

跨模态身份不一致的交叉分类素描行人重识别

黄勃淳^{1,2}, 李凡^{1,2}, 汪淑娟^{1,2*}

¹昆明理工大学信息工程与自动化学院, 云南 昆明 650500;

²云南省人工智能重点实验室, 云南 昆明 650500

摘要 素描行人重识别旨在从可见光行人图像库中查找与给定素描行人图像身份相同的图像。虽然已经有较多的跨模态检索算法可以应用于该类问题,但这些算法的背景设定较为单一,没有考虑到训练集中部分身份的行人仅有一个模态样本,即跨模态身份不一致,这极大限制了算法在实际场景下的应用。为此,提出了基于交叉分类的素描行人重识别网络。该网络包括交叉分类和基于距离的身份信息对齐两部分。其中,交叉分类利用单一模态数据训练的分类器引导编码器从另一模态提取到模态不变的信息。而基于距离的身份信息对齐能够将同身份不同模态间的特征距离减小,同时抑制跨模态身份不一致的影响,进而强化了特征的判别性和鲁棒性。为验证跨模态身份不一致时模型的性能表现,基于 Market-1501 数据集生成了新的素描行人重识别数据集 S-Market1501,并在该数据集上将 Rank-1 指标提升了 11.0 个百分点。同时模型在公开数据集 Sketch Re-ID 上 Rank-1 指标达到了 60%,所设计的数据集将开源在“https://github.com/huangdaichui/Sketch_dataset”。

关键词 素描行人重识别; 模态不变; 模态身份不一致; 交叉分类

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP212820

Cross-Classification-Based Sketch Person Re-Identification in Inconsistent Cross-Modal Identity Scenes

Huang Bochun^{1,2}, Li Fan^{1,2}, Wang Shujuan^{1,2*}

¹Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, Yunnan, China;

²Yunnan Key Laboratory of Artificial Intelligence, Kunming 650500, Yunnan, China

Abstract Sketch person re-identification aims to identify images with identities similar to those of sketched person images located in an RGB image gallery. Although several cross-modal retrieval algorithms can be adopted for this purpose, the background settings of such algorithms are relatively simple and fail to consider that certain identities have only one modal sample in the training set, that is, the cross-modal identity is inconsistent. This significantly limits the application of such algorithms in practical scenarios. In this paper, a sketch re-identification network based on cross-classification is proposed. The network consists of two parts: cross-classification and identity information alignment based on distance. Among these, cross-classification guides the encoder to extract modal-invariant information from one modal using constraints of the classifier trained using other modal data. The alignment of identity information based on distance can reduce the feature distance between different modals of the same identity, suppress the influence of cross-modal identity inconsistencies, and strengthen the discrimination and robustness of features. To verify the performance of the re-identification network when the cross-modal identity is inconsistent, a new sketch re-identification dataset is generated based on Market-1501. The Rank-1 is improved by 11.0 percentage points on this dataset. Simultaneously, the model also achieves a Rank-1 of 60% on the public dataset Sketch Re-ID. The dataset used in this study is an open-source dataset available on “https://github.com/huangdaichui/Sketch_dataset”.

Key words sketch person re-identification; modal invariance; cross modal identity inconsistency; cross classification

收稿日期: 2021-10-27; 修回日期: 2021-11-30; 录用日期: 2021-12-21; 网络首发日期: 2022-01-09

基金项目: 国家自然科学基金(61966021)、云南省重大科技专项计划(202002AD080001)、云南省基础研究计划(202101AT070136)

通信作者: *478263823@qq.com

1 引言

素描行人重识别^[1](Sketch re-identification)任务旨在从彩色行人图像库中检索与提供的素描行人图像同身份的图像。作为行人重识别任务的一个重要分支,该任务与社区安防、犯罪预防、走失人口查询等社会需求密切相关,所以一经提出便受到了广泛的关注。传统的行人重识别任务查询图像和图库图像均为彩色图像^[2-3],无需考虑模态间信息的差异。众多研究人员利用表征学习的 Re-ID 方法^[4-5]在封闭世界下产生了优秀的性能效果。然而随着行人重识别任务与实际需求接轨,该任务需要对跨域^[6-7]或跨模态数据^[1,8]进行检索。其中,与刑侦密切相关的素描行人重识别问题受到了广泛关注。

素描行人重识别模型需克服模态间信息差异并进行精确的检索,这对素描行人重识别网络提出了更高的要求。未经特殊设计网络难以在不同模态的行人图像中提取与身份相关且不随模态变化的信息(即模态不变信息)。尽管现阶段有大量的针对跨模态行人重识别任务的研究^[9-11],也存在一定数量的针对素描行人重识别的研究^[1,8],但这些方法仍面临着诸多问题。一方面,虽然相关的研究使用生成对抗网络来生成行人另一模态的图像并同原数据一起训练^[9-12],以此约束网络提取模态不变信息,缓解模态间数据差异导致的网络性能下降。但该类方法难以进行端到端训练,且引入判别器等参数量较大的神经网络,提高了训练成本。而基于解纠缠的思想所设计的跨模态网络^[11-13],虽然可以进行端到端的训练,但依旧需要进行图像级别的生成和判断,这依旧有着不小的计算成本。另一方面,现阶段的素描行人重识别研究所使用的数据集中所有身份的行人均同时具有 2 个模态的训练集样本。在这一条件下,以往的方法利用了三元组损失或其他损失函数来约束网络能够拉近同身份样本之间特征的距离,进而提取模态不变的特征,这一点和实际应用场景相差较大。在训练集制作时,同时获取同身份的素描和可见光图像难度较高。

