

文章编号: 2095-2163(2022)08-0196-04

中图分类号: TP391

文献标志码: A

# 基于 ResNet 模型的回环检测算法

陈 勇, 党淑雯, 聂 铃

(上海工程技术大学 航空运输学院, 上海 201620)

**摘 要:** 回环检测是 SLAM 系统中关键部分, 是减少移动机器人在运行过程中产生的累积误差的重要步骤。传统的回环检测算法使用人工建立特征, 易忽略有用信息。本文提出了基于卷积神经网络(CNN)的回环检测算法, 对预先训练的卷积神经网络模型(ResNet50、ResNet101、ResNet152)进行性能对比; 通过改进关键帧的选取策略, 将筛选出的关键帧输入预先训练好的 ResNet 模型, 输出高维特征向量; 利用主成分分析(PCA)白化来降低特征向量的维数, 并通过计算特征向量间的欧式距离; 最后计算相似矩阵, 验证回环的准确率。通过对比实验可知, 本文算法性能优于其它回环检测算法。

**关键词:** 视觉 SLAM; 回环检测; ResNet; PCA

## Loop detection algorithm based on ResNet model

CHEN Yong, DANG Shuwen, NIE Ling

(School of Air Transportation, Shanghai University of Engineering Science, Shanghai 201620, China)

**[Abstract]** This paper mainly analyzes the problems in loop detection in visual SLAM framework. Loop detection is a key part of SALM system and an important step to reduce the cumulative error of mobile robot in the process of operation. The traditional loop detection algorithm uses artificial features, which is easy to ignore the useful information. A loop detection algorithm based on Convolutional Neural Network (CNN) is proposed in this paper. Firstly, the performance of the pretrained convolutional neural network models (ResNet50, ResNet101 and ResNet152) is compared. By improving the selection strategy of key frames, the selected key frames are input into the pretrained ResNet model, the high-dimensional feature vector is output, the dimension of the feature vector is reduced by whitening with Principal Component Analysis (PCA), and the Euclidean distance between the feature vectors is calculated. Finally, the similarity matrix is calculated to verify the accuracy of the loop. Through comparative experiments, the performance of this algorithm is better than other loop detection algorithms.

**[Key words]** visual SLAM; loop detection; ResNet; PCA

## 0 引 言

随着人工智能的发展, 移动机器人已成为热门研究课题。为实现自主导航任务, 移动机器人需要执行实时定位、路径规划等操作, 在此过程中即时定位和建图(SLAM)起着至关重要的作用。按传感器不同, SLAM 可分为激光 SLAM 和视觉 SLAM。由于 RGB-D 相机的更新换代, 兼且还具有体积小、重量轻、造价低廉等诸多优点, 就使视觉 SLAM 引起了学界的广泛关注与浓厚兴趣。回环检测作为视觉 SLAM 的关键组成部分, 可以识别机器人以前经过的地方, 从而减少机器人移动过程中产生的累积误差。现有的回环检测方法可以分为基于图像特征和基于外观不变性两种。其中, 基于外观的回环检测目前获得了较为广泛应用, 该方法从本质上比较了视觉系统中 2 幅相邻图像之间的相似性。

在基于外观的方法中, 较为通用的是基于词袋模型的回环检测算法<sup>[1]</sup>。基于词袋模型的回环检测算法的主要运行步骤是: 首先利用算法(SIFT、SURF、ORB)从采集到的环境信息中提取特征向量作为“单词”, 即通过提取图像中具有局部不变的特征点作为视觉词汇; 其次, 根据 K 均值聚类算法与相近的单词结合构成词表, 并分组到一个词典中; 最后, 统计单词在采集环境信息中出现的次数, 用  $k$  维数值向量表示图像。但 BOW 模型中手工标记的特征有一定局限性, 即这些特征忽略了图像中的一些有用信息, 因此可能会产生精度较低的回环检测。

## 1 基于 ResNet 模型的回环检测算法

虽然传统的回环检测算法能在视觉 SLAM 系统中检测出回环, 但需要经过大量训练来提取特征, 且在不同场景下准确率较低。近些年, 研究人员尝试

**作者简介:** 陈 勇(1994-), 男, 硕士研究生, 主要研究方向: 图像处理; 党淑雯(1980-), 女, 博士, 副教授, 主要研究方向: 航电及飞控系统相关技术及理论研究; 聂 铃(1997-), 女, 硕士研究生, 主要研究方向: 机器学习。

