

融合卷积神经网络与 VLAD 的闭环检测方法

李昂¹ 阮晓钢^{1,2} 黄静^{1,2*} 朱晓庆^{1,2}

¹(北京工业大学信息学部 北京 100124)

²(计算智能与智能系统北京市重点实验室 北京 100124)

摘要 传统的闭环检测方法大多采用人工设计的特征,很容易受到环境的影响。卷积神经网络通过提取层次化特征,更好地应对了光照变化,但忽略了图像的局部空间特性。针对该问题,提出一种融合 VGG16 与 VGG-NetVLAD 的闭环检测算法。该网络保留了 VGG16 的部分结构,并在最后一层引入了基于局部聚合描述符向量(VLAD)思想的池化层 NetVLAD,使提取的特征更适用于闭环检测。实验表明,相较于传统的视觉词袋模型及其他几种深度学习方法,该算法具有更强的泛化性,可以在闭环检测中达到更高的准确率并满足实时性的要求。

关键词 闭环检测 卷积神经网络 特征提取 词袋模型

中图分类号 TP3

文献标志码 A

DOI:10.3969/j.issn.1000-386x.2021.01.023

LOOP CLOSURE DETECTION ALGORITHM BASED ON CONVOLUTIONAL NEURAL NETWORK AND VLAD

Li Ang¹ Ruan Xiaogang^{1,2} Huang Jing^{1,2*} Zhu Xiaoqing^{1,2}

¹(Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China)

²(Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing 100124, China)

Abstract The traditional methods for loop closure detection are vulnerable to environmental influence, because they always adopt features which are low-level and designed artificially. Convolutional neural network can better deal with light changes by extracting hierarchical features, but also ignores the local spatial characteristics of images. In view of this, we proposed a loop closure detection algorithm which combined VGG16 with VLAD (VGG-NetVLAD). This network retained part of the structure of VGG16, and added the pooling layer NetVLAD based on the idea of VLAD to the last layer, making the extracted features more suitable for loop closure detection. The experiment results show that our model can achieve better precision compared with bag-of-words and other deep learning methods, and also has stronger generalization and meet the requirement of real-time performance.

Keywords Loop closure detection Convolutional neural network Feature extraction Bag-of-words

0 引言

目前,闭环检测已经成为了移动机器人导航领域的关键问题和研究热点,是视觉同时定位与地图构建(SLAM)的重要环节。视觉 SLAM 主要由视觉里程计、后端图优化、闭环检测和建图四个部分组成^[1]。其中

闭环检测也称为位置识别,指机器人在导航过程中使用视觉传感器提供的图像来识别是否经过了先前到达的位置^[2]。假设在当前时刻和较早时刻相机捕获了两幅图像,那么闭环检测的任务就是根据这两幅图像的相似性来判断这两个时刻的位置是否相同。正确的闭环检测可以为后端优化的姿态图中增加边缘约束,进一步优化机器人的运动估计,消除视觉里程计产生的

累计位姿误差,而错误的闭环检测会导致地图绘制失败。因此,一个好的闭环检测算法对于构建全局一致性地图乃至整个 SLAM 系统都至关重要。

近年来,学者们在闭环检测方面做了大量的研究。经典的闭环检测算法大致可以分为两类:词袋模型(BoW)算法^[2]和全局描述子算法。BoW 从图像中提取局部特征,然后聚类形成不同的单词,最终用单词向量的形式来描述图像,将闭环检测转化成了图像单词向量的相似性度量问题。词袋模型算法的关键在于如何选择最优的局部特征,目前多采用传统的 SIFT^[3]、SURF^[4]和 ORB^[5]等特征作为图像的表达。而全局描述子算法的主要思想是直接计算整幅图像的描述子,从而表达图像的整体属性。GIST 是一种非常有效的常规图像描述子,它能够以紧凑的方式去捕捉不同类型场景的基本结构。在此基础上,Liu 等^[6]使用 GIST 提取图像的全局特征,并采用 PCA 对描述子进行降维,提高了闭环检测的计算效率和图像描述子的识别能力。但以上两种方法各有优缺点。Furgale 等^[7]证明了 BoW 相较于全局描述子能更好地应对相机视角变化,而 Milfold 等^[8]和 Naseer 等^[9]提出全局描述子方法在光照变化的情况下更鲁棒。McManus 等^[10]将这两种方法结合起来,提出了一种使用场景标签的无监督系统,该系统可以为不同的视觉元素产生广域检测器,提高了闭环检测的准确性。然而,这些方法所使用的都是低层特征,是人为设计的,它们对光照、天气等因素的影响很敏感,缺乏必要的鲁棒性。

随着大规模数据集的公开(如 Imagenet)以及各种硬件的升级(如 GPU),深度学习^[11-12]得到了迅速发展。深度学习能够通过多层神经网络对输入的图像提取抽象的高层特征,从而更好地应对环境的变化^[13],这种优势使其在图像分类和图像检索中得到了广泛应用。考虑到闭环检测问题与图像检索十分相似,学者们开始尝试将深度学习应用于闭环检测。Gao 等^[14]利用自编码器提取图像特征并采用相似性度量矩阵对闭环进行检测,但忽略了图像的空间特性。何元烈等^[15]设计了一种快速、精简的卷积神经网络(FLCNN),在保证闭环检测准确率的情况下提高了算法的计算效率。Xia 等^[16]利用 PCANet 对图像进行特征提取,证明了该网络提取的特征优于传统的手工设计特征,但限制了输入图像的尺寸。Hou 等^[17]利用 PlaceCNN 进行闭环检测,该方法即使在光照变化时也具有较高的检测精度,但算法运算时间较长。