为解决以上问题,本文提出了一种交叉分类素描行人重识别网络。该网络包含两部分:交叉分类和基于距离的身份信息对齐。首先,交叉分类设计了 2 个分类器,在第一阶段分别对 2 种模态的图像特征进行分类,约束网络获得编码行人身份相关特征的能力。而在第二阶段将交换 2 个分类器的分类任务,并固定分类器的参数,约束编码器能够提取模态不变的特征。其次,基于距离的身份对齐考虑了消极样本之间的关系,约束网络能够缓解模态间信息差异和跨模态身份不一致的影响,提取模态不变的信息。利用 Market-1501^[14]生成了一个新素描图像数据集 S-Market1501,该数据集可以模拟真实的工程环境,更好地验证模型在贴合实际场景下的性能表现。同时,本文设计的模

型在 S-Market1501 上的实验结果相比于 Baseline 的首位命中率(Rank-1)和平均正确率(mAP)分别提升了 11.1% 和 9.1%,并在 Sketch Re-ID 数据集上取得了 Rank-1 高达 60% 的出色效果。

2 相关工作

现阶段针对素描行人重识别文献较少。但由于素描行人重识别是跨模态行人重识别的子任务,且存在的问题与跨模态行人重识别较为相似,同时,素描行人重识别也是素描图像检索的子任务,所以本文将从模态不变特征提取的行人重识别与素描图像检索两方面展开相关工作介绍。

在本节,将对引言中阐述的问题进行图文并茂的说明。如图 1 所示,第一行为可见光模态的样本,第二行为素描模态的样本。在理想情况下,如图 1(a)框中样本,训练集应当包含每个行人的 2 种模态的样本。但实际上,尽管美术学院等机构可以提供大量素描的作品,却无法保证每一幅素描肖像都有便于获取的匹配可见光图像。这说明数据集存在着跨模态身份不一致现象,如图 1(b)所示。在跨模态身份不一致的场景下,跨模态任务中常用的三元组损失无法通过拉近同身份不同模态样本特征之间距离的方式来提升网络提取模态不变信息的能力。



图 1 跨模态身份不一致现象。(a)理想场景;(b)真实场景
Fig. 1 Cross modal identity inconsistency. (a) Perfect scene; (b) real scene

2.1 模态不变特征

约束网络提取模态不变特征是缓解模态信息差异对网络性能影响的重要方法。而对抗学习^[12]是引导网络提取模态不变特征的常用技术路线,在跨模态行人

重识别和域自适应行人重识别网络中都有大量的应用。Wang 等^[9]利用生成对抗网络生成了对应模态下的图像,并利用该图像与原图像一同训练。Zhang 等^[10]利用生成网络转化得到的另一模态的图像作引导,并使用教师网络帮助学生网络更快地获得提取红外模态图像身份信息的能力。虽然生成的图像参与训练可以帮助网络提取模态不变的信息,但是训练一个生成对抗网络需要较大的训练成本,并且训练得到的网络通常只能进行两种特定模态之间的信息转化。生成网络在任务涉及的两种模态发生变化时,需重新训练。Wang 等^[11]利用解纠缠的方法将图像中信息编码为模态相关和模态无关的信息,并强化模态相关信息的判别性。与之类似,Choi 等^[13]利用了层次编码的方式,缓解了模态相关信息的影响,也缓解了行人姿态变化对特征鲁棒性的影响。在解纠缠方法的基础上,Lu 等^[15]在两种模态间传播信息,以达到降低模态间图像检索难度的目的。这些解纠缠方法试图将图像中蕴含的信息分为身份相关和身份无关特征,但是在测试阶段网络得到的身份无关特征身份没有使用,这造成了算力浪费,也意味着该类方法需要引入至少两个编码器,还有大量的鉴别器,训练成本依旧较大。

与上述需要进行生成对抗的方法不同,本文提出的交叉分类的网络模型仅使用一个编码器,这有效降低了训练成本。同时利用两个参数较少的全连接层分类器进行交叉分类,引导网络获得提取模态不变信息的能力,这也极大地降低了训练成本。

2.2 素描-RGB 图像检索

素描-RGB 图像检索在电商推荐等领域有着广泛应用,存在着大量的研究小商品素描-RGB 图像检索的文献。Yu 等^[16]针对鞋子设计了基于三元组的训练方法。Liu 等^[17]利用语义信息对素描图像进行检索,

强化了模型的泛化性。Lei 等^[18]利用边缘图作为两种模态之间的桥梁,弱化了网络受到模态信息差异的影响。而 Bhunia 等^[19]基于强化学习构建了一种探究以最少笔画的素描图进行检索的模型。Pang 等^[20]利用自监督学习和拼图的手段使网络理解图像中图像块的意义,借此强化了模型的鲁棒性。

