

基于特征学习的双路径红外-可见光行人重识别算法

朱松豪, 吕址函, 宋 杰

(南京邮电大学 自动化学院 人工智能学院, 江苏 南京 210023)

摘要:由于拍摄视角、行人姿态的变化以及不同的相机光谱造成的额外跨模态差异,RGB 图像和红外图像之间存在着明显的差异,提取有效的模态共享特征是红外-可见光行人重识别中的难点。本研究提出一种双路径学习算法来识别特征,利用改进的 BNNeck 模块来提取 RGB 和红外图像的特征信息,改善算法的识别性能。该算法首先将注意力机制引入双路径特征学习网络,获取 RGB 图像在空间维度和通道维度上的特征信息,实现红外特征信息匹配;然后,将 BNNeck 模块引入至跨模态行人重识别算法,减少模态特征信息差异,加快算法收敛速度;最后,在异质中心损失函数和交叉熵损失函数的基础上,引入跨模态下行人身份损失函数,提高行人识别的准确性。SYSU-MM01 和 RegDB 数据集的实验结果表明,相对于目前大多数已有算法,所提算法具有更好的泛化能力和鲁棒性,Rank-1/mAP 分别达到 59.39%/85.44% 和 57.81%/73.19%,比最新算法分别提高 2.43%/2.86% 和 2.44%/1.19%。

关键词:行人重识别;跨模态;注意力机制;双路径网络;模态共享

中图分类号:TP391.41

文献标志码:A

Dual-path infrared-visible person re-identification algorithm based on feature learning

ZHU Songhao, LÜ Zhihan, SONG Jie

(College of Automation and College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu 210023, China)

Abstract: Due to the changes of shooting angle, pedestrian posture and additional cross-modality differences caused by different camera spectra, there are obvious discrepancy between RGB images and infrared images. A difficulty in infrared-visible person re-identification is how to extract effective modal shared features. A novel dual-path algorithm was proposed to learn recognition features and an improved BNNeck module was utilized to extract the feature information of RGB and infrared images, which can improve the recognition performance of the whole algorithm. Firstly, the attention mechanism was introduced into the dual-path feature learning network to obtain the feature information of RGB images in spatial and channel dimensions, which achieved infrared feature information matching. Then, the BNNeck module was introduced into the cross-modality person re-identification algorithm to reduce the cross-modality differences and accelerate the convergence speed of the algorithm. Finally, on the basis of hetero-center loss function and cross entropy loss function, the pedestrian identity loss function under different modes was introduced to improve the accuracy of person re-identification. The experimental results of SYSU-MM01 and RegDB datasets show that compared with most of the existing algorithms, the proposed algorithm has better generalization ability and robustness, with the mAP/Rank-1 reaching 59.39%/85.44% and 57.81%/73.19% respectively, 2.43%/2.86% and 2.44%/1.19% higher than the latest algorithm respectively.

Key words: person re-identification; cross-modality; attention mechanism; dual-path network; modal share

收稿日期:2021-05-10

基金项目:国家自然科学基金青年科学基金项目(62001247)

作者简介:朱松豪(1973—),男,江苏如皋人,副教授,博士,主要从事深度学习、人工智能等方面研究。

E-mail: zhush@njupt.edu.cn

行人重识别是指利用计算机视觉技术来确定一个目标行人是否在图像或视频中存在。近年来,随着社会对公共安全问题的关注越来越多,行人重识别引起了人们极大的研究兴趣。许多行人重识别算法被提出,包括手工特征算法^[1]、度量特征算法^[2]以及深度学习算法^[3],这些算法一定程度上解决了行人重识别问题。深度学习算法目前已成为行人重识别的主流算法,包括单模态和多模态行人重识别算法。

现有的大部分研究^[4-6]都专注于 RGB 模态中的行人重识别,主要集中在 RGB 模态中外部条件变化的挑战,包括光照条件、视点变化、不对称等方面。文献[4]利用排序优化框架中的相似性和差异性线索进行行人识别。为提高行人重识别的精度,文献[5]研究了跨模态支持一致性和跨模态投影一致性的算法。Chen 等^[6]利用自然语言描述作为有效的视特征,识别不同身份的额外训练监督。然而,现有的单模态行人重识别算法只能解决 RGB 图像中的识别问题,不能很好地解决跨模态行人重识别问题。

如图 1 所示,虽然图 1(a)和 1(c)是不同的行人图像,但是图 1(a)所对应的红外图像图 1(b)与图 1(c)非常相似,说明拍摄角度、行人姿态和光照变化等因素严重影响对红外-可见光行人重识别的准确性。双模态摄像机很好地弥补了单模态 RGB 摄像机的缺陷,并且将单模态行人重识别研究延伸至跨模态领域,目前已提出多种针对跨模态行人重识别的方法。文献[7]提出 SYSY-MM01 大规模跨模态行人重识别数据集,得出零填充有利于提高识别精度的结论。文献[8]引入一种分层跨模态解调算法,有效减少模态间差异。文献[9]提出双向中心约束顶级排序算法,通过引入模态间约束和模态内约束,刻画尺寸变形和类内变化。文献[10]通过在网络中添加 X 模态,辅助学习不同模态特征,提高识别性能。

