计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.191100061

基于改进困难三元组损失的跨模态行人重识别框架

李 灏 唐 敏 林建武 赵云波
浙江工业大学信息工程学院 杭州 310023
(li_hao_056@qq.com)

摘 要 为了提升跨模态行人重识别算法的识别精度,提出了一种基于改进困难三元组损失的特征学习框架。首先,改进了传统困难三元组损失,使其转换为全局三元组损失。其次,基于跨模态行人重识别中存在模态间变化及模态内变化的问题,设计了模态间三元组损失及模态内三元组损失,以配合全局三元组损失进行模型训练。在改进困难三元组损失的基础上,首次在跨模态行人重识别模型中设计属性特征来提高模型的特征提取能力。最后,针对跨模态行人重识别中出现的类别失衡问题,首次将 Focal Loss 用于替代传统交叉熵损失来进行模型训练。相比现有算法,在 RegDB 数据集实验中,所提框架在各项指标中均有 1.9%~6.4%的提升。另外,通过消融实验证明了 3 种方法均能提升模型的特征提取能力。

关键词:跨模态;行人重识别;困难三元组损失;属性特征;类别失衡
中图法分类号 TP391.41

Cross-modality Person Re-identification Framework Based on Improved Hard Triplet Loss

LI Hao, TANG Min, LIN Jian-wu and ZHAO Yun-bo
College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China

Abstract In order to improve the recognition accuracy of cross-modality person re-identification, a feature learning framework based on improved hard triplet loss is proposed. Firstly, traditional hard triplet loss is converted to a global one. Secondly, intra-modality and cross-modality triplet losses are designed to match the global one for model training based on the intra-modality and cross-modality variations. On the basis of improving the hard triplet loss, for the first time the attribute features are designed to increase the ability of the model to extract features in the cross-modality person re-identification model. Finally, for the category imbalance problem, Focal Loss is used to replace the traditional Cross Entropy loss for model training. Compared with existing algorithms, the proposed approach behaves the best on the publicly available RegDB dataset, with an increase of 1.9%~6.4% in all evaluation indicators. In addition, ablation experiments also show that all the three methods can improve the feature ability extraction of the model.

Keywords Cross-modality, Person re-identification, Hard triplet loss, Attribution feature, Category imbalance

1 引言

行人重识别方法[1-2]旨在解决不重叠监控设备下行人图像的匹配问题,在罪犯定位识别、特定人物跟踪等方面具有重要作用。该类方法的关键难点是由行人姿势、光照强度、摄像头视角等造成的类内变化和类间变化。现有方法往往使用可见光相机(Visible cameras)获取图像,通过设计行人图片表征特征[3]或通过度量学习方法提取具有明显区分性的行人特征[4]来提高准确率,但在夜晚或光线较差的情况下,可见光相机往往难以获取理想的图像,从而影响算法的准确度。

针对可见光相机在光线较差的环境下无法获取有效信息这一问题,Ye 等[5]提出了使用对光线强度依赖更小的热成像相机(Thermal cameras)获取人物的外观信息来进行行人重识别的 VT-REID(Visible-Thermal person Re-identification)方法。VT-REID 弥补了传统行人重识别方法在光线不充足的情况下无法进行有效识别的不足,同时引入了不同模态间图像匹配的问题。目前,Nguyen 等[6]提出的 RegDB 数据集是 VT-REID 公开数据集,如图 1 所示,数据集中包含大量不同模态、不同视角的行人照片,保证了对跨模态方法进行验证的有效性。鉴于跨模态行人重识别在智能安防领域的重要性,有必要在现有研究的基础上对跨模态识别方案进行深入研究。

VT-REID 面临的挑战不仅包括行人个体在模态内存在视角变化、姿态变化等传统 REID 问题,还包括行人个体存在

到稿日期:2019-11-08 返修日期:2020-04-03 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家自然科学基金(61673350)
This work was supported by the National Natural Science Foundation of China (61673350).
通信作者:赵云波(ybzhao@ieee.org)

模态间差异性问题。如图 2(b)所示，相同行人在不同模态间的图片存在巨大差异，模型从可见光模态中学习的颜色、衣服图案等信息无法运用到热成像模态中进行识别，因此当前跨模态行人重识别研究专注于如何对两种模态图片的特征提取进行学习。目前，VT-REID 的研究方法主要使用双流结构进行可见光图片和热成像图片的特征提取，并学习辨别性特征。Ye 等[5]提出了双流卷积网络(Two-stream CNN Network，TONE)来学习跨模态特征表达；Hao 等[7]提出了具有分类和识别约束的端到端双流超球体流型嵌入网络(HyperSphere Manifold Embedding Network，HSME)；Wang 等[8]提出了校准生成对抗网络(Alignment Generative Adversarial Network AlignGAN)进行模态间的特征对齐和像素对齐。然而当前研究无法在公有数据集上达到理想效果，这表明跨模态行人重识别问题极具挑战性。

