计算机科学
COMPUTER SCIENCE

基于向量注意力机制 GoogLeNet-GMP 的行人重识别方法

孟月波，穆思蓉，刘光辉，徐胜军，韩九强

引用本文

孟月波，穆思蓉，刘光辉，徐胜军，韩九强. 基于向量注意力机制GoogLeNet-GMP的行人重识别方法[J]. 计算机科学，2022, 49(7): 142-147.

MENG Yue-bo, MU Si-rong, LIU Guang-hui, XU Sheng-jun, HAN Jiu-qiang. Person Re-identification Method Based on GoogLeNet-GMP Based on Vector Attention Mechanism [J]. Computer Science, 2022, 49(7): 142-147.

Software Self-admitted Technical Debt Identification with Bidirectional Gate Recurrent Unit and Attention Mechanism

计算机科学, 2022, 49(7): 212-219. https://doi.org/10.11896/jsjkx.210500075

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.210600198

基于向量注意力机制 GoogLeNet-GMP 的行人重识别方法

孟月波[1,2]　穆思蓉[1]　刘光辉[1]　徐胜军[1,2]　韩九强[1,2]
1 西安建筑科技大学信息与控制工程学院　西安 710055
2 人工智能与数字经济广东省实验室(华南理工大学)　广州 510000
(mengyuebo@163.com)

摘　要　为了提高行人重识别(Re-ID)的准确率和适用性,提出了一种基于向量注意力机制 GoogLeNet 的 Re-ID 方法。首先,将 3 组图像(锚、正、负)输入到 GoogLeNet-GMP 网络中,获得分段式特征向量。然后,利用空间金字塔池化(Spatial Pyramid Pooling,SPP)对来自不同金字塔等级的特征进行聚合,并引入注意力机制,通过对代表目标视觉信息的多尺度池化区域进行整合,获得多个语义等级上的可区分性特征。同时,将两个不同损失函数的混合形式作为最终损失函数。在 Market-15012 和 Duke-MTMC3 数据集上进行实验,结果表明,相比其他优秀方法,所提方法在 Rank-1 和 mAP 指标方面表现更优。

关键词:行人重识别;注意力机制;GoogLeNet;空间金字塔池化;损失函数
中图法分类号　TP391

Person Re-identification Method Based on GoogLeNet-GMP Based on Vector Attention Mechanism

MENG Yue-bo[1,2],MU Si-rong[1],LIU Guang-hui[1],XU Sheng-jun[1,2] and HAN Jiu-qiang[1,2]
1 School of Information and Control Engineering,Xi'an University of Architecture and Technology,Xi'an,710055,China
2 Artificial Intelligence and Digital Economy Guangdong Provincial Laboratory,South China University of Technology,Guangzhou 510000,China

Abstract　In order to improve the accuracy and applicability of person re-identification(Re-ID),a Re-ID method based on vector attention mechanism GoogLeNet is proposed. Firstly,three groups of images(anchor,positive and negative) are input into the GoogLeNet-GMP network to obtain segmented feature vectors. Then,spatial pyramid pooling(SPP) is used to aggregate the features from different pyramid levels,and attention mechanism is introduced. By integrating the multi-scale pooling regions which represent the visual information of the target,the distinguishable features on multiple semantic levels are obtained. At the same time,the mixed form of two different loss functions is taken as the final loss function. Experiments on Market-15012 and Duke-MTMC3 data set show that the proposed method performs better in Rank-1 and mAP indicators than other excellent methods.