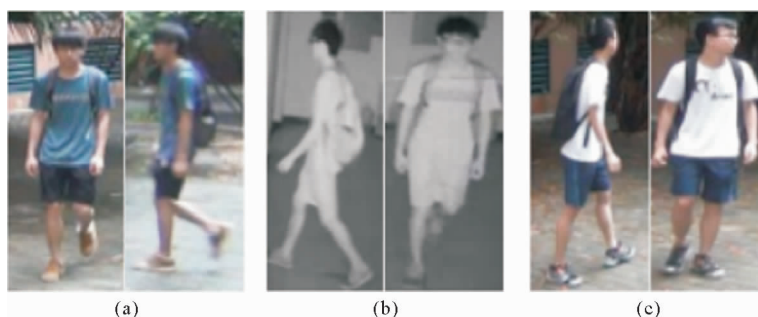


图 1 红外-可见光行人重识别中复杂的变化

Fig. 1 Complex variations in infrared-visible person re-identification

上述算法着重于缩小 RGB 与红外模态间差异,在模态信息不共享情况下,识别精度不太理想。本研究提出一种基于特征学习的双路径红外-可见光行人重识别算法,在引入注意力机制刻画模态内和模态间特征差异的同时,利用身份损失函数、身份异质中心损失函数和交叉熵损失函数,提高行人重识别准确性。

1 基于特征学习的双路径红外-可见光行人重识别算法

本研究提出的红外-可见光行人重识别的双路径学习算法,通过在 RGB 分支中引入注意机制来获得更多行人信息,在特征嵌入阶段引入 BNNeck(batch normalization neck)模块加快身份损失收敛速度,其双路径网络结构如图 2 所示。

1.1 双路径局部特征网络

在 RGB 单模态行人图像重识别中,常用算法是对行人图像进行水平分割,提取局部特征,然后对行人图像进行特征匹配。由于红外图像只有行人的外观信息和姿态信息,没有颜色和光照等信息,故不能解决红外-可见光行人重识别问题。本研究采用的双路特征学习算法主要包括特征提取和特征嵌入两部分。

1) 特征提取。首先,利用卷积注意力机制提取 RGB 图像特征,并结合红外图像特征构建特征描述子;然后,水平切割具有高层语义信息的红外-RGB 行人特征,并将其作为共享层的输入。由于训练数据有限,网络预训练参数通过大规模数据集 ImageNet 训练得到。

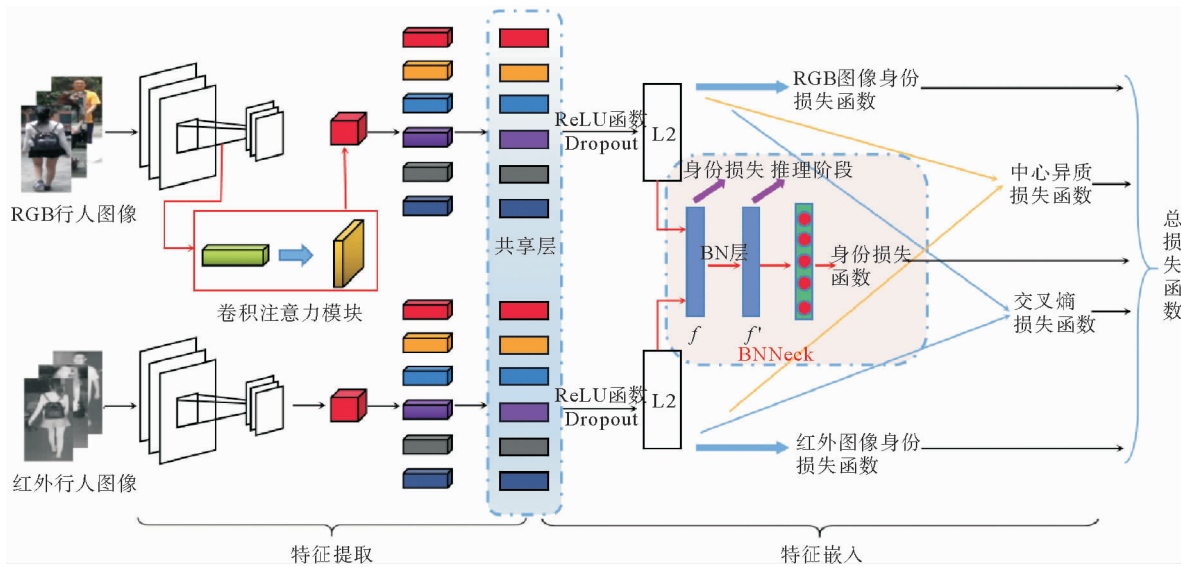


图2 本研究算法的双路径网络结构

Fig. 2 Architecture of the proposed dual-path network

2) 特征嵌入。首先,为更好学习模态间可区分的低维嵌入特征,将提取的特征引入共享层;然后,使用L2正则化方法平滑特征表示;最后,利用提出的BNNeck身份损失函数,以及RGB图像身份损失函数、红外图像身份损失函数、交叉熵、损失函数和中心异质损失函数,得到准确的行人身份。

