

## 基于精简卷积神经网络的快速闭环检测方法

何元烈, 陈佳腾, 曾 碧

(广东工业大学 计算机学院, 广州 510006)

**摘 要:** 基于深度学习的闭环检测方法在复杂光照下能取得较好的检测效果, 但存在提取场景特征维度高、难以满足闭环检测实时性的问题。为此, 基于精简深度卷积神经网络, 提出一种闭环检测方法。结合级联修正线性单元、批规范化和深度残差模块完成网络模型的设计, 并利用大型场景识别的数据集(Places365-Standard)完成网络模型的训练, 用训练好的网络模型提取场景特征, 通过计算场景特征的相似性得到闭环区域。测试结果表明, 与基于位置卷积神经网络与自编码的闭环检测方法相比, 该方法在保证较高准确率的同时提高了检测速度。

**关键词:** 闭环检测; 视觉同步定位与地图构建; 卷积神经网络; 深度学习; 场景识别; 特征提取

**中文引用格式:** 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]. 计算机工程, 2018, 44(6): 182-187.

**英文引用格式:** HE Yuanlie, CHEN Jiateng, ZENG Bi. Fast closed loop detection method based on simplification convolutional neural network[J]. Computer Engineering, 2018, 44(6): 182-187.

## Fast Closed Loop Detection Method Based on Simplification Convolutional Neural Network

HE Yuanlie, CHEN Jiateng, ZENG Bi

(School of Computers, Guangdong University of Technology, Guangzhou 510006, China)

**[Abstract]** Loop closure detection methods based on deep learning can achieve good detection performance in complicate illumination environment. But the dimension of extracted scene feature is too high to achieve the real-time detection requirement for closed loop. For that, propose a loop closure detection method based on fast and lightweight convolutional neural network. Concatenated rectified linear unit, batch normalization and deep residual module are combined to design the fast and lightweight network model which is trained with the Places365-Standard data set. The scene feature is extracted by trained neural network model and the regions of closed loop are obtained by measuring the similarities of them. Results from the test show that compared with loop closure detection algorithms based on PlaceCNN and Autoencoder, the proposed method not only achieves the high correct rate but also increases the implementation speed for detection.

**[Key words]** closed loop detection; visual Simultaneous Localization and Mapping Construction (vSLAM); convolutional neural network; deep learning; scene recognition; feature extraction

**DOI:** 10.19678/j.issn.1000-3428.0047182

### 0 概述

视觉同步定位与地图构建 (visual Simultaneous Localization and Mapping Construction, vSLAM) 是机器人领域的重要基础研究。vSLAM 在未知环境中利用视觉信息对机器人进行定位, 并同时构建环境地图。vSLAM 算法主要由视觉里程计、后端图优化、闭环检测以及构图组成。其中闭环检测的功能是通过判断当前位置是否处于已访问过的区域来消除构图过程中的累积误差, 进而提高后端图优化的

准确性。

目前, 大部分视觉闭环检测算法将当前位置图像与先前位置图像进行匹配, 在这种情况下, 闭环检测可以看成图像匹配的问题, 即包含图像描述和相似性度量两部分。视觉词袋模型 (Bag of Visual Word, BoVW)<sup>[1]</sup> 是闭环检测中最常用的图像描述方法。文献[2]使用 FAST 算子提取图像的局部特征, 然后聚类构成视觉词典, 最后用视觉词典中的单词表征图像。文献[3]将图像的二值特征应用到增量式的闭环检测中, 对闭环检测的效率有所提

**基金项目:** 广东省科技计划重大专项项目(2016B010108004); 东莞市产学研合作项目(2015509109107)。

**作者简介:** 何元烈(1976—), 男, 副教授、博士, 主研方向为图形图像处理、计算机视觉; 陈佳腾, 硕士研究生; 曾 碧, 教授、博士。

**收稿日期:** 2017-05-12      **修回日期:** 2017-07-09      **E-mail:** heyuanlie@163.com

高。文献[4]利用混合高斯模型建立视觉单词的概率模型,并应用贝叶斯滤波提高闭环检测的成功率。为了进一步提高特征匹配的速度,文献[5]使用图像的 GIST 描述子提取图像的全局特征。BoVW 和 GIST 在闭环检测中取得了不错的效果,但是构造 BoVW 和 GIST 的图像特征都属于手工设计的特征,它们对环境中光照变化非常敏感,因此,在复杂光照下,它们检测闭环的成功率并不高。

