文章编号: 1001-0920(2019)05-0981-08 DOI: 10.13195/j.kzyjc.2017.1514

基于栈式卷积自编码的视觉SLAM闭环检测

张云洲[1,2], 胡 航[1†], 秦 操[1], 楚 好[2], 吴运幸[1]

(1. 东北大学 信息科学与工程学院，沈阳 110004；2. 东北大学 机器人科学与工程学院，沈阳 110004)

摘 要: 同时定位与构图(SLAM)主要用于解决移动机器人在未知环境中进行地图构建和导航的问题,是移动机器人实现自主移动的基础. 闭环检测是视觉SLAM的关键步骤,对构建一致性地图和减少位姿累积误差具有重要作用. 当前的闭环检测方法通常采用传统的SIFT、SURF等特征,很容易受到环境影响,为了提高闭环检测的准确性和鲁棒性,提出基于无监督栈式卷积自编码(CAEs)模型的特征提取方法,运用训练好的CAEs卷积神经网络对输入图像进行学习,将输出的特征应用于闭环检测. 实验结果表明:与传统的BoW方法及其他基于深度学习模型的方法相比,所提出的算法能够有效降低图像特征的维数并改善特征描述的效果,可以在机器人SLAM闭环检测环节获得更好的精确性和鲁棒性.

关键词: 机器人；同时定位与构图；闭环检测；深度学习；无监督学习；栈式卷积自编码

中图分类号: TP24 **文献标志码:** A

Loop closure detection for visual SLAM based on stacked convolutional autoencoder

ZHANG Yun-zhou[1,2], *HU Hang*[1†], *QIN Cao*[1], *CHU Hao*[2], *WU Yun-xing*[1]

(1. College of Information Science and Engineering，Northeastern University，Shenyang 110004，China；2. Faculty of Robot Science and Engineering，Northeastern University，Shenyang 110004，China)

Abstract: As the foundation to realize the autonomous movement of mobile robots, simultaneous localization and mapping(SLAM), which is mainly used to solve the problem of mobile robots mapping and navigation in unknown environment, has been paid more attention in recent years. Loop closure detection, one of the key steps of visual SLAM, plays an important role to make a globally consistent map and reduce accumulated error of robot pose. Current methods for loop closure detection are vulnerable to environmental influence because they always adopt traditional features such as SIFT and SURF. To improve the accuracy and robustness of loop closure detection, a method based on unsupervised Stacked Convolutional Autoencoders(CAEs) model is proposed. The trained CAEs convolution neural network is used to learn from input images, while the output features are used for loop closure detection. The results of experiment show that the proposed method, compared with traditional BoW-based methods and other methods based on deep learning model, can effectively reduce the dimension of image features and improve the effect of feature description. Thus, it can attain better accuracy and robustness in loop closure detection of robot SLAM.

Keywords: robot；SLAM；loop closure detection；deep learning；unsupervised learning；stacked convolutional autoencoders

0 引 言

同时定位与构图(Simultaneous localization and mapping, SLAM)是机器人实现自主移动的关键基础之一,包括特征提取与匹配、数据配准、闭环(Loop closure)检测和全局优化等步骤[1-4]. 其中闭环检测可以判断当前位置是否已被移动机器人访问过,是SLAM过程的关键环节. 准确地检测出闭环可以有效减少机器人位姿估计的累积误差,有利于构建更加精确的地图[2],保证生成地图的一致性.

闭环检测在本质上是数据关联问题. 仅依靠机

收稿日期: 2017-11-10；**修回日期:** 2018-02-23.

基金项目: 国家自然科学基金项目(61471110, 61733003)；国家重点研发计划项目(2017YFC080500015005)；中央高校基本科研业务费专项基金项目(N172608005, N160413002).

责任编委: 高会军.

作者简介: 张云洲(1974−), 男, 教授, 博士, 从事智能机器人、计算机视觉领域等研究；胡航(1993−), 男, 硕士生, 从事计算机视觉的研究.

†通讯作者. E-mail: nickhoo@stumail.neu.edu.cn.

器人内部传感器进行位姿估计产生的误差较大,难以实现准确的闭环检测. 对于视觉SLAM,常规做法是将当前时刻的场景图像与之前采集到的场景图像序列进行匹配,当相似度高于阈值时即为机器人的闭环位置. 采用RGB-D传感器并利用深度图像可以在一定程度上避免普通摄像机受到光照、 阴影和色度等环境因素的影响,因而被广泛应用. 然而,当前的闭环检测大多还是采用基于传统特征(如SIFT[5]、SURF[6])的视觉词袋模型[7](Bag of words, BoW)方法,在场景或感知条件发生变化时难以提供准确且鲁棒的图像特征描述,容易导致SLAM闭环检测失败.

