自动化学报
ACTA AUTOMATICA SINICA

基于多对多生成对抗网络的非对称跨域迁移行人再识别

梁文琦 王广聪 赖剑煌

Asymmetric Cross-domain Transfer Learning of Person Re-identification Based on the Many-to-many Generative Adversarial Network

LIANG Wen-Qi, WANG Guang-Cong, LAI Jian-Huang

在线阅读 View online: https://doi.org/10.16383/j.aas.c190303

基于多对多生成对抗网络的非对称跨域迁移行人再识别

梁文琦 [1] 王广聪 [1] 赖剑煌 [1, 2, 3, 4]

摘 要 无监督跨域迁移学习是行人再识别中一个非常重要的任务. 给定一个有标注的源域和一个没有标注的目标域, 无监督跨域迁移的关键点在于尽可能地把源域的知识迁移到目标域. 然而, 目前的跨域迁移方法忽略了域内各视角分布的差异性, 导致迁移效果不好. 针对这个缺陷, 本文提出了一个基于多视角的非对称跨域迁移学习的新问题. 为了实现这种非对称跨域迁移, 提出了一种基于多对多生成对抗网络 (Many-to-many generative adversarial network, M2M-GAN) 的迁移方法. 该方法嵌入了指定的源域视角标记和目标域视角标记作为引导信息, 并增加了视角分类器用于鉴别不同的视角分布, 从而使模型能自动针对不同的源域视角和目标域视角组合采取不同的迁移方式. 在行人再识别基准数据集 Market1501、DukeMTMC-reID 和 MSMT17 上, 实验验证了本文的方法能有效提升迁移效果, 达到更高的无监督跨域行人再识别准确率.

关键词 行人再识别, 多对多跨域迁移, 非监督迁移学习, 生成对抗网络

引用格式 梁文琦, 王广聪, 赖剑煌. 基于多对多生成对抗网络的非对称跨域迁移行人再识别. 自动化学报, 2022, **48**(1): 103−120

DOI 10.16383/j.aas.c190303

开放科学(资源服务)标识码(OSID):

Asymmetric Cross-domain Transfer Learning of Person Re-identification Based on the Many-to-many Generative Adversarial Network

LIANG Wen-Qi[1] WANG Guang-Cong[1] LAI Jian-Huang[1, 2, 3, 4]

Abstract Unsupervised cross-domain transfer learning is an extremely important task in person re-identification (ReID). Given a labeled source domain and an unlabeled target domain, the key to the unsupervised cross-domain transfer learning is to transfer the knowledge from the source domain to the target domain as much as possible. However, current cross-domain transfer learning methods cannot obtain desired performance because they ignore the distribution differences between different views within domains. Therefore, we propose a new problem of view-based asymmetric cross-domain transfer learning for ReID. To address this problem, we propose a novel transfer learning method based on the many-to-many generative adversarial network (M2M-GAN). The M2M-GAN embeds source view labels and target view labels as the guide information, and adds view classifiers to identify different view distributions, so that the model can automatically adopt different transferring ways according to different source views or target views. Experiments on three ReID benchmark datasets Market1501, DukeMTMC-reID and MSMT17 verify that the proposed method can improve the performance of transfer learning and achieve higher recognition rate of unsupervised cross-domain ReID.

Key words Person re-identification, many-to-many cross-domain transfer learning, unsupervised transfer learning, generative adversarial network

Citation Liang Wen-Qi, Wang Guang-Cong, Lai Jian-Huang. Asymmetric cross-domain transfer learning of person re-identification based on the many-to-many generative adversarial network. *Acta Automatica Sinica*, 2022, **48**(1): 103−120

收稿日期 2019-04-16 录用日期 2019-09-02
Manuscript received April 16, 2019; accepted September 2, 2019
国家自然科学基金 (61573387, 62076258),广东省重点研发项目 (2017B030306018), 广东省海洋经济发展项目 (粤自然资合 [2021] 34) 资助
Supported by National Natural Science Foundation of China (61573387, 62076258), Key Research Projects in Guangdong Province (2017B030306018), and Contract of Department of Natural Resources of Guangdong Province ([2021] 34)
本文责任编委 刘青山
Recommended by Associate Editor LIU Qing-Shan

1. 中山大学计算机学院 广州 510006 2. 广州新华学院 广州 510520 3. 广东省信息安全技术重点实验室 广州 510006 4. 机器智能与先进计算教育部重点实验室 广州 510006

1. School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510006 2. Guangzhou Xinhua University, Guangzhou 510520 3. Guangdong Province Key Laboratory of Computational Science, Guangzhou 510006 4. Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, Guangzhou 510006

行人再识别[1−7] 是指在非重叠的摄像头视角下检索特定的目标行人图片或视频片段, 它是多摄像机跟踪、搜索取证等重要应用中的关键技术, 广泛应用于智能视频监控网络中[8]. 行人再识别最初的

研究方法是先设计一种能够描述行人图片的手工视觉特征, 再建立一个鲁棒的距离度量模型来度量视觉特征之间的相似性[9−15]. 近年来, 随着深度学习的发展, 大部分研究者转向使用深度学习来处理行人再识别问题. 文献 [16−18] 分别提出了基于分类损失、验证损失、三元组损失的行人再识别基本框架. 为了处理行人图像不对齐的问题, 文献 [19−20] 分别提出全局区域和局部区域的对齐方法, 文献 [21] 提出动态的特征对齐方法. 为了处理摄像头之间的差异, 文献 [22] 提出使用多组生成对抗网络在同域内的多个视角之间进行迁移, 以此缩小域内不同视角之间的差别. 为了进一步提高识别准确率, 最近有很多文献尝试使用额外的标注信息作为辅助. 例如文献 [23] 提出人体姿势驱动的深度卷积模型, 文献 [24] 引入行人属性标记, 文献 [25−26] 加入了人体掩模, 文献 [27] 提出在检索过程中加入时空约束.

得益于深度学习的发展, 如今行人再识别任务在大规模数据集上已经取得了良好的效果, 但需要大量带标注的训练数据. 然而, 与其他检索任务不同, 收集带标注的行人再识别训练数据更加困难. 标注数据的难点在于, 行人再识别数据集没有固定的类别, 多人合作标注很困难; 而且图像分辨率低, 不容易辨别. 为了更符合实际场景的应用需求, 科研人员开始研究如何在目标数据集没有标注信息的前提下实现行人再识别. 在这种背景下, 非监督行人再识别 (Unsupervised person re-identification) 成为新的研究热点.

目前, 非监督行人再识别有两类主要的研究方法. 第 1 类是基于聚类的非监督学习方法. 文献 [28] 提出一种基于聚类的非对称度量学习方法, 利用非对称聚类学习把不同视角的数据投影到共享空间中. 文献 [29] 提出基于聚类和微调的非监督深度学习框架. 该方法先使用预训练的神经网络模型提取目标数据集的特征, 然后通过聚类算法得到目标数据集的伪标签, 再利用伪标签对预训练的网络进行微调 (Fine-tune). 文献 [30] 在文献 [29] 的框架上再进行改进, 提出一种自底向上逐层合并最相近簇的聚类方法.

非监督行人再识别的第 2 类研究方法是跨域迁移学习方法 (Cross-domain transfer learning). 这类方法通常都有带标注的行人再识别数据集作为辅助, 这个辅助的数据集称为源数据集或源域 (Source domain), 实际应用场景对应的无标注数据集称为目标数据集或目标域 (Target domain). 由于只有源域是有标注的, 所以这类方法的关键之处在于尽可能地把从源域中学习到的知识迁移到目标域中. 文献 [31] 通过添加域分类器和梯度反传网络层来实现域适应. 文献 [32] 提出一种跨域自适应的 Ranking SVM (Support vector machine) 方法, 利用了源域的正负样本、目标域的负样本和目标域估计的正样本均值来训练. 文献 [33−34] 则提出两阶段的跨域迁移学习方法: 首先利用生成对抗网络实现源域数据分布到目标域数据分布的变换, 根据变换前源域数据的标签对变换后的图片进行标注; 然后使用变换后的图片及其对应的标注进行有监督训练.

跨域迁移学习方法对目标域训练集 (无标注数据) 的数据分布限制更少, 应用范围更广泛, 更加适合实际的行人再识别应用场景. 但是现有的跨域迁移学习方法没有考虑视角偏差 (View-specific bias) 问题, 源域中不同视角 (摄像机) 的数据以完全相同的迁移方式变换到目标域中. 也就是说, 目前的迁移方式都是对称的 (对称迁移). 然而在智能监控视频网络中, 不同拍摄地点的光照条件、拍摄角度以及摄像机本身的参数都可能存在明显的差别, 不同摄像头拍摄到的图片往往服从不同的分布. 在跨域迁移学习时, 忽略摄像头的分布差异一方面会导致迁移效果不佳, 另一方面会导致迁移后的数据无法体现出目标域多个视角子分布的情况, 从而不利于训练跨视角匹配模型.

基于以上分析, 本文提出基于多视角 (摄像机) 的非对称跨域迁移学习方法. 在基于生成对抗网络的两阶段跨域迁移学习方法[33−34] 基础上, 本文针对视角之间的差异问题进行建模. 为了对每种源域–目标域视角组合使用不同的迁移方式 (称为非对称迁移), 一个最简单直观的想法是把每个视角的数据看成是各自独立的, 然后训练多组互不相干的生成对抗网络模型, 每个模型分别把知识从源域的某个视角迁移到目标域的某个视角. 然而, 这种不同视角组合使用不同网络参数的非对称迁移方式非常消耗训练时间和存储空间. 假如源域有 M 个视角, 目标域有 N 个视角, 则一共需要训练 $M \times N$ 组生成对抗网络. 大型智能监控网络涉及的摄像头数目非常多, 显然这种方法是不切实际的. 除此之外, 单独使用每对视角的数据来训练生成对抗网络无法利用数据集内不同视角数据之间的相关性. 为了解决独立训练而造成成本太高的问题, 并尽可能地利用不同视角数据的相关性, 本文提出把非对称迁移学习嵌入到一组生成对抗网络中. 为此, 我们设计了一个多对多生成对抗网络 (Many-to-many generative adversarial network, M2M-GAN), 同时实现源域任意视角子分布到目标域任意视角子分布的转换. 实验表明, 与现有的对称迁移方法 (不考虑视角

差异, 且仅有一组生成对抗网络网络) 相比, 我们的方法只需增加少量训练时间和空间成本就能有效提升识别准确率. 与单独训练多组生成对抗网络这种简单的建模方式 (考虑视角差异, 但需 $M \times N$ 组生成对抗网络) 相比, 我们的方法在训练成本和识别准确率两方面都取得更优的性能.

本文的主要贡献: 1) 针对源域或者目标域存在多个具有差异性的子分布问题, 本文提出一种多对多的跨域迁移模型来区别对待源域不同的子分布到目标域不同的子分布的迁移. 本文将这种区分性的迁移模式称为非对称迁移. 为了更好地优化非对称迁移学习模型, 本文提出了一种基于多对多生成对抗网络 (M2M-GAN) 的迁移学习方法, 同时实现把源域任意子分布的图像风格转变成目标域任意子分布的图像风格. 2) 视角偏差或摄像机差异是跨域迁移行人再识别领域被忽略的一个关键问题. 本文将 M2M-GAN 方法应用于该领域, 生成了具有视角差异且服从目标域各个视角子分布的行人图片, 进而使得模型学习到的特征具有视角偏差不变性, 有效提升了无监督跨域迁移行人再识别的准确率. 3) 在 Market-1501, DukeMTMC-reID 和 MSMT17 三个大规模多摄像头行人再识别基准数据集上, 实验结果验证了 M2M-GAN 的有效性.

