

## 跨模态行人重识别的对称网络算法\*

张艳,相旭,唐俊,王年,屈磊  
(安徽大学电子信息工程学院,安徽合肥 230601)

**摘要:**针对模态间差异,提出基于对称网络的跨模态行人重识别算法,该网络将基于概率分布的模态混淆与对抗学习结合,通过对称网络产生模态不变特征,从而达到模态混淆的目的;针对外观差异和模态内差异,使用不同隐藏层的网络卷积特征构造混合三元损失,提高网络的特征表征能力。RegDB和SYSU-MM01数据集上的大量实验结果表明了该方法的有效性。

**关键词:**跨模态;行人重识别;对称网络;对抗学习;混合三元损失

中图分类号:TP391.4 文献标志码:A 文章编号:1001-2486(2022)01-122-07

## Cross-modality person re-identification algorithm using symmetric network

ZHANG Yan, XIANG Xu, TANG Jun, WANG Nian, QU Lei

(School of Electronic and Information Engineering, Anhui University, Hefei 230601, China)

**Abstract:** For the difference between modalities, a cross-modality person re-identification algorithm which based on symmetric network was proposed. The network combined the modal confusion based on probability distribution with adversarial learning, and generated modal-invariant features through symmetric network to achieve modal confusion. To deal with appearance differences and intra-modality differences, the network constructed a mixed-triplet loss using convolution features of different hidden layers, which can improve the characterization capability of the network. Numerous experimental results on the RegDB and SYSU-MM01 datasets demonstrate the effectiveness of the method.

**Keywords:** cross-modality; person re-identification; symmetric network; adversarial learning; mixed-triplet loss

行人重识别旨在多摄像头多场景下寻找同一个目标人物,因其在目标追踪、公共安防、视频监控等领域起到重要的作用,行人重识别受到了学者们的普遍关注。行人重识别难点主要在于视角变化、遮挡、行人姿态变化、光照差异和背景杂乱等,目前已经提出了大量行人重识别算法<sup>[1-6]</sup>,推动了行人重识别的快速发展。传统的行人重识别属于单模态的检索任务,是仅在可见光图像之间进行检索的技术。但随着社会发展,传统行人重识别并不能满足社会需求,跨模态行人重识别便从行人重识别衍生出来,它比传统行人重识别更具有挑战性。跨模态行人重识别旨在红外光图像与可见光图像之间匹配同一个目标人物的技术,跨模态行人重识别作为一个新兴的研究课题,已成为当前的一个研究热点。

在夜间,可见光摄像机无法捕获足够的行人外观信息,行人的信息由红外摄像机或深度

摄像机来获取。由于红外摄像机和可见光摄像机成像机制不同,两种模态之间存在巨大的模态差异。可见光图像和红外光图像示例如图1所示,两类图像在外观上有很大的不同,可见光图像比红外光图像包含更多的颜色信息,模态差异成为跨模态行人重识别需要解决的另一个难题。

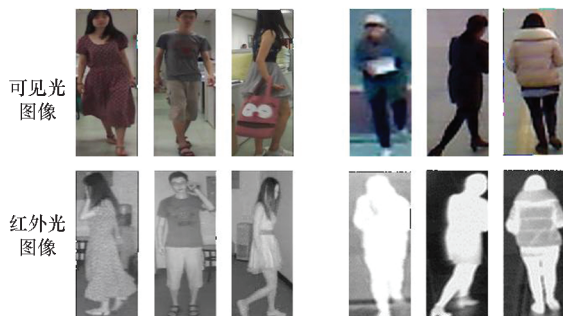


图1 可见光图像和红外光图像示例

Fig. 1 Examples of visible images and infrared images

\* 收稿日期:2020-07-22

基金项目:国家自然科学基金资助项目(61772032,61871411);国家重点研发计划资助项目(2018YFC0807302)

作者简介:张艳(1982—),女,安徽宣城人,副教授,博士,硕士生导师,E-mail:zhangyan@ahu.edu.cn;

唐俊(通信作者),男,教授,博士,博士生导师,E-mail:tangjunahu@163.com

可见光模态和红外光模态之间的模态差异可分为特征差异和外观差异。为了减小特征差异,文献[7-9]试图利用统一的嵌入空间来对齐跨模态特征,但忽略了两个模态之间巨大的外观差异。文献[7,10]使用生成对抗网络(Generative Adversarial Networks, GAN)来实现跨模态的图像转换,从而减少外观差异的影响。虽然由GAN生成的虚拟图像与原始图像相似,但是并不能保证生成与身份相关的细节信息。文献[11]采用双流网络分别处理不同模态的图像,并将隐藏层卷积特征与网络输出特征进行融合以增强特征的鉴别能力。这表明隐藏层卷积特征具有描述结构和空间信息的能力,利用隐藏层卷积特征来缩小两种模态间差异是一种有效的解决方案。