为减少模态间差异,本算法采用操作:①为表征不同模态的低级特征,RGB特征网络和红外特征网络的ResNet50的第二个卷积块参数不完全相同;②为表征模态的中层特征,每个特征网络共享深层卷积块参数;③在自适应池化卷积层后,使用共享的BN(batch normalization)层学习共享特征。

1.2 注意力机制

在RGB分支中加入CBAM(convolution block attention module)注意力机制,学习不同图像区域权值获得更有鉴别性的信息。

1) 通道注意力模块。不同场景的不同通道刻画不同的重要性,通过引入通道注意力机制选择不同通道。如图3所示,通道注意力机制以RGB图像特征作为输入,先使用全局平均池化和最大池化聚合特征图的空间信息,生成空间上下文描述图 F_{av}^c 和 F_{max}^c ;再将两个描述图通过两个全连接层做进一步处理并相加,得到通道注意力 $M_c \in i^{c \times 1 \times 1}$ (c 为通道数):

$$M_c = \sigma(W_1(W_0(F_{av}^c)) + W_1(W_0(F_{max}^c))) \quad (1)$$

式中, σ 表示sigmoid函数, W_0 和 W_1 分别为两个全连接层的参数。该通道注意力机制能够较好地把握RGB图像的特征信息,减小RGB图像与红外图像的差异,可视为不同的半动态语境选择语义属性的过程。

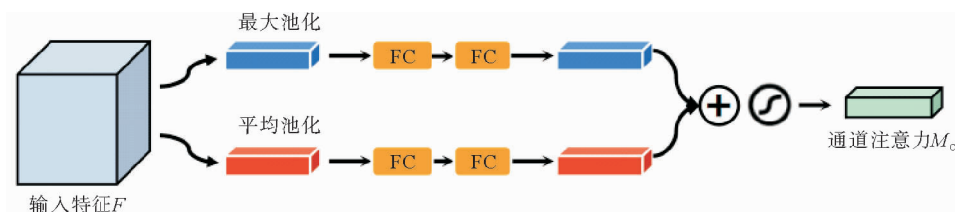


图3 通道注意力模块结构

Fig. 3 Channel attention module structure

2) 空间注意力模块。由于摄像机的角度、视野等原因,行人图像会呈现远近大小不一问题,导致算法无法准确定位行人。由于通道注意力机制忽略了行人在图像中的空间结构,导致行人位置信息不够精准,影响

了图像语义描述的准确性。为刻画行人空间结构信息,引入空间注意力机制,弥补通道注意力机制的不足。空间注意力机制关注的是“在哪里”,利用特征图空间信息的内在关系,构建聚焦于信息丰富的空间注意力图,实现行人准确定位。如图4所示,空间注意力机制先在通道维度上使用全局平均池化和最大池化聚合特征图的通道信息,生成特征描述图 $F_{av}^S \in i^{1 \times H \times W}$ 和 $F_{max}^S \in i^{1 \times H \times W}$ (H 、 W 分别为图像的高度和宽度);再使用 concat 连接两幅特征描述图;最后,使用 sigmoid 函数做进一步处理,生成空间注意力机制

$$M_s = \sigma(\varphi_{conv}^{7 \times 7}(F_{av}^S; F_{max}^S))。 \quad (2)$$

式中, σ 表示 sigmoid 函数, $\varphi_{conv}^{7 \times 7}$ 表示卷积核大小为 7×7 的卷积操作。

红外行人图像是一个只包含语义结构和形状信息的单通道图像^[1],而 RGB 图像是一个包括外观和颜色等高级语义信息的多通道图像。为避免在特征提取中损失重要信息,尽量获得更多行人信息,减少模态间差异,本算法在 RGB 分支加入注意力机制(如图5所示)进行特征提取。具体的特征提取过程为:①在输入 $288 \times 144 \times 3$ 的 RGB 行人图像后,利用注意力机制,针对 Conv5_x 和平均池化层间的高层特征进行加权,得到 $18 \times 9 \times 2048$ 的特征映射,有效扩大感受野范围;②利用提出的平直流行结构,融合 $18 \times 9 \times 2048$ 的特征映射和加权后的特征映射,加强全局跨通道交互。

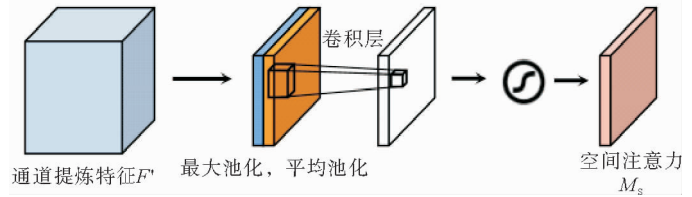


图4 空间注意力模块结构

Fig. 4 Spatial attention module structure

①在输入 $288 \times 144 \times 3$ 的 RGB 行人图像后,利用注意力机制,针对 Conv5_x 和平均池化层间的高层特征进行加权,得到 $18 \times 9 \times 2048$ 的特征映射,有效扩大感受野范围;②利用提出的平直流行结构,融合 $18 \times 9 \times 2048$ 的特征映射和加权后的特征映射,加强全局跨通道交互。

