融合 Gist 特征与卷积自编码的闭环检测算法

邱晨力[1,2]，黄东振[1,2]，刘华巍[1]，袁晓兵[1]，李宝清[1*]

[1]中国科学院上海微系统与信息技术研究所微系统技术重点实验室，上海 201800；

[2]中国科学院大学，北京 100049

摘要 闭环检测算法可消除视觉同时定位与建图(VSLAM)系统的累计误差，并对构建全局一致性地图有重要作用。针对现有传统闭环检测算法在视角与场景外观变化下准确率与稳健性降低，及部分基于深度学习方法特征提取与闭环识别实时性不佳的问题，设计了一种融合 Gist 特征与卷积自编码的闭环检测算法，将 Gist 特征作为卷积自编码网络重构目标，可增强模型在外观变化下的场景特征表达能力；同时通过透视变换构造视角变化训练图像对，以提升模型在视角变化下闭环检测的准确率与稳健性。所设计的模型较精简，可实现实时关键帧特征提取与闭环检测。在 Gardens Point 与 Nordland 数据集的实验结果表明，相较于传统视觉词袋模型(BoVW)、Gist 算法及现有部分深度学习方法，本文算法可以达到更高的准确率和稳健性。

关键词 机器视觉；同时定位与建图；闭环检测；卷积自编码；深度学习

中图分类号 TP242 **文献标识码** A **doi**：10.3788/LOP56.181501

Loop Closure Detection Algorithm Based on Convolutional Autoencoder Fused with Gist Feature

Qiu Chenli[1,2], Huang Dongzhen[1,2], Liu Huawei[1], Yuan Xiaobing[1], Li Baoqing[1*]

[1] *Science and Technology on Microsystem Laboratory, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai* 201800, *China*;

[2] *University of Chinese Academy of Sciences, Beijing* 100049, *China*

Abstract Loop closure detection algorithm is essential for the visual simultaneous localization and mapping (VSLAM) systems to reduce accumulative error and build a globally consistent map. When detecting loops under the change of viewpoint and scene appearance, the precision and robustness of traditional loop closure detection algorithms decline and some algorithms based on deep learning are difficult to extract features and perform loop closure detection in real time. To overcome these problems, we propose a novel loop closure detection algorithm based on convolutional autoencoder fused with Gist feature, forcing the encoder to reconstruct the Gist feature to enhance the expressive ability of the model when the scene appearance changes. In the same time, we warp images with randomized projective transformations to make the training pairs to improve the precision and robustness of the model when the viewpoint changes. Our model is relatively lightweight which is capable of extracting keyframe features and detecting loops in real time. The results of experiments on Gardens Point and Nordland datasets show that our model can achieve better precision and robustness compared with traditional methods, like bag of visual word (BoVW), Gist, and some other methods based on deep learning.

Key words machine vision; visual simultaneous localization and mapping (VSLAM); loop closure detection; convolutional autoencoder; deep learning

OCIS codes 150.4065; 200.4260; 330.7310

收稿日期：2019-01-12；**修回日期**：2019-01-31；**录用日期**：2019-04-09

基金项目：微系统技术国防科技重点实验室基金(CXJJ-17S072)

* **E-mail**：sinoiot@mail.sim.ac.cn

1 引　　言

近年来，基于视觉的同时定位与建图(VSLAM)在机器人和智能驾驶等领域得到了广泛应用。当前 VSLAM 算法主要由视觉里程计[1-2]、后端优化、闭环检测和建图四个模块构成。若 VSLAM 系统中缺少闭环检测模块，则系统就相当于一个视觉里程计，由于前端视觉里程计只考虑局部位姿约束，所以必然会产生累积误差，而系统中存在闭环检测模块就可以判断移动机器人当前是否回到曾经访问过的位置，从而为后端优化提供有效的数据，使得整个 VSLAM 系统得到全局一致的估计结果。在大范围 VSLAM 系统的长时间测量过程中，若有效检测出真阳性闭环就可以显著降低系统的累积误差，但假阳性闭环则会使系统后端优化算法收敛到完全错误的值。因此，对于 VSLAM 系统而言，做到稳定有效的闭环检测至关重要。

当前主流的闭环检测是基于外观实现，并脱离前后端估计，只依靠关键帧图像的相似性来确定闭环检测，故闭环检测的核心问题分为图像描述和相似性度量两个部分。而图像场景描述方法主要包括基于局部特征描述子、基于全局描述子及基于深度学习的特征描述子三类。其中基于局部特征描述子的场景描述方法如视觉词袋模型(BoVW)[3]，被广泛应用在现有的 VSLAM 系统中进行闭环检测工作。BoVW 算法首先对图像的关键帧提取人为设计特征[如加速稳健特征(SURF)[4]、定向快速二进制特征(ORB)[5]等]，并对提取的特征描述子进行聚类以构造字典树，如基于外观的快速建图算法(FAB-MAP)[6]构造的 Chou-Liu 树。BoVW 算法利用字典树对每个图像的关键帧作图像描述，进而度量距离检测闭环。而 Mur-Artal 等[7]提出的基于 ORB 特征的 VSLAM 系统就借助了 BoVW 算法，获得较为稳定且准确的闭环检测效果。由于字典树的视觉单词需要预训练得到，所以视觉词袋模型的泛化能力尤为重要，尽管局部特征描述子对视角变化及光照变化具有一定的稳健性，但面对大范围的外界场景变化仍难以保持较高的闭环检测准确率和稳健性。基于全局特征描述子的图像描述方法则有 Gist 算子[8-9]，通过将整体图像划分网格，并与不同方向和尺度的 Gabor 滤波器组进行滤波，构造全局特征描述子以进行闭环检测，若直接利用 Gist 特征进行闭环检测，发现其对视角变化及大范围外部环境变化较为敏感，且检测准确率欠佳。

