基于图像和特征联合约束的跨模态行人重识别

张玉康 [1, 2] 谭 磊 [1, 2] 陈靓影 [1, 2]

摘 要 近年来, 基于可见光与近红外的行人重识别研究受到业界人士的广泛关注. 现有方法主要是利用二者之间的相互转换以减小模态间的差异. 但由于可见光图像和近红外图像之间的数据具有独立且分布不同的特点, 导致其相互转换的图像与真实图像之间存在数据差异. 因此, 本文提出了一个基于图像层和特征层联合约束的可见光与近红外相互转换的中间模态, 不仅实现了行人身份的一致性, 而且减少了模态间转换的差异性. 此外, 考虑到跨模态行人重识别数据集的稀缺性, 本文还构建了一个跨模态的行人重识别数据集, 并通过大量的实验证明了文章所提方法的有效性, 本文所提出的方法在经典公共数据集 SYSU-MM01 上比 D2RL 算法在 Rank-1 和 mAP 上分别高出 4.2% 和 3.7%, 该方法在本文构建的 Parking-01 数据集的近红外检索可见光模式下比 ResNet-50 算法在 Rank-1 和 mAP 上分别高出 10.4% 和 10.4%.

关键词 跨模态, 行人重识别, 中间模态, 联合约束

引用格式 张玉康, 谭磊, 陈靓影. 基于图像和特征联合约束的跨模态行人重识别. 自动化学报, 2021, **47**(8): 1943−1950

DOI 10.16383/j.aas.c200184

Cross-modality Person Re-identification Based on Joint Constraints of Image and Feature

ZHANG Yu-Kang[1, 2] TAN Lei[1, 2] CHEN Jing-Ying[1, 2]

Abstract In recent years, the research of person re-identification based on visible and near-infrared has attracted widespread attention from the industry. The existing methods mainly use the mutual conversion between them to reduce the difference between their modalities. However, due to the problem of data independence and different distribution between visible image and near-infrared image, there is a large difference between the converted image and the real image, which leads to further improvement of this method. Therefore, this paper proposes a middle modality of conversion between visible and near-infrared modality. So visible and near-infrared can be seamlessly transferred, realizing the identity consistency of person and reducing the difference of conversion between modalities. In addition, considering the scarcity of cross modality person re-identification dataset, this paper also constructs a cross modality person re-identification dataset, and proves the effectiveness of the proposed method through a large number of experiments. In the All-Search Single-shot mode on the SYSU-MM01 dataset, the result of the proposed method is 4.2% and 3.7% higher than Rank1 and mAP using the D2RL algorithm, respectively. Compared with ResNet-50 algorithm, the result of the proposed method on the Parking-01 dataset constructed in this paper is 10.4% and 10.4% higher in Rank-1 and mAP respectively.

Key words Cross modality, person re-identification, middle modality, joint constraint

Citation Zhang Yu-Kang, Tan Lei, Chen Jing-Ying. Cross-modality person re-identification based on joint constraints of image and feature. *Acta Automatica Sinica*, 2021, **47**(8): 1943−1950

近年来, 随着城市监控网络的不断完善, 行人重识别技术由于其巨大的应用潜力而受到越来越多的关注. 给定一个需要检索的行人图像, 行人重识别的任务是检索出一段时间内由非重叠区域下的摄像机所拍摄到的所有该行人图像, 其在智能监控、行人追踪、行为分析等计算机视觉应用及公共安全领域扮演着十分重要的角色[1−4].

当前行人重识别研究方法大多都专注于解决在可见光条件下人体姿态、背景、光照等的问题. 因此, 此类方法主要采用行人特征提取、相似性判别或基于生成式[5−8] 的方式来实现行人重识别. 例如, Zhao 等[5] 提出了一种基于人体区域引导的多级特征分解和树状结构竞争特征融合的 Spindle-Net 网络, 其主要用于对齐人体语义区域来解决行人重识别问题; Sun 等[6] 提出采用基于注意力机制的方式, 把行人水平均匀分割成六个子块, 并对其进行局部调整对

收稿日期 2020-04-03 录用日期 2020-10-19
Manuscript received April 3, 2020; accepted October 19, 2020
国家自然科学基金面上项目 (61977027), 湖北省科技创新重大专项 (2019AAA044) 资助
Supported by General Program of National Natural Science Foundation of China (61977027), Major scientific and Technological Innovation Projects in Hubei Province (2019AAA044)
本文责任编委 黄庆明
Recommended by Associate Editor HUANG Qing-Ming
1. 华中师范大学教育大数据国家工程实验室 武汉 430072 2. 华中师范大学国家数字化学习工程技术研究中心 武汉 430072
1. National Engineering Laboratory for Big Data for Education, Central China Normal University, Wuhan 430072 2. National Engineering Research Center for E-Learning, Central China Normal University, Wuhan 430072

齐, 极大地改善了行人重识别的效果; Hermans 等[7] 提出了一种改进的三元组损失函数, 其约束条件在于除要求行人类内距离小于类间距离, 还使其小于某个阈值来提升行人重识别的效果; PTGAN 算法[8] 提出一种保持行人图像前景不变而将背景迁移为目标图像背景的方法, 极大地缓解了行人重识别研究所面临的数据标注困难的问题.