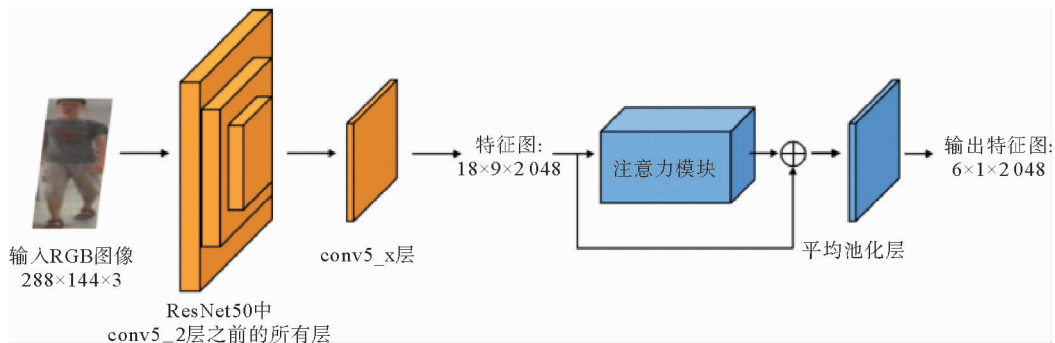


图5 基于注意力机制的特征提取过程示意图

Fig. 5 Feature extraction process based on attention mechanism

基于注意力机制提取的特征映射,为后续处理提供包含重要信息和次要信息的全局特征,有助于从模态层面把握 RGB 图像的特征信息,减少两种模态之间的差异。

1.3 跨模态 BNNeck 和损失函数

1.3.1 概述

本研究利用 RGB 图像丰富的信息来学习模态特征,缩小模态间的差异。如图2所示,利用特定模态的网络和身份损失提取模态相关信息,并在 RGB 分支中引入卷积注意力模块,对行人图像进行特征提取,使其更具有判别特征;将 BNNeck 模块与模态共享层相结合,提高深度模态共享的协调性。得到的最终总损失函数为:

$$L = L_{id}^V + L_{id}^I + L_{id}^{BNNeck} + \lambda_1 L_{HC} + \lambda_2 L_{CE}。 \quad (3)$$

式中: L_{id}^V 和 L_{id}^I 分别代表 RGB 和红外分支上的身份损失, L_{id}^{BNNeck} 为跨模态的共同身份损失, L_{HC} 为跨模态中心异质损失, L_{CE} 为交叉熵损失, λ_1 和 λ_2 为惩罚项。

1.3.2 身份损失

1) 特定模态身份损失。由于 RGB 图像和红外图像中行人的视觉特征差别很大,需要在提取每个分支的特征时使用不同的权重来学习每个模态的单一网络。每个网络分支采用特定的身份损失来学习其特征,

本研究采用 Softmax 损失进行身份预测,表示为:

$$\begin{cases} L_{id}^V = - \sum_{n=1}^{N^V} \lg \frac{e^{(\mathbf{w}_{y_n^V}^V)^T x_n^V + b_{y_n^V}^V}}{\sum_{m=1}^M e^{(\mathbf{w}_m^V)^T x_n^V + b_m^V}}, \\ L_{id}^I = - \sum_{n=1}^{N^I} \lg \frac{e^{(\mathbf{w}_{y_n^I}^I)^T x_n^I + b_{y_n^I}^I}}{\sum_{m=1}^M e^{(\mathbf{w}_m^I)^T x_n^I + b_m^I}}. \end{cases} \quad (4)$$

式中: x_n^V 和 x_n^I 分别表示属于 y_n^V 和 y_n^I 类的第 n 个 RGB 和红外图片的特征; \mathbf{W}_m^V 和 \mathbf{W}_m^I 表示上一全连通层中权重为 \mathbf{W}^V 和 \mathbf{W}^I 的第 m 列; b^V 和 b^I 是指特定模态的偏置, M 表示总的身份个数; N^V 和 N^I 表示在一个批次中训练样本的数量。

为更好提取 RGB 图像的特征并学习其全局不变性,在 RGB 分支中加入注意力机制,将式(4)改写为:

$$\begin{cases} L_{id}^V = - \sum_{n=1}^{N^V} \lg \frac{e^{(\mathbf{w}_{y_n^V}^V)^T x_n^V + b_{y_n^V}^V}}{\sum_{m=1}^M e^{(\mathbf{w}_m^V)^T x_n^V + b_m^V}} + \lambda L_c, \\ L_{id}^I = - \sum_{n=1}^{N^I} \lg \frac{e^{(\mathbf{w}_{y_n^I}^I)^T x_n^I + b_{y_n^I}^I}}{\sum_{m=1}^M e^{(\mathbf{w}_m^I)^T x_n^I + b_m^I}}. \end{cases} \quad (5)$$

则卷积注意力模块在 RGB 分支特征提取中造成的偏置。

$$L_c = \frac{1}{\omega \times h} \sum_{i=1}^{\omega} \sum_{j=1}^h \| X_c(i, j) - X'(i, j) \|. \quad (6)$$