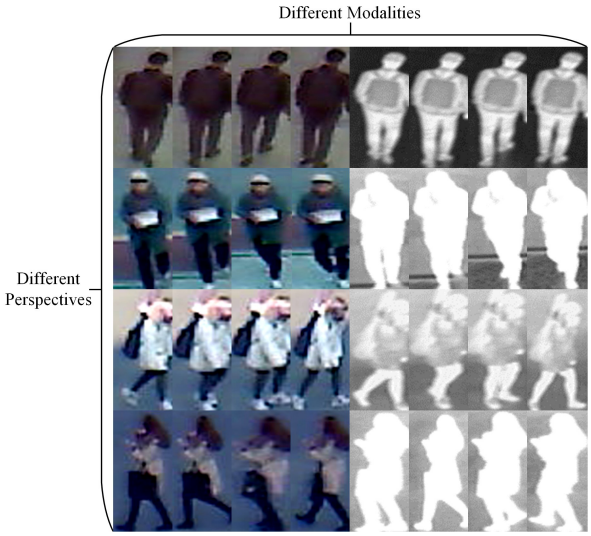


图 1　RegDB 数据集图片

Fig. 1　Picture of RegDB dataset

(a)同一行人模态内变化　(b)同一行人模态间变化

(c)不同行人模态内变化　(d)不同行人模态间变化

图 2　RegDB 数据集的模态和变化

Fig. 2　Modality and variation in RegDB dataset

基于跨模态行人重识别存在的模态间差异问题及传统 ReID 中姿态、视角等差异问题，本文提出了多种特征学习方法，并在 RegDB 数据集上证明了所提方法的有效性。本文使用 Resnet50m[9]作为骨干网络，对传统困难三元组损失函数(Hard Triplet Loss)[10]进行改进，提出了全局三元组损失、模态间三元组损失以及模态内三元组损失，通过组合方式在跨模态行人重识别数据集实验中证明了困难七元组损失函数的效果最好。本文首次将属性特征[11]加入跨模态行人重识别模型中进行训练，提高了模型的特征提取能力，并首次将 Focal Loss[12]损失函数应用于跨模态行人重识别方法中，以替代传统交叉熵损失函数，缓解了类别失衡问题。在 RegDB 数据集实验中，本文方法相比现有的跨模态行人重识别方法在各项指标中均取得了最优效果。

2　相关工作

2.1　单模态行人重识别

单模态行人重识别算法解决了在不重叠的有色光摄像机之间匹配 RGB 行人图像的问题。与易于从低分辨率图片中获取步态特征的识别算法[13]相比，单模态行人重识别对图片质量的要求较高，但低于人脸识别等小范围生物特征识别方法对图片质量的要求。单模态行人重识别算法主要分为设计手工特征方法、度量学习方法以及深度学习算法。设计手工特征方法主要是提取行人辨别性特征[11,14]，例如性别、头发、衣着等属性。度量学习方法旨在通过设计损失函数，通过训练卷积网络模型，让同一行人的不同图片特征向量之间的距离减小，让不同行人的图片特征向量之间的距离增大[4,12]。深度学习算法采用深度卷积网络以端到端的方式学习特征[9]。现有的单模态行人重识别方法迁移至跨模态行人重识别问题上无法保持原有效果，因此跨模态行人重识别问题亟待有效解决。

2.2　跨模态行人重识别

跨模态行人重识别主要包括可见光-深度跨模态行人重识别 (RGB-D Re-ID)[15]、可见光-热成像跨模态行人重识别 (VT-REID)[16]以及可见光-红外跨模态行人重识别 (RGB-IR Re-ID)[17]。RGB-D Re-ID 用于匹配 RGB 图像和深度图像间的人物信息，深度信息提供不变的身体形状和轮廓信息，并且不受光线和颜色变化的影响。VT-REID 和 RGB-IR Re-ID 用于匹配 RGB 图像和红外图像间的人物信息，不同的是 VT-REID 中的红外图像采用红外热成像技术来获取热成像图像信息，RGB-IR Re-ID 的红外图像采用红外摄像机主动发射红外光并通过自带的收集器收集返回的红外光获得相关的红外图像。跨模态行人重识别的研究较少。Wu 等[17]首次提出 Zero Padding 方法，并发布了 RGB-IR 数据集 SYSU-MM01。Nguyen 等[6]收集并提出 VT 数据集 RegDB。本文通过在 RegDB 数据集上进行实验，证明了本文方法的有效性。

3　本文方法

本文所提方法的整体框架如图 3 所示。整体框架包括深度卷积神经网络模块以及特征学习模块。与传统跨模态行人重识别方法的双流结构相比，本文使用单流结构利用卷积神经网络模型强大的特征提取能力同时提取两种模态嵌入特征。特征提取完成后，使用多重损失函数来减少模态间差异和模态内差异，保证模型学习到具有辨别性的特征。

