

李国民,张鹏超,朱代先. 基于双特征和 HNSW 网络的视觉 SLAM 闭环检测算法[J]. 西安科技大学学报, 2026, 46(2): 256-266.

LI Guomin, ZHANG Pengchao, ZHU Daixian. Loop closure detection algorithm of visual SLAM based on dual features and HNSW network[J]. Journal of Xi'an University of Science and Technology, 2026, 46(2): 256-266.

基于双特征和 HNSW 网络的视觉 SLAM 闭环检测算法

李国民, 张鹏超, 朱代先

(西安科技大学 通信与信息工程学院, 陕西 西安 710054)

摘要: 针对当前闭环检测算法采用单一特征表征图像时存在表征不足, 闭环搜索时存在耗时过长, 进而出现闭环检测准确率低和实时性差的问题, 提出一种基于 ResNet18 卷积特征和传统 ORB 人工特征并结合层次化可导航小世界 HNSW 网络的闭环检测算法。首先对检测图像提取 ResNet18 卷积特征和 ORB 人工特征, 通过构建的视觉词典利用 VLAD 算法对提取特征进行编码, 对 2 种 VLAD 编码利用最优融合权重系数进行融合; 其次利用层次化可导航小世界 HNSW 网络构建 VLAD 索引结构, 对融合 VLAD 编码进行加速检索, 并在闭环检测公开数据集上进行验证。结果表明: 相对于单一 ResNet18 卷积特征的 VLAD 编码, 基于双特征融合 VLAD 编码的算法在 NewCollege 和 CityCentre 数据集上的闭环检测准确率分别提高了 29% 和 20%; 相比于原始算法的暴力搜索, 采用 HNSW 网络的 VLAD 索引结构搜索算法在 NewCollege 数据集上的闭环搜索总时间下降了 51%, 同时闭环检测准确率高达 85%。基于双特征和 HNSW 网络的闭环检测算法具有闭环检测准确率高和搜索时间短, 为高精度的实时闭环检测系统提供了一种新的设计方案。

关键词: 视觉 SLAM; 闭环检测; 视觉词典; VLAD 融合; HNSW 网络

中图分类号: TP 391

文献标志码: A

文章编号: 1672-9315(2026)02-0256-11

DOI: 10.13800/j.cnki.xakjdxxb.2026.0203

Loop closure detection algorithm of visual SLAM based on dual features and HNSW network

LI Guomin, ZHANG Pengchao, ZHU Daixian

(College of Communication and Information Engineering, Xi'an University of Science and Technology, Xi'an 710054, China)

Abstract: When the current loop closure detection algorithm uses a single feature to characterize the image, the effective information of the image is insufficiently characterized, and there is an excessively long time-consuming in the image loop closure search, resulting in low loop closure detection accuracy and poor realtime performance. Therefore, this paper proposes a loop closure detection algorithm based on ResNet18 convolution features and traditional ORB artificial features combined with hierarchical navigable small world networks. Firstly, the ResNet18 convolution features and ORB artificial features were extracted from the detected image. The extracted features were encoded by the constructed visual

收稿日期: 2025-09-15

基金项目: 国家自然科学基金项目(51774235)

通信作者: 李国民, 男, 湖南宁乡人, 教授, E-mail: liguomin@163.com

dictionary using the VLAD algorithm, the two VLAD codes were fused by the optimal fusion weight coefficient. Secondly, the VLAD index structure constructed by the hierarchical navigable small world HNSW network was used to accelerate the retrieval of the fused VLAD coding. The results show that: Compared with the VLAD coding based on single ResNet18 convolution features, the algorithm based on dual features fusion VLAD coding improves the accuracy of loop closure detection on NewCollege and CityCentre datasets by 29% and 20% respectively. Compared with the global traversal search of the original algorithm, the total loop closure search time on the NewCollege dataset using the VLAD index structure search algorithm of the HNSW network is reduced by 51%, and the accuracy of loop closure detection was as high as 85%. The loop closure detection algorithm based on dual features and HNSW network has high loop closure detection accuracy and short search time, which provides a new design scheme for high precision realtime loop closure detection system.

Key words: visual SLAM; loop closure detection; visual dictionary; VLAD fusion; HNSW network

0 引言

随着人工智能技术的快速发展,视觉同时定位与地图构建技术(Visual Simultaneous Location And Mapping, VSLAM)已成为移动机器人领域的研究热点^[1]。基于 VSLAM 技术的机器人在移动过程中,由于传感器测量误差的逐渐累积,会造成构建地图精度下降,最终使得构建的地图和实际存在偏差。闭环检测作为 VSLAM 中的重要环节^[2],主要作用是使机器人能够正确识别出之前经过的地方,当由于传感器误差而出现累计漂移^[3]时,利用闭环检测可以对累计漂移进行纠正^[4-5],从而构建全局一致的轨迹和地图。

在 VSLAM 闭环检测方法中,通常采用的是图像特征的相似度比较检索方法^[6-7],即首先对图像提取特征,利用特征向量间的相似距离在历史图像中查找和检测图像相似度最大的图像,检索结果就是检测图像的闭环候选图像。当前,对图像提取的特征是利用人工特征算法或者预训练的深度学习网络提取,存在对图像有效信息表征不足^[8-9]的问题,会直接导致闭环检测精确率的下降。此外,目前相似图像搜索的方法通常采用的是全域暴力搜索和基于概率的区域搜索,但暴力的搜索方法随着图像的增多耗时会增大^[10],基于概率的区域搜索通用性差。

针对上述问题,一些研究者进行改进。在提高图像有效信息表征方面,李晓等提出了一种利用预训练的 VGG16 卷积神经网络末端池化层,对图像提取特征,但仅使用图像的全局语义特征来表征图像^[11];刘玮等提出对 YOLOv5 网络的损失

函数进行改进,接着对图像提取语义特征,但对图像的特征质量还有提升的空间^[12];胡正南等提出将预训练的 Vision Transformer 模型与 ResNet34 模型进行模型融合,利用融合模型对图像进行特征提取,增强了图像有效信息的表征能力,有效提高了闭环检测的准确性^[13]。