1 生成对抗网络和跨域迁移学习行人再识别

1.1 循环生成对抗网络

生成对抗网络 (Generative adversarial network, GAN)[35] 主要用于图像风格转换, 例如把照片变成油画、把马变成斑马等. 它由生成器和鉴别器两部分组成, 利用博弈论观点来训练. 生成器试图生成能够以假乱真的图片来 “欺骗”鉴别器, 而鉴别器则尽可能地把生成器生成的 “假冒”图片鉴别出来. 生成器和鉴别器相互对抗, 交替训练, 直到鉴别器无法判断生成器生成图片的真假, 这个过程可以用对抗损失表示.

用于图片风格转换的生成对抗网络一般需要成对的训练样本, 但实际应用中很难收集到足够的成对样本. 循环生成对抗网络 (Cycle-GAN)[36] 通过在普通生成对抗网络的基础上添加循环一致约束来实现非配对图像之间的转换. 假设需要在两个域 X 和 Y 之间转换, 循环一致约束要求对于域 X 的一张真实的图片 x, 通过生成器 G 生成图片 $G(x)$, $G(x)$ 再经过生成器 $\bar{G}$ 生成重构图片 $\bar{G}(G(x))$, 这个重构图片需要和原始的真实图片 x 保持像素级别的一致. 同理, 对于域 Y 也是如此.

1.2 基于循环生成对抗网络的跨域迁移行人再识别

跨域迁移学习方法是当前在目标数据集无标注的情况下解决行人再识别最常用的方法. 除了没有标注信息的目标数据集外, 这类方法还会使用其他场景下有标注的行人再识别数据集作为辅助. 跨域迁移学习方法的关键之处在于尽可能地把有标注的辅助数据集 (即源域 (Source domain)) 的知识迁移到无标注的目标数据集 (即目标域 (Target domain)). 跨域行人再识别方法的难点在于不同的行人再识别数据集之间存在较大的差异 (Dataset bias), 因而极大增加了从源域到目标域知识迁移的难度. 这些数据集差异包括图像背景、光照、拍摄角度、图像分辨率等. 如果能减小这种数据集之间的差异, 那么在有标注数据集上训练的模型就可以适用于无标注的目标数据集. 而循环生成对抗网络恰好适合用于减小数据集之间的差异. 如第 1.1 节所述, 循环生成对抗网络可以把一种风格的图片变成另一种风格的图片, 并且不需要使用两个域中配对的数据来训练. 利用循环生成对抗网络, 可以把源域的图片风格转变成目标域的图片风格. 因此, 基于循环生成对抗网络的跨域迁移行人再识别可分为两阶段: 第 1 阶段是训练循环生成对抗网络, 用源域图片生成具有目标域图片风格的新数据集, 新数据集使用源域的身份标注; 第 2 阶段是利用新数据集及对应的身份标注进行有监督的行人再识别.

2 基于多对多生成对抗网络的非对称跨域迁移行人再识别

目前跨域迁移行人再识别的课题研究忽略了域内不同视角的数据分布差异问题, 对于所有的视角数据采用完全相同的迁移方式. 我们重点研究了多个视角子分布的差异在迁移学习中的重要性, 并提出了非对称跨域迁移行人再识别问题. 为了能高效地实现非对称迁移, 我们设计了多对多生成对抗网络 (M2M-GAN). 本节将详细介绍我们提出的方法.

2.1 非对称跨域迁移行人再识别问题

在智能视频监控网络中, 不同摄像头被布置在光照条件、拍摄角度不同的位置, 再加上摄像头自身参数的差异, 所以实际应用中不同摄像头拍摄得到的行人图片数据会呈现不同的分布情况. 图 1 列出了 4 个常见的行人再识别数据集的例子, 其中每张子图的不同列代表该数据集内不同摄像头拍摄得到的行人图片. 从图 1 可以看出, 摄像头分布差异在行人再识别任务中是普遍存在的.

然而, 现有的行人跨域迁移框架忽略了这种摄像头分布差异性, 会造成精度的损失. 我们用图 2 对此进行直观的解释. 图 2(a) 和 2(b) 两幅图描述了源数据集和目标数据集整体上存在较大差异, 每个数据集内部也存在一定的子分布差异的现象. 每幅图内不同的曲线分别代表了数据集内某个特定视角的分布. 不同的视角子分布都近似于整体分布, 具有相似性, 但是各自又存在一定的偏差. 图 2(c) 和 2(d) 分别是图 2(a) 和 2(b) 中各个数据集内所有视角子分布的平均值, 图 2(c) 和 2(d) 两幅图只描述了源数据集和目标数据集差异. 现有的迁移框架主要研究如何减小数据集差异, 在迁移过程中把源域和目标域都分别看作一个整体, 即从图 2(c) 到 2(d) 的迁移. 但是, 使用平均分布来估计具有多子分布的真实数据集, 即用图 2(c) 和 2(d) 来估计图 2(a) 和 2(b), 会带来精度的损失. 所以我们重点研究迁移过程中的视角差异问题, 提出用源数据集多视角到目标数据集多视角迁移 (即 (a) → (b)) 代替传统的整体迁移 (即 (c) → (d)) 方式.

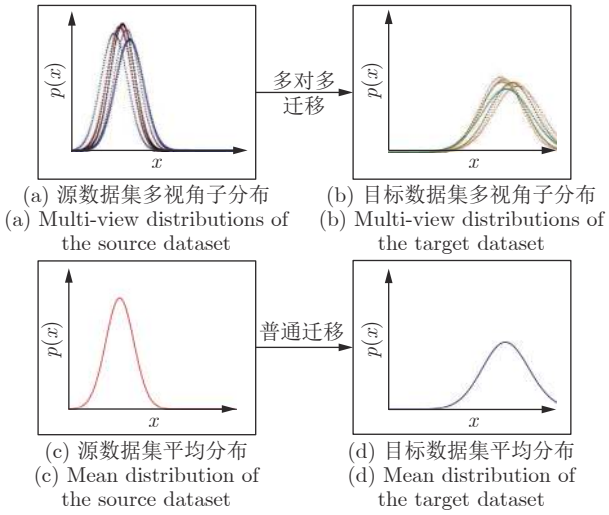
我们从两方面详细地分析多对多迁移的好处. 一方面, 在迁移过程中加入视角信息可以提高生成图片质量. 假设生成器 G 用于从源域到目标域迁移, 鉴别器 D 用于鉴别输入图片是否属于目标域. 生成器 G 的输入是具有多个子分布的图片 (图 2(a)), 假如忽略视角差异, 那么生成器就需要针对所有视角子分布采用相同的生成方式. 但是同一种映射方式很难同时适用于多种视角分布. 而如果在生成器中引入视角信息, 使生成器针对不同的子分布选择不同的生成方式, 就可以提高生成图片的质量. 同样地, 鉴别器的输入也是具有多个视角子分布的真实

(a) PRID (b) iLIDS (c) DukeMTMC-reID

(d) MSMT17

图 1 摄像机分布差异举例

Fig. 1 Examples of distribution differences between different views



(a) 源数据集多视角子分布 (a) Multi-view distributions of the source dataset
(b) 目标数据集多视角子分布 (b) Multi-view distributions of the target dataset
(c) 源数据集平均分布 (c) Mean distribution of the source dataset
(d) 目标数据集平均分布 (d) Mean distribution of the target dataset

图 2　本文提出的多视角对多视角迁移方式与现有迁移方式的比较

Fig. 2　Comparison of our M2M transferring way and the existing methods

图片或生成图片. 如果在鉴别器中融入视角信息, 使鉴别器针对不同视角子分布采取不同的鉴别方式, 就可以提高鉴别器的鉴别能力, 从而间接提高生成器的效果.

另一方面, 强调视角子分布可以使生成的数据更好地模拟出目标域多视角分布的现实情况, 更有利于解决目标域跨视角匹配问题. 假如在迁移时不区分视角, 那么迁移的结果只包含了目标域的总体统计特性 (图 2(d)), 没有包含目标域各个视角的统计特性 (图 2(b)). 而目标域行人再识别的一个关键问题是要实现跨视角行人图片匹配, 即要训练一个对视角鲁棒的行人再识别模型. 为了训练对视角鲁棒的行人再识别模型, 就需要具有不同视角分布风格的训练图片. 所以, 如果能生成服从目标域不同视角子分布的图片 (图 2(b)) 作为训练数据, 而不仅仅生成服从目标域平均分布 (图 2(d)) 的图片, 将会大大提升模型对视角的鲁棒性.

以上分析体现了迁移过程中结合视角信息的重要性. 因此, 我们提出了非对称跨域迁移行人再识别问题, 强调针对源域的不同视角或目标域的不同视角采取不同的迁移方式.

具体地, 非对称跨域迁移行人再识别问题可以描述为: 令 S 表示一个有标注的源域, T 表示一个没有标注的目标域. S 包含 M 个视角, 记为 $S_1, S_2, \cdots, S_i, \cdots, S_M$. T 包含 N 个视角, 记为 $T_1, T_2, \cdots, T_j, \cdots, T_N$. 每张图片来自哪个摄像头是容易收集的标注信息, 所以每张图片的视角标记 S_i 或 T_j 可视为已知信息. 非对称跨域迁移行人再识别的目标是把源域的多个子域分别迁移到目标域的多个子域. 即对于任意的 $i \in [1, M]$, $j \in [1, N]$, 要实现从 S_i 到 T_j 的迁移. 给定源域视角 S_i 的一张真实图片 x_{s_i} 和相应的身份标注 y_{s_i}, 利用 x_{s_i} 生成具有目标域视角 T_j 风格的图片 $x^*_{t_j}$. 然后, 使用生成图片 $x^*_{t_j}$ 和身份标注 y_{s_i} 训练行人再识别模型.

2.2　多对多生成对抗网络

由于同一个行人不会同时出现在源域和目标域中, 我们无法获取配对的训练样本用于源域–目标域行人图片风格迁移. 而循环生成对抗网络恰好不需要成对训练样本, 所以目前的研究方法通常采用循环生成对抗网络来实现源域–目标域行人图片风格迁移. 但是直接应用循环生成对抗网络, 只能实现源域整体与目标域整体风格之间的迁移.

想要实现源域多视角与目标域多视角之间的非对称迁移, 一种简单的方案是针对每对源域–目标域视角组合 (S_i, T_j) 分别训练一组循环生成对抗网络, 共需 $M \times N$ 组循环生成对抗网络用于实现源域 M 个视角和目标域 N 个视角之间的迁移. 显然这种方法是不切实际的, 因为在一个大型智能视频监控网络中可能存在成百上千个摄像头, 训练 $M \times N$ 组深度网络会带来巨大的训练时间和存储空间损耗. 我们希望只用一组生成对抗网络实现源域多个视角和目标域多个视角图片风格迁移. 因此, 我们在循环生成对抗网络基础上进行改进, 设计了多对多生成对抗网络, 使得模型不仅局限于两个域整体风格的迁移, 还能细化到多个视角的图像风格迁移.

为了实现两个域的不同视角之间的图片风格迁移, 首先我们需要把视角信息输入到模型中. 假如没有子视角信息, 模型就无法针对不同的视角组合采取不同的迁移方式. 为此, 我们设计了视角嵌入模块 (第 2.2.2 节), 把原始输入图片、源域视角标记、目标域视角标记整合在一起, 形成新的 “嵌入图”作为生成器的输入, 为生成器提供视角信息.