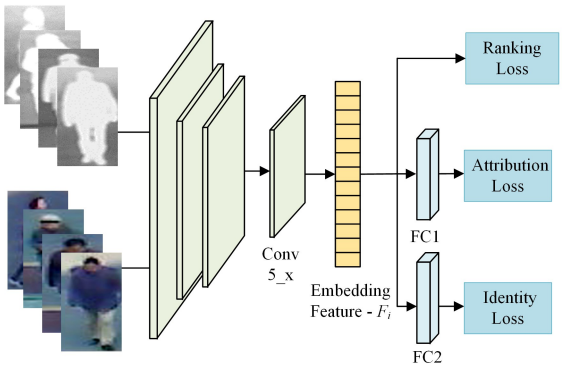


图 3 框架结构

Fig. 3 Framework structure

3.1 Ranking Loss

3.1.1 困难三元组损失函数

三元组损失函数(Triplet Loss)[4]是一种常见的排序损失(Ranking Loss),广泛应用于图像检索领域,比如人脸识别、行人重识别和车辆检索等。三元组损失函数不仅有减小类内距离的特性,还有增大类间距离的特性。这两种特性使得三元组损失函数相比身份损失函数更加适用于行人重识别网络训练。三元组损失函数的公式如下:

$$L_{Tri}=\sum_{i}^{N}[\parallel f(x_i^a)-f(x_i^p)\parallel_2^2-\parallel f(x_i^a)-f(x_i^n)\parallel_2^2+\alpha]_+ \tag{1}$$

其中,N 为一个 Batch 大小,Batch 中的每一张图片都会作为锚点从 Batch 中采样另外两张图片组成三元组,共有 N 个三元组;x_i^a 表示当前图片为锚点,x_i^p 和 x_i^n 分别表示与锚点图片身份相同的图片和与锚点图片身份不同的图片;$f(x_i^a)$ 为锚点图片的特征向量表示;α 为人为设定的阈值参数,当 x_i^a 与 x_i^n 之间欧氏距离的平方和以及 x_i^a 与 x_i^p 之间欧氏距离的平方之差小于 α 时,$[b]_+=b$;当其大于 α 时,$[b]_+=0$,其中 $[b]_+=\max(b,0)$。

Hermans 等[10]提出了困难三元组损失函数来对样本进行困难采样,与传统三元组损失相比,困难三元组损失提升了模型的训练速度和检索任务的准确率。为了提升训练速度,在训练集中随机采样 P 个行人类别,从每个类别中随机挑选 K 张图片,每一个 Batch 的行人图片数量为 PK。困难三元组损失函数公式如下:

$$L_{Htri}=\frac{1}{P*K}\sum_{i=1}^{P}\sum_{a=1}^{K}[\max_{p=1,\cdots,K}D(f(x_a^i),f(x_p^i))-\min_{\substack{j=1,\cdots,P\\ n=1,\cdots,K\\ j\neq i}}D(f(x_a^i),f(x_n^j))+\alpha]_+ \tag{2}$$

$$D(f(x_a^i),f(x_n^i))=\parallel f(x_a^i)-(x_n^i)\parallel_2 \tag{3}$$

其中,累加 PK 次是因为在训练中需要对 Batch 里的所有图片进行遍历,针对每一张图片寻找一个困难三元组,所以共有 PK 个困难三元组,困难三元组损失函数每次选取与锚点相比的最难(距离最远)正样本以及最难(距离最近)负样本作为困难三元组。由于每次训练时都专注于难样本,所以困难三元组损失函数会在训练速度以及模型精度上优于三元组损失函数。

3.1.2 改进方法

困难三元组损失思想适用于所有图像检索场景。然而在跨模态行人重识别中,如图 2 所示,与模态内差异相比,模态间差异对跨模态行人重识别模型的影响更大。跨模态行人重识别模型需要更加关注模态间变化,但是也不能忽略模态内变化。针对跨模态行人重识别数据集中不同模态行人图片差异巨大的特点,本文对困难三元组损失进行改进,提出了三组跨模态行人重识别三元组损失函数,分别是全局三元组损失、模态间三元组损失以及模态内三元组损失。

3.1.3 全局三元组损失

全局三元组损失是基于困难三元组损失的思想进行改进的。首先,由于跨模态行人重识别中存在多种模态,所以本文对 Batch 进行修改,每一个 Batch 随机采样 P 个行人类别,从每个行人类别中随机挑选 K 张可见光图片,K 张热成像图片,总共 $2PK$ 张图片。其次,将困难样本选择范围扩展至两个模态,全局三元组损失的公式如下:

$$L_{Gtri}=\frac{1}{2*P*K}\sum_{i=1}^{P}\sum_{a=1}^{2K}[\max_{p=1\cdots 2K}D(f(x_a^i),f(x_p^i))-\min_{\substack{j=1,\cdots P\\ n=1\cdots 2K\\ j\neq i}}D(f(x_a^i),f(x_n^j))+\alpha]_+ \tag{4}$$

其中,锚点图片 x_a^i 的选择范围是两种模态图片集合,x_p^i 为与锚点图片类别相同的同模态图片或跨模态图片,x_n^j 为与锚点图片类别不同的同模态图片或跨模态图片。

全局三元组损失的特点在于没有限定难样本图片模态,该损失训练模型的最终目的是将与锚点图片类别相同的所有模态的行人图片的特征向量之间的距离尽可能减小,与锚点图片类别不同的所有模态的行人图片的特征向量之间的距离尽可能增大。由式(4)可知,当锚点图片和与锚点图片类别相同的所有行人图片特征向量的最大距离加上阈值 α 小于与锚点图片类别不同的所有行人图片特征向量的最小距离时,锚点图片能与 Batch 中所有行人图片正确匹配,此时损失值为 0。当锚点和与锚点相同行人图片的最大距离加上 α 大于锚点和与锚点不同行人图片的最小距离时,说明模型还未完全收敛,需要计算损失值继续训练。