尽管素描图像检索有着巨大成功,但是素描-RGB 图像检索的方法直接用于素描行人重识别任务效果不佳。这是因为行人图像的构成相比鞋、茶杯等小商品的图像构成更加复杂。Pang 等^[11]针对素描行人重识别问题设计了基于对抗的素描行人重识别网络,并首次提出了 Sketch Re-ID 素描行人重识别数据集。Gui 等^[8]利用灰度图像辅助网络提取模态不变特征,并加入注意力机制强化了网络对行人特征的提取能力。但文献[1]、文献[8]没有考虑行人素描图像的收集工作相比于小商品难度更大,在训练集中同一身份的行人难以同时具备两种模态的样本。这对模态之间行人身份信息对齐提出了更高的要求。

同时,常用的三元组损失无法通过拉近同身份不同模态样本特征之间距离的方式,来提升网络提取模态不变信息的能力。图 2 中生动地展示了这一点,图中每一个小图形代表着一张图像样本,图 2(a)左侧为可见光图像,右侧为素描模态图像。而同样的图形表示样本属于同一身份。菱形与圆形所代表的两个身份不存在另一模态的样本。由于在批中负样本与锚点极有可能属于不同的模态,二者距离本就很大。此时,即使类内样本距离较大,三元组损失也会为 0,从而失去约束网络的能力。正样本难以被拉近,同时不同模态之间的信息将难以对齐。本文设计的基于距离的身份信息对齐避免了损失函数对跨模态正样本的依赖,能够有效地缓解跨模态身份不一致带来的影响,更好地对齐身份信息,提取模态不变的特征。

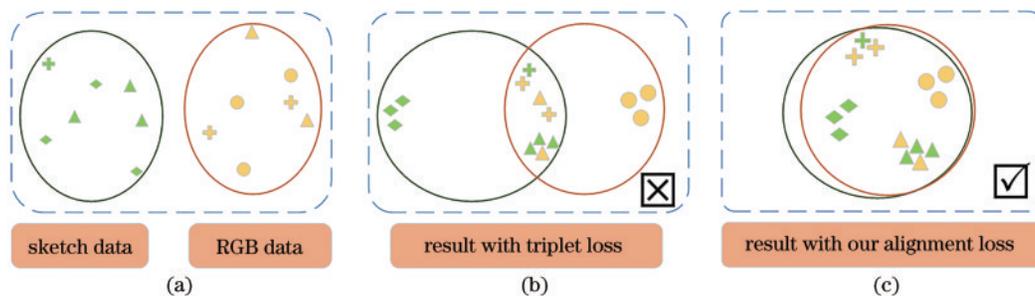


图 2 三元组损失与基于距离的身份信息对齐效果图。(a)原始数据;(b)加入三元组损失的结果;(c)本文对齐损失的结果

Fig. 2 Triplet loss and modal information alignment based on distance. (a) Original data; (b) result with triplet loss; (c) result with our alignment loss

3 跨模态身份不一致数据集

3.1 旧有数据集的问题

Sketch Re-ID^[4]是现阶段唯一的素描行人重识别

数据集。对数据驱动的深度学习方法而言,该数据集的 300 张训练图像过少。同时该数据集没有收集跨模态身份不一致的数据,不能很好地验证模型在现实场景下的表现。为此需要一个更加贴合现实场景的大数

据量数据集。

3.2 设计的新的数据集

受制于数据收集的成本,在已有的数据集上进行转化,以保证图像的数量。在 Market-1501^[14]数据集基

础上进行调整得到 S-Market1501,将数据集中的 query 图像和所有训练集图像利用 OpenCV 进行处理,经过高斯模糊和二值化操作后得到类似素描的图像,图 3 展示了转化前后的图像。



图 3 新构造的数据集样本图像

Fig. 3 Sample images of the newly constructed dataset

该数据集包括 7 个以不同比例将素描图和原图组合的训练集,用以分析不同程度的身份不一致对性能的影响,具体比例及数据量如表 1 所示。在后续

训练中,将分别对比提出的模型在不同程度的跨模态身份不一致的情况下的性能,以证明本文模型有效性。

表 1 S-Market1501 构成
Table 1 Composition of S-Market1501

Training set number	Identity inconsistent sketch image	Identity inconsistent RGB image	Identity consistent image	All identity inconsistent image	Proportion of inconsistent identity
1	0	0	12936	0	0
2	1618	1616	9702	3234	1/4
3	2157	2155	8624	4312	1/3
4	3235	3233	6468	6468	1/2
5	4313	4311	4312	8624	2/3
6	4852	4850	3234	9702	3/4
7	6468	6468	0	12936	1

4 提出的模型

4.1 模型概览

提出的模型如图 4 所示,其中特征提取网络 E (亦称编码器)能够将输入图像编码为特征向量,分类器 C_r 、 C_s 和 C 能够促使编码器提取模态不变的特征,基于距离的身份信息对齐能够在距离层面约束网络,在跨模态身份不一致条件下能够提取模态不变的信息。