然而, 在实际的监控系统中, 特别是在光照不足的条件下, 摄像机通常需要从可见光模式切换到近红外模式来应对这种情况. 因此, 在将此类方法应用于实际场景之前, 有必要考虑可见光与近红外跨模态下的行人重识别问题.

基于跨模态下的行人重识别已成为近两年来业内人士的一个重要关注点. 其研究目标是对可见光状态下 (自然状态) 和近红外状态下 (摄像机所捕捉到行人不同光谱的状态) 的行人进行匹配[9−13]. 目前, 该方向主要有两种思路: 一种是基于近红外和可见光模态下的行人特征提取方法; 另一种是基于生成式的方式 (Generative adversarial networks, GANs) 将两种跨模态下的行人转换成同一种模态, 以实现行人重识别过程.

针对前者, Wu 等[9] 提出了一种基于深度零填充的方式将两种模态以参数共享的方法进行训练来解决行人重识别问题. Ye 等[10] 提出了一种基于双向双约束 Top-ranking 损失的双路网络来提取行人特征. 此外, 在其另一项工作中[11], 他们提出了一个层级跨模态匹配模型来联合优化行人在特定模态和共享模态下的特征描述. 对于后者, Dai 等[12] 设计一个基于判别器的生成对抗训练模型, 从不同的模态中学习具有判别力的特征. 为了减少模态差异, Wang 等[13] 提出一种将近红外图像和可见光图像进行相互转换的方法, 并提取相应模态下的行人图像特征. Wang 等[14] 认为灰度图像比彩色图像的识别效果高, 将彩色图像全部转换为灰度图像并用于行人重识别中.

上述提到的方法虽然在一定程度上提升了跨模态行人重识别的精度, 但由于可见光图像和近红外图像具有数据独立且分布不同的特点, 导致其相互转换的图像与真实图像之间存在数据差异. 基于此, 本文设计了一种新颖的中间模态生成器, 通过将两种模态分别进行特征提取后, 以自适应的方式解码在一个共享的潜在特征空间, 进而转化为中间模态的图像, 利用其潜在的特征空间来实现可见光与近红外之间的迁移, 从而提升行人重识别的效果. 实验表明, 本文方法不仅可以减少跨模态行人重识别的模态差异, 而且还能保持行人外貌特征的一致性, 极大地提升了跨模态行人重识别的精度.

在此基础上, 为了保留生成图像与真实图像之间行人身份的一致性, 本文提出特征约束模块和图像约束模块, 从特征层和图像层分别对中间模态生成器进行约束.

另外, 在基于监督的行人重识别中, 数据集的标注是一个耗时耗力的工作, 而跨模态的数据集的标注更加困难, 加剧了行人重识别算法设计的复杂度. 因此, 本文提出了一个用于评估实际监控场景下的跨模态行人重识别数据集, 本数据集仅用于测试, 而不用于训练, 详细见下文的第 3.2 节.

综上所述, 本文贡献主要包括以下三个方面:

1) 本文提出了一种新的中间模态生成器, 用于解决近红外与可见光状态下的行人重识别过程中所存在模态差异性的问题.

2) 为了保持生成器在生成过程中行人身份的一致性, 本文提出了一个特征约束模块和图像约束模块, 分别用于特征层和图像层的联合约束.

3) 针对跨模态行人重识别数据集的匮乏, 本文提出了一个用于评估实际监控场景下基于跨模态的行人重识别数据集.

实验结果证明了该方法的有效性, 相对于当前跨模态的行人重识别方法, 本文所提出的算法取得了较大的性能优势.

1 方法

1.1 总体框架

在此小节中, 将介绍本文所提出的基于图像和特征联合约束的中间模态行人重识别方法, 如图 1 所示, 本文所提出的方法的完整结构包括: 中间模态生成器 (Middle modality generator, MMG)、特征约束模块 (Feature constraint module, FCM) 和图像约束模块 (Image constraint module, ICM). MMG 模块是为了解决跨模态行人重识别中由于图像成像的变化而导致的模态差异问题, 通过加入 MMG 模块, ICM 模块可以更好地关注不同行人之间的距离; 而 ICM 模块也可以反过来对 MMG 模块进行约束, 促进 MMG 模块寻找更加合适的图像和特征, 因此本文所提出的 MMG 模块和 ICM 模块两个可以相互促进、共同优化.

