

doi: 10.12052/gdutxb.180068

# 基于卷积神经网络的视觉闭环检测研究

杨孟军, 苏成悦, 陈 静, 张洁鑫

(广东工业大学 物理与光电工程学院, 广东 广州 510006)

**摘要:** 闭环检测是视觉SLAM中很重要的一部分, 成功地检测出闭环能减小定位算法所产生的累积里程漂移, 鉴于深度卷积神经网络在分类问题上的优越表现, 本文首次将应用于图像分类的vgg16-places365卷积神经网络模型应用于视觉SLAM闭环检测中, 将配准数据输入训练好的该卷积神经网络, 其各个隐藏层的输出对应于图像特征表示, 然后通过实验比较选用匹配精度较高的中间层完成场景特征提取, 通过计算场景特征的相似性得到闭环区域; 最后在闭环检测数据集上进行实验测试. 测试结果表明, 相比于传统的闭环检测方法, vgg16-places365卷积神经网络模型在相同召回率条件下准确率要高约3%; 对于特征提取时间, 在CPU上要快约5~10倍, 而在GPU上更是比传统人工设计特征的闭环检测快近100倍.

**关键词:** 视觉SLAM; 闭环检测; 卷积神经网络; 特征提取; 相似度

中图分类号: TP242

文献标志码: A

文章编号: 1007-7162(2018)05-0031-07

## Loop Closure Detection for Visual SLAM Using Convolutional Neural Networks

Yang Meng-jun, Su Cheng-yue, Chen Jing, Zhang Jie-xin

(School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China)

**Abstract:** The detection of loop closure is a very important part of visual slam. Successful detection of loop closure can reduce the accumulated mileage drift generated by positioning algorithms. In view of the superior performance of deep convolutional neural networks in classification, the network of VGG16-Places 365 is used, which is widely used in image classification to the area of loop closure detection for the first time. The registration data are input into a trained convolutional neural network, and the output of each hidden layer corresponds to the image feature representation. Then, experiments are implemented to get an intermediate layer with higher matching accuracy, which is used to complete scene feature extraction, and then the loop closure region is obtained by calculating the similarity of the scene feature; finally, experimental tests are performed on loop closure detection dataset. Test results show that the accuracy rate of the VGG16-Places 365 convolutional neural network model is about 3% higher than the traditional ways under the same recall rate; and the the feature extraction time is about 5 to 10 times faster on the CPU and 100 times on the GPU.

**Key words:** visual simultaneous location and mapping (vSLAM); loop closure detection; convolutional neural network; deep learning; similarity

视觉同步定位与地图构建(visual Simultaneous Localization and Mapping, vSLAM)是机器人实现真正全自主移动的关键, 其描述的是机器人在未知环境中从一个未知位置开始移动, 在移动过程中根据位置估计和地图进行自身定位, 同时在自身定位的

基础上建造增量式地图, 实现机器人的自主定位和导航<sup>[1]</sup>. vSLAM算法主要由以下模块组成: 视觉里程计、后端非线性优化、闭环检测以及建图. 闭环检测判断机器人是否曾经到达过先前的位置, 本文将对基于视觉的闭环检测进行研究, 前端提供特征点的

收稿日期: 2018-03-20

基金项目: 国家自然科学基金青年科学基金资助项目(61305069); 广东省信息产业发展专项现代服务业项目(2150510)

作者简介: 杨孟军(1988-), 男, 硕士研究生, 主要研究方向为机器人控制技术、机器视觉.

通信作者: 苏成悦(1961-), 男, 教授, 硕士生导师, 主要研究方向为应用物理、机器视觉. E-mail: scy.gdut@163.com

提取和轨迹、地图的初值,而后端负责对所有的数据进行优化,然而如果要视觉里程计仅考虑相邻时间上的关联,那么之前产生的误差将累积到下一时刻,使整个系统出现累积误差,长期估计的结果将不准确.而如果检测到回环,将其信息提供给后端进行处理,能显著地减少机器人移动过程中的累积误差从而提高后端优化处理的准确性,因此闭环检测对于整个SLAM系统精度与鲁棒性的提升具有重要意义<sup>[2]</sup>.