近年来,新兴的深度卷积神经网络在图像识别和分类[8-9]中显示出巨大的优势,受到了国内外研究者的高度关注. 深度学习尤其在图像内容表征方面表现出了优异的性能. 例如,在国际大规模视觉识别大赛(ILSVRC)中,采用深度学习的分类任务Top-5错误率已下降至4.8%[10-11],超过了人类的识别能力(Top-5错误率为5.1%). 此外,无监督学习方法不需要带有标签的数据集,能够适用于大量数据,减少人工标注的工作量.

基于上述因素,本文提出一种基于无监督卷积神经网络模型的算法,用于提高移动机器人SLAM闭环检测的精度和鲁棒性. 本文通过选取训练图片对栈式卷积自编码(Stacked convolutional autoencoder, CAEs)[12]进行训练,运用训练好的网络对原始图像进行测试,得到的特征是对该图像的描述. 利用这些特征度量图像的相似性建立帧到帧的特征关联,并且这些特征描述是图像与机器人位姿一对一的平滑映射,可以在不跟踪机器人位姿的前提下完成闭环检测. 在衡量特征向量距离时,本文通过余弦距离的方法判断图像间的相似度,以获得较高的准确率和可靠性.

1 相关工作

在视觉SLAM闭环检测领域,现有的方法可以大致分为两类:传统特征的方法和基于深度学习的方法.

传统方法主要是基于手工特征提取图片的特征表达. 其中,局部特征提供了对一定光照和仿射变换的不变性,适合解决图像匹配、检索等问题,已广泛应用于视觉闭环检测. 基于局部特征提出的视觉词袋模型(BoW)具有海量图片高效检索的特点,是视觉SLAM中最常用的图像表示模型. 为了形成BoW描述符,需要提取诸如SIFT或SURF的图像特征,通过K-means聚类方法形成若干聚类,将其中心作为视觉单词. 通过计算每个中心的特征点数目可以获得视觉词汇的频率. 因此,一个图像可以由频率列表(矢量化)表示. 然后,估计两个向量的距离,可以判断它们是否为同一个位置. 例如,Cummins等[13]提出的FAB-MAP算法利用已探索区域的外部数据图像特征形成视觉词袋,并通过比较由两个位置生成的向量概率来决定是否实现闭环,得到了较好的效果.

基于词典的方法也具有局限性,该方法经常会把不同位置获取的相似场景判断为同一场景. 为了克服该问题,Gálvez等[14]提出了以FAST+BRIEF特征构建BoW进行场景识别的方法;Mei等[15]提出了类似的方法,并且引入了动态词袋的概念,避免了大多数基于图像的闭环算法中常见的机器人轨迹空间的任意离散化;Murphy等[16]提出了一种在线场景探索和识别的方法,融合时间信息拓扑地创建独特的场景,能够提高场景识别的准确率和召回率;Kawewong等[3]提出了一种快速的在线增量式闭环检测方法,利用K均值方法加快特征匹配速度,并改善了图像相似性计算方法;Kejriwal等[17]将空间共生信息加入到词典中,解决了感知混淆的缺陷;Khan等[18]提出的方法增量式地生成二元词汇表,不需要前期的词汇学习阶段,其词汇生成过程是基于连续的图像之间的特征跟踪;李博等[19]提出了移动机器人闭环检测的视觉字典树金字塔TF-IDF得分匹配方法,解决了视觉字典场景外观表征性能受制于有限单词个数以及算法效率低的问题.

近年来,深度学习技术的快速发展为闭环检测问题的解决提供了新的思路[4]. 对于视觉场景识别及分类、深度学习具有极大的优势. 尤其是对于存在一定变化的视觉场景,深度学习可以提供鲁棒的图像描述[20-21]. 考虑到移动机器人在视觉SLAM闭环检测环节同样需要可靠的特征,深度学习的方法逐渐引起了研究者的关注. 然而,深度卷积神经网络模型在场景识别领域通常采用有监督的学习方法,需要大量含有标签的数据,在实际操作时存在很大的困难. 因此,采用不依赖于数据集的自编码方法具有更高的可行性. Gao等[22]提出了一种利用堆栈式去噪稀疏自编码模型进行闭环检测的方法,但没有考虑图像的空间局部特性;Xia等[23]应用PCA-Net深度网络提取图像特征作为图像的描述,网络简单并且便于调试,但输入的图像像素太小,导致实际操作时不容易判断图像间的相似性.

考虑到当前研究存在的缺陷,本文采用CAEs网络. CAEs属于无监督深度学习模型,考虑了图像的空间局部特性,在泛化性、鲁棒性方面表现优异. 与卷

积神经网络(CNN)相比,CAEs不需要含有标签的数据训练,能够有效减少标记的工作量.

2　基于栈式卷积自编码的闭环检测

本文的SLAM系统是基于图优化的方法,主要组成部分为SLAM前端、基于CAEs特征的闭环检测和SLAM后端,具体架构如图1所示.