**通讯作者:** 党淑雯 Email: swdang@126.com

收稿日期: 2021-11-01

将神经网络应用于计算机视觉领域,且有大量实验已经证明了在计算机视觉中神经网络有着良好表现。Lecun 等人<sup>[2]</sup>将卷积神经网络(CNN)应用于视觉中。实验证明,CNN 能有效提取特征,应用深度学习网络,能成功检测出视觉 SLAM 的回路,为解决回环检测问题提供了一种选择。Gao 等人<sup>[3]</sup>提出了一种基于良好训练的神经网络的深度特征提取方法。但该方法训练整个网络耗时长、效率低,不适用于实时的 SLAM 系统。Hou 等人<sup>[4]</sup>使用一个预先训练过的 CNN 模型,生成一个适合于视觉 SLAM 中的回环检测的图像表示。结果表明,Conv3 和 Pool5 的性能最佳。但该模型 CNN 描述符的维度非常高。Xia 等人<sup>[5]</sup>比较了数种 CNN 模型(PCANet、CaffeNet、AlexNet、GoogLeNet)和传统方法(BoW、GIST)在回环检测中的性能。结果表明,该 CNN 模型能更好地适用于回环检测。但是,由于模型中使

用支持向量机(SVM)来检测循环,不能很好地满足视觉 SLAM 系统中实时性要求。Zhang 等人<sup>[6]</sup>使用一个开源预训练的 CNN 模型,即 OverFeat 来提取特征。近年来由于需要大量的标签来训练数据,所以就很少用到该 CNN 模型。

针对上述算法存在的问题,本文提出了基于 ResNet 模型回环检测算法。总体流程如图 1 所示,算法实现步骤如下:

(1) 根据选取策略,筛选关键帧集。

(2) 通过预先训练好的 ResNet 模型,提取筛选出的关键帧集特征,生成高维特征向量。

(3) 利用 PCA 白化来降低向量的维数,提高检测效率。

(4) 通过计算特征向量间的欧式距离,再计算相似矩阵,验证回环的准确率。



图 1 算法流程图

Fig. 1 Flow chart of the algorithm

## 1.1 关键帧筛选策略

由于移动机器人在运动过程中拍摄的连续帧存在较多冗余信息,造成计算资源浪费,因此引入关键帧很有必要。选择合适的关键帧,能有效提高定位和建图的精准性和实时性。常见的关键帧筛选方法有:根据时间间隔对数据帧进行采样、根据空间距离对数据帧进行采样、根据图像相似性进行采样等。以上这些方法都是使用单一准则来筛选关键帧,无法适应不同的环境变化,因此关键帧倾向于多重融合选择。

本文通过设定相对运动量来筛选关键帧,具体步骤如下:

(1) 关键帧集合  $I$ , 第一帧、即为  $F_0$ , 将其归入集合  $I$ 。

(2) 对于新的一帧  $F_x$ , 计算  $I$  中最后一帧与  $F_x$  的运动, 并估计其运动的大小。若  $M \geq M_{\max}$ , 则表明 2 帧运动变化大, 离得较远, 应剔除; 或  $M \leq M_{\min}$ , 则表明 2 帧运动变化小, 离得太近, 应剔除。这里,  $M$  表示 2 个相邻帧的运动量,  $M_{\max}$ 、 $M_{\min}$  为设定的阈值。只有 2 帧的运动估计正确、能匹配, 且两者存在

一定距离, 则把该帧  $F_x$  加入到关键帧集合  $I$  中。对应数学模型可表示为:

$$M = \omega_1 \|\Delta t\| + \omega_2 \|\alpha, \beta, \gamma\|_2 \quad (1)$$

其中,  $\alpha, \beta, \gamma$  表示帧间的三轴方向的旋转量;  $\omega_1, \omega_2$  分别表示平移所占的权重和旋转所占的权重;  $\Delta t$  是 2 帧间的平移量。

## 1.2 ResNet 网络模型

2015 年, 何凯明团队提出 ResNet 网络<sup>[7]</sup>。该网络的发现不仅影响了学术界和工业界深度学习的发展方向, 且在图像检测、图像分割和图像识别等领域获得广泛应用。ResNet 网络结构如图 2 所示。基本设计原理是通过残差结构来解决神经网络中出现的退化和梯度消失问题, 提高了整个网络性能, 且在 ImageNet 数据集上获得了良好的分类结果。

由图 2 可知, 当输入  $x$  后, 输出为:

$$F = W_{2\sigma}(W_1 x) \quad (2)$$

其中,  $\sigma$  表示非线性函数  $relu$ 。

$x$  通过一个捷径(shortcut), 和图 2 中的第 2 个  $relu$  相结合, 最终输出:

$$y = F(x, \{W_i\}) + x \quad (3)$$

当需要对输入和输出维数进行变化时(如改变通道数目),可以在通过“捷径”时,对  $x$  加以线性变换,数学公式具体如下:

$$y = F(x, \{W_i\}) + W_s x \quad (4)$$

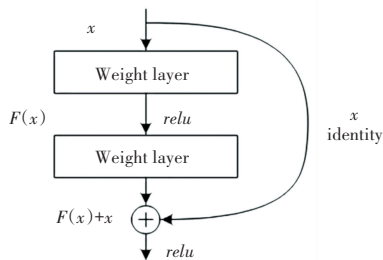


图2 ResNet 网络结构图

Fig. 2 ResNet network structure

### 1.3 PCA+白化降维算法

研究中,可以从预先训练过的网络模型中提取出整个图像特征,但却具有高维数。本文通过主成分分析(PCA)和白化算法进行降维,同时也降低了关键帧的冗余程度。

PCA 是可以提升无监督特征学习速度的数据降维算法。算法的设计原理是将  $n$  维特征投影到  $m$  维( $n > m$ )。输入样本集  $x_1, x_2, \dots, x_n$ ,使其映射到空间  $m$ ,并实施中心化,选取  $m$  个最大特征值,输出映射矩阵。中心化时需用到的数学公式为:

$$x_i - \frac{1}{n} \sum_{i=1}^n x_i \rightarrow x_i \quad (5)$$

降维后,通过白化算法对每一维除以其标准差,公式如下:

$$x_{PCAwhite,i} = \frac{x_{rot,i}}{\sqrt{\lambda_i}} \quad (6)$$

### 1.4 相似度计算

通过降维处理后的 CNN 特征,使用欧氏距离来计算图像  $i$  和  $j$  之间的差异。并绘制相似矩阵图,用于度量图像之间的相似度。欧式距离公式如下:

$$D_{i,j} = \left\| \frac{X_w^i}{\|X_w^i\|_2} - \frac{X_w^j}{\|X_w^j\|_2} \right\|_2 \quad (7)$$

计算图像  $i$  和  $j$  之间的归一化相似性:

$$S_{i,j} = 1 - \frac{D_{i,j}}{\max(D_{i,j})} \quad (8)$$

通过定义相似度矩阵,可以度量图像之间的相似度。相似度矩阵示意如图3所示。由图3可知,矩阵的每一行都包含相似度的值,其范围为(0,1)。不同的值用不同的颜色表示,值越高,图像间的相似性就越高。若值为1,则表示检测是一个循环闭合。

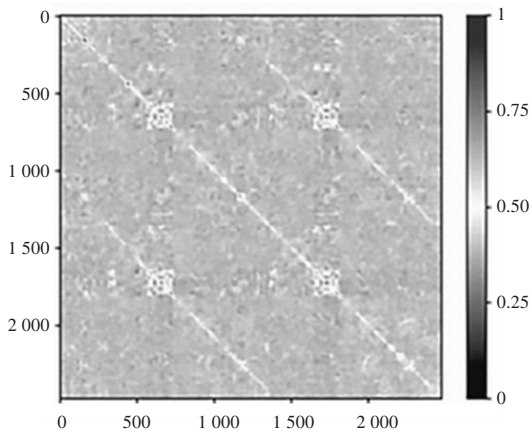


图3 相似度矩阵示意图

Fig. 3 Schematic diagram of similarity matrix

## 2 实验结果与分析

### 2.1 数据集

实验测试数据集为 City Centre 和 New College,即当机器人穿过室外城市环境时,每 1.5 m 用摄像头采集一次图像,照明条件稳定。这 2 个数据集分别是 1 237 对和 1 073 对图像,数据集的更多细节见表 1。

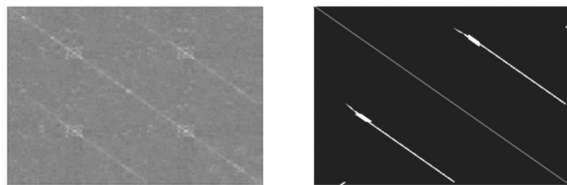
表 1 数据集参数

Tab. 1 Data set parameters

	Number	Image size	Image type	Dataset open
City Centre	2 474	640×480	RGB	Yes
New College	2 146	640×480	RGB	Yes

### 2.2 不同模型相似性矩阵对比实验

本文使用预先训练过的 ResNet50、ResNet101 和 ResNet152 模型来提取特征。根据公式(8),可以计算出图像的相似得分,并求出相似性矩阵。研究中得到的基于 ResNet 50 网络模型的回环检测算法的相似矩阵图和 Ground Truth 图即如图 4 所示。图 4 中,Ground Truth 图是相似矩阵图的“掩码”。



(a) 相似度矩阵图

(b) Ground Truth 图

图4 基于 ResNet50 网络模型的回环检测算法的相似矩阵图和 Ground Truth 图

Fig. 4 Schematic diagram of similarity matrix and Ground Truth based on ResNet50 network model

由图4中可知,较冷的颜色对图像对的相似程度较小,而较暖的颜色对图像对的相似程度更高。