检测闭环的关键点在于决定观察到的相似性,大部分视觉闭环检测算法是将当前位置图像与先前位置图像进行配准,在这种情况下闭环检测可以看成图像配准的问题,对于大多数人为设计特征进行图像配准的算法采用的是视觉词袋模型(BoVW)<sup>[3]</sup>.词袋模型对图像中视觉特征描述子进行聚类,建立词典,然后对于给定的图像在词袋中找到对应的单词.常见的视觉特征SIFT、Surf、Orb等<sup>[4-6]</sup>被广泛应用,并取得了显著的效果,如在FAB-MAP<sup>[7]</sup>中引入了BoVW,由于SIFT和SURF等本地图像特征在构建BoVW描述符时具有不变性,FAB-MAP实现了优异的性能,成为闭环检测研究的标准基线算法之一.Cummins等<sup>[8]</sup>使用FAST算子提取图像的局部特征,然后应用到增量式的闭环检测中,提高了闭环检测的效率.Liu Y等<sup>[9]</sup>使用图像的GIST描述子提取图像的全局特征,在最近的视觉SLAM研究中受到欢迎.这些技术都是使用人为设计的特征,为了实现所需的特征,其中人类专业知识和见解在开发过程中占主导地位.这将导致很大的应用局限性,特征的提取耗费大量的时间,同时在光照变化明显的场景中时,这些方法忽略了环境中有用的信息,造成闭环检测的准确度不高.

深度学习技术的目的是从可用于分类的原始数据中学习表示数据的方法,闭环检测本质上来说很像一个分类问题,这为典型的闭环检测问题带来了新的方法.从2015年以来,学者们尝试将深度学习的方法应用到闭环检测中.Gao X等<sup>[10-11]</sup>使用Autoencoder提取图像特征,并使用相似度矩阵检测闭环,在公开的数据集上取得了很好的效果.卷积神经网络的最新进展启发研究者们将CNN作为现有图像描述符弱点的潜在解决方案.在许多研究中,CNN从视觉数据特征提取抽象层次的能力已经超过基于人为设计特征解决方案的性能<sup>[12-13]</sup>.特别是CNN在图像分类<sup>[14]</sup>和图像检索任务<sup>[15]</sup>方面的卓越成就是非常令人鼓舞的,考虑到视觉闭环检测类似于图像分类和图像检索,因此基于CNN设计视觉闭环检测问题的解决方案是合理的.He Y等<sup>[16]</sup>使用FLCNN提取图像特征,并

计算相似度矩阵,进一步提高闭环检测的实时性和准确性.Xia Y等<sup>[17]</sup>利用级联深度学习模型PCANet提取图像特征应用到闭环检测中,效果优于传统手工设计的特征.Hou Y等<sup>[18]</sup>利用PlaceCNN提取图像特征进行闭环检测,在复杂的光照情况下取得了很好的检测效果.

为了进一步提高闭环检测的速度和准确性,本文首次提出了将深度模型vgg16-places365卷积神经网络模型应用于视觉SLAM的闭环检测.将该模型作为图像特征提取器,在数据集New College<sup>[7]</sup>上进行验证,在闭环检测的速度和准确性上取得了很好的检测效果,为视觉SLAM闭环检测探索出一种新的方法.

# 1 实验模型

## 1.1 vgg16-places365卷积神经网络结构框架

本文使用开源的Caffe<sup>[19]</sup>深度学习框架来提取基于vgg16-places365CNN的特征.在表1中概述了在Caffe中该模型的体系结构.该CNN模型是一个多层神经网络,主要由3层类型组成:5个卷积层(conv1, conv2, conv3, conv4, conv5),5个最大池化层(pool1, pool2, pool3, pool4, pool5)和3个完全连接的层(fc6, fc7, fc8).最大池化层为相关特征提供平移不变性并同时减小其尺寸.事实上,它也是通过合并底层本地信息来构建抽象表示的过程.而对于完全连接层,前一层中的所有神经元都完全连接到当前层的每个单个神经元.全连接层主要应用在神经网络的高层网络中,这是因为其有利于图像分类和图像检索应用.

