, 100 个不同身份的行人图像(总共 965 张图像作为数据库图像和查询图像)用于测试.

Partial REID 数据集包括 60 个不同身份的行人. 每个身份的行人包括 5 张全身图像和 5 张部分遮挡的图像.

我们使用 30 个不同身份的行人图像(总共 300 张图像)用于训练,其余用于测试。

我们使用标准的评价指标来衡量行人重识别的性能,包括平均正确率均值(表示为 mAP)和累积匹配特性曲线(即 CMC)在第 1 匹配率的值(表示为 Rank-1)。

**实验环境.**我们使用 Pytorch 平台进行代码编写,在配置 Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10GHz 和 GTX TITAN GPU 的服务器上进行实验。

**多尺度 GAN.**在 Market1501、DukeMTMC-reID 和 CUHK03 数据集上分别训练我们的多尺度 GAN。我们分别选择了 7 508 张、7 956 张和 13 132 张干净(非遮挡)的训练图像,进行随机遮挡。具体实现细节如下:随机选择不同大小的矩形框(矩形框大小与原图大小比例  $s$  控制在 0.3 以内),用来替换矩形框内的像素值为(0,255)的随机值,从而生成相应的遮挡图像,形成遮挡-非遮挡图像对。由于 Partial REID 数据集中训练图像数目过少,因此我们首先在 Market1501 数据集上预训练多尺度生成器和判别器,然后利用该数据集上的训练图像进行微调。多尺度生成器输入的图像对大小都是  $128 \times 64$ ,输出图像大小是  $256 \times 128$ 。多尺度 GAN 判别器的输入图像大小都是  $256 \times 128$ 。我们采用 Adam 优化器优化整个生成对抗网络。动量、衰减率和初始学习率分别设置为 0.5、0.000 5 和 0.000 1。对于多尺度 GAN,总共训练 150 个迭代次数。

需要说明的是,由于目前现有的数据集中无法提供足够多的真实遮挡图像与对应的非遮挡图像数据,本文首先利用多尺度生成器对随机遮挡的训练图像(即对无遮挡图像添加随机遮挡)进行去遮挡,以生成新的训练样本,并添加到原始训练集中,从而构建一个扩增训练集。该做法有两方面好处。一方面,随机遮挡图像中随机遮挡出现的区域是不固定的,这与真实遮挡场景中遮挡区域不固定的特点是一致的。另一方面,随机遮挡图像有效去除了行人图像中遮挡区域内的各种变化,使得多尺度 GAN 可以将该去遮挡任务转化为一个相对简单的图像生成任务。并且,本文联合使用内容损失和特征损失,用来保证去遮挡后的生成图像的真实性。尽管这些生成的遮挡图像与真实场景下的遮挡图像相比,存在较大的风格差异,但在遮挡图像和对应非遮挡图像数据不足的情况下,实验结果(见第 3.2 节和第 3.3 节)表明了随机遮挡的有效性和可行性。

**分类识别网络.**该网络的输入图像大小是  $256 \times 128$ 。我们采用 Adam 优化器优化整个分类识别网络。动量、衰减率和初始学习率分别设置为 0.5、0.000 5 和 0.000 3。分类识别网络在 Market1501、DukeMTMC-reID 和 CUHK03 数据集上,我们分别训练 60 个迭代次数。而在 Partial REID 数据集上,由于训练图像数目过少,只训练 30 个迭代次数。

**测试设置.**为了充分评估遮挡场景下的行人重识别性能,我们测试了原始图、真实遮挡图、随机遮挡图这 3 种不同设置条件下的识别效果。原始图是指采用原始数据集提供的查询图像,这些图像包含有遮挡和无遮挡的图像。真实遮挡图是指从原始图中筛选的包含真实遮挡的查询图像,并选取遮挡区域(本文采用人工的方式,也可以利用检测方法自动检测遮挡区域)将区域内的像素值替换为(0,255)的随机值。随机遮挡图是指对所有原始图进行随机遮挡,从而得到比较多的遮挡图像来模拟遮挡场景下的图像。原始图直接基于原始图像提取特征并进行识别。真实遮挡图像和随机遮挡图像则需要经过多尺度生成器进行去遮挡处理后再提取特征进行识别,能够用于比较不同遮挡场景下行人重识别的性能。

在第 3.2 节多尺度生成对抗网络的实验中,我们利用 DukeMTMC-reID 数据集和 Partial REID 数据集进行实验。在 DukeMTMC-reID 数据集上测试了原始图、真实遮挡图、随机遮挡图的识别效果。其中,在 DukeMTMC-reID 数据集上,我们从原始图中(2 268 张)筛选出包含真实遮挡的图像(758 张)进行测试。而 Partial REID 数据集中的原始图(所有查询集)即为真实遮挡图像,总共 150 张。