为了进一步提高闭环检测算法的准确率和鲁棒

性,本文采用了融合 VGG16 与 VLAD 的网络结构 VGG-NetVLAD。该网络采用了基于 VLAD^[18]思想的 NetVLAD 池化层,充分考虑了图像的局部空间特性,以提高闭环检测的准确性和鲁棒性。不同于典型的 CNN,VGG-NetVLAD 训练数据的标签仅为各自的地理位置,能够进行弱监督学习。

1 融合 VGG16 与 VLAD 的网络模型

1.1 传统 VLAD 的编码方法

局部聚合描述符向量(VLAD)是一种常用于实例检索和图像分类的描述子池化方法,可以捕获局部特征在图像中聚合的统计信息。词袋模型保存了图像中不同种类单词的数量信息,而 VLAD 储存的是每个单词与对应聚类中心的残差和。

假设 VLAD 算法的输入为单幅图像的 N 个 D 维特征向量 \mathbf{x}_i ,参数为 K 个聚类中心 \mathbf{c}_k ,则 VLAD 会输出一个 $K \times D$ 维的图像特征向量,将其写成矩阵的形式,并记作 \mathbf{V} ,计算公式如下:

$$\mathbf{V}(j,k) = \sum_{i=1}^N a_k(\mathbf{x}_i)(\mathbf{x}_i(j) - \mathbf{c}_k(j)) \quad (1)$$

式中: $\mathbf{x}_i(j)$ 和 $\mathbf{c}_k(j)$ 分别代表第 i 个特征向量和第 k 个聚类中心的第 j 个元素; $a_k(\mathbf{x}_i)$ 表示第 i 个特征向量对应第 k 个聚类中心的权重,当该特征属于这个聚类中心时,权重为1,否则为0。由此看出,矩阵 \mathbf{V} 的第 k 个 D 维列向量代表所有特征向量对应第 k 个聚类中心的残差和 $(\mathbf{x}_i - \mathbf{c}_k)$ 。因此,VLAD 方法提取的图像特征表达了聚类中心周围的特征分布,包含了一定的语义信息。

1.2 可训练的 VLAD 编码模型 NetVLAD

在 VLAD 中权重 $a_k(\mathbf{x}_i)$ 只有两种取值,并且特征向量在不同聚类中心的权重之和为 1,是一个离散函数,所以无法通过反向传播进行训练。为了将 VLAD 的思想引入到 CNN,本文采用了 Arandjelovic 等^[19]提出的 NetVLAD 池化层。它采用了一种近似的方式,将 $a_k(\mathbf{x}_i)$ 软分配到多个聚类中心,使其可微:

$$\bar{a}_k(\mathbf{x}_i) = \frac{e^{-\alpha \|\mathbf{x}_i - \mathbf{c}_k\|^2}}{\sum_{k'} e^{-\alpha \|\mathbf{x}_i - \mathbf{c}_{k'}\|^2}} \quad (2)$$

$\bar{a}_k(\mathbf{x}_i)$ 取值范围在 0 到 1 之间,距离 \mathbf{x}_i 最近的聚类中心权重最高,距离 \mathbf{x}_i 最远的聚类中心权重最低。 α 是一个正常数,控制响应随距离大小的衰减。当 $\alpha \rightarrow \infty$ 时,式(2)就与原 VLAD 的方法相同。

将式(2)中的平方项展开,分子分母同时约掉 $e^{-\alpha \|x_i\|^2}$,得到如下的软分配(soft-assignment)形式:

$$\bar{a}_k(x_i) = \frac{e^{w_k^T x_i + b_k}}{\sum_{k'} e^{w_{k'}^T x_i + b_{k'}}} \quad (3)$$

式中: $w_{k'} = 2\alpha c_{k'}$; $b_k = -\alpha \|c_k\|^2$,将式(3)代入式(1)最终得到 NetVLAD 层输出的特征向量:

$$V(j, k) = \frac{e^{w_k^T x_i + b_k}}{\sum_{k'} e^{w_{k'}^T x_i + b_{k'}}} (x_i(j) - c_k(j)) \quad (4)$$

NetVLAD 层通过将 $\bar{a}_k(x_i)$ 软分配到不同聚类中心的方式,有效地聚合了不同特征空间的一阶残差 $(x_i - c_k)$ 的加权和。该层具有三个参数 w_k 、 b_k 、 c_k ,可以通过端到端学习的方式进行优化。NetVLAD 与原始 VLAD 只有一个参数相比,具有更好的灵活性,如图 1 所示。

