

基于残差网络的图像序列闭环检测

占浩^{1,2,3*}, 朱振才^{1,2,3}, 张永合^{1,2,3}, 郭明^{1,2}, 丁国鹏^{1,2}

¹中国科学院微小卫星创新研究院, 上海 201203;

²中国科学院微小卫星重点实验室, 上海 201203;

³中国科学院大学, 北京 100049

摘要 机器人在大尺度场景下开展同时定位与建图(SLAM)任务时,其闭环检测环节会出现较严重的错匹配或者漏匹配问题,因此,采用残差网络(ResNet)对图像序列进行特征提取,并提出一种新的闭环检测算法。通过预训练的 ResNet 提取输入图像的全局特征,并对该帧图像及之前具有一定长度的图像序列的特征按照下采样的方式进行拼接,将结果作为当前帧图像的特征,保证图像特征的丰富性与准确性。同时,设计一种双层查询的方法以获得最相似的图像帧,并对最相似图像进行一致性检验,确保闭环的准确性。在闭环检测主流公开数据集 New College 和 City Centre 上,所提算法在 100% 准确率下的召回率为 83%,在 99% 准确率下的召回率为 85%。与传统的词袋方法和 VGG16 方法相比,所提算法具有显著的提升。

关键词 成像系统; 闭环检测; 残差网络; 同时定位与建图; 机器视觉

中图分类号 TP242

文献标志码 A

doi: 10.3788/LOP202158.0411003

Loop-Closure Detection Using Image Sequencing Based on ResNet

Zhan Hao^{1,2,3*}, Zhu Zhencai^{1,2,3}, Zhang Yonghe^{1,2,3}, Guo Ming^{1,2}, Ding Guopeng^{1,2}

¹Innovation Academy for Microsatellite, Chinese Academy of Sciences, Shanghai 201203, China;

²Key Laboratory of Microsatellites, Chinese Academy of Sciences, Shanghai 201203, China;

³University of Chinese Academy of Sciences, Beijing 100049, China

Abstract When robots conduct simultaneous localization and mapping (SLAM) tasks in large-scale scenes, there is serious mismatching or missed matching in loop-closure detection. Focused on this problem, this study proposes a new closed-loop detection algorithm based on a residual network (ResNet) to extract features of image sequences. The global features of an input image are extracted using a pretrained ResNet. The features of the frame image and previous image sequenced with a certain length are stitched by the down sampling method, and the results are taken as the features of the current frame image to ensure the richness and accuracy of the image features. Then, a double-layered query method is designed to obtain the most similar image frame, and the consistency of the most similar image is checked to ensure the accuracy of the loop-closure. The proposed algorithm can achieve an 83% recall rate under 100% accuracy and an 85% recall rate under 99% accuracy in the loop-closure detection mainstream public datasets of New College and City Centre, which is significantly improved compared with the traditional bag of words method and VGG16 method.

Key words imaging systems; loop-closure detection; ResNet; simultaneous localization and mapping; machine vision

OCIS codes 110.2970; 100.4996; 150.5758

收稿日期: 2020-07-15; 修回日期: 2020-08-07; 录用日期: 2020-08-13

基金项目: 中国科学院战略性先导科技专项(XDA15020305)、中国科学院国防科技重点实验室基金(CXJJ-19S012)

*E-mail: zh1226@mail.ustc.edu.cn

1 引言

复杂、非结构未知环境下的自主导航问题一直是机器人研究领域的热点,其中,实现高效的实时定位与地图构建(SLAM)技术是解决该问题的重要途径之一^[1]。近年来,随着计算机视觉技术的不断发展,相比其他感知方法,视觉同时定位与地图构建(VSLAM)凭借更为丰富的视觉信息越来越是现在SLAM研究领域的主流之一。一个完整VSLAM系统一般包含特征提取和配准、相机位姿和地图点的同时优化、曾经到达地点的检测与优化等步骤,分别对应前端跟踪、后端优化及闭环检测3个SLAM系统的典型环节,可以对周围环境进行建模,并估计自身位姿和校准跟踪运动轨迹^[2-3]。在大尺度场景下,仅仅依靠前端跟踪和后端优化的SLAM系统往往存在无法优化和忽略累积误差的问题。闭环检测技术旨在在机器人运动并构建周围环境的过程中辨识出曾经到访过的区域^[4]。准确地检测曾经访问过的地点并匹配,让机器人可以重定位,再通过相应全局优化算法,一般都可以有效消除闭环发生前的累积误差^[5-6],即使在大尺度长时间任务中也能保证SLAM系统的精度。

所以闭环检测一直都是大尺度SLAM问题中的研究重点,并且已经部署在了目前许多主流的SLAM框架中,如基于位姿图理论的ORB-SLAM^[7]、VINS-Mono^[8]等。闭环检测的场景与图像检索很相似,一般首先需要抽象并描述图像。基于图像描述符的提取方法可以分为3类:第一种,一般离线提取大量图像中的特征点及其描述子

