doi:10.3969/j.issn.1673-9833.2023.05.002

结合多重注意力机制的 V-SLAM 闭环检测特征匹配算法

伍宣衡[1]，高 贵[1,2]，王忠美[1]，薛子豪[1]，龙永红[1]

（1. 湖南工业大学 轨道交通学院，湖南 株洲 412007；2. 西南交通大学 地球科学与环境工程学院，四川 成都 610000）

摘 要：为了在复杂环境下对 V-SLAM 闭环检测的准确率 - 召回率有更好的鲁棒性，提出一种在图神经网络中结合多重注意力机制的局部特征匹配算法，并在闭环检测上进行应用。首先，采用 SuperPoint 检测器获取图像序列中的关键点，再将提取出来的特征点输入关键点编码器内，通过多层感知器将其升维到与局部描述子维度一样；然后，同时经过多重注意力机制网络中重复 9 次，得到更具有代表信息的局部描述子；其次，在最优匹配层中采用 SinkHorn 算法求解出最优匹配矩阵，通过对阈值的合理设定，得到闭环检测结果；最后，在 New College 和 City Centre 两个公共数据集上与 5 种其他闭环检测基准算法进行实验，结果表明该算法在召回率一定的情况下，其准确率比其他实验算法的要高，有更强的鲁棒性，满足闭环检测要求。

关键词：同步定位与建图；闭环检测；图神经网络；多重注意力机制

中图分类号：TP311 **文献标志码**：A **文章编号**：1673-9833(2023)05-0009-08

引文格式：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.

V-SLAM Loop Closure Detection Feature Matching Algorithm Combined with Multiple Attention Mechanisms

WU Xuanheng[1]，GAO Gui[1,2]，WANG Zhongmei[1]，XUE Zihao[1]，LONG Yonghong[1]

（1. College of Railway Transportation，Hunan University of Technology，Zhuzhou Hunan 412007，China；
2. Faculty of Geosciences and Environmental Engineering，Southwest Jiaotong University，Chengdu 610000，China）

Abstract：In order to obtain an improved robustness to the accuracy recall of V-SLAM loop closure detection in complex environments, a local feature matching algorithm, combined with multiple attention mechanisms in graph neural network, has been proposed with an application to the loop closure detection. Firstly, the SuperPoint detector is used to obtain the key points in the image sequence, followed by an input of the extracted feature points into the key point encoder, with its dimension raised to the same as the local descriptor sub-dimension by using a multi-layer perceptron. Then, a more representative local description can be obtained after being repeated 9 times in a multiple attention mechanism network. Next, the SinkHorn algorithm is used to solve the optimal matching matrix in the optimal

收稿日期：2023-02-01

基金项目：国家级创新创业基金资助项目（S202111535505，S202111535041）；湖南省教育厅科研基金资助项目（22A0391，22B0586）

作者简介：伍宣衡（1995-），男，湖南耒阳人，湖南工业大学硕士生，主要研究方向为 SLAM，图像处理，E-mail：wuxuanheng424@163.com

通信作者：王忠美（1984-），男，湖北荆州人，湖南工业大学讲师，博士，硕士生导师，主要研究方向为智能信息处理，E-mail：ldwangzm2008@163.com

matching layer, thus obtaining the loop closure detection result by setting the threshold reasonably. Finally, experiments are conducted, alongside with five other loop closure detection benchmark algorithms, on two common datasets of New College and City Centre. The results show that the proposed algorithm is characterized with a higher accuracy and a stronger robustness than other experimental algorithms under a certain recall rate, meeting the requirements of closed-loop detection.

Keywords：simultaneous localization and mapping（SLAM）；loop closure detection；graph neural network；multiple attention mechanism

1 研究背景

同步定位与建图（simultaneous localization and mapping，SLAM）技术，在过去几十年的快速发展过程中日益成熟，被认为是机器人实现自主导航的关键技术之一。经典的 SLAM 系统架构，主要由传感器数据的读取、视觉里程设计（visual odometry，VO）、后端优化、闭环检测和地图构建 5 个部分组成 [1]。同时，通过所搭载传感器的不同，又可以分为视觉 SLAM（vision simultaneous localization and mapping，V-SLAM）、激光 SLAM 和多传感器融合的 SLAM。闭环检测作为 SLAM 中的一部分，一直是一个热点研究问题 [2]，因为精确的闭环检测可以对 SLAM 系统进行重定位，并能减少在前端视觉里程计中所带来的漂移误差，从而提高地图精度 [3]。