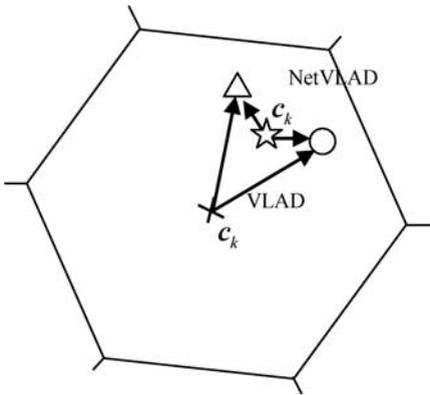


图 1 NetVLAD 相对于 VLAD 的自适应性

图 1 中圆圈和三角形分别代表两个不同图像的特征描述符,被分配到同一个集群,聚类中心为 c_k 。当采用余弦距离计算图像相似度时,在 VLAD 编码下,由于 VLAD 的特征向量进行了 L2 归一化,所以相似性得分等价于两幅图像残差向量 $(x_i - c_k)$ 的点积,聚类中心 c_k 位于“x”处是为了均匀分布所有的残差。而在 NetVLAD 编码下,如果两幅图像并不相似,它会学习到一个更优的聚类中心“☆”,使得残差向量间的点积更小,从而得到更加准确的相似度。

1.3 融合 NetVLAD 层的卷积神经网络设计

为了使卷积神经网络提取到的图像特征更适用于闭环检测,本文对 VGG16 网络进行了裁剪,去掉了最后一个卷积层(conv5_3)之后的池化层和全连接层,包括 ReLU 激活函数,并将 NetVLAD 层连接到卷积层 conv5_3 之后,作为新的池化层。将图像输入到 VGG-NetVLAD 网络中,提取 Layer_5 层的特征作为池化层 NetVLAD 的输入。网络结构如图 2 所示。

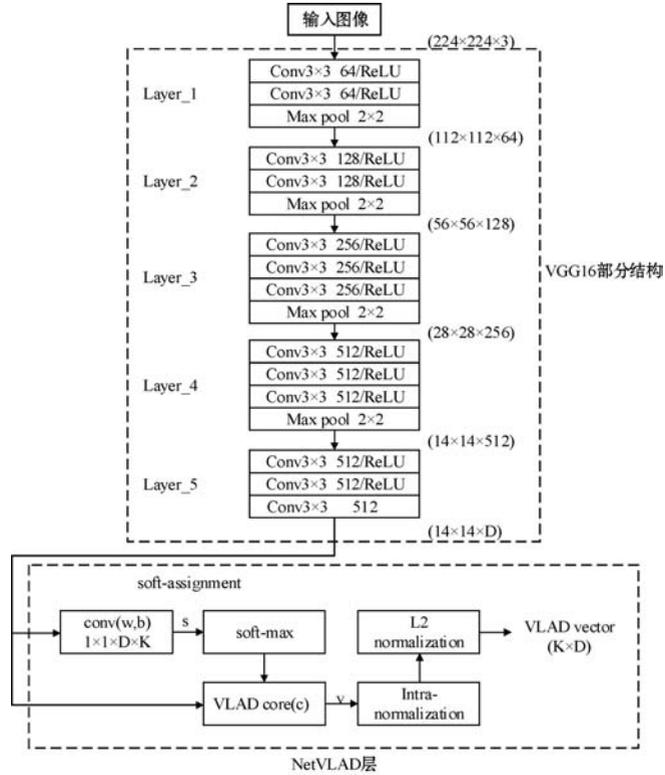


图 2 融合 VGG16 与 VLAD 的网络模型

NetVLAD 层可以分解成几个基本的 CNN 层,通过连接形成一个有向无环图。其中 soft-assignment 过程可以分为两个步骤:1) 将最后一个卷积层输出的特征 x_i 通过一个含有 K 个 1×1 卷积核的卷积层,得到输出: $s_k(x_i) = w_k^T x_i + b_k$; 2) 将 $s_k(x_i)$ 通过 soft-max 函数得到 $\bar{a}_k(x_i)$ 。当得到矩阵 V 后,需将 V 中每一列 D 维向量进行 L2 归一化,并将矩阵 V 转换成向量,再对向量整体进行一次 L2 归一化,最后输出的特征向量维度为 $K \times D$ 。

1.4 VGG-NetVLAD 的训练

为了获得 VGG-NetVLAD 的最优参数,便于进行闭环检测,使用 Pittsburgh (Pitts250k) 数据集^[19]进行训练。它包含了 250 000 幅从 Google 街景下载的数据库图像和 24 000 幅与数据库拍摄时间不同(间隔为几年)的测试图像,可用于地点识别。将 Pitts250k 平均分为三部分用于训练、验证和测试,每部分包含 83 000 幅数据库图像和 8 000 幅查询图像。通过对拍摄于不同季节但地点相同的图像进行描述,能够使算法学习到更优的特征表达,以准确地识别图像所处的位置。

不同于图像分类数据集,Google 街景时光机器拍摄的图片为全景图,每幅全景图由一组不同方向的透视图组成,每幅透视图只有代表其在地图上大致位置的 GPS 标签。这种弱监督信息会导致两个地理位置很近的查询图像不一定描述了相同的场景,因为它们