视角嵌入模块明确告诉生成器需要从源域的哪一个视角迁移到目标域的哪一个视角, 但只有信息嵌入并不足以引导生成器按照期望的方向生成图片. 除了视角信息输入, 我们还需要为生成器提供监督信号, 引导生成器利用输入的视角信息. 也就是说, 我们需要判断生成图片与期望视角的差别有多大, 这样才能把缩小差距作为训练目标来优化生成器的参数. 为此, 我们设计了视角分类模块 (第 2.2.3 节), 该模块利用一个视角分类器来预测生成图片与期望视角的差距, 然后用预测的结果监督生成器, 引导生成器生成与期望视角差异尽可能小的图片, 如图 3 所示.

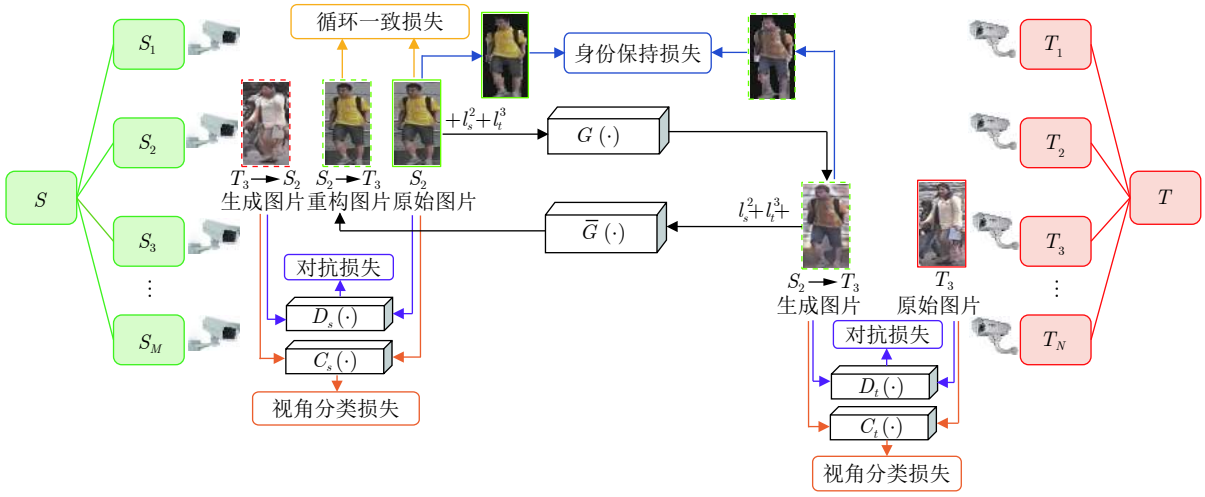


图 3 多对多生成对抗网络框架 (省略了目标域 → 源域的生成过程、循环一致损失和身份保持损失)

Fig. 3 Framework of our M2M-GAN (The generation process, the cycle consistency loss, and the identity preserve loss of target domain → source domain are omitted)

视角嵌入和视角分类两个模块, 前者为生成器提供了视角信息, 后者监督生成器充分地利用输入的视角信息. 这两个模块配合使用, 就可以在同一个网络中实现两个域多个视角组合的图片风格迁移.

下面将详细介绍多对多生成对抗网络, 先回顾循环生成对抗网络 (第 2.2.1 节), 然后介绍视角嵌入模块 (第 2.2.2 节) 和视角分类模块 (第 2.2.3 节), 最后给出总目标函数 (第 2.2.4 节) 并说明模型结构设计和训练的细节 (第 2.2.5 节).

2.2.1 循环生成对抗网络

多对多生成对抗网络在循环生成对抗网络基础上进行改进, 本节简要描述循环生成对抗网络算法. 循环生成对抗网络[36] 由两个生成器和两个鉴别器组成. 其中两个生成器 (G 和 $\bar{G}$) 分别用于源域迁移到目标域和目标域迁移到源域的图片生成, 两个鉴别器 (D_t 和 D_s) 分别用于判断生成图片是否属于源域和目标域的真实分布. 循环生成对抗网络包含对抗损失和循环一致损失. 为了统一符号, 我们把对抗损失和循环一致损失改写成多视角形式, 与原本的循环生成对抗网络略有不同.

1) 对抗损失. 源域视角 S_i 迁移到目标域视角 T_j 的对抗损失可以表示为

$$L_{\text{GAN}}(G, D_t, S_i, T_j) = \mathrm{E}_{x_{t_j}}\left[\log D_t(x_{t_j})\right] + \mathrm{E}_{x_{s_i}}\left[\log\left(1 - D_t\left(G(x_{s_i}, l_s^i, l_t^j)\right)\right)\right] \tag{1}$$

其中, x_{s_i} 表示源域视角 S_i 的真实图片, x_{t_j} 表示目标域视角 T_j 的真实图片, l_s^i 表示源域视角标记, l_t^j 为期望的目标域视角标记, $G(x_{s_i}, l_s^i, l_t^j)$ 表示源域视角 S_i 迁移到目标域视角 T_j 的生成图片, $D_t(x_{t_j})$ 和 $D_t(G(x_{s_i}, l_s^i, l_t^j)$ 分别表示真实图片和生成图片属于目标域 T 的概率. 生成器要最小化对抗损失, 鉴别器要最大化对抗损失.

源域 S 迁移到目标域 T 的对抗损失是所有源域–目标域视角组合的对抗损失的平均值

$$L_{\text{GAN}}(G, D_t) = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} L_{\text{GAN}}(G, D_t, S_i, T_j) \tag{2}$$

由式 (1) 和式 (2) 得:

$$L_{\text{GAN}}(G, D_t) = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\mathrm{E}_{x_{t_j}}\left[\log D_t(x_{t_j})\right] + \mathrm{E}_{x_{s_i}}\left[\log\left(1 - D_t\left(G(x_{s_i}, l_s^i, l_t^j)\right)\right)\right]\right) \tag{3}$$

其他损失函数与此类似, 两个域之间的损失等于源域–目标域所有视角组合损失的平均值, 不再对此详细说明.

类似地, 目标域 T 迁移到源域 S 的对抗损失为

$$L_{\text{GAN}}(\bar{G}, D_s) = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\mathrm{E}_{x_{s_i}}\left[\log D_s(x_{s_i})\right] + \mathrm{E}_{x_{t_j}}\left[\log\left(1 - D_s\left(\bar{G}(x_{t_j}, l_t^j, l_s^i)\right)\right)\right]\right) \tag{4}$$

2) 循环一致损失. 循环一致约束在多对多生成对抗网络里的表现形式是: 给定源域视角 S_i 的一张图片 x_{s_i}, 通过生成器 G 生成目标域视角 T_j 的图片 $G(x_{s_i}, l_s^i, l_t^j)$; $G(x_{s_i}, l_s^i, l_t^j)$ 再通过生成器 $\bar{G}$ 重构源域

视角 S_i 的图片 $\bar{G}(G(x_{s_i}, l_s^i, l_t^j), l_t^j, l_s^i)$, 这张重构图片要与原始图片 x_{s_i} 保持一致. 目标域到源域的迁移同理, 即 $x_{s_i} \to G(x_{s_i}, l_s^i, l_t^j) \to \bar{G}(G(x_{s_i}, l_s^i, l_t^j), l_t^j, l_s^i) \approx x_{s_i}$, $x_{t_j} \to \bar{G}(x_{t_j}, l_t^j, l_s^i) \to G(\bar{G}(x_{t_j}, l_t^j, l_s^i), l_s^i, l_t^j) \approx x_{t_j}$. 这两个过程用以下损失函数来约束, 式中 $\|\cdot\|_1$ 表示 L1 范数:

$$L_{\text{cyc}}(G, \bar{G}) = \frac{1}{MN} \times \sum_{i=1}^{M} \sum_{j=1}^{N} \left(\mathrm{E}_{x_{s_i}} \left[\left\| x_{s_i} - \bar{G}\left(G(x_{s_i}, l_s^i, l_t^j), l_t^j, l_s^i \right) \right\|_1 \right] + \mathrm{E}_{x_{t_j}} \left[\left\| x_{t_j} - G\left(\bar{G}(x_{t_j}, l_t^j, l_s^i), l_s^i, l_t^j \right) \right\|_1 \right] \right) \tag{5}$$

2.2.2　源域和目标域视角嵌入

为了让生成器能明确当前迁移方向是从源域 (或目标域) 的哪个视角迁移至目标域 (或源域) 的哪个视角, 生成器的输入应包含视角标记. 我们把视角标记作为额外的输入通道 (Channel), 这样就可以把视角信息输入到生成器神经网络中. 但是图像和视角标签的维度不一致, 无法直接结合, 要对视角标记进行转换, 转换过程如图 4 所示. 我们先用 one-hot 编码方式对视角标记进行编码. 然后对每一位 one-hot 编码值, 如果其值为 1, 就生成一张值全为 1 的二维图; 如果其值为 0, 就生成一张值全为 0 的二维图. 通过这种方式, 视角标记可以转换为与输入图片具有相同图片尺寸的多通道图片. 于是输入图片、输入图片对应的视角标记、期望输出图片对应的视角标记三者就可以依次叠加, 一起输入到生成器中.

以源域迁移到目标域为例, 目标域迁移到源域同理. 对于每张真实图片 x_{s_i}, 我们把图片 x_{s_i}、图片所属的源域视角标记 l_{s_i}、待生成的目标域视角标记 l_{t_j} 整合在一起, 形成嵌入图 $x_{\text{embed}}^{s_i t_j}$:

$$x_{\text{embed}}^{s_i t_j} = \left[x_{\text{rgb}}, B_s^i, B_t^j \right] \tag{6}$$

其中, x_{rgb} 表示大小为 $(3, h, w)$ 的 RGB 图片, B_s^i 表示大小为 (M, h, w) 的二进制张量, B_t^j 表示大小为 (N, h, w) 的二进制张量. M, N 分别是源域和目标域的视角数目. B_s^i 的第 i 通道 (即大小为 (h, w) 的张量) 的值全都设为 1, 表明 x_{rgb} 来自源域的第 i 个视角, 其余 $M-1$ 个通道全都设为 0. B_t^j 的第 j 通道全都设为 1, 表明 x_{rgb} 将要被变换到目标域的第 j 个视角, 其余 $N-1$ 个通道全都设为 0. $[\cdot, \cdot]$ 表示通道串联操作, 把 x_{rgb}、B_s^i、B_t^j 相应的通道依次叠加, 得到大小为 $(3+M+N, h, w)$ 的嵌入图 $x_{\text{embed}}^{s_i t_j}$. 嵌入图 $x_{\text{embed}}^{s_i t_j}$ 被输入到生成器中, 可以引导生成器生成期望的图片 $x_{t_j}^*$. 类似地, 我们可以获得目标域迁移到源域的嵌入图 $x_{\text{embed}}^{t_j s_i}$.