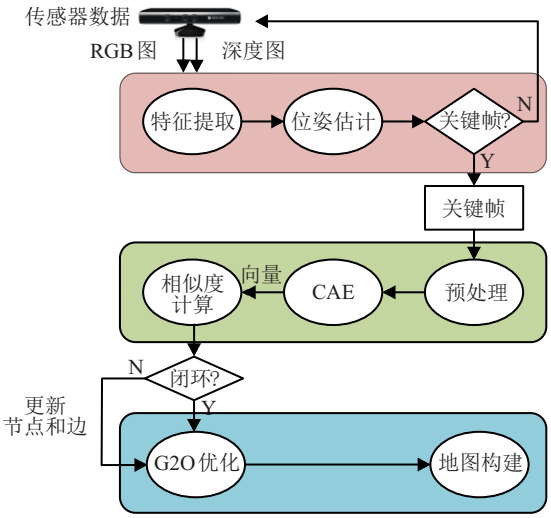


图1　本文的SLAM系统框架

2.1　SLAM前端

本文以RGB-D传感器获取彩色和深度图像,对连续的图像提取ORB特征,再结合深度图像,利用随机采样一致性(Ransac)算法求解PnP问题得到相机的位姿估计[24]. 由于连续的图片存在大量的冗余信息,本文采取基于姿态约束的策略对关键帧进行筛选,基本思路为:如果机器人在采集两帧图像时的相对位姿变化较大,则新帧就被判定为关键帧. 这些关键帧将以节点和边的方式送给G2O环节进行SLAM后端优化.

2.2　基于CAEs的闭环检测

CAE是一种利用卷积操作将一系列简单信号进行编码,尝试重构输入的神经网络模型. 它与自编码器(AE)和卷积神经网络(CNN)关系密切,但又不同于这两种模型. 作为无监督模型,在AE中图像必须被展开为一个向量,这迫使每个特征都是全局特征,忽略了自然图像中物体的结构信息;CAE则更侧重局部特征的学习,更符合视觉与目标识别的发展趋势. 与CNN相似,CAE以二维图像作为输入,采用多种卷积滤波器提取图像特征,但CNN属于监督学习,需要获得大量标签才能训练,CAE的无监督特性提高了其实用价值. 文献[22]提到的堆栈式去噪稀疏自编码模型(SDA)与AE结构类似,但本文无需通过传统特征提取图像块,而采用深度学习的方法处理闭环检测问题.

CAE包括两个部分:编码和解码. CAE的编码部分由卷积层与最大池化层构成. 单一的卷积滤波器无法实现多种类的模式学习. 因此,每个卷积层由n个(超完备参数)卷积滤波器组成. 假设输入图像为$I\{I_1, I_2, \cdots, I_N\}$,编码过程中的卷积滤波器为$F^1 = \{F_1^{(1)}, F_2^{(1)}, \cdots, F_n^{(1)}\}$,每个滤波器的通道数目与输入图像的深度相同. 经过卷积操作,每张图像可以得到n组特征激活图(Activation map)

$$h_m(i,j) = a\Big(\sum_{u=-k}^{k}\sum_{v=-k}^{k} F_m^{(1)}(u,v)I(i-u,j-v) + b_m^{(1)}\Big),$$
$$m = 1, 2, \cdots, n. \tag{1}$$

其中:$h_m(i,j)$是第m个激活图中像素(i,j)处的激活值;a是激活函数;k是与正方形卷积滤波器相关的变量,$2k+1$是滤波器的大小;$b_m^{(1)}$是第m个激活图的偏置.

为了提高网络的泛化能力,每次都采用一个非线性激活函数a作用于网络,卷积后的结果记为

$$h_m = a(I \times F_m^{(1)} + b_m^{(1)}),\ m = 1, 2, \cdots, n. \tag{2}$$

解码操作是从特征激活图中重建输入图像I. CAE是全卷积网络,因此解码过程也是反卷积操作(即转置卷积). 考虑到编码后得到的激活图大小会比输入图像小,再通过解码过程的转置卷积不能重建输入图像的大小,因此需要对输入图像进行补零padding操作,以便在编码解码后产生与原输入图像大小相同的图像. 将编码后的结果作为解码器的输入,然后与卷积滤波器$F^{(2)}$做卷积,即可得到重构图像$\tilde{I}$,即

$$\tilde{I} = a(H \times F_m^{(2)} + b^{(2)}),\ m = 1, 2, \cdots, n, \tag{3}$$

其中H是n个h特征激活图的集合. 代价函数采用均方误差(MSE)[25],定义输入图像与重构图像之间的差异$L(I, \tilde{I})$,即

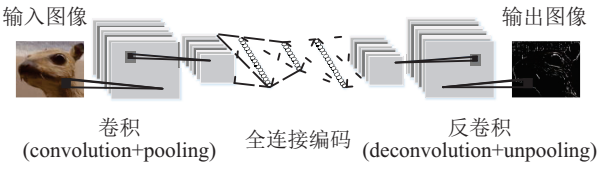
$$L(I, \tilde{I}) = \frac{1}{2}\|I - \tilde{I}\|_2^2. \tag{4}$$

由RGB-D相机得到的图像大小为640×480像素,其较高的维度会带来很高的计算复杂度. 因此,将图像进行缩小,预处理后的图像大小为320×240像素,然后将缩小的图片批量进行图像增强处理.