可能处于不同的拍摄方向,所以针对弱监督训练的问题,本文采用了一种弱监督排序的损失函数^[20],其目标为:给定一幅查询图像 q ,要在数据集所有图像 I_i 中找到与 q 位置距离最近的图像 I_{i^*} 。对每一幅图像 q 进行查询时,数据集会根据 GPS 信息将与其距离相近(10 米以内)的图像作为正样本集合 $\{p_i^q\}$,距离很远(超过 25 米)的图像作为负样本集合 $\{n_j^q\}$,构建一个新的三元组^[21]数据集 $(q, \{p_i^q\}, \{n_j^q\})$ 。在这个三元组中,正样本 $\{p_i^q\}$ 中至少包含一幅能与查询图像匹配的图像,但事先并不清楚具体是哪一幅。为了解决这种模糊性,采用如下形式得到最佳的匹配图像 $p_{i^*}^q$:

$$p_{i^*}^q = \operatorname{argmin}_{p_i^q} d_\theta(q, p_i^q) \quad (5)$$

对于每一个训练三元组,目标都是要学习一种最优的图像表示方法 f_θ ,使得查询图像 q 与最佳匹配图像 $p_{i^*}^q$ 的距离小于查询图像 q 与任何一个负样本图像的距离:

$$\forall j \quad d_\theta(q, p_{i^*}^q) < d_\theta(q, n_j^q) \quad (6)$$

在三元组的基础上,定义弱监督排序损失函数的形式为:

$$L_\theta = \sum_j l(\min_i d_\theta^2(q, p_{i^*}^q) + m - d_\theta^2(q, n_j^q)) \quad (7)$$

式中: l 为 hinge loss 函数, $l(x) = \max(x, 0)$; m 为一个附加常数; L_θ 代表了所有负样本图像的损失之和,对于每一个负样本图像,当其与查询图像的距离大于查询与最佳匹配图像的距离与 m 之和,则损失为 0,否则其损失值与 m 成正比。

VGG-NetVLAD 在反向传播过程中,使用随机梯度下降法对参数进行优化,设置每个训练批次的样本数量为 4 个三元组, K 值设为 64,初始学习速率设为 0.000 1, m 设为 0.1,冲量系数设为 0.9,每迭代 5 次学习速率减半,共迭代 30 次。用训练好的 VGG-NetVLAD 模型在 Pitts250k 数据集上进行实验,结果显示:当准确率为 100% 时,该算法可以达到 85% 的召回率。

2 基于 VGG-NetVLAD 的闭环检测算法

闭环检测的任务是识别当前时刻位置是否曾经到达过,算法流程可分为图像特征提取与相似性度量两个步骤,具体流程如图 3 所示。本文将在 Pittsburgh 数据集上训练好的 VGG-NetVLAD 模型作为图像的特征提取器,将提取到的特征用于闭环检测。

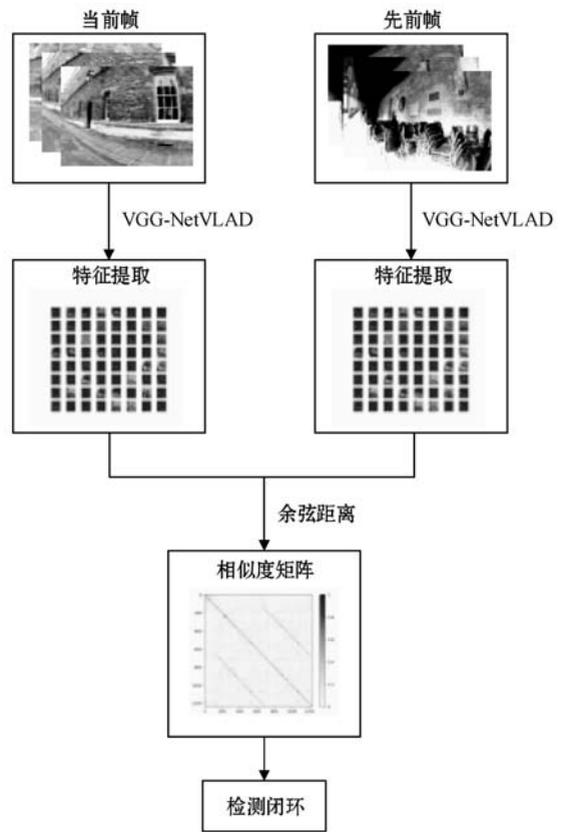


图 3 闭环检测算法流程

2.1 图像特征提取

图像的原始尺寸为 640×480 ,为了减小算法的计算复杂度,将图像大小缩小为 224×224 ,并进行减均值处理后输入到网络中。从图 4 所示的特征图可以看出,VGG-NetVLD 中池化层 NetVLAD 输出的特征图更加关注静态的建筑物,而不是汽车这种有利于图像分类的物体,因此更适合应用于闭环检测。因此,将 NetVLAD 层提取的特征作为图像的最终表述,每个图像的特征向量的维度为 4 096。

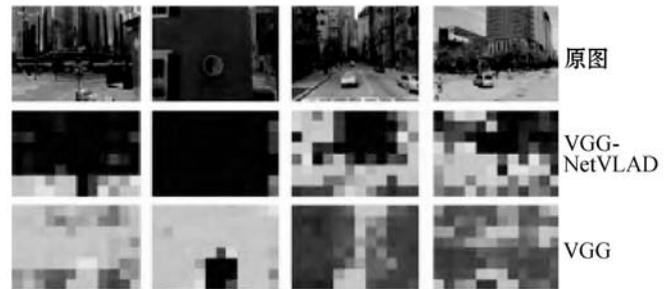


图 4 两种算法生成的特征图

2.2 相似性度量

假设当前图像为 I_n ,之前某一时刻图像为 I_p ,分别输入到 VGG-NetVLAD 模型中得到特征向量 v_n 和 v_p 。为了计算两图像的相似程度,采用向量间的余弦距离作为度量标准,余弦相似度计算公式如下:

$$S_{n,p} = \frac{\sum_{i=1}^{K \times D} \mathbf{v}_n \mathbf{v}_p}{\sqrt{\sum_{i=1}^{K \times D} (\mathbf{v}_n)^2} \sqrt{\sum_{i=1}^{K \times D} (\mathbf{v}_p)^2}} \quad (8)$$