在第 3.3 节与现有先进方法的对比实验中,考虑到 Market1501 和 CUHK03 数据集的查询集中包含真实遮挡图的数量较少( $<100$  幅),因此没有进行真实遮挡图的实验测试(如上所述,DukeMTMC-reID 和 Partial REID 上真实遮挡图的实验结果在第 3.2 节给出)。在 3 个较大规模的数据集上(DukeMTMC-reID、Market1501 和 CUHK03),我们只比较了原始图(即遮挡比例为  $s=0$ )和随机遮挡图(包括遮挡比例为  $s=0.3$  和  $s=0.6$ )的实验结果。

### 3.2 多尺度生成对抗网络的模型简化测试实验结果

在本小节中,我们从生成图像的视觉效果和对行人重识别性能的影响两个方面来评估多尺度生成对抗网络的

有效性.我们与本文提出方法的两个变形方法进行对比和分析.这两个变形方法分别为 GD 和 EGD.它们的主要不同在于生成器的网络结构,而判别器的网络结构都相同.具体来说,GD 只采用一个生成器(生成器的网络模型和图 2 中 G 模块的参数一致),并利用对抗损失和特征损失联合训练整个 GAN 网络.生成器的输入和输出图像大小都是  $256 \times 128$ .EGD 采用两个生成器,这两个生成器的网络结构与我们的方法(如图 2 所示)相同,但是两个生成器的输入图像都是相同尺度(即  $256 \times 128$ ).而 mGD 为本文提出的方法,采用多尺度 GAN.综上,我们将对比和分析 3 种不同的方法(即 GD、EGD 和 mGD)

### 3.3 定性评价

首先,我们给出本文提出方法采用的多尺度 GAN 在 DukeMTMC-reID 训练集中对随机遮挡图进行去遮挡后得到的生成图像,如图 3 所示.



Fig.3 The de-occluded images of the proposed mGD method on the randomly occluded images on the Duke MTMC-reID training set

图 3 本文提出的方法 mGD 在 DukeMTMC-reID 训练集中对随机遮挡图上的去遮挡结果示例图

在图 3 中我们可以看出,对于训练集中的每一张行人图像,我们首先进行随机遮挡,再利用多尺度生成器进行去遮挡,最后得到与原始图像相似(保留行人身份信息)但不完全一样的行人图像.这些生成图像相比于原始图像,包含不同的姿态变化,如第 1 列的 3 张生成图,其下半身姿态相对原始图像有比较明显的不同.也有一些是去除了原始图像中的一些背包,手提包等,如第 3 列的前两张生成图.这些生成图像添加到原始训练数据集中,可以增加训练数据的多样性.因此,在训练分类识别网络的阶段,多尺度 GAN 可以起到数据增强的作用.

其次,我们给出了 3 个不同的方法在 DukeMTMC-reID 查询集中对真实遮挡图进行去遮挡得到的生成图像,如图 4 所示.从第 1 行到第 5 行分别是原始图、遮挡图、GD 生成的图像、EGD 生成的图像和 mGD 生成的图像.

从第 1 列~第 5 列,查询图像中的遮挡类型越来越复杂(包括交通标志,车辆,行人,建筑物,书包等),相应的遮挡区域面积也随之增加.可以看出,我们提出的 mGD 方法,可以基于遮挡图像,生成缺失的身体部件,从而增加有效的目标信息,提高遮挡图像的识别性能.相比于 GD 和 EGD 方法,我们的方法 mGD 得到的行人图像在身体部件的生成能力明显更强并且生成图像的细节表现上更加真实,比如第 1 列图像和第 4 列图像,生成的腿部件看

起来比较真实.在第 4 列图像中,当出现下半身缺失的时候,GD 和 EGD 的生成质量明显变差,而 mGD 可以较好地生成缺失的下半身.但是需要指出的是,对于一些严重遮挡的图像(比如遮挡区域超过原图面积比例 0.7 时),这 3 种方法还是对遮挡区域还是难以恢复出有效图像信息.比如第 5 列的遮挡图像,遮挡区域非常大,可视信息只剩下头部.所有方法均会失效.