深度学习的出现为图像描述提供了新的思路。深度学习模型可以由浅到深,逐层提取图像特征,已经在图像分类<sup>[6]</sup>、图像去噪<sup>[7]</sup>等领域取得较好的研究成果。自 2015 年以来,学者们开始尝试将深度学习的方法应用到闭环检测中。文献[8]使用自编码(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。文献[9]利用主成分分析网络(Principal Component Analysis Network, 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。文献[10]利用位置卷积神经网络(Place Convolutional Neural Network, 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,该算法虽然改善了光照变化对闭环检测的影响,但是提取特征的维度高,难以满足闭环检测实时性的要求。

为了进一步提高闭环检测的实时性和准确性,本文设计了一种快速、精简的卷积神经网络用于闭环检测。使用 Places365-Standard 场景数据集<sup>[11]</sup>训练网络模型,采用迁移学习的思想,将训练好的模型作为图像特征提取器,并利用提取到的特征完成闭环检测。

## 1 快速、精简的卷积神经网络

### 1.1 快速、精简的卷积神经网络设计

为了满足闭环检测算法实时性的要求,设计了一种快速、精简的卷积神经网络模型(Fast and Lightweight Convolutional Neural Network, FLCNN)用于闭环检测。FLCNN 由 1 个输入层 X、20 个卷积层(Conv1、Conv2、Conv3、Conv4、Conv5)、2 个池化层(Pool1、Pool2)、2 个全连接层(Fc6、Fc7)和 1 个输出层 O 组成,如图 1 所示。

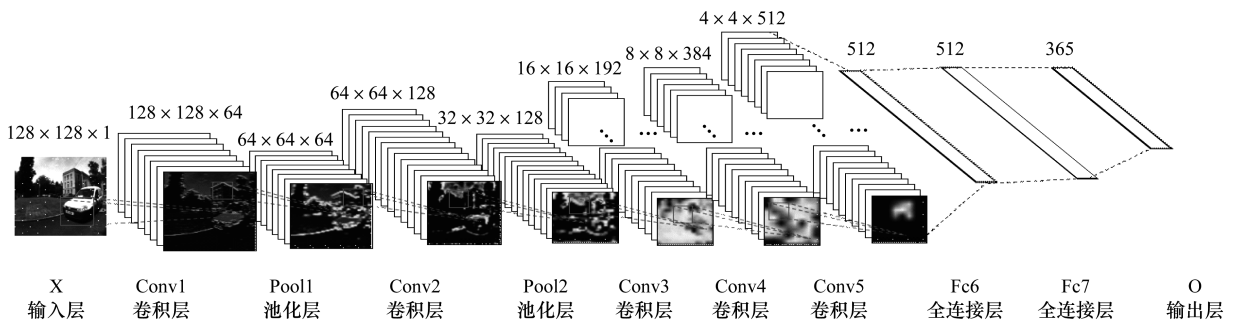


图 1 快速、精简的卷积神经网络模型示意图

FLCNN 的输入层是一幅  $128 \times 128$  的单通道图像,卷积层和全连接层的激活函数均采用修正线性单元<sup>[12]</sup>(Rectified Linear Unit, ReLU)。为了加快模型收敛,在卷积层和全连接层后均进行批规范化<sup>[13]</sup>(Batch Normalization, BN)。Conv1 和 Conv2 采用了级联修正线性单元模块(Concatenated Rectified Linear Unit, C. ReLU)<sup>[14]</sup>设计,C. ReLU 模块如图 2 所示。