近年来，由于深度学习技术在视觉场景下分类与识别精度两个方面的优异表现，基于深度学习的方法越来越频繁地出现在闭环检测的解决方案中。Hou 等[10]利用基于 Places[11]数据集与 AlexNet 预训练的 Places-CNN 模型进行闭环检测，并对比 BoVW 及 Gist 下的闭环检测方法，可发现在光照变化的场景下，深度学习方法较 BoVW 与 Gist 算法具有明显优势。Sünderhauf 等[12]利用 AlexNet 网络评估了场景视角变化及外观变化对闭环检测的影响，发现中间层对外观变化具有更好的稳健性，而高层特征则对视角变化的稳健性更佳。Gao 等[13-14]利用堆栈式自编码器构建闭环检测模型以去除解码层，并将隐藏层的输出作为特征提取器，以构建相似度矩阵进行闭环检测，发现该模型的特征维度高且对视角及大范围场景变化的稳健性较弱。Lopez-Antequera 等[15]构造了同一地点的一组存在视角及场景外观变化的三元训练组，并结合 Hinge 损失函数对卷积神经网络进行训练，发现在全新场景环境中，算法网络泛化性能有待提高。Merrill 等[16]结合方向梯度直方图特征(HOG)构建深度自编码模型，发现其在光照与视角变化剧烈的测试集中性能良好，但 HOG 特征在场景识别中的准确率受限。鲍振强等[17]则结合 YOLO[18]目标检测算法来过滤场景干扰项，进而利用 VGG19 提取场景特征进行闭环检测，发现其具有较高的准确率与稳健性，但干扰项过滤且关键帧特征提取较为耗时。

为提高当前闭环检测算法在视角及大范围场景外观变化下的准确率与稳健性，同时满足闭环检测的实时性要求，本文先对原始 Places365-Standard[11]数据集进行视角变化的数据增强，再构造视角变化训练图像对，并融合 Gist 特征构建一种深度卷积自编码模型，最后将训练得到的卷积自编码模型的编码层作为 VSLAM 关键帧的特征提取器，从而实现实时的关键帧特征提取与闭环识别。该模型作为一种无监督学习方法，与传统方法相比，既可保证模型精简，关键帧特征提取与闭环识别速度较快，也能达到优于对比方法的准确率与稳健性。

2 融合 Gist 特征的卷积自编码模型

本文提出的融合 Gist 特征与卷积自编码的闭环检测模型，主要包括视角变化训练图像对的构建、VSLAM 关键帧的 Gist 特征提取、融合 Gist 特征的卷积自编码模型的搭建训练及 VSLAM 场景下关键帧闭环检测流程。

2.1 Places 视角变化训练图像对构建

Places365-Standard[11]数据集是美国麻省理工学院(MIT)的研究团队为实现场景视觉感知而构造的一个超过 180 万张场景图片及 365 种场景类别的数据集。尽管在每一种场景类别中都包含 5000～30000 张场景图像数据,但为使训练得到的卷积神经网络具有一定的抗视角变化能力,将参考 Merrill 等[16]的方法对原始数据集构建视角变化训练图像对。假设原始图像 I 的大小为 $W\times H$,则四个角点坐标分别为$(0,0)$,$(0,H)$,$(W,0)$,(W,H)以构成坐标点组 P_o。如图 1 所示,左侧为原始图像,通过对原始图像四个边角取长和高分别为 $W/4$ 与 $H/4$ 的边缘框(图 1 方框),并在边缘框中随机取点(x_1,y_1),(x_2,y_2),(x_3,y_3),(x_4,y_4)构成坐标点组 P_r,P_o 与 P_r 坐标点组对应点如图 1 实线所示。计算 P_o 与 P_r 间的透视变换矩阵,即

$$[x,y,z]=\boldsymbol{A}[u,v,1], \tag{1}$$

式中:(x,y,z)为 P_r 中的目标点三维坐标;$(x/z,y/z)$为 P_r 中的目标点二维坐标;$\boldsymbol{A}$ 为一个 3×3 的透视变换矩阵;(u,v)为 P_o 中的源目标点。因此可通过坐标点组 P_o 与 P_r 的代入求解透视变换矩阵 $\boldsymbol{A}$,并利用透视变换矩阵 $\boldsymbol{A}$ 对原始图像重投影即可得到视角变化图像 I_w。图 1 即为对 Places365-Standard 数据集样例图像进行视角变化后得到的训练图像对$\langle I,I_w\rangle$。在本文算法中,将利用$\langle I,I_w\rangle$作为训练模型的两路输入,以提升训练模型的抗视角变化性能。