在相似图像搜索效率上,一些研究者也提出很多优化策略,莫桂江等对图像提取特征后,利用欧几里得距离结合暴力的搜索方法,查找检测图像的相似图像,但搜索耗时大^[14];赵珊等使用基于 ORB(Oriented FAST and Rotated BRIEF)特征生成的 Binary-VLAD(Vector of Locally Aggregated Descriptors)描述符,利用改进的倒排索引结构在全局进行粗搜索,接着利用 ORB 特征对闭环候选项进行验证,但查找闭环图像效率较低^[15];杨馨竹等在 Darknet 主干网络中添加了 NetVLAD 池化层,对图像提取特征后,采用概率预测结合贪心算法的相似图像搜索,提高了闭环准确率,但算法通用性存在一定局限性^[16]。以上相似图像搜索算法,存在搜索结果准确率低和搜索耗时大的问题,进而会影响闭环检测的性能。

目前,主流的闭环检测算法主要采用单一特征对图像进行表征,存在对图像有效信息表征不足的情况,而采用双特征来表征图像的闭环检测算法研究较少;同时,主流的闭环检测算法在进行相似图像搜索时,存在闭环检测实时性差和准确率低,不能较好在高精度的实时闭环检测系统上进行应用。基于此,提出了一种基于卷积特征和传统人工特征的 VLAD 融合并结合层次化可导航小世界网络的快速闭环检测算法。首先,利用

ResNet18 卷积网络和 ORB 特征算法对图像进行双特征提取,构建 ResNet18 卷积特征和 ORB 特征的视觉词典,对提取的 ResNet18 特征和 ORB 特征进行 VLAD 编码;对基于 ResNet18 特征的 VLAD 和基于 ORB 特征的 VLAD 进行融合,研究使用不同权重融合系数进行融合的 VLAD 编码对图像有效信息表征质量的影响;最后,针对闭环图像搜索存在耗时大的问题,引入层次化可导航小世界 HNSW (Hierarchical Navigable Small World) 网络^[17],利用 HNSW 网络的图索引构建 VLAD 索引结构,对融合后的 VLAD 编码进行加速检索,进而实现准确率高且快速的闭环检测。

1 闭环检测基础

闭环检测 (Loop Closure Detection, LCD) 作为视觉 SLAM 系统中至关重要的一部分,可以有效消除因机器人长时间移动而造成的累积误差,避免误差的传递。同时,闭环检测可以对机器人进行重定位,降低视觉里程计所产生的累计漂移对系统的影响,将检测到的闭环信息送入后端进行优化处理,提高系统的建图和定位精度,从而构建全局一致的轨迹和地图。闭环检测的方法主要是基于图像特征比较的方法,即通过对图像特征之间利用距离公式进行相似性比较,从而得到检测图像的闭环候选图像。闭环检测方法的示意图如图 1 所示,首先对图像序列进行图像特征提取,然后利用特征间的相似性来对图像进行闭环搜索。

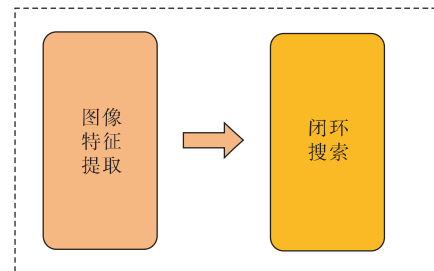


图 1 闭环检测方法

Fig. 1 Loop closure detection method

图像特征提取通常利用卷积神经网络末端的卷积层对输入图像提取特征,或者利用传统的人工特征 SURF (Speeded Up Robust Features) 特征算法和 ORB 特征算法对输入图像提取特征。通过利用图像的提取特征来表征图像的有效信息,然后将提取的图像特征用于闭环搜索。

闭环搜索是对当前输入图像在历史图像中进行相似图像搜索,进而判断是否出现了闭环。通常搜索相似图像采用的是以蛮力搜索方式通过利用欧式距离或余弦距离来比较图像特征向量之间的距离大小,从而得到可能性最大的闭环关系图像。

2 基于双特征和 HNSW 网络的闭环检测算法

闭环检测算法整体框架,如图 2 所示,由图像特征提取及 VLAD 融合和 HNSW 图搜索 2 部分组成。

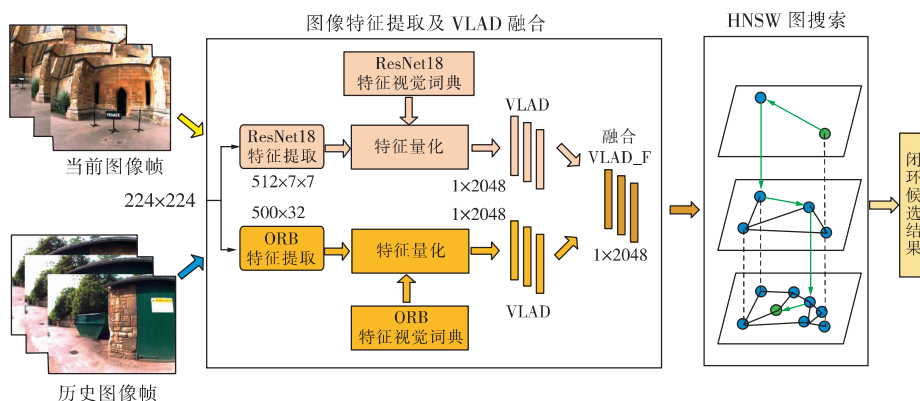


图 2 文中算法整体框图

Fig. 2 Overall block diagram of the proposed algorithm

2.1 图像双特征提取与特征的 VLAD 编码及融合

2.1.1 ResNet18 卷积特征提取

ResNet18 网络是一种深度残差卷积神经网络,与其他卷积网络^[18]如 VGG 和 AlexNet 网络相

比,ResNet18 卷积网络通过残差连接可以有效提取图像中的多层次特征^[19],并且其残差连接可以对网络中的不同层特征进行跨层融合,因此 ResNet18 卷积网络具有很好的多尺度融合特征提取

能力。Places365-Standard 数据集是一种包含 365 种不同的室内外视觉场景数据集,在此数据集上训练的网络模型提取图像特征的质量更好,为此采用在 Places365-Standard 数据集上预训练的 ResNet18 网络来进行图像特征的提取。

ResNet18 网络末端的全连接层 FC 和均值池化层 avgpool 是对网络最后一个卷积层的输出进行均值和维度处理,使用网络末端的 FC 层和 avgpool 层来提取特征,会降低图像特征的代表质量。因此,采用 ResNet18 网络末端卷积层的输出,作为输入图像的卷积特征。