Keywords　Person re-identification,Attention mechanism,GoogLeNet,Spatial pyramid pooling,Loss function

1　引言

行人重识别(Re-IDentification,Re-ID)[1]是对不同时刻、不同相机获得的同一个目标进行关联的任务,是典型的图像视频问题以及图像检索的子问题。Re-ID 在视频监控、交通安全等领域具有重要应用,是一个热门的研究课题,其目前存在视角变化、背景混淆和遮挡等问题[2]。

已有一些研究成果致力于解决行人重识别中的难题,这些研究一般可以分为特征表征方法[2]、最优匹配度量方法[3]以及深度学习方法[4]。

特征表征方法通过设计区分性外观特征描述子来解决 Re-ID 问题。其可以在以分块匹配策略[5]、显著性学习[6]以及相机网络为导向的方案中结合多个局部特征和全局特征。此类方法一般结构较为简单、容易实现,其缺点是固有特征表示难以捕捉非常多的细节特征。

最优匹配度量方法通过学习最优的非欧氏相异性度量来增强 Re-ID。例如,利用零空间或低秩方案[7],通过强化正半定条件来引入度量学习,捕捉相似性空间的全局结构,通过度量学习来增加人体相似度度量的精确性。也有一些方法将相似性空间的全局结构和局部结构相结合[8],以增强最优匹配。

到稿日期:2021-06-24　返修日期:2021-12-26
基金项目:国家自然科学基金面上项目(51678470);陕西省自然科学基金面上项目(2020JM-473,2020JM-472);西安建筑科技大学基础研究基金(JC1703);西安建筑科技大学自然科学基金(ZR19046)
This work was supported by the National Natural Science Foundation of China(51678470),Nature Science Foundation of Shaanxi,China(2020JM-473,2020JM-472),Natural Science Basic Research of Xi'an University of Architecture and Technology(JC1703) and Natural Science Foundation of Xi'an University of Architecture and Technology(ZR19046).
通信作者:穆思蓉(m_srong0413@163.com)

深度学习方法是基于深度学习的解决方案，通过深度学习得到的表征中包含高度区分性信息。文献[9]提出了基于人体姿态估计的密集神经网络，采用了基于区域的网络特征。但姿态估计和 Re-ID 问题之间存在概念分歧，无法确保检测到的人体部位对于 Re-ID 任务来说是最优的。为此，一些方法舍弃了人体分区估计器，转而考虑固定人体部分。文献[10]使用生成对抗网络无监督学习方法来生成无标签图像，然后将生成的图像与原数据集联合后进行半监督卷积神经网络训练。构建一个 Siamese 网络，结合分类模型和验证模型的特点进行训练。文献[11]使用深度卷积生成对抗网络来生成额外的无标签行人图片，并采用标签平滑正则化来处理这些新生成的无标签行人图片，提出了一种新的无监督重排序框架。文献[12]提出了一种称为 KDAS-ReID 的人体重识别方法，该方法基于知识特征提取，采用了 ResNet-50 框架。文献[13]将图像切块折叠为通过身份损失优化的单个紧凑表征，略去了 Re-ID 过程中单个切块的重要程度，但该方法仅考虑了骨干网络的输出，且未使用注意力机制。

本文尝试利用金字塔方法[14]解决上述难题，构建了一个 Siamese 架构，并考虑了水平方向的人体切块的特征以及从不同语义等级提取出的信息。其主要创新总结如下：1）通过提取不同细节等级（即不同网络深度）的信息，捕捉属于输入的不同语义概念；利用从架构的不同深度得到的特征，提供特定语义概念的信息；2）对来自不同等级的细节点进行聚合，通过注意力机制，从每个多级表征中选择最优 Re-ID 特征，从而获得更优的表征信息。

2 基于 GoogLeNet-GMP 的特征提取

相关研究证明，若包含接近输入层与接近输出层之间的连接较短，则可对神经网络进行更具深度、更准确的高效训练。本文以 GoogLeNet 架构[15]为基础，利用由每个架构组成的计算块来生成结果，将这些输出考虑为基本特征，以提取与 Re-ID 最相关的信息。

GoogLeNet 可以解决深度卷积神经网络的过拟合问题。不同于一般卷积神经网络，GoogLeNet 引入 1×1 卷积，采用了 Inception 模块，其具体结构如图 1 所示。其优点是将稀疏矩阵转变为较为密集的子矩阵，减少了参数的数量，节省了计算开销，提高了运行速度。Inception 有多个不同的版本。其中，Inception v3 较为常用，该模块将二维卷积进行非对称拆分，获得了两个较小的卷积核，增加了整个网络的非线性空间特征。因此，本文采用 Inception v3 结构。