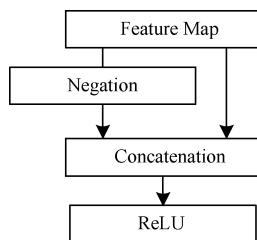


图 2 C. ReLU 模块示意图

从图 2 可以看出,C. ReLU 将卷积得到原始特征图的值取反,再把原始特征图和取反后的特征图级联,最后使用修正线性单元。因此,使用 C. ReLU 可以将前几层的卷积核的数量减半,减少模型参数。Pool1 和 Pool2 均采用最大池化(Max Pooling),分别接在 Conv1 和 Conv2 后面减少特征图的大小。Conv $x$ ( $x=3,4,5$ )由 Conv $x$ \_1 和 Conv $x$ \_2 组成,Conv $x$ \_1 和 Conv $x$ \_2 都是由 3 个卷积层组成,卷积核的大小分别为  $1 \times 1$ ,  $3 \times 3$  和  $1 \times 1$ ,其中,  $1 \times 1$  卷积核用于控制特征图的维度,减少  $3 \times 3$  卷积核的输入、输出维度,进而减少模型的参数量。Conv $x$ \_1 和 Conv $x$ \_2 均采用深度残差模块<sup>[15]</sup>以解决深层卷积神经网络难优化的问题。

Conv $x$ \_1 和 Conv $x$ \_2 的结构如图 3 所示。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 512。输出层含有 365 个神经元,表示数据集的类别数目。

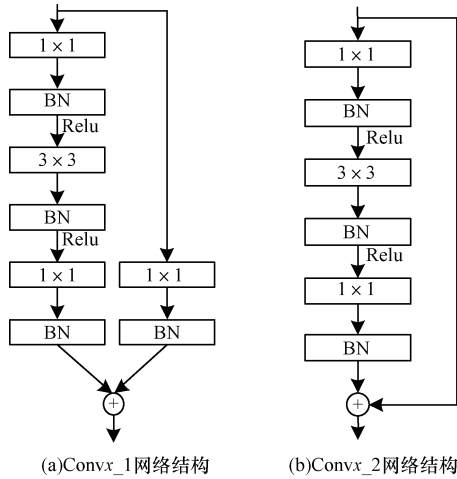


图 3 Convx 网络结构

## 1.2 快速、精简的卷积神经网络训练

为了使 FLCNN 学习到丰富的场景语义信息,便于后续的闭环检测,使用 Places365-Standard 数据集训练 FLCNN。Places365-Standard 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。

由于场景识别是一个图像多分类问题,因此采用了 Softmax 分类器对输入图像进行分类。Softmax 分类器对于输入样本  $x^{(i)}$ ,利用函数  $h(x^{(i)})$  计算该样本属于每个类别  $j$  的概率  $p(y=j|x)$ 。具体来说,对于第  $i$  个样本  $x^{(i)}$ ,函数  $h(x^{(i)})$  定义如下:

$$h(x^{(i)}) = \begin{bmatrix} p(y^{(i)}=1|x^{(i)};\theta) \\ p(y^{(i)}=2|x^{(i)};\theta) \\ \vdots \\ p(y^{(i)}=k|x^{(i)};\theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^k e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \vdots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix} \quad (1)$$

其中,  $\theta$  为网络模型的参数矩阵,  $k$  为分类数。对于输入的图像  $x^{(i)}$ ,对其进行灰度化和减均值的预处理并将其缩小为  $128 \times 128$ 。将处理后的图像输入到 FLCNN 中,最终会得到一个  $k$  维的概率向量  $P_i$ ,进而通过式(2)可预测  $x^{(i)}$  的类别  $\hat{k}$ :

$$\hat{k} = \arg \max_{1 \leq i \leq k} \{P_i\} \quad (2)$$

Softmax 分类器对应的损失函数为交叉熵损失函数,因此,将其作为网络的损失函数,损失函数为:

$$L(\theta) = -\frac{1}{m} \left[ \sum_{i=1}^m \lg \frac{e^{\theta_{y^{(i)}}^T x^{(i)}}}{\sum_{k=1}^k e^{\theta_k^T x^{(i)}}} \right] \quad (3)$$

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  $i$  个样本,  $y^{(i)}$  为第  $i$  个样本真实标签,  $k$  为分类数。在网络的训练过程中,对于每