表 1 vgg16-places365CNN模型在caffe中的架构以及每层的特征维度  
Tab.1 The architecture and the feature dimension of each layer of the vgg16-places365 CNN model in Caffe

层次	维度
conv1	3211264
pool1	802816
conv2	1605632
pool2	401408
conv3	802816
pool3	200704
conv4	401408
pool4	100352
conv5	100352
pool5	25088
fc6	4096
fc7	4096
fc8	365

输入层是一幅 $224 \times 224 \times 3$ 的三通道图像,卷积层和全连接层的激活函数均采用修正线性单元<sup>[20]</sup> (Rectified Linear Unit, ReLU). ReLU是目前使用最多的激活函数,主要因为其收敛更快,并且能保持同样效果. 标准的ReLU函数为

$$f(x) = \max(x, 0). \quad (1)$$

当 $x > 0$ 时,输出 $x$ ;当 $x \leq 0$ 时,输出为0. convx( $x=1, 2$ )由convx\_1和convx\_2组成,convx( $x=3, 4, 5$ )由convx\_1和convx\_2和convx\_3组成,poolx( $x=1, 2, 3, 4, 5$ )均采用最大池化(Max Pooling)分别接在conv1\_2、

conv2\_2、conv3\_3、conv4\_3、conv5\_3后面减少特征图的大小. 借助深层架构,CNN能够在各种抽象层次上学习高级语义特征. 当输入一幅RGB图像到该模型后,可以提取到每层特征可视化图如图1所示. 此外,与浅层次的卷积和池化层相比,pool5这样的更深层次的池层对于视觉环路闭合检测效果特别突出,因为它仍然保留输入图像的大部分空间信息并且导出输入图像的更丰富的语义表示. fc6和fc7对前面卷积层和池化层提取到的特征进行融合和分类,维度均为4 096. 输出层含有365个神经元,表示数据集的类别数目.

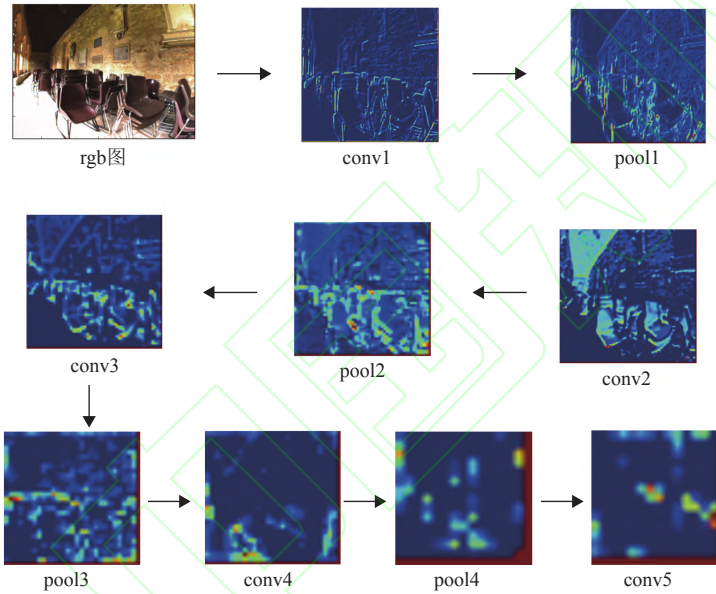


图1 每层特征的可视化图

Fig.1 Visualization of each feature

## 1.2 vgg16-places365卷积神经网络的训练

本文使用vgg16-places365<sup>[21]</sup>卷积神经网络这一模型来作为整个图像的特征提取器,该模型通过Places365-Standard数据集来训练,该数据集是由麻省理工学院收集的用于完成场景识别和场景理解任务的数据集,整个数据集包含180多万张场景图片,分为365个场景类别. 当输入图像通过该网络时,每层的输出被认为是一个特征向量,随着网络层次的不断加深,获得的表达场景语义信息也越丰富,用这些特征向量来实现后续的闭环检测. 因为场景识别是一个图像多分类问题,所以该网络最后采用了Softmax分类器对输入图像进行分类. Softmax是一个分类器,计算的是类别的概率,是Logistic Regression的一种推广. Logistic Regression只能用于二分类,而Softmax可以用于多分类,其计算公式如下:

$$p_j = \frac{e_j^o}{\sum_k e_k^o}, \quad (2)$$

其中 $o$ 为网络模型的参数矩阵; $k$ 为分类数.