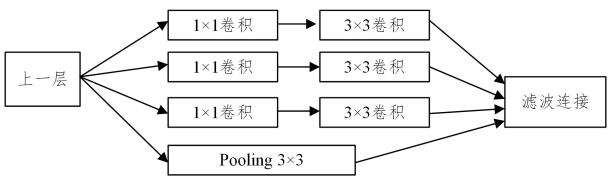

图 1 GoogLeNet 中的 Inception 结构

Fig. 1 Inception construction in GoogLeNet

为了获得优质的语义相关特征，本文的 GoogLeNet 网络设计参考了全局最大池化（Global Maximum Pooling，GMP）网络，在 Inception v3 的基础上，增加了 GMP 层。其具体网络参数如表 1 所列。此外，使用 Sigmoid 全连接（Full Connection，FC）层替换原有的稀疏 FC 层。GoogleNet-GMP 与分段特征提取结构如图 2 所示。

表 1 本文 GoogLeNet-GMP 的结构参数

Table 1 Structure parameters of GoogLeNet-GMP in this paper

类型	大小/步长	尺寸
Conv-1	3×3/2	299×299×3
Conv-2	3×3/2	149×149×32
Pool	3×3/2	147×147×64
Conv-3	3×3/1	73×73×64
Conv-4	3×3/2	73×73×80
Pool	3×3/2	71×71×192
Inception v3	—	35×35×192
Inception×2	—	35×35×288
Inception×5	—	17×17×768
Inception v3	—	17×17×768
GMP	—	8×8×2048
Sigmiod	分类	1×1×2048

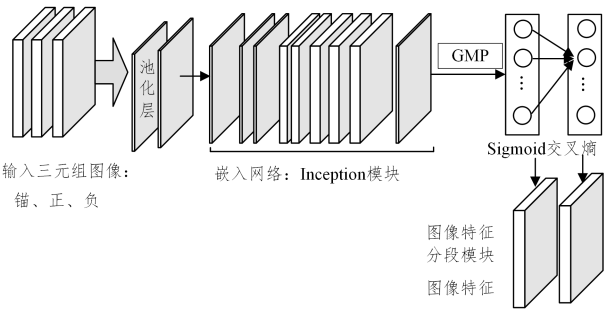

图 2 GoogLeNet-GMP 与分段特征提取

Fig. 2 GoogLeNet-GMP and segmented feature extraction

图 2 中的锚、正、负元组分别用 $(\boldsymbol{I}^a, y^a)$，$(\boldsymbol{I}^p, y^p)$ 和 $(\boldsymbol{I}^n, y^n)$ 表示，其中 y 表示目标对象标识。选择 3 个元组，使 $y^a = y^p$，$y^a \neq y^n$。

GoogLeNet-GMP 生成与所考虑的深度等级具有语义相关性的特征（即第一层对应于颜色、边和角，更深的层则由更复杂和抽象的特征来激活[16]）。因此，对共享特定特征的目标进行检索时，不同网络层的重要程度不同。