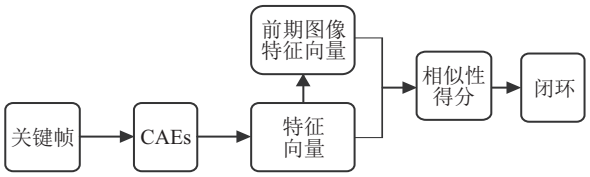
Oxford数据集中的New College和City Centre数据集广泛用于评价视觉SLAM闭环检测的性能[11,26]. 两个数据集分别包含2 146帧、2 474帧图像,采集的距离分别达到1.9 km、2.0 km. 本文按照均匀

图像间隔序列从New College和City Centre中挑选60%的图片作为训练集.

为了将维度较高的原图像降到合适的维度,并且能够很好地表征图像特征,以便高效地计算图像相似度,本文采用CAEs网络模型,即多层的CAE叠加而成的网络. CAE网络的解码和编码过程通过卷积层和反卷积层实现,其中编码层包括padding层、卷积层、tanh层、池化层、dropout层,解码层包括padding层、卷积层、tanh层. 整个CAE网络共8层,单个CAE网络模型结构如图2(a)所示.



(a) CAE模型结构



(b) CAEs闭环检测流程

图2 CAE模型结构及闭环检测流程

具体地,本文采用的CAEs网络模型包含5个结构相同的CAE网络,总共由40(5×8)层组成,最后一个CAE网络中编码层输出的结果是768维特征向量,即为提取到的图像特征. 采用此模型可以将预处理后的320×240像素的图像降维到可用768维向量表达,从而在提取特征的同时有效地降低维度. 本文CAEs网络模型中的CAE编码层参数如表1所示. CAE解码层的padding层、卷积层、激活函数层与表1中的相同. 最后将训练集里面的图片加载到Tensorflow搭建的CAEs网络进行训练,优化器选用AdamOptimizer,不断迭代优化均方误差至收敛,训练得到网络的参数.

表1 CAE网络参数

参数	数值	描述
Padding	VALID	填充方式
F	3×3	卷积核大小
m	32	卷积核数量
a	tanh	激活函数
池化	2×2	池化大小
Dropout	0.5	Dropout参数

SLAM中的闭环检测问题就是寻找机器人运动过程中的相同场景. 对于输入的序列图像数据,本文通过训练好的CAEs和优化后的特征判断场景的相似性,进而判断是否形成闭环.

假设存在两个关键帧 f_i 和 f_j,每个关键帧通过CAEs之后可以用如下 t 个特征来表达:

$$f_n = \{p_1^{(n)}, p_2^{(n)}, \cdots, p_t^{(n)}\},\ n = i, j. \tag{5}$$

定义一个相似性函数 δ,利用夹角余弦衡量特征的相似度 s,即

$$s = \delta(p^{(i)} - p^{(j)}) = \frac{\sum_{k=1}^{t} p_k^i p_k^j}{\sqrt{\sum_{k=1}^{t} (p_k^i)^2}\sqrt{\sum_{k=1}^{t} (p_k^j)^2}}. \tag{6}$$

闭环检测阈值选取原则是: 首先计算当前关键帧图像与上一关键帧图像之间先验相似度 $s(f_t, f_{t-\Delta t})$; 然后,计算当前帧与之前某关键帧之间的相似度,并根据该先验相似度进行归一化

$$\eta(f_{t_i}, f_{t_j}) = \frac{s(f_{t_i}, f_{t_j})}{s(f_t, f_{t-\Delta t})}. \tag{7}$$

本文认为当前帧与之前某关键帧的相似度超过当前帧与上一个关键帧的相似度的3倍时,形成闭环. 图2(b)给出了CAEs闭环检测的基本流程.

2.3 SLAM后端

前端估计得到的变换矩阵存在局部误差,随着时间推移会出现位姿漂移现象,需要通过后端优化方法处理. 基于图优化的误差均衡包含两个方面: 1) 构建图结构,在位姿图中,节点表示机器人位姿,节点间的边表示位姿之间的空间约束关系; 2) 利用构建的图结构计算达到全局一致性的节点位姿,将累积误差进行均衡处理,求解最优的位姿序列.

节点 x_i 和 x_j 表示机器人不同时刻的两个位姿, z_{ij} 表示从节点 x_i 观测到节点 x_j 的测量值, Ω_{ij} 表示信息矩阵(协方差矩阵的逆),协方差矩阵表示观测的不确定性. e_{ij} 表示观测值与节点真实的位姿变换间的误差,误差函数 $e_{ij}(x_i, x_j)$ 可表示为

$$e_{ij}(x_i, x_j) = z_{ij} - \hat{z}_{ij}(x_i, x_j). \tag{8}$$

其中: z_{ij} 表示真实的观测值, $\hat{z}_{ij}$ 表示估计值.