本文对输入图像进行减均值的预处理,将处理后的图像输入到vgg16-places365,最终得到一个 $k$ 维的概率向量,进而通过公式(2)可预测输入图像的类别 $\hat{k}$ :

$$\hat{k} = \arg \max_{1 \leq i \leq k} \{p_i\}. \quad (3)$$

Softmax分类器对应的损失函数为交叉熵损失函数(loss function),因此该网络将其作为网络的损失函数,损失函数为

$$L(\theta) = -\frac{1}{m} \left[ \sum_{i=1}^m \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^k e^{\theta_l^T x^{(i)}}} \right], \quad (4)$$



其中 $m$ 为每个训练批次的样本数量; $\theta$ 为网络模型的参数矩阵; $x^{(i)}$ 为第 $i$ 个图片样本; $y^{(i)}$ 为第 $i$ 个样本真实标签; $k$ 为分类数.由公式(4)可知损失函数是非凸的,没有解析解,因此需要通过优化方法来求解.solver的主要作用就是交替调用前向算法和后向算法来更新参数,从而最小化损失,实际上就是一种迭代的优化算法.本文使用随机梯度迭代算法(Stochastic Gradient Descent),来求解最优参数.该网络中的权重和偏置等参数由以下公式更新:

$$V_t = \eta \cdot V_{t-1} - \lambda \cdot \nabla L(\theta_{t-1}), \quad (5)$$

$$\theta_t = \theta_{t-1} + V_t, \quad (6)$$

其中 $\lambda$ 是负梯度的学习率, $\eta$ 是冲量系数为上一次梯度值的权重, $V_t$ 为第 $t$ 次迭代的参数更新值, $\theta_t$ 是第 $t$ 次迭代的参数值.本文重新对模型进行了训练,将该模型中基准学习率 $\eta$ 设为0.01,在迭代次数达到每30万次时将学习率乘以0.1,在冲量系数上做了调整,将冲量系数 $\eta$ 由0.9设为0.95,这样可以使用SGD的深度学习方法更加稳定以及快速.

## 2 闭环检测方法

用训练好的vgg16-places365卷积神经网络模型来提取待验证图片的特征.将图片输入该模型后,就得到隐藏单元每层中数据为该图片的抽象特征,由

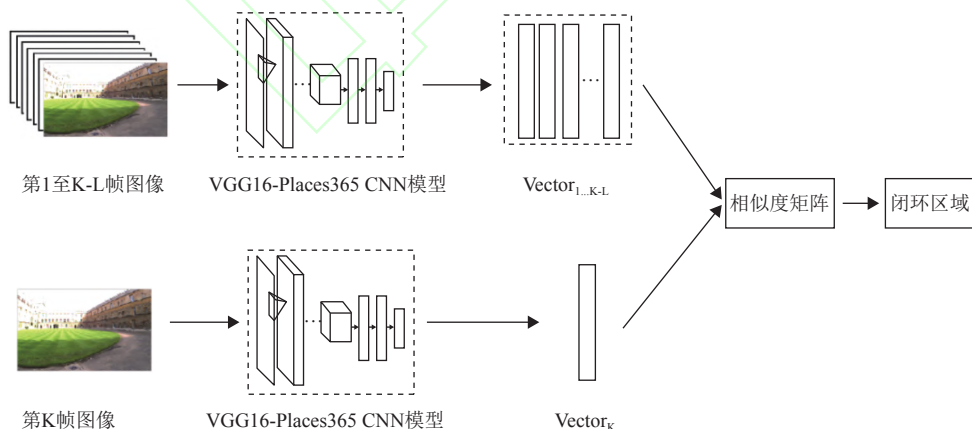


图2 闭环检测方法示意图

Fig.2 Sketch map of loop closure detection

## 3 实验结果与分析

为了验证基于vgg16-places365卷积神经网络闭环检测方法的性能,本文将其与基于传统人工设计特征的BoVW以及其余几种基于深度模型的闭环检测方法,就准确率、召回率以及提取所花的时间等性