视觉 SLAM 的闭环检测问题，是指移动机器人在导航过程中，通过摄像头传感器输入的图像数据来识别前期去过的地方，在一定程度上可将其视为图像识别与检索问题。传统的 V-SLAM 闭环检测主要依赖于人工设计的特征进行设计，对于图像特征的提取，常采用尺度不变特征变换（scale-invariant feature transform，SIFT）、加速稳健特征 (speeded up robust features，SURF)、加速分段试验特征（features from accelerated segment test，FAST），以及定向 fast 特征点提取和旋转 brief 描速子（oriented fast and rotated brief，ORB）图像局部特征等。在目前较为成熟的传统视觉闭环检测中，为减少图像匹配间的计算量，有学者开发出了基于视觉单词包（bag of visual word，BoVW）[4] 和费希尔向量（Fisher vector，FV）[5] 等的闭环检测方法。词袋模型的关键，在于如何选取最优的局部特征，故在传统方法中，需要用不同方法将图像特征提取后再进行相应匹配。例如，在快速外观映射（fast appearance-based mapping，FAB-MAP）[6] 中引入 BoVW，由于其对局部图像特征的提取中所用的 SIFT 和 SURF 等描述子具有尺度不变性，因此 FAB-MAP 在闭环检测中有较好的性能，基于 ORB 特征提取的主流传统 ORB-SLAM2[7] 系统在相机运动视角变换时有较强的鲁棒性，且系统的实时性较高，但是对于外界光照变化的影响不够鲁棒。由于这些基于人工设计的特征是一种低层特征 [8]，因此图像在真实的复杂环境中，易受到光照、视点变换等因素的影响，且会严重影响算法效果，缺乏必要的鲁棒性，从而使得移动机器人的闭环检测精确性下降。

近年来，随着深度学习技术在图像中的不断发展，在 SLAM 中也有相应的应用，利用深度神经网络在特征提取上与闭环检测相结合，例如将 VGG16（visual geometry group）与 NetVLAD 池化层相融合的闭环检测 [9]，可以提高算法鲁棒性，但在图像描述子提取上较为耗时。V-SLAM 中的闭环检测可被视为图像识别与检索问题，故在一系列图像中，找出相似度最高的图像是闭环检测目标。深度学习多采用欧式空间数据的特征提取，而图神经网络（graph neural network，GNN）可处理非欧式空间，例如节点分类、链接预测和聚类等问题。本文拟在基于 SuperGlue 架构基础上，通过前期将图像的局部特征提取出来，采用图神经网络训练一个中间端，对不同图像进行相似度匹配，从而实现 V-SLAM 闭环检测。

2 SuperGlue 架构及原理

S. Paule 等人于 2020 年提出来的 SuperGlue[10]，是一种基于图神经网络的特征匹配算法，其主要采用基于空间方法的图注意力网络，通过前端输入的关键点和描述子，将不同图像之间的匹配关系输出。其主要构架由注意力图神经网络（attentional graph neural network，AGNN）[11] 和最优匹配层（optimal matching layer）[12] 两部分组成，图 1 是 SuperGlue 的基础框架结构示意图。由图 1 可以看出，SuperGlue 是一个中间端，其主要是将局部特征点的匹配转化为可微最优传输问题。

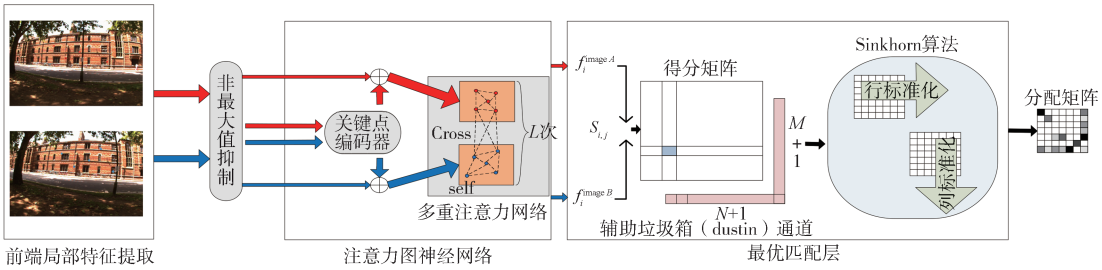


图 1　SuperGlue 的基础架构图

Fig. 1　SuperGlue's infrastructure diagram

2.1　注意力图神经网络

SuperGlue 中的注意力图神经网络模块中，前端输入是两幅图的关键点位置信息 $\boldsymbol{p}$ 和描述子 $\boldsymbol{d}$，通过关键点编码器中的多层感知机（multilayer perceptron，MLP）对关键点位置信息进行升维，并与描述子进行耦合，得到各特征点初始信息 ${}^{(0)}\boldsymbol{x}_i$，

$$ {}^{(0)}\boldsymbol{x}_i = \boldsymbol{d}_i + MLP_{\text{enc}}\left(\boldsymbol{p}_i\right)。 \tag{1} $$