3 网络模块与学习

本文方法的整体框架如图 3 所示。

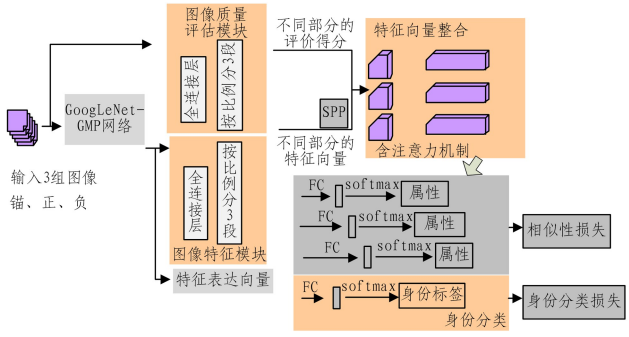

图 3 本文方法整体框架示意图

Fig. 3 Overall framework diagram of the proposed method

3.1 质量评估模块

文献[17]提了出一种较为鲁棒可靠的方法来评估行人每帧图像的质量，本文采用类似方法。由于人类是直立行走的，约束了图像特征的相对位移，例如腿部特征不应在躯干之上。根据人类身体部位的构造，将目标行人按照 3∶2∶2 的比例进行分割操作，如图 4 所示。根据已有方法，将所有的目标图像输入到两个不同的模块中。一个模块用于输出图像表征向量，该模块是一个完整的神经网络；另一个模块用于评估分割后图像的 3 个部分。经过这两个模块之后，将得到的特征向量按照之前的分割方法分为上、中、下 3 个部分，每个部分的特征向量对应一个评估分数。令 $Seq=\{fr_1, fr_2, \cdots, fr_n\}$ 表示某个指定行人的视频序列；$\boldsymbol{f}_u(fr_i)$ 表示目标行人第 i 帧的上部分特征向量，$\boldsymbol{f}_m(fr_i)$ 表示目标行人第 i 帧的中部分特征向量；$\boldsymbol{f}_l(fr_i)$ 表示目标行人第 i 帧的下部分特征向量；$\mu_u(fr_i)$，$\mu_m(fr_i)$ 和 $\mu_l(fr_i)$ 分别表示 3 个不同部分的评估分数，该分数可以按照比例归一化到区间[0,1]。最终特征表示形式为[17]：

$$F_{\text{part}}(Seq)=\sum_{i=1}^{n}\mu_{\text{part}}(fr_i)f_{\text{part}}(fr_i) \tag{1}$$

其中，$F_{\text{part}}(Seq)$ 表示给定视频序列的区域特征，该区域特征分为上、中、下 3 个区域；视频序列的帧数为 n；评估分数满足如下条件：

$$\sum_{i=1}^{n}\mu_{\text{part}}(fr_i)=1 \tag{2}$$

图 4 行人按比例分割示意图

Fig. 4 Pedestrian segmentation diagram in proportion

3.2 对特征执行注意力机制

通常情况下，仅从单一尺寸进行分析可能会造成其他尺寸信息的丢失，因此从不同尺度对图片进行分析非常重要。图像金字塔方法提供了灵活的多分辨率框架，用于捕捉不同尺寸的对象和特征。因此，本文架构充分利用了多尺度特征属性。

通过 GoogLeNet-GMP，在不同尺度等级生成的输出向量，提供了多个隐性语义特征表征，利用了每个空间金字塔池化[18]机制；为每个特征执行注意力机制。但由于该处理被单独应用到每个细节等级，因此无法确保通过这些特征的比较能得到最优的整体表征。虽然高级语义信息可能对 Re-ID 任务更有用，但其他网络层也可能携带有助于 Re-ID 任务的相关信息。为了捕捉此类特征信息，并利用不同等级的语义特征，本文引入了额外的注意力步骤，对不同细节等级处生成的表征进行联合分析和加权。如图 5 所示，为了对不同细节等级处生成的金字塔池化表征进行加权，通过残差网络块的初始应用，生成注意力向量。然后，通过全连接层，生成一个与所有金字塔池化特征的串联相同大小的向量。最后，应用批标准化(Batch Normalization，BN)和 Sigmoid 函数将向量数值归一化到[0,1]中，将加权向量表示为 $\boldsymbol{w}=[\boldsymbol{w}_1, \cdots, \boldsymbol{w}_L]^{\mathrm{T}}$，其中每个 $\boldsymbol{w}_l$ 也可以通过金字塔等级和切块来表示，然后进行如下计算：

$$\hat{\boldsymbol{x}}_l^{p,i}=\boldsymbol{x}_l^{p,i}\odot\boldsymbol{w}_l^{p,i} \tag{3}$$

其中，$\odot$ 为 hadamark 积。在式(3)中对所有 $p\in\{1,\cdots,m_l\}$ 和 $i\in\{1,\cdots,p\}$ 进行评估，得：

$$\hat{\boldsymbol{x}}_l=[\boldsymbol{x}_l^{p,i}\odot\boldsymbol{w}_l^{p,i}\mid p\in\{1,\cdots,m_l\}\wedge i\in\{1,\cdots,p\}]^{\mathrm{T}} \tag{4}$$