输入图像 ResNet18 卷积特征提取步骤如下。

1) 输入层卷积和下采样。输入大小为 224×224 的 RGB 图像,对输入图像进行卷积和下采样处理。其中,卷积层是一个卷积核大小为 7×7 、步长为 2 的卷积操作构成的;下采样处理是通过一个卷积核大小为 3×3 、步长为 2 最大池化(Max Pooling)构成的。 224×224 的输入图像经过卷积和下采样处理之后得到大小为 56×56 的特征图。

2) 卷积层。网络包含 layer 1, layer 2, layer 3 和 layer 4 共 4 个卷积层,每个 layer 都是由 2 个残差块组成的,每个残差块又由 2 个卷积核大小为 3×3 的卷积处理以及一个跳跃连接组成。这样由输入层得到大小为 56×56 的特征图经过 layer1 处理之后大小仍然为 56×56 ,在依次经过 layer2 - layer4 的处理后特征图的大小依次减半,得到的特征图大小分别为 $28 \times 28, 14 \times 14, 7 \times 7$ 。

3) 输出特征。输入图像经过 ResNet18 网络处理后,在网络最后一个卷积层得到大小为 $512 \times 7 \times 7$ 的图像表征特征。

2.1.2 ORB 特征提取

ORB 特征具有计算速度快和鲁棒性高的特点,因而在计算机视觉领域被广泛应用。与其他传统特征相比较,ORB 特征^[20]能够提取图像中更多的局部几何特征,并且可以对特征点进行高效检测。ORB 特征是由 FAST 角点和 BRIEF 特征描述子组成的,该特征具有对尺度光照和旋转的良好不变性,利用 OpenCV 库中的 ORB 算法进行特征提取。

提取 ORB 特征分为以下 2 个步骤。

1) FAST 角点提取。在图像中选取像素 P ,检测以像素 P 为圆心,半径为 3 至 4 个像素的圆上是否有连续 n 个像素的亮度值 I_n ,满足圆心所在的亮

度值 I_p 与 I_n 差的绝对值都是大于阈值 T ,如果有那么认为该像素点为特征点,判断公式如下

$$|I_p - I_n| > T \quad (1)$$

式中 阈值 T 通常设置为 I_p 的 0.25 倍。

2) 特征描述子生成。利用图像的矩来得到图像的灰度质心方向,并按照灰度质心方向对原图像进行旋转,然后将特征点范围内 n 个随机点的像素进行对比,来得到二进制码,即 ORB 特征描述子。

2.1.3 提取特征的 VLAD 编码和融合

局部聚合描述符^[21] (Vector of Locally Aggregated Descriptors, VLAD) 是一种通过局部迭代的方式聚集特征描述子的细致图像特征编码方法,在提高特征表征的精度同时还可以降低内存的消耗。因此,利用 VLAD 算法对提取的 ResNet18 特征和 ORB 特征进行编码,然后对这 2 个 VLAD 编码进行融合。

1) 图像特征的 VLAD 编码。为了实现提取图像特征的 VLAD 编码,需要构建 ResNet18 特征和 ORB 特征的视觉词典。为视觉词典质量更好,在 Places365-Standard 数据集上进行训练,构建 VLAD 编码需要的离线视觉词典。视觉词典的构建过程^[22]:利用 ResNet18 网络(ORB 特征算法)提取数据集上图像的特征,然后利用 K-means 算法对图像特征向量进行 K 聚类,将聚类中心 C_i 作为视觉词汇,将所有的视觉词汇进行组合,最后构成视觉词典 $C = \{C_1, C_2, \dots, C_k\}$ 。如图 3 所示,白色小圈代表从数据集中提取的特征,整个特征空间利用 K-means 聚类算法被划分为 K 值是 7 的子空间,每个子空间的红色圈表示聚类中心,将其作为视觉词汇,组合所有的视觉词汇就构建成视觉词典。

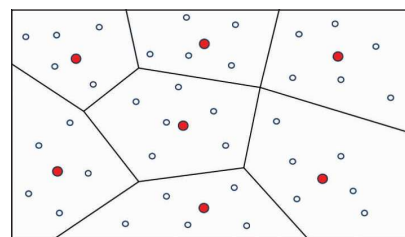


图3 视觉词典示意图

Fig. 3 Schematic diagram of visual dictionary

在构建视觉词典后,输入图像 I 的提取特征就可以用相应的 VLAD 编码进行表示^[23],步骤如下。

Step 1: 使用 ResNet18 网络 (ORB 算法) 提取输入图像的特征描述 $X = \{X_1, X_2, \dots, X_n\}$, 在视觉词典 $C = \{C_1, C_2, \dots, C_k\}$ 上进行特征量化, 获得每个特征描述子最近邻的视觉单词, 表示如下

$$NV(X_i) = \arg_k \min \|X_i - C_i\| \quad (2)$$

式中 X_i 为每个特征的描述子向量; C_i 为特征描述子向量最近邻的视觉单词, $i \in 1, \dots, k$ 。

Step 2: 对于每个特征的描述子向量, 计算该向量与视觉词典中每一个词汇间的距离, 进一步确定描述子向量所属的词汇簇, 对于同一个簇, 根据描述子向量的均值重新确定聚类中心, 然后计算全部描述子向量 X_i 与各自所属类别词汇 C_i 之间的残差和 V_i , 见下式

$$V_i = \sum_{i: NN(i)=k} (X_i - C_i) \quad (3)$$

将视觉词典上的所有残差和串联起来, 得到特征向量 $V = \{V_1, V_2, \dots, V_k\}$, 其维度为 $k \times d$, d 为特征描述子的维数。

Step 3: 对向量 V 进行带符号的平方根运算得到 V' , 接着对 V' 作 L2 范数归一化, 最后得到一个长度为 $k \times d$ 的 VLAD 描述向量 V , 将此描述向量 V 作为图像特征的 VLAD 编码。

$$V' = |V_i|^a \cdot \text{sign}(V_i); 0 \leq a \leq 1 \quad (4)$$

$$V = \frac{V'}{\|V'\|_2} \quad (5)$$

2) 图像特征的 VLAD 融合。为了充分利用这 2 种不同类型特征的 VLAD 对图像的有效描述信息, 将这 2 种 VLAD 编码进行加权融合。对输入图像由 ResNet18 特征生成的 VLAD, 表示为 V_R , 对输入图像由 ORB 特征生成的 VLAD, 表示为 V_O 。将 V_R 编码和 V_O 编码进行加权融合, 融合公式如下

$$V_F = [\alpha \cdot V_R + \beta \cdot V_O]; \alpha, \beta \in [0, 1] \quad (6)$$