定义数据集为 D , x_i^r 和 x_j^s 分别为训练集的第 i 张可见光图像和第 j 张素描图像, y_i 和 y_j 分别为图像 x_i^r 和 x_j^s 的身份标签。利用特征提取器 E 从可见光图像 x_i^r 中提取信息并编码为特征图 $F_i^r \in \mathbb{R}^{c \times 16 \times 8}$, c 为特征图的通道数, 16 与 8 为特征图的高和宽。

$$F_i^r = E(x_i^r) \tag{1}$$

与之类似, E 从素描图像 x_j^s 中提取信息并编码为特征图 $F_j^s \in \mathbb{R}^{c \times 16 \times 8}$ 。利用全局平均池化(GAP),可将

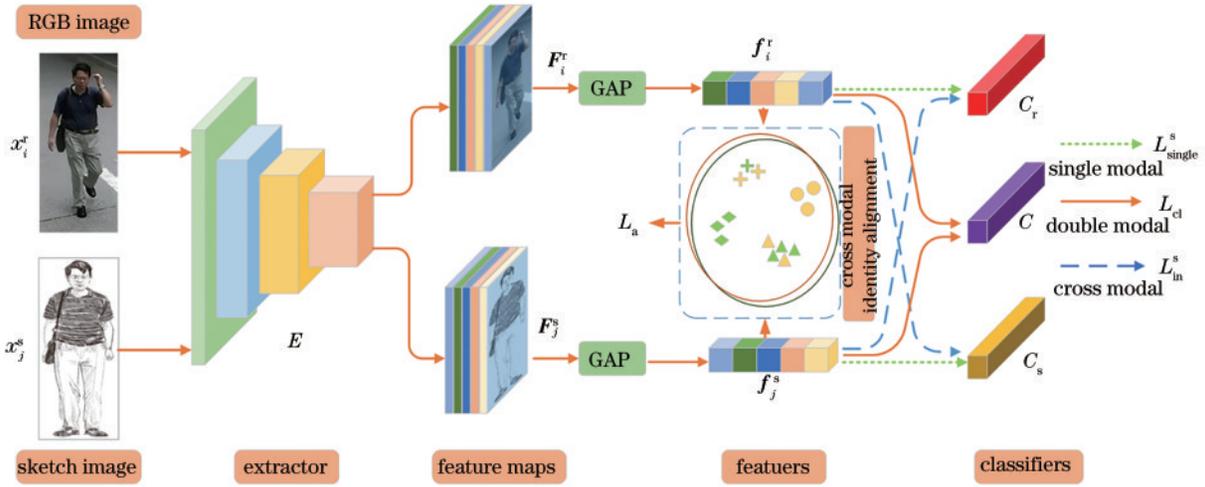


图 4 模型总架构

Fig. 4 General structure of the model

特征图 F_i^r, F_j^s 分别聚合为特征向量 $f_i^r = \text{GAP}(F_i^r)$ 和特征向量 $f_j^s = \text{GAP}(F_j^s)$, 其中 $f_i^r, f_j^s \in \mathbb{R}^c$ 。

4.2 交叉分类

为提取行人模态不变特征, 提出了交叉分类学习策略。与基于对抗方法来提取模态不变信息的网络不同, 这里利用了 2 个分类器之间的交替训练, 引导特征提取器获得提取模态不变信息的能力, 避免了过大的计算量和不能端到端学习的问题。交叉分类分为两步: 第 1 步是单模态特征学习; 第 2 步是模态不变特征学习。在训练中两步将持续迭代直到最大轮次。

4.2.1 单模态特征学习

在单模态特征学习中, 编码器 E 将在分类器 C_r 和 C_s 的引导下获得编码单一模态下身份相关信息的能力。分类器 C_r 和 C_s 将对输入的特征向量对应图像的身份进行预测, 并分别得到预测向量 \hat{y}_i^r 和 \hat{y}_j^s 。该向量的每一维度的数值代表图像 x_i^r 和 x_j^s 是该维度对应身份的概率, 具体公式为

$$\hat{y}_i^r = C_r(f_i^r), \quad (2)$$

$$\hat{y}_j^s = C_s(f_j^s), \quad (3)$$

式中: \hat{y}_i^r, \hat{y}_j^s 分别为分类器 C_r, C_s 对可见图像光 x_i^r 特征向量 f_i^r 和素描图 x_j^s 特征向量 f_j^s 的预测结果。该预测过程在图 4 中以单模态线 (single modal) 标出, 利用交叉熵损失对编码得到的特征向量进行约束, 进而迫使编码器编码具有身份判别性更强的特征。利用标签和预测向量进行损失计算, 可得到单模态下的分类损失。

$$L_{\text{single}}^s = \sum_{i=1}^R \text{CE}(y_i, \hat{y}_i^r) + \sum_{j=1}^S \text{CE}(y_j, \hat{y}_j^s), \quad (4)$$

式中: $\text{CE}(\cdot, \cdot)$ 为交叉熵损失函数。 R 和 S 分别为训练集中具有同身份素描样本的可见光样本总数和具有同身份可见光样本的素描样本总数。通过最小化式 (4) 能够优化编码器 E 和分类器 C_r, C_s 参数, 强化身份相关信息的编码能力。

但仅依赖 2 个分类器并不能约束网络编码模态不变的特征, 这是因为 2 个分类器在损失函数的约束下, 仅能够提取单一模态下的身份相关信息, 却不能消除模态信息的干扰。特别是在跨模态身份不一致的情况下, 网络更难以提取不随模态变化的身份相关信息。因此引入第 2 步的模态不变特征学习。