在训练阶段, 每个输入图像都被用于训练采用了生成对抗网络的近红外与可见光模态编码器. 同时, 共享解码器将利用这两个编码器中的中间特征来解码到中间模态图像, MMG 所生成的中间模态图像作为了行人重识别 (ICM) 的输入. 由于 MMG

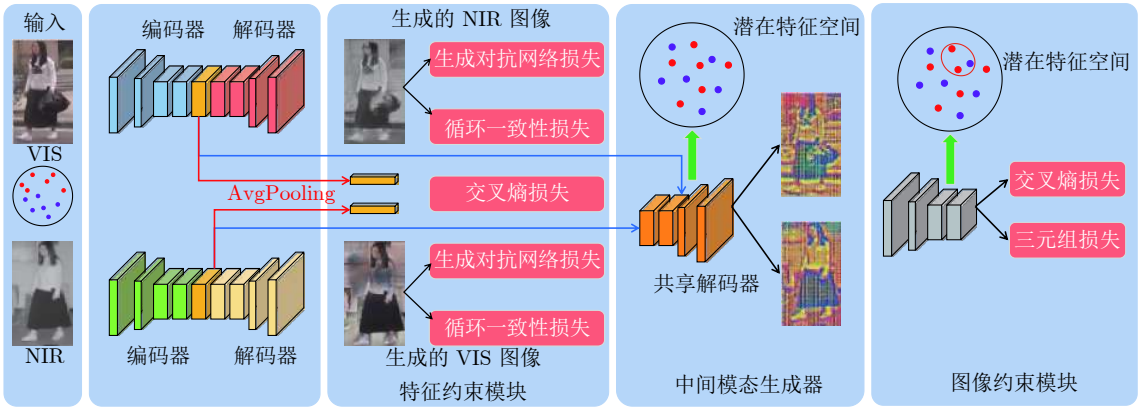

图 1　本文方法的总体框架

Fig. 1　The overall framework of this method

是一个独立的模块, 该部分较容易地嵌入到一个设计良好的 ICM 模块中, 并进行端到端的训练. 在这项工作中, 本文采用基于 ResNet-50 网络[15]训练所提出的 MMG.

本文所提出的基于图像和特征联合约束的跨模态行人重识别方法, 在图像层面通过行人重识别约束模块 (ICM) 对生成对抗网络的中间模态生成器模块 (MMG) 所生成的中间模态图像进行约束, 在特征层面通过特征约束模块 (FCM) 对生成对抗网络的中间模态生成器模块 (MMG) 的编码器进行特征层面的约束.

1.2　中间模态生成器

尽管利用 GAN 从近红外和可见光图像相互迁移在行人重识别的性能上取得了一定的进展, 但由于潜在传输空间巨大, 瓶颈明显. 因此, 本文试图在普通 GAN 的基础上, 在近红外和可见光之间的迁移空间中找到一个潜在的中间模态. 受启发于基于 CycleGAN[16] 的生成器的结构可分为两部分: 编码器和解码器. 根据这一结构, 本文在近红外和可见光的生成器中添加了另一个共享的解码器, 以从近红外和可见光图像中得到一种潜在的中间模态图像.

在本文中, 本文定义 x_{vis} 表示来自于可见光模态 X_{vis} 的输入图像, x_{nir} 表示来自于近红外模态 X_{nir} 的输入图像, 本文采用了 CycleGAN[12] 的近红外−可见光循环一致性结构来建立本文的中间模态生成器 MMG 模块, 在本文方法中, MMG 模块包括近红外和可见光两个模态生成器用以促进中间模态的生成, 其中生成器 G_{vis} 和判别器 D_{vis} 用来从近红外模态生成和判别可见光模态的图像, 生成器 G_{nir} 和判别器 D_{nir} 用来从可见光模态生成和判别近红外模态图像. 因此, 本文的 MMG 模块可以通过以下对抗损失来进行训练:

$$L_{GAN}^{vis} = E_{x_{vis}\sim X_{vis}}[\ln D_{vis}(x_{vis})]+ E_{x_{nir}\sim X_{nir}}[\ln(1-D_{vis}(G_{vis}(x_{nir})))] \quad (1)$$

$$L_{GAN}^{nir} = E_{x_{nir}\sim X_{nir}}[\ln D_{nir}(x_{nir})]+ E_{x_{vis}\sim X_{vis}}[\ln(1-D_{nir}(G_{nir}(x_{vis})))] \quad (2)$$

其中, 判别器 D_{vis} 和 D_{nir} 的作用是通过最大化上述等式来区分生成图像和真实目标图像, 生成器 G_{vis} 和 G_{nir} 的作用是通过最小化上述等式来生成更加真实的图像.

此外, 受到 CycleGAN[12] 循环一致性损失的启发, 使得生成器 G_{vis} (G_{nir}) 所生成的可见光 (近红外) 图像可以被生成器 G_{nir} (G_{vis}) 还原为原始近红外 (可见光) 图像, 本文通过 L_{cyc} 损失来约束 MMG 网络:

$$L_{cyc}(G_{nir},G_{vis}) = E_{x_{ir}\sim X_{ir}}[||G_{nir}(G_{vis}(x_{nir}))-x_{nir}||]+ E_{x_{vis}\sim X_{vis}}[||G_{vis}(G_{nir}(x_{vis}))-x_{vis}||] \quad (3)$$

通过在 MMG 模块中使用可见光−近红外生成器来促进中间模态图像的生成, 进一步缓解了模态间的差异性.