(SURF、BRIEF等)并通过聚类的方法生成视觉字典,称之为视觉词袋,然后将单张图像的所有特征抽象为词袋中显示各个词语数量的直方图向量,但词袋模型有明显的局限性,在真实的环境下,往往有很多类似于窗户和桌椅这样的重复物体,从而有严重的错误匹配现象^[9-13];第二种,直接基于图像像素统计信息,将一张图像压缩成一个描述符(GIST等),通过该描述符进行匹配^[14-16];第三种,借助在机器视觉领域性能优越的深度卷积网络来提取图像的特征,并设计相应方法压缩输出特征维度^[17-21]。

针对上述方法在高准确率下都存在的召回率过低情况,本文提出了一种改进的残差网络(ResNet)^[22]提取图片特征,并对一定长度的图片序列的特征进行组合,采用间隔搜索和组匹配的双层匹配方法进行闭环查询。在牛津大学设计的开源闭环检测数据集New College和City Centre^[11]上验证所提算法的性能,与其他方法相比,所提算法显著提高了召回率。

2 算法结构

所提算法流程如图1所示。首先对输入图像进行预处理,改进的ResNet对预处理图像提取特征,并将该特征存储到地图容器中;再对具有一定长度序列中的每一帧图像特征按一定规则进行组合,先对当前时刻的序列特征与历史时刻序列特征进行间隔查询,得到候选帧,然后对候选帧进行当前时刻序列特征与历史时刻序列特征的组匹配,得到最高分的闭环组,并进行一致性检验,最终得到正确闭环。

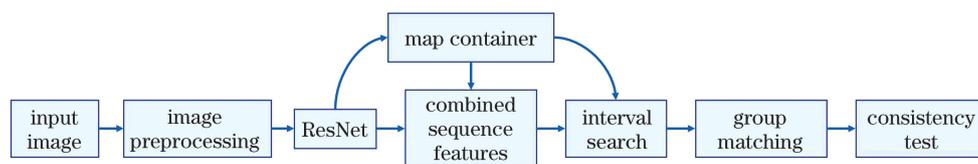


图1 所提算法流程

Fig. 1 Flowchart of the proposed algorithm

2.1 ResNet 结构

ResNet是一种针对图像识别和分类问题的深度卷积网络。ResNet与此前卷积网络最大的不同在于引入了残差机制,即每2到3层为一个单元,该单元的输出为内部卷积层前向输出和该单元输入之和:

$$\mathbf{y} = f(\mathbf{x}, \mathbf{W}_l) + \mathbf{x}, \quad (1)$$

式中: \mathbf{x} 和 \mathbf{y} 分别为网络的输入和输出; \mathbf{W}_l 为第 l 层参数。一个基础的ResNet18结构如图2所示,

即一个18层的残差网络。

像VGG^[23]这样一贯而下的超深卷积网络,当深度达到一定时,不可避免地就出现梯度消失、爆炸问题。而残差机制的引入可以保证在网络训练过程中,当较深的冗余卷积层不能带来更优秀的性能时,浅网络传入的残差可以在一定程度上抑制深网络卷积层的前向输出,避免出现通过削弱具有优秀性能浅网络的参数下降梯度来优化不必要深网络参数的

情况,有效解决了梯度消失或者爆炸的问题,提高训练效果。并且,ResNet 可以在层数相同的情况下,大大降低整个网络的参数量。一个完整的 VGG16 网络有 138357544 个参数,而采用 BasicBlock 的完整 ResNet18 网络的参数量仅有 33161024 个,这是因为 ResNet 的卷积核结构设计更小,数量也更少。网络中,更少的参数量代表更小的存储空间,具有更

快的运算效率。

在 ResNet 中,往往把多个具有相同结构的卷积核组成的卷积网络称为一个 layer,比如在图 2 中,4 个连续 $3 \times 3 \times 64$ 的卷积核组成 layer 1。本文在 ResNet18 的基础上进行改进,去除最后的 layer 4 及全连接(FC)层,只用到 3 个 layer 层,如图 3 所示,这进一步减少了网络的参数量。