式中 V_F 为由 V_R 编码和 V_O 编码进行加权融合后的 VLAD 编码; α, β 为权重参数, $\alpha + \beta = 1$ 。

2.2 基于 HNSW 网络的闭环候选图像搜索

在输入图像被表示成融合的 VLAD 编码后, 要判断当前输入图像和历史图像是否出现了可能的闭环, 就要对当前输入图像在历史图像中进行相似图像搜索, 即检索当前输入图像的闭环候选图像。目前, 闭环候选图像的搜索通常采用全域暴力搜索和基于概率的区域搜索来得到闭环候选图像, 但存在通用性差和闭环候选图像搜索效率低的问题, 而 HNSW 网络是一种效率高通用性强的搜索结构, 为此通过引入 HNSW 网络, 利用

HNSW 网络的图索引构建 VLAD 索引结构, 使用构建的 VLAD 索引结构对融合后的 VLAD 编码进行加速检索, 从而实现快速的图像闭环候选搜索。

2.2.1 HNSW 网络图索引构建 VLAD 索引结构

HNSW 网络是一种完全基于图的近似最近邻搜索结构, 具有增量图构建的性质, 其本质是把待索引向量构造成图结构 $G = \{V_n, E\}$, V_n 为待索引向量形成的节点集合; E 为所有节点连接关系组成的边集合^[24]。将输入图像融合后的 VLAD 编码特征作为网络的节点, 利用 HNSW 网络的图索引构建 VLAD 索引结构。

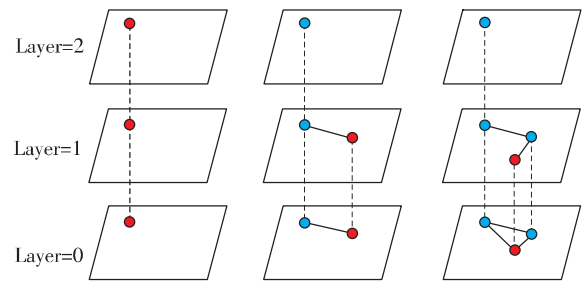


图4 VLAD索引结构构建示意图

Fig. 4 Schematic diagram of VLAD index structure construction

利用 HNSW 图索引构建 VLAD 索引结构, 以 3 层索引结构为例, 构建过程如图 4 所示。在构建开始时, 索引结构图中没有节点, 需要向索引结构图中插入新节点。图中红色节点均作为插入到索引结构图中的新节点, 插入的新节点通过指数衰减的概率分布确定该节点所在的最顶层数, 同时低于这个层数的其他层也会存在该节点, 将所有层的该节点存入该节点构造的图结构 $G = \{V_n, E\}$ 的 V_n 中。图 4 的左图是插入第 1 个新节点, 中间图是插入第 2 个新节点, 每插入的新节点从最顶层开始, 使用贪婪搜索算法利用 L2 距离在该层寻找与其最近邻的若干邻居节点, 并且与邻居节点建立节点连接关系。该过程自上层到下层逐层进行, 直到新插入的该节点在第零层上也建立连接, 同时将所有邻居节点和连接关系边, 存入新插入节点构造的图结构 $G = \{V_n, E\}$ 中。这样通过不断的向索引图中插入新节点, 重复按照上述流程, 就可以构建出一定规模的基于 HNSW 图索引的 VLAD 索引结构。其中每个节点的最大邻居数量 M , 是重要的构造参数, 决定索引图构建的时间成本。

2.2.2 基于 HNSW 网络的闭环候选图像搜索

对输入图像进行 HNSW 图搜索, 其实质上是

在索引结构图的所有 VLAD 节点中,找到输入 VLAD 节点的最近似节点。首先将得到的输入图像 VLAD 节点,按照索引结构构建步骤插入到索引图中,同时在每一层中与其最近邻的若干邻居节点建立连接关系边。接着进行搜索过程,首先在输入图像的 VLAD 节点的最高层网络图中随机选择一个节点,从选择节点开始在该层网络图中依据贪婪搜索算法利用 L2 距离搜索其最近邻节点,在搜索到最近邻节点后跳转到下一层的此节点,从此节点开始继续在这一层网络图中搜索其最近邻节点,在搜索到最近邻节点后继续跳转到下一层,重复上述过程直到在第零层网络也进行最近邻节点搜索,最终在第零层网络搜索到的最近邻节点就是输入 VLAD 节点的最近似节点^[25]。

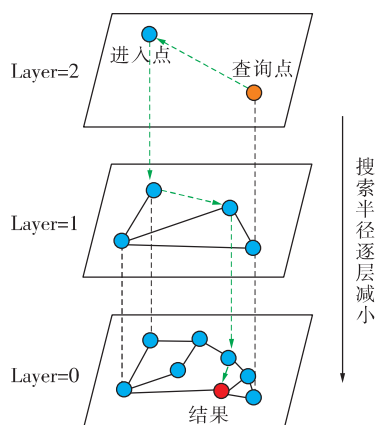


图5 HNSW 图搜索示意图

Fig. 5 Schematic diagram of HNSW graph search

图5是3层的HNSW图搜索示意图,其中橙色节点代表查询节点,即当前时刻待检索图像对应的VLAD节点,在查询节点插入到索引结构图之后,对查询节点进行搜索的过程如下:搜索从查询节点所在的最高层Layer2中随机选择的“进入点”开始,依据贪婪搜索算法利用L2距离来搜索查询节点的最近邻节点,在搜索到Layer2层查询节点的最近邻节点后,跳转到Layer1层的此节点,从此节点开始在Layer1层中继续搜索其最近邻节点,最后跳转到Layer0层中对应的节点进行最终搜索。随着在搜索过程中不断跳转到下一层中进行搜索,搜索半径也在逐层减小,最后在Layer0层查找到查询节点的最近似节点,其标注为“结果”。通过找到查询节点的最近似节点,进而最近似节点所对应的图像就是当前检索图像的闭环候选图像。

3 试验与分析

为了验证文中提出闭环检测算法的性能,在闭环检测公开数据集CityCentre和NewCollege上进行测试。CityCentre数据集是在城市中心拍摄的,包含树木草丛,房屋建筑以及道路车辆等视觉场景。NewCollege数据集是在大学校园内拍摄的,包含室内走廊,室外建筑以及室外树丛等视觉场景。数据集的相关信息见表1。

表1 数据集的相关信息

Table 1 Information about the datasets

数据集	总长度/km	图像数量/张	图像尺寸/px
CityCentre	2.0	2 474	640 × 480
NewCollege	1.9	2 146	640 × 480