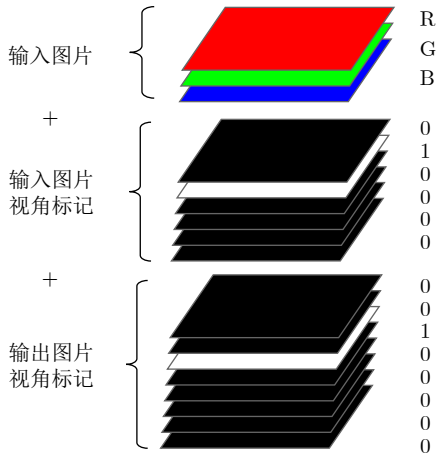


图 4　视角嵌入

Fig. 4　View embedding

2.2.3　视角分类器

通过视角嵌入模块, 视角信息被输入到生成器中. 但是仅仅增加输入信息并不能约束生成器, 还需要为生成器提供监督信号, 引导生成器利用输入的视角信息. 生成器的目标是尽可能缩小生成图片与期望视角的差距, 所以我们可以把生成图片属于期望视角的概率值作为监督信号. 假如生成器生成的图片偏离期望的视角分布, 就需要惩罚生成器.

以源域视角 S_i 迁移到目标域视角 T_j 为例. 把嵌入图 $x_{\text{embed}}^{s_i t_j}$ 输入到生成器中, 得到生成图片 $x_{t_j}^*$. 在对抗损失约束下, 可以认为生成图片 $x_{t_j}^*$ 近似服从目标域整体分布. 但是, $x_{t_j}^*$ 不一定服从目标域 N 个视角中的 T_j 这一特定视角分布. 我们需要计算 $x_{t_j}^*$ 属于目标域视角 T_j 的概率有多大, 并且把 $x_{t_j}^*$ 属于 T_j 的概率作为监督信号对生成器进行监督, 优化生成器参数使得 $x_{t_j}^*$ 属于 T_j 的概率尽可能高. 这样就可以约束生成器生成尽可能服从期望视角分布的图片.

接下来, 我们将叙述如何计算生成图片与期望视角的差距 (视角类别估计), 以及如何利用这一信息监督生成器.

1) 估计视角类别. 估计生成图片属于域内某个视角的概率, 其实就是视角分类任务. 所以, 我们可以训练视角分类器, 然后利用视角分类器来预测生成图片属于域内各个视角的概率.

源域视角分类和目标域视角分类是两个独立的任务, 我们设计了视角分类器 C_s 和 C_t, 分别用于源域 M 个视角和目标域 N 个视角分类. 训练视角

分类需要训练样本和样本对应的视角标记, 我们利用了数据集中真实的图片和真实的视角标记作为训练样本. 训练 C_s 时, 使用源域真实的训练图片和视角标记. 训练 C_t 时, 使用目标域真实的训练图片和视角标记. 值得一提的是, 这里只使用了视角标记, 并不会使用任何身份标记. 为了训练 C_s 和 C_t, 采用图像分类中最常用的交叉熵损失函数

$$L_{\text{view}}^{C}(C_s,C_t)=\frac{1}{M}\sum_{i=1}^{M}\mathrm{E}_{x_{s_i}}\left[-\log\left(C_s(x_{s_i})^{(i)}\right)\right]+\frac{1}{N}\sum_{j=1}^{N}\mathrm{E}_{x_{t_j}}\left[-\log\left(C_t(x_{t_j})^{(j)}\right)\right] \tag{7}$$

其中, 第 1 项是源域的视角分类器损失, 第 2 项是目标域的视角分类器损失. $C_s(x_{s_i})^{(i)}$ 表示源域视角分类器输出的概率向量 $C_s(x_{s_i})$ 的第 i 位, 即 x_{s_i} 正确分类到视角 S_i 的概率值. 类似地, $C_t(x_{t_j})^{(j)}$ 表示 x_{t_j} 正确分类到视角 T_j 的概率值.

2) 监督视角生成. 以视角 S_i 迁移到视角 T_j 为例, 利用视角分类器可以估计生成图片属于目标域 N 个视角中的视角 T_j 的概率. 接下来, 对生成器进行约束, 使得生成图片属于视角 T_j 类别的概率尽量接近 1, 属于其他 $N-1$ 个视角类别的概率尽量接近 0. 这个目标与分类任务类似, 不同点只在于此时需要优化的参数是生成器, 而视角分类器只用来估计生成图片属于各个视角类别的概率, 参数是固定不变的. 因此, 也可以用以下交叉熵损失对生成器进行约束:

$$L_{\text{view}}^{G}(G,\bar{G})=\frac{1}{M}\times\sum_{i=1}^{M}\mathrm{E}_{x_{s_i}}\left[-\log\left(C_t\left(G(x_{s_i},l_s^i,l_t^j)\right)^{(j)}\right)\right]+\frac{1}{N}\sum_{j=1}^{N}\mathrm{E}_{x_{t_j}}\left[-\log\left(C_s\left(\bar{G}(x_{t_j},l_t^j,l_s^i)\right)^{(i)}\right)\right] \tag{8}$$

其中, $C_t(G(x_{s_i},l_s^i,l_t^j))$ 表示目标域视角分类器 C_t 对生成图片 $G(x_{s_i},l_s^i,l_t^j)$ 预测的概率向量, $C_t(G(x_{s_i},l_s^i,l_t^j))^{(j)}$ 表示概率向量的第 j 位, 也就是生成图片预测为属于类别 T_j 的概率. $C_t(G(x_{s_i},l_s^i,l_t^j))^{(j)}$ 越接近 1, 代表生成图片的分布越接近期望的目标域视角 T_j. 同样地, $C_s(G(x_{t_j},l_t^j,s_i^i))^{(i)}$ 表示生成图片预测为属于类别 S_i 的概率.

2.2.4 总目标函数

目标函数由 4 部分组成, 包括对抗损失 (式 (3) 和式 (4))、循环一致损失 (式 (5))、视角分类损失 (式 (7) 和式 (8))、身份保持损失 (式 (9)). 除了前文已经介绍的前三种损失, 还需要身份保持损失, 用来保证迁移过程中身份信息不变, 否则无法得到生成后图片的正确身份标注. 我们使用文献 [34]提出的基于前景掩模的身份保持损失函数. 在多对多生成对抗网络中, 该损失可以表示为

$$L_{\text{id}}(G,\bar{G})=\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\mathrm{E}_{x_{s_i}}\left[\left\|x_{s_i}\circ M(x_{s_i})-G(x_{s_i},l_s^i,l_t^j)\circ M(x_{s_i})\right\|_2\right]+\mathrm{E}_{x_{t_j}}\left[\left\|x_{t_j}\circ M(x_{t_j})-\bar{G}(x_{t_j},l_t^j,l_s^i)\circ M(x_{t_j})\right\|_2\right]\right) \tag{9}$$

其中, $M(\cdot)$ 表示图片的前景掩模, 实验中使用在 COCO (Common objects in context) 数据集[37] 训练过的 Mask R-CNN (Region convolutional neural network)[38] 模型来提取行人的前景. $\|\cdot\|_2$ 表示 L2 范数, "$\circ$"表示逐像素乘法操作. 式 (9) 表示生成前和生成后的图片前景应保持一致, 即 x_{s_i} 与 $G(x_{s_i},l_s^i,l_t^j)$、x_{t_j} 与 $\bar{G}(x_{t_j},l_t^j,l_s^i)$ 前景保持一致.

鉴别器和视角分类器有很强的联系, 鉴别器分辨输入图片是否属于某个域, 视角分类器分辨输入图片属于该域的哪个视角, 它们之间是粗粒度分类和细粒度分类的关系. 于是, 类似于多任务学习, 我们令鉴别器和视角分类器共享一部分参数并让它们共同优化.

综合以上 4 种损失函数, 我们将总目标分为两部分. 生成器的目标函数为

$$L_G=L_{\text{adv}}(G,D_t)+L_{\text{adv}}(\bar{G},D_s)+\lambda_1L_{\text{view}}^{G}(G,\bar{G})+\lambda_2L_{\text{id}}(G,\bar{G})+\lambda_3L_{\text{cyc}}(G,\bar{G}) \tag{10}$$

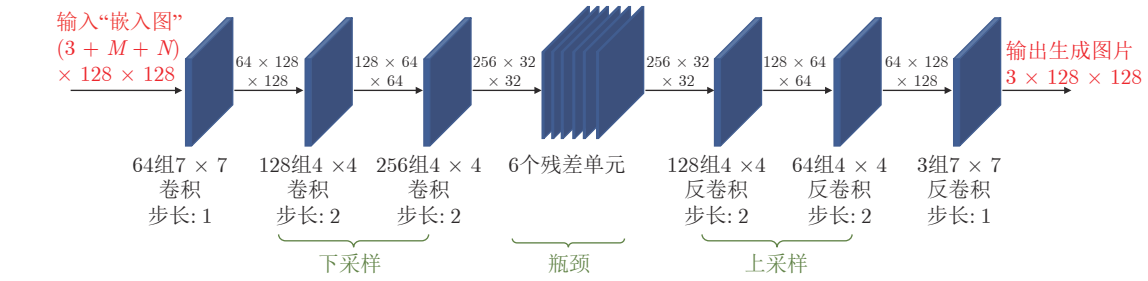
鉴别器和视角分类器的目标函数为

$$L_D=-L_{\text{adv}}(G,D_t)-L_{\text{adv}}(\bar{G},D_s)+\lambda_1L_{\text{view}}^{C}(C_s,C_t) \tag{11}$$

其中, λ_1, λ_2, λ_3 分别用于控制视角分类损失、身份保持损失和循环一致损失的相对重要性.

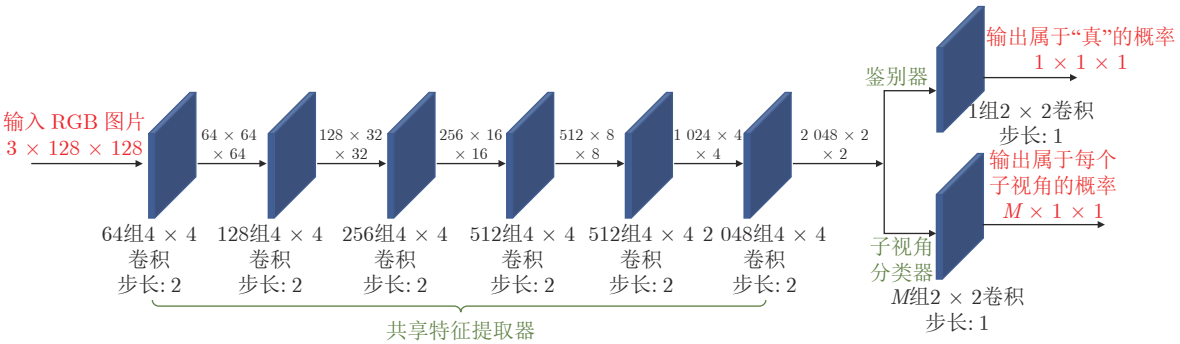
2.2.5 实现细节

在循环生成对抗网络的基础上, 多对多生成对抗网络增加了两个视角分类器. 所以多对多生成对抗网络一共由两个生成器、两个鉴别器和两个视角分类器组成. 图 3 是多对多生成对抗网络的框架图 (图中只完整描述了源域迁移至目标域的全过程, 省略了目标域迁移至源域的生成过程、循环一致损失和身份保持损失). 图 5 是网络结构示意图. 网络结构图中长方体部分表示网络层、箭头上方是每一网



(a) 生成器网络结构

(a) Network structures of the generator



(b) 鉴别器、子视角分类器网络结构

(b) Network structures of the discriminator and the sub-view classifier

图 5　多对多生成对抗网络结构图

Fig. 5　Network structures of our M2M-GAN

络层的输出维度、M 和 N 分别表示源域和目标域子视角类别数. 图 5 中展示的网络结构以源域为例, 目标域的网络结构与此类似, 唯一不同点是子视角类别数不同.