图 1 视角变化训练图像对$\langle I,I_w\rangle$

Fig. 1 Training pairs with viewpoint change $\langle I,I_w\rangle$

2.2 VSLAM 关键帧 Gist 特征提取

Gist 特征[8-9]是一种生物启发特征,根据场景图像的空间包络模型来描述场景全局特征,其中利用 Gabor 滤波器组可获取场景图像在不同空间频率、位置与方向选择性的局部结构信息,因此常用于场景识别分类及图像检索任务中。VSLAM 关键帧的 Gist 特征提取流程及相应的参数设置如下。

1) 将一幅大小为 $w\times h$ 的关键帧图像 $f(x,y)$ 进行网格划分,划分为 $r\times c$ 个规则的网格,r 与 c 分别为网格行数与列数,子网格的规格大小均一致,可得到相应个数的网格子区域 P_i,i 为子网格编号,P_i 大小为 $w'\times h'$,其中 $w'=w/c$,$h'=h/r$,取 $r=c=4$,则 $i=1,2,3,\cdots,16$。

2) 对单通道图像构建 m 尺度 n 方向的 Gabor 滤波器组,对 $r\times c$ 个网格子区域进行卷积滤波,得到每个网格子区域的局部 Gist 特征,即

$$G_i(x,y)=\mathrm{cat}[f(x,y)*g_{mn}(x,y)],\quad (x,y)\in P_i, \tag{2}$$

式中:cat(·)为网格子区域经过 Gabor 卷积滤波后的级联运算;* 表示卷积运算;Gabor 滤波器组 $g_{mn}(x,y)$是基于 Gabor 函数的自相似性函数,通过对母小波进行多尺度与多方向扩展变换得到;m 为滤波器组尺度数;n 为滤波器组方向数。构造一个 3 个尺度方向数量分别为 8,8,4 的 Gabor 滤波器组,因此局部网格子区域的 Gist 特征维数为 20。

3) 对网格子区域计算出的局部 Gist 特征取平均值,得到局部子区域 Gist 特征分别为 $\bar{G}_1,\bar{G}_2,\cdots,\bar{G}_{16}$。依次级联各个网格子区域的 Gist 特征向量,构造全局 Gist 特征向量 $\boldsymbol{G}$,其表达式为

$$\boldsymbol{G}=[\bar{G}_1,\bar{G}_2,\cdots,\bar{G}_{16}]。 \tag{3}$$

对 RGB 三通道图像进行 Gist 特征提取,得到全局 Gist 特征维数为 $20\times4\times4\times3=960$。

2.3 融合 Gist 特征与卷积自编码的闭环检测模型

结合 2.1 节中视角变化训练图像对$\langle I,I_w\rangle$与 2.2节中的关键帧 Gist 全局特征,设计了一种融合 Gist 特征的卷积自编码模型。Gao 等[13-14]也提出了一种无监督闭环检测方法,首先对原始图像进行分块并序列化,然后将其作为堆栈式自编码器的输入,堆叠多个去噪自编码器,将上层隐藏层输出作为下层输入,并通过重构原始输入的误差来逐层对网络进行训练。但基于堆栈式自编码网络的闭环检测对视角变化及大范围场景外观变化的稳健性较差。本研究一方面构造视角变化训练图像对$\langle I,I_w\rangle$,以提升模型在视角变化下的闭环检测性能;另一方面将 Gist 特征作为深度卷积自编码网络重构目标,使得深度卷积自编码网络中间层特征融合 Gist 描述子的场景特征表达能力。图 2 为所提网络结构示意图,融合 Gist 特征的深度自编码网络的输入为两路,即 2.1 节中构建的视角变化训练图像对$\langle I,I_w\rangle$,并采用随机算法选择训练图像对$\langle I,I_w\rangle$中的原始图像 I 与视角变化图像 I_w 分别作为训练模型的两路输入,其中一路对输入图像进行 2.2 节中所述的全局Gist特征提取(CalcGist),得到固定长度为

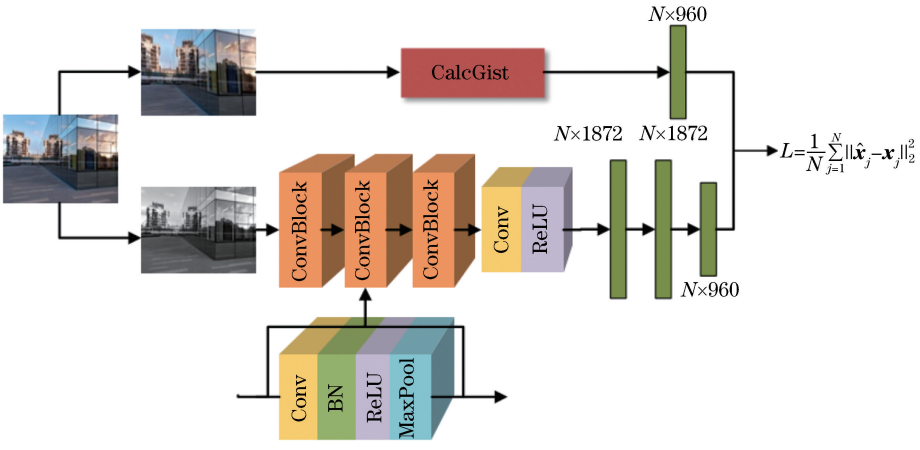

图 2 融合 Gist 特征的深度卷积自编码模型结构

Fig. 2 Training pipeline of convolutional autoencoder fused with Gist feature

960 维关键帧特征向量；而另外一路输入图像则预处理为灰度图像以精简网络模型大小，卷积自编码网络编码器部分采用轻量化设计，为融合 Gist 特征的场景特征表达能力，将通过三层全连接层作为解码器重构 960 维 Gist 特征表达。