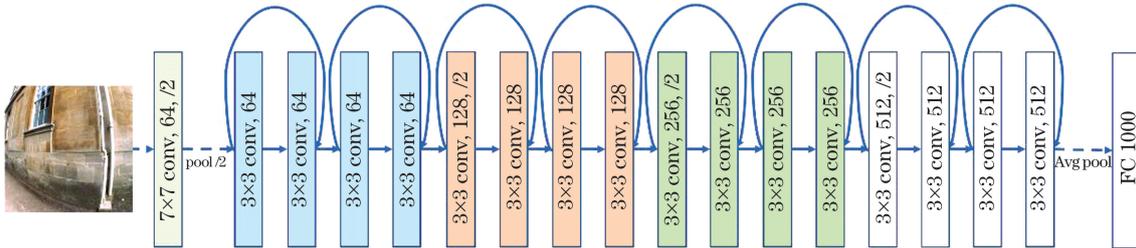


图 2 典型的 ResNet18 结构

Fig. 2 Typical ResNet18 structure

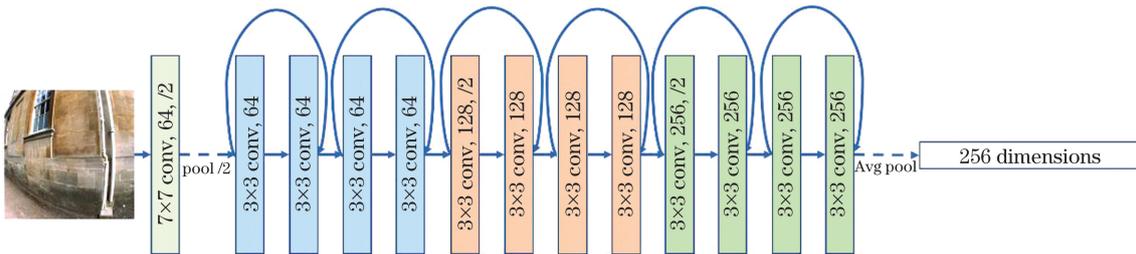


图 3 所设计的 ResNet 结构

Fig. 3 Designed ResNet structure

2.2 图像预处理和特征提取

首先需要对图像进行预处理,即对图像进行归一化。首先将图像统一为 224×224 的正方形图像,然后对图像进行去均值处理,表达式为

$$p'(I) = \frac{p(I) / 255 - m}{v}, \quad (2)$$

式中:图像 RGB 三个通道值 $p(I) = [R, G, B]$; $m = [0.485, 0.456, 0.406]$ 和 $v = [0.229, 0.224, 0.225]$ 分别为基于 ImageNet^[24] 统计出来的自然图像的 RGB 三通道的均值和方差。归一化的目的主要是使网络训练的数据集是归一化的图像。使用归一化图像训练网络能避免神经网络中激活函数梯度过小的问题,加快训练过程中的收敛。ImageNet 基于大量自然图像统计出来的均值 m 和方差 v 有着更好的泛化性能,被广泛用于各类深度网络训练场景中。

将通过 Place365 数据集^[25] 预训练的 ResNet18 网络参数部分载入到所设计的 ResNet 网络中。Place365 是麻省理工学院制作的一个场景图像数据集,用于训练基于图像的场景分类任务下的各种模型。在该数据集下,训练的 ResNet18 的全连接层

输出维度为 365,对应于 Place365 中 365 种户内、户外场景。在卷积网络中,网络的层次越深,提取的特征层次就越丰富,抽象程度越高,往往层次更高的特征对应更高的语义信息。所设计的 ResNet 网络的直接输出形式如图 4 所示,是一个 $256 \times 28 \times 28$ 的三维矩阵。为了压缩矩阵维度,采用卷积网络中的一般方法,取每个 28×28 方阵的均值作为一个特征维度,此过程即 average pool,最终网络输出一个 256 维的特征向量。

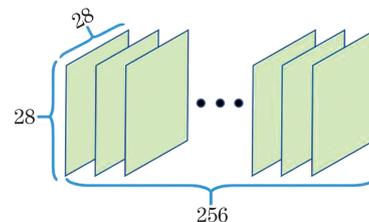


图 4 所设计网络的直接输出形式

Fig. 4 Direct output form of designed network

2.3 图像序列特征组合

当每一帧图像经 ResNet 提取特征后,就将这一帧图像的特征及其对应的图像序号传入到地图容

器中。为了提高网络在闭环图像搜索过程中的鲁棒性,对当前传入图像及其之前数帧的图像进行特征组合。记第 i 帧图像特征为 \mathbf{V}_i , 是一个 256 维的向量。组合后的第 i 帧图像特征称为序列特征, 表达式为

$$\mathbf{f}_i = [\mathbf{V}_i | \mathbf{V}_{i-1} | \cdots | \mathbf{V}_{i-j}], \quad (3)$$

式中: j 为组合序列的长度, 序号小于 j 的图像不具备序列特征。

这样的组合, 相当于将一帧图像的特征扩展为 $j+1$ 帧视频的特征。在 SLAM 问题中, 采集图像往往具有连续性, 当访问曾经到达过的场景时, 某一帧图像的相邻几帧图像往往也包含该场景信息。对于大尺度户外场景, 场景与场景间往往非常相似, 比如都会存在树木、窗户等语义上非常相似的图像, 如果仅仅靠一帧的场景信息, 往往存在误匹配。而融合了多帧信息的序列特征丰富了场景及其附近的信息, 能够极大地提高匹配准确率。

序列长度的增加会不可避免地增加序列特征的维数, 会占用大量的数据存储空间, 在进行进一

步的闭环搜索环节中, 维数的增加会线性提高计算代价。针对这样的情况, 采用下采样图像序列的组合特征:

$$\mathbf{f}_i = [\mathbf{V}_i | \mathbf{V}_{i-k_1} | \cdots | \mathbf{V}_{i-k_t} | \mathbf{V}_{i-j}], \quad (4)$$