2个公开数据集都有其相应的GroundTruth矩阵,里面标注了哪些图像对是真实闭环,分别用 $2\,474 \times 2\,474$ 和 $2\,146 \times 2\,146$ 的二维矩阵进行存储标注。比如行序号的图像*i*和列序号的图像*j*在GroundTruth矩阵中对应的(*i*,*j*)元素值为1,则代表图像*i*和图像*j*是闭环关系,若为0,则是代表不是闭环关系。因此,利用数据集的GroundTruth矩阵,对当前输入图像的闭环候选图像进行验证,进而反应算法的闭环检测准确性。

采用在Places365-Standard数据集上预训练的ResNet18网络模型,对图像进行ResNet18特征提取,利用OpenCV库中的ORB算法来提取输入图像的特征,将尺寸调整为 224×224 。VLAD字典规模取值过低,会降低编码表征信息的精度,取值过高又会增加搜索时长,因此折中选取字典规模为64。HNSW最大邻居数量在32、64和128中通过对比发现,使用ResNet18特征进行试验时,取值为64的闭环准确率比取值32高出15%,而取值为128的闭环准确率比取值64只高出3%,同时耗时增加41%,为此最终折中选取HNSW最大邻居数量为64。试验设备为Intel(R)Core(TM)i7-12700H,内存16GB,显存GTX3050 4G,深度学习环境Anaconda3,Pytorch,操作系统为Ubuntu 18.04LTS。关键参数见表2。

3.1 图像VLAD编码融合

V_R 编码和 V_O 编码的融合是通过通过对 V_R 编码和 V_O 编码分别赋值 $[0,1]$ 范围内的不同权重系数,通过在闭环检测数据集上使用暴力搜索与余弦距离

相结合的图像闭环衡量方式,利用平均准确率^[26](Average-Precision) AP 对不同权重系数的 VLAD 融合效果进行衡量比较,从而得到一组最优的融合权重系数,使融合后的 VLAD 编码特征能够表征更多的图像有效信息。

表 2 关键参数与取值

Table 2 Key parameters and values

参数	取值
输入图像大小	224 × 224
归一化(均值、标准差)	(0.485, 0.456, 0.406) (0.229, 0.224, 0.225)
VLAD 字典规模	64
VLAD 向量维度	2 048
VLAD 距离度量	L2-norm
HNSW 最大邻居数量	64

平均准确率 AP,是指精确率-召回率曲线在横纵坐标轴上围成的面积。通常,AP 值越大,则代表闭环检测的准确性越好。在每一组权重系数的 VLAD 融合试验中,由于统一使用暴力搜索结合余弦相似距离的图像闭环衡量方式,所以当 AP 值越大,则代表使用相应组的权重系数生成融合后的 VLAD 编码对图像有效信息的表征质量更好。为此,按照公式(6),在闭环检测数据集上进行融合测试对比试验,见表 3。

表 3 NewCollege 和 CityCentre 数据集上 VLAD 融合试验

Table 3 VLAD fusion experiment on NewCollege and CityCentre datasets

权重参数(α, β)	AP 值(NC)	AP 值(CC)
(1.0, 0.0)	0.61	0.60
(0.9, 0.1)	0.65	0.62
(0.8, 0.2)	0.71	0.66
(0.7, 0.3)	0.79	0.70
(0.6, 0.4)	0.86	0.73
(0.5, 0.5)	0.83	0.75
(0.4, 0.6)	0.78	0.72
(0.3, 0.7)	0.71	0.68
(0.2, 0.8)	0.63	0.62
(0.1, 0.9)	0.57	0.55
(0.0, 1.0)	0.55	0.53

α, β 分别为 V_R 编码和 V_O 编码的权重参数, AP 值(NC)与 AP 值(CC)分别代表在 NewCollege 和 CityCentre 数据集上进行 VLAD 融合试验计算的 AP 值。在 2 种闭环检测数据集上以不同系数

融合 V_R 编码和 V_O 编码进行试验时的 AP 值,均比单独使用 V_R 编码或 V_O 编码时要高。由此说明,卷积特征和传统人工特征的 VLAD 编码都没有充分表征图像的有效信息,而使用 ResNet18 卷积特征和 ORB 特征的 VLAD 融合编码,可以表征更多的图像有效信息。最终通过比较,在 NewCollege 和 CityCentre 数据集上选择最优权重系数分别为 (0.6, 0.4) 和 (0.5, 0.5),通过最优权重系数将 V_R 编码和 V_O 编码进行加权融合,从而使用对图像有效信息表征质量好的融合 VLAD 编码特征进行闭环检测。

在确定最优权重融合系数时,采用预设的权重系数组合进行试验,需要重复进行多次试验,这样会降低算法的实用性。然而最优权重系数也可以通过学习机制来自动确定,首先设置基于平均准确率的目标损失函数 $L(\alpha, \beta)$, η 为学习率,对权重参数 α, β 进行初始化,接着开始进行迭代,每次迭代根据目标损失函数的损失值,利用梯度下降法更新权重,在进行多次迭代之后,就可以自动得到最优的权重融合系数,从而进一步提高算法的实用性。

3.2 评价指标

采用 PR 曲线和闭环搜索时间 2 个指标,对文中算法和其他算法进行对比试验。

3.2.1 PR 曲线

在闭环检测中,如果 2 幅来自于不同场景的图像被错误判断为来自于同一场景,则称这种情况为假阳性(False Positive, FP);如果 2 幅来自于同一场景的图像被错误判断为来自于不同场景,则称这种情况为假阴性(False Negative, FN)。那么相反,如果 2 幅来自于同一场景的图像被正确判断为来自于同一场景,则称这种情况为真阳性(True Positive, TP),如果 2 幅来自于不同场景的图像被正确判断为来自于不同场景,则为真阴性(True Negative, TN)。因此,2 个统计量精确率和召回率定义如下

$$\text{precision} = \frac{TP}{TP + FP}, \text{recall} = \frac{TP}{TP + FN} \quad (7)$$

精确率表示的是算法提取的所有闭环中确实是真实闭环的概率,召回率则表示在所有真实闭环中被正确检测出来的概率。为了评价闭环检测算法的好坏,通常会计算出算法检测过程中的精确率 P 和召回率 R ,然后绘制 PR 曲线。当使用召回率为横轴,精确率为纵轴时,通常以整条 PR 曲