1) 生成器结构. 我们使用两个不同的生成器, 分别用于从源域到目标域、从目标域到源域的迁移, 两个生成器结构一致. 与循环生成对抗网络类似, 我们也采用经典的编码器–解码器 (Encoder-decoder) 模式搭建生成器.

生成器的输入为集成了源域视角和目标域视角信息的嵌入图 (式 (6)), 嵌入图大小为 $(3+M+N)\times 128\times 128$, M 和 N 分别表示源域和目标域的视角数目. 由于源域视角和目标域视角数目不固定, 嵌入图的通道数不是一个确定的值. 所以在对输入数据进行特征编码前, 首先用一个有 64 组滤波器的卷积层对嵌入图进行卷积操作, 把通道数统一调整为 64.

卷积后得到的特征图大小为 $3\times 128\times 128$, 经过连续两次下采样操作得到 $256\times 3\times 3$ 的特征图. 每一次下采样操作会把特征图的宽和高变为原来的一半, 通道数变为原来的两倍. 下采样把图像从高维空间转换到潜在的低维空间, 便于后续对特征进行处理. 下采样可以通过卷积来实现. 接下来, 特征图输入到由 6 个残差单元组成的瓶颈层. 瓶颈层会对特征进行加工, 筛选出能够用于构造目标图片的特征. 最后, 特征图经过连续两次上采样操作, 得到 $256\times 3\times 3$ 的特征图. 上采样和下采样是一一对应的, 上采样把低维特征图重新变成高维图像, 可以通过反卷积实现. 为了使网络最终输出 3 通道 RGB 图片, 网络最后一层是一个由 3 组滤波器组成的卷积层, 把图片通道数重新调整为 3.

2) 鉴别器和视角分类器结构. 鉴别器和视角分类器的学习目标有相似之处, 前者判断输入图片是否服从某个域的整体分布, 后者判断输入图片服从某个域的哪一个视角子分布, 所以我们令鉴别器和视角分类器共享一部分参数. 具体来说, 鉴别器和视角分类器共享特征提取器, 该特征提取器由 6 个卷积层组成. 输入图片经过共享的特征提取器, 提取到对应的特征图, 然后特征图流向两个分支, 进行多任务学习.

其中一个分支是鉴别器, 鉴别器需要对 “真假” 两个类别进行分类, 是二分类问题. 所以我们选择用 1 组滤波器构成 “真假” 分类器, 该滤波器用于计算图片属于真实分布的概率.

另一个分支是视角分类器, 视角分类器需要对

源域 (或目标域) M 个视角 (或N 个视角) 进行分类, 是多分类问题. 所以我们选择用 M 组 (或N 组) 滤波器来构造多分类器. 每个滤波器分别计算图片属于其中某个视角的概率, 多组滤波器输出的是图片属于源域 (或目标域) M 个视角 (或N 个视角) 的概率.

3) 训练方式. 多对多生成对抗网络训练时, 多视角组合是共同训练和优化的. 同一数据集不同视角的图片包含相同的人群、相同的时间段和相似的背景, 这种数据高相关性使得同时优化 $M \times N$ 对视角组合的迁移并不需要 $M \times N$ 倍训练时间 (对比单独优化一对视角的训练时间). 下面举个简单的例子进行说明. 假设源域 S 有 S_1, S_2 两个视角, 目标域 T 有 T_1, T_2 两个视角. 有三种不同的训练方式: a) 用 S_1, S_2, T_1, T_2 所有数据训练一个不区分视角的生成对抗网络 (基本的迁移方法, 不区分视角). b) 单独使用一对视角组合数据训练一组生成对抗网络, 即用 (S_1, T_1)、(S_1, T_2)、(S_2, T_1)、(S_2, T_2) 分别训练一组生成对抗网络 (区分视角, 单独优化). c) 用 S_1, S_2, T_1, T_2 所有数据训练一个区分视角的生成对抗网络 (本文提出的多对多迁移方法, 区分视角, 共同优化). 由于不同视角数据分布相似, 再加上新增的网络结构和损失函数对整体收敛速度影响不大, 所以无论是用一对视角数据 (如 S_1, T_1) 还是全部视角数据 (即 S_1, S_2, T_1, T_2), 无论是训练普通的生成对抗网络还是多对多生成对抗网络, 对于一组生成对抗网络的训练开销是相近的. 因此方式 1、方式 2 和方式 3 训练一组生成对抗网络需要的训练开销近似, 但方式 2 需要训练 4 组生成对抗网络, 因此训练开销大约是其他两种方式的 4 倍. 实验结果可以验证, 多对多迁移不需要大量的训练开销.

多对多生成对抗网络的具体训练过程是: 每次迭代都随机从源域和目标域分别选择 Q 张图片 (Q 等于训练批次大小的一半), 每张图片有两个视角标记, 分别是图片自身的视角标记以及随机生成的另一个域的视角标记. 然后用这 $2Q$ 张图片交替训练鉴别器和生成器. 训练鉴别器和视角分类器时, 生成器的参数固定不变, 用式 (11) 计算损失, 并反向传播更新鉴别器和视角分类器的网络参数; 训练生成器时, 鉴别器和视角分类器的参数固定不变, 用式 (10) 计算损失, 并反向传播更新生成器的网络参数.

2.3 行人特征学习

多对多生成对抗网络训练完成后, 我们就可以利用训练好的生成器把源域的图片变换到目标域, 然后使用变换后的数据集训练行人再识别模型. 因为经过迁移的图片可以使用迁移前的身份标记, 这样就把无监督问题变成了有监督的行人再识别问题, 然后可以采取任意一种有监督的行人再识别方法. 特别地, 在本文中我们采用最常见的基于分类损失的行人再识别框架, 把每个行人看作一个类别, 使用分类损失函数 (交叉熵) 训练

$$L_{\text{cross}} = -\sum_{k=1}^{K} \log\left(p(k)\right) q(k) \tag{12}$$

其中, K 是行人类别数, $p(k)$ 是深度模型预测的样本属于第 k 类行人的概率. $q(k)$ 表示真实的概率分布. 当 k 为输入图片真实类别时, $q(k)$ 值为 1, 否则 $q(k)$ 为 0. 训练结束后, 提取深度模型分类层前一层的网络层输出作为特征描述子, 最后用欧氏距离度量查询图片和所有候选图片的相似性.

2.4 整体流程总结

为了便于理解, 我们对整体流程总结如下:

步骤 1. 训练多对多生成对抗网络. 使用源域的训练集和目标域的训练集训练一个多对多生成对抗网络, 该步骤不使用任何行人身份标注.

步骤 2. 生成新数据集. 多对多生成对抗网络训练完成后, 对于源域的每一张图片, 我们利用训练好的生成器都生成出目标域 N 个视角的新图片, 新图片的身份标注使用对应的原始图片的身份标注.

步骤 3. 训练行人再识别模型. 使用生成的数据集和对应的身份标注, 以监督学习的方式训练一个行人再识别的深度神经网络.

3 实验结果与分析

为了验证本文提出的基于多对多生成对抗网络的迁移方法的有效性, 我们设计了多个实验.

1) 设计了 M2M-GAN 的可视化实验, 定性分析 M2M-GAN 是否能把源域各个视角的图片风格转换成目标域各个视角的图片风格. 该实验的目的是验证 M2M-GAN 是否能够模拟出目标域多视角分布的实际情况.

2) 将 M2M-GAN 生成的图片用于行人再识别的效果. 如果 M2M-GAN 生成的图片能更好地体现目标域多个视角分布的特性, 那么生成的图片就能够提高目标域内跨视角匹配的准确率. 为了验证此观点, 我们设置了两种跨域行人再识别基准实验作为对比, 分别是: a) 无迁移; b) 不区分视角的迁移. 把 M2M-GAN 与这两种基准实验进行对比, 可

以验证我们提出的区分视角的迁移算法的有效性.

3) 设计了训练时间和模型参数量大小分析实验, 对比了 M2M-GAN 和两种迁移方式: a) 经典的不区分视角的迁移模型; b) 针对多种源域−目标域视角组合训练多组生成对抗网络的模型. 与这两种方法对比, 目的是验证第 2.2 节提出的用多组生成对抗网络实现多视角迁移的不可行性和 M2M-GAN 的可实现性的观点.

4) 设计了消融实验, 用来分析 M2M-GAN 模型中的视角嵌入模块、视角分类模块和身份保持模块这 3 个模块的作用. 另外, 该实验也能分析生成过程和鉴别过程中引入视角信息所起的作用. 假如引入多视角信息有助于提高生成图片用于行人再识别的准确率, 则可验证第 2.1 节提出的 “在迁移时引入视角信息是有利的”的观点.

5) 与其他无监督方法进行对比, 从而验证我们方法的先进性.

3.1　实验数据集和算法性能评测指标

本文实验使用了 3 个公开的大规模跨视角行人再识别基准数据集, 包括 Market1501[39] 数据集、DukeMTMC-reID[40] 数据集和 MSMT17[34] 数据集.

Market1501 数据集包含了 6 个监控摄像头数据. 数据集中有 1 501 个行人共 32 668 张图片. 其中训练集包含 751 个行人共 12 936 张图片, 测试集包含 750 个行人共 19 732 张图片, 测试集中有3 368 张图片被随机选为查询集, 剩下的图片作为候选集.

DukeMTMC-reID 数据集包含了 8 个监控摄像头数据. 数据集中一共有 1 812 人, 其中 408 人只有一个摄像机拍摄的图片. 训练集包含 702 个行人共 16 522 张图片, 测试集包含 702 个行人. 测试集中有 2 228 张行人图片被选为查询图片, 其余的 17 661 张图片 (包括 702 个作为测试的行人及 408 个作为干扰的行人) 作为候选集.

MSMT17 数据集是 2018 年新发布的行人再识别数据集, 它更符合实际应用场景. 例如, 数据集包含的摄像头数目更多, 由 15 个监控摄像头组成; 数据集包含的行人数目更多, 共 4 101 人; 拍摄时间和地点跨度更大, 包含了室内室外四天内早中晚三个时段的数据, 光照和背景变化更丰富. 训练集中有 1 041 个行人共 32 621 张图片, 测试集中有 3 060 人共 93 820 张图片. 其中测试集里有 11 659 张图片被随机选为查询图片, 其余 82 616 张图片作为候选集.

本文采用常见的行人再识别评测指标, 包括累积匹配曲线 (Cumulative matching characteristics, CMC) 和平均准确率均值 (Mean average precision, mAP). CMC 表示前 r 个匹配结果中正确匹配的比例, $r = 1$ 表示首位匹配准确率. CMC 主要反映模型的准确率, mAP 则兼顾准确率和召回率. 其中 AP (Average precision) 是某个类别所有返回结果的排列序号 (Rank) 倒数的加权平均值, mAP 就是所有类别 AP 值的平均值.