于相似的输入会导致类似的特征,两个任意场景( $m, n$ )的差异可以用它们的隐藏层的响应来表示.在基于外观的闭环检测算法中,核心问题是如何计算图像之间的相似性,本文采用余弦相似度来计算场景( $m, n$ )的相似度,余弦相似度衡量的是2个向量间的夹角大小,通过夹角的余弦值表示结果,因此2个向量的余弦相似度为:

$$\cos \theta = \frac{m \cdot n}{\|m\| * \|n\|}. \quad (7)$$

分子为向量 $m$ 与向量 $n$ 的点乘,分母为二者各自的 $L_2$ 相乘,即将所有维度值的平方相加后开方.余弦相似度的取值为 $[-1, 1]$ ,值越大表示越相似.

闭环检测处理的图像在相邻图像帧的特征相关性较大,相似度往往很高,容易被误检测成闭环.因此,在寻找闭环区域的过程中,需要设置相似度比较的图像范围.设 $l$ 是与第 $k$ 帧图像相邻帧的数目,则与第 $k$ 帧图像相似度比较的图像范围应该为第1帧到第 $k-l$ 帧.本文提出的基于vgg16-places365卷积神经网络闭环检测方法如图2所示.对于采集的第 $k$ 帧图像,首先对其进行减均值等预处理,然后将图像输入到vgg16-places365卷积神经网络模型中,提取fc6层的输出作为图像的特征向量( $Vector_k$ ),最后将提取到的特征向量与先前图像的特征向量( $Vector_{1...K-L}$ )比较相似度,若相似度大于设定的阈值,则判断出现闭环区域.

能进行了比较.实验数据是闭环检测数据集New College.计算机硬件配置为:CPU-3.5 GHz, RAM-32 GB, 显卡GTX-1070.软件平台系统为ubuntu,在其上使用caffe这一框架进行网络模型的搭建;除此之外还安装了CUDA8.0,使用CUDA并行计算结构,这充分利用了CPU和GPU各自的优点;安装cuDNN加速库用

来给GPU加速,最后使用python程序来进行特征提取与相似度计算。

### 3.1 实验数据集

New College数据集是由牛津大学移动机器人团队收集的专门用于视觉SLAM的闭环检测算法的评估验证数据集。这个标准数据集包含了1 073对图像,



(a) 左



(b) 右

图3 New College数据集的同一位置左、右摄像头采集的示例图像

Fig.3 Sample images of the left and right cameras collected at the same position in New College

### 3.2 准确率和召回率的比较

对于闭环检测的难度在于是否能正确检测到闭环,常见的两种错误的闭环为:(1)假阳性又称感知偏差,看起来很像回环但实际并不是回环。(2)假阴性又称感知变异,实际上是回环,但是因为光照变化等原因看起来不像是回环<sup>[11]</sup>。闭环检测算法应该能够分辨出这些假阳性和假阴性的回环。为了评价算法的好坏,一般会测试它的 $P$ 和 $R$ 值,其中 $P$ 值为准确率(precision), $R$ 值为召回率(recall),然后做出一条precision-recall曲线,其计算公式如下:

$$P = \frac{TP}{TP + FP}, \quad (8)$$

$$R = \frac{TP}{TP + FN}, \quad (9)$$

其中 $TP$ 代表正确检测出的闭环数目, $FP$ 称假阳性代表真实上不是闭环但是检测的结果是闭环的数目; $FN$ 又称假阴性代表没有检测到的真实的闭环数目;准确率描述的是算法所提取的所有闭环中,确实是真实闭环的概率。召回率则描述的是在所有真实的闭环中,被正确检测出来的概率。

为了找出最适用于视觉环路闭合检测的基于vgg16-places365卷积神经网络的特征,本文评估了来自所有层(conv1、conv2、conv3层除外)的基于该网络特征的性能。图4显示了在New College数据集上基于