式中: ω 和 h 分别表示图像的宽度和高度, λ 用来限制损失。

2) 跨模态 BNNeck 身份损失。身份损失将识别任务视为一个分类任务,并通过余弦距离优化身份损失。如果直接将两个模态的特征相结合并计算身份损失,很可能导致两种模态下的目标行人身份不一致,损失函数无法收敛。为解决这一问题,在双分支网络的全链接层后引入名为跨模态 BNNeck 的模块,在共享全连接分类层后增加一个 BN 层。从网络中提取的原始特征为 f , 经过 BN 层后变为 f' 。在训练阶段,分别使用 f 和 f' 计算身份损失,并通过 BN 层得到正则化后的特征。BNNeck 减少了身份损失的限制,使其更容易收敛,同时,正则化进一步缩小了同一行人 RGB 图像特征和红外图像特征间的差异。跨模态 BNNeck 身份损失公式为:

$$L_{id}^{BNNeck} = \frac{1}{2N} \sum_N \frac{1}{K_i^V K_i^I} \sum_{p=1}^{K_i^V} \sum_{q=1}^{K_i^I} \| x_{ip}^V - x_{iq}^I \|^2. \quad (7)$$

式中: $1 \leq i \leq N$, $1 \leq p \leq K_i^V$, $1 \leq q \leq K_i^I$, K_i^V 和 K_i^I 分别表示不同模态下第 i 个行人的身份, x_{ip}^V 和 x_{iq}^I 分别表示不同模态下第 i 个行人的第 p 个和第 q 个身份。

1.3.3 中心异质损失和交叉熵损失

为缩小内在差异,引入中心异质损失和交叉熵损失。中心异质损失公式为

$$\begin{cases} L_{HC} = \sum_{i=1}^U (\| c_{i,1} - c_{i,2} \|^2), \\ c_{i,1} = \frac{1}{M} \sum_{j=1}^M x_{i,1,j}, \\ c_{i,2} = \frac{1}{N} \sum_{j=1}^N x_{i,2,j}. \end{cases} \quad (8)$$

式中: $c_{i,1}$ 和 $c_{i,2}$ 分别为第 i 个行人的 RGB 模态和红外模态特征分布中心, U 为行人个数, M 和 N 为第 i 个行人的 RGB 图像和红外图像数, $x_{i,1,j}$ 和 $x_{i,2,j}$ 分别为第 i 个行人的第 j 个 RGB 图像和红外图像的特征。

由于中心异质损失仅通过约束每个行人的中心距离提高模态内相似度,并不能通过网络学习判别特征

来扩大模态间差异。因此,采用中心异质损失和交叉熵损失联合监管,交叉熵损失函数为:

$$L_{CE} = - \sum_{i=1}^K \lg \frac{e^{w_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^n e^{w_j^T x_i + b_j}} \quad (9)$$

式中: K 为批次数, x_i 是第 y_i 类中第 i 个特征, w_j 是 W 中第 j 行参数, b 表示模态的偏置。

2 实验

实验环境为 Ubuntu 16.04, GPU 选用 NVIDIA GeForce RTX 2070, 深度学习框架 Pytorch 的版本为 1.0.0。输入图像大小设置为 288×144 。为提升训练和测试精度,采用随机裁剪和随机水平翻转对数据集进行扩充, batch size 的大小为 32, 设置总训练周期为 60 个 epoch。

2.1 数据集

本研究选用公开的数据集 SYSU-MM01^[5] 和 RegDB^[11], 使用首位排序精度(Rank-1, R1)、前 10 位排序精度(Rank-10, R10)、前 20 位排序精度(Rank-20, R20)和平均精度(mean average precision, mAP)共同对算法性能进行评估。

SYSU-MM01 共有 491 位行人, 其中 395 个行人用于训练, 96 个行人用于测试; 同时, 22 258 张 RGB 图像和 11 909 张红外图像用于训练, 301 张 RGB 图像和 3 803 张红外图像用于测试。由于单镜头全景搜索方式更接近现实, 因此采用该搜索方式进行性能评估。

RegDB 共有 412 位行人, 其中 206 个行人用于训练, 206 个行人用于测试。测试过程中, RGB 图像作为查询示例, 红外图像作为图库集用以验证算法性能。

2.2 SYSU-MM01 数据集上的测试结果

与本研究所提算法进行性能比较的算法包括方向梯度直方图(histogram of oriented gradient, HOG)等在内的传统算法以及包括 cmGAN 等在内的深度学习算法共 14 种, 从全景搜索/室内搜索和单镜头/多镜头两个角度进行评估, 4 个方面的比较数据见表 1。可以看出, 室内搜索的性能优于全景搜索的性能, 这是因为室内环境较为单一, 无明显光线变化, 行人匹配更容易; 全搜索下, 多镜头 R1 的准确率高于单镜头 R1, 但多镜头 R1 的 mAP 低于单镜头排序 R1, 这是因为多镜头可以从各个角度收集更多的行人照片, 更容易进行匹配, 而单镜头模式则相反。