本模型的训练图像输入大小为 160 pixel×120 pixel，深度卷积自编码网络部分由编码器与解码器两个部分组成。模型编码器部分由四个卷积单元构成，前三层卷积单元（ConvBlock）均由卷积（Conv）、批量归一化[19]（BN）、线性修正单元[20]（ReLU）及最大池化（MaxPool）构成。第四层单元则经过 Conv 与 ReLU 线性输入解码器。考虑到本文模型的轻量化需求，编码器部分均采用较小尺寸卷积核进行深度特征的提取，在保证局部感受野大小的同时，以减小模型参数量，而编码器各层 ConvBlock 中第一层采用了 32 个卷积核大小为 5×5、步长为 1、补零为 2 的 Conv 层，而后续两个 ConvBlock 则均采用了 64 个卷积核大小为 3×3、步长为 1、补零为 1 的 Conv 层。为加速网络训练过程，避免训练过程中梯度弥散，及网络权重初始化问题，在前三层卷积层后都依次级联 BN 和 ReLU 以提升网络性能。同时为降低编码器特征输出维度，并简化网络整体参数，缩减 VSLAM 关键帧特征提取与闭环查询时间，前三层 ConvBlock 最后均加入窗口大小为 3×3、步长为 2 的 MaxPool 层。编码器第四层由 8 个卷积核大小为 3×3、步长为 1 的 Conv 层与 ReLU 构成，编码器第四层的输出经过 Flatten 操作即为本文闭环检测模型的最终关键帧特征向量。

卷积自编码网络解码器则由三层全连接网络构成，将编码器输出重构为 960 维 Gist 特征向量，三层全连接层维度分别为 1872，1872 和 960。由于 Gist 特征向量分布在[0，1]之间，为更好地重构 Gist 特征分布，将训练模型在全连接层后加入 Sigmoid 激活层，并使特征输出分布在[0，1]之间。在训练过程中，全连接层输出 960 维特征 $\hat{\boldsymbol{x}}$ 与另一路匹配图像对的 960 维 Gist 特征 $\boldsymbol{x}$ 进行欧氏损失（Euclidean loss）计算，欧氏损失函数定义为

$$L=\frac{1}{N}\sum_{j=1}^{N}\|\hat{\boldsymbol{x}}_j-\boldsymbol{x}_j\|_2^2, \tag{4}$$

式中：L 为损失值；N 为网络每次迭代输入的训练图像对数量；$\boldsymbol{x}_j$ 为第 j 组训练图像对全局 Gist 特征向量；$\hat{\boldsymbol{x}}_j$ 为第 j 组训练图像对深度自编码网络输出特征向量。网络采用随机梯度下降法对损失函数进行优化。

2.4 闭环检测流程

如图 3 所示，对 VSLAM 场景中的第 k 个关键帧进行分析，将卷积自编码网络的编码器作为特征提取器得到关键帧特征向量 $\boldsymbol{v}_k$，并将其与关键帧数据库中历史关键帧特征向量进行相似性得分计算，$\boldsymbol{v}_{N_q}$ 为最后一帧历史关键帧特征向量。

假设存在两个关键帧 f_{k1} 与 f_{k2}（$k1$ 和 $k2$ 为关键帧编号），每个关键帧通过深度自编码网络编码器可以得到长度为 t 的特征向量 $\boldsymbol{v}_{k1}$ 与 $\boldsymbol{v}_{k2}$。采用余弦相似度作为关键帧相似性度量标准，余弦相似度计算式为

$$s=\frac{\sum_{l=1}^{t}\boldsymbol{v}_{k1}^{(l)}\boldsymbol{v}_{k2}^{(l)}}{\sqrt{\sum_{l=1}^{t}(\boldsymbol{v}_{k1}^{(l)})^2}\sqrt{\sum_{l=1}^{t}(\boldsymbol{v}_{k2}^{(l)})^2}}, \tag{5}$$

式中：s 为关键帧 f_{k1} 与 f_{k2} 的相似性得分；l 为特征

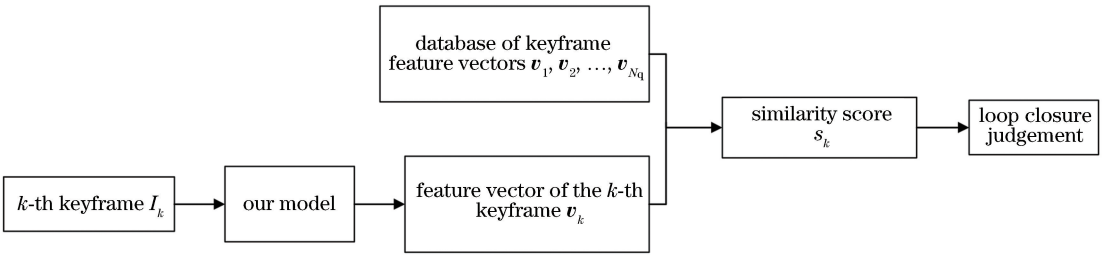

图 3 闭环检测流程示意图

Fig. 3 The flowchart of loop closure detection

向量元素序号。在闭环检测过程中，计算当前第 k 个关键帧与历史关键帧最大相似性得分 s_k，设定相似性阈值 δ，若 s_k 大于 δ，则判断关键帧对形成闭环；否则判断未检测出闭环。在实际工程应用中，闭环检测算法的目的是在保证高准确率的前提下，尽可能提高召回率。