次输入的  $m$  个样本,经网络模型的输出和样本的真实标签按式(3)计算损失值,然后通过带冲量的随机梯度下降更新网络参数。参数更新方式如下:

$$V_t = \lambda \cdot V_{t-1} - \eta \cdot \nabla L(\theta_{t-1}) \quad (4)$$

$$\theta_t = \theta_{t-1} + V_t \quad (5)$$

其中,  $t$  为迭代次数,  $\lambda$  为冲量系数,矩阵  $V_t$  为第  $t$  次迭代的参数更新值,矩阵  $\theta_t$  为第  $t$  次迭代的参数值,  $\eta$  为学习率,  $\nabla L(\theta_{t-1})$  为损失函数对参数的导数。采用以下方式初始化网络的的权值:

$$w = \text{Uniform}\left(-\sqrt{\frac{6}{m+n}}, \sqrt{\frac{6}{m+n}}\right) \quad (6)$$

其中,  $w$  为网络权值,  $m$  为输入维度,  $n$  为输出维度,  $\text{Uniform}$  表示  $w$  服从均匀分布。网络偏置初值设为 0。采用上述初始化方式,可以保证每层输出的方差近似相等,从而保持每层的数据分布不变,有利于网络中信息的流动。将冲量系数  $\lambda$  设为 0.9,基准学习率  $\eta$  设为 0.01,分别在迭代次数达到 10 万、18 万、24 万、30 万次时将学习率乘以 0.1,共迭代 32 万次。最终在 Place365-Standard 数据集上 top1 准确率为 50.16%, top5 准确率为 80.03%。

## 2 基于 FLCNN 的闭环检测

闭环检测所要解决的问题是判断当前位置是否处于已访问过的区域,这个问题和场景识别问题相似,即两者都需要对场景图像提取合适的特征。两者的不同之处在于闭环检测处理的数据是连续的视频帧,并且没有类别标签。但是 CNN 模型的训练是有监督学习的过程,如果数据没有标签信息,那么无法完成模型的训练。针对上述问题,采用在大型有标签的场景数据集上完成 FLCNN 模型的训练,将训练好的 FLCNN 模型当做场景图像的特征提取器,最后将提取到的特征应用于闭环检测。

FLCNN 模型中的全连接层用来融合学习到的图像特征,全连接层中包含了图像的大量语义信息。虽然网络层数越深,网络提取到的特征越抽象、表征能力也就越强,但迁移能力也随之下降。因此,将网络 Fc6 层的输出作为图像的特征向量。

闭环检测处理的数据具有时间连续性,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。具体地,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<sup>[16]</sup>。本文提出的基于 FLCNN 闭环检测算法流程如图 4 所示,对于采集的第  $K$  帧图像,首先对其进行预处理包括灰度化、减均值并调整其小为  $128 \times 128$ ,然后将图像输入到 FLCNN 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( $Vector_{1,2,\dots,K-L}$ )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。

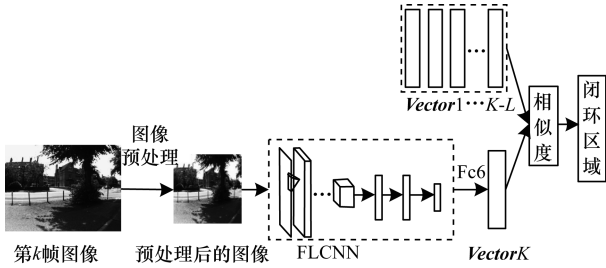


图4 闭环检测算法流程

### 3 实验结果与分析

为了验证 FLCNN 的性能,将其与基于 BoVW、GIST、PlaceCNN 和 Autoencoder 闭环检测算法进行了比较,比较的指标包括准确性和时间性能。实验共采用了 2 个闭环检测数据集 City Center<sup>[2]</sup>和 New College<sup>[2]</sup>。进行实验的计算机配置为:CPU-3.5 GHz, RAM-16 GB,显卡 GTX-1070。

#### 3.1 实验数据集

City Center 和 New College 是由牛津大学移动机器人团队收集的用于 vSLAM 的研究和闭环检测算法的评估验证。这 2 个数据集分别包含 1 237 对和 1 073 对图像,它们通过以下方式收集: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 1.5 m 采集一次图像。图 5 给出了 2 个数据集的示例图像。



图5 数据集的示例图像

由图 5 可知,同一位置两边摄像头采集到的图像是完全不同的。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。除了在整个数据集(City Center All、New College All)上进行实验,还分别在数据集的左半部分(City Center Left、New College Left)和数据集的右半部分(City Center Right、New College