表 1 SYSU-MM01 数据集的实验对比

Tab. 1 Comparison with other works on SYSU-MM01 dataset

算法	全景搜索								室内搜索							
	单镜头				多镜头				单镜头				多镜头			
	R1	R10	R20	mAP	R1	R10	R20	mAP	R1	R10	R20	mAP	R1	R10	R20	mAP
HOG ^[12]	2.76	18.30	32.00	4.24	3.82	22.80	37.70	2.16	3.22	24.70	44.30	7.25	4.75	29.10	49.40	3.51
LOMO ^[13]	3.64	23.20	37.30	4.53	4.70	28.30	43.10	2.28	5.75	34.40	54.90	10.20	7.36	40.40	60.40	5.64
Two-Stream ^[7]	11.70	48.00	65.50	12.90	16.40	58.40	74.50	8.03	15.60	61.20	81.10	21.50	22.5	72.30	88.70	14.00
One-Stream ^[7]	12.10	49.70	66.80	13.70	16.30	58.20	75.10	8.59	17.00	63.60	82.10	23.00	22.70	71.80	87.90	15.10
Zero-padding ^[7]	14.08	54.12	71.33	15.95	19.13	61.40	78.41	10.89	20.58	68.38	85.79	26.92	24.43	75.86	91.32	18.64
HCML ^[8]	14.32	53.16	69.17	16.16	—	—	—	—	—	—	—	—	—	—	—	—
BDTR ^[14]	17.01	55.43	71.69	19.66	—	—	—	—	—	—	—	—	—	—	—	—
cmGAN ^[15]	26.97	67.51	80.56	27.80	31.49	72.74	85.01	22.27	31.63	77.23	89.18	42.19	37.00	80.94	92.11	32.76
D ² RL ^[11]	28.90	70.60	82.40	29.20	—	—	—	—	—	—	—	—	—	—	—	—
DGD+MSR ^[3]	37.35	83.40	93.34	38.11	43.86	86.94	95.68	30.48	39.64	89.29	97.66	50.88	46.56	93.57	98.80	40.08
DPMBN ^[16]	37.02	79.46	89.87	40.28	—	—	—	—	44.47	87.12	95.24	54.51	—	—	—	—
X Modality ^[10]	49.92	89.79	95.96	50.73	—	—	—	—	—	—	—	—	—	—	—	—
DDAG ^[17]	54.75	90.39	95.81	53.02	—	—	—	—	61.02	94.06	98.41	67.98	—	—	—	—
TSLFN+HC ^[18]	56.96	91.50	96.82	54.95	62.09	93.74	97.85	48.02	59.74	92.07	96.22	64.91	69.76	95.85	98.90	57.81
本算法	59.39	93.35	97.89	57.81	65.64	95.57	98.70	50.92	61.80	95.27	98.58	68.54	72.34	95.66	99.57	61.62

由表1中可见,HOG/LOMO在R1和mAP中分别只有2.76%/3.64%和4.24%/4.53%,说明传统算法难以在跨行人重识别中取得良好效果;单流(One-Stream)、双流(Two-Stream)和零填充(Zero-Padding)将R1和mAP的性能提高了近8%,说明分类损失有助于特征的学习;在R1和mAP的识别性能上,相比于零填充,cmGAN和D²RL的性能均提高了近10%,说明通过网络优化和训练,算法的有效性得到了很大提高;DDAG和TSLFN+HC的R1和mAP均达55%左右,同时R10和R20也分别高出1.85%和1.07%,这证明了双路径网络的有效性。综上,本研究算法在R1和mAP分别达到59.39%和57.81%,R10和R20分别达到93.35%和97.89%,证明了算法的有效性。

2.3 RegDB数据集上的测试结果

在数据集RegDB上使用visible2 thermal搜索方式,以RGB图像作为搜索图像库,红外图像作为待搜索图像库,与其他9个算法的测试结果对比见表2。

由表2可见,本算法的R1和mAP分别为85.44%和73.19%,比X Modality算法的相应值分别高出23.23%和13.01%,比性能最好的TSLFN+HC算法的相应值分别高出1.44%和1.19%;同时R10和R20也达到了最高性能。在RegDB数据集上,本算法的R1和mAP优于其他算法。

2.4 注意力机制的可视化

为直观显示注意力机制的性能,比较SYSU-MM01数据集上的红外图像的热力图,如图6所示。由图6可以看出,卷积注意力机制能够很好地保留行人的特征(如行人姿态),在RGB图像中与在红外图像中保持相同的身份,表明融合注意力机制可以使网络更好地提取行人特征,提高识别精度;与DSCSN+CCN^[19]和RNPR^[20]两种算法相比,RNPR算法因只能识别一小部分行人信息而降低了行人重新识别的准确性,而本研究的注意力机制可以更准确地提取行人特征,更好地保留行人特征信息。

2.5 各模块消融实验

本节评估本算法在全搜索和室内搜索方式SYSU-MM01数据集上每个模块的有效性(如表3),并评估RegDB数据集上每个模块的有效性(如表4)。表中,D表示双路径网络的基线网络结果,B表示BNNeck模块,C表示添加到RGB分支中的CBAM注意力机制。

由表3和表4可知:①双路径网络具有明显优势,卷积块共享双路径网络在跨模态行人重识别中可以获得优良的性能。②B的有效性。在两个数据集上,BNNeck模块(D+B)的加入显著提高了算法的性能。③C的有效性。当在网络中加入CBAM注意模块(D+C)时,通过注意力机制将两种模态之间的差异进一步缩小,从而提高算法的性能。④两个模块的有效性。将BNNeck模块和CBAM模块合并到网络中(D+B+C),整体性能进一步提高,说明两个模块相互配合,共同促进。