3 实验与结果分析

实验仿真的计算机硬件配置 CPU 为 Intel Xeon E5-2620 V4，GPU 为 Nvidia GeForce GTX 1080Ti，操作系统为 Ubuntu 16.04，使用 Caffe 作为深度学习训练框架。算法均基于 Python2.7 软件实现，准确率-召回率曲线绘制基于 Matplotlib 库。

3.1 实验数据集及评价标准

实验选取 Gardens Point 与 Nordland 数据集作为闭环检测算法测试数据集，如图 4 所示，Gardens Point 数据集是在澳洲昆士兰科技大学采集的三个视频序列，其中图 4(a)和图 4(b)为两个白天光照良好条件下，分别靠行道左侧与右侧记录的连续视频关键帧(day_left，day_right)，图 4(c)为一个夜间弱光条件下记录的连续视频关键帧(night_right)，三个视频序列关键帧的记录长度均为 200 frame，实验选取图 4(a)和图 4(b)两组图像测试对比算法对视角变化的准确率和稳健性；选取图 4(a)和图 4(c)两组图像测试视角与光照剧烈变化下对比算法的准确率和稳健性。如图 4(d)和图 4(e)所示，Nordland 数据集包含四个季节的火车旅程视频序列，对比序列包含显著的季节变化，本文选取文献[16]中提供的两组共 344 frame 的春冬两季的视频序列 Nordland_spring 与 Nordland_winter 进行算法性能测试。

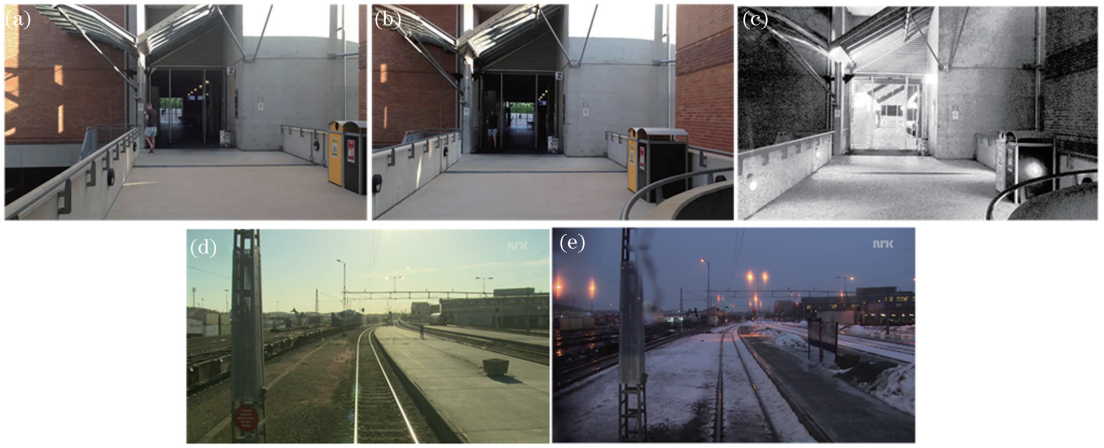

图 4 Gardens Point 和 Nordland 测试数据集的部分图像。(a)白天_左侧；(b)白天_右侧；(c)夜晚_右侧；(d)春季序列；(e)冬季序列

Fig. 4 Some pictures from Gardens Point and Nordland test datasets. (a) Day_left; (b) day_right; (c) night_right; (d) spring sequence; (e) winter sequence

把准确率-召回率曲线的 AUC(area under the curve)值及算法关键帧特征提取时间作为闭环检测算法性能评价的两个指标。AUC 值为准确率-召回率曲线下方面积，AUC 值越接近于 1 则说明算法平均准确率越高，性能越好。而准确率是闭环检测模块检测到的真阳性闭环数量(正确检测出的闭环数量)与所有检测出的闭环数量的比值；召回率是真阳性闭环数量与数据集中真实闭环数量的比值，准确