针对模态间和模态内差异问题,本文提出基于对称网络的跨模态行人重识别算法。针对模态间差异,受文献[12]的启发,本文算法将基于概率分布的模态混淆这一思想与对抗学习结合,为可见光模态和红外光模态分别构造分类器,构成了对称网络,通过最小化两个分类器输出概率分布的差异来产生模态不变特征,从而达到模态混淆的目的;本文还提出混合三元损失,并在对称网络的不同深度对齐特征,以减少模态内差异带来的影响;针对可见光模态和红外光模态之间存在的外观差异,本文利用隐藏层卷积特征增强网络学习空间结构信息的能力,以减少外观差异带来的影响。

## 1 方法实现

### 1.1 问题描述

从可见光图像集  $V$  中选取一个样本  $V_i$ , 再从红外光图像集  $T$  中选取一个样本  $T_j$ , 其中  $V_i$  和  $T_j$  是同一个人  $y$  在不同模态下的图像,  $y \in Y$ ,  $Y$  是行人身份 (IDentification, ID) 的集合。将  $V_i$  和  $T_j$  输入到主干网络中提取模态特定特征, 然后将特征嵌入公共空间中, 获得特征  $X$ , 最后输入到对应的分类器中, 获得相应的概率分布  $D_i, D_j$ 。在检索过程中, 给定一个待检索图像, 通过网络提取图像特征  $X$ , 然后将待检索行人图像特征和检索库中的图像特征逐一进行距离相似性度量计算, 根据距离的大小排序得到前  $k$  张与待检索行人最为相似的图像。

### 1.2 网络结构

本文提出了一个对称网络, 网络由生成器和鉴别器组成, 网络结构如图2所示。生成器由两列独立的 ResNet50 和两列独立的全连接层构成, 生成器模块通过两列 ResNet50 提取特定模态下的特征, 全连接层将特定模态特征嵌入公共空间中, 学习两种模态下的公共特征, 以减小模态间差异的影响; 生成器通过学习数据分布来减小模态差异, 混淆鉴别器。鉴别器由可见光分类器和红外光分类器组成, 通过各分类器得到相应的类别概率分布以区分数据来自哪一模态。本文算法通过对鉴别器与生成器进行交替训练, 减小可见光模态和红外光模态间的模态差异。