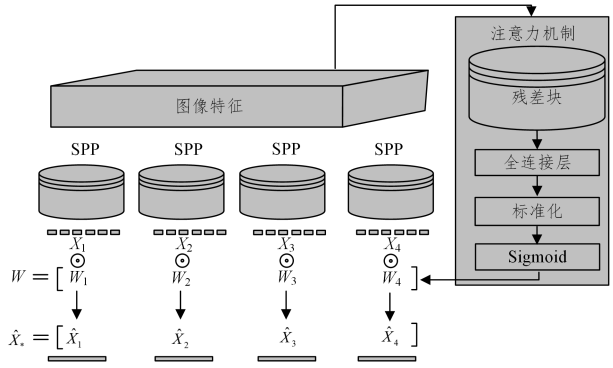

图 5 本文的注意力机制

Fig. 5 The proposed attention mechanism

由此生成纠缠态图像表征，该特征考虑了所有注意力加权的金字塔等级特征以及不同骨干块的输出。该纠缠态图像表征为 $\hat{\boldsymbol{x}}_*=[\hat{\boldsymbol{x}}_1, \cdots, \hat{\boldsymbol{x}}_L]^{\mathrm{T}}$。

3.3 基于混合方案的学习策略

很多方法利用身份损失对当前架构进行微调，学习到更具稳健性的表征信息。此外，也可以通过 Siamese 架构来分析图像的成对三元组，以学习相似性度量[19]。本文将这两种方案合并，形成一个联合损失。

具体地，通过从不同骨干块生成切块特征，来计算特定的身份损失。由此，网络单独考虑从不同图像切块中得到的特征，学习对输入数据的分类。同时，通过所有块和每个骨干块的特征串联得到表征 $\hat{\boldsymbol{x}}_*$，以进行相似性学习，从而提供有区分性的图像表征。

(1)身份损失。给定骨干 blocks，利用 $p\in\{1,\cdots,m_l\}$，$i\in\{1,\cdots,p\}$ 和生成的每个 $\hat{\boldsymbol{x}}_l^{p,i}$ 反馈到 FC 层，获得身份标签 $\hat{y}_l^{p,i}$。将其与真实标签 y 一起考虑，以计算交叉熵损失 $L_{\text{ce}}(\hat{y}_l^{p,i}, y)$。将所有骨干 blocks、金字塔等级和切块一起考虑，则网络优化的身份损失定义为：

$$L_{id}(\boldsymbol{I})=\sum_{l=1}^{L}\sum_{p=1}^{m_l}\sum_{i=1}^{p}L_{\text{ce}}(\hat{y}_l^{p,i}, y) \tag{5}$$

(2)相似性损失。锚、正、负图像三元组的纠缠态图像表征之间的边界排序损失(Boundary Sorting Loss，BSL)为：

$$L_{\text{sim}}(\boldsymbol{I}^a, \boldsymbol{I}^p, \boldsymbol{I}^n)=\max(\|\hat{\boldsymbol{x}}_*^a-\hat{\boldsymbol{x}}_*^p\|-\|\hat{\boldsymbol{x}}_*^a-\hat{\boldsymbol{x}}_*^n\|+\alpha, 0) \tag{6}$$

上述公式中的损失函数确保了与负样本相比，正样本的表征与锚样本的距离更接近至少一个余量 α。

(3)混合损失。合并上述两个损失函数，得到混合损失为：

$$L=\lambda L_{\text{sim}}(\boldsymbol{I}^a,\boldsymbol{I}^p,\boldsymbol{I}^n)+(1-\lambda)/3\sum_{\pi\in\{a,p,n\}}L_{\text{id}}(\boldsymbol{I}^\pi) \tag{7}$$