4.2.2 模态不变特征学习

尽管第一阶段编码器不能够消除模态信息干扰, 但是第一阶段分类器 C_r, C_s 获得了区分特定模态样本的能力。如果素描模态对应的分类器 C_s 能够将可见光图像 x_i^r 经网络 E 编码得到的特征 f_i^r 进行正确分类, 意味着编码器 E 具有了编码模态不变特征的能力, 反之亦然。

为获得此编码能力, C_s 的参数将被冻结, 并将可见光图像特征 f_i^r 送入分类器 C_s 输出预测结果 \tilde{y}_i^r , 具体为

$$\tilde{y}_i^r = C_s(f_i^r). \quad (5)$$

类似地, 素描图像 x_j^s 对应的特征 f_j^s 将被送入分类器 C_r 进行分类, 并得到分类预测结果 \tilde{y}_j^s :

$$\tilde{y}_j^s = C_r(f_j^s). \quad (6)$$

该交叉预测流程在图 4 中以交叉模态线条 (Cross modal) 标出。与式 (4) 类似, 利用交叉熵损失优化编码器网络 E 使其具有提取模态不变特征的能力。

$$L_{\text{in}}^s = \sum_{i=1}^R \text{CE}(y_i, \tilde{y}_i^r) + \sum_{j=1}^S \text{CE}(y_j, \tilde{y}_j^s). \quad (7)$$

在 2 个分类器的交叉优化下, 本文的网络能以较小的训练成本获得提取模态不变特征的能力。

4.3 基于距离的身份信息对齐

在跨模态身份不一致的场景下, 使用三元组损失在距离层面对编码器进行引导优化, 不利于身份信息对齐, 如图 2 所示。但如果能够约束不同模态之间负样本的距离与同一模态负样本对的距离, 则有助于网络提取模态不变信息。换言之, 无论负样本是何种模

态和何种身份的数据,其与锚点图像之间的距离应当近乎相等。基于此本文设计了一种损失函数,称作跨模态身份对齐损失。该损失函数考虑锚点特征 f_{a_i} 、最难正样本特征 f_{p_i} 、与锚点不同模态的难负样本特征 $f_{n_i}^{\text{diff}}$ 、与锚点同一模态的难负样本 $f_{n_i}^{\text{same}}$ 以及与锚点不同模态的任意负样本 $f_{n_i}^{\text{diff}}$ 之间的关系。

$$L_a = \max(d_{a_i, p_i} - d_{a_i, \text{diff}} + m, 0) + \max(d_{a_i, \text{diff}} - d_{a_i, \text{same}} + m, 0) + \max(d_{n_i, n_i'} - d_{a_i, \text{diff}} + m, 0), \quad (8)$$

式中: d_{a_i, p_i} 为样本的特征向量 f_{a_i} 与 f_{p_i} 之间的欧氏距离; $d_{a_i, \text{diff}}$ 为样本特征向量 f_{a_i} 和 $f_{n_i}^{\text{diff}}$ 的距离; $d_{a_i, \text{same}}$ 为向量 f_{a_i} 与 $f_{n_i}^{\text{same}}$ 之间的距离; $d_{n_i, n_i'}$ 为 $f_{n_i}^{\text{diff}}$ 和 $f_{n_i'}^{\text{diff}}$ 之间的距离; m 为

阈值。

该损失函数的第1项能够推动网络拉近类内样本距离,推远类间样本距离,第2项能够利用2个不同模态的负样本引导网络提取模态不变的信息,而第3项能够利用2个不同身份的负样本促进模态不变特征的提取。不仅如此,该损失函数没有利用任何跨模态的正样本做引导,这意味着本文提出的模型在跨模态身份不一致的情况下更为适用。在实验中,使用不同的损失约束所产生的数据分布如图5所示,本节所提出的损失效果如该图中图(c)所示,其中每个图形代表一个样本,同形状的图形表示同一身份的样本,不同模态的身份特征得以对齐。