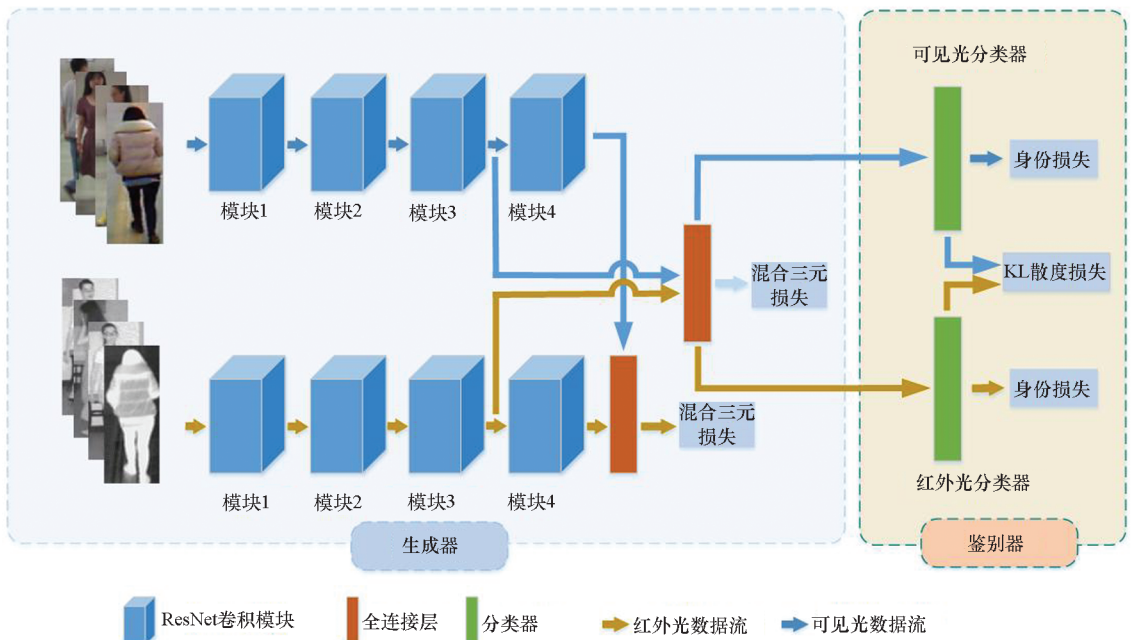


图2 本文提出的网络结构

Fig. 2 Network structure proposed in this paper

### 1.3 基于类别概率分布的模式混淆

从  $V, T$  中分别选取一个样本  $V_i$  和  $T_j$ , 并将它们输入到对应的 ResNet50 分支中提取模态特定特征, 然后输入到对应的分类器中, 获得  $V_i$  和  $T_j$  的概率分布  $D_i$  和  $D_j$ 。如果将  $T_j$  的特征输入到可见光分类器中获得  $D'_j$ , 则概率分布  $D'_j$  与  $D_i$  是不同的, 甚至有很大的偏差, 因为可见光分类器无法很好地对红外光特征进行分类。据此, 本文认为特定模态分类器只能对特定模态特征进行准确的分类, 在已知  $V_i$  和  $T_j$  以及  $y$  的情况下, 分类器能够通过  $D'_j$  与  $D_i$  概率分布的差异来判定数据来自哪个模态。本文使用 KL 散度来衡量  $D'_j$  与  $D_i$  之间的概率分布差异, 定义 KL 散度损失为:

$$\mathcal{L}_{kl} = KL(D'_j, D_i) \quad (1)$$

式中,  $D'_j$  与  $D_i$  是分类器输出的概率分布。

本文采用交叉熵损失作为分类器的身份损失, 则两个分类器的身份损失为:

$$\mathcal{L}_{id} = -\frac{1}{N_1} \sum_{i=1}^{N_1} \lg(y_i | V_i) - \frac{1}{N_2} \sum_{j=1}^{N_2} \lg(y_j | T_j) \quad (2)$$

式中,  $N_1$  和  $N_2$  是该训练批次下相应模态下的样本数量, 等号右边第一项是可见光分类器身份损失, 第二项是红外光分类器身份损失。

### 1.4 混合三元损失

为了解决模态间差异和模态内差异问题, 常见的思路是提取特定模态下特征, 然后通过三元损失来缩小模态间差异和模态内差异。图 3 为模态间差异和模态内差异示意图, 其中模态间差异指可见光域图像和红外光域图像因成像原理不同而导致成像后图像上的差异; 模态内差异是指在同一模态下由于行人的姿态不同、行人类别不同、遮挡、摄像机视角变化等造成的差异。图 3 中,  $V_a$  代表样本 1 的可见光图像,  $V_p$  代表样本 1 姿态不同的另一可见光图像,  $V_n$  代表样本 2 的可见光图像,  $V_n$  与  $V_a$  包含的 ID 信息不同; 而  $T_a, T_p$  代表样本 1 的红外光图像,  $T_n$  代表样本 2 的红外光图像。可见光域图像  $V$  与红外光域图像  $T$  之间存在模态间差异;  $V_a, V_p, V_n$  之间, 以及  $T_a, T_p, T_n$  之间存在模态内差异。为解决这些差异, 文献[11, 13]将三元损失分为模态间损失和模态内损失两部分。

与之前方法不同, 本文将三元损失与对抗学习结合, 提出了一种混合三元损失来减小模态间的差异。当通过对抗学习达到模态混淆时, 不再需要区分样本来自哪个模态, 即在不区分模态的

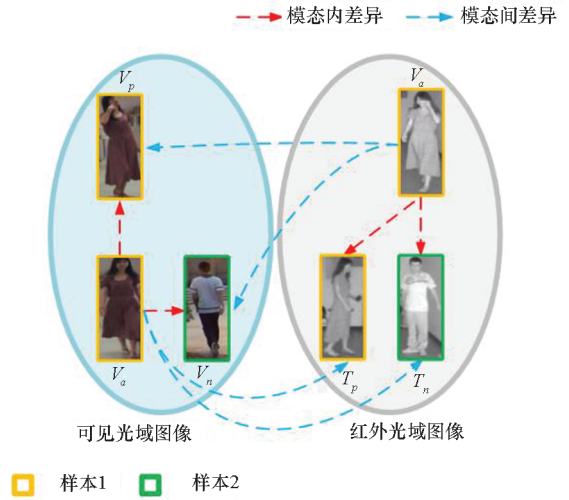


图 3 模态间差异和模态内差异示意图  
Fig. 3 Schematic diagram of the inter-modal differences and the intra-modal differences