Fig.4 The de-occluded images obtained by different methods on the occluded query set of Duke MTMC-reID

图 4 不同的对比方法在 DukeMTMC-reID 查询集的真实遮挡图上的去遮挡结果示例图

### 3.4 定量评价

在本小节中,我们主要从性能评价方面,分析不同的方法对于行人重识别性能的影响.首先,我们对比 3 种不同的方法在 DukeMTMC-reID 数据集上的实验结果.具体来说,这 3 种方法分别采用 GD、EGD、mGD 中的生成器(进行去遮挡)生成扩充训练数据集,然后基于扩充训练数据集(生成数据与真实数据的比例设置为 1:1),分别训练相同的分类识别网络 R(我们用 GD+R,EGD+R,mGD+R 来表示这 3 种方法).同时,为了验证我们提出方法的有效性,我们也给出了基准方法(即只使用原始训练数据集训练分类识别网络 R)的实验结果.我们得到如图 5(a)所示的识别结果,包括原始图、真实遮挡图和随机遮挡图上的 Rank-1 识别结果.图 5 中,DukeMTMC-reID 数据集上包括原始图、真实遮挡图以及随机遮挡图的识别结果;Partial REID 数据集上的原始图(查询集图像)即为遮挡图像,可以直接用于比较遮挡场景下的识别性能.

从图 5(a)中可以得到以下结论.首先,基准方法 R 在所有测试图像中取得的性能最差,尤其是在随机遮挡图上的识别结果有明显的下降.这主要是由于测试阶段出现的遮挡样本携带大量的噪音,而 R 中训练的分类识别网络并没有针对性对这些噪音样本进行学习.这说明了遮挡是限制行人重识别性能的一个重要因素.

其次,从原始查询图上的识别结果,可以看出相比于只采用原始训练数据集(Rank-1 达到 76.6%),基于去遮挡模型产生增强训练数据的识别方法(包括 GD+R,EGD+R 和 mGD+R)可以明显提高识别结果(分别为

79.0%,78.6%和 79.8%).其中,mGD 对于识别结果的提升是最多的(Rank-1 的准确率提升了 3.2%).而且,mGD+R 得到的 Rank-1 准确率相比于 GD+R 和 EGD+R,分别提高了 0.8%和 1.2%.这主要是由于我们的方法有效地融合多个尺度的图像细节信息,通过由粗到细的方式来逐步去除遮挡,从而进一步提高行人图像的去遮挡性能.因此,经过 mGD 方法生成的图像有效地增加了训练数据的多样性,充分地提高了模型的泛化性并提升最终的识别效果.

在真实遮挡图上,相比于基准方法 R,我们的方法 Rank-1 准确率提升了 1.6%;相比于 GD+R,EGD+R,我们的方法也有 0.4%和 0.5%的提升.可以看出,mGD 可以有效提高真实遮挡场景下的行人重识别效果.真实遮挡场景下的图像,往往遮挡区域更大.

在随机遮挡图上的测试实验中,采用数据增强的 3 种方法对于行人重识别性能的提升更加明显,Rank-1 准确率分别提升了 15.1%,15.4%,16.2%.可以看出,通过对遮挡图像进行去遮挡,有效地减少遮挡物体的干扰,扩大目标区域的图像信息,从而提升遮挡场景下的行人重识别性能.

不同的对比方法在 Partial REID 数据集上的识别结果,如图 5(b)所示.正如第 3.1 节所述,由于该数据集上包含的训练集数量偏少,因此对于这些对比方法,我们利用在 Market1501 数据集上训练得到的生成器和判别器进行微调.Partial REID 的查询集即为真实遮挡图像,基准方法 R 在原始图(即遮挡图)上的识别结果仅有 72.3%,而这 3 种对比方法对原始图进行去遮挡后的识别结果都有明显的提升.这主要得益于这些方法的去遮挡预处理操作以及数据增强作用.相比于 DukeMTMC-reID 数据集,该数据集上的遮挡区域会小一些并且遮挡类型也少一些,因此我们提出的方法在该数据集上的真实遮挡场景下的识别结果有明显地提升.这也充分说明了多尺度 GAN 的有效性.