3.1.4 模态间三元组损失

模态间三元组损失相比全局三元组损失更注重模态间变化。在跨模态行人重识别问题中,模型需要更加关注模态间变化,因为模态间变化不能通过颜色、服装样式等进行区分,只能通过轮廓、边界等特征进行区分。因此本文提出模态间三元组损失,增加了模态间变化学习,将其作为全局三元组损失的补充。模态间三元组损失公式如下:

$$L_{Ctri}=\frac{1}{2*P*K}\sum_{i=1}^{P}\sum_{a=1}^{2K}[\max_{cp\in CB}D(f(x_a^i),f(x_{cp}^i))-\min_{\substack{j=1,\cdots,P\\ cn\in CB\\ j\neq i}}D(f(x_a^i),f(x_{cn}^j))+\alpha]_+ \tag{5}$$

其中,当 $a\leqslant K$ 时,$CB=\{K+1,K+2,\cdots,2K\}$;当 $a>K$ 时,$CB=\{1,2,\cdots,K\}$。图片 x_a^i 中 $a\in\{1,2,\cdots,K\}$ 表示锚点图片是可见光图片,$a\in\{K+1,K+2,\cdots,2K\}$ 表示锚点图片是热成像图片。x_{cp}^i 为与锚点图片类别相同的跨模态图片,x_{cn}^i 为与锚点图片类别不同的跨模态图片。。

3.1.5 模态内三元组损失

模态内三元组损失相比全局三元组损失更注重模态内变化,它是对全局三元组损失的补充。跨模态行人重识别中,虽然模态间差异会大于模态内差异,但是正确识别模态内差异

也是检验行人重识别算法的重要部分。如图 2(a)和图 2 (c)所示，同一行人相同模态的不同图片存在较大差异，不同行人相同模态的图片非常相似，这些都会影响模型的识别效果。因此本文提出模态内三元组损失，以增加模态内变化学习。模态内三元组损失的公式如下：

$$L_{Itri}=\frac{1}{2*P*K}\sum_{i=1}^{P}\sum_{a=1}^{2K}[\max_{ip\in IB}D(f(x_a^i),f(x_{ip}^i))-\min_{\substack{j=1\cdots P\\in\in IB\\j\neq i}}D(f(x_a^i),f(x_{in}^j))+\alpha]_+ \tag{6}$$

其中，当 $a\leqslant K$ 时，$IB=\{1,2,\cdots,K\}$；当 $a>K$ 时，$IB=\{K+1,K+2,\cdots,2K\}$。x_{ip}^i 为与锚点图片类别相同的相同模态图片，x_{in}^i 为与锚点图片类别不同的相同模态图片。

3.2　属性损失

Lin 等[11]提出将属性特征应用于行人重识别中指引行人重识别模型学习表征特征，并对传统可见光行人重识别数据集进行额外属性标注，最后通过实验证明了属性特征在单模态行人重识别方法中的有效性。由于跨模态行人重识别中不同模态行人图片间存在着巨大差异，数据集中原生特征信息无法让模型提取出区别于其他类别行人图片的行人特征。然而通过设计图片的属性特征可以在训练中丰富模型对行人图片的特征表征，提升模型的识别能力。基于属性特征在行人重识别模型训练中的作用，本文首次将属性特征用于跨模态行人重识别模型的训练中，将 RegDB 数据集中标注了行人性别的属性信息加入模型中进行训练。

本文在模型中增加了属性损失(Attribution Loss)模块，如图 3 所示。对于 Batch 中的单张图片 x_i，卷积网络输出图片嵌入特征向量 F_i，对特征向量 F_i 构建全连接层 FC1，用于计算两种性别属性得分 S_i，最后通过交叉熵损失函数计算 x_i 的属性损失 L_{SG_i}。本文对 Batch 中的所有图片计算属性损失并累加，得到属性损失 L_{AG}，具体公式如下：

$$L_{SG_i}=-\sum_{i=1}^{M}y_i\log\hat{p}_i \tag{7}$$

$$\hat{p}_i=softmax(S_i) \tag{8}$$

$$L_{AG}=\frac{1}{2*P*K}\sum_{i=1}^{2PK}L_{SG_i} \tag{9}$$

其中，y_i 为类别真实概率分布；$\hat{p}_i$ 表示经过 softmax 函数计算的 x_i 类别预测概率；M 为性别特征类别个数，取值为 2。

3.3　身份损失

身份损失(Identity Loss)被广泛应用于行人重识别网络训练，用于减少类内距离。现有跨模态行人重识别方法中，身份损失主要使用交叉熵损失，交叉熵损失通过提取特定分类信息减少类内变化，但是由于热成像图片提供的特征信息较少，跨模态行人重识别模型无法获取颜色特征，只能获取行人轮廓特征，导致其在分类过程中难以识别热成像人物图片，因此大部分热成像图片属于难样本，会影响交叉熵损失在训练阶段提升模型准确率的效率。