基于图优化的SLAM的主要思想是: 保留所有的传感器数据和这些数据之间的空间约束关系,然后以最大似然方法估计机器人的位姿. 优化目标函数可表示为

$$F(x) = \sum_{ij} e_{ij}^{\mathrm{T}}(x_i, x_j)\Omega_{ij}e_{ij}(x_i, x_j). \tag{9}$$

这样,SLAM的优化问题就转化为求目标函数 $F(x)$ 的最小值,即

$$x^* = \arg\min_x F(x). \tag{10}$$

由式(10)可知,SLAM的图优化问题本质上是最小二乘问题,通过迭代法进行求解. 构建全局一致性的节点位姿有3个前提: 1) 闭环检测准确; 2) 有足够的约束条件; 3) G2O准确求解. 前提2)主要是指相机位姿之间的约束关系,约束越多越准确,估计的位姿也会越准确. 优化之后完成对位姿的更新,得到全局一致性位姿信息. 因此,将关键帧之间的准确约束和准确的闭环检测约束送给G2O可求解出全局一致性的节点位姿,从而得到全局一致性地图.

3　实验结果及分析

为验证所提出的算法,本文在3个公开数据集(Oxford、TUM和NYU-Depth V2)上进行测试. 其中: Oxford数据集和TUM数据集是用于评价闭环检测算法的标准数据集; NYU-Depth V2数据集用于测试基于CAEs闭环检测的SLAM系统构图效果. 同时,通过Kinect实时采集图像进行SLAM构图,为了比较构图效果,本文在实际环境中开展基于深度学习网络闭环检测的实验. 本文训练CAEs网络的平台是深度学习服务器,其配置为: 双Intel Xeon E5 CPU, 2 400 MHz,128 GB 内存,4 块 GTX1 080GPU. 机载处理器为Intel Xeon E3 CPU,主频为3.40 GHz,内存容量为8 GB.

3.1　闭环检测算法性能分析

本文通过Oxford数据集中的New College和City Centre数据集测试闭环检测算法的性能. 实验时,分别以左目相机、右目相机进行图像采集,并通过预处理方法进行图像缩小和增强处理.

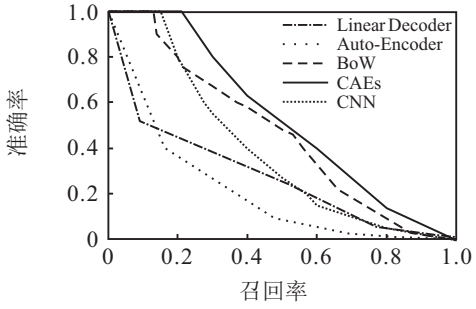
为验证本文算法的效果,将本文方法与其他方法进行对比,包括传统的方法(如BoW)和深度学习的方法,例如CNN、自编码(Auto-Encoder)[27]、线性解码(Linear Decoder)[28]. BoW是基于手工特征的方法,在训练字典时提取64维SURF描述子. 在深度学习方法中,CNN采用了GoogLeNet模型框架,自编码和线性编码则基于UFLDL Tutorial[29]中的模型. 由于各种算法是以不同的语言和软件实现的(如线性解码器、自编码器通过Matlab程序实现,BoW以C++语言在Linux环境下实现,CNN、CAEs通过Python实现并在Jupyter下测试),很难对时间效率进行对比. 本文主要采用两个指标来评估算法的性能: 准确率和召回率. 这里,准确率(Precision)是检测到的真实闭环(TP)与检测到的闭环总量(TP和FP)的比值; 召回率(Recall)是检测到的真实闭环(TP)与数据集中存在的闭环总量(TP和FN)的比值.

$$\text{Precision}: P = \frac{\text{TP}}{\text{TP}+\text{FP}}, \tag{11}$$

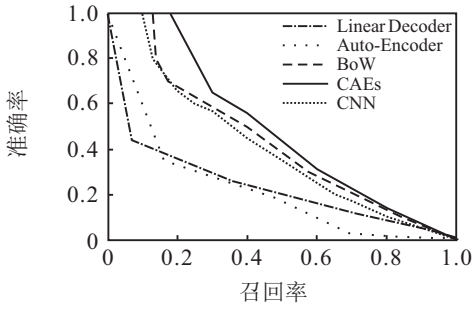
$$\text{Recall}: R = \frac{\text{TP}}{\text{TP}+\text{FN}}. \tag{12}$$

在实验之前,将New College和City Centre的数据分成两部分,分别包括左目相机采集的图像和右目相机采集的图像. 然后,将4组图像分别进行试验.