1.3　特征约束模块

虽然循环一致性损失和对抗性损失有助于图像在两种模态之间进行迁移, 但在迁移过程中保持行人身份一致性也是必不可少的. 以前的工作是大都利用基于 Re-ID 骨干网的损失约束生成对抗网络使其产生行人身份的一致性. 目前最先进的 Re-ID 方法是基于 ResNet-50 网络[15] 的, 它在特征提取方面显示出强大的能力. 尽管这种能力有助于解决许多计算机视觉任务, 但在可见光−近红外识别问题中, 需要一种更强的损失对行人身份进行约束. 受

TP-GAN[17] 的启发, 本文提出了一种特征约束模块(FCM), 该模块在编码器之后采用身份损失来在图像生成阶段进行特征级约束. 如图 1 所示, FCM 模块通过具有全连接层的平均池化来构造, 通过交叉熵损失进行约束, 该损失公式为:

$$L_{jcm} = -\frac{1}{N}\sum_{i=1}^{N}\ln\left(p_{jcm}\left(e_{mid}\right)\right) \tag{4}$$

其中, N 表示输入网络的一个批次的图像的数量, $p(e_{mid})$ 表示输入编码器特征 e_{mid} 的概率分布.

综上, 本文所提出的 MMG 模块的损失函数表示如下:

$$L_{MMG} = \lambda_1 L_{cyc} + \lambda_2 L_{GAN} + \lambda_3 L_{jcm} \tag{5}$$

其中, 遵循最初的 CycleGAN 参数配置, 本文设定 λ_1 和 λ_2 分别为 10 和 1, 设定 λ_3 为 0.5.

1.4 图像约束模块

由于上文中 MMG 模块已经将可见光和红外图像迁移到了潜在的中间模态, 因此跨模态的行人重识别问题已经转为为单模态的识别问题. 由于基于 ResNet-50 网络[15] 的可见光下 Re-ID 方法在该领域取得了很大进展, 这里本文采用了相同的设定, 以交叉熵损失和三元组损失为约束损失的 ResNet-50 来完成最后一步:

$$L_{ide} = -\frac{1}{N}\sum_{j=1}^{N}\ln(p_{ide}(x_{mid})) \tag{6}$$

$$L_{tri} = \sum_{f_m^a, f_m^p + f_m^n}\left[D\left(f_m^a, f_m^p\right) - D\left(f_m^p, f_m^n\right) + \xi\right] \tag{7}$$

其中, 对于交叉熵损失, N 是输入网络的每一批次的图像数量, $p(x_{mid})$ 是中间模态图像的概率分布; 对于三元组损失, f_m^a 和 f_m^p 表示来自于同一正样本对的行人的中间模态的特征, f_m^a 和 f_m^n 表示来自于负样本对的不同行人的中间模态的特征. D 表示对特征向量计算其欧氏距离, ξ 表示提前设定好的阈值, $[z]_+ = \max(z, 0)$.

综上所述, 本文所提出的模型的总体损失表示如下:

$$L = l_{MMG} + \alpha_1 L_{ide} + \alpha_2 L_{tri} \tag{8}$$

本文按照经验设定 α_1 和 α_2 分别为 1 和 1.

2 数据集

2.1 Parking-01 数据集

2.1.1 Parking-01 数据集介绍

如图 2 所示, 本数据集采集于冬季下午某地一个路口拐弯处的 9 个摄像机下. 共 103 个行人的 2008 张图像, 其中可见光图像为 1 409 张, 近红外图像为 599 张. 与现有 SYSU-MM01 数据集[5] 相比, 本数据集主要有以下特点:

图 2 数据集图像示例

Fig. 2 Example of dataset images

1) 由于受到光照等环境因素的影响, 这 9 个摄像机并非同一时间全部转换模态, 而是有一定的时间差, 这个时间差会对行人重识别的应用产生较大影响, 本数据集是第一个基于这个时间差构建的跨模态行人重识别数据集.

2) 数据集更能反映出现实世界中跨模态行人数据本身分布的特点, 对研究现实场景下的跨模态行人重识别问题具有重要意义;

3) 基线方法在此数据集上的效果只有 15.5 %, 表明了本数据集具有较大的挑战性, 因此本数据集具有一定的研究意义和学术价值.

这里以图像 “0001_c2_400_nir.jpg” 为例来介绍本数据集图像的命名规则, “0001” 表示 ID, “c2”表示该图像来自于第二个摄像机下, “400” 表示这段视频以 1 s 时间采集时的帧率, “nir” 表示该图像为近红外图像, 如果是 “vis” 则表示该图像为可见光图像. 本文将该数据集命名为 Parking-01 数据集.