式中: S 为图像 I_n 与 I_p 的相似性得分,也是特征向量 \mathbf{v}_n 和 \mathbf{v}_p 夹角的余弦值。在闭环检测时,对数据中所有图像两两分别进行相似性比较,最终得到一个相似性矩阵,矩阵中第 i 行第 j 列的值代表图像 I_i 与图像 I_j 的相似性得分。设定相似性阈值,当图像间的相似性得分超过阈值时,认为两图像为同一地点,产生了闭环,否则就判定不是闭环。

2.3 时间约束

在进行闭环检测时,输入的图像是由相机在连续时间内拍摄的,因此相邻图像的内容重复性很高,容易被检测成同一位置,但实际的运动轨迹并未形成闭环。为了避免产生大量的错误结果,需考虑图像相似度对比的时间阈值,对于当前图像,不考虑与其拍摄的间隔时间小于阈值的图像进行比较。由于难以获取每个图像具体的拍摄时间,所以采取设定图像比较范围的方式,假设当前检测的图像为 I_t ,在时间阈值内与 I_t 相邻的图像数量为 d ,则图像 I_t 的相似性比较范围为第 1 帧至第 $t-d$ 帧。

3 实验

为了验证基于 VGG-NetVLAD 的闭环检测算法的性能,与其他闭环检测算法进行对比,其中包括经典的 BoW 方法(在训练词典时提取 ORB 特征),以及三种基于深度学习的方法:自编码器(Autoencoder)、Inception 和 VGG16^[22],均通过 ImageNet 数据集进行训练。评价算法的性能指标为准确性和计算时间。实验采用的环境为:Ubuntu16.04, Tensorflow。计算机配置为:内存 32 GB,一块 GTX1080 显卡,处理器为 Intel Xeon E5-2603 V3。

3.1 实验数据集

闭环检测实验在 NewCollege 和 City Center 两个数据集上进行,它们是由牛津大学发布,应用最广泛的闭环检测验证数据集。NewCollege 数据集主要为自然物体的图像序列,如公园里的花草树木,而 CityCenter 数据集则主要为城市环境的图像序列,如在街道上行走的人和车。这两个数据集的共同点是在同一个地点存在左右两个不同视角的图像,其获取方式为:在移动机器人的云台上放置左右两个相机,当机器人每移动 1.5 m 时采集一次图像。表 1 为两个数据集的详细信

息,图 5 为数据集的部分示例图。

表 1 数据集详细信息

数据集	总长度	图片	尺寸	闭环个数
NewCollege	1.9 km	2 146	640 × 480	14 832
CityCenter	2.0 km	2 474	640 × 480	26 976

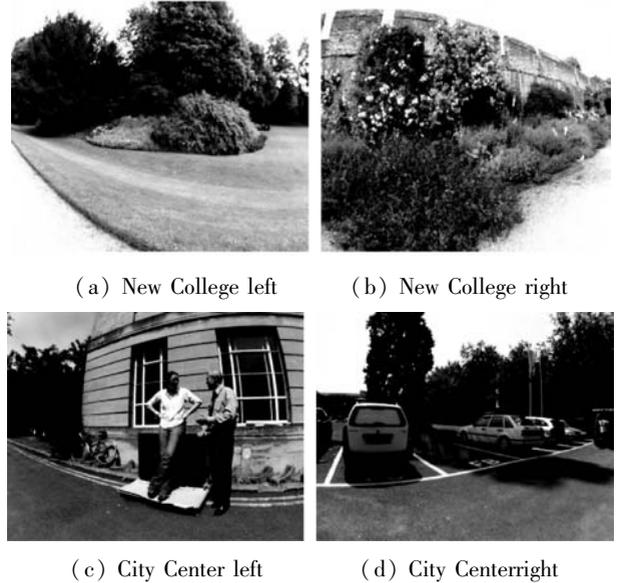


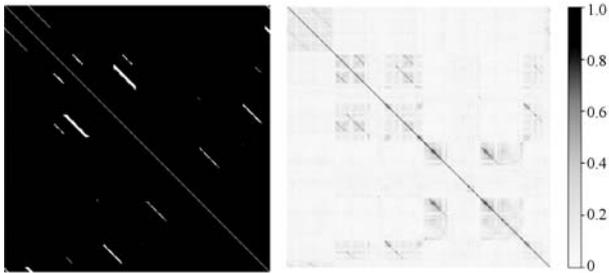
图 5 数据集部分示例图

数据集没有将左右两个相机拍摄的图片进行划分,而是分别以奇偶序号命名。两个数据集都提供了图像对是否形成闭环的真实标注,分别储存为 $2\ 146 \times 2\ 146$ 、 $2\ 474 \times 2\ 474$ 的二维矩阵形式。当图像 i 和图像 j 为同一地点时(形成闭环区域),矩阵中 (i,j) 处的数值为 1,否则为 0。由图 5 可以看出,左右两个相机在同一地点拍摄的两幅图像完全不同,而真实的标注数据将很多同一地点拍摄但场景完全不同的图像对 (i,j) 也归类为闭环区域,这种情况下仅仅通过比较图像间的相似性很难获得与真实标注数据一样的结果。为了解决上述问题,首先对两个数据集各自进行归类,将数据集中左相机采集的图像和右相机采集的图像分为两部分,对标注数据也采取相同的操作,然后再分别进行 4 组闭环检测实验。参照文献[23]中 d 值的设置方式,将分类后 New College 两个数据集的 d 值设为 50, City Center 两个数据集的 d 值设为 400。