式中 MLP_{enc} 为关键点编码器的多层感知机。

多重图神经网络 [13] 中主要采用自注意力和交叉注意力两种机制，对于前端输入的两幅图 image A 和 image B 中所有关键点上使用无向图，且将边拆分为两个独立集合，一个边连接单幅图像中所有关键点集合 $\varepsilon_{\text{self}}$，另一个边则连接跨图像关键点集合 $\varepsilon_{\text{cross}}$，故更新后图像 image A 或 image B 第 l 层特征点信息为

$$ {}^{(l+1)}\boldsymbol{x}_i^{\text{image } A \text{ or image } B} = {}^{(l)}\boldsymbol{x}_i^{\text{image } A \text{ or image } B} + MLP\left(\left[{}^{(l)}\boldsymbol{x}_i^{\text{image } A \text{ or image } B} \,\middle\|\, \boldsymbol{m}_{\varepsilon\to i}\right]\right), \tag{2} $$

式中：$[\cdot\|\cdot]$ 为串联操作，且其中 $\varepsilon\in\{\varepsilon_{\text{self}}, \varepsilon_{\text{cross}}\}$；$\boldsymbol{m}_{\varepsilon\to i}$ 为通过自注意力和交叉注意力机制处理后，聚合所有特征点 $\{j:(i:j)\in\varepsilon\}$ 的信息，且

$$ \boldsymbol{m}_{\varepsilon\to i} = \sum_{j:(i,j)\in\varepsilon} \alpha_{ij} v_j。 \tag{3} $$

其中 α_{ij} 为注意力权重，v_j 为元素值。

消息传递机制如下：在奇数层，即 l=1 时，信息使用自边缘传播；偶数层，即 l=2 时，使用交叉边缘传播。在多头注意力机制中，对特征信息的匹配类似于数据库检索，创建 3 个向量 $\boldsymbol{q}_i$、$\boldsymbol{k}_i$ 和 $\boldsymbol{v}_i$，即通过查询基于元素 $\boldsymbol{q}_i$ 的属性 $\boldsymbol{k}_i$ 键盘，检索到某些元素的值 $\boldsymbol{v}_i$。注意力权重

$$ \alpha_{ij} = \text{soft}\max_j\left(\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{k}_j\right)。 \tag{4} $$

按照 Paul-Edouard Sarlin 理解，采用自注意力和交叉注意力机制是模仿人眼来回浏览两幅图像间不同处，自注意力机制可使得特征具有匹配特异性，而交叉注意力机制则利用特异性的特征点做图像间的相似度比较。利用两种注意力机制来回增强，重复 L 次，所得匹配描述子再经过一个线性投影输出后为

$$ f_i^{\text{image } A \text{ or image } B} = W\cdot {}^{(L)}\boldsymbol{x}_i^{\text{image } A \text{ or image } B} + b。 \tag{5} $$

式中：$f_i^{\text{image } A \text{ or image } B}$ 为在图像 A 或者 B 上 i 特征点的匹配描述子；W 为权重；b 为偏差。

在某种程度上，该操作可以理解为将图像中所有的边都近似去除，使得所有的节点之间相互独立，这样可以在后续对相互独立的节点进行计算等操作。

2.2　最优匹配层

最优匹配层表示将每个可能对应的匹配概率进行一个分配矩阵 $\boldsymbol{P}$ 计算，根据输入图像 A 中每个关键点只能与图像 B 的关键点匹配这一准则，构建一个软分矩阵 $\boldsymbol{S}$ 来计算两幅图像间的匹配分数，即

$$ \boldsymbol{S}_{i,j} = < f_i^{\text{image } A}, f_j^{\text{image } B} >, \ \forall(i,j)\in \text{image } A\times\text{image } B, \tag{6} $$