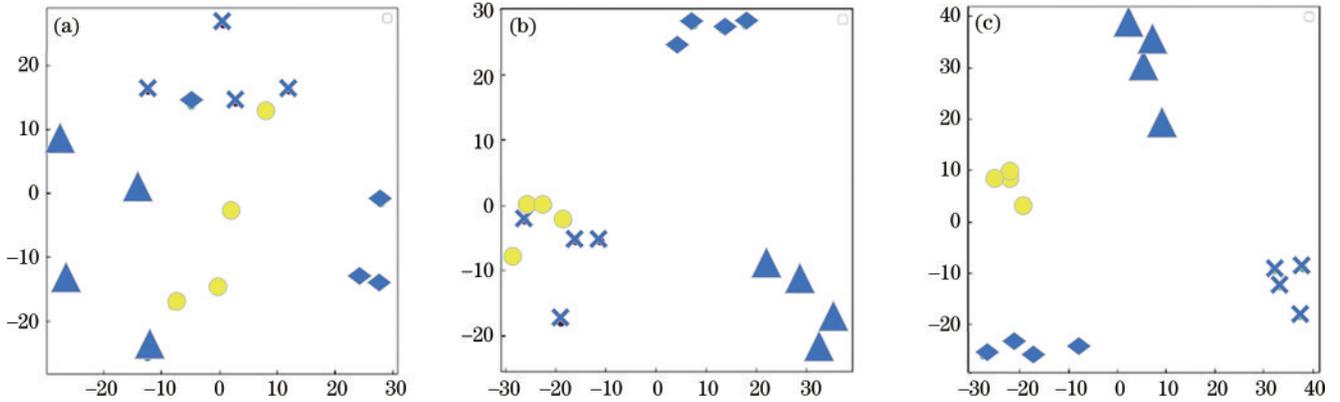


图5 特征向量的降维可视化。(a)无三元组损失结果;(b)三元组损失结果;(c)本文三元组损失结果

Fig. 5 Dimensionality reduction and visualization of eigenvectors. (a) Result without triplet loss; (b) result with triplet loss; (c) result with our alignment loss

4.4 总损失函数

为进一步约束编码器 E 编码模态不变的信息,引入了分类器 C 。如图4双模态线(double modal)所示, C 对两种模态图像 x_i^r 和 x_j^s 的身份进行预测,得到预测结果 \hat{y}_i 和 \hat{y}_j 。使用该预测结果,对分类器 C 以及编码器 E 进行优化。

$$L_{cl} = \sum_{i=1}^{R'} \text{CE}(y_i, \hat{y}_i) + \sum_{j=1}^{S'} \text{CE}(y_j, \hat{y}_j). \quad (9)$$

式中: R' 为训练集中可见光图片总数; S' 为训练集中素描图片总数。定义单模态分类损失为 $L_{cl} + L_{\text{single}} + L_a$,模态不变性损失为 $L_{cl} + L_{\text{in}}^s + L_a$ 。在训练网络时,这两种损失将交替约束网络,并优化编码器 E 和分类器 C_r, C_s 和 C 。

在表2的算法流程下,在每代的Step1内,编码器 E 和3个分类器中的参数将被优化,更好地提取与行人身份相关的特征;在Step2中, C_r, C_s 的参数将被固定,编码器 E 将着重学习模态不变的特征;在达到最终迭代次数后,编码器 E 将获得提取模态不变特征的能力;在测试阶段,图像经过编码器 E 后获得的特征向量将用于最终的检索。

表2 训练算法

Table 2 Training algorithm

Input: Initialize encoder E , classifier C_r, C_s and C , the maximum number of iterations T , dataset D
for $t=1, 2, 3, \dots, T$ do:
Step1:
Calculate f_i^r, f_j^s by Eq (1)
Calculate \hat{y}_i^r, \hat{y}_j^s by Eqs (2) and (3)
Update E, C_r, C_s and C by minimizing Eqs (4), (8) and (9)
Step2:
Extractor features with E
Calculate $\tilde{y}_i^r, \tilde{y}_j^s$ by Eqs (5) and (6)
Update E, C by minimizing Eqs (7), (8) and (9) while freeze C_r, C_s
end for
Output: Encoder E

5 实 验

为验证本文模型的有效性和相较于现有方法的优越性,本节进行了详尽的对比实验和消融实验。

5.1 实现细节

实验以 Bag of tricks^[4]作为 Baseline,以 Resnet50 网络作为特征提取器(编码器),并利用全连接层分类器进行分类。设定首轮学习率为 0,在后续 10 轮逐轮匀速上涨至 0.0002。设定总优化轮次为 115。实验于一张 RTX 2080Ti 显卡,PyTorch 1.6 学习框架下完成。测试中使用余弦相似度计算相似性。

5.2 数据集及评价指标

在广泛使用的数据集 Sketch Re-ID^[1]数据集和新构造的数据集 S-Market1501 上进行实验。Sketch Re-ID 数据集共包含 200 个身份行人的 600 张图像。每个行人包含 2 张 RGB 图像和 1 张素描图像。其中 150 个身份的图像用于训练,剩余 50 个身份的图像用作测试。S-Market1501 数据集由 Market-1501^[14]转化而来。该数据集共包括身份 1501 个,与 Market-1501 数据集的测试协议一致,指定的 750 个身份用于训练,指定的另外 751 个身份用于测试。数据集中的素描图像由可见光图片经高斯模糊与二值化处理得到。数据集中训练集共 7 个,分别验证在不同程度的跨模态身份不一致影响下模型的鲁棒性。这些训练集分别约有占总身份数的 0、1/4、1/3、1/2、2/3、3/4、1 的身份存在跨模态身份不一致现象。测试集中的 query 图像均为 Market-1501 中 query 经过转化的素描图像,而 Gallery 中的图像均为 Market-1501 数据集中的原始 Gallery 图像。

评价指标:和主流行人重识别研究一致^[12],采用累

加匹配特性(CMC(Rank-1))与 mAP 两种评价指标。CMC 能反映出检索排序中特定名次的正确率。而 mAP 反映了所有 query 图像的平均正确率。

5.3 对比实验

5.3.1 S-Market1501 上的性能

在提出的数据集 S-Market1501 上进行了对比实验。由于相关的行人重识别代码尚未开源,将 CycleGan^[21]进行扩充后的行人图像与 S-Market1501 数据集中的原始图像数据作为 Baseline 训练集并与本文模型的性能进行对比,如表 3 所示。