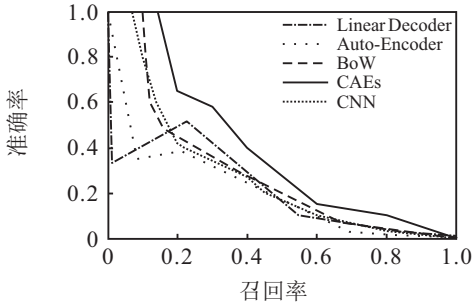
对CAEs网络进行训练,得到网络的参数. 利用整个数据集的数据做测试,提取图片特征向量,通过余弦相似度函数计算向量之间的值. 设定合适的阈值,当得分超过这一阈值时认为检测到了闭环. 改变阈值可得精度召回曲线. New College数据集、City Centre数据集的准确率召回率测试结果如图3所示.



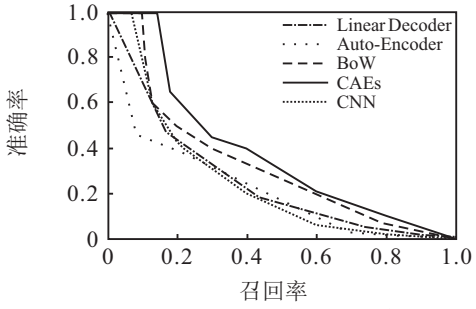
(a) New College 奇数号图像



(b) New College 偶数号图像



(c) City Centre 奇数号图像



(d) City Centre 偶数号图像

图3　公开数据集测试获得的准确率召回率

在NewCollege数据集上,当准确率为80%时,CAEs、自编码器、线性解码器、CNN和BoW的平均召回率依次为27.5%、6%、2.5%、16.5%和17%;在City Centre数据集上,准确率为80%时,相应的数据依次为17.5%、3%、3.5%、10%和11%. 可以看出,与传统的BoW方法,或与线性解码器、自编码器和CNN等深度学习方法相比,CAEs能够在准确率较高的情况下保持较高的召回率.

3.2 检测闭环序列

本文采用TUM的freiburg2_pioneer_slam数据集,它包含2 921帧RGB-D图像,并且提供了真实位姿信息,路径长度为40.380 m.

由于数据集没有提供闭环信息,需要自行利用真实位姿增加闭环. 本文采用文献[22]的方法为数据集增加闭环信息. 利用SLAM中的配准算法对数据集提取关键帧,然后对于每一对关键帧通过其位姿相对变换的程度衡量其是否形成闭环.

每一帧关键帧表示为$f_i(i = 1,2,\cdots,N)$,则每对关键帧位姿之间的距离可表示为

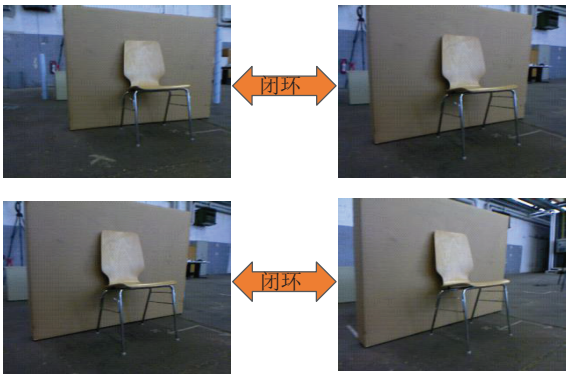
$$D_{i,j} = \mathrm{dis}({f_i}^{-1}f_j) + \mathrm{angle}({f_i}^{-1}f_j), \qquad (13)$$

其中函数$\mathrm{dis}(\cdot)$和$\mathrm{angle}(\cdot)$分别衡量转换矩阵${f_i}^{-1}f_j$的平移部分和旋转部分. 如果$D_{i,j}$小于设定的阈值,则证明机器人在获取两帧关键帧时的位置和姿态是相近的,然后将此关键帧对标记为闭环.

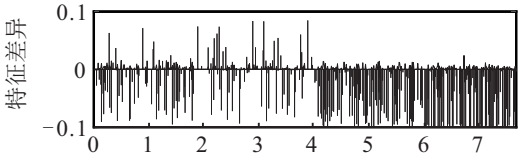
训练好CAEs后,可以利用其提取图像的特征. 考虑到本文所用数据集的图像分辨率为640×480,图像特征维数过高,在图像预处理之后,本文采用栈式卷积自编码实现特征提取,并且达到降维的目的,最终每帧图像用768维的向量描述. 此方法的优势在于:以较低维数的向量表征图像特征的同时,能够较准确地检测区分闭环与非闭环.

在CAEs提取特征向量之后,通过相似性函数公式(6)计算得到图像与图像之间的相似度得分. 在freiburg2_pioneer_slam数据集中,图像序号376与588构成闭环,图像序号591与2 899构成闭环,如图4(a)所示.