式中: k_t 为采样的序号, t 为采样的数量。采样的标准为

$$\Delta_s = \frac{1 - \text{similarity}(\mathbf{V}_i, \mathbf{V}_{i-j+1})}{t + 1}, \quad (5)$$

$$\mathbf{V}_{i-k_t} = \arg \min [\text{similarity}(\mathbf{V}_n, \mathbf{V}_{i-j}) + t \times \Delta_s - 1],$$

$$\mathbf{V}_n \in \{\mathbf{V}_i, \cdots, \mathbf{V}_{i-j}\}, \quad (6)$$

式中: $\text{similarity}(\cdot)$ 表示特征间相似度。这样的采样要求希望采样的图像特征与当前帧图像特征的相似度尽量平均。

相似度函数是描述图片之间相关性的关键, 即两个序列特征的距离描述。不同的特征往往采用不同的距离描述, 比如文献[15]使用加权的汉明距离和欧氏距离的组合来描述两个图片的相似性, 文献[19-21]则使用余弦距离。所提算法采用 L_1 范数的倒数作为相似度函数, 表达式为

$$\text{similarity}(\mathbf{f}_i, \mathbf{f}_{i'}) = \frac{1}{\|\mathbf{f}_i - \mathbf{f}_{i'}\|} = \frac{1}{\|\mathbf{V}_i - \mathbf{V}_{i'}\| + \|\mathbf{V}_{i-j} - \mathbf{V}_{i'-j}\| + \sum_{a=1}^{a=m} \|\mathbf{V}_{i-k_a} - \mathbf{V}_{i'-k_a}\|}. \quad (7)$$

相似度函数值越大, 对应的两帧图像越相似。连续的图像里一般都存在较多的冗余信息, 基于平均相似度的采样可以减少这些冗余信息, 同时最大程度保证序列整体信息可以描述场景及其附近的一定范围, 可以压缩序列特征长度又不至于损失信息。后续的实验也证明了这一点, 下采样可以降低匹配时间, 而不会对匹配精度产生较大影响。

2.4 闭环检测方法

按照上述方法得到序列特征后, 采用归一化的相似度对输入图像和地图数据库进行匹配, 表达式为

$$s(\mathbf{f}_i, \mathbf{f}_{i'}) = \frac{\text{similarity}(\mathbf{f}_i, \mathbf{f}_{i'})}{\text{similarity}(\mathbf{f}_i, \mathbf{f}_{i-1})}, \quad (8)$$

式中: $s(\cdot)$ 表示在搜索过程中当前第 i 帧与地图数据库中第 i' 帧的评价分数。在闭环检测环节中, 相邻图像帧的特征相关性往往较大, 相似度往往很高, 一般误匹配往往都是发生在距离最近的那一帧图像, 为了避免这样的情况发生, 一般不对距离当前帧最近的一帧图像进行检测。评价分数需要大于一个阈值 α , 大于阈值的图像帧即为候选帧。查询闭环方法并非全部查询, 而是间隔一定帧数 l_s 来查询, 即查询完第 i 帧后, 下一次应当查询第 $i+l_s$ 帧。

当得到候选帧之后, 后续的闭环检测采用组匹配的方法, 当图像 I_i 与 $I_{i'}$ 是匹配帧时, 往往 I_i 与 $I_{i'+\Delta t}$ 、 $I_{i'+2\Delta t}$ 等也是匹配帧, 因此在进行检索时, 比较的是当前图像帧与地图图像帧序列的结果。记地图图像帧序列为 \mathbf{M}_{T_i} , 包含对应图像帧的特征向量 $[\mathbf{f}_{a_i}, \cdots, \mathbf{f}_{b_i}]$, 当前图像帧与某一地图图像帧序列 $\mathbf{M}_{T_{i'}}$ 的评分表达式为

$$S(\mathbf{f}_i, \mathbf{M}_{T_{i'}}) = \sum_{k=a_{i'}}^{b_{i'}} s(\mathbf{f}_i, \mathbf{f}_k). \quad (9)$$

图像的分组以候选帧为中心, 分组长度为 l_w 。分组长度应当大于闭环图像间隔帧数的 2 倍, 这样可以保证在查询图像过程中中间的图像也能被包括在分组中, 从而进一步得到确认。在候选帧中, 选择组匹配分数最大的组作为候选闭环序列。

2.5 一致性检验

一致性检验主要包括时间一致性检验和空间一致性检验。时间一致性检验要求闭环帧符合 SLAM 问题里图像帧连续的特点, 空间一致性检验要求闭环帧在空间位置上与当前图像帧相近。

时间检验一致性, 即当前帧的匹配应当与当前帧最近的 k 帧的匹配有一定程度的重叠, 即

(f_i, M_T) 对应的 M_T 应当与 $(f_{i-1}, M_{T_1}), (f_{i-2}, M_{T_2}), \dots, (f_{i-k}, M_{T_k})$ 对应的地图序列存在一定的重叠数 N_o 。有 N_o 张相同的图像。对满足时间检验一致性的候选闭环序列,计算出其中与当前图像帧之间归一化相似度最大的图像帧 I_p 。

空间一致性检验,即判断当前图像帧 I_i 与经过时间检验一致性的图像帧 I_p 的几何关系,对两张图像进行特征点提取并匹配,使用 RANSAC 算法^[26]和基础矩阵进行解算,当它们之间满足要求的特征点达到 N_d 对时,即通过空间一致性检验。