情况下选择正样本特征和负样本特征, 以进行特征对齐, 减小模态差异。根据文献[14], 混合三元损失定义为:

$$\mathcal{L}_{tri}(X) = \sum_{i=1}^P \sum_{a=1}^K [\rho + \max_{p=1,2,\dots,K} D(X_a^i, X_p^i) - \min_{\substack{j=1,2,\dots,P \\ n=1,2,\dots,K \\ j \neq i}} D(X_a^i, X_n^i)] \quad (3)$$

式中,  $X_a$  代表锚点样本的特征,  $X_p$  代表正样本的特征,  $X_n$  代表负样本的特征,  $D$  代表样本特征之间的欧式距离,  $\rho$  为混合三元损失中预定义的最小间隔。设置每个训练批次大小为  $P \cdot K$ , 其中选取  $P$  个行人 ID 身份, 并从训练集中随机为每个行人 ID 选取  $K$  张可见光图像和  $K$  张红外光图像, 在将所有的样本图像输入网络后, 获得相应特征, 本文采用欧氏距离度量的方式来衡量样本特征之间相似度的大小。首先选取一个特征作为锚点样本特征; 然后从中选取与锚点样本特征 ID 信息相同, 但是与锚点样本特征欧氏距离最远的正样本特征作为最不相似的正样本特征; 最后从中选取与锚点样本特征 ID 信息不同, 且与锚点样本特征欧氏距离最近的负样本特征作为最相似的负样本特征, 以此组成混合三元损失来训练网络。

根据文献[11]的思想, 为了使网络能够学习到更多的空间结构信息, 减小颜色差异的影响, 本文采用隐藏层卷积特征即来自 ResNet50 模块 3 的特征作为后面全连接层的输入。在 ResNet50 的模块 3 和模块 4 的输出层使用混合三元损失来优化网络, 则总的混合三元损失定义为:

$$\mathcal{L}_{total} = \mathcal{L}_{tri1} + \mathcal{L}_{tri2} \quad (4)$$

式中,  $\mathcal{L}_{tri1}$  是表示模块 3 的混合三元损失,  $\mathcal{L}_{tri2}$  是表

示模块 4 的混合三元损失。

## 1.5 训练细节

本文的网络属于端到端的对称网络结构,具体的算法步骤如下所示。

**Step 1:** 训练网络的生成器和鉴别器,使网络能对样本正确分类。使用身份损失和混合三元损失训练网络,使网络能够捕获模态可鉴别特征。则该步骤总损失为:

$$\mathcal{L}_1 = \mathcal{L}_{id} + \beta \mathcal{L}_{trial} \quad (5)$$

**Step 2:** 训练鉴别器,固定生成器的参数,通过最大化  $\mathbf{D}'_j$  与  $\mathbf{D}_i$  的概率分布差异,提升网络分辨图片模态属性的能力。该步骤总损失为:

$$\mathcal{L}_2 = \mathcal{L}_{id} - \alpha \mathcal{L}_{kl} \quad (6)$$

**Step 3:** 训练生成器以最小化  $\mathbf{D}'_j$  与  $\mathbf{D}_i$  的概率分布差异。此时冻结鉴别器的参数,在训练过程中,该步骤循环三次以平衡生成器和鉴别器。总损失为:

$$\mathcal{L}_3 = \alpha \mathcal{L}_{kl} + \beta \mathcal{L}_{trial} \quad (7)$$

在对样本正确分类前提下,通过对生成器和鉴别器的交替训练,使  $\mathbf{D}'_j$  与  $\mathbf{D}_i$  的概率分布差异最小化,达到模态混淆的目的。

## 2 实验设置

本节首先介绍实验所用的数据集和评价标准,其次介绍实验环境及参数设置。

### 2.1 数据集及评价标准

#### 2.1.1 跨模态行人重识别数据集介绍

SYSU-MM01<sup>[15]</sup>是由 4 个可见光摄像机和 2 个红外光摄像机收集的大规模跨模态数据集。该数据集一共有室内和室外两种场景,训练集包含 395 个行人 ID 数据,其中包括 11 909 张红外光图像和 22 258 张可见光图像。测试集包含 96 个行人 ID 数据,本文采用文献[15]的设置,将测试集 3 803 张红外图像作为查询集,并随机选择可见光图像作为测试集。