率与召回率的表达式为

$$P=\frac{T_{\mathrm{P}}}{T_{\mathrm{P}}+F_{\mathrm{P}}},\tag{6}$$

$$R=\frac{T_{\mathrm{P}}}{T_{\mathrm{P}}+F_{\mathrm{N}}},\tag{7}$$

式中：P 与 R 分别为准确率与召回率；T_{P} 为正确检测到的真实闭环数量；F_{P} 为检测到的错误闭环数量，T_{P} 与 F_{P} 之和即为模型所检测到的闭环总量；F_{N} 为未检测到的真实闭环数量，T_{P} 与 F_{N} 之和即为测试数据集中存在的真实闭环总量。对每一个查询关键帧 f_k，由(5)式计算它在历史关键帧中的最大相似性得分 s_k，构成最大相似性得分向量 $\boldsymbol{s}_{\mathrm{scores}}=[s_1,s_2,\cdots,s_k,\cdots,s_{N_{\mathrm{q}}}]$；判断最大相似性得分历史关键帧与原始查询关键帧是否匹配闭环得到闭环判断值 c_k(匹配时 c_k 为 1，不匹配时 c_k 为 0)，构成闭环判断向量 $\boldsymbol{c}=[c_1,c_2,\cdots,c_k,\cdots,c_{N_{\mathrm{q}}}]$，其中 N_{q} 为总查询关键帧数。通过动态调整闭环检测相似性阈值 δ，查询最大相似性得分向量 $\boldsymbol{s}_{\mathrm{scores}}$ 与闭环判断向量 $\boldsymbol{c}$，计算对应阈值 δ 下的 T_{P}、F_{P} 和 F_{N}，进而得到准确率与召回率即可绘制模型的准确率-召回率曲线。

3.2 实验细节

模型训练数据集来源于 Places365-Standard 数据集的 365 个场景类别中超过 180 万张分辨率为 160 pixel×120 pixel 的场景图像，并经过视角变化得到的训练图像对$\langle I,I_{\mathrm{w}}\rangle$作为输入。通过 2.3 节构造融合 Gist 特征的卷积自编码模型并结合欧氏距离损失及随机梯度下降法对模型进行训练，将训练得到的深度自编码网络编码器部分作为特征提取器，获得 1872 维关键帧特征向量。训练初始学习率为 0.0009，权重衰减值为 0.0005，训练 batch size 为 256，epoch 设置为 42。

对比算法有传统的 BoVW[3] 算法、Gist[8-9] 算法及基于深度学习方法的 CALC[16] 算法与 Places-CNN[12] 算法。而 BoVW 算法基于开源 DBoW2[3] 及 ORB 特征[5]得以实现。Gist 算法则采用与本文模型所融合的相同 Gist 特征，即对 RGB 图像划分子网格数量为 4×4，并利用 3 个尺度方向数量分别为 8、8、4 的 Gabor 滤波器组构造 960 维 Gist 特征。CALC 模型训练数据集同样采用 Places365-Standard 数据集，测试图像输入大小为 160 pixel×120 pixel，文献[16]中通过编码器第三层卷积单元激活层输出构造 1064 维关键帧特征向量。Places-CNN 则基于 AlexNet 卷积神经网络，通过 Places205 数据集对原始网络进行训练，测试图像输入大小为 227 pixel×227 pixel，由于文献[12]中发现 Conv3 层在场景识别问题上稳健性最佳，因此选取 Conv3 层作为关键帧特征提取器，输出 64896 维关键帧特征向量。

3.3 算法性能对比分析

对本文提出的闭环检测算法与传统 BoVW 算法[3]、Gist 算法[8-9]及基于深度学习方法的 CALC 算法[16]与 Places-CNN 算法[12]进行性能比较。根据 3.1 节中的评价指标及 3.2 节中的实验细节，基于 Gardens Point 与 Nordland 测试数据集计算对比算法的准确率与召回率，绘制五种对比算法的准确率-召回率曲线，如图 5 所示。

如图 5(a)所示，在 Gardens Point 数据集白天光照环境下的关键帧序列 day_left 与 day_right 性能测试中，闭环场景中主要存在场景视角变化，融合 Gist 特征的深度自编码模型所呈现的准确率-召回率曲线 AUC 值达到 0.89，而对比算法 CALC、Places-CNN、BoVW 及 Gist 的 AUC 则分别为0.84，0.83，0.84，0.63，表明本文模型在视角变化场景下平均准确率上相较于对比算法效果最佳。当闭环检测召回率达到 80%时，融合 Gist 特征的深度自编码模型准确率达到 86.5%，而对比算法 CALC、Places-CNN、BoVW 及 Gist 分别为 71.0%，75.0%，71.1%，60.6%，说明本文模型在高召回率条件下仍然保持了较高的检测准确率。如图 5(b)所示，在 day_left 与 night_right 性能测试中，闭环场景中包括了视角及剧烈光照场景环境变化，可以发现在视角及大范围场景光照环境变化情况下，本文模型的平均准确率在对比算法中达到最高，准确率-召回率曲线 AUC 达到 0.77，相较于排名第二的 CALC 算法提升了 0.05。如图 5(c)所示，Nordland 春冬两季火车旅程测试序列包含显著的季节场景变化，本文算法的测试平均准确率同样优于其余四种对比算法，准确率-召回率曲线 AUC 达到 0.74，相较于排名第二的 CALC 算法也提升了 0.05。特别需要注意的是，本文模型相对于直接采用相同参数的 Gist 特征闭环检测方法，在三组闭环测试中准确率-召回率曲线的平均 AUC 提升了 0.43，说明了本文融合 Gist 特征的卷积自编码模型的有效性。由图 5 实验结果可知，在光照乃至季节场景外观变化下的三组对比实验中，本文模型平均准确率均优于四类对比算法，并且本文算法在三类场景中准确率-召回率曲线 AUC 最大差值仅 0.12，稳健性最佳。