训练数据集中存在难易样本，而造成模型训练效果无法达到最优的原因就是类别失衡。在模型训练阶段，易分辨样本导致损失过小，造成模型学习效率降低，同时过多的易分辨样本会造成模型无法学习正确的特征，从而降低模型提取特征的能力。

针对类别失衡问题，传统算法如 OHEM 的处理方式是增加难分类样本的权重，忽略易分类样本。Lin 等[12]对传统类别失衡方法进行了改进，提出了 Focal Loss 来处理类别失衡问题。Focal Loss 通过改进交叉熵损失函数，降低但不消除易分类样本的权重，使模型在训练时更加专注于难样本的分类。

类别失衡问题在跨模态行人重识别模型的训练阶段非常普遍，导致模型训练低效，准确度不高。因此本文首次将 Focal Loss 引入跨模态行人重识别框架中，以替代传统交叉熵损失进行模型训练。Focal Loss 的公式如下：

$$L_{Foc}=\frac{1}{2*P*K}\sum_{i=1}^{2PK}\sum_{j=1}^{N}-(1-p_{ij})^{\gamma}\log(p_{ij}) \tag{10}$$

其中，$(1-p_{ij})^{\gamma}$ 为交叉熵损失函数调节因子；p_{ij} 表示 Batch 中第 i 张照片第 j 个类别对应正确匹配的概率；N 表示验证集类别总数；γ 是可调参数，本文统一设置为 2。

本文将排序损失、属性损失和身份损失融合到模型训练中，模型训练的损失函数公式如下：

$$L_{Ours}=L_{Gtri}+L_{Ctri}+L_{Itri}+L_{AG}+L_{Foc} \tag{11}$$

4　实验结果与分析

4.1　数据集描述以及评价指标

本文使用跨模态行人重识别数据集 RegDB 来评估所提方法。RegDB 数据集收集了在同一时刻下通过可见光相机以及热成像相机拍摄的人物图片信息。数据集共含有 412 个行人，每个行人有 10 张可见光图片以及对应的 10 张热成像图片。图片中存在正面拍摄以及背面拍摄的情况。由于拍摄期间存在人物移动情况，每张图片都会存在一定姿态、光线条件的差异。

Ye 等[5]重新整理了 RegDB 数据集，并提出一套公开验证方法，用于验证跨模态行人重识别方法。为了提升验证方法的可靠性，对数据集使用 10 折交叉验证方式，每次从 412 个行人中选择 206 个行人共 4 120 张图片作为训练集，其余 4 120 张图片用于验证，重复进行 10 次实验，并对实验数据取平均值作为单个实验的最终指标。

在实验验证阶段，Ye 等使用平均精度均值(mean Average Presicion，mAP)以及累计匹配曲线(Cumulative Match Characteristic curve，CMC 曲线)作为评价指标。mAP 指对全部待检索图片分别在验证集中计算得到的平均精度值进行求和后再取平均所得到的值，是用于衡量多标签图像分类的常见评价指标。平均精度(Average presicion，AP)通过单张待检索图片在验证集中正确匹配的精确率(Precision)获得。CMC 曲线用于评估行人重识别算法的性能，评价指标包括 rank1，rank10 等。mAP 的计算公式如下：

$$mAP=\frac{1}{N}\frac{1}{K_{q_i}}\sum_{i=1}^{N}\sum_{j=1}^{K_{q_i}}(\hat{r}_j/r_j) \tag{12}$$

其中，N 为待检索图片个数，K_{q_i} 为待检索图片 q_i 的 ID 在验证集中的出现个数，其中验证集记为 G。图片 q_i 提取特征后依次与 G 中的图片特征计算距离，按照升序排列并记为 $\hat{G}$，将 $\hat{G}$ 中与 q_i 相同 ID 的图片按照升序排列并另记为 $\hat{G}_{q_i}$，$\hat{r}_j$ 表示 ID 与 q_i 相同的图片 $\hat{g}_j$ 在 $\hat{G}_{q_i}$ 中的位置，r_j 表示 $\hat{g}_j$ 在 $\hat{G}$ 中的位置。Rank-n 表示 $\hat{G}$ 中前 n 个搜索结果中包含正确样本的概率。

本文采用 Ubuntu16.04 服务器训练跨模态行人重识别

模型，显卡为 NVIDIA GeForce 1080Ti，显存 11 GB。算法使用 Pytorch 实现。

在实验参数设置中，本文随机选择 P 个行人，每个行人随机选择 K 个可见光模态图片和 K 个热成像模态图片，一个 Batch 选择 $2PK$ 张图片。本文设定 $P=8$，$K=4$，总共 64 张图片。在训练阶段本文对图片进行预处理，包括随机水平翻转等。训练 Epoch 为 90，每一个结果需要进行十折交叉验证实验，最后取 10 个实验的平均值作为结果，以保证实验的可靠性。

4.2 对比其他跨模态行人重识别方法

本文选取多个已在 RegDB 数据集上评估有效的跨模态行人重识别算法。为了评估本文方法的性能，评估指标采用 mAP，以及 CMC 曲线中的 rank-1，rank-10 和 rank-20。