RegDB<sup>[16]</sup>共有 412 个行人 ID 数据,这些数据由双摄像头系统捕获。每个 ID 包含 10 张可见光图像和 10 张红外光图像。采用文献[8,13]的设置,将数据集随机分为两个部分,每部分随机选择一半数据分别用于训练和测试。

#### 2.1.2 评价标准

本文采用首位命中率(Rank- $k$ , R- $k$ )、平均准确率均值(mean Average Precision, mAP)和累积匹配特征曲线(Cumulative Matching Curve, CMC)作为评价指标。CMC 为在检索库图像集合

中与待查询图像匹配相似度最高的前  $k$  张图片命中查询图像的概率,其常以 R- $k$  的形式体现。其中 mAP 的定义为:

$$mAP = \frac{AP}{Q} \quad (8)$$

式中, $Q$  代表查询集数量,CMC 可认为是 Rank list 的可视化。

$$CMC(r) = \frac{m(r)}{Q} \quad (9)$$

### 2.2 实验环境及参数设置

本文实验环境如下:操作系统为 ubuntu16.04,深度学习框架为 Pytorch1.1,编程语言为 python3.6, GPU 为 NVIDIA GeForce RTX 2080Ti。本文采用 ResNet50 作为骨干网络,并采用在 ImageNet 上预训练的参数初始化网络权重,行人的特征维度设置为 1 024。在数据处理阶段,将输入图像大小调整为 288 像素  $\times$  144 像素,然后对边缘进行 10 层补零填充,将其随机左右翻转并裁剪为 288 像素  $\times$  144 像素进行数据增广,设置  $K=4$ ,  $P=16$ ,  $N=P \cdot K=2N_1=2N_2$  张图像构成一个批次。本文使用自适应梯度优化器对网络模型进行优化,其中指数衰减率被设置为  $b1=0.9$ ,  $b2=0.999$ 。网络在训练时,遍历整个数据集总次数为 80 次,初始学习速率设置为 0.000 01,然后在第 50 次中衰减为原来的 0.1 倍。对于本文提出的混合三元损失,预定义的最小间隔均设置为  $\rho=0.5$ ,对于 KL 系数  $\alpha$  和混合三元损失系数  $\beta$ ,设置为  $\alpha=1$  和  $\beta=1.4$ 。

## 3 实验结果及分析

本节首先对比分析本文算法与现有的跨模态行人重识别算法,最后,通过全面的消融研究,来分析模中每个组成部分对算法性能的影响。

### 3.1 与现有算法对比

将本文方法与跨模态行人重识别算法<sup>[7-9,11,13,15,17-18]</sup>等实验结果进行对比,对比结果见表 1,表 1 中加粗数值是最大值。由表 1 可知,本文算法提取到的特征更具有表征能力,这证明了本文算法的有效性。

对比跨模态算法<sup>[8-9,15,17]</sup>,本文算法有很大的提升。本文算法在 SYSU-MM01 数据集的 R-1 结果比文献[13]的跨模态行人重识别算法的 R-1 结果提升 10.98%,比文献[7]的跨模态行人重识别算法的结果提升 8.96%;mAP 结果也比文献[7]的算法结果提升 10.02%,与文献[7,13]的两种对抗学习相关的方法对比,证明了本

文算法的有效性。本文在这两个数据集上的结果比文献[11]的跨模态行人重识别算法有所提升,在 RegDB 的 R - 1 结果比文献[11]的算法提升 1.16%, R - 20 提升 1.59%,对比文献[11]的算法通过跨层融合的方式来使用隐藏层特征,本文使用隐藏层特征的策略更为有效。本文算法在 SYSU - MM01 数据集上 mAP 的结

果略低于文献[11]的算法,但检索算法在实际应用中更侧重于 R - k 指标,R - k 代表在样本按照相似度排序后,前 k 个样本的命中概率,更能体现检索算法的实用性和检索准确度。本文算法在 R - 1, R - 10, R - 20 的结果均优于其他跨模态算法,通过 R - k 的结果足以说明本文算法的有效性和鲁棒性。