3.2　实验参数设置

我们使用 Adam 优化器[41] ($\beta_1 = 0.5$, $\beta_2 = 0.999$) 训练 M2M-GAN. 训练数据为源域和目标域的训练集图片和对应的视角标记 (摄像头编号), 不使用任何行人类别标记. 初始学习率设为 0.0001, 经历 100 000 次迭代训练后学习率开始线性递减, 直到第 200 000 次迭代时学习率递减为 0. 网络输入的图像尺寸重新调整为 128×128 像素, 批次大小设为 16. 式 (10) 和式 (11) 中有 3 个超参数 λ_1, λ_2, λ_3, 其中 λ_3 控制循环一致损失的比重, 本文参照文献 [36] 提供的设置, 将其设为 10. 图 6 显示了参数 λ_1 和 λ_2 对识别率的影响, 可以看出当 λ_1 取值为 0.5 ~ 10.0, λ_2 取值为 50 ~ 100 时, 模型都能取得良好的识别率. 特别地, 当 λ_1 和 λ_2 取 1 和 100 时性能最好. 因此, 将 λ_1 和 λ_2 分别设为 1 和 100.

行人特征学习网络使用在 ImageNet 上预训练过的 ResNet50[42] 网络参数 (替换掉最后一层全连接层) 作为模型的初始化参数, 然后使用生成的数据集及对应的行人类别标记来微调. 特征学习网络用随机梯度下降法 (Stochastic gradient descent, SGD) 优化器训练, 最后一层全连接层的初始学习率设为 0.1, 其余层的初始学习率设为 0.01, 训练 30 回合 (Epochs) 后学习率变为原来的 1/10. 网络输入的图像尺寸重新调整为 256×128 像素, 批次大小设为 64.

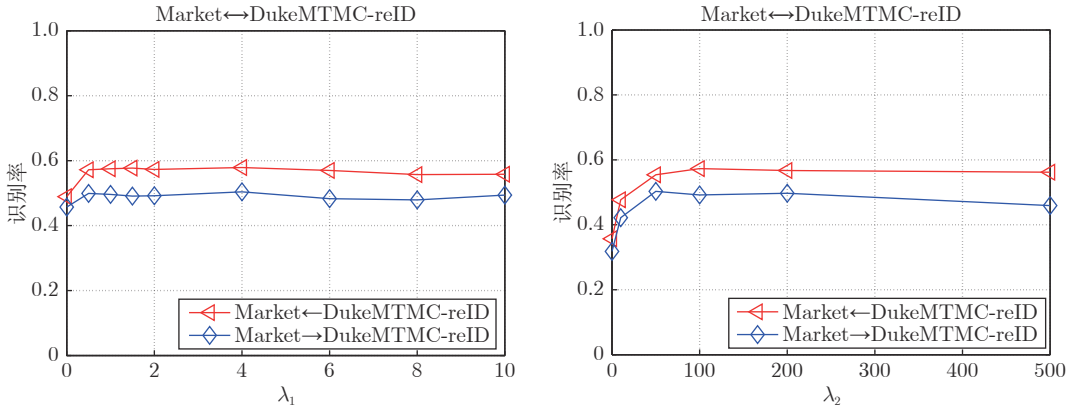
3.3　不同数据集上的实验结果

我们在 3 个数据集上进行实验. 当使用某个数据集作为目标域时, 其余两个数据集分别作为源域来评估该源域迁移到目标域的行人再识别的性能.

3.3.1　迁移到 Market1501 数据集的实验结果

本节实验选择 Market1501 数据集作为目标域, 分别评估 DukeMTMC-reID 和 MSMT17 数据集迁移到 Market1501 数据集的结果.

首先分析 M2M-GAN 生成图片的视觉效果. 图 7(a) 是 Market1501 数据集 6 个视角真实图片的例子, 图 7(b) 是 DukeMTMC-reID → Market1501



(a) 参数 λ_1 对识别率的影响
(a) Effect of parameter λ_1 on the matching rate

(b) 参数 λ_2 对识别率的影响
(b) Effect of parameter λ_2 on the matching rate

图 6 不同参数对识别率的影响

Fig. 6 Influence of different parameters on the matching rate

的效果, 图 7(c) 是 MSMT17 → Market1501 的效果. 图 7(b) 或图 7(c) 最左列是源数据集的真实图片, 右边几列是左边图片变换到目标数据集 (Market1501 数据集) 各个视角的生成图片, 每一列代表一个视角. 可以看出, 生成图片在视觉上更接近于 Market1501 数据集的风格, 而身份信息 (行人外观) 也没有丢失. 各个视角的生成图片之间的差异不显著, 这是因为 Market1501 数据集各个摄像头的地理位置相距很近, 拍摄到的数据分布相差不够明显. 第 3.3.2 节和第 3.3.3 节展示的另外两个数据集更贴近实际应用场景.

然后定量分析迁移的效果, 即把迁移后的数据用于行人再识别任务. 表 1 是分别从 DukeMTMC-reID 数据集、MSMT17 数据集迁移到 Market1501 数据集的跨域迁移行人再识别结果. 为了公平对比, 所有实验使用相同的参数设置. 表中 "Pre-training" 表示直接把从源域真实数据训练得到的模型用于目标域测试. "CycleGAN" 和 "M2M-GAN" 都是先进行图像风格迁移, 再用变换后的新数据集训练行人再识别模型. 其中 "CycleGAN" 相当于文献 [34] 提出的 PTGAN, 迁移时忽略了不同视角的差异性. "M2M-GAN" 是我们提出的基于多视角的非对称迁移方法. 根据表 1 的实验结果可知, "Pre-training" 和 "CycleGAN" 这两种不使用迁移或只用对称迁移算法的效果较差. 与 "CycleGAN" 相比, "M2M-GAN" 的 rank1 提高了 11.7% 和 11.8%, mAP 提高了 8.1% 和 7.7%, 这验证了 "M2M-GAN" 的有效性.

特别地, "CycleGAN" 比 "Pre-training" 的效果有所降低. "Pre-training" 没有使用迁移方法, "CycleGAN" 使用了迁移方法, 但是却比不使用迁移 ("Pre-training") 效果还差, 这说明并不是所有迁移算法都能获得效果提升. 虽然迁移算法能把部分源域的知识迁移到目标域, 但是迁移过程中可能会有信息损失. 仔细观察真实图片和对应的生成图片 (图 7), 会发现生成图片丢失了一些行人细节信息. 例如, 生成图片中行人的鞋子、背包、五官等纹理变得模糊. 使用这些损失了部分细节信息的图片来训练, 会降低模型学习行人判别性特征的能力. 但另一方面, 图片风格迁移又会提高模型对目标域的适应程度. 所以风格迁移算法既有提升 (减小数据集差异) 也有损失 (迁移过程中的信息丢失), 提升大于损失的迁移算法才能比无迁移算法的效果好.

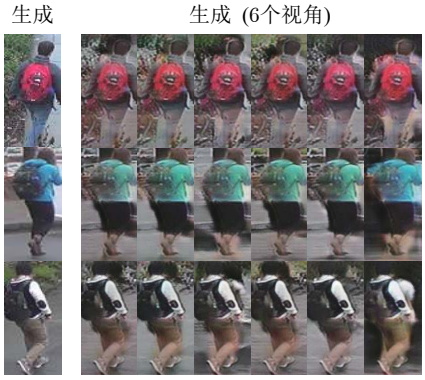
表 1 不同风格迁移方法在 Market1501 数据集上的识别率 (%)

Table 1 Matching rates of different style translation methods on the Market1501 dataset (%)

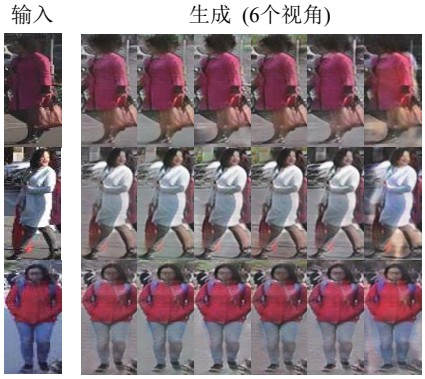
方法 (源域数据集)	DukeMTMC-reID		MSMT17	
	Rank1	mAP	Rank1	mAP
Pre-training	50.4	23.6	51.5	25.5
CycleGAN	47.4	21.5	46.1	21.1
M2M-GAN (本文)	**59.1**	**29.6**	**57.9**	**28.8**

接下来是训练时间和模型参数量的对比. 表 2 是不同方法在 Market1501 数据集上的训练时间和模型参数量对比, 源数据集为 DukeMTMC-reID 数据集. 行人特征学习网络是完全一样的, 所以只需对比生成对抗网络这一阶段. 表 2 中 "CycleGAN"、"M2M-GAN" 的含义与前文解释一致. "$M \times N$ CycleGAN" 是本文设置的基准实验, 在 "CycleGAN" 基础上考虑了多视角问题, 对每种源域–目标

(a) Market1501数据集6个视角真实图片
(a) Real images from 6 views of Market1501



(b) DukeMTMC-reID→Market1501



(c) MSMT17→Market1501

图 7　其他数据集迁移到 Market 数据集的可视化例子
Fig. 7　Visual examples of translations from other datasets to the Market1501 dataset

域视角组合都单独训练一组生成对抗网络. 从表 2 可知, “$M \times N$ CycleGAN” 和 “M2M-GAN”都能提升准确率, 说明非对称迁移的重要性, 但两者训练开销相差较大, “M2M-GAN”开销更少. 另外, “M2M-GAN”准确率略高于 “$M \times N$ CycleGAN”, 说明同时用多对视角组合数据训练比单独用各对视角组合数据训练更能利用不同视角数据之间的相关性, 使得迁移效果更好.

为了验证不同模块的重要性, 我们进行了消融实验. 本文在循环生成对抗网络基础上添加了视角嵌入模块、视角分类模块和身份保持模块 (身份保持模块参考文献 [34]), 我们分别验证这些模块的作用. 表 3 是在生成阶段使用不同的网络模块, 最终得到的行人再识别准确率. 源数据集是 DukeMTMC-reID 数据集, 目标数据集是 Market1501 数据集. 由于身份保持模块是基于图片风格迁移的跨域行人再识别这类算法的基本模块, 所以我们在保留了身份保持模块的基础上再对视角嵌入、视角分类这两个模块进行分析. 对比表 3 的第 1 行和第 2 行, 可以验证身份保持的重要性. 对比表 3 的第 2～4 行, 可以发现单独使用视角嵌入模块或者视角分类模块都能提高准确率, 但准确率提升不明显. 对比表 3 的第 3～5 行, 可以发现同时使用视角嵌入模块和视角分类模块可以获得更显著的性能提升. 这是因为视角嵌入模块为生成器提供了辅助信息输入, 视角分类模块为生成器提供了监督信号, 两个模块配合使用才能取得理想的效果. 另外, 视角分类损失和身份保持损失的权重分析如图 6 所示, 当 λ_1 取值为 0.5～10.0, λ_2 取值为 50～500 时, 模型都能取得良好的识别率.