它们是由在移动平台的左边和右边各放置一个摄像头,移动平台每行进1.5 m采集一次图像来进行收集。图3给出了数据集的示例图像。数据集中还给出了形成闭环区域的真实标注,标注以矩阵的形式给出,若图像 $i$ 与图像 $j$ 形成闭环区域,则 $(i, j)$ 对应的数值为1,否则为0。闭环检测时 $L$ 值的设置方式与文献[5]相同,将New College数据集的 $L$ 值设为100。

vgg16-places365卷积神经网络中的某些层比较的实验结果(precision-recall曲线)。对于卷积层和最大池化层来说,可以看出性能在逐层递增;而最终完全连接的层性能在不断下降,那是因为全连接层会丢失空间信息,所以用高层次的全连接层提取的特征视觉环路闭合检测的效果不是很好,表明深度学习的深度和编码图像时空信息的深度的重要性。由图可以看出fc6、pool5、与conv5三层的特征在闭环检测上的性能优于其他层。其中以fc6层性能最佳,当召回率低于40%时其准确率为100%,当召回率达到40%时其准确率也达到90%左右。

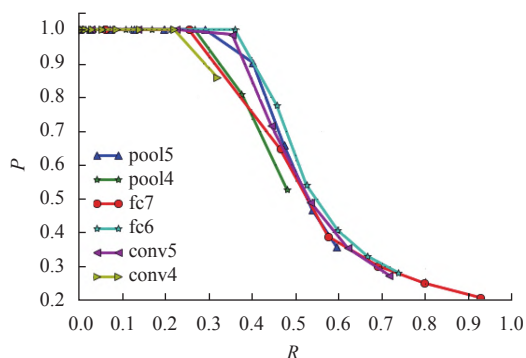


图4 基于vgg16-places365 CNN模型的各个层次特征的闭环检测的Precision-Recall曲线比较

Fig.4 Comparison of Precision-Recall curve for loop closure detection of various layer features of vgg16-places365 CNN Model

为了比较不同算法对视觉闭环检测的有效性,在New College数据集上对基于vgg16-places365卷积神经网络的描述符的两个最优表现层特征以及基于人工设计的特征(BoVW, GIST)的性能进行了比较,其实验结果如图5所示. 由图5可以看到,基于vgg16-places365卷积神经网络fc6层的描述符在这两个评估标准的表现优于人工设计的描述符的表现,在相同的召回率下其准确率要高约3%~4%. 这是因为人工设计的特征主要依靠设计者的先验知识,很难利用大数据的优势. 由于依赖手工调参数,特征的设计中只允许出现少量的参数,而利用深度模型学习得到的数据特征对大数据的丰富内在信息更有代表性,深度学习可以从大数据中自动学习特征的表示,其中可以包含成千上万的参数;除此之外,利用人工设计的特征得到的是局部特征,而利用深度学习得到的是全局特征.

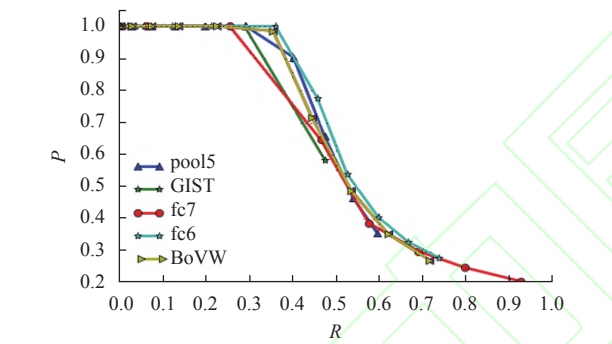


图 5 基于vgg16-places365 CNN模型与传统人工设计特征的PR曲线对比

Fig.5 Comparison of PR Curve of vgg16-places365 Model and Traditional Artificial Design Features

本文除了对比传统的人工设计特征视觉闭环检测的有效性外,还对比了其他最新深度学习在闭环检测性能上的对比,实验同样在New College数据集上对基于vgg16-places365卷积神经网络的描述符的两个最优表现层特征以及最近基于深度学习的PlaceCNN、Autoencoder、FLCNN等的性能进行了比较,其实验结果如图6所示. 由图知,根据两个评估标准,基于vgg16-places365卷积神经网络fc6层的描述符的表现略微优于其余基于卷积神经网络最优层的描述符(FLCNN、PlaceCNN)的表现,当召回率相同时有着更高的准确率. 这是因为基于vgg16-places365卷积神经网络结构更复杂,每层特征维度更多,因此其更能比较好的体现特征. 除此之外,还可知基于卷积神经网络的描述符要优于基于Autoencoder在闭环检测上的性能,这是因为基于卷积神经网络的模型对

于场景分类问题表现更优.