Right)上进行实验。 $L$  值的设置方式与文献[16]相同,将 City Center 数据集的  $L$  值设为 800, New College 数据集的  $L$  值设为 100。

#### 3.2 准确性比较

采用准确率-召回率曲线和平均准确率验证算法的准确性。准确率(precision)和召回率(recall)是评价闭环检测算法经常使用的 2 个指标。准确率和召回率的计算公式如下:

$$P_{\text{precision}} = \frac{TP}{TP + FP} \quad (7)$$

$$R_{\text{recall}} = \frac{TP}{TP + FN} \quad (8)$$

其中, $TP$  代表检测到正确的闭环数目, $FP$  代表检测到错误的闭环数目, $FN$  代表没有检测到真实的闭环数目。利用本文算法提取数据集中图像特征,通过计算余弦距离得到相似度矩阵,并调整阈值的大小,按照式(7)、式(8)计算相应的准确率和召回率,绘制准确率-召回率曲线,与其他 4 种闭环检测算法进行比较,比较结果如图 6 和图 7 所示。

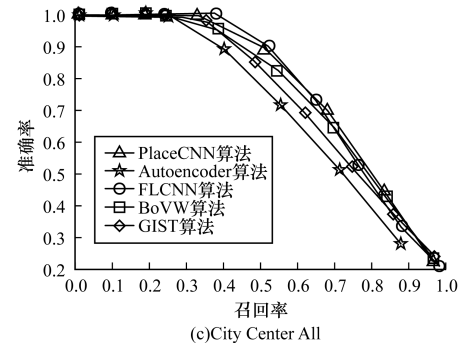
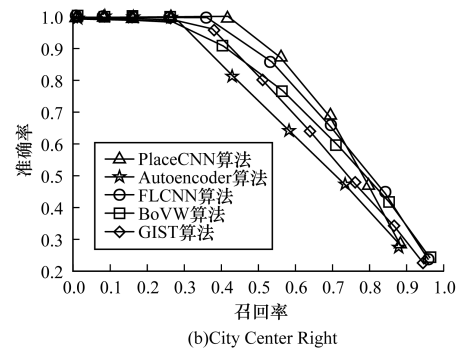
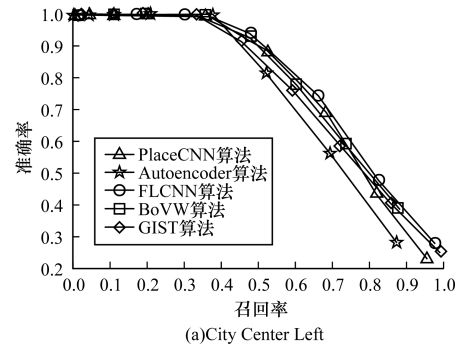