表 2　不同方法在 Market1501 数据集上的训练时间和模型参数量
Table 2　Training time and model parameters of different methods on the Market1501 dataset

方法	训练时间	模型参数量	Rank1 (%)
CycleGAN	**16 h**	**106.3 M**	47.4
$M \times N$ CycleGAN	14 h×8×6	106.3 M×8×6	58.0
M2M-GAN (本文)	17 h	106.6 M	**59.1**

表 3　不同模块在 Market1501 数据集上的准确率分析 (%)
Table 3　Accuracy of different modules on the Market1501 dataset (%)

视角嵌入模块	视角分类模块	身份保持模块	Rank1	mAP
×	×	×	35.7	12.5
×	×	√	47.4	21.5
√	×	√	48.0	22.0
×	√	√	48.6	22.1
√	√	√	**59.1**	**29.6**

最后我们还与目前最先进的无监督行人再识别方法进行了对比. 将对比的方法分成 3 组, 第 1 组是基于手工特征和欧氏距离的方法; 第 2 组是基于聚类的无监督方法; 第 3 组是基于跨域迁移学习的方法. 这些方法都没有用到目标域的标注信息, 属于无监督学习范畴. 由表 4 可知, 本文的方法在

表 4 不同无监督方法在 Market1501 数据集上的识别率 (%) (源数据集为 DukeMTMC-reID 数据集)

Table 4 Matching rates of different unsupervised methods on the Market1501 dataset (%) (The source dataset is the DukeMTMC-reID dataset)

类型	方法	Rank1	mAP
手工特征	LOMO[12]	27.2	8.0
	Bow[39]	35.8	14.8
基于聚类的无监督学习	PUL[29]	45.5	20.5
	CAMEL[28]	54.5	26.3
跨域迁移学习	PTGAN[34]	38.6	—
	SPGAN+LMP[33]	57.7	26.7
	TJ-AIDL[43]	58.2	26.5
	ARN[44]	70.2	39.4
	M2M-GAN (本文)	59.1	29.6
	M2M-GAN (本文)+LMP[33]	63.1	30.9

Market1501 数据集上的 Rank1 达到 63.1%, mAP 达到 30.9%, 高于大多数算法, 特别是高于其他基于生成对抗网络的方法. 另外, TJ-AIDL[43] 使用了行人属性信息, 但我们的结果依然比该方法要好. 虽然本文的方法低于 ARN (Adaptation and re-identification network)[44], 但是本文与 ARN 是不同类型的算法. 基于生成对抗网络的这类方法单独考虑迁移和行人再识别问题, 生成的风格迁移图像可以直接输入到各种有监督行人再识别模型中, 不需要调整行人再识别模型结构. 所以基于生成多抗网络的跨域迁移方法具有很强的灵活性和发展前景. 并且当图片生成技术或者有监督行人再识别技术有所提升时, 基于生成对抗网络的跨域行人再识别性能也将得到进一步提高. 因此, 本文提出的方法虽然低于 ARN, 但在基于生成对抗网络的这一类方法中取得最好的效果, 同样具有应用前景和研究价值.

3.3.2 迁移到 DukeMTMC-reID 数据集的实验结果

本节实验选择 DukeMTMC-reID 数据集作为目标域, 所有图表的含义与第 3.3.1 节实验完全相同.

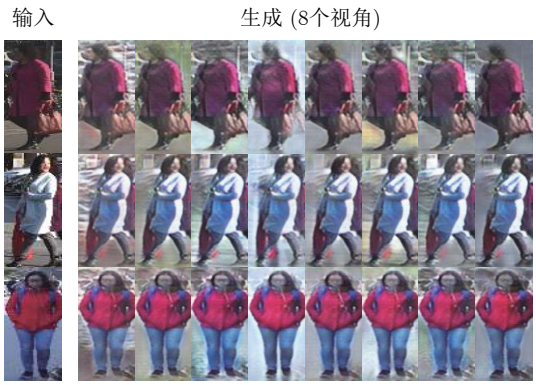
图 8 是 M2M-GAN 生成图片的视觉效果, 可以看出生成图片在视觉上更接近于 DukeMTMC-reID 数据集的风格, 并且各个视角的生成图片之间有明显的差别, 主要差别是图片的背景. 迁移到 DukeMTMC-reID 数据集不同视角的图片有不同的背景, 这与 DukeMTMC-reID 数据集实际的数据分布是相符的.

然后是定量分析迁移的效果. 从表 5 可以得出与 Market 数据集实验类似的结论. 与 “Pre-training”和 “CycleGAN”相比, “M2M-GAN” 有效提高了行人再识别准确率.

(a) DukeMTMC-reID数据集8个视角真实图片

(a) Real images from 8 views of DukeMTMC-reID

(b) Market→DukeMTMC-reID



(c) MSMT17→DukeMTMC-reID

图 8 其他数据集迁移到 DukeMTMC-reID 数据集的可视化例子

Fig. 8 Visual examples of translations from other datasets to the DukeMTMC-reID dataset

接下来是训练时间和模型参数量的对比. 由表 6 可知, “M2M-GAN”的准确率比 “CycleGAN”和 “$M\times N$ CycleGAN”高, 而训练时间和网络参数量只略微高于 “CycleGAN”, 远低于 “$M\times N$ CycleGAN”. 再一次表明了 “M2M-GAN”有效提升了识别准确率, 同时不需要大量的训练开销.

表 7 是消融实验结果. 与第 3.3.1 节 Market1501

表 5　不同风格迁移方法在 DukeMTMC-reID 数据集上的识别率 (%)

Table 5　Matching rates of different style translation methods on the DukeMTMC-reID dataset (%)

方法 (源域数据集)	Market1501		MSMT17	
	Rank1	mAP	Rank1	mAP
Pre-training	38.1	21.4	53.5	32.5
CycleGAN	43.1	24.1	51.1	30.0
M2M-GAN (本文)	**52.0**	**29.8**	**61.1**	**37.5**

表 6　不同方法在 DukeMTMC-reID 数据集上的训练时间和模型参数量

Table 6　Training time and model parameters of different methods on the DukeMTMC-reID dataset

方法	训练时间	模型参数量	Rank1 (%)
CycleGAN	**16 h**	**106.3 M**	43.1
$M\times N$ CycleGAN	14 h$\times$6$\times$8	106.3 M$\times$6$\times$8	49.9
M2M-GAN (本文)	17 h	106.6 M	**52.0**

表 7　不同模块在 DukeMTMC-reID 数据集上的准确率分析 (%)

Table 7　Accuracy of different modules on the DukeMTMC-reID dataset (%)

视角嵌入模块	视角分类模块	身份保持模块	Rank1	mAP
×	×	×	31.8	12.6
×	×	√	43.1	24.1
√	×	√	45.0	25.3
×	√	√	43.5	24.1
√	√	√	52.0	29.8

数据集上的消融实验结果类似, 分别加入视角嵌入模块或者视角分类模块都能使行人再识别准确率略微提升. 但是视角嵌入和视角分类两个模块同时使用, 可以取得更显著的提升. 这说明了本文方法的每个模块都是有效的, 并且同时使用能够取得更好的效果.

最后是与其他方法进行对比. 由表 8 可知, 本文的方法在 Market1501 数据集上的 Rank1 达到 54.4%, mAP 达到 31.6%, 超过了其他基于生成对抗网络的方法, 仅低于 ARN[44].

3.3.3　迁移到 MSMT17 数据集的实验结果

本节实验选择 MSMT17 数据集作为目标域, 所有图表的含义与第 3.3.1 节实验完全相同. 图 9 是其他两个数据集迁移到 MSMT17 数据集的可视化例子. 可以看出, 生成的图片在保持身份信息不变的同时视觉上更接近于 MSMT17 数据集的风格, 并且各个视角的生成图片服从不同的分布. 最明显

表 8　不同无监督方法在 DukeMTMC-reID 数据集上的识别率 (%) (源数据集为 Market1501 数据集)

Table 8　Matching rates of different unsupervised methods on the DukeMTMC-reID dataset (%) (The source dataset is the Market1501 dataset)

类型	方法	Rank1	mAP
手工特征	LOMO[12]	12.3	4.8
	Bow[39]	17.1	8.3
基于聚类的无监督学习	UMDL[45]	18.5	7.3
	PUL[29]	30.0	16.4
跨域迁移学习	PTGAN[34]	27.4	–
	SPGAN+LMP[33]	46.4	26.2
	TJ-AIDL[43]	44.3	23.0
	ARN[44]	60.2	33.4
	M2M-GAN (本文)	52.0	29.8
	M2M-GAN (本文)+LMP[33]	54.4	31.6

的差别是光照不同, 其次是背景信息不同, 这与 MSMT17 数据集的实际情况完全相符, 也验证了多对多生成对抗网络能同时进行多个视角的迁移, 而且迁移效果更好.

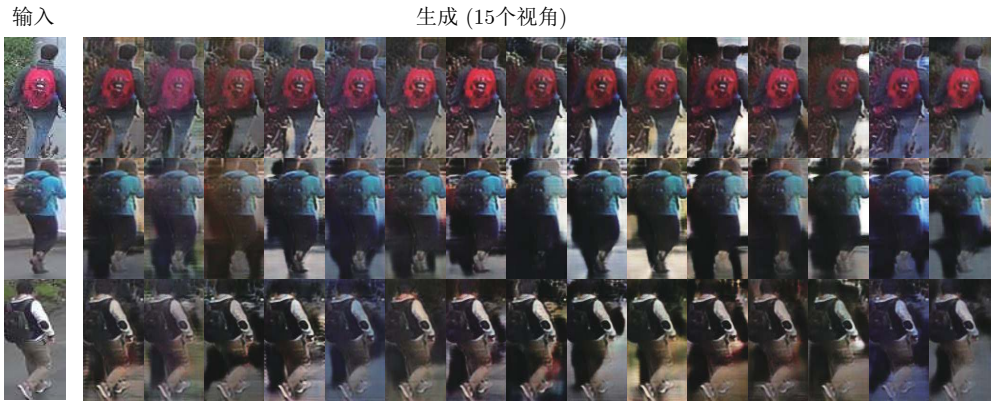
定量分析行人再识别的准确率也得出与前两个数据集类似的结果. 由表 9 可得, 在 MSMT17 数据集上, “M2M-GAN”比 “CycleGAN”和 “Pre-training”这两种方法的效果好很多. 与 “Pre-training”相比, “M2M-GAN”的 Rank1 提升了 17.7% 和 16.6%. 与 “CycleGAN”相比, “M2M-GAN”的 Rank1 提升了 9.2% 和 12.1%. 与前两个数据集相比, “M2M-GAN”在 MSMT17 数据集的优势最明显, 原因在于 MSMT17 数据集摄像头数目多、摄像头数据分布差异大. 这说明了我们提出的 “M2M-GAN”比其他方法更能适用于复杂的现实场景. 表 10 与其他方法对比也体现 “M2M-GAN”的良好效果.

4　结论

目前跨域迁移行人再识别的方法忽略了域内多个视角子分布的差异性, 导致迁移效果不好. 本文提出了基于多视角 (摄像机) 的非对称跨域迁移的新问题, 并针对这一问题设计了多对多生成对抗网络 (M2M-GAN). M2M-GAN 考虑了源域多个视角子分布的差异性和目标域多个视角子分布的差异性, 并共同优化所有的源域–目标域视角组合. 与现有不考虑视角差异的迁移方法相比, M2M-GAN 取到更高的识别准确率. 在 3 个大规模行人再识别基准数据集上, 实验结果充分验证了本文提出的 M2M-GAN 方法的有效性. 未来的研究工作将考虑除视角以外的其他划分子域的方式, 例如按照不同的光照或行人背景进行划分, 这样能更好地刻画出

(a) MSMT17数据集15个视角真实图片
(a) Real images from 15 views of MSMT17

(b) Market1501→MSMT17



(c) DukeMTMC-reID→MSMT17

图 9 其他数据集迁移到 MSMT17 数据集的可视化例子

Fig. 9 Visual examples of translations from other datasets to the MSMT17 dataset

表 9 不同风格迁移方法在 MSTM17 数据集上的识别率 (%)

Table 9 Matching rates of different style translation methods on the MSTM17 dataset (%)

方法 (源域数据集)	Market1501		DukeMTMC-reID	
	Rank1	mAP	Rank1	mAP
Pre-training	14.2	4.5	20.2	6.7
CycleGAN	22.7	7.6	24.7	7.8
M2M-GAN (本文)	**31.9**	**10.8**	**36.8**	**11.9**

表 10 不同无监督方法在 MSMT17 上的识别率 (%) (源数据集为 Market1501 数据集)

Table 10 Matching rates of different unsupervised methods on the MSMT17 dataset (%) (The source dataset is the Market1501 dataset)

类型	方法	Rank1	mAP
跨域迁移学习	PTGAN[34]	10.2	2.9
	M2M-GAN (本文)	31.9	**10.8**
	M2M-GAN (本文)+LMP[33]	**32.2**	9.7

一个域内的多个子分布的统计特性, 进一步提高迁移的性能.