在 I_i 对应的地图容器中标注其对应闭环帧 I_p , 到这里,一个完整的闭环检测过程完成了。一个正确的闭环要满足该闭环帧与当前帧的匹配分数足够高的要求,闭环帧附近的图像帧也都与当前帧相关

性较高,并且该候选回环最终要同时通过时间一致性检验和空间一致性检验。

3 仿真验证

为了验证所提算法,在 FAB-MAP 论文中开源的 City Centre 和 New College 数据集上进行相应的测试,数据集提供了标注对应闭环的真实值矩阵,因此被广泛用来测试闭环检测方法的性能。车载相机每隔约 1.5 m 拍摄场景图像,两个数据集分别有 1237 和 1073 对由一对车载摄像头拍摄的图片, City Centre 拍摄于市中心, New College 拍摄于校园内,其中一些场景如图 5 所示。两个数据集的全球定位系统(GPS)定位数据可视化结果如图 6 所示,较细的数据线为初始运动轨迹,较粗的数据线表示存在回环的轨迹。



图 5 不同数据集的典型图像。(a)City Centre; (b)New College

Fig. 5 Typical images in different datasets. (a) City Centre; (b) New College

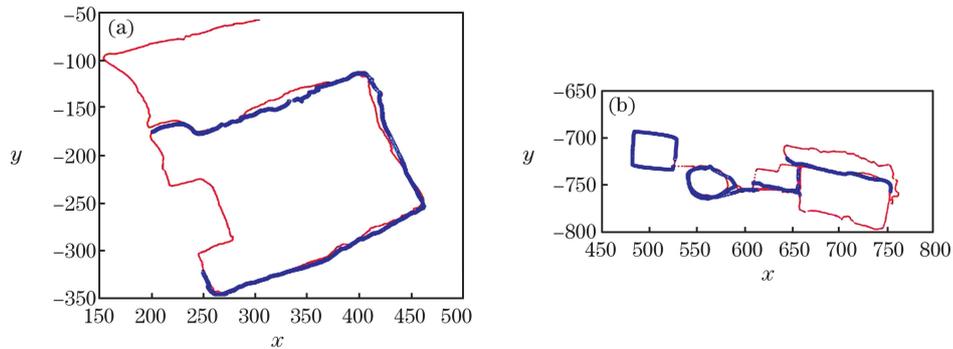


图 6 GPS 数据。(a)City Centre; (b)New College

Fig. 6 GPS data. (a) City Centre; (b) New College

表 1 为实验中用到的部分参数,主要涉及单帧图片查询、组匹配及一致性检验过程中的一些参数。

表 1 部分参数的设置

Table 1 Setting of some parameters

| Parameter | Symbol | Value |
|----------------------------------|--------|-------|
| Number of frames not detected | l | 50 |
| Number of frames in the interval | l_s | 7 |
| Group length/frame | l_w | 15 |
| Overlap length/frame | N_o | 3 |
| Number of matching points | N_d | 12 |

3.1 算法评估

闭环检测常用的指标为准确率(P)和召回率(R),设置不同的阈值 α 得到算法不同的准确率和召回率,并绘制相应的 P-R 曲线。准确率和召回率的计算公式分别为

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}}, \quad (10)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}}, \quad (11)$$

式中： N_{TP} 为正确匹配的闭环的数量； N_{FP} 为错误匹配的闭环数； N_{FN} 为漏匹配的闭环数。P-R 曲线中，曲线下方面积越大，结果越好，曲线越靠近 (1,1)，结果越好。要得到更高的准确率和更高的召回率，对于 SLAM 闭环检测的情况，应当是优先保证准确率下要求更高的召回率。

为了验证所提算法的性能，对所提算法、文献 [9] 中经典的 FAB-MAP 算法及文献 [21] 中的 VGG16+APF 算法进行比较。为了证明改进的 ResNet 性能，对其与典型 ResNet18 其他层的输出结果进行比较。之后对序列组合特征的结果与仅当

前帧特征的结果进行比较。

基于图 7 的结果，不论是场景更广阔的 City Centre 数据集还是视野更狭小的 New College 数据集，综合来看：相比 ResNet18 其他层的输出特征，改进的 ResNet 网络输出的特征的性能更为出色；同时所提算法的结果显著好于其他算法；ResNet18 不同层的输出特征经过组合(组合方法如图 8 所示)后能在一定程度上提高精度，但同样会提高时间复杂度。相比较而言，改进的 ResNet 网络的输出结果综合性能是最优越的。