2.1.2 Parking-01 数据集评估协议

Parking-01 数据集仅用于网络模型的测试评估, 不用于训练网络. 考虑到本数据集采集的特殊性, 在测试过程中包括两种测试模式, 一种为可见光检索近红外模态, 另一种为近红外检索可见光模态. 在可见光检索近红外模式时, 以可见光图像为 query 库, 近红外图像中每个 ID 在每个摄像机下任取一张图像为 gallery 库. 在近红外检索可见光图像时, 以近红外图像为 query 库, 可见光图像中每个 ID 在每个摄像机下任取一张图像为 gallery 库. 数据集最终测试结果为测试 10 次求其平均值.

2.2 SYSU-MM01 数据集介绍

本研究在 SYSU-MM01 数据集[9] 上做了大量的评估实验. SYSU-MM01 数据集是跨模态行人重识别的公认权威数据集. 包含由 4 个可见光相机拍摄的 287 628 张可见光图像和 2 个近红外相机拍摄的 15 792 张近红外图像, 一共有 491 个行人. SYSU-MM01 数据集分为训练集和测试集, 分别包含 395 个和 96 个行人. 根据其标准评估协议, 数据集包括 all-search 模式和 indoor-search 检索模式. 对于 all-search 模式, 可见光相机 1、2、4 和 5 用于 gallery 集, 红外相机 3 和 6 用于 query 集. 对于 indoor-search 模式, 可见光摄像机 1 和 2 (不包括室外摄像机 4 和 5) 用于 gallery 集, 红外摄像机 3 和 6 用于 query集. 对于这两种模式, 本文都采用 single-shot 和 multi-shot 设置进行测试.

本文提出了在 SYSU-MM01 数据集进行训练后, 在 Parking-01 数据集上按照 2.1.2 的数据集评估协议进行评估.

3 实验测试与结果

3.1 实验实施细节

评估协议. 以标准累积匹配特性 (CMC) 曲线和平均精度 (mAP) 作为性能评价指标. 在测试阶段, 可见光相机的样本用于 gallery 集, 近红外相机的样本用于 query 集.

实验细节. 本方法中的训练图像大小首先被设定为 128×128, 并且使用了水平翻转的图像增强方式, 之后再输入 Re-ID 网络时图像大小调整为 256×128. 在本文中 Re-ID 模块所使用的 ResNet-50 网络模型是经过 ImageNet[18] 预训练的, 然后在 SYSU-MM01 数据集进行微调. 本方法输入网络的图像批次为 32. MMG 模块的学习率为 0.0002, FCM 和 Re-ID 模块的学习率设定为 0.1, 经过 100 次的训练后, 该学习率被衰减为 0.01, 模型一共训练 150 次. MMG 模块的优化器采用的是 Adam, 其数值设定为 (0.5, 0.999). FCM 和 ICM 模块的优化器是 SGD. 三元组损失的 P 和 K 值分别设定为 8 和 4.

3.2 与现有方法的比较

本文提出的方法与 SYSU-MM01 数据集上的 13 种最新方法进行了比较, 包括 HOG[19]、LOMO[20]、Two-Stream[9]、One-Stream[9]、Zero-Padding[9]、BCTR[10]、BDTR[10]、D-HSME[21]、MSR[22]、cmGAN[12]、ResNet-50、CMGN[23]、D2RL[13] 和 AlignGAN[14], 其中 ResNet-50* 为本文所测出的结果. 为了公平比较, 上述方法分别在 SYSU-MM01 数据集上的 all-search single-shot、indoor-search single-shot、all-search multi-shot 和 indoor-search multi-shot 四种模式下进行实验, 实验结果分别如下表 1 ~ 4 所示, 表中 R1、R10、R20 分别代表 Rank-1、Rank-10、Rank-20. 其中, "*" 表示本文测出的结果.

表 1　SYSU-MM01 数据集 all-search single-shot 模式实验结果

Table 1　Experimental results in all-search single-shot mode on SYSU-MM01 dataset

方法	All-Search Single-shot			
	R1	R10	R20	mAP
HOG[19]	2.8	18.3	32.0	4.2
LOMO[20]	3.6	23.2	37.3	4.5
Two-Stream[9]	11.7	48.0	65.5	12.9
One-Stream[9]	12.1	49.7	66.8	13.7
Zero-Padding[9]	14.8	52.2	71.4	16.0
BCTR[10]	16.2	54.9	71.5	19.2
BDTR[10]	17.1	55.5	72.0	19.7
D-HSME[21]	20.7	62.8	78.0	23.2
MSR[22]	23.2	51.2	61.7	22.5
ResNet-50*	28.1	64.6	77.4	28.6
cmGAN[12]	27.0	67.5	80.6	27.8
CMGN[23]	27.2	68.2	81.8	27.9
D2RL[13]	28.9	70.6	82.4	29.2
本文方法	**33.1**	**73.9**	**83.7**	**32.9**