References

1 Li You-Jiao, Zhuo Li, Zhang Jing, Li Jia-Feng, Zhang Hui. A survey of person re-identification. *Acta Automatica Sinica*, 2018, **44**(9): 1554−1568
(李幼蛟, 卓力, 张菁, 李嘉锋, 张辉. 行人再识别技术综述. 自动化学报, 2018, **44**(9): 1554−1568)

2 Qi Mei-Bin, Tan Sheng-Shun, Wang Yun-Xia, Liu Hao, Jiang Jian-Guo. Multi-feature subspace and kernel learning for person re-identification. *Acta Automatica Sinica*, 2016, **42**(2): 299−308
(齐美彬, 檀胜顺, 王运侠, 刘皓, 蒋建国. 基于多特征子空间与核学习的行人再识别. 自动化学报, 2016, **42**(2): 299−308)

3 Liu Yi-Min, Jiang Jian-Guo, Qi Mei-Bin, Liu Hao, Zhou Hua-Jie. Video-based person re-identification method based on GAN and pose estimation. *Acta Automatica Sinica*, 2020, **46**(3): 576−584
(刘一敏, 蒋建国, 齐美彬, 刘皓, 周华捷. 融合生成对抗网络和姿态估计的视频行人再识别方法. 自动化学报, 2020, **46**(3): 576−584)

4 Wang G C, Lai J H, Xie X H. P2SNet: Can an image match a video for person re-identification in an end-to-end way? *IEEE Transactions on Circuits and Systems for Video Technology*, 2018, **28**(10): 2777−2787

5 Feng Z X, Lai J H, Xie X H. Learning view-specific deep networks for person re-identification. *IEEE Transactions on Image Processing*, 2018, **27**(7): 3472−3483

6 Zhuo J X, Chen Z Y, Lai J H, Wang G C. Occluded person re-identification. In: Proceedings of the 2018 IEEE International Conference on Multimedia and Expo. San Diego, USA: IEEE, 2018. 1−6

7 Chen Y C, Zhu X T, Zheng W S, Lai J H. Person re-identification by camera correlation aware feature augmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, **40**(2): 392−408

8 Gong S G, Cristani M, Yan S C, Loy C C. *Person Re-identification*. London: Springer, 2014. 139−160

9 Chen Y C, Zheng W S, Lai J H, Pong C Y. An asymmetric distance model for cross-view feature mapping in person reidentification. *IEEE Transactions on Circuits and Systems for Video Technology*, 2017, **27**(8): 1661−1675

10 Chen Y C, Zheng W S, Lai J H. Mirror representation for modeling view-specific transform in person re-identification. In: Proceedings of the 24th International Conference on Artificial Intelligence. Buenos Aires, Argentina: AAAI Press, 2015. 3402−3408

11 Zheng W S, Li X, Xiang T, Liao S C, Lai J H, Gong S G. Partial person re-identification. In: Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015. 4678−4686

12 Liao S C, Hu Y, Zhu X Y, Li S Z. Person re-identification by local maximal occurrence representation and metric learning. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015. 2197−2206

13 Wu A C, Zheng W S, Lai J H. Robust depth-based person re-identification. *IEEE Transactions on Image Processing*, 2017, **26**(6): 2588−2603

14 Köstinger M, Hirzer M, Wohlhart P, Roth P M, Bischof H. Large scale metric learning from equivalence constraints. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE, 2012. 2288−2295

15 Prosser B, Zheng W S, Gong S G, Xiang T. Person re-identification by support vector ranking. In: Proceedings of the British Machine Vision Conference. Aberystwyth, UK: British Machine Vision Association, 2010. 1−11

16 Zheng L, Bie Z, Sun Y F, Wang J D, Su C, Wang S J, et al. Mars: A video benchmark for large-scale person re-identification. In: Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands: Springer, 2016. 868−884

17 Yi D, Lei Z, Liao S C, Li S Z. Deep metric learning for person re-identification. In: Proceedings of the 22nd International Conference on Pattern Recognition. Stockholm, Sweden: IEEE, 2014. 34−39

18 Cheng D, Gong Y H, Zhou S P, Wang J J, Zheng N N. Person re-identification by multi-channel parts-based CNN with improved triplet loss function. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016. 1335−1344

19 Zheng Z D, Zheng L, Yang Y. Pedestrian alignment network for large-scale person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, 2019, **29**(10): 3037−3045

20 Zhao L M, Li X, Zhuang Y T, Wang J D. Deeply-learned part-aligned representations for person re-identification. In: Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017. 3239−3248

21 Luo H, Jiang W, Zhang X, Fan X, Qian J J, Zhang C. AlignedReID++: Dynamically matching local information for person re-identification. *Pattern Recognition*, 2019, **94**: 53−61

22 Zhong Z, Zheng L, Zheng Z D, Li S Z, Yang Y. Camera style adaptation for person re-identification. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 5157−5166

23 Su C, Li J N, Zhang S L, Xing J L, Gao W, Tian Q. Pose-driven deep convolutional model for person re-identification. In: Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017. 3980−3989

24 Su C, Yang F, Zhang S L, Tian Q, Davis L S, Gao W. Multi-task learning with low rank attribute embedding for person re-identification. In: Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015. 3739−3747

25 Song C F, Huang Y, Ouyang W L, Wang L. Mask-guided contrastive attention model for person re-identification. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 1179−1188

26 Kalayeh M M, Basaran E, Gökmen M, Kamasak M E, Shah M. Human semantic parsing for person re-identification. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 1062−1071

27 Wang G C, Lai J H, Huang P G, Xie X H. Spatial-temporal person re-identification. In: Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Hawaii, USA: AAAI, 2019. 8933−8940

28 Yu H X, Wu A C, Zheng W S. Cross-view asymmetric metric learning for unsupervised person re-identification. In: Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017. 994−1002

29 Fan H H, Zheng L, Yan C G, Yang Y. Unsupervised person re-identification: Clustering and fine-tuning. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2018, **14**(4): Article No. 83

30 Lin Y T, Dong X Y, Zheng L, Yan Y, Yang Y. A bottom-up clustering approach to unsupervised person re-identification. In: Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Hawaii, USA: AAAI, 2019. 8738−8745

31 Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, et al. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 2016, **17**(1): 2096−2030

32 Ma A J, Li J W, Yuen P C, Li P. Cross-domain person reidentification using domain adaptation ranking SVMs. *IEEE Transactions on Image Processing*, 2015, **24**(5): 1599−1613

33 Deng W J, Zheng L, Ye Q X, Kang G L, Yang Y, Jiao J B. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018.

994−1003

34 Wei L H, Zhang S L, Gao W, Tian Q. Person transfer GAN to bridge domain gap for person re-identification. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 79−88

35 Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: Proceedings of the 27th Conference on Neural Information Processing Systems. Quebec, Canada: NIPS, 2014. 2672−2680

36 Zhu J Y, Park T, Isola P, Efros A A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017. 2242−2251

37 Lin T Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. Microsoft COCO: Common objects in context. In: Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: Springer, 2014. 740−755

38 He K M, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. In: Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017. 2980−2988

39 Zheng L, Shen L Y, Tian L, Wang S J, Wang J D, Tian Q. Scalable person re-identification: A benchmark. In: Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015. 1116−1124

40 Zheng Z D, Zheng L, Yang Y. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro. In: Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017. 3774−3782

41 Kingma D P, Ba J. Adam: A method for stochastic optimization. In: Proceedings of the 3rd International Conference on Learning Representations. San Diego, USA: ICLR, 2014. 1−13

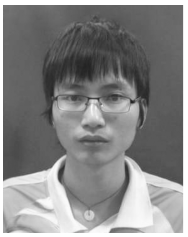
42 He K M, Zhang X Y, Ren S Q, Sun J. Deep residual learning for image recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016. 770−778

43 Wang J Y, Zhu X T, Gong S G, Li W. Transferable joint attribute-identity deep learning for unsupervised person re-identification. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 2275−2284

44 Li Y J, Yang F E, Liu Y C, Yeh Y Y, Du X F, Wang Y C F. Adaptation and re-identification network: An unsupervised deep transfer learning approach to person re-identification. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Salt Lake City, USA: IEEE, 2018. 172−178

45 Peng P X, Xiang T, Wang Y W, Pontil M, Gong S G, Huang T J, et al. Unsupervised cross-dataset transfer learning for person re-identification. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016. 1306−1315

梁文琦 中山大学计算机学院硕士研究生. 2018 年获中山大学计算机科学与技术学士学位. 主要研究方向为行人再识别和深度学习.
E-mail: liangwq8@mail2.sysu.edu.cn
(**LIANG Wen-Qi** Master student at the School of Computer Science and Engineering, Sun Yat-sen University. She received her bachelor degree in intelligence science and technology from Sun Yat-sen University in 2018. Her research interest covers person re-identification and deep learning.)



王广聪 中山大学计算机学院博士研究生. 2015 年获吉林大学通信工程学院学士学位. 主要研究方向为行人再识别和深度学习.
E-mail: wanggc3@mail2.sysu.edu.cn
(**WANG Guang-Cong** Ph.D. candidate at the School of Computer Science and Engineering, Sun Yat-sen University. He received his bachelor degree in communication engineering from Jilin University in 2015. His research interest covers person re-identification and deep learning.)

赖剑煌 中山大学教授. 1999 年获得中山大学数学系博士学位. 目前在 *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, *IEEE Transactions on Neural Networks and Learning Systems (TNNLS)*, *IEEE Transactions on Image Processing (TIP)*, *IEEE Transactions on Systems*, *Man*, and *Cybernetics Part B — Cybernetics (TSMC-B)*, *Pattern Recognition (PR)*, *IEEE International Conference on Computer Vision (ICCV)*, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, *IEEE International Conference on Data Mining (ICDM)* 等国际权威刊物发表论文 200 多篇. 主要研究方向为图像处理, 计算机视觉, 模式识别. 本文通信作者. E-mail: stsljh@mail.sysu.edu.cn
(**LAI Jian-Huang** Professor at Sun Yat-sen University. He received his Ph.D. degree in mathematics from Sun Yat-sen University in 1999. He has published over 200 scientific papers in international journals and conferences including *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, *IEEE Transactions on Neural Networks and Learning Systems (TNNLS)*, *IEEE Transactions on Image Processing (TIP)*, *IEEE Transactions on Systems*, *Man*, and *Cybernetics Part B — Cybernetics (TSMC-B)*, *Pattern Recognition (PR)*, *IEEE International Conference on Computer Vision (ICCV)*, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, *IEEE International Conference on Data Mining (ICDM)*. His research interest covers digital image processing, computer vision, and pattern recognition. Corresponding author of this paper.)