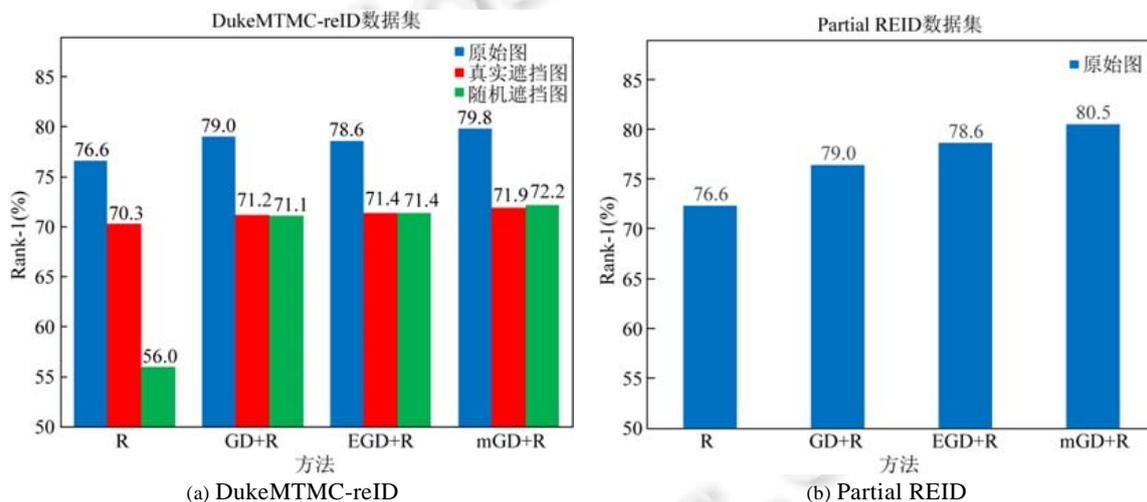


Fig.5 Rank-1 results obtained by different methods (i.e., R, GD+R, EGD+R, mGD+R) on the DukeMTMC-reID and the Partial REID datasets

图 5 不同的方法(R,GD+R,EGD+R,mGD+R)在 DukeMTMC-reID 和 Partial REID 数据集上的 Rank-1 识别结果

此外,为了分析不同数量的生成数据(利用 mGD 生成)对行人重识别性能的影响,我们在 DukeMTMC-reID 测试集上对扩增训练集中生成数据与真实数据的比例选取不同的值(选取了 0.5:1、1:1、3:1、5:1)进行识别网络 R 的训练.在原始图、真实遮挡图和随机遮挡图上的 Rank-1 准确率和 mAP 识别结果见表 1.

从表 1 可以看出,当生成数据的数量不断增加,行人重识别的性能略有提升.当生成数据与真实数据的比例为 5:1 的情况下,Rank-1 准确率和 mAP 的值最大.这充分说明了利用本文提出的多尺度生成网络可以生成类型多样的训练样本,提高网络的训练能力.但是,由于生成数据的分布和真实数据的分布存在一定的偏差.如果生成数据数量过多,分类网络主要学习生成数据的分布,而忽略真实数据,可能会导致分类网络对生成数据过拟合.另外,由于训练数据增多,网络的训练复杂度变大.因此,为了防止过拟合问题和降低计算复杂度,在下面的实

验中,本文对生成数据与真实数据的比例仍然固定为 1:1.

**Table 1** Influence of different ratios of real data and synthetic data on the performance (the rank-1 (%) accuracy and mAP (%) on the DukeMTMC-reID dataset

**表 1** 在 DukeMTMC-reID 测试集上生成数据与真实数据的比例对行人重识别性能 (Rank-1(%)准确率和 mAP(%))的影响

比例	原始图		真实遮挡图		随机遮挡图	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
0.5:1	79.3	62.4	60.7	46.0	69.7	53.5
1:1	79.5	62.9	61.7	47.9	70.6	54.1
3:1	79.8	63.5	63.2	48.8	72.6	55.8
5:1	80.0	64.1	63.7	49.3	72.6	56.6

### 3.5 与现有方法的比较

在本小节中,我们在 3 个广泛使用的公开行人重识别数据集上,包括 Market1501,DukeMTMC-reID,CUHK03 数据集,将本文提出的方法与现有的行人重识别方法进行比较.对比方法包括传统手工特征和度量学习的方法 (NPD<sup>[34]</sup>,XQDA<sup>[3]</sup>) 和最近提出的深度学习的方法 (IDE<sup>[30]</sup>,TriNet<sup>[35]</sup>,Quad<sup>[36]</sup>,P2S<sup>[37]</sup>,RandEra<sup>[18]</sup>,GLAD<sup>[6]</sup>,PAN<sup>[38]</sup>,SVDNet<sup>[39]</sup>,DPFL<sup>[40]</sup>,AACN<sup>[5]</sup>,RNLSTMA<sup>[22]</sup>).我们分别比较了不同算法在原始图(即  $s=0$ )和随机遮挡图(包括  $s=0.3$  和  $s=0.6$ )上的识别结果,见表 2~表 4.其中,最好的结果用粗体表示.