其中，π 表示三元组的锚、正、负元素，$\lambda\in[0,1]$控制了身份损失与相似性损失之间的权衡。

4 实验结果与分析

4.1 实验设置与数据集

本文方法在配置为 intel 双核 i5 的个人台式机上进行，采用 pytorch 实现，GPU 为 GTX 1080。在单个 GPU 上利用 SGD 优化器执行 100 代的训练。初始学习率设为 0.1，每 40 代降低至原先的 1/10。对所有网络层应用 0.9 的动量和 0.0005 的权重衰减惩罚。在两个行人重识别基准数据集即 Market-15012[20] 和 Duke-MTMC3[21] 上进行了实验评估，部分样例图像如图 6 所示。使用 Rank-1 和平均精度均值(mAP)指标来评估性能，其中 Rank-5 和 Rank-10 用于参考评估，Rank-1 和 mAP 用于主评估。

(1)Market-15012 数据集包含 6 部不同相机捕捉到的 1501 个行人的 32668 张图像。利用先进探测器得到同一个人的多张图像。其中训练集和测试集分别包含 750 和 751 个行人标识。涉及对应行人的图像作为一个短视频存于数据集中，训练集包含 12936 张图像，测试集分别包含 19732 张图库图像和 3368 张查询图像。

(2)Duke-MTMC3 数据集通过 8 部相机来捕捉图像，并人工标注了包围框。该数据集共涉及 1812 个行人，其中 1404 个行人在一部以上相机捕捉的图像中出现。将涉及对应行人的图像作为一个短视频存于数据集中，训练集和测试集各包含 702 个行人。训练集包含 16522 张图像，测试集分别包含 17661 张图库图像和 2228 张查询图像。

(a)Market-15012 数据集上的部分样例图像

(b)Duke-MTMC3 数据集上的部分样例图像

图 6 数据集样例图像

Fig. 6 Sample images in datasets

4.2 方法比较分析

在 Market-15012 数据集和 Duke-MTMC3 数据集上的测试结果比较分别如表 2 和表 3 所列。可以看出，*Rank*-1 准确度低于 *Rank*-5 准确度，*Rank*-5 准确度低于 *Rank*-10 准确度，这是因为取 5 个最大概率的类别的准确度一般高于取 1 个最大概率的类别的准确度，以此类推。由表 2 和表 3 可知，在 Market-15012 数据集中，所提方法的 *Rank*-1 准确度至少提高了 4.2%，*mAP* 至少提高了 8.6%；在 Duke-MTMC3 数据集中，所提方法的 *Rank*-1 准确度至少提高了 1.6%，*mAP* 至少提高了 10.3%。文献[9]采用密集网络，仅考虑了骨干网络的输出，未使用注意力机制，因此难以捕捉不同细节等级(包括语义和视觉)的图像相关性。文献[10]的单尺度分析会造成其他尺度的相关信息缺失。文献[12]采用 ResNet 框架，其在捕捉图像特征的相对位移的相关信息方面，弱于所提方法。本文方法明显提高了人体重识别的准确度，这主要得益于所提方法的多尺度金字塔表征能够捕捉不同的相关图像细节，且注意力机制能有效选择对 Re-ID 最合适的多尺度表征，有效地合并了来自所有网络层的信息，从而明显地提升了性能。

表 2 Market-15012 数据集上的测试结果比较

Table 2 Comparison of test results on Market-15012 dataset

(单位：%)

方法	*Rank*-1	*Rank*-5	*Rank*-10	*mAP*
文献[9]方法	91.8	96.1	97.8	78.9
文献[10]方法	89.0	93.6	97.3	73.2
文献[12]方法	92.1	96.3	97.9	79.1
本文方法	93.2	96.5	97.9	87.5

表 3 Duke-MTMC3 数据集上的测试结果比较

Table 3 Comparison of test results on Duke-MTMC3 dataset

(单位：%)

方法	*Rank*-1	*Rank*-5	*Rank*-10	*mAP*
文献[9]方法	84.1	90.7	91.3	80.9
文献[10]方法	73.9	86.1	87.9	73.1
文献[12]方法	83.5	90.5	92.3	81.9
本文方法	85.7	91.9	93.7	91.2