表 2　SYSU-MM01 数据集 all-search multi-shot 模式实验结果

Table 2　Experimental results in all-search multi-shot mode on SYSU-MM01 dataset

方法	All-Search Multi-shot			
	R1	R10	R20	mAP
HOG[19]	3.8	22.8	37.7	2.16
LOMO[20]	4.70	28.3	43.1	2.28
Two-Stream[9]	16.4	58.4	74.5	8.03
One-Stream[9]	16.3	58.2	75.1	8.59
Zero-Padding[9]	19.2	61.4	78.5	10.9
ResNet-50*	30.0	66.2	75.7	24.6
cmGAN[12]	31.5	**72.7**	**85.0**	22.3
本文方法	**33.4**	70.0	78.7	**27.0**

从表 1 可以看出, 本文方法在 SYSU-MM01 数据集上 all-search single-shot 模式下达到了目前最好的效果, 在 Rank-1、Rank-10、Rank-20 以及 mAP 上分别超过最排在第二位的 D2RL 4.2 %、3.3 %、1.3 %、3.7 %.

表 3 SYSU-MM01 数据集 indoor-search single-shot 模式实验结果

Table 3 Experimental results in indoor-search single-shot mode on SYSU-MM01 dataset

方法	indoor-search single-shot			
	R1	R10	R20	mAP
HOG[19]	3.2	24.7	44.6	7.25
LOMO[20]	5.8	34.4	54.9	10.2
Two-Stream[9]	15.6	61.2	81.1	21.2
One-Stream[9]	17.0	63.6	82.1	23.0
Zero-Padding[9]	20.6	68.4	85.8	27.0
CMGN[23]	30.4	74.2	87.5	40.6
ResNet-50*	31.0	78.2	**90.3**	41.9
cmGAN[12]	**31.7**	77.2	89.2	**42.2**
本文方法	31.1	**79.5**	89.1	41.3

表 4 SYSU-MM01 数据集 indoor-search multi-shot 模式实验结果

Table 4 Experimental results in indoor-search multi-shot mode on SYSU-MM01 dataset

方法	indoor-search multi-shot			
	R1	R10	R20	mAP
HOG[19]	4.8	29.1	49.4	3.51
LOMO[20]	7.4	40.4	60.4	5.64
Two-Stream[9]	22.5	72.3	88.7	14.0
One-Stream[9]	22.7	71.8	87.9	15.1
Zero-Padding[9]	24.5	75.9	91.4	18.7
ResNet-50*	29.9	66.2	75.7	24.5
cmGAN[12]	37.0	**80.9**	**92.3**	32.8
本文方法	**37.2**	76.0	83.8	**33.8**

从表 2 可以看出, 本文方法在 SYSU-MM01 数据集上 all-search multi-shot 模式下达到了较好的效果, 其中在 Rank-1 和 mAP 上达到了最好的效果.

从表 3 可以看出, 本文方法在 SYSU-MM01 数据集上 indoor-search single-shot 模式下达到了较好的效果, 其中在 Rank10 上达到了最好的效果.

从表 4 可以看出, 本文方法在 SYSU-MM01 数据集上 indoor-search multi-shot 模式下达到了较好的效果.

如表 3 和表 4 所示, 在 SYSY-MM01 数据集的测试中本文发现, 在 indoor-search 两种模式下本方法略低于现有最好方法, 这主要是由于对于中间模态的约束不佳, 而造成生成结果反而比原始的图像更加难以检索. 这在一定程度上是基于生成对抗网络的跨模态行人重识别方法所存在的普遍问题. 但是相比于其他方法, 本文所提出的方法仍然在 Rank1 和 mAP 上处于领先的地位.

表 5 和表 6 为本文所提出的方法在所构建的 Parking-01 数据集上的实验效果, 其中表 5 为近红外检索可见光模式的实验结果, 表 6 位可见光检索近红外模式的实验结果.

表 5 近红外检索可见光模式的实验结果

Table 5 Experimental results of near infrared retrieval visible mode

方法	近红外–>可见光			
	R1	R10	R20	mAP
ResNet-50*	15.5	39.7	51.9	19.3
本文方法	**25.9**	**53.8**	**62.8**	**29.7**

表 6 可见光检索近红外模式的实验结果

Table 6 Experimental results of visible retrieval near infrared mode

方法	可见光–>近红外			
	R1	R10	R20	mAP
ResNet-50*	20.2	45.6	50.0	14.7
本文方法	**31.6**	**48.2**	**56.1**	**19.7**

从上述两个表格可以看出一下问题: 1) Rank-1 和 mAP 的效果较低, 显示出跨模态行人重识别在本文所构建的 Parking-01 数据集上具有很大的挑战, 因此所构建的数据集有着巨大的研究意义和研究价值; 2) 本文所提方法在近红外检索可见光模式下, 比 ResNet-50 网络[11] 的基准线分别在 Rank-1、Rank-10、Rank-20 以及 mAP 上分别高出了 10.4 %、14.1 %、10.9 %、10.4 %; 在可见光检索近红外模式下, 比 ResNet-50 的基准线分别在 Rank-1、Rank-10、Rank-20 以及 mAP 上分别高出了 11.4 %、2.6 %、6.1 %、5.4 %, 这个也证实了本文所提出方法的有效性.