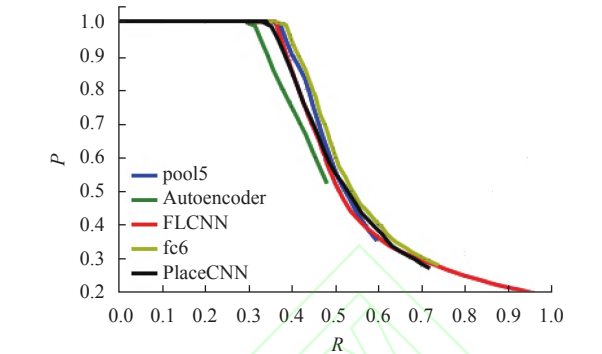


图 6 基于vgg16-places365 CNN模型与基于其余深度学习模型的PR曲线对比

Fig.6 Comparison of PR curve of vgg16-places365 model and other deep learning models

3.3 时间性能比较

评估图像描述符的另一个重要考虑因素是其计算效率. 这个效率是通过提取描述符所需的时间和描述符的长度来衡量的. 本文比较了基于vgg16-places365卷积神经网络的闭环检测与其他两种人工设计特征的闭环检测以及其他3种基于深度模型闭环检测算法所需要的时间,得到结果如表2所示. 这里的时间是1 000张图像的平均值,其中深度模型只包括特征提取时间,不包括加载输入图像或卷积神经网络模型的时间. 由表2可以看出在CPU上时,基于vgg16-places365描述符比所有人工设计的描述符更有效,每个图像平均需要0.139 s,比其他两种最先进的手工特征快约5~10倍. 如果在GPU上提取CNN描述符,则速度还要快一个数量级达到0.014 s. 由表2还可知道基于vgg16-places365卷积神经网络特征提取速度要快于基于Autoencoder和PlaceCNN两模型的,但是要稍微慢于基于FLCNN模型的速度. 这是因为基于FLCNN是一种专门设计的快速、精简的卷积网络模型,有着很好的实时性.

表 2 用不同特征描述子来计算所得的平均每幅图像所花费时间

Tab.2 The cost time spent on the average of each image with different feature descriptors

特征	时间/s
BoVW	0.567
GIST	1.787
VGG(CPU)	0.139
VGG(GPU)	0.014
Autoencoder	0.020
PlaceCNN	0.052
FLCNN	0.013



## 4 结论

本文使用大型的场景识别数据集训练好的vgg16-places365卷积神经网络模型作为场景图像的特征提取器,最后使用所提取的特征解决闭环检测问题.实验对该网络模型的某些层的特征在闭环检测上的性能进行对比,由实验结果的precision-recall曲线图可以明显观察到fc6在检测精度和表示的紧凑性方面表现最佳;同时本文还对基于最先进的人工设计的特征和基于vgg16-places365卷积神经网络表现比较好的层特征在视觉SLAM闭环检测性能进行比较研究,观察到基于vgg16-places365模型闭环检测的性能要优于人工设计的特征以及其余3种深度模型在闭环检测上的性能,在同样的召回率下其基于准确率更高.另外本文还在在时间性能上进行了对比,基于vgg16-places365卷积神经网络模型的图像特征提取比基于人工设计特征提取在CPU上的速度要快一个数量级,而在GPU上提取速度更是比传统人工设计特征快约两个数量级,同时在GPU上也比基于Autoencoder和PlaceCNN要快2到3倍,虽速度上略慢于基于FLCNN模型的方法,但差距不是很大,也有着良好的实时性.