**Table 2** The rank-1 (%) accuracy and mAP (%) obtained by the proposed method and the state-of-the-art methods against the different levels of occlusions on the Market1501 datasets

**表 2** 本文方法与其他方法在 Market1501 数据集上不同遮挡比例下的 Rank-1(%)准确率和 mAP(%)

方法	$s=0$		$s=0.3$		$s=0.6$	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
XQDA <sup>[3]</sup> , 2015	43.0	21.7	28.3	28.3	28.3	12.0
NPD <sup>[34]</sup> , 2016	55.4	30.0	39.6	19.1	32.5	16.1
IDE <sup>[30]</sup> , 2016	81.9	61.0	62.4	48.2	45.6	36.4
TriNet <sup>[35]</sup> , 2017	83.2	64.9	68.6	54.7	47.9	38.9
Quad <sup>[36]</sup> , 2017	86.3	72.2	65.6	54.8	44.8	38.2
P2S <sup>[37]</sup> , 2017	69.9	50.1	36.2	27.0	35.8	25.9
RandEra <sup>[18]</sup> , 2017	85.8	68.4	73.8	58.7	51.4	38.9
GLAD <sup>[6]</sup> , 2017	89.9	71.0	59.5	40.7	46.0	33.6
PAN <sup>[38]</sup> , 2017	81.0	63.4	52.0	36.5	43.2	30.0
SVDNet <sup>[39]</sup> , 2017	81.4	61.2	62.3	46.9	52.0	40.3
DPFL <sup>[40]</sup> , 2017	88.6	72.6	-	-	-	-
AACN <sup>[5]</sup> , 2018	85.9	66.9	-	-	-	-
RNLSTMA <sup>[22]</sup> , 2019	90.6	76.9	77.0	64.0	53.1	45.1
mGD+R	89.3	74.3	82.1	66.9	60.9	48.6
mGD+RNLSTMA	<b>91.3</b>	<b>77.9</b>	<b>85.1</b>	<b>71.2</b>	<b>65.8</b>	<b>53.2</b>

**Table 3** The rank-1 (%) accuracy and mAP (%) obtained by the proposed method and the state-of-the-art methods against the different levels of occlusions on the DukeMTMC-reID datasets

**表 3** 本文方法与其他方法在 DukeMTMC-reID 数据集上不同遮挡比例下的 Rank-1(%)准确率和 mAP(%)

方法	$s=0$		$s=0.3$		$s=0.6$	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
XQDA <sup>[3]</sup> , 2015	31.2	17.2	20.5	10.6	17.4	9.4
NPD <sup>[34]</sup> , 2016	46.7	27.3	33.7	17.7	29.7	15.7
IDE <sup>[30]</sup> , 2016	66.3	45.2	57.9	41.6	39.0	28.4
TriNet <sup>[35]</sup> , 2017	71.4	51.6	56.0	40.8	39.0	28.4
Quad <sup>[36]</sup> , 2017	73.4	58.0	61.6	47.7	43.1	33.9
P2S <sup>[37]</sup> , 2017	58.7	40.5	45.2	31.5	33.5	22.9
RandEra <sup>[18]</sup> , 2017	73.3	57.0	62.9	47.4	47.9	35.1
PAN <sup>[38]</sup> , 2017	71.6	51.5	44.7	29.0	39.9	25.9
SVDNet <sup>[39]</sup> , 2017	75.9	56.3	59.1	43.5	50.6	37.9
DPFL <sup>[40]</sup> , 2017	79.2	60.6	-	-	-	-
AACN <sup>[5]</sup> , 2018	76.8	59.3	-	-	-	-
RNLSTMA <sup>[22]</sup> , 2019	77.4	62.5	70.2	58.6	52.3	41.5
mGD+R	79.5	62.9	71.4	55.5	58.6	44.4
mGD+RNLSTMA	<b>80.8</b>	<b>63.9</b>	<b>74.1</b>	<b>58.8</b>	<b>63.0</b>	<b>47.7</b>

**Table 4** The rank-1 (%) accuracy and mAP (%) obtained by the proposed method and the state-of-the-art methods against the different levels of occlusions on the CUHK03 datasets

**表 4** 本文方法与其他方法在 CUHK03 数据集上不同遮挡比例下的 Rank-1(%)准确率和 mAP(%)

方法	s=0		s=0.3		s=0.6	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
XQDA <sup>[3]</sup> , 2015	44.2	—	36.9	—	32.3	—
NPD <sup>[34]</sup> , 2016	53.7	—	39.5	—	33.8	—
IDE <sup>[30]</sup> , 2016	68.2	62.7	65.1	59.8	46.2	43.6
TriNet <sup>[35]</sup> , 2017	79.1	76.4	68.0	66.9	48.1	49.2
Quad <sup>[36]</sup> , 2017	84.3	82.0	75.3	72.4	58.1	56.5
P2S <sup>[37]</sup> , 2017	54.5	51.3	47.3	45.3	37.2	36.5
GLAD <sup>[6]</sup> , 2017	82.2	—	55.8	—	48.3	—
PAN <sup>[38]</sup> , 2017	85.4	<b>90.9</b>	61.0	66.5	53.0	57.6
SVDNet <sup>[39]</sup> , 2017	81.2	84.5	71.2	66.8	63.9	62.1
DPFL <sup>[40]</sup> , 2017	82.0	78.1	—	—	—	—
AACN <sup>[5]</sup> , 2018	89.5	—	—	—	—	—
RNLSTM <sub>A</sub> <sup>[22]</sup> , 2019	86.6	83.8	77.3	75.8	60.3	60.1
mGD+R	86.5	82.3	80.5	77.5	75.1	71.7
mGD+RNLSTM <sub>A</sub>	<b>88.0</b>	84.2	<b>83.4</b>	<b>80.1</b>	<b>77.7</b>	<b>74.2</b>

我们可以得到以下的观察结果.