本文使用 Resnet50m 作为骨干网络，并使用式(11)对模型进行训练，实验结果如表 1 所列。可以看出，本文方法在 RegDB 数据集上比其他方法更优越。特别地，本文方法在两个指标上比最新提出的 AGAN 表现优异，并且与其他跨模态行人重识别方法(如 D-HSME，D2LR)相比，在 rank-1 以及 mAP 上分别至少有 8.95%和 12.4%的提升。

表 1 RegDB 数据集对比结果

Table 1 Comparison results of RegDB dataset

(单位：%)

方法	rank-1	rank-10	rank-20	*mAP*
GSM[18]	17.28	34.47	45.26	15.06
MLAPG[19]	17.82	40.29	49.73	18.03
XQDA[20]	21.94	45.05	55.73	21.80
HCML[5]	24.44	47.53	56.78	20.80
BDTR[21]	33.47	58.42	67.52	31.83
D2LR[22]	43.30	66.10	76.30	44.10
D-HSME[7]	50.85	73.76	81.66	47.00
AGAN[8]	57.90	—	—	53.60
Ours	**59.80**	**80.30**	**87.90**	**59.40**

为了进一步证明本文方法的有效性，本文使用 Grad-CAM 方法可视化本文模型提取的特征。Grad-CAM[23] 被广泛用于解释卷积网络模型对图片分类的机理，通过热力图的形式展现模型关注的辨别性特征。

由图 4 可以看出，相比于传统行人重识别模型关注颜色、纹理等特征，跨模态行人重识别模型更加注重行人轮廓特征以及在热成像图片中具有区分性的边界特征。本文认为，VT-REID 的关键是从热成像图片中获得可辨别性特征。因为可见光图片包含大量可辨别性特征，而热成像图片不存在颜色、图案等可辨别性特征。

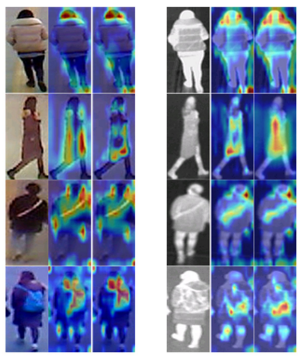


图 4 特征可视化

Fig. 4 Feature visualization

4.3 困难三元组损失改进方法评估

本文对传统困难三元组损失在跨模态上进行改进，提出了全局三元组损失、模态间三元组损失和模态内三元组损失。为了验证以上 3 种损失函数有效性，本文分别对传统困难三元组损失以及本文 3 种损失的组合方式进行实验。为了保证实验的一致性，本实验使用 Resnet50m 作为骨干网络，不添加属性损失，将身份损失设置为交叉熵损失，困难三元组损失函数 Batch 大小设置为 PK，其余损失函数 Batch 大小设置为 $2PK$，实验结果如表 2 所列。

表 2 改进方法的对比结果

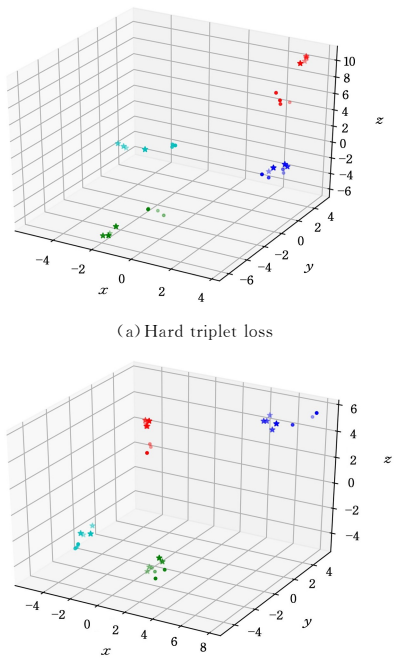
Table 2 Comparison results of improved method

(单位：%)

损失	rank-1	rank-10	rank-20	*m*AP
L_{Tri}	48.2	70.8	80.4	48.9
L_{Gtri}	49.5	71.0	79.6	48.6
$L_{Gtri}+L_{Itri}$	53.9	75.7	84.2	54.3
$L_{Gtri}+L_{Ctri}$	53.9	74.5	82.9	54.1
$L_{Gtri}+L_{Ctri}+L_{Itri}+$	**56.3**	**76.8**	**85.2**	**56.1**

实验结果表明，本文提出的 3 种三元组损失方法在跨模态行人重识别实验中相比传统困难三元组损失更具优势。基于全局三元组损失，作为补充的模态间三元组损失及模态内三元组损失也具有提升模型性能的效果。特别是当 3 个损失函数融合时，训练模型的效果最明显，本文将此损失函数称为困难七元组损失函数。

为了进一步证明困难七元组损失函数的作用，本文使用传统困难三元组损失函数和困难七元组损失函数分别训练模型，并将嵌入特征映射至 3 维空间。特征分布如图 5 所示，其中不同颜色代表不同身份的行人图片，两种不同形状的点表示两种模态。