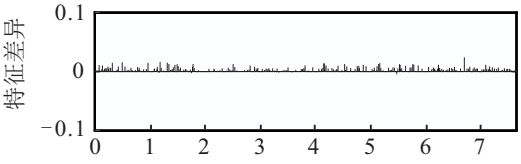
若两个关键帧形成闭环,则经CAEs优化后得到的特征向量各元素基本是相等的,而未形成闭环的两帧特征向量各元素差异很大. 本文通过CAEs提取特征得到768维的特征向量,将两张图像的CAEs特征向量中的各元素作差,得到一个新的768维向量,将该向量可视化得到特征差异. 将闭环对与非闭环对的关键帧图像特征向量中各元素作差,得到如图4(b)和图4(c)所示的非闭环处和闭环处的特征差异.



(a) 检测出的闭环



(b) 非闭环处的特征对



(c) 闭环处特征对

图4 图像特征差异对比

3.3 基于CAEs闭环检测的SLAM系统性能评测

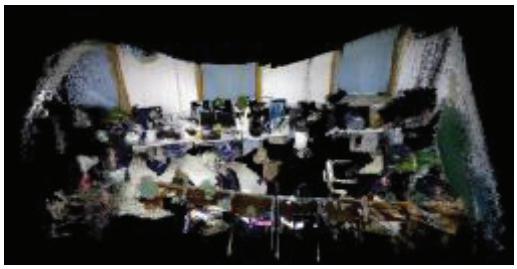
本文的室内场景RGB图和深度图来源于NYU-Depth V2数据集,为测试CAEs闭环检测在SLAM系统中的性能,提取782张照片构建SLAM三维点云地图. 通过位姿策略筛选关键帧进行三维点云的构建. 为了评估CAEs模型在整个SLAM系统中的有效性,本文分别采用基于位姿的闭环检测(传统算法)、基于自编码(Auto-Encoder)的闭环检测(深度学习算法)和本文提出的基于CAEs模型闭环检测进行SLAM地图构建实验. 从图5(a)和图5(c)可以看出,基于位姿的闭环检测方法构建的地图存在明显的误差,

基于AE的闭环检测方法构建的地图出现了点云的缺失,而经过CAEs模型闭环检测构建的地图很好地消除了这种误差和缺失.

(a) 基于位姿闭环检测的NYU数据集构图效果

(b) 基于AE闭环检测的NYU数据集构图效果

(c) 基于CAEs闭环检测的NYU数据集构图效果



(d) 基于CAEs模型的实际环境构图效果

图5　NYU数据集构图效果

为了验证本文算法的实用性,利用已经训练完成的CAEs网络在实际环境(实验室)中进行SLAM构图实验. 通过RGB-D传感器(Kinect)采集数据,筛选关键帧之后将其送入CAEs检测闭环,再交给G2O进行优化处理,最后进行SLAM地图构建,具体效果如图5(d)所示. 可以看出,通过CAEs检测闭环能够有效地消除累积误差,其构图效果较好.

本文卷积自编码的网络模型和提取图像特征的相关核心代码链接: https://github.com/Hoototo/CAE-loop-closure-of-SLAM.

4　结　论

为了提高移动机器人SLAM闭环检测的效果,本文提出了一种新的基于CAEs模型的视觉SLAM闭环检测方法. 不同于传统的基于手工特征的方法,本文直接利用深度神经网络从图像提取特征,该方法具有更好的准确性和鲁棒性. 为验证所提出的算法的有效性,本文分别在数据集和实际环境中开展了实验测试,结果表明,基于CAEs模型的闭环检测具有更好的精度和SLAM构图效果.

参考文献(References)

[1] 张亮, 蒋荣欣, 陈耀武. 移动机器人在未知环境下的同步定位与地图重建方法[J]. 控制与决策, 2010, 25(4): 515-520.
(Zhang L, Jiang R X, Chen Y W. An improved fast SLAM algorithm for mobile robots' simultaneous localization and mapping in unknown environments[J]. Control and Decision, 2010, 25(4): 515-520.)

[2] Williams B, Klein G, Reid I. Automatic relocalization and loop closing for real-time monocular SLAM[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2011, 33(9): 1699-1712.

[3] Kawewong A, Tongprasit N, Hasegawa O. A speeded-up online incremental vision-based loop-closure detection for long-term SLAM[J]. Advanced Robotics, 2013, 27(17): 1325-1336.

[4] 赵洋, 刘国良, 田国会, 等. 基于深度学习的视觉SLAM综述[J]. 机器人, 2017, 39(6): 889-896.
(Zhao Y, Liu G L, Tian G H, et al. A survey of visual SLAM based on deep learning. robot[J]. Robot, 2017, 39(6): 889-896.)

[5] Ng P C, Henikoff S. SIFT: Predicting amino acid changes that affect protein function[J]. Nucleic Acids Research, 2003, 31(13): 3812-3814.

[6] Bay H, Ess A, Tuytelaars T, et al. Speeded-up robust features(SURF)[J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.

[7] Shekhar R, Jawahar C V. Word image retrieval using bag of visual words[C]. IEEE 10th IAPR Int Workshop on Document Analysis Systems(DAS). Gold Coast: IEEE, 2012: 297-301.