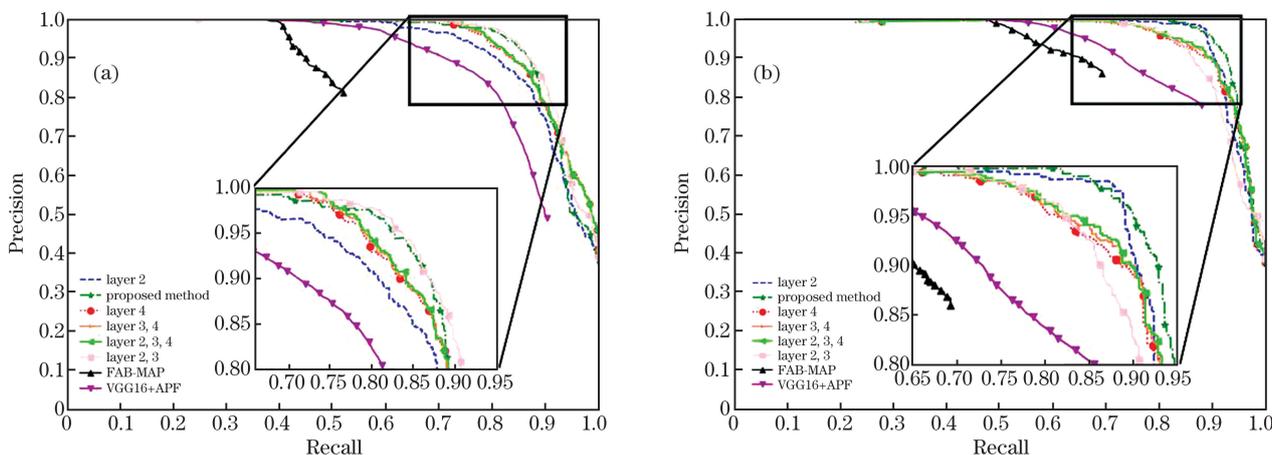


图 7 所提算法、ResNet18 不同层及其他算法的 P-R 曲线比较。(a) City Centre；(b) New College

Fig.7 Comparison of P-R curves of proposed algorithm, different layers in ResNet18, and other algorithms.

(a) City Centre；(b) New College

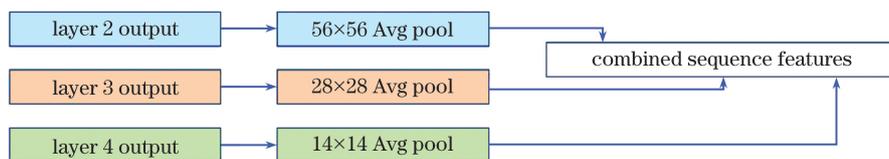


图 8 输出组合的形式

Fig. 8 Form of output combination

图 9 中， $seqX$ 表示长度为 X 的序列特征对应的结果，即 j 值。可以看出：在 City Centre 数据集上，通过组合，序列特征可以较大幅度提高结果的精度，序列长度在小于 15 时，准确率会越来越高，而大于 15 之后，准确率开始下降，但是召回率开始上升；在 New College 数据集上，所提算法在序列长度为 2 时能取得比单图像特征略好的结果，而大于 2 时准确率会逐渐降低。这主要是因为 New College 数据集中场景更为紧凑，运动的范围更小，闭环的场景更小，数据集中的闭环更为密集，如图 5(b) 所示，其转动曲率更大，较短的序列就可以完整描述该场景，而序列较长时，检测到的闭环帧与其他闭环场景有

较多重叠，会存在误匹配的情形。

图 10 是 City Centre 序列组合中下采样数量 m 对结果的影响。结果显示，下采样数量 m 合适时，不会对序列组合的结果造成影响，在一定程度上会比直接组合的序列特征结果要好。结果表明，采样后，特征长度 j 最好在序列长度的 1/5 到 1/3，此时所提算法能兼具性能与运算效率。

综合上述结果显示，长度 15 的序列、采样数量为 5 的方法可以在 City Centre 上达到 100% 准确率下 83% 召回率，在 New College 上所提算法不进行序列组合可以达到 99% 准确率下 83% 召回率；序列长度为 2 时可以达到 99% 准确率下 85% 召回率。

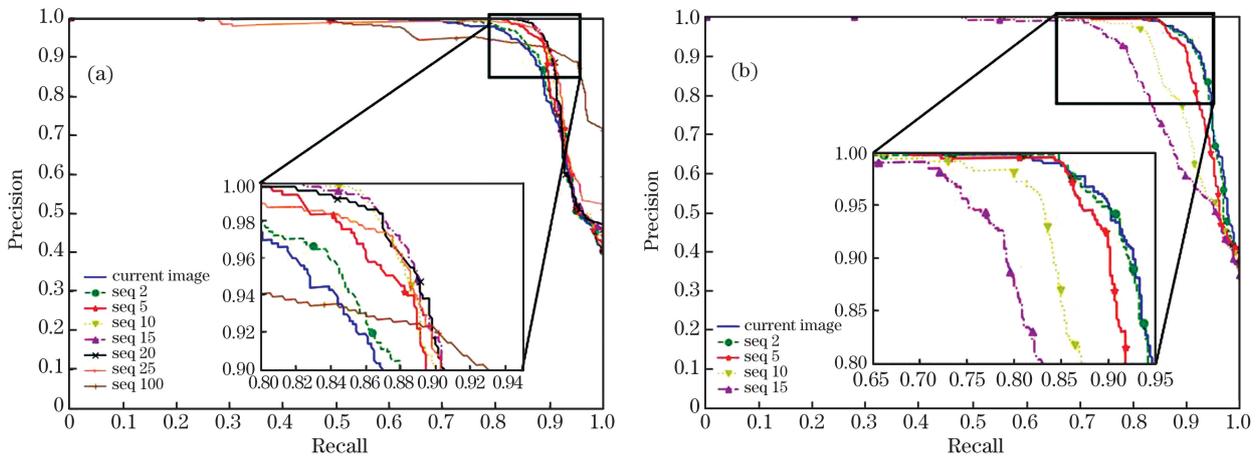


图 9 不同序列长度对结果的影响。(a) City Centre; (b) New College

Fig. 9 Influence of different sequence lengths on the results. (a) City Centre; (b) New College