(1) 相比于传统手工特征和度量学习的方法,例如 NPD<sup>[34]</sup>和 XQDA<sup>[3]</sup>,其他的行人重识别方法都是基于深度学习相关的方法,这类方法的识别性能有大幅度的提升,这也证明了深度学习在特征表示学习方面的优越性.

(2) 相比于基准的分类识别方法 R,我们的方法 mGD+R,在 3 个数据集的原始图上的识别结果都有不同程度的提高,这证明了我们提出的方法 mGD 有效地增强了数据的多样性,从而提高行人重识别的性能.

(3) 相比于现有的基于深度学习的方法,比如基于局部特征学习的方法(例如 GLAD<sup>[6]</sup>,PAN<sup>[38]</sup>和 AACN<sup>[5]</sup>),我们提出的多尺度生成对抗网络结合简单的分类识别网络(mGD+R),就可以达到与现有的方法相近的结果.并且与现有的行人重识别方法,比如与 RNLSTM<sub>A</sub><sup>[22]</sup>结合就可以取得最好的结果.实验结果表明 mGD 对于最终的识别有明显的促进作用.主要原因就是本文的方法不仅可以有效增强训练样本,而且去除遮挡带来的干扰.

(4) 值得注意的是,相比于利用随机遮挡样本进行数据增强的方法 RandEra<sup>[18]</sup>,本文提出的方法利用多尺度 GAN 进行去遮挡预处理,在面对遮挡场景时候的识别性能可以保持比较满意的识别结果.此外,本文提出的方法 mGD+R,在非遮挡场景下(s=0)的识别结果与 RNLSTM<sub>A</sub><sup>[22]</sup>方法相近.但是需要指出的是,mGD+R 利用数据增强的方式,只使用了简单的分类识别网络,而不需要设计特定的网络用于鲁棒特征的学习.同时,在不同比例的遮挡场景下的识别结果远优于 RNLSTM<sub>A</sub>方法.这主要是由于 mGD 利用多尺度生成对抗网络进行去遮挡减少了遮挡物体的干扰,从而提高了遮挡行人重识别的识别效果.通过结合两种方法,即 mGD+RNLSTM<sub>A</sub>,可以在非遮挡场景下与遮挡场景下均达到最好的识别效果,这主要是结合了去遮挡和鲁棒特征学习的优势.

## 4 结论

本文详细介绍了一种新型的遮挡行人重识别方法,并设计了一个简单的多尺度生成对抗网络,其不仅可以用于生成多样的训练数据起到数据增强的作用;并且可以通过恢复遮挡区域的图像信息,进而提高遮挡场景下的行人重识别精度.本文提出的多尺度生成对抗网络可以直接与现有的行人重识别方法结合使用,充分利用去遮挡和鲁棒特征学习的优势.在多个有挑战性的行人重识别数据集上,广泛的实验设计与分析验证了我们提出方法的有效性.目前本文采用人工的方式选取遮挡区域.在未来工作中,可以利用注意力机制对遮挡的区域进行定位,并判断遮挡的严重程度,从而达到对遮挡区域干扰信息的去除以及有用信息的保留.同时,还将研究如何有效处理行人重识别剧烈光照变化带来的影响.