(a) Hard triplet loss

(b) Hard heptaplet loss

图 5 损失函数对比

Fig. 5 Comparison of Loss Function

由图 5 可以看出，只使用 Hard Triplet Loss 损失时，类内

点不能很好地聚合，同时不同类别区分不明显。相比于图5(a)，使用困难七元组损失能使类内点聚合集中，同时不同类别点距离较远，存在明显区分。图5的实验结果证明，相比困难三元组损失，困难七元组损失在跨模态行人重识别中能够更有效地减小类内距离和增大类间距离。

4.4　属性损失评估

本文首次将属性特征用于跨模态行人重识别方法，以提升模型提取跨模态行人特征的能力。为了验证属性特征在跨模态行人重识别数据集中的作用，本文基于3.1节提出的排序损失改进方法，增加属性特征来验证属性损失的有效性。实验选用4.3节实验所使用的5种损失函数组合数据作为对比，以验证仅增加属性损失的效果，实验结果如图6所示。

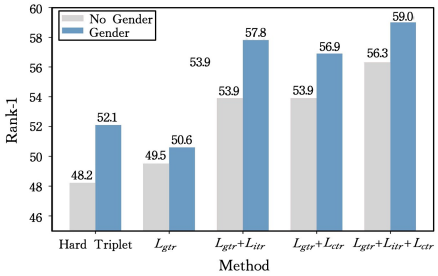


图6　属性损失对比图

Fig. 6　Comparison of attribute loss

由图6可以看出，属性损失在所有排序损失组合中均有提升模型特征提取能力的效果，证明了属性特征在跨模态行人重识别模型训练中的有效性。由于本文方法只利用跨模态图片中的性别特征进行学习，在未来工作中可以尝试在RegDB数据集中加入其他属性特征进行学习。

4.5　Focal Loss评估

Focal Loss用于缓解类别失衡问题，本文首次将Focal Loss用于跨模态行人重识别中，以替代传统交叉熵损失函数。为了证明Focal Loss的有效性，本文基于3.1节和3.2节的改进方法，仅通过修改七元组损失函数的超参数α，进行交叉熵损失和Focal Loss的对比实验，实验结果如图7所示。

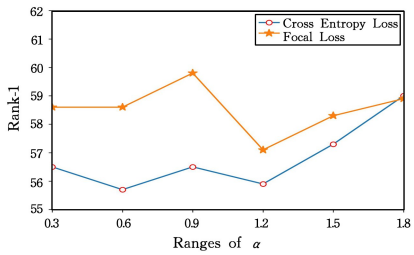


图7　Focal Loss对比图

Fig. 7　Comparison of Focal Loss

由图7可以看出，除了$\alpha=1.8$时Focal Loss的效果不明显外，Focal Loss在其余情况下的效果均优于交叉熵损失函数，特别是当$\alpha=0.9$时，跨模态行人重识别模型达到最好效果，并且在RegDB数据集上领先于现有的跨模态行人重识别算法。Focal Loss效果显著的原因在于，跨模态行人重识别数据集中存在类别失衡，由于热成像图片包含的信息较少，属于难样本，导致模型无法有效地识别热成像行人身份。实验结果证明，在3.1节和3.2节改进的基础上，Focal Loss还能提升跨模态行人重识别模型提取特征的能力。

4.6　消融实验

本文提出了3个改进方法，分别是将困难三元组损失改进为困难七元组损失、增加属性损失、使用Focal Loss替换身份损失。为了单独展现每种改进方法的效果，本文使用消融实验证明其有效性。

消融实验是深度学习研究中确定某种方法是否有效的最直接方式。消融实验本质上是通过删除部分网络结构来验证被删除部分对模型指标的影响。本文对提出的3种方法进行消融实验，每次剔除其中一种方法以查看效果。评价指标包括rank-1，rank-10，rank-20和mAP。实验结果如表3所列。

表3　消融实验结果

Table 3　Results of ablation experiment

Hepta Loss	Gender Loss	Focal Loss	rank-1/%	rank-10/%	rank-20/%	*mAP*/%
×	√	√	53.4	74.9	84.4	53.5
√	×	√	57.4	78.6	86.4	57.7
√	√	×	57.3	77.7	86.0	57.3
√	√	√	**59.8**	**80.3**	**87.9**	**59.4**

由表3可以看出，同时使用3种方法时，模型在处理跨模态行人重识别问题上具有最好效果。其他3个消融实验在各种指标上都略逊于3种方法同时使用的效果，证明了3种方法在跨模态行人重识别上具有提升模型特征提取的能力。另外，当不使用困难七元组损失函数时，模型准确度有较大程度的下降，说明在跨模态行人重识别中，困难七元组损失函数相比另外两种方法更能提升模型特征提取的能力。

结束语　针对跨模态行人重识别问题，本文提出了一种改进困难三元组损失的特征学习框架。考虑到跨模态行人重识别中行人图片的模态间变化和模态内变化，将困难三元组损失改造成困难七元组损失，并首次将属性特征加入跨模态行人重识别模型的训练中。基于跨模态行人重识别中类别失衡的问题，在VT-REID中首次使用Focal Loss替换传统交叉熵损失函数，将困难七元组损失、属性损失、Focal Loss相结合，加入跨模态行人重识别模型的训练中，并且在公开数据集RegDB上获得了比现有跨模态行人重识别算法更好的效果。

然而，本文方法仍存在一些不足，需要进一步探究：1)困难七元组损失函数在训练阶段存在冗余情况，会对其中一种模态样本重复提取计算损失；2)本文改进框架由于添加多重损失函数，在模型训练阶段耗时较长，但是在验证实验中的耗时与原方法差异不大，今后可以研究如何提高模型的训练效率；3)本文仅使用了性别属性特征，今后可以尝试加入多种属性特征进行学习。

参 考 文 献

[1]　SONG W R, ZHAO Q Q, CHEN C H, et al. Survey on pedestrian re-identification research[J]. CAAI transactions on intelligent systems, 2017, 12(6): 770-780.