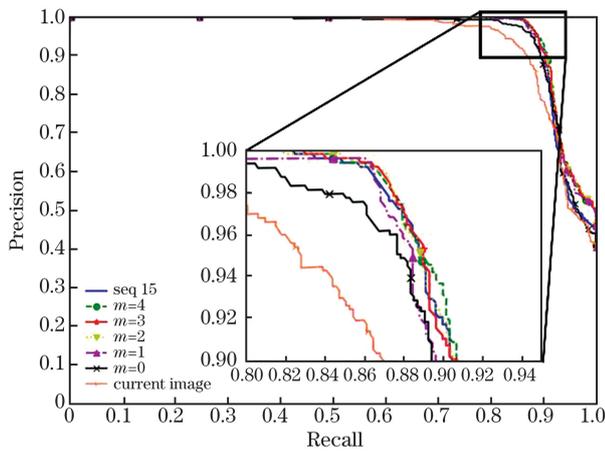


图 10 City Centre 数据集中序列长度为 15 时,不同 m 对结果的影响

Fig. 10 Influence of different m on the results when the sequence length is 15 in City Centre dataset

3.2 讨论

图 11 是第 1000 帧图像的序列组合特征和单图

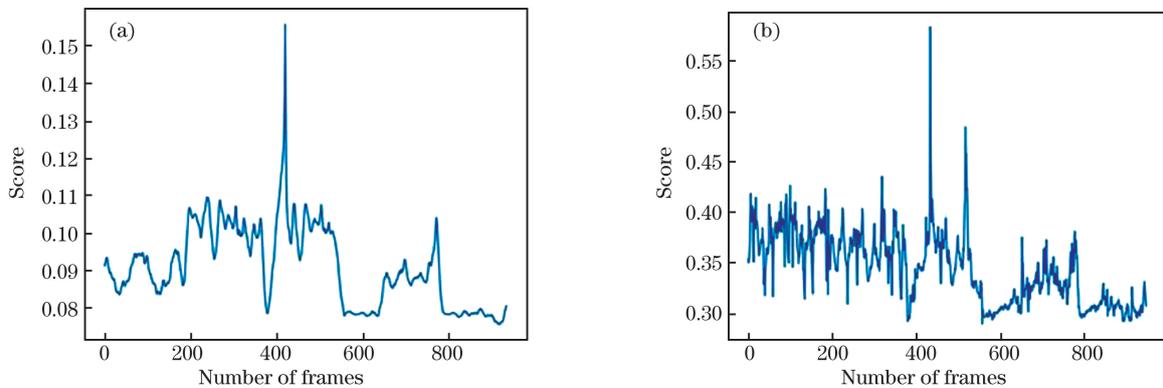


图 11 长度为 15 的序列组合前后的相似度曲线。(a)序列组合特征与其他帧的相似度; (b)不进行序列组合的相似度
Fig. 11 Similarity curves before and after combination of 15 length sequences. (a) Similarity between sequence combination features and other frames; (b) similarity of sequence without combination

像特征分别与之前图像的对应特征的相似度曲线。通过对比序列组合前后的相似度曲线可以发现,序列组合后,有效削弱了其他的相似度高峰相对于真正高峰的高度,并让曲线更加平滑。主要原因就在于通过图像序列特征的叠加,本就不相似的场景变得更不相似,即相比于那些相似的场景,不相似场景之间的 L_1 距离变化更大。

根据实验结果, j 大小的设计原则与场景有关,根据仿真经验,在开阔或者运动变化缓慢的场景下, j 应当选择在 10 到 20 之间,对于紧凑或者运动变化快速的场景,应当 j 小于 5。这是一种保守的可以让结果更好的选择。

4 结论

基于 Place365 预训练的 ResNet,设计了一种组合序列特征的闭环检测方法。将预训练模型作为场景的特征提取器,并将相应的序列图像特征组合作为当前帧的特征,使用间隔搜索和组匹配的方法