在表 3 中,Baseline 的 Rank-1 与 mAP 较低,这是由于 Baseline 并没有设计相关的模块来约束网络提取模态不变的特征。Baseline+CycleGan 利用了生成网络生成的图像,所以即使在身份不一致图像占比较大的情况下进行训练,其性能也与身份不一致图像占比较小时的 Baseline 性能相接近。但是,在跨模态身份完全不一致的场景下,GAN 网络由于缺少一致身份图像的引导,不能由可见光模态的身份信息生成素描模态的身份信息。这导致模型无法学习到模态不变的信息,所以该方法在跨模态身份完全不一致时正确率较低。而相比于 Baseline,本文模型在跨模态身份不一致图像占比很高的情况下依旧有着稳定的性能表现。这体现了本文设计的模型在跨模态身份不一致情境下的鲁棒性。同时,模型性能相比于 Baseline+CycleGan 有着极大的提升,证明了本文模型在素描行人重识别任务中性能的优越性。

表 3 模型在 S-Market1501 上的性能
Table 3 Performance of the model on S-Market1501

Method		Proportion of inconsistent identity						
		0	1/4	1/3	1/2	2/3	3/4	1
Baseline	Rank-1 / %	37.4	36.4	35.9	32.6	30.8	29.2	18.0
	mAP / %	21.1	20.6	19.3	17.1	15.2	15.7	11.8
Baseline+CycleGan	Rank-1 / %	37.7	36.4	37.5	37.3	35.9	30.9	22.7
	mAP / %	21.0	21.1	20.6	20.1	20.8	17.3	13.6
Ours	Rank-1 / %	41.8	41.6	40.7	40.5	39.7	39.9	29.1
	mAP / %	23.5	23.2	23.1	22.0	21.1	21.4	20.9

5.3.2 Sketch Re-ID 上的性能

将本文模型的实验结果与其他 5 个文献提出方法的结果进行对比。其中,包括 1 个手工设计特征的方法 Dense-HOG^[22]+LBP+rankSVM,以及 4 个利用深度学习的方法 Triplet SN^[23],GN Siamese^[24],AFL Net^[1],LMDIF^[8]。Triplet SN 以及 GN Siamese 被设计用于手绘图像检索,而 AFL Net 和 LMDIF 是为素描行人重识别任务专门设计的。表 4 展示了本文设计模型的性能,“-”代表相关文章未提及该指标,提出的方法相比于手工建模的方法有着极大的提升,Rank-1 提升了 54.9 个百分点。相比于 Triplet SN 和 GN Siamese 至少提升了 31.1 个百分点。而相比于近年针

表 4 Sketch Re-ID 性能对比

Table 4 Performance comparison of Sketch Re-ID

Method	Rank-1 / %	mAP / %
Dense-HOG+LBP+rankSVM	5.1	-
Triplet SN	9.0	-
GN Siamese	28.9	-
AFL Net	34.0	-
LMDIF	49.0	-
Ours	60.0	54.4

对素描行人重识别的方法,本文方法的 Rank-1 提升了 11.0 个百分点,这体现了本文方法的优越性。

5.4 消融实验

使用 S-Market1501 数据集中的 4 号训练集训练模型,以此独立验证模型中每一个模块所起的作用。在 Baseline 的基础上,本文所设计的模块独立地加入训练,并将模型性能记录在表 5 中。

表 5 消融实验
Table 5 Ablation experiment

Method	Rank-1/%	mAP/%
Baseline	32.6	17.1
Baseline+C	36.1	19.7
Baseline+D	34.9	19.9
Baseline+C+D(Ours)	40.5	22.0

在模型仅使用交叉熵损失和三元组损失优化网络时(Baseline),模型 Rank-1 和 mAP 仅有 32.6% 和 17.1%。将交叉分类(Baseline+C)加入训练后,模型性能有着显著的提升。而将基于距离的身份信息对齐(Baseline+D)加入后,模型的性能也得到提升。这证明本文提出的模型中,每个部分都对网络性能有着明显提升。而当两个模块共同作用于网络(Baseline+C+D)时,本模型达到了最佳效果。

为进一步说明本文设计网络每个模块起到的效果,在图 5 中展示了特征向量降维可视化结果,其中每个图形代表一个样本,同形状的图形表示同一身份的样本。图 5(a)为 Baseline 编码得到特征的可视化结果,各身份之间没有清晰的决策边界;图 5(b)为 Baseline+C 网络编码得到的特征,可以发现同叉状图形样本与圆形样本虽然有着清晰的边界,却依旧受到跨模态身份不一致现象的影响;当 Baseline+C+D 对网络进行训练时,特征向量的可视化结果如图 5(c)所示。每一类均有较为清晰的决策边界。

5.5 S-Market1501 对真实场景素描图检索的贡献

为展示本文提出的 S-Market1501 对真实场景下的素描图检索的贡献,设计了相关实验。具体地,分别利用 Market-1501 的训练集和 S-Market1501 中 1 号训练集训练 Baseline,并在真实的素描数据集 Sketch Re-ID 上进行测试。其结果如表 6 所示。

表 6 Market-1501 和 S-Market1501 1 号训练的 Baseline 在 Sketch Re-ID 测试集上性能

Table 6 Performance of Baseline trained by Market-1501 and S-Market1501 (1) and test on Sketch Re-ID