[2]　ZHENG L, YANG Y, HAUPTMANN A G. Person re-identification: Past, present and future[J]. arXiv: 1610.02984, 2016.

[3]　MATSUKAWA T, SUZUKI E. Person re-identification using CNN features learned from combination of attributes[C]// 2016

23rd International Conference on Pattern Recognition (ICPR). Cancun: IEEE Press, 2016: 2428-2433.
[4] SCHROFF F, KALENICHENKO D, PHILBIN J. Facenet: A unified embedding for face recognition and clustering[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE Press, 2015: 815-823.
[5] YE M, LAN X, LI J, et al. Hierarchical discriminative learning for visible thermal person re-identification[C]// Thirty-Second AAAI Conference on Artificial Intelligence. New Orleans AAAI Press, 2018.
[6] NGUYEN D, HONG H, KIM K, et al. Person recognition system based on a combination of body images from visible light and thermal cameras[J]. Sensors, 2017, 17(3): 605.
[7] HAO Y, WANG N, LI J, et al. HSME: Hypersphere Manifold Embedding for Visible Thermal Person Re-Identification[C]// Proceedings of the AAAI Conference on Artificial Intelligence. Hawaii: AAAI Press, 2019, 33: 8385-8392.
[8] WANG G, ZHANG T, CHENG J, et al. RGB-Infrared Cross-Modality Person Re-Identification via Joint Pixel and Feature Alignment[C]// Proceedings of the IEEE International Conference on Computer Vision. Seoul: IEEE Press, 2019: 3623-3632.
[9] YU Q, CHANG X, SONG Y Z, et al. The devil is in the middle: Exploiting mid-level representations for cross-domain instance matching[J]. arXiv: 1711. 08106, 2017.
[10] HERMANS A, BEYER L, LEIBE B. In defense of the triplet loss for person re-identification[J]. arXiv: 1703. 07737, 2017.
[11] LIN Y, ZHENG L, ZHENG Z, et al. Improving person re-identification by attribute and identity learning[J]. Pattern Recognition, 2019, 95: 151-161.
[12] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]// Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE Press, 2017: 2980-2988.
[13] BEN X Y, XU S, WANG K J. Review on Pedestrian Gait Feature Expression and Recognition[J]. Pattern Recognition and Artificial Intelligence, 2012, 25(1): 71-81.
[14] BAO Z M, GONG S R, ZHONG S, et al. Person Re-identification Algorithm Based on Bidirectional KNN Ranking Optimization [J]. Computer Science, 2019, 46 (11): 267-271.
[15] WU A, ZHENG W S, LAI J H. Robust depth-based person re-identification[J]. IEEE Transactions on Image Processing, 2017, 26(6): 2588-2603.
[16] YE M, WANG Z, LAN X, et al. Visible Thermal Person Re-Identification via Dual-Constrained Top-Ranking[C]// the 27th International Joint Conference on Artificial Intelligence. Stockholm: Morgan Kaufmann Press, 2018: 1092-1099.
[17] WU A, ZHENG W S, YU H X, et al. Rgb-infrared cross-modality person re-identification[C]// Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE Press, 2017: 5380-5389.
[18] LIN L, WANG G, ZUO W, et al. Cross-domain visual matching via generalized similarity measure and feature learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(6): 1089-1102.
[19] LIAO S, LI S Z. Efficient psd constrained asymmetric metric learning for person re-identification[C]// Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE Press, 2015: 3685-3693.
[20] LIAO S, HU Y, ZHU X, et al. Person re-identification by local maximal occurrence representation and metric learning[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Santiago: IEEE Press, 2015: 2197-2206.
[21] YE M, WANG Z, LAN X, et al. Visible Thermal Person Re-Identification via Dual-Constrained Top-Ranking[C]// the 27th International Joint Conference on Artificial Intelligence. Stockholm: Morgan Kaufmann Press, 2018: 1092-1099.
[22] WANG Z, WANG Z, ZHENG Y, et al. Learning to reduce dual-level discrepancy for infrared-visible person re-identification [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE Press, 2019: 618-626.
[23] SELVARAJU R R, COFSWELL M, DAS A, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization[C]// Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE Press, 2017: 618-626.

LI Hao, born in 1995, postgraduate. His main research interests include deep learning and person reidentification.

ZHAO Yun-bo, born in 1981, Ph.D, professor. His main research interests include networked control systems, AI enabled automation, human-machine integrated intelligence, and systems biology.