图6 City Center 数据集上的准确率-召回率曲线

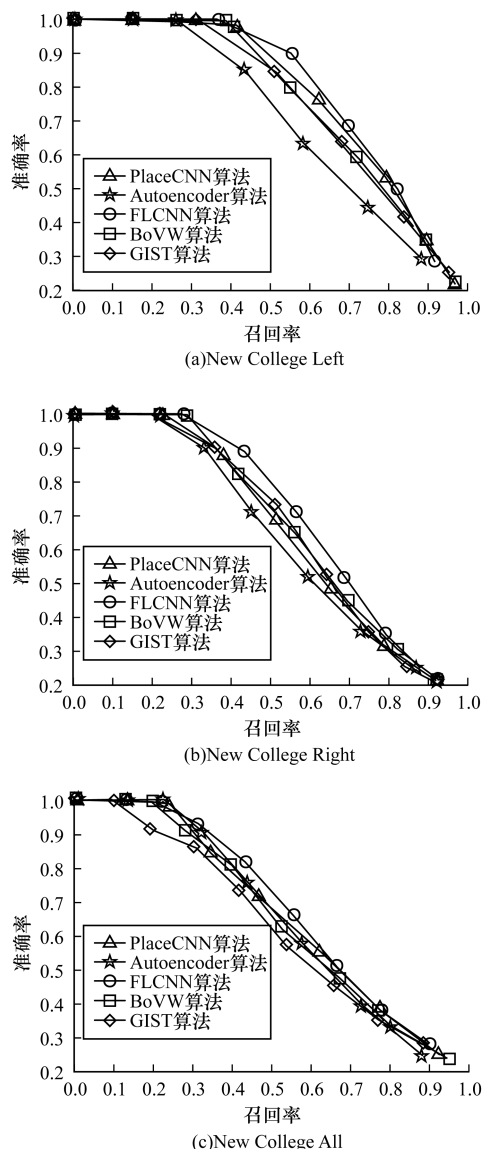


图7 New College 数据集上的准确率-召回率曲线

为了进一步验证本文算法的准确性,分别计算了不同算法的平均准确率,结果如表1所示。

表1 不同数据集上算法平均准确率 %

数据集	BoVW	GIST	Autoencoder	PlaceCNN	FLCNN
City Center All	83.92	83.60	74.81	85.98	85.21
City Center Left	86.97	86.91	82.78	87.71	88.17
City Center Right	82.60	82.39	78.27	84.35	84.11
New College All	74.81	74.16	71.64	73.61	80.12
New College Left	74.04	71.49	75.38	76.42	78.03
New College Right	78.87	78.04	76.09	81.08	83.16

从图6和图7可以看出,当召回率较小时,各种算法的准确率都是1。当召回率增大到某个值时,各种算法的准确率开始随着召回率的增加而逐渐下降。在图6(a)中,当召回率小于0.58时,FLCNN的准确率不低于其他算法;当召回率大于0.58时,

FLCNN的曲线下降的速度比其他算法快。在图6(b)中,FLCNN和PlaceCNN取得相似的准确率;在召回率大于0.81时,两者的准确率均低于BoVW和GIST。在图6(c)中,当召回率在0.41和0.74之间时,FLCNN的准确率低于PlaceCNN;但召回率大于0.74时,FLCNN的准确率高其他算法。在图7(a)和图7(b)中,当召回率相同时,FLCNN的准确率均不低于其他算法。在图7(c)中,当召回率小于0.59时,FLCNN的准确率均不低于其他算法;但当召回率大于0.59时,FLCNN的准确率低于PlaceCNN。

从表1可以看到,在New College数据集上,无论是New College All还是New College Left,New College Right,FLCNN的平均准确率均高于其他算法。这是因为FLCNN层数较深,可以充分提取图像特征。在City Center数据集上,FLCNN仅在City Center Left平均准确率高于其他算法,在City Center All和City Center Right平均准确率略低于PlaceCNN。原因在于City Center Right中出现了大量分散的物体如人和汽车,而PlaceCNN在ImageNet<sup>[17]</sup>上训练过,可以识别这些物体。

### 3.3 时间性能比较

为了验证本文方法的时间性能,比较了FLCNN与其他4种闭环检测算法的特征匹配时间。特征匹配时间指提取完图像特征后,与1000张图像比较余弦相似度所需要的时间。实验结果如表2所示。从表2可以看出,FLCNN特征匹配速度最快,原因在于FLCNN提取特征的维度最低。考虑到闭环检测对算法的实时性要求较高,提出的算法更适合使用。

表2 不同算法的时间性能

分类	特征维度	特征匹配时间/s
BoVW	1 024	18.59
GIST	800	15.84
Autoencoder	2 500	20.02
PlaceCNN	9 126	51.91
FLCNN	512	13.53

### 3.4 算法分析

使用了2个标准的闭环检测数据集验证提出的FLCNN与其他闭环检测方法。实验结果表明,提出的检测算法能够在保证较高准确率的同时提高了检测速度。在文献[18]提出了卷积神经网络的时间复杂度计算公式为:

$$O\left(\sum_{l=1}^d n_{l-1} \cdot s_l^2 \cdot n_l \cdot m_l^2\right) \quad (9)$$

其中, $l$ 为卷积层的下标, $d$ 为卷积层的数目, $n_l$ 为第 $l$ 层卷积核的数目, $s_l$ 为卷积核的尺寸, $m_l$ 为输出特征图的尺寸。虽然FLCNN的网络层数较多,相比8层的PlaceCNN,FLCNN共有22层,但是FLCNN使