图 7 给出了所提方法人体重识别的部分结果，其中共有 4 行、11 列的图像，第 1 列图像用于查询，后几列图像是查询后的结果。正确结果采用绿色边框标记，错误结果采用黄色边框标记。可以看出，所提方法可以较好地适应人体姿态变化和物体遮挡，验证了所提方法的有效性。这表明本文方法能够利用有限的数据，学习有意义的特征表征和相似性度量，获得了良好的重识别结果。

图 7 部分行人重识别的结果展示(电子版为彩图)

Fig. 7 Results of partial pedestrian re-identification

4.3 消融实验分析

为了分析不同网络模块和一些关键损失函数对 Re-ID 性能的影响，本文以 Market-15012 数据集为对象进行消融实验，测试不同消融模块的平均准确率。

消融实验主要包括 5 个不同的部分：特征提取模块、注意力机制、身份损失函数、相似性损失函数和分类模块。由于分类模块对于所有网络框架都是必要的，因此需保留。在消融实验中，分别删除特征提取模块、去除注意力机制、移除身份损失函数、移除相似性损失函数，然后进行不同实验，其结果如表 4 所列。由表 4 可以看出，当移除其中一项时，与完整网络相比，mAP 均有一定程度的下降。在所有情况下，没有特征提取模块时，mAP 下降得最厉害，只有 49.1%。通常情况下，框架的初始特征相当简单、粗糙，直接进入下一步处理会严重影响方法的性能。因此，特征提取模块是必需的。注意力机制是本文框架特有的，其也是一个重要模块，移除特征向量的注意力机制同样会导致 mAP 的大幅下降，这说明了特征向量的注意力机制的重要性。由表 4 还可以看出，只采用一个损失函数会造成 mAP 一定程度的下降，因此，通过混合的身份损失、相似性损失对提出的模型进行训练，可以进一步增强 Re-ID 的性能，具有“锦上添花”的效用。综上，每个模块对最终输出结果都有一定的促进作用。

表 4 Market-15012 数据集上的消融实验结果

Table 4 Ablation results in Market-15012 dataset

（单位：%）

消融实验条目	mAP
没有特征提取模块	49.1
没有注意力机制	70.7
只采用身份损失函数	85.1
只采用相似性损失函数	84.3
完整网络结构	91.2

此外，金字塔块的选择对本文方法的性能也具有一定的影响，因为需要通过金字塔级数（m_l）和 1×1 卷积的数量（$\hat{k}_l$）来控制 SPP 和注意力机制。图 8 给出了在 Market-15012 数据集上通过变更这两个参数得到的 mAP 性能。仅保留少数特征图（即 32 个）特征图得到了较好性能，但该结果很大程度上受金字塔级数的影响。使用大量特征图能够提高性能，但超过 64 个特征图后性能开始下降，这表明使用过多特征图时，共享的冗余信息会限制泛化性能。此外，在金字塔级数方面，实验结果表明，将图像特征分割为大量切块（即 $m_l>4$）并不能提高性能，这可能是池化区域的空间分辨率造成的。事实上，使用深度较大的层会降低特征图的空间分辨率，因此金字塔级数较多时，池化的图像切块可能仅包含很少的垂直像素，由此生成的高维向量取决于图像内视觉概念的垂直位置。

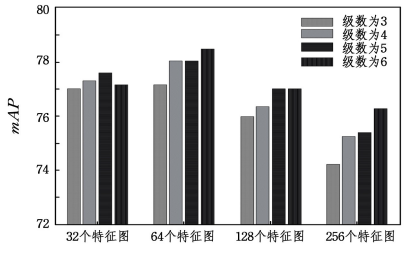

图 8 不同金字塔级数对 mAP 的影响

Fig. 8 Influence of different pyramid series on mAP

结束语 本文提出了一种改进 GoogLeNet 的 Re-ID 方法，利用 SPP 捕捉图像的多级信息，以捕捉不同的语义细节，以图像金字塔理念为基础，利用不同人体部位的细节等级进行信息提取。使用通过注意力加权后的特征向量，来捕捉身份和相似性损失，以显著提升性能。在两个基准数据集上的实验结果证明，所提方法的性能优于其他优秀方法。

在 GoogLeNet 中引入 SPP 和注意力机制会增加计算成本，因此需要额外的训练时间，这可能会限制所提方法在大规模环境中的应用。若能够在推理过程中将较深层的信息迁移到第一层，则可能不需要非常深的网络即可达到较好的结果。

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] WU Z, LI Y, RADKE R J. Viewpoint invariant human re-identification in camera networks using pose priors and subject-discriminative features[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(5): 1095-1108.