式中：$<\cdot,\cdot>$ 为向量的内积；软分矩阵为 $M\times N$ 阶，M 和 N 分别为图像 A 和图像 B 中的关键点个数。

但是移动机器人在运动过程中，由于机器视点变化或者动态目标的遮挡会导致特征点不匹配这一实际问题，SuperGlue 的最优匹配层在输入图像特征点的提取上，增加了一个辅助垃圾箱（dustin）通道，以此匹配其他图像中的任何不匹配关键点，即当图像 A 中的 M 个特征点都无法与图像 B 中的 N 个特征点进行相应匹配时，就可以将 M 个特征点视为与在 N 个特征点后再加一层辅助垃圾箱层，故

$$ \overline{S}_{i,N+1} = \overline{S}_{M+1,j} = \overline{S}_{M+1,N+1} = z\in\mathbf{R}。 \tag{7} $$

通过式（6）和（7）可以看出，如果特征点 i 和 j 真实匹配，则软分矩阵 $\boldsymbol{S}_{i,j}$ 的值最大，于是在加入辅助垃圾箱通道后，需要找到最佳匹配点的问题可以转化为在（M+1, N+1）中找出各点的映射分配矩阵 $\boldsymbol{P}$，使得软分矩阵 $\boldsymbol{S}_{i,j}$ 最大，故约束条件如下：

$$ \max \boldsymbol{S}_{i,j}\cdot P_{i,j},\quad \text{s.t.}\begin{cases} a = \left[\mathbf{1}_M^{\mathrm{T}} \quad N\right]^{\mathrm{T}}, \\ b = \left[\mathbf{1}_N^{\mathrm{T}} \quad M\right]^{\mathrm{T}}, \\ \boldsymbol{P}\mathbf{1}_{N+1} = a, \\ \boldsymbol{P}^{\mathrm{T}}\mathbf{1}_{M+1} = b。 \end{cases} \tag{8} $$

这样就将匹配问题转换为最优传输问题，采用 Sinkhorn[14] 算法进行求解。因 Sinkhorn 算法能在确保精准分配的同时，在熵正则化的作用下使得分配矩阵偏向均匀化，且关键点与辅助垃圾箱通道之间的匹

配分数是一个可学习参数，于是使用 Sinkhorn 算法来计算部分分配矩阵，当经过 T 次迭代后，将辅助垃圾箱通道丢弃，且恢复分配矩阵 $\boldsymbol{P}=\boldsymbol{P}_{1:M,\ 1:N}$。

3 基于 SuperGlue 的闭环检测

V-SLAM 闭环检测的最终目的是将当前帧图像与之前所有帧图像进行匹配，找出匹配度最高的相似图像，从而实现一个闭环过程。故本文在进行 V-SLAM 的闭环检测时，输入为当前帧与之前的每一帧，并采用 SuperPoint[15] 网络模型提取其局部特征点，这在 SuperGlue 架构的前端输入可视为相同。

3.1 前端局部特征点提取

SuperGlue 的前端局部特征点检测算法采用响应分数选择关键点时，会出现具有最高响应的关键点集中在图像中的某一小部分的现象，一旦与顶部响应关键点过滤相结合后，会在图像中留下一大块几乎没有关键点的区域。于是在对 SuperGlue 进行训练时，由于可用资源的限制，对于前端局部特征点的选取，采用固定数量的关键点，以便进行高效批处理。

SuperPoint 在对关键点的选取上采用非最大抑制（non-maximum suppression，NMS）。NMS 从检测阶段就过滤相应候选关键点，并仅保留其邻近区域中响应最大的关键点。用于非最大抑制的内核大小选择 9，固定 2 048 个响应最高的关键点进行提取，经前端局部特征点提取实验后，在匹配方面有较明显的改进效果。首先，在整个前端局部特征点训练过程中，每张图像的关键点数量会随着裁剪增强处理而减少；其次，顶部响应后的过滤使得每张图像上的关键点传递数量不超过 1 024 个。因一个批次中可包含有不同数量检测到的关键点图像，故需在批次中选择最小数量的关键点进行堆叠，从而过滤掉得分最低的关键点。