**References:**

- [1] Yi D, Lei Z, Liao S, Li SZ. Deep metric learning for person re-identification. In: Proc. of the IEEE Int'l Conf. on Pattern Recognition (ICPR). 2014. 34–39. [doi: 10.1109/ICPR.2014.16]
- [2] Song W, Zhao Q, Chen C, Gan Z, Liu F. Survey on pedestrian re-identification research. CAAI Trans. on Intelligent Systems, 2017,12(6):770–780 (in Chinese with English abstract). <http://tis.hrbeu.edu.cn/oa/darticle.aspx?type=view&id=201706084> [doi: 10.11992/tis.201706084]
- [3] Liao S, Hu Y, Zhu X, Li SZ. Person re-identification by local maximal occurrence representation and metric learning. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2015. 2197–2206. [doi: 10.1109/CVPR.2015.7298832]
- [4] Zhao H, Tian M, Sun S, Shao J, Yan J, Yi X, Wang X, Tang X. Spindle net: person re-identification with human body region guided feature decomposition and fusion. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2017. 1077–1085. [doi: 10.1109/CVPR.2017.103]
- [5] Xu J, Zhao R, Zhu F, Wang H, Ouyang W. Attention-aware compositional network for person re-identification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2018. 2119–2128. [doi: 10.1109/CVPR.2018.00226]
- [6] Wei L, Zhang S, Yao H, Gao W, Tian Q. Glad: Global-local-alignment descriptor for pedestrian retrieval. In: Proc. of the 25th ACM Int'l Conf. on Multimedia (ACM MM). 2017. 420–428. [doi: 10.1145/3123226.3123279]
- [7] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. In: Proc. of the Advances in Neural Information Processing Systems (NIPS). 2014. 2672–2680. [doi: 10.1007/978-1-4842-3679-6\_8]
- [8] Tang X, Du Y, Liu Y, Li J, Ma Y. Image recognition with conditional deep convolutional generative adversarial networks. Acta Automatica Sinica, 2018,44(5):855–864 (in Chinese with English abstract). <http://www.aas.net.cn/CN/Y2018/V44/I5/855> [doi: 10.16383/j.aas.2018.c170470]
- [9] Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks. In: Proc. of the Int'l Conf. on Machine Learning (ICML). 2017. 214–223.
- [10] Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville AC. Improved training of Wasserstein gans. In: Proc. of the Advances in Neural Information Processing Systems (NIPS). 2017. 5767–5777.
- [11] Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proc. of the IEEE Int'l Conf. on Computer Vision (ICCV). 2017. 223–2232. [doi: 10.1109/ICCV.2017.244]
- [12] Choi Y, Choi M, Kim M, Ha JW, Kim S, Choo J. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2018. 8789–8797. [doi: 10.1109/CVPR.2018.00916]
- [13] Li D, Chen X, Zhang Z, Huang K. Learning deep context-aware features over body and latent parts for person re-identification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2017. 384–393. [doi: 10.1109/CVPR.2017.782]
- [14] Ahmed E, Jones M, Marks TK. An improved deep learning architecture for person re-identification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2015. 3908–3916. [doi: 10.1109/CVPR.2015.7299016]
- [15] Varior RR, Haloi M, Wang G. Gated siamese convolutional neural network architecture for human re-identification. In: Proc. of the European Conf. on Computer Vision (ECCV). 2016. 791–808. [doi: 10.1007/978-3-319-46484-8\_48]
- [16] Shi H, Yang Y, Zhu X, Liao S, Lei Z, Zheng W, Li SZ. Embedding deep metric for person re-identification: A study against large variations. In: Proc. of the European Conf. on Computer Vision (ECCV). 2016. 732–748. [doi: 10.1007/978-3-319-46448-0\_44]
- [17] Cheng D, Gong Y, Zhou S, Wang J, Zheng N. Person re-identification by multi-channel parts-based cnn with improved triplet loss function. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2016. 1335–1344. [doi: 10.1109/CVPR.2016.149]
- [18] Zhong Z, Zheng L, Kang G, Li S, Yang Y. Random erasing data augmentation. arXiv Preprint arXiv:1708.04896, 2017.
- [19] Huang H, Li D, Zhang Z, Chen X, Huang K. Adversarially occluded samples for person re-identification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2018. 5098–5107. [doi: 10.1109/CVPR.2018.00535]
- [20] Zhuo J, Chen Z, Lai J, Wang G. Occluded person re-identification. In: Proc. of the IEEE Int'l Conf. on Multimedia and Expo (ICME). 2018. 1–6. [doi: 10.1109/ICME.2018.8486568]