根据文献[21]的卷积神经网络时间复杂度公式

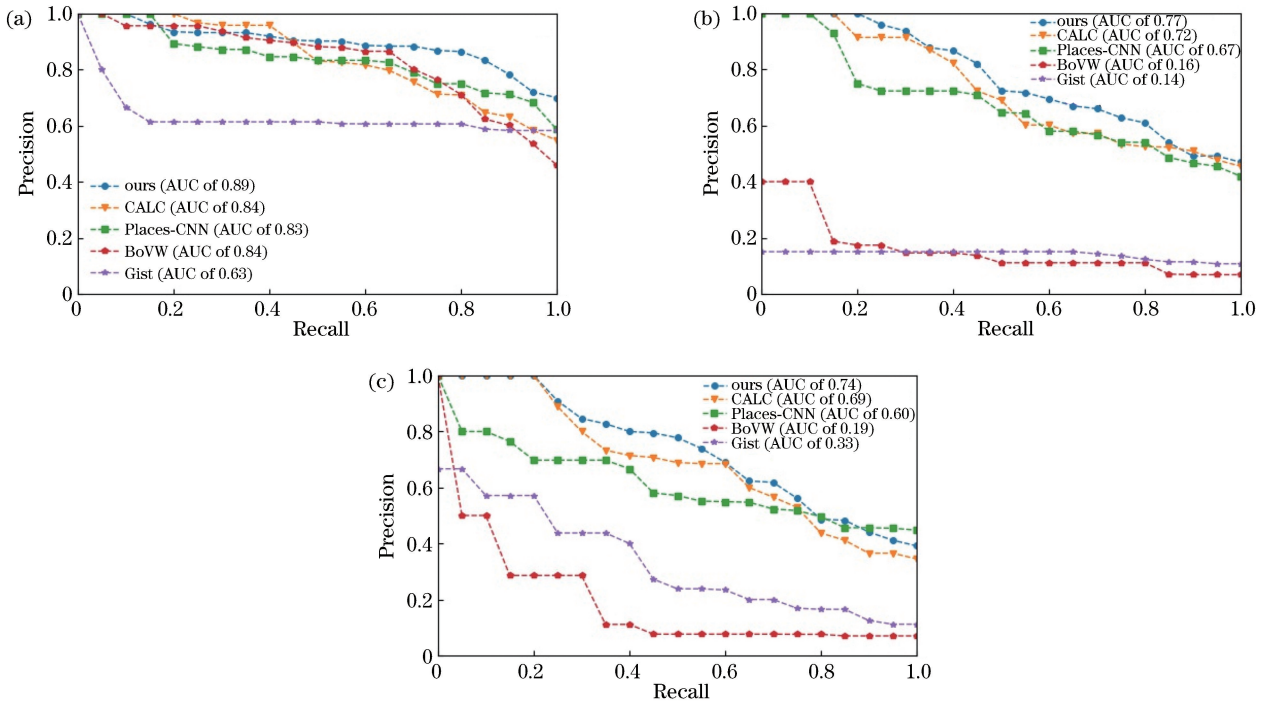

图 5 Gardens Point 与 Nordland 测试数据集闭环检测准确率-召回率曲线。(a) Gardens Point 白天_左侧与白天_右侧；(b) Gardens Point 白天_左侧与夜晚_右侧；(c) Nordland 春季序列与冬季序列

Fig. 5 Precision-recall curves for Gardens Point and Nordland datasets. (a) Gardens Point: day_left versus day_right; (b) Gardens Point: day_left versus night_right; (c) Nordland: spring versus winter

$O(\sum_{l=1}^{d} n_{l-1} f_l^2 n_l M_l^2)$，对三类卷积神经网络方法特征提取阶段进行时间复杂度分析。其中，d 为卷积层总层数；n_l 为第 l 层输出特征图数量；f_l 为第 l 层卷积核边长；M_l 为输出特征图边长。基于深度学习的闭环检测算法 Places-CNN、CALC，及本文算法时间复杂度如表 1 所示。受益于轻量化卷积自编码网络结构，本文模型相对于 Places-CNN 的闭环检测方法特征提取时间和复杂度大幅度减小；由于编码器层数相较于 CALC 算法加深，本文算法的网络整体特征提取时间和复杂度略高于 CALC 算法。

表 1 算法特征提取与闭环查询时间

Table 1 Time to extracting features and querying the database on Gardens Point dataset

Method	Extract features time /ms	Query time / μs	Complexity / 10^6
Ours(GPU)	0.81	5.88	217.15
CALC(GPU)	**0.64**	5.13	**186.14**
Places-CNN(GPU)	2.31	42.98	479.45
BoVW(CPU)	20.27	17.00	—
Gist(CPU)	41.43	**5.10**	—

实验在 Gardens Point 数据集 day_left，day_right 关键帧序列上计算特征提取平均时间及闭环查询平均时间用来具体评判算法的时间性能。但由于 BoVW 与 Gist 算法基于 CPU 实现，而本文模型、CALC 与 Places-CNN 基于 GPU 实现，因此在特征提取时间对比上需要考虑硬件因素的影响。从表 1 可以看出，本文模型关键帧特征提取时间显著小于 Places-CNN，略高于 CALC 模型。而在闭环查询时间方面，由于本文算法编码器层输出特征维度 1872 维，对比 Places-CNN 的 Conv3 层特征输出 64896 维的闭环查询时间具有显著优势，而 CALC 特征为 1064 维，Gist 特征为 960 维，因此在闭环查询时间上本文模型高于这两种算法，整体上本文模型在特征提取及闭环查询总耗时满足闭环检测的实时性要求。