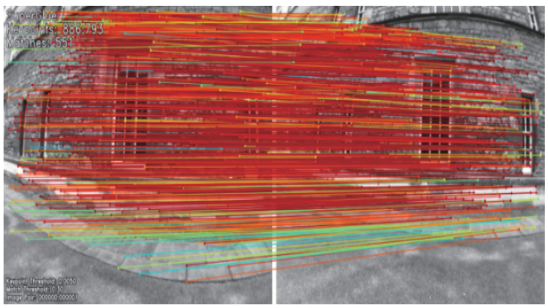
3.2 图神经网络匹配

在 SuperGlue 中，对于关键点的位置信息采用多层感知器进行编码，这种关键点编码器可以与视觉信息即描述子相结合，在训练过程中用于前向传播。故前馈网络为图像中的每个关键点生成的位置编码，在 V-SLAM 闭环检测中，将层数设置为 3 层，且全连接层之后是 RELU 激活和批标准化。

多重注意力网络和最优匹配层在整体架构中是可以反向传播的，故在网络训练中采用监督学习方式。将前端输入的两张图像 A 和 B 所构成的真值匹配矩阵视为学习目标，一旦给定真值标签后，最小化分配矩阵 $\boldsymbol{P}$ 的负对数似然函数为 [16]

$$Loss=-\sum_{(i,j)\in M}\log P_{i,j}-\sum_{i\in I}\log P_{i,N+1}-\sum_{j\in J}P_{M+1,j}\text{。}\quad(9)$$

图 2 所示为一个场景匹配效果图，这两张图像看起来相似，但实际上却并不是同一场景，即假阳性，故在最后的识别中不能构成一个闭环的判别。



a）图像 A　　b）图像 B

图 2　图像 A 和 B 的匹配效果图

Fig. 2　Matching renderings of images A and B

3.3 闭环检测算法流程

基于多重注意力机制的图神经网络闭环检测算法流程如图 3 所示。

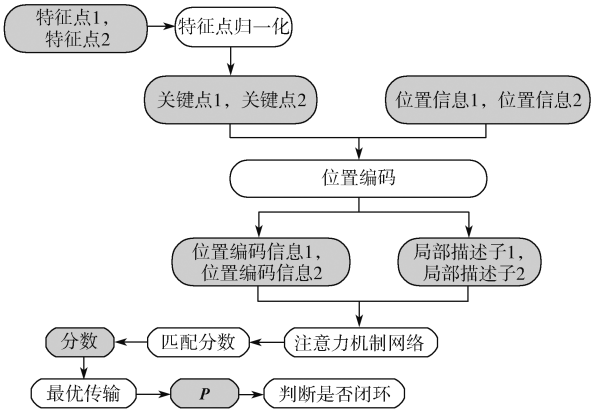


图 3　基于图神经网络的闭环检测算法流程图

Fig. 3　Flow chart of loop closure detection algorithm based on graph neural network

具体检测步骤如下：

步骤 1　前端输入的两张图像分别为查询图像和数据集图像，通过训练好的局部特征提取网络模型，得到相应的特征点。

步骤 2　对特征点进行非最大抑制处理后，再对其进行归一化处理，使其取值范围为 [−1, 1]。

步骤 3　将关键点和位置信息输入关键点位置编码器内，经过多层感知器升维到与局部特征的描述子维度一样。

步骤 4　将位置编码信息与局部描述子同时输入多重注意力机制网络中，重复 L 次。

步骤 5　以多层注意力机制网络处理后得到的匹配描述子构建软分矩阵，再经过 Sinkhorn 算法得到分配矩阵 $\boldsymbol{P}$。

步骤 6　根据所得到的分配矩阵进行阈值设定，判断是否形成闭环。

4　实验与结果分析

4.1　实验设置

本实验中，所用的计算机环境为 Python3.2 和 PyTorch1.10.1 等，且为了验证基于 SuperGlue 闭环检测算法性能，将其与基于 BoW、VGG16、FAB-MAP、AutoEncoder 和 PlaceCNN 的闭环检测算法进行比较。实验中，将描述子设为 256 维，用 5 个多层感知机将位置信息与关键点进行编码，编码后的位置编码信息映射到（3, 32, 64, 128, 256），多种注意力机制网络中的自注意力机制和交叉注意力机制间的重复层数设为 9 层。之所以设置为 9 层，是因为通过对比每层可视化关键点匹配后的结果，发现在第 9 层中能将较难的关键点进行匹配。SuperGlue 在 MegaDepth 数据集中进行训练，这是一种包含大量深度室外图像的数据集，方便后续图像跟踪等系列任务，使用 Adam 优化器，且最优传输算法 Sinkhorn 的迭代次数为 100。