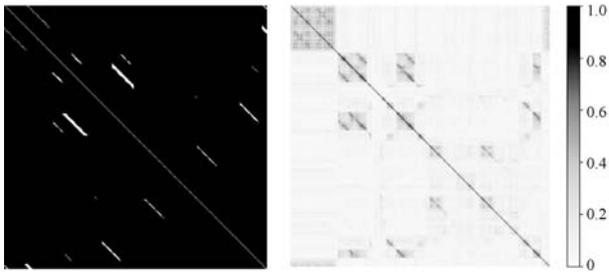
3.2 相似性矩阵

在四组实验中,分别按照 3.1 节和 3.2 节的方法提取输入图像的特征向量,以余弦距离作为度量标准计算每对图像间的相似性得分,最终生成整个数据集的相似度矩阵。图 6 对四组实验的相似度矩阵进行了可视化,左侧为真实标注数据的热力图,右侧为基于 VGG-NetVLAD 方法生成的热力图。图中颜色越亮的

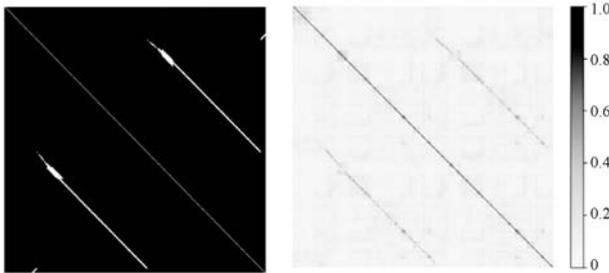
地方代表图像的相似度越高,颜色越暗的地方图像相似度越低,其中对角线的值为图像对*(i, i)*的相似性得分,因此对角线的颜色最亮。可以看出,基于 VGG-NetVLAD 的闭环检测方法可以检测出大部分的真实闭环区域,并且与非闭环区域有一定的区分度。



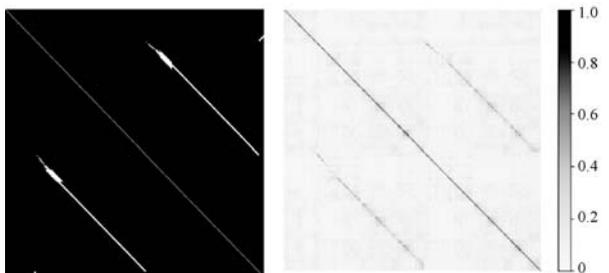
(a) New College left



(b) New College right



(c) City Center left



(d) City Center right

图 6 相似性矩阵的可视化

3.3 准确性比较

为了验证算法的准确性,将算法的预测值与真实值进行比较,分为 4 种结果,如表 2 所示。

表 2 闭环检测的结果分类

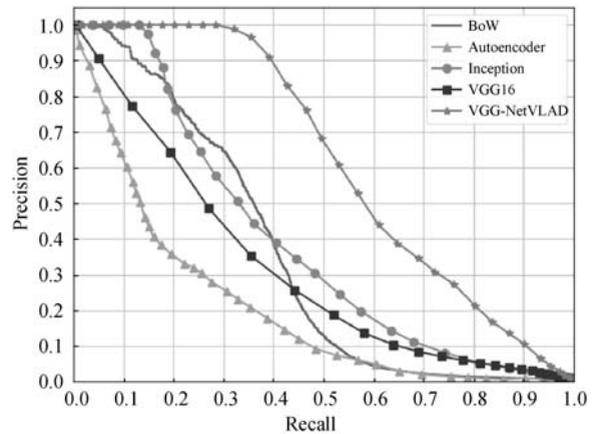
结果分类	是闭环	不是闭环
判断是闭环	True Positive(TP)	False Positive(FP)
判断不是闭环	False Negative(FN)	True Negative(TN)

准确性指标主要为准确率 (precision) 和召回率 (recall),准确率代表算法预测出的所有闭环中正确闭环所占的比例;而召回率则代表算法预测出的正确闭环占有真实闭环的比例。分别统计表 2 中 TP、FP、FN 的预测个数,根据式(9) - 式(10)可以计算准确率 *P* 和召回率 *R*。通过不断改变阈值大小,获得不同的预测结果,最终绘制准确率-召回率曲线。