进行闭环检索,并进行一致性检验,最终完成闭环检测。对不同场景下的不同算法的性能进行了测试和评估,并比较了 ResNet18 其他层及其组合对实验结果的影响,重点比较了不同的序列长度和不同的采样数量对结果的影响。实验结果证明,改进的 ResNet 提取特征的性能要显著好于传统手工方法提取的特征和 VGG16 提取的特征。组合序列特征的方法更适用于像 City Centre 数据集这样场景更广阔、变化更缓慢、视场更大的场景,可以较大程度地提高闭环匹配的准确率。但是对于 New College 数据集这样紧凑的场景,可能并不需要组合序列特征也能取得不错的性能,这归功于 ResNet 提取图像特征的能力和闭环检测环节中的多重保险,但是较短长度的序列特征组合也能在一定程度上提高闭环检测的准确率,所以可以认为组合序列特征的方法对提高闭环检测性能有较为显著的帮助。同样,针对直接组合序列特征带来的特征向量维度过高等问题,下采样实验的结果表明,合适的下采样不会影响闭环检测的准确率,甚至可以略微优化算法结果。

参 考 文 献

- [1] Boal J, Sánchez-Miralles Á, Arranz Á. Topological simultaneous localization and mapping: a survey[J]. *Robotica*, 2014, 32(5): 803-821.
- [2] Konolige K, Agrawal M. FrameSLAM: from bundle adjustment to real-time visual mapping [J]. *IEEE Transactions on Robotics*, 2008, 24(5): 1066-1077.
- [3] Lin F C, Liu Y H, Zhou J F, et al. Optimization of visual odometry algorithm based on ORB feature[J]. *Laser & Optoelectronics Progress*, 2019, 56(21): 211507.
林付春, 刘宇红, 周进凡, 等. 基于 ORB 特征的视觉里程计算法优化[J]. *激光与光电子学进展*, 2019, 56(21): 211507.
- [4] Ho K L, Newman P. Detecting loop closure with scene sequences [J]. *International Journal of Computer Vision*, 2007, 74(3): 261-286.
- [5] Williams B, Klein G, Reid I. Automatic relocalization and loop closing for real-time monocular SLAM[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(9): 1699-1712.
- [6] Lu S D, Tu M Y, Luo X Y, et al. Laser SLAM pose optimization algorithm based on graph optimization theory and GNSS [J]. *Laser & Optoelectronics Progress*, 2020, 57(8): 081024.
陆世东, 涂美义, 罗小勇, 等. 基于图优化理论和 GNSS 激光 SLAM 位姿优化算法[J]. *激光与光电子学进展*, 2020, 57(8): 081024.
- [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] Qin T, Li P L, Shen S J. VINS-mono: a robust and versatile monocular visual-inertial state estimator[J]. *IEEE Transactions on Robotics*, 2018, 34(4): 1004-1020.
- [9] 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.
- [10] 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*. 2006, 3951: 404-417.
- [11] 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.
- [12] Salas-Moreno R F, Newcombe R A, Strasdat H, et al. SLAM + +: simultaneous localisation and mapping at the level of objects [C]//2013 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2013, Portland, OR, USA. New York: IEEE Press, 2013: 1352-1359.
- [13] Calonder M, Lepetit V, Strecha C, et al. BRIEF: binary robust independent elementary features [M]//Daniilidis K, Maragos P, Paragios N. *Computer vision-ECCV 2010. Lecture notes in computer science*. Berlin, Heidelberg: Springer, 2010, 6314: 778-792.
- [14] Liu G Z, Hu Z Z. Fast loop closure detection based on holistic features from SURF and ORB[J]. 2017, 39(1): 36-45.
刘国忠, 胡钊政. 基于 SURF 和 ORB 全局特征的快速闭环检测[J]. *机器人*, 2017, 39(1): 36-45.
- [15] 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 Press, 2012: 1051-1056.
- [16] Friedman A. Framing pictures: the role of knowledge in automatized encoding and memory for gist [J]. *Journal of Experimental Psychology. General*, 1979, 108(3): 316-355.
- [17] 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.
- [18] Qiu C L, Huang D Z, Liu H W, et al. Loop closure detection algorithm based on convolutional

- autoencoder fused with GIST feature [J]. *Laser & Optoelectronics Progress*, 2019, 56(18): 181501.
- 邱晨力, 黄东振, 刘华巍, 等. 融合 Gist 特征与卷积自编码的闭环检测算法 [J]. *激光与光电子学进展*, 2019, 56(18): 181501.
- [19] 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.
- [20] Zhang X D, Gu Z Q, Qin X F. VGG16 model-based fast loop closure detection algorithm [J]. *Optical Instruments*, 2019, 41(3): 20-26.
- 张学典, 顾璋琦, 秦晓飞. 基于 VGG16 模型的快速闭环检测算法 [J]. *光学仪器*, 2019, 41(3): 20-26.
- [21] 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 Press, 2015: 2238-2245.
- [22] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [EB/OL]. (2015-12-10) [2020-07-14]. <https://arxiv.org/abs/1512.03385>.
- [23] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2015-04-10) [2020-07-14]. <https://arxiv.org/abs/1409.1556>.
- [24] Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database [C]//2009 IEEE Conference on Computer Vision and Pattern Recognition, June 20-25, 2009, Miami, FL, USA. New York: IEEE Press, 2009: 248-255.
- [25] Zhou B L, Lapedriza A, Khosla A, et al. Places: a 10 million image database for scene recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(6): 1452-1464.
- [26] Fischler M A, Bolles R C. rando sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography [J]. *Communications of the ACM*, 1981, 24(6): 381-395.