用的卷积核尺寸较小,输出特征图的尺寸也较小,并且卷积层的通道数较少。因此,相比于 PlaceCNN, FLCNN 网络结构更精简,参数量更少,时间复杂度更低。按式(9)计算出 FLCNN 的时间复杂度约为 PlaceCNN 的 0.29 倍。FLCNN 训练结束后可实时地完成在线闭环检测任务,即对于采集的待检测的图片,可以立即完成图像特征的提取和闭环区域的确定。与 PlaceCNN 相比,本文算法的实用价值更高。

#### 4 结束语

如何利用深度学习方法解决视觉 SLAM 的闭环检测问题是一个新兴的研究方向,结合最新的网络设计方法,本文设计了一种快速、精简的卷积神经网络闭环检测方法。以大型场景识别数据集训练网络模型,将训练好的模型作为场景图像的特征提取器,使用所提取的特征解决闭环检测问题。实验结果表明,该方法在兼顾检测效果的同时,提高了闭环检测的速度。目前,闭环检测深度学习方法大多为提取图像的全局特征,它们对机器人运动过程中的视角变化较为敏感,因此,下一步将对此进行深入研究,以取得更好的检测效果。

#### 参考文献

- [1] LI Feifei, PERONA P. A Bayesian hierarchical model for learning natural scene categories [C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Computer Society, 2005: 524-531.
- [2] 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.
- [3] ZHANG Gangcong, MASON J, PATRICIO A. Learning binary features online from motion dynamics for incremental loop closure detection and place recognition [C]//Proceedings of International Conference on Robotics and Automation. Washington D. C., USA: IEEE Press, 2016: 765-772.
- [4] 梁志伟,陈燕燕,朱松豪,等.基于视觉词典的单目视觉闭环检测算法[J].模式识别与人工智能,2013,26(6):561-570.
- [5] LIU Yang, ZHANG Hong. Visual loop closure detection with a compact image descriptor [C]//Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. Washington D. C., USA: IEEE Press, 2012: 1051-1056.
- [6] 何希平,张琼华,刘 波.基于 HOG 的目标分类特征深度学习模型[J].计算机工程,2016,42(12):176-180.
- [7] 李传朋,秦品乐,张晋京.基于深度卷积神经网络的图像去噪研究[J].计算机工程,2017,43(3):253-260.
- [8] GAO Xiang, ZHANG Tao. Unsupervised learning to detect loops using deep neural networks for visual SLAM system [J]. Autonomous Robots, 2017, 41(1): 1-18.
- [9] XIA Yifan, LI Jie, QI Lin, et al. Loop closure detection for visual SLAM using PCANet features [C]//Proceedings of International Joint Conference on Neural Networks. Washington D. C., USA: IEEE Press, 2016: 1109-1116.
- [10] HOU Yi, ZHANG Hong. Convolutional neural network-based image representation for visual loop closure detection [C]//Proceedings of International Conference on Information and Automation. Washington D. C., USA: IEEE Press, 2015: 2238-2245.
- [11] ZHOU Bolei, KHOSLA A, LAPEDRIZA A, et al. Places: an image database for deep scene understanding [J]. Journal of Vision, 2017, 17(10): 611-622.
- [12] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]//Proceedings of International Conference on Neural Information Processing Systems. Washington D. C., USA: IEEE Press, 2012: 1097-1105.
- [13] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C]//Proceedings of International Conference on Machine Learning. New York, USA: MLR Press, 2015: 448-456.
- [14] SHANG Wenling, SOHN K, ALMEIDA D, et al. Understanding and improving convolutional neural networks via concatenated rectified linear units [C]//Proceedings of International Conference on Machine Learning. New York, USA: MLR Press, 2016: 2217-2225.
- [15] HE Kaiming, ZHANG Xiaoyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]//Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 770-778.
- [16] BAI Dongdong, WANG Chaochun, ZHANG Bo, et al. Matching-range-constrained real-time loop closure detection with CNNs features [J]. Robotics and Biomimetics, 2016, 3(1): 15-21.
- [17] JIA Deng, WEI Dong, SOCHER R, et al. ImageNet: a large-scale hierarchical image database [C]//Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2009: 248-255.
- [18] HE Kaiming, SUN Jian. Convolutional neural networks at constrained time cost [C]//Proceedings of Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2015: 5353-5360.

编辑 刘 冰