3.3 算法分析

3.3.1 不同模态转换性能分析

为了与近红外转可见光模式、可见光转近红外模式两种方法进行比较, 本文进行将 ResNet-50、转近红外模式、转可见光模式与本文所提出的方法在 SYSU-MM01 数据集上进行了实验比较, 实验结果如表 7 所示. 从表 7 的实验结果可以看出, 本文所提出的中间模态转换比转到近红外模式或者转到可见光模式在 Rank-1 上分别高出了 2.3 %、3.5 %, 显示了本文所提方法有着较大的性能优势.

3.3.2 算法时间复杂度

本文的方法实施框架为 Pytorch, 算法在两个 GeForce 1080Ti GPU 上训练时间约为 40 个小时.

表 7　不同模态转换的实验结果
Table 7　Experimental results of different mode conversion

方法	R1	R10	R20	mAP
ResNet-50*	28.1	64.6	77.4	28.6
转为可见光	29.6	69.8	80.5	30.7
转为近红外	30.8	71.5	83.2	31.2
本文提出的方法	**33.1**	**73.9**	**83.7**	**32.9**

算法模型一共训练 150 个 epoch, 一个 epoch 时间为 16 分钟, 一个 epoch 训练的图像张数为 32 451 张图像, 当 epoch = 1 时, 随机选取 100 张图像进行测试, 单张图像平均测试时间约为 1.69 秒.

3.4　循环一致性损失分析

为了检测算法中循环一致性损失的效果, 本文在去掉了循环一致性损失并对网络进行训练与测试, 仅用和相同的判别网络设置去约束近红外和可见光图片送入编码器后解码并输入到图像约束模块, 并且与本文所提出的中间模态方法进行比较. 从表 8 的实验结果可以看出, 本文所提出的中间模态转换比没有循环一致性损失在 Rank-1 上高出了 3.5 %, 显示了较大的优势.

表 8　有无循环一致性损失的实验结果
Table 8　Experimental results with or without loss of cycle consistency

方法	R1	R10	R20	mAP
无循环一致性	29.6	67.1	78.3	31.1
有循环一致性	**33.1**	**73.9**	**83.7**	**32.9**

3.5　中间模态图像可视化及分析

为了更好地观察实验中间模态的结果, 本文将中间模式生成器生成的图像以图 3 中的图像形式可视化. 从图中可以看出, 本文提出的方法可以通过潜在的特征空间将可见光和红外模态转换为一种模态, 减少了模态之间的差异, 提高了行人重识别的效果.

4　结论

本文针对近红外和可见光之间数据分布存在差异性的问题, 不同于以往使用生成对抗网络进行单向转换为近红外或者可见光的方法, 提出了通过生成对抗网络寻找一种在其相互转换过程中潜在的中间模态, 以提升此种模态下的行人重识别效果. 本文提出的特征约束模型和行人重识别约束模型对生成对抗网络的中间模态生成器进行约束, 进一步压缩了可见光和近红外图像及特征间的转换空间. 此外, 考虑到跨模态行人重识别数据集的稀缺性, 本文还构建了一个基于跨模态的行人重识别数据集, 为进一步开展此方向的研究提供了有效的评估策略和依据.

图 3　中间模态生成器所生成的中间模态图像
Fig. 3　Middle modality image generated by middle modality generator

References

1　Ye Yu, Wang Zheng, Liang Chao, Han Zhen, Chen Jun, Hu Rui-Min. A survey on multi-source person re-identification. *Acta Automatica Sinica*, 2020, **46**(9): 1869−1884
(叶钰, 王正, 梁超, 韩镇, 陈军, 胡瑞敏. 多源数据行人重识别研究综述. 自动化学报, 2020, **46**(9): 1869−1884)

2　Luo Hao, Jiang Wei, Fan Xing, Zhang Si-Peng. A survey on deep learning based person re-identification. *Acta Automatica Sinica*, 2019, **45**(11): 2032−2049
(罗浩, 姜伟, 范星, 张思朋. 基于深度学习的行人重识别研究进展. 自动化学报, 2019, **45**(11): 2032−2049)

3　Zhou Yong, Wang Han-Zheng, Zhao Jia-Qi, Chen Ying, Yao Rui, Chen Si-Lin. Interpretable attention part model for person re-identification. *Acta Automatica Sinica*, 2020, **41**: 1−13
(周勇, 王瀚正, 赵佳琦, 陈莹, 姚睿, 陈思霖. 基于可解释注意力部件模型的行人重识别方法. 自动化学报, 2020, **41**: 1−13)