[3] WANG J, WANG Z, LIANG C, et al. Equidistance constrained metric learning for person re-identification[J]. Pattern Recognition, 2018, 74(Feb.): 38-51.

[4] YANG F, XU Y, YIN M X, et al. Review on deeplearning-based pedestrian re-identification [J]. Journal of Computer Applications, 2020, 40(5): 1243-1252.

[5] RUI Z, OUYANG W, WANG X. Unsupervised salience learning for person re-identification[C]// Proceeding of IEEE Conference Computer Vision Pattern Recognition. 2013: 3586-3593.

[6] LIU J. Research on human weight recognition technology based on local features[D]. Beijing: Beijing Jiaotong University, 2019.

[7] LIU H. Research on pedestrian re inspection technology for video surveillance[D]. Beijing: University of Chinese Academy of Sciences, 2014.

[8] MARTINEL N. Accelerated low-rank sparse metric learning for person re-identification[J]. Pattern Recognition Letters, 2018, 112(Sep. 1): 234-240.

[9] GÜLER R A, NEVEROVA N, KOKKINOS I. DensePose: Dense Human Pose Estimation In The Wild[C]// Conference on Computer Vision and Pattern Recognition (CVPR). 2018: 7297-7306.

[10] XIONG W, FENG C, XIONG Z J, et al. Improved pedestrian recognition technology based on CNN[J]. Computer Engineering and Science, 2019, 41(4): 95-102.

[11] DAI C C, WANG H Y, NI T G, et al. Person re-identification based on deep convolutional generative adversarial network and expanded neighbor reranking[J]. Journal of Computer Research and Development, 2019, 56(8): 1632-1641.

[12] LEI Z, YANG K, JIANG K, et al. KDAS-ReID: Architecture search for person re-identification via distilled knowledge with dynamic temperature[J]. Algorithms, 2021, 14(5): 137-148.

[13] YAN Y,NI B,LIU J,et al. Multi-level attention model for person re-identification[J]. Pattern Recognition Letters, 2018, 127(4):156-164.

[14] LI C, ZHAO S L, ZHAO J P, et al. Scaling-up algorithm of multi-scale association rules[J]. Computer Science, 2017, 44(8):285-289.

[15] CHEN C, QI F. Review on development of convolutionalneural network and its application in computer vision[J]. Computer Science,2019,46(3):69-79.

[16] MAHENDRAN A, VEDALDI A. Visualizing deep convolutional neural networks using natural pre-images[J]. International Journal of Computer Vision,2015,120(3):75-83.

[17] SONG G,LENG B,YU L,et al. Region-based quality estimation network for large-scale person re-identification[J]. arXiv:1711.08766v2,2017.

[18] LU D,MA W Q. Gesture recognition based on improved YOLOv4 tiny algorithm[J]. Journal of Electronics & Information Technology,2021,43(6):1-9.

[19] GUO Y,CHEUNG N M. Efficient and deep person re-identification using multi-level similarity[C]// Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018: 2335-2344.

[20] CHANG H,QU D,WANG K,et al. Attribute-guided attention and dependency learning for improving person re-identification based on data analysis technology[J]. Enterprise Information Systems,2021,47(5):1-26.

[21] RISTANI E,SOLERA F,ZOU R S,et al. Performance measures and a data set for multi-target, multi-camera tracking[C]// European Conference on Computer Vision. Cham: Springer, 2016:17-35.

MENG Yue-bo, born in 1979, Ph.D, associate professor. Her main research interests include computer vision perception and understanding, intelligent architecture and artificial intelligence.

MU Si-rong, born in 1995, postgraduate. Her main research interests include visual processing, artificial intelligence, and image processing.

(责任编辑:何杨)