线偏向右上方的程度,作为算法的评价指标。

3.2.2 闭环搜索时间

闭环搜索时间是对输入图像在历史图像中进行闭环搜索,从开始搜索到得到输入图像闭环候选图像之间所用的时间。

3.3 试验结果分析

3.3.1 PR 曲线对比

选取以下 6 种算法进行对比试验:①基于传统视觉词袋模型的 ORB-DBoW2 算法;②(基准算法)基于 ResNet18 网络的 ResNet18-VLAD 算法;③基于 ResNet18 网络和 ORB 特征的 ResNet18-VLAD_F 算法;④(文中算法)基于 ResNet18 网络、ORB 特征和 HNSW 网络的 ResNet18-VLAD_F-HNSW 算法;⑤基于 Darknet 网络的 Darknet-NVLAD-PGreedy 算法^[16];⑥近 2 年基于主流深度模型 Vision Transformer 的 Transformer-ResAlex-KPCA 算法。对应的 PR 曲线如图 6 和图 7 所示。

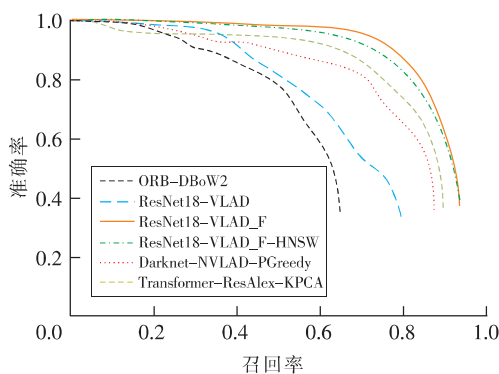


图 6 NewCollege 数据集上的 PR 曲线

Fig. 6 PR curves on NewCollege dataset

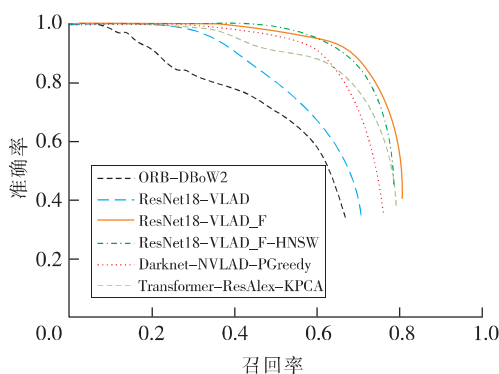


图 7 CityCentre 数据集上的 PR 曲线

Fig. 7 PR curves on CityCentre dataset

图 6 和图 7 展示了以上 6 种算法分别在 NewCollege 和 CityCentre 这 2 个闭环检测数据集上进行试验对应的 PR 曲线,以图 6 在 NewCollege 数据

集上的 PR 曲线为例进行分析,ResNet18-VLAD 基准算法的 PR 曲线效果比传统闭环检测算法 ORB-DBoW2 要好,在高召回率的时候保持更高的精确率。基于 ResNet18 特征和 ORB 特征的 ResNet18-VLAD_F 算法,PR 曲线效果有了很大的提升,与 ResNet18-VLAD 算法相比,其整体精确率-召回率提高了 29%。而引入了 HNSW 网络的 ResNet18-VLAD_F-HNSW 算法与之前的 ResNet18-VLAD_F 算法相比,PR 曲线效果略微下降,但其仍然具有极好的整体精确率-召回率检测特性。此外,与近 2 年闭环检测性能好 Darknet-NVLAD-PGreedy 算法^[16]和基于主流深度模型 Vision Transformer 的 Transformer-ResAlex-KPCA 算法进行比较,文中算法的整体精确率-召回率也分别高出了 6% 和 4%。

相比较于基准算法,使用 ResNet18 和 ORB 双特征的 ResNet18-VLAD_F 算法和 ResNet18-VLAD_F-HNSW 算法的整体精确率-召回率有很大的提高,这主要是由于文中算法采用双特征来对图像的有效信息进行表征。ResNet18 特征是利用卷积网络对图像中的多层次特征进行提取,代表图像的全局特征;而 ORB 特征主要对图像中的局部几何特征进行提取,代表图像的局部特征。通过将输入图像 ResNet18 全局特征和 ORB 局部特征的 VLAD 编码利用最优融合权重系数进行融合,使得 VLAD 融合编码不仅包含输入图像的全局特征有效信息,也包含输入图像的局部特征有效信息,从而提升 VLAD 编码对图像有效信息的表征质量。通过 VLAD 融合试验和 PR 曲线对比可进一步验证,相比较单一特征的 VLAD 编码,双特征的 VLAD 融合编码表征图像有效信息的能力更好。

图 8 展示部分被基准算法忽略但被文中算法检测到的闭环图像。图像序号 899 与图像序号 2025 为一组闭环图像,这组闭环图像的视角发生了较大变化,文中算法可以正确识别这组闭环,说明该算法对视角变化有较好的鲁棒性。图像序号 621 与图像序号 465 为另一组闭环图像,这组闭环图像不仅视角发生了较大变化,而且还存在动态干扰物(行人),但文中算法仍可以正确识别这组闭环,说明文中算法也对动态干扰物有较好的鲁棒性。

3.3.2 搜索时间对比

在 PR 曲线对比部分可以看出,与 ResNet18-



图8 闭环检测示例

Fig. 8 Illustration of loop closure detection

VLAD_F 算法 PR 曲线效果相比,引入了 HNSW 网络的 ResNet18-VLAD_F-HNSW 算法 PR 曲线效果却出现了精确率略微下降的情况,然而 HNSW 网络是一种速度快的近似最近邻搜索结构,为此通过对比全部图像闭环搜索时间和单幅图像平均闭环搜索时间,来衡量文中提出算法的实时性。在 NewCollege 数据集上,选取右侧相机拍摄的共 1073 张图像进行搜索时间指标对比,结果见表 4。

表 4 闭环搜索总时间及平均时间对比

Table 4 Comparison between total and average search time for loop closure detection

算法	总时间/s	平均时间/ms
ORB-DBoW2	3.14	3.10
ResNet18-VLAD	3.63	3.58
ResNet18-VLAD_F	4.38	4.32
Darknet-NVLAD-PGreedy	2.89	2.86
文中算法	2.14	2.11

ORB-DBoW2 算法采用优化的倒排索引结构进行闭环搜索,而 ResNet18-VLAD 和 ResNet18-VLAD_F 算法采用暴力搜索的闭环检索方式,结合算法对应的 PR 曲线图可以看出,与倒排索引的闭环搜索方式相比,暴力搜索的方式虽然在一定程度上提高了闭环检测的准确性,但闭环搜索时间明显增多。而 ResNet18-VLAD_F-HNSW 算法采用 HNSW 图索引构建的 VLAD 索引结构来加速检索,从而实现快速的图像闭环搜索,与暴力搜索的 ResNet18-VLAD_F 算法相比,文中算法 PR 曲线的精确率虽然稍微有所下降,但闭环搜索总时间减