表 1 在 RegDB 数据集和 SYSU-MM01 数据集上的比较实验结果

Tab. 1 Comparison experiment results on RegDB dataset and SYSU-MM01 dataset

方法	RegDB				SYSU-MM01				%
	R - 1	R - 10	R - 20	mAP	R - 1	R - 10	R - 20	mAP	
Zero-Padding <sup>[15]</sup> (CVPR2017)	17.75	34.21	44.35	18.90	14.80	54.12	71.33	15.95	
TONE <sup>[8]</sup> (AAAI2018)	16.87	34.03	44.10	14.92	12.52	50.72	69.60	14.42	
TONE + HCML <sup>[8]</sup>	24.44	47.53	56.78	20.80	14.32	53.16	69.17	16.16	
BCTR <sup>[9]</sup>	32.67	57.64	66.58	30.99	16.12	54.90	71.47	19.15	
BDTR <sup>[9]</sup>	33.47	58.42	67.52	31.83	17.01	55.43	71.96	19.66	
cmGAN <sup>[13]</sup>	—	—	—	—	26.97	67.51	80.56	27.80	
MAC <sup>[17]</sup>	36.43	62.36	71.63	37.03	33.26	79.04	90.09	36.22	
D <sup>2</sup> RL <sup>[7]</sup> (CVPR2019)	43.40	66.10	76.30	44.10	28.90	70.60	82.40	29.20	
HSME <sup>[18]</sup> (AAAI2019)	50.85	73.16	81.66	47.00	20.68	62.74	77.95	23.12	
EDFL <sup>[11]</sup>	52.58	72.10	81.47	52.98	36.94	84.52	93.22	<b>40.77</b>	
本文方法	<b>53.74</b>	<b>74.71</b>	<b>83.06</b>	<b>53.67</b>	<b>37.86</b>	<b>84.56</b>	<b>94.52</b>	39.22	

### 3.2 消融研究

首先,研究参数  $\alpha$  和  $\beta$  取值对模型性能的影响。不同  $\alpha$  值在 RegDB 数据集上的结果如图 4 所示。当  $\alpha = 1$  时,网络的性能达到了最优,R - 1 结果为 53.74%,mAP 结果为 53.67%。不同  $\alpha$  值在 SYSU-MM01 数据集上的结果如图 5 所示。当  $\alpha = 1$  时, R - 1 结果最高为 37.86%,mAP 结果为 39.22%,此时在该数据集上的结果达到了最优。

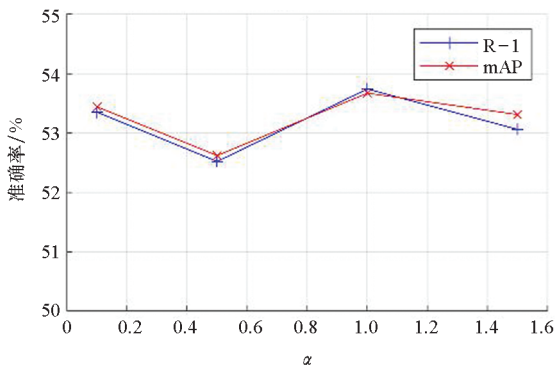


图 4 不同  $\alpha$  值在 RegDB 数据集上的结果  
Fig. 4 Results of different  $\alpha$  values on the RegDB dataset

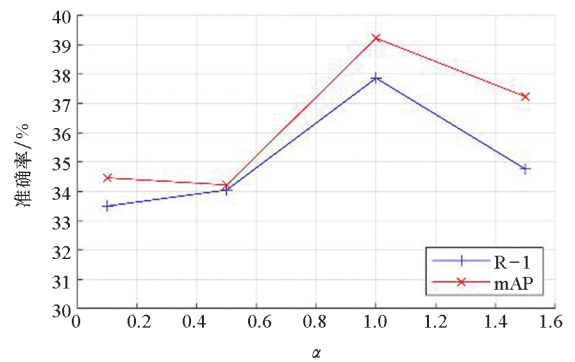


图 5 不同  $\alpha$  值在 SYSU-MM01 数据集上的结果  
Fig. 5 Results of different  $\alpha$  values on the SYSU - MM01 dataset

图 6 和图 7 给出了在  $\beta$  不同取值下,网络在两个数据集上的性能,当  $\beta = 1.4$  时,网络性能达到最优。当  $\beta = 1.4$  时,在 RegDB 数据集上, R - 1 为 53.40%,mAP 为 53.75%;在 SYSU-MM01 上,R - 1 为 37.86%,mAP 为 39.22%。从以上实验还可以看出,本文提出的模型在  $\alpha$  和  $\beta$  较宽泛的取值范围内都能取得较好的结果,这也反映了模型的鲁棒性。