表2 RegDB数据集的实验对比

Tab. 2 Comparison with other works on RegDB dataset

算法	R1	R10	R20	mAP
One-Stream ^[7]	13.11	32.98	42.51	14.2
Two-Stream ^[7]	12.43	30.36	40.96	13.42
Zero-Padding ^[7]	17.75	34.21	44.35	18.90
HCML ^[8]	24.44	47.53	56.78	20.08
BDTR ^[14]	33.41	58.42	67.52	31.83
D ² RL ^[11]	43.40	66.10	76.30	44.10
DGD+MSR ^[3]	48.43	70.32	79.95	48.67
X Modality ^[10]	62.21	83.13	91.72	60.18
TSLFN+HC ^[18]	83.00	—	—	72.00
本算法	85.44	96.17	98.20	73.19

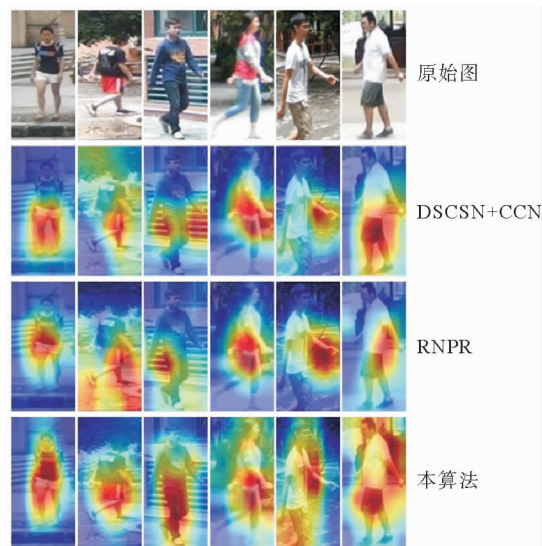


图6 不同算法的注意力机制热力图

Fig. 6 Heat map of attention mechanism for different algorithms

表 3 各模块在 SYSU-MM01 数据集下的实验结果

Tab. 3 Evaluation of each component on the large-scale SYSU-MM01 dataset

算法	全景单镜头搜索(SYSU-MM01)				室内单镜头搜索(SYSU-MM01)			
	R1	R10	R20	mAP	R1	R10	R20	mAP
D	49.14	88.71	95.25	48.00	52.04	89.78	95.69	59.83
D+B	57.83	92.62	97.33	56.06	58.40	93.96	98.08	66.10
D+C	57.59	92.81	97.62	56.50	60.11	95.21	98.66	67.82
D+B+C	59.39	93.35	97.89	57.81	61.80	95.27	98.58	68.54

2.6 与基准损失函数的比较

将损失函数与基准损失函数进行比较,验证所提算法的性能。比较 4 种不同的损失函数:Softmax 损失、结合图注意力机制的 Softmax 损失^[17]、异中心损失以及交叉熵损失,实验结果见表 5。

可以看到,与其他基线损失函数相比,本研究提出算法的性能最好。图注意力机制的 Softmax 函数比单一的 Softmax 函数好,说明图注意力机制对于红外-可见光行人重识别性能的重要性。交叉熵的异中心损失增加了模态内相似性,改善了性能。在使用特定模态 Softmax 和模态共享 Softmax 进行扩展后,在两个数据集上的识别性能均得到了提高。

表 4 各模块在 RegDB 数据集下的实验结果

Tab. 4 Evaluation of each component on the RegDB dataset

算法	R1	R10	R20	mAP
D	80.78	94.90	97.57	69.71
D+B	85.19	96.55	98.59	73.02
D+C	85.39	96.60	98.45	73.08
D+B+C	85.44	96.79	98.20	73.19

表 5 双路径网络下,本算法与基准算法间不同损失函数的性能比较

Tab. 5 Performance comparison of different loss functions between the proposed algorithm and the baseline algorithm under the dual-path network

损失函数类型	SYSU-MM01(全景单镜头搜索)				RegDB			
	R1	R10	R20	mAP	R1	R10	R20	mAP
Softmax	47.50	—	—	47.65	70.05	—	—	66.37
Softmax+图注意力	54.75	90.39	95.81	53.02	69.34	86.19	91.49	63.46
交叉熵+中心异质损失	59.39	91.50	96.82	57.81	83.00	—	—	72.00
Softmax+交叉熵+中心异质损失	59.39	93.35	97.89	57.81	85.44	—	—	73.19

2.7 收敛速度测试

数据集 SYSU-MM01 上的训练阶段损失变化趋势如图 7 所示,可见由于网络参数较少,同时采用端到端的设计,使得训练过程中部分参数能够及时得到释放,避免参数量过高的问题,因而使得本算法模型规模小,训练周期短,收敛速度快。