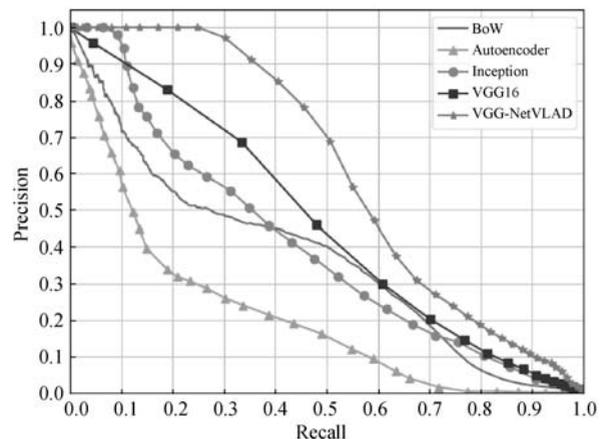
$$P = \frac{TP}{TP + FP} \tag{9}$$

$$R = \frac{TP}{TP + FN} \tag{10}$$

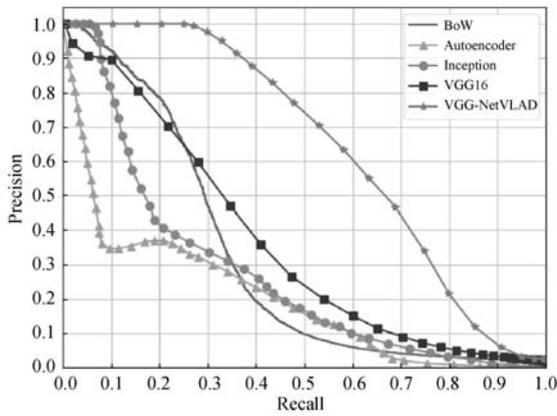
实验的准确率-召回率比较结果如图 7 所示。当召回率接近 0 时,五种算法的准确率都等于 1。随着阈值的减小,算法能够检测到更多的闭环数量,因此召回率不断上升,但同时算法的准确率随之下降。从图 7(a)、图 7(c)和图 7(d)可以看出,当召回率分别小于 0.43、0.24 和 0.17 时,VGG16(fc6 层的输出)的准确率均低于传统的 BoW,而 VGG-NetVLAD 要优于其他四种算法。在图 7 中,当准确率为 100% 时,VGG-NetVLAD 的召回率依次达到 0.284、0.247、0.258 和 0.274,在保证高准确率的情况下提高了召回率,有效地增加了检测出真实闭环的数量。



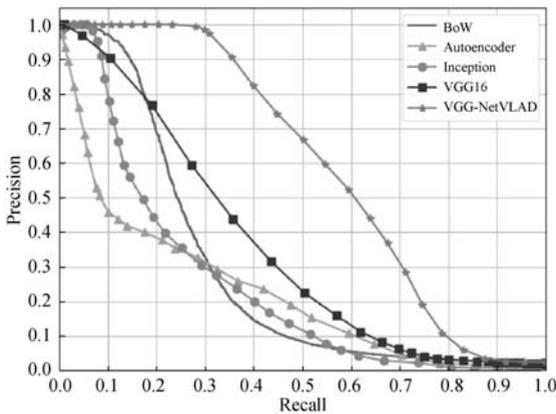
(a) New College left



(b) New College right



(c) City Center left



(d) City Center right

图 7 准确率-召回率曲线

另一个衡量算法准确性的指标为平均准确率:

$$P_{ap} = \sum_{k=1}^N P(k) \Delta r(k) \quad (11)$$

式中: N 代表数据集中图像的总个数; $P(k)$ 表示检测到 K 个闭环区域时的准确率; $\Delta r(k)$ 表示当调整阈值时,检测到的闭环区域从 K 个增加到 $K + 1$ 时召回率的变化值。不同闭环检测算法在四组实验下的平均准确率如表 3 所示,结果显示无论是在 New College left、New College right 数据集还是 City Center left、City Center right 数据集进行实验,VGG-NetVLAD 的平均准确率均高于 BoW、VGG16 和其他两种深度学习方法。由于 VGG-NetVLAD 使用的训练数据集为 Pitts250k,其获取的场景中人、汽车和道路出现的次数更多,所以算法在 CityCenter 数据集下的平均准确率相对更高。

表 3 不同数据集上算法的平均准确率 %

数据集	BoW	Auto-encoder	Inception	VGG16	本文方法
New College left	49.84	33.46	50.35	46.69	73.79
New College right	46.73	32.14	47.87	56.67	73.95
City Centerleft	43.29	35.58	41.29	51.01	78.95
City Centerright	40.77	36.22	42.37	50.56	76.18

3.4 时间性能比较

除了对算法的准确性进行比较,还通过计算不同算法的特征提取时间来衡量算法的实时性。实验在 New College 数据集上计算提取单幅图像特征的平均时间,基于深度学习算法的提取过程均在 GPU 下实现。从表 4 的时间性能比较可以看出,本文模型在特征提取时间上要远小于传统的 BoW,但由于 VGG-NetVLAD 需要对特征进行降维,略高于 VGG16,但大体上满足了闭环检测的实时性的需求。

表 4 时间性能比较

算法	特征维度	特征提取时间/s
BoW	1 024	0.465
Autoencoder	2 500	0.024
Inception	1 024	0.085
VGG16	4 096	0.019
本文方法	4 096	0.042

4 结 语

本文提出了一种融合 VGG16 与 NetVLAD 池化层的闭环检测方法。NetVLAD 层将 VLAD 的思想引入到了卷积神经网络,其参数可以通过端到端的学习获得。本文使用 Pittsburgh 数据集作为网络的训练集,并采用弱监督排序损失函数训练网络,将训练好的网络作为图像的特征提取器,在 New College 和 City Center 数据集上进行闭环检测实验。实验结果表明,本文的闭环检测算法与传统的 BoW 以及其他三种深度学习方法相比,在保证高召回率的情况下提升了准确率,同时基本满足了闭环检测的实时性。下一步研究会着重考虑机器人在移动过程中视角变化、光照变化对闭环检测的影响,对算法进行改进以达到更好的鲁棒性。

参 考 文 献

[1] Fuentes-Pacheco J, Ruiz-Ascencio J, Rendon-Mancha J M. Visual simultaneous localization and mapping: A survey[J]. ArtificialIntelligence Review,2015,43(1):55-81.