少了 51%。Darknet-NVLAD-PGreedy 算法^[16]采用概率预测结合贪心算法的闭环检索方式,与其闭环搜索总时间相比,文中算法的闭环检测搜索总时间下降了 25%。ResNet18-VLAD_F-HNSW 算法与前 4 种闭环检测算法相比,文中算法的闭环搜索总时间及平均时间最少,在保持高召回率和高精确率的前提下,文中算法也具有快速闭环检测性能。

4 结 论

1) 相比较于单一 ResNet18 卷积特征和 ORB 人工特征的 VLAD 编码,双特征的 VLAD 融合编码有更高的闭环检测准确率,其闭环检测准确率最高可提升 29%,双特征的 VLAD 融合编码可有效提高图像有效信息的表征质量。

2) 相比于原始算法的暴力搜索,使用 HNSW 网络图索引构建的 VLAD 索引结构进行闭环搜索时,在确保高准确率结果的同时可以实现加速检索,其闭环搜索总时间可减少 51%。

3) 基于双特征并使用 HNSW 网络构建的 VLAD 索引结构进行搜索的闭环检测算法具有闭环检测准确率高、搜索时间短的优势,为构建实时高精度的视觉 SLAM 系统提供一定参考。

参考文献(References):

- [1] 高沛林,高贇. 移动机器人多信标 SLAM 技术[J]. 西安科技大学学报,2019,39(5):905-911.
GAO Peilin,GAO Yun. SLAM navigation of mobile robots based on multi-beacons[J]. Journal of Xi'an University of Science and Technology,2019,39(5):905-911.
- [2] 高贵,伍宣衡,王忠美,等. V-SLAM 深度学习闭环检测研究进展与展望[J]. 计算机工程与应用,2022,58(11):47-59.
GAO Gui,WU Xuanheng,WANG Zhongmei, et al. Research progress and prospect of V-SLAM deep learning loop closure detection[J]. Computer Engineering and Applications,2022,58(11):47-59.
- [3] ZHANG X,WANG L,SU Y. Visual place recognition: A survey from deep learning perspective[J]. Pattern Recognition,2021,113(12):1077-1089.
- [4] 刘凌云,蔡成林,吴芊. 室外光照易变场景下的回环检测方法[J]. 传感器与微系统,2022,41(4):121-124.
LIU Lingyun,CAI Chenglin,WU Qian. Loop closure de-

- tection method in outdoor scene with variable illumination [J]. *Transducer and Microsystem Technologies*, 2022, 41(4):121-124.
- [5] 张冰战,尹晨晨,李志远,等. 基于点云特征全局搜索的闭环检测算法[J]. *电子测量与仪器学报*, 2024, 38(3):176-186.
ZHANG Bingzhan, YIN Chenchen, LI Zhiyuan, et al. Loop closure detection algorithm based on global search of point cloud features [J]. *Journal of Electronic Measurement and Instrumentation*, 2024, 38(3):176-186.
- [6] 杨骏,胡为,朱文福. 基于改进 MobileNetV3 的视觉 SLAM 闭环检测算法[J]. *计算机与现代化*, 2024, 21(10):21-26.
YANG Jun, HU Wei, ZHU Wenfu. Visual SLAM loop closure detection algorithm based on improved MobileNetV3 [J]. *Computer and Modernization*, 2024, 21(10):21-26.
- [7] 赵燕成,房桐,杜保帅,等. 移动机器人视觉 SLAM 闭环检测现状研究[J]. *无线电工程*, 2023, 53(1):129-139.
ZHAO Yancheng, FANG Tong, DU Baoshuai, et al. Research on the current situation of loop closure detection in visual SLAM for mobile robots [J]. *Radio Engineering*, 2023, 53(1):129-139.
- [8] 蒋经纬,吉月辉,刘俊杰,等. 基于轻量级 CNN 的视觉 SLAM 快速闭环检测算法[J]. *计算机仿真*, 2024, 41(8):182-188.
JIANG Jingwei, JI Yuehui, LIU Junjie, et al. Fast loop closure detection algorithm for visual SLAM based on lightweight CNN [J]. *Computer Simulation*, 2024, 41(8):182-188.
- [9] 胡章芳,冯淳一,罗元,等. 基于 CNN 多层特征加权融合的闭环检测算法[J]. *华中科技大学学报(自然科学版)*, 2021, 49(8):75-80.
HU Zhangfang, FENG Chunyi, LUO Yuan, et al. Loop closure detection algorithm based on CNN multilayer feature weighted fusion [J]. *Journal of Huazhong University of Science and Technology (Natural Science Edition)*, 2021, 49(8):75-80.
- [10] 唐灿,唐亮贵,刘波. 图像特征检测与匹配方法研究综述[J]. *南京信息工程大学学报*, 2020, 12(3):261-273.
TANG Can, TANG Lianggui, LIU Bo. A review of image feature detection and matching methods [J]. *Journal of Nanjing University of Information Science & Technology*, 2020, 12(3):261-273.
- [11] 李晓,马社祥,李啸. 基于卷积神经网络的闭环检测算法[J]. *现代电子技术*, 2022, 45(1):72-76.
LI Xiao, MA Shexiang, LI Xiao. Loop closure detection algorithm based on convolutional neural network [J]. *Modern Electronics Technique*, 2022, 45(1):72-76.
- [12] 刘玮,温显斌. YOLO-NKLT 视觉 SLAM 闭环检测方法[J]. *光电子·激光*, 2023, 34(4):405-412.
LIU Wei, WEN Xianbin. YOLO-NKLT visual SLAM loop closure detection method [J]. *Journal of Optoelectronics · Laser*, 2023, 34(4):405-412.
- [13] 胡正南,胡立坤. 基于 Vision Transformer 多模型融合的视觉闭环检测算法[J]. *激光杂志*, 2024, 45(6):75-81.
HU Zhengnan, HU Likun. Visual loop closure detection algorithm based on vision transformer multimodel fusion [J]. *Laser Journal*, 2024, 45(6):75-81.
- [14] 莫桂江,颜增显. 基于 OBS-CNN 的视觉 SLAM 闭环检测方法研究[J]. *火力与指挥控制*, 2022, 47(12):93-97.
MO Guijiang, YAN Zengxian. Research on loop closure detection method of visual SLAM based on OBS-CNN [J]. *Fire Control & Command Control*, 2022, 47(12):93-97.
- [15] 赵珊,管启,丁德锐,等. COVFast-LCD:一种组合 ORB 和 VLAD 特征的快速闭环检测算法[J]. *小型微型计算机系统*, 2023, 44(6):1318-1323.
ZHAO Shan, GUAN Qi, DING Derui, et al. COVFast-LCD: A fast loop closure detection algorithm combining ORB and VLAD features [J]. *Journal of Chinese Computer Systems*, 2023, 44(6):1318-1323.
- [16] 杨馨竹,张建勋,郭纪志. Darknet-NVPP 视觉 SLAM 快速闭环检测方法[J]. *小型微型计算机系统*, 2023, 44(4):832-837.
YANG Xinzhu, ZHANG Jianxun, GUO Jizhi. Darknet-NVPP visual SLAM fast loop closure detection method [J]. *Journal of Chinese Computer Systems*, 2023, 44(4):832-837.
- [17] MALKOV Y, YASHUNIN D. Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(4):824-836.
- [18] 窦慧,张凌茗,韩峰,等. 卷积神经网络的可解释性研究综述[J]. *软件学报*, 2024, 35(1):159-184.
DOU Hui, ZHANG Lingming, HAN Feng, et al. Survey on interpretability of convolutional neural networks [J]. *Journal of Software*, 2024, 35(1):159-184.
- [19] SHAFIQ M, GU Z Q. Deep residual learning for image recognition: A survey [J]. *Applied Sciences*, 2022, 12(18):8972-8972.