## 参考文献:

- [1] BAILEY T, DURRANT-WHYTE H. Simultaneous localization and mapping: part I [J]. IEEE Robotics & Automation Magazine, 2006, 13(2): 99-110.
- [2] WANG H, HOU Z, CHENG L, TAN M. Online mapping with a mobile robot in dynamic and unknown environments [J]. International Journal of Modelling Identification & Control, 2008, 4(4): 415-423.
- [3] FILLIAT D. A visual bag of words method for interactive qualitative localization and mapping [C]//Robotics and Automation, IEEE International Conference. Roma: IEEE, 2007: 3921-3926.
- [4] LOWE D G. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [5] BAY H, TUYTELAARS T, GOOL L. SURF: Speeded up robust features [C]//Computer Vision-ECCV 2006 [J]. Computer Vision-ECCV, 2006, 3951: 404-417.
- [6] RUBLEE E, RABAUD V, KONOLIGE K. Orb: an efficient alternative to sift or surf [C]//Computer Vision, IEEE International Conference. Barcelona, Spain: IEEE, 2011: 2564-2571.
- [7] CUMMINS M, NEWMAN P. Highly scalable appearance-only SLAM - FAB-MAP 2.0 [M]//Proceedings of Robotics: Science and Systems. Seattle, 2009: 1-8.
- [8] CUMMINS M, NEWMAN P. FAB-MAP: Probabilistic localization and mapping in the space of appearance [J]. International Journal of Robotics Research, 2008, 27(6): 647-665.
- [9] LIU Y, ZHANG H. Visual loop closure detection with a compact image descriptor [J]. IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, Vilamoura-Algarve, Portugal, 2012: 1051-1056.
- [10] GAO X, ZHANG T. Unsupervised learning to detect loops using deep neural networks for visual SLAM system [J]. Autonomous Robots, 2017, 41(1): 1-18.
- [11] GAO X, ZHANG T. Loop closure detection for visual slam systems using deep neural networks [C]//Technical committee on control theory, Chinese control conference. Hangzhou: Chinese Association of Automation, 2015: 5851-5856.
- [12] CHATFIELD K, SIMONYAN K, VEDALDI A, *et al.* Return of the devil in the details: delving deep into convolutional nets [J]. Computer Science, 2014: 1-11.
- [13] WAN J, WANG D Y. Deep learning for content-based image retrieval: a comprehensive study [C]//Multimedia, ACM International Conference. Istanbul: ACM, 2014: 157-166.
- [14] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks [C]//International Conference on Neural Information Processing. Doha, Qatar: ACM, 2012, 60(2): 1097-1105.
- [15] BABENKO A, SLESAREV A, CHIGORIN A. Neural codes for image retrieval [C]//Computer Vision, European Conference. Zurich: Springer, 2014: 584-599.
- [16] 何元烈, 陈佳腾, 曾碧. 基于精简卷积神经网络的快速闭环检测方法[J/OL]. 计算机工程, 2017: 1-6. <http://kns.cnki.net/kcms/detail/31.1289.TP.20170807.1156.002.html>.  
HE Y L, CHEN J T, ZENG B. A fast loop closure detection method based on lightweight convolutional neural network [J]. Computer Engineering, 2017: 1-6. <http://kns.cnki.net/kcms/detail/31.1289.TP.20170807.1156.002.html>.
- [17] XIA Y, LI J, QI L, *et al.* Loop closure detection for visual SLAM using PCANet features [C]//Neural Networks, IEEE International Joint Conference. Vancouver, Canada: IEEE, 2016: 2274-2281.
- [18] HOU Y, ZHANG H, ZHOU S. Convolutional neural network-based image representation for visual loop closure detection [C]//Information and Automation, IEEE International Conference. Lijiang, China: IEEE, 2015: 2238-2245.
- [19] JIA Y Q, SHELHAMER E, JEFF D. Caffe: convolutional architecture for fast feature embedding [C]//Multimedia, ACM International Conference. Istanbul: ACM, 2014: 675-678.
- [20] SHANG W, SOHN K, ALMEIDA D, *et al.* Understanding and improving convolutional neural networks via concatenated rectified linear units [C]//Machine Learning, IEEE International Conference. New York: IEEE, 2016: 1-17.
- [21] ZHOU B, KHOSLA A, LAPEDRIZA A, *et al.* Places: An image database for deep scene understanding [J]. Journal of Vision, 2016, 17(10): 1-12.