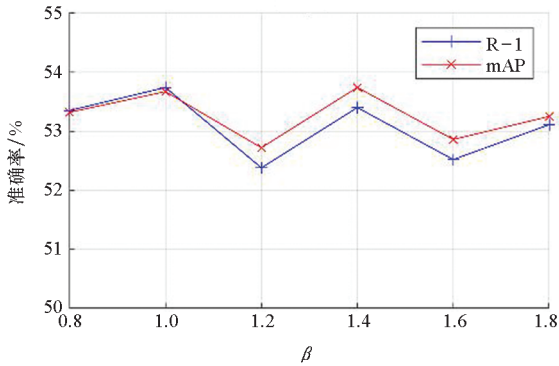


图 6 不同  $\beta$  值在 RegDB 数据集上的结果

Fig. 6 Results of different  $\beta$  values on the RegDB dataset

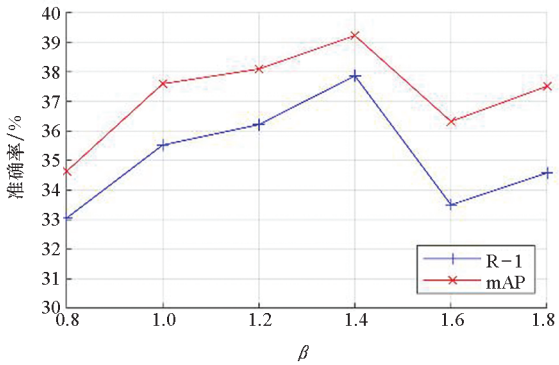


图 7 不同  $\beta$  值在 SYSU-MM01 数据集上的结果

Fig. 7 Results of different  $\beta$  values on the SYSU-MM01 dataset

本文方法在 SYSU - MM01 数据集上的消融实验结果见表 2。基准网络 (baseline) 由两个并行独立的 ResNet50 以及两列共享的全连接层构成, baseline 仅采用身份损失。baseline 的 R - 1 结果为 12.96%, mAP 为 17.12%。第二组实验仅增加混合三元损失模块 (即 mtri, 模块 4 的特征), 与 baseline 结果相比, 增加混合三元损失模块后, R - 1 结果提升了 10.07%, mAP 结果提升了 9.13%, 这证明了混合三元损失的有效性。此外, 将第二组实验与第三组实验 (baseline + AL + mtri(4), AL 代表基于概率分布的模态混淆) 进行对比, 当增加基于概率分布的模态混淆模块后, R - 1 的结果提升了 3.87%, mAP 也提升了 5.86%, 这也证明了基于概率分布的模态混淆模块的有效性。

baseline + AL + tri(4,3) 从不同深度优化网络 (即从模块 4,3 优化网络但不使用隐藏层特征), 对比 baseline + AL + mtri(4) 的结果, 其 R - 1 结果提升了 6.02%, mAP 提升了 2.72%; 对比 baseline + AL + mtri(3,4) (即从模块 3,4 优化网络, 使用隐藏层特征) 和 baseline + AL + mtri(3) 实

表 2 在 SYSU-MM01 数据集上的消融实验结果

Tab. 2 Ablation results on the SYSU-MM01 dataset

	%			
SYSU-MM01	R - 1	R - 10	R - 20	mAP
baseline	12.96	55.98	73.97	17.12
baseline + mtri(4)	23.03	72.15	87.67	26.25
baseline + AL + mtri(4)	26.90	73.70	86.08	32.11
baseline + AL + mtri(3)	30.27	81.01	93.72	34.17
baseline + AL + mtri(4,3)	32.92	81.78	93.79	34.83
baseline + AL + mtri(3,4)	37.86	84.56	94.52	39.22

验结果, 在都采用隐藏层特征前提下, 证明了从不同深度使用混合三元损失优化网络, 可以使网络获得更好的表征能力。对比 baseline + AL + mtri(3) 和 baseline + AL + mtri(4) 的结果, 当网络采用隐藏层特征时, baseline + AL + mtri(3) 的 R - 1 结果提升了 3.37%, mAP 结果提升了 2.06%; 对比 baseline + AL + mtri(3,4) 和 baseline + AL + mtri(4,3) 结果, 从不同深度优化网络的前提下, (3,4) 使用隐藏层特征, (4,3) 不使用隐藏层特征, 证明了隐藏层特征可以使网络更好地学习空间结构信息。