Datasets	Rank-1 / %	mAP / %
Market-1501	2.0	1.4
S-Market1501 (1)	16.0	10.2

由表 6 可知,S-Market1501 训练得到的 Baseline 有着更高的性能,意味着本文设计的转化图在真实应用中有一定的贡献。

6 结 论

针对素描行人重识别中的跨模态身份不一致现象进行了分析,并提出了交叉分类的素描行人重识别网络。该网络能够有效地提取模态不变特征,缓解因跨模态身份不一致现象引起的性能下降,且无需引入参数量较大的判别器进行辅助训练,具有更低的计算成本。对比实验体现了本文所关注的跨模态身份不一致现象对模型性能的严重影响,也体现了本文提出的模型性能的优越性。消融实验体现了本文模型中每一个模块的有效性。在今后的工作中,将进一步探索并解决行人重识别任务在现实中所受到的挑战,为行人重识别技术的应用创造条件。

参 考 文 献

- [1] Pang L, Wang Y W, Song Y Z, et al. Cross-domain adversarial feature learning for sketch re-identification [C]//Proceedings of the 26th ACM international conference on Multimedia, October 22-26, 2018, Seoul, Republic of Korea. New York: ACM Press, 2018: 609-617.
- [2] Wang F Q, Zuo W M, Lin L, et al. Joint learning of single-image and cross-image representations for person re-identification[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1288-1296.
- [3] 刘莎,党建武,王松,等.结合一阶和二阶空间信息的行人重识别[J].激光与光电子学进展,2021,58(2):0215005. Liu S, Dang J W, Wang S, et al. Person re-identification based on first-order and second-order spatial information[J]. Laser & Optoelectronics Progress, 2021, 58(2): 0215005.
- [4] Luo H, Gu Y Z, Liao X Y, et al. Bag of tricks and a strong baseline for deep person re-identification[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 16-17, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 1487-1495.
- [5] 刘可文,房攀攀,熊红霞,等.基于多层级特征的行人重识别[J].激光与光电子学进展,2020,57(8):081503. Liu K W, Fang P P, Xiong H X, et al. Person re-identification based on multi-layer feature[J]. Laser & Optoelectronics Progress, 2020, 57(8): 081503.
- [6] Li H F, Pang J, Tao D P, et al. Cross adversarial consistency self-prediction learning for unsupervised domain adaptation person re-identification[J]. Information Sciences, 2021, 559: 46-60.
- [7] Li H F, Chen Y W, Tao D P, et al. Attribute-aligned domain-invariant feature learning for unsupervised domain adaptation person re-identification[J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 1480-1494.
- [8] Gui S J, Zhu Y, Qin X X, et al. Learning multi-level domain invariant features for sketch re-identification[J].

- Neurocomputing, 2020, 403: 294-303.
- [9] Wang G A, Zhang T Z, Cheng J, et al. RGB-infrared cross-modality person re-identification via joint pixel and feature alignment[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 3622-3631.
- [10] Zhang Z Y, Jiang S, Huang C, et al. RGB-IR cross-modality person ReID based on teacher-student GAN model [J]. Pattern Recognition Letters, 2021, 150: 155-161.
- [11] Wang G A, Zhang T Z, Yang Y, et al. Cross-modality paired-images generation for RGB-infrared person re-identification[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12144-12151.
- [12] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.
- [13] Choi S, Lee S M, Kim Y, et al. Hi-CMD: hierarchical cross-modality disentanglement for visible-infrared person re-identification[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 10254-10263.
- [14] Zheng L, Shen L Y, Tian L, et al. Scalable person re-identification: a benchmark[C]//2015 IEEE International Conference on Computer Vision, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1116-1124.
- [15] Lu Y, Wu Y, Liu B, et al. Cross-modality person re-identification with shared-specific feature transfer[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 13376-13386.
- [16] Yu Q, Liu F, Song Y Z, et al. Sketch me that shoe[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 799-807.
- [17] Liu Q, Xie L X, Wang H Y, et al. Semantic-aware knowledge preservation for zero-shot sketch-based image retrieval[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 3661-3670.
- [18] Lei J J, Song Y X, Peng B, et al. Semi-heterogeneous three-way joint embedding network for sketch-based image retrieval[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(9): 3226-3237.
- [19] Bhunia A K, Yang Y X, Hospedales T M, et al. Sketch less for more: on-the-fly fine-grained sketch-based image retrieval[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 9776-9785.
- [20] Pang K Y, Yang Y X, Hospedales T M, et al. Solving mixed-modal jigsaw puzzle for fine-grained sketch-based image retrieval[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 10344-10352.
- [21] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2242-2251.
- [22] Hu R, Collomosse J. A performance evaluation of gradient field HOG descriptor for sketch based image retrieval[J]. Computer Vision and Image Understanding, 2013, 117(7): 790-806.
- [23] Yu Q, Liu F, Song Y Z, et al. Sketch me that shoe[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 799-807.
- [24] Sangkloy P, Burnell N, Ham C, et al. The sketchy database: learning to retrieve badly drawn bunnies[J]. ACM Transactions on Graphics, 2016, 35(4): 1-12.