实验结果表明，本文算法在视角、光照及季节场景外观变化条件下闭环检测平均准确率及稳健性优于传统的 BoVW、Gist 算法、Places-CNN 及 CALC 模型，同时满足闭环检测的实时性要求，具有较高的实用性。

4 结 论

针对当前闭环检测算法在面对视角及大范围场

景环境变化下准确率与稳健性下降的问题，设计了一种无监督模型，通过透视变换对原始训练数据进行数据增强以构造视角变化训练图像对，并结合Gist特征构造了一个深度卷积自编码网络实现无监督训练，模型在准确率-召回率曲线性能优异的情况下，由于模型的精简可实现快速的特征提取及闭环查询，可满足闭环检测的实时性要求，且具有较高的实用性。

从实验结果看出，在光照季节此类外观剧烈变化的条件下，模型整体平均准确率明显低于视角变化闭环检测，因此后续工作可对加强光照季节此类场景外观变化条件下的闭环检测算法稳健性进行深入探究。

参 考 文 献

[1] Lin Z L, Zhang G L, Yao E L, *et al*. Stereo visual odometry based on motion object detection in the dynamic scene[J]. Acta Optica Sinica, 2017, 37(11): 1115001.
林志林, 张国良, 姚二亮, 等. 动态场景下基于运动物体检测的立体视觉里程计[J]. 光学学报, 2017, 37(11): 1115001.

[2] Wang K, Jia S M, Li X Z, *et al*. Mobile robot monocular visual odometry algorithm based on ground features[J]. Acta Optica Sinica, 2015, 35(5): 0515002.
王可, 贾松敏, 李秀智, 等. 基于地面特征的移动机器人单目视觉里程计算法[J]. 光学学报, 2015, 35(5): 0515002.

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

[4] Bay H, Tuytelaars T, van Gool L. SURF: speeded up robust features[M]//Leonardis A, Bischof H, Pinz A. Computer vision-ECCV 2006. Lecture notes in computer science. Berlin Heidelberg: Springer, 2006, 3951: 404-417.

[5] Rublee E, Rabaud V, Konolige K, *et al*. ORB: an efficient alternative to SIFT or SURF[C]//2011 International Conference on Computer Vision, November 6-13, 2011, Barcelona, Spain. New York: IEEE, 2011: 2564-2571.

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

[7] Mur-Artal R, Montiel J M M, Tardós J D. ORB-SLAM: a versatile and accurate monocular SLAM system[J]. IEEE Transactions on Robotics, 2015, 31(5): 1147-1163.

[8] Oliva A, Torralba A. Building the gist of a scene: the role of global image features in recognition[J]. Progress in Brain Research, 2006, 155: 23-36.

[9] Liu Y, Zhang H. Visual loop closure detection with a compact image descriptor[C]//2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, October 7-12, 2012, Vilamoura, Portugal. New York: IEEE, 2012: 1051-1056.

[10] Hou Y, Zhang H, Zhou S L. Convolutional neural network-based image representation for visual loop closure detection[C]//2015 IEEE International Conference on Information and Automation, August 8-10, 2015, Lijiang, China. New York: IEEE, 2015: 2238-2245.

[11] Zhou B L, Lapedriza A, Xiao J X, *et al*. Learning deep features for scene recognition using places database[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems, December 8-13, 2014, Montreal, Canada. Cambridge: MIT Press, 2014: 487-495.

[12] Sünderhauf N, Shirazi S, Dayoub F, *et al*. On the performance of ConvNet features for place recognition[C]//2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), September 28-October 2, 2015, Hamburg, Germany. New York: IEEE, 2015: 4297-4304.

[13] Gao X, Zhang T. Loop closure detection for visual SLAM systems using deep neural networks[C]//2015 34th Chinese Control Conference (CCC), July 28-30, 2015, Hangzhou, China. New York: IEEE, 2015: 5851-5856.

[14] 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.

[15] Lopez-Antequera M, Gomez-Ojeda R, Petkov N, *et al*. Appearance-invariant place recognition by discriminatively training a convolutional neural network[J]. Pattern Recognition Letters, 2017, 92: 89-95.

[16] Merrill N, Huang G Q. Lightweight unsupervised deep loop closure[J/OL]. (2018-05-24)[2018-12-29]. https://arxiv.org/abs/1805.07703.

[17] Bao Z Q, Li A H, Cui Z G, *et al*. Loop closure detection algorithm based on multi-level convolutional

neural network features[J]. Laser & Optoelectronics Progress, 2018, 55(11): 111507.
鲍振强, 李艾华, 崔智高, 等. 融合多层次卷积神经网络特征的闭环检测算法[J]. 激光与光电子学进展, 2018, 55(11): 111507.

[18] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI. New York: IEEE, 2017: 6517-6525.

[19] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift[J/OL]. (2015-03-02)[2018-12-29]. https://arxiv.org/abs/1502.03167.

[20] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks[C]//14th International Conference on Artificial Intelligence and Statistics, April 11-13, 2011, Fort Lauderdale, USA. Cambridge: PMLR, 2011, 15: 315-323.

[21] He K M, Sun J. Convolutional neural networks at constrained time cost[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 5353-5360.