3 结论

本研究提出一种双路径深度学习算法,通过双路径分别提取 RGB 图像和红外图像特征,获得不同行人的身份信息。引入注意力机制提取 RGB 图像

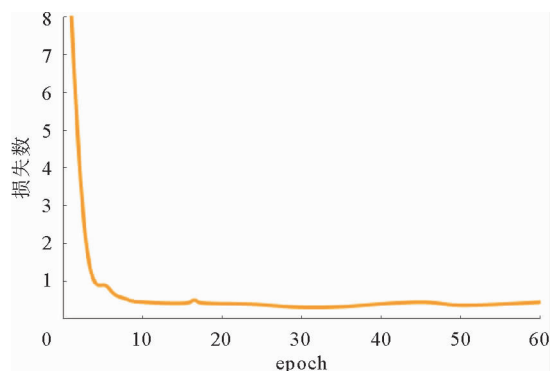


图 7 训练阶段损失变化趋势

Fig. 7 Trend of loss during training

特征,通过共享层减少 RGB 图像与红外图像差异,并通过融合跨模态 BNNeck 得到共享身份信息;通过多个损失函数交互作用,有效提高算法性能。本算法模型规模小、训练周期短,能有效减小红外图像与 RGB 图像模态间差异,算法识别精度优于现有算法。但由于数据集较少,训练度不够,本算法仅适用于训练周期较短的识别任务。

参考文献:

- [1]LIAO S C,LI S.Efficient PSD constrained asymmetric metric learning for person re-identification[C]//IEEE International Conference on Computer Vision,Santiago,Dec.13-16,2015;3685-3693.
- [2]YI D,LEI Z,LIAO S C,et al.Deep metric learning for person re-identification[C]//IEEE International Conference on Pattern Recognition,Stockholm,Aug. 24-28,2014;34-39.
- [3]FENG Z X,LAI J H,XIE X H.Learning modality-specific representations for visible-infrared person re-identification[J].IEEE Transactions on Image Processing,2020,29;579-590.
- [4]YE M,LIANG C,YU Y,et al.Person re-identification via ranking aggregation of similarity pulling and dissimilarity pushing [J].IEEE Transactions on Multimedia,2016,18(12):2553-2566.
- [5]WANG Z,HU R M,LIANG C,et al.Zero-shot person re-identification via cross-view consistency[J].IEEE Transactions on Multimedia,2016,18(2):260-272.
- [6]CHEN D P,LI H S,LIU X H,et al.Improving deep visual representation for person re-identification by global and local image-language association[C]//European Conference on Computer Vision,PartXVI.Munich,Sep. 8-14,2018;56-73.
- [7]WU A,ZHENG W S,YU H X,et al.RGB-infrared cross-modality person re-identification[C]//IEEE International Conference on Computer Vision,Venice,Oct. 22-29,2017;5390-5399.
- [8]CHOI S,LEE S,KIM H,et al.Hi-CMD:hierarchical cross-Modality disentanglement for visible-infrared person re-identification[C]//IEEE Conference on Computer Vision and Pattern Recognition.Seattle,Jun. 16-18,2020;10254-10263.
- [9]YE M,LAN X Y,WANG Z,et al.Bi-directional center-constrained top-ranking for visible thermal person re-identification [J].IEEE Transactions on Information Forensics and Security,2020,15;407-419.
- [10]LI D G,WEI X,HONG X P,et al.Infrared-visible cross-modal person re-identification with an X modality[C]//The Association for the Advance of Artificial Intelligence.New York,Feb. 7-12,2020;4610-4617.
- [11]WANG Z X,WANG Z,ZHENG Y Q,et al.Learning to reduce dual-level discrepancy for infrared-visible person re-identification[C]//IEEE Conference on Computer Vision and Pattern Recognition.Los Angeles,Jun. 15-20,2019;618-626.
- [12]DALAL N,TRIGGS B.Histograms of oriented gradients for human detection[C]//IEEE Conference on Computer Vision and Pattern Recognition.San Diego,Jun. 20-26,2005;886-893.
- [13]LIAO S,HU Y,ZHU X Y,et al.Person re-identification by local maximal occurrence representation and metric learning[C]//IEEE Conference on Computer Vision and Pattern Recognition.Boston,Jun. 7-12,2015;2197-2206.
- [14]ZHANG Z Z,LAN C L,ZENG W J,et al.Relation-aware global attention for person re-identification[C]//IEEE Conference on Computer Vision and Pattern Recognition.Seattle,Jun. 14-19,2020;3183-3192.
- [15]DAI P Y,JI R R,WANG H B,et al.Cross-modality person re-identification with generative adversarial training[C]//27th International Joint Conference on Artificial Intelligence.Stockholm,Jul. 13-19,2018;677-683.
- [16]XIANG X,LÜ N,YU Z,et al.Cross-modality person re-identification based on dual-path multi-branch network[J].IEEE Sensors Journal,2019,19(23);11706-11713.
- [17]YE M,SHEN J B,DAVID J,et al.Dynamic dual-attentive aggregation learning for visible-infrared person re-identification [C]//16th European Conference on Computer Vision.Glasgow,Aug. 23-28,2020;229-247.
- [18]ZHU Y,YANG Z,WANG L,et al.Hetero-center loss for cross-modality person re-identification[J].Neurocomputing, 2020,386;97-109.
- [19]ZHANG S Z,YANG Y F,WANG P,et al.Attend to the difference:Cross-modality person re-identification via contrastive correlation[J].IEEE Transactions on Image Processing,2021,30;8861-8872.
- [20]PARK H,HAM B.Relation network for person re-identification[C]//34th The Association for the Advance of Artificial Intelligence.New York,Feb. 7-12,2020;11839-11847.

(责任编辑:傅 游)