- [21] Su C, Li J, Zhang S, Xing J, Gao W, Tian Q. Pose-driven deep convolutional model for person re-identification. In: Proc. of the IEEE Int'l Conf. on Computer Vision (ICCV). 2017. 3960–3969. [doi: 10.1109/ICCV.2017.427]
- [22] Yang W, Yan Y, Chen S. Adaptive deep metric embeddings for person re-identification under occlusions. *Neurocomputing*, 2019,340:125–132. [doi: 10.1016/j.neucom.2019.02.042]
- [23] Hou R, Ma B, Chang H, Gu X, Shan S, Chen X. VRSTC: Occlusion-free video person re-identification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2019. 7183–7192.
- [24] Zhong Z, Zheng L, Zheng Z, Li S, Yang Y. Camera style adaptation for person reidentification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2018. 5157–5166. [doi: 10.1109/CVPR.2018.00541]
- [25] Liu J, Ni B, Yan Y, Zhou P, Cheng S, Hu J. Pose transferrable person reidentification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2018. 4099–4108. [doi: 10.1109/CVPR.2018.00431]
- [26] Qian X, Fu Y, Xiang T, Wang W, Qiu J, Wu Y, Jiang Y, Xue X. Pose-normalized image generation for person re-identification. In: Proc. of the European Conf. on Computer Vision (ECCV). 2018. 650–667. [doi: 10.1007/978-3-030-01240-3\_40]
- [27] Deng W, Zheng L, Ye Q, Kang G, Yang Y, Jiao J. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2018. 994–1003. [doi: 10.1109/CVPR.2018.00110]
- [28] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2016. 770–778. [doi: 10.1109/CVPR.2016.90]
- [29] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: Proc. of the Int'l Conf. on Learning Representations (ICLR). 2015.
- [30] Zheng L, Yang Y, Hauptmann AG. Person re-identification: Past, present and future. arXiv Preprint arXiv:1610.02984, 2016.
- [31] Ristani E, Solera F, Zou R, Cucchiara R, Tomasi C. Performance measures and a data set for multi-target, multi-camera tracking. In: Proc. of the European Conf. on Computer Vision (ECCV). 2016. 17–35. [doi: 10.1007/978-3-319-48881-3\_2]
- [32] Li W, Zhao R, Xiao T, Wang X. Deepreid: Deep filter pairing neural network for person re-identification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2014. 152–159. [doi: 10.1109/CVPR\_2014.27]
- [33] Zheng W, Li X, Xiang T, Liao S, Lai J, Gong S. Partial person re-identification. In: Proc. of the IEEE Int'l Conf. on Computer Vision (ICCV). 2015. 4678–4686. [doi: 10.1109/ICCV.2015.531]
- [34] Zhang L, Xiang T, Gong S. Learning a discriminative null space for person reidentification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2016. 1239–1248. [doi: 10.1109/CVPR.2016.139]
- [35] Hermans A, Beyer L, Leibe B. In defense of the triplet loss for person reidentification. arXiv Preprint arXiv:1703.07737, 2017.
- [36] Chen W, Chen X, Zhang J, Huang K. Beyond triplet loss: A deep quadruplet network for person re-identification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2017. 403–412. [doi: 10.1109/CVPR.2017.145]
- [37] Zhou S, Wang J, Wang J, Gong Y, Zheng N. Point to set similarity based deep feature learning for person re-identification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2017. 3741–3750. [doi: 10.1109/CVPR.2017.534]
- [38] Zhao L, Li X, Zhuang Y, Wang J. Deeply-learned part-aligned representations for person re-identification. In: Proc. of the IEEE Int'l Conf. on Computer Vision (ICCV). 2017. 3219–3228. [doi: 10.1109/ICCV.2017.349]
- [39] Sun Y, Zheng L, Deng W, Wang S. Svdnet for pedestrian retrieval. In: Proc. of the IEEE Int'l Conf. on Computer Vision (ICCV). 2017. 3800–3808. [doi: 10.1109/ICCV.2017.410]
- [40] Chen Y, Zhu X, Gong S. Person re-identification by deep learning multi-scale representations. In: Proc. of the IEEE Int'l Conf. on Computer Vision (ICCV). 2017. 2590–2600. [doi: 10.1109/ICCVW.2017.304]

#### 附中文参考文献:

- [2] 宋婉茹,赵晴晴,陈昌红,干宗良,刘峰.行人重识别研究综述.智能系统学报,2017,12(6):770–780. <http://tis.hrbeu.edu.cn/oa/darticle.aspx?type=view&id=201706084> [doi:10.11992/tis.201706084]
- [8] 唐贤伦,杜一铭,刘雨微,李佳歆,马艺玮.基于条件深度卷积生成对抗网络的图像识别方法.自动化学报,2018,44(5):855–864. <http://www.aas.net.cn/CN/Y2018/V44/I5/855> [doi: 10.16383/j.aas.2018.c170470]



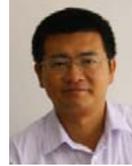
杨婉香(1993-),女,硕士,主要研究领域为行人重识别.



张小康(1996-),男,学士,主要研究领域为行人重识别.



严严(1984-),男,博士,副教授,CCF 专业会员,主要研究领域为计算机视觉,模式识别.



王菡子(1973-),男,博士,教授,博士生导师,CCF 专业会员,主要研究领域为计算机视觉,模式识别.



陈思(1987-),女,博士,副教授,CCF 专业会员,主要研究领域为计算机视觉,模式识别.

www.jos.org.cn

www.jos.org.cn