4　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)

5　Zhao H, Tian M, Sun S, et al. Spindle net: Person re-identification with human body region guided feature decomposition and fusion. In: Proceedings of the IEEE CVPR. Hawaii, USA: IEEE, 2017. 1077−1085

6　Sun Y, Zheng L, Yang Y, et al. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In: Proceedings of the ECCV. Munich, Germany: Springer, 2018. 480−496

7　Hermans A, Beyer L, Leibe B. In defense of the triplet loss for person re-identification. arXiv preprint arXiv: 1703.07737, 2017

8　Wei L, Zhang S, Gao W, et al. Person transfer gan to bridge domain gap for person re-identification. In: Proceedings of the IEEE CVPR. Salt Lake City, UT, USA: IEEE, 2018. 79−88

9　Wu A, Zheng W S, Yu H X, et al. RGB-infrared cross-modality person re-identification. In: Proceedings of the IEEE ICCV. Honolulu, USA: IEEE, 2017. 5380−5389

10 Ye M, Wang Z, Lan X, et al. visible Thermal person re-identification via dual-constrained top-ranking. In: Proceeding of IJCAI. Stockholm, Sweden, 2018, 1: 2

11 Ye M, Lan X, Li J, et al. Hierarchical discriminative learning for visible thermal person re-identification. In: Proceeding of AAAI. Louisiana, USA: IEEE, 2018. 32(1)

12 Dai P, Ji R, Wang H, et al. Cross-Modality person re-identification with generative adversarial training. In: Proceeding of IJCAI. Stockholm, Sweden, 2018. 1: 2

13 Wang Z, Wang Z, Zheng Y, et al. Learning to reduce dual-level discrepancy for infrared-visible person re-identification. In: Proceedings of the IEEE CVPR. California, USA: IEEE, 2019. 618−626

14 Wang G, Zhang T, Cheng J, et al. Rgb-infrared cross-modality person re-identification via joint pixel and feature alignment. In: Proceedings of the IEEE ICCV. Seoul, Korea: IEEE, 2019. 3623−3632

15 He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE CVPR. Las Vegas, USA: IEEE, 2016. 770−778

16 Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE ICCV. Honolulu, USA: IEEE, 2017. 2223−2232

17 Huang R, Zhang S, Li T, et al. Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis. In: Proceedings of the IEEE ICCV. Honolulu, USA: IEEE, 2017. 2439−2448

18 Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database. In: Proceedings of the IEEE CVPR, Miami, FL, USA: IEEE, 2009. 248−255

19 Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Proceedings of the IEEE CVPR. San Diego, CA, USA: IEEE, 2005: 886−893

20 Liao S, Hu Y, Zhu X, et al. Person re-identification by local maximal occurrence representation and metric learning. In: Proceedings of the IEEE CVPR. Boston, USA: IEEE, 2015: 2197−2206

21 Hao Y, Wang N, Li J, et al. HSME: Hypersphere manifold embedding for visible thermal person re-identification. In: Proceedings of the AAAI. Hawaii, USA: IEEE, 2019, 33: 8385−392

22 Kang J K, Hoang T M, Park K R. Person re-identification between visible and thermal camera images based on deep residual cnn using single input. *IEEE Access*, 2019: 1−1

23 B J J A, B K J, B M Q A, et al. A cross-modal multi-granularity attention network for rgb-ir person re-identification. *Neurocomputing*, 2020, **406**: 59−67

张玉康 华中师范大学国家数字化学习工程技术研究中心硕士研究生. 主要研究方向为行人重识别, 生成对抗网络.
E-mail: zhangyk@mails.ccnu.edu.cn
(**ZHANG Yu-Kang** Master student at the National Engineering Research Center for E-Learning, Central China Normal University. His research interest covers person re-identification and generative adversarial networks.)

谭 磊 华中师范大学国家数字化学习工程技术研究中心硕士研究生. 主要研究方向为模式识别和计算机视觉.
E-mail: lei.tan@mails.ccnu.edu.cn
(**TAN Lei** Master student at the National Engineering Research Center for E-Learning, Central China Normal University. His research interest covers pattern recognition and computer vision.)

陈靓影 华中师范大学国家数字化学习工程技术研究中心教授. 2001 年获得南洋理工计算机科学与工程系博士学位. 主要研究方向为图像处理, 计算机视觉, 模式识别, 多媒体应用. 本文通信作者.
E-mail: chenjy@mail.ccnu.edu.cn
(**CHEN Jing-Ying** Professor at the National Engineering Research Center for E-Learning, Central China Normal University. She received her Ph. D. degree from the School of Computer Engineering, Nanyang Technological University, Singapore in 2001. Her research interest covers image processing, computer vision, pattern recognition, and multimedia applications. Corresponding author of this paper.)