- [20] LUO C, YANG W, HUANG P, et al. Overview of image matching based on ORB algorithm[J]. Journal of Physics: Conference Series, 2019, 1237(3): 29–31.
- [21] OKAWAM. From BoVW to VLAD with KAZE features: Offline signature verification considering cognitive processes of forensic experts[J]. Pattern Recognition Letters, 2018, 113(1): 75–82.
- [22] 李邨, 马慧宇, 李海燕, 等. 基于视觉词袋模型提取胶痕特征的卷烟真伪鉴别[J]. 包装工程, 2023, 44(15): 252–259.
LI Dan, MA Huiyu, LI Haiyan, et al. Cigarette authenticity identification based on visual bag-of-words model to extract glue mark features[J]. Packaging Engineering, 2023, 44(15): 252–259.
- [23] 张丽娜. 复杂环境下的vSLAM闭环检测算法研究[D]. 西安: 西安科技大学, 2021.
ZHANG Lina. Research on vSLAM loop closure detection algorithm in complex environment[D]. Xi'an: Xi'an University of Science and Technology, 2021.
- [24] 姜三, 马一尘, 李清泉, 等. 无序无人机影像的并行化SfM三维重建方法[J]. 测绘学报, 2024, 53(5): 946–958.
JIANG San, MA Yichen, LI Qingquan, et al. Parallel SfM three dimensional reconstruction method for unordered drone imagery[J]. Acta Geodaetica et Cartographica Sinica, 2024, 53(5): 946–958.
- [25] NALLAPPAN M, VELSWAMY R. Exploring deep learning based content-based video retrieval with hierarchical navigable small world index and ResNet-50 features for anomaly detection[J]. Expert Systems with Applications, 2024, 247(12): 3197–3217.
- [26] 伍宣衡, 高贵, 王忠美, 等. 结合多重注意力机制的V-SLAM闭环检测特征匹配算法[J]. 湖南工业大学学报, 2023, 37(5): 9–16.
WU Xuanheng, GAO Gui, WANG Zhongmei, et al. V-SLAM loop closure detection feature matching algorithm combined with multiple attention mechanisms[J]. Journal of Hunan University of Technology, 2023, 37(5): 9–16.

(责任编辑: 刘洁)

(上接第255页)

- [40] 陶静. 液氮预注后页岩压裂的损伤破裂机理研究[D]. 徐州: 中国矿业大学, 2020.
TAO Jing. Study on damage mechanics of shale nitrogen fracturing after liquid nitrogen pre-conditioning[D]. Xuzhou: China University of Mining and Technology, 2020.
- [41] 张向向. 层理页岩水/气压裂破裂机制与损伤演化规律研究[D]. 徐州: 中国矿业大学, 2019.
ZHANG Xiangxiang. Fracturing mechanism and damage evolution of water/gas fracturing in bedding shale[D]. Xuzhou: China University of Mining and Technology, 2019.
- [42] 程远平, 刘洪永, 郭品坤, 等. 深部含瓦斯煤体渗透率演化及卸荷增透理论模型[J]. 煤炭学报, 2014, 39(8): 1650–1658.
CHENG Yuanping, LIU Hongyong, GUO Pinkun, et al. A theoretical model and evolution characteristic of mining-enhanced permeability in deeper gassy coal seam[J]. Journal of China Coal Society, 2014, 39(8): 1650–1658.
- [43] 中国煤炭工业协会. 煤和岩石物理力学性质测定方法第7部分 单轴抗压强度测定及软化系数计算方法: GB/T 23561. 7—2009[S]. 北京: 中国标准出版社, 2009.
- [44] 景锋, 盛谦, 张勇慧, 等. 中国大陆浅层地壳实测地应力分布规律研究[J]. 岩石力学与工程学报, 2007, 26(10): 2056–2062.
JING Feng, SHENG Qian, ZHANG Yonghui, et al. Research on distribution rule of shallow crust of China mainland[J]. Chinese Journal of Rock Mechanics and Engineering, 2007, 26(10): 2056–2062.
- [45] 康红普, 伊丙鼎, 高富强, 等. 中国煤矿井下地应力数据库及地应力分布规律[J]. 煤炭学报, 2019, 44(1): 23–33.
KANG Hongpu, YI Bingding, GAO Fuqiang, et al. Database and characteristics of underground in-situ stress distribution in Chinese coal mines[J]. Journal of China Coal Society, 2019, 44(1): 23–33.
- [46] 梁鑫. 岩石水/气压裂分形破裂机理与分形离散裂隙网络研究[D]. 徐州: 中国矿业大学, 2019.
LIANG Xin. Study on fractal fracture mechanism of rock under water & gas fracturing and fractal discrete fracture networks[D]. Xuzhou: China University of Mining and Technology, 2019.

(责任编辑: 刘洁)