[2] Shekhar R, Jawahar C V. Word image retrieval usingbag of visual words[C]//IEEE 10th IAPR InternationalWorkshop on Document Analysis Systems (DAS). IEEE,2012:297-301.

[3] Ng P C, Henikoff S. SIFT: Predicting amino acid changesthat affect protein function[J]. Nucleic Acids Research,

- 2003,31(13):3812-3814.
- [4] Bay H, Ess A, Tuytelaars T, et al. Speeded-up robust features (SURF) [J]. *Computer Vision and Image Understanding*, 2008, 110(3): 346-359.
- [5] Mur-Artal R, Tardos J D. ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras [J]. *IEEE Transactions on Robotics*, 2017, 33(5): 1255-1262.
- [6] Liu Y, Zhang H. Visual loop closure detection with a compact image descriptor [C] // 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2012: 1051-1056.
- [7] Furgale P, Barfoot T D. Visual teach and repeat for long-range rover autonomy [J]. *Journal of Field Robotics*, 2010, 27: 534-560.
- [8] Milford M J, Wyeth G F. SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights [C] // 2012 IEEE International Conference on Robotics and Automation. IEEE, 2012: 1643-1649.
- [9] Naseer T, Spinello L, Burgard W, et al. Robust visual robot localization across seasons using network flows [C] // Proceedings of the 28th AAAI Conference on Artificial Intelligence. ACM, 2014: 2564-2570.
- [10] McManus C, Upcroft B, Newmann P. Scene signatures: Localised and point-less features for localisation [C] // Proceedings of Robotics: Science and Systems, 2014.
- [11] 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.
- [12] 刘强, 段富海, 桑勇, 等. 复杂环境下视觉 SLAM 闭环检测方法综述 [J]. *机器人*, 2019, 41(1): 112-123, 136.
- [13] Chatfield K, Simonyan K, Vedaldi A, et al. Return of the devil in the details: Delving deep into convolutional nets [EB]. arXiv:1405.3531, 2014.
- [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] 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.
- [16] Xia Y F, Li J, Qi L, et al. Loop closure detection for visual SLAM using PCANet features [C] // 2016 International Joint Conference on Neural Networks (IJCNN). IEEE, 2016: 2274-2281.
- [17] 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. IEEE, 2015: 2238-2245.
- [18] Huang Y, Sun F C, Guo Y. VLAD-based loop closure detection for monocular SLAM [C] // 2016 IEEE International Conference on Information and Automation (ICIA). IEEE, 2016: 511-516.
- [19] Arandjelovic R, Gronat P, Torii A, et al. NetVLAD: CNN architecture for weakly supervised place recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 40(6): 1437-1451.
- [20] Loquercio A, Dymczyk M, Zeisl B, et al. Efficient descriptor learning for large scale localization [C] // 2017 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2017: 3170-3177.
- [21] 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.
- [22] 张云洲, 胡航, 秦操, 等. 基于栈式卷积自编码的视觉 SLAM 闭环检测 [J]. *控制与决策*, 2019, 34(5): 981-988.
- [23] Bai D D, Wang C C, Zhang B, et al. Matching-range-constrained real-time loop closure detection with CNNs features [J]. *Robotics and Biomimetics*, 2016, 3(1): 15-21.
- ~~~~~
- (上接第 92 页)**
- [14] 巫威眺, 沈旅欧, 靳文舟. 基于速度截断分布和流量的车队离散模型 [J]. *西南交通大学学报*, 2014, 49(1): 126-133.
- [15] 左洪福, 蔡景, 吴昊. *航空维修工程学* [M]. 北京: 科学出版社, 2011: 313-315.
- ~~~~~
- (上接第 115 页)**
- [5] Golfarelli A, Zagnoni M, Proli P, et al. Acquisition system for pressure sensor network [C] // Proceedings of 2004 IEEE SENSORS. IEEE, 2004: 579-582.
- [6] 常莹, 马炳和, 邓进军, 等. 基于微型压力传感器阵列的翼面压力分布直接测量系统 [J]. *实验流体力学*, 2008, 22(3): 89-93.
- [7] 胡为东. CAN FD 总线介绍及力科的测试解决方案 [J]. *中国集成电路*, 2015, 24(8): 50-53.
- [8] 杨辉. 基于 CAN 总线的智能传感器网络系统的研制 [D]. 杭州: 浙江大学, 2012.
- [9] 刘永涛, 徐大诚, 郭述文. 一种硅压阻式压力传感器温度补偿算法及软件实现 [J]. *现代电子技术*, 2013(12): 18-20.
- [10] 李智, 张景宇, 牛军浩, 等. 基于 IEEE1588 的时钟同步技术研究 [J]. *仪表技术与传感器*, 2012(12): 99-102.
- [11] 曹禹, 邹磊, 时维铎, 等. 基于 IEEE1588 时间同步系统的硬件实现 [J]. *电测与仪表*, 2013, 50(3): 107-112.