[8] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]. Advances in neural information processing systems. Lake Tahoe: Curran Associates Inc, 2012:

1097-1105.

[9] Szegedy C, Toshev A, Erhan D. Deep neural networks for object detection[C]. Advances in Neural Information Processing Systems. Lake Tahoe: Curran Associates Inc, 2013: 2553-2561.

[10] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. Proc of the 32nd Int Conf on Machine Learning, 2015, 37: 448-456.

[11] He K, Zhang X, Ren S, et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification[C]. Proc of the IEEE Int Conf on Computer Vision. Santiago: IEEE Computer Society Washington, 2015: 1026-1034.

[12] Masci J, Meier U, Cirean D, et al. Stacked convolutional auto-encoders for hierarchical feature extraction[J]. Artificial Neural Networks and Machine Learning-ICANN, 2011, DOI: 10.1007/978-3-642-21735-7_7.

[13] Cummins M, Newman P. Appearance-only SLAM at large scale with FAB-MAP 2.0[J]. The Int J of Robotics Research, 2011, 30(9): 1100-1123.

[14] Gálvez-López D, Tardos J D. Bags of binary words for fast place recognition in image sequences[J]. IEEE Trans on Robotics, 2012, 28(5): 1188-1197.

[15] Mei C, Sibley G, Newman P. Closing loops without places[C]. IEEE/RSJ Int Conf on IEEE Intelligent Robots and Systems(IROS). Taipei: IEEE Computer Society Washington, 2010: 3738-3744.

[16] Murphy L, Sibley G. Incremental unsupervised topological place discovery[C]. IEEE Int Conf on Robotics and Automation(ICRA). Hong Kong: IEEE, 2014: 1312-1318.

[17] Kejriwal N, Kumar S, Shibata T. High performance loop closure detection using bag of word pairs[J]. Robotics and Autonomous Systems, 2016, 77: 55-65.

[18] Khan S, Wollherr D. IBuILD: Incremental bag of binary words for appearance based loop closure detection[C]. IEEE Int Conf on Robotics and Automation(ICRA). Seattle: IEEE, 2015: 5441-5447.

[19] 李博, 杨丹, 邓林. 移动机器人闭环检测的视觉字典树金字塔TF-IDF得分匹配方法[J]. 自动化学报, 2011, 37(6): 665-673.
(Li B, Yang D, Deng L. Visual vocabulary tree with pyramid TF-IDF scoring match scheme for loop closure detection[J]. Acta Automatica Sinica, 2011, 37(6): 665-673.)

[20] Sünderhauf N, Shirazi S, Dayoub F, et al. On the performance of convnet features for place recognition[C]. IEEE/RSJ Int Conf on Intelligent Robots and Systems(IROS). Hamburg: IEEE, 2015: 4297-4304.

[21] Xu Yan, Tao Mo, Qiwei Feng, et al. Deep learning of feature representation with multiple instance learning for medical image analysis[C]. Int Conf on Acoustics, Speech and Signal Processing(ICASSP). Florence: IEEE, 2014: 1626-1630.

[22] 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.

[23] Xia Y, Li J, Qi L, et al. Loop closure detection for visual SLAM using PCANet features[C]. IEEE Int Joint Conf on Neural Networks. Vancouver: IEEE, 2016: 2274-2281.

[24] 罗杨宇, 刘宏林. 基于光束平差法的双目视觉里程计研究[J]. 控制与决策, 2016, 31(11): 1936-1944.
(Luo Y Y, Liu H L. Research on binocular vision odometer based on bundle adjustment method[J]. Control and Decision, 2016, 31(11): 1936-1944.)

[25] 季秀才, 郑志强, 张辉. SLAM问题中机器人定位误差分析与控制[J]. 自动化学报, 2008, 34(3): 323-330.
(Ji X C, Zheng Z Q, Zhang H. Analysis and control of robot position error in SLAM.[J]. Acta Automatica Sinica, 2008, 34(3): 323-330.)

[26] Cummins M, Newman P. FAB-MAP: Probabilistic localization and mapping in the space of appearance[J]. The Int J of Robotics Research, 2008, 27(6): 647-665.

[27] Gao X, Zhang T. Loop closure detection for visual slam systems using deep neural networks[C]. The 34th Chinese Control Conf(CCC). Hangzhou: IEEE, 2015: 5851-5856.

[28] Huang F J, Boureau Y L, LeCun Y. Unsupervised learning of invariant feature hierarchies with applications to object recognition[C]. IEEE Conf on Computer Vision and Pattern Recognition(CVPR). Minneapolis: IEEE, 2007: 1-8.

[29] Ng A, Ngiam J, Foo C Y, et al. UFLDL tutorial[EB/OL]. http://deeplearning.stanford.edu/wiki/index.php/UFLDL_Tutorial.

(责任编辑：闫 妍)