## 4 结论

本文根据基于概率分布的模态混淆思想, 提出了基于对称网络的跨模态行人重识别算法。将基于概率分布的模态混淆、对抗学习以及混合三元损失结合在一起, 解决模态间差异和模态内差异, 提高网络对空间结构的关注度, 降低外观差异带来的负面影响。此外, 充分利用了隐藏层特征, 提高了特征对空间结构的表征能力。在 RegDB 和 SYSU-MM01 的实验结果证明了本文方法的有效性。

## 参考文献 (References)

[1] 吴彦丞, 陈鸿昶, 李邵梅, 等. 基于行人属性先验分布的行人再识别[J]. 自动化学报, 2019, 45(5): 953 - 964.  
WU Y C, CHEN H C, LI S M, et al. Person-re-identification using attribute priori distribution [J]. Acta Automatica Sinica, 2019, 45(5): 953 - 964. (in Chinese)

[2] ZHONG Z, ZHENG L, ZHENG Z D, et al. Camera style

- adaptation for person re-identification [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018; 5157 – 5166.
- [3] YANG J W, SHEN X, TIAN X M, et al. Local convolutional neural networks for person re-identification[C]//Proceedings of the 26th ACM International Conference on Multimedia, 2018; 1074 – 1082.
- [4] SUN X X, ZHENG L. Dissecting person re-identification from the viewpoint of viewpoint [C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019; 608 – 617.
- [5] QIAN X L, FU Y W, XIANG T, et al. Pose-normalized image generation for person re-identification [C]//Proceedings of European Conference on Computer Vision, 2018; 661 – 678.
- [6] LUO C C, CHEN Y T, WANG N Y, et al. Spectral feature transformation for person re-identification[C]//Proceedings of IEEE/CVF International Conference on Computer Vision, 2019; 4975 – 4984.
- [7] WANG Z X, WANG Z, ZHENG Y Q, et al. Learning to reduce dual-level discrepancy for infrared-visible person re-identification[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019; 618 – 626.
- [8] YE M, LAN X Y, LI J W, et al. Hierarchical discriminative learning for visible thermal person re-identification [C]//Proceedings of the 32nd AAAI Conference on Artificial Intelligence, 2018; 7501 – 7508.
- [9] YE M, WANG Z, LAN X Y, et al. Visible thermal person re-identification via dual-constrained top-ranking [C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence, 2018; 1092 – 1099.
- [10] WANG G A, ZHANG T Z, CHENG J, et al. RGB-infrared cross-modality person re-identification via joint pixel and feature alignment [C]//Proceedings of IEEE/CVF International Conference on Computer Vision, 2019; 3622 – 3631.
- [11] LIU H J, CHENG J, WANG W, et al. Enhancing the discriminative feature learning for visible-thermal cross-modality person re-identification[J]. *Neurocomputing*, 2020, 398: 11 – 19.
- [12] SAITO K, WATANABE K, USHIKU Y, et al. Maximum classifier discrepancy for unsupervised domain adaptation [C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018; 3723 – 3732.
- [13] DAI P Y, JI R R, WANG H B, et al. Cross-modality person re-identification with generative adversarial training [C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence, 2018; 677 – 683.
- [14] HERMANS A, BEYER L, LEIBE B. In defense of the triplet loss for person re-identification [EB/OL]. (2017 – 9 – 21) [2020 – 07 – 14]. <https://arxiv.org/abs/1703.07737>.
- [15] WU A C, ZHENG W S, YU H X, et al. RGB-infrared cross-modality person re-identification [C]//Proceedings of IEEE International Conference on Computer Vision, 2017; 5390 – 5399.
- [16] NGUYEN D T, HONG H G, KIM K W, et al. Person recognition system based on a combination of body images from visible light and thermal cameras [J]. *Sensors (Basel, Switzerland)*, 2017, 17(3): 605.
- [17] YE M, LAN X Y, LENG Q M. Modality-aware collaborative learning for visible thermal person re-identification [C]//Proceedings of the 27th ACM International Conference on Multimedia, 2019; 347 – 355.
- [18] HAO Y, WANG N N, LI J, et al. HSME: hypersphere manifold embedding for visible thermal person re-identification[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 8385 – 8392.