关于闭环检测实验的数据集，选用牛津大学公开的 City Centre 和 New College 两个数据集进行测试。其中 City Centre 数据集中包含较多的行人和车辆等动态对象；而 New College 数据集不仅包含动态对象，还保留了较多会导致闭环检测出现误判的复杂视觉元素，例如相似度较高的墙壁和草地等。两个数据集都是采用布置在一左一右的车载相机所拍摄的图像。拍摄时间戳为每约隔 1.5 m 拍摄 1 次，分别拍摄尺寸为 640 × 480 的 2 474, 2 146 张图像，图像保存格式为 .jpg，且在图像命名编号中，编号为奇数表示左侧车载相机拍摄图像，偶数表示右侧车载相机拍摄图像。数据集中同时给定图像轨迹真实坐标信息，若图像 i 和图像 j 所示为同一地点形成的闭环区域，则二维矩阵（i, j）为 1，否则为 0，故该数据集是一种应用最广泛的闭环检测验证数据集，具体信息见表 1。

表 1　数据集详细信息

Table 1　Dataset details

数据集	全程长度 /km	尺寸 /mm	图片数量 / 帧	闭环数量 / 个
City Centre	2.0	640 × 480	2 474	26 976
New College	1.9	640 × 480	2 146	14 832

数据集中存在一左一右两组场景图像，区分时，本文并没有将其分组标注选出单独实验，而是在程序中设置间隔跳跃图像序列采集。值得注意的是，由于图像序列是每间隔 1.5 m 采集图像 1 次，这在一定程度上存在图像序列 N 到 L 之间容易造成闭环检测的误判出现，从而降低了算法性能，这样是无意义的检测。故在图像的选择上，采用类似于连续跳跃间隔序列方法 [17]，分别在 City Centre 和 New College 两个数据集中将 L 设置为 200 和 50。

4.2　算法性能评价

为能更好地对不同算法进行对比，在 V-SLAM 中，确定对闭环检测的评价指标为准确率 – 召回率和平均准确率。因在闭环检测中会出现感知混叠问题，如同一地方拍摄的图像可能会在不同时刻受到光照影响，导致图像辨识度低，称为假阴性（false negative）；还可能出现感知偏差情景，即两个不同地方所拍摄的照片在视觉上看起来相似，称为假阳性（false positive），得到的闭环检测结果分类见表 2。

表 2　闭环检测分类结果

Table 2　Classification results of loop closure detection

检测	实际	
	闭环	非闭环
闭环	真阳性（true positive）	假阳性（false positive）
非闭环	假阴性（false negative）	真阴性（true negative）

检测中准确率和召回率的计算公式如下：

$$\begin{cases}\eta_{\text{precision}}=TP/(TP+FP),\\ \eta_{\text{recall}}=TP/(TP+FN)。\end{cases}\tag{10}$$

式中：$\eta_{\text{precision}}$ 为准确率；η_{recall} 为召回率；TP 为真阳性（true positive，TP）；FP 为假阳性（false positive，FP）；FN 为假阴性（false negative，FN）；TN 为真阴性（true negative，TN）。

闭环检测的准确率即在检测出所有的闭环中得到真实的闭环概率，召回率即在所有真实的闭环中能正确被检测出闭环的概率。两者间存在一种矛盾关系，即当随着闭环检测召回率增大时，其准确率会随之下降，这是因当提高闭环检测算法中某个阈值时，会使得检测算法变得更严谨，导致所检测出的闭环个数减少，从而使得准确率提高。但正因为所检测闭环个数下降，可能导致原来是闭环的地方被遗漏，令其召回率下降。如果选择宽松的算法配置环境，会使算法所检测出闭环的个数增加，召回率提高，但容易出现一些不是闭环的情况也被算法检测出来，导致准确率下降。值得注意的是，在 V-SLAM 闭环检测中，所更多注重的是闭环检测的准确率，对召回率的要求相对宽松，因此希望在召回率较大的同时其准确率可保持好的表现，故采用准确率 – 召回率曲线反映闭环检测中的综合性能指标。在闭环检测数据集中，通过检查统计出这 4 个值，在一定程度上希望 TP 和 TN 的值尽量高，而 FP 和 FN 的值尽量低 [18]。

平均准确率是指准确率 – 召回率曲线在坐标轴上围成的面积，也是衡量算法的重要指标，在一定程度上，曲线所围面积越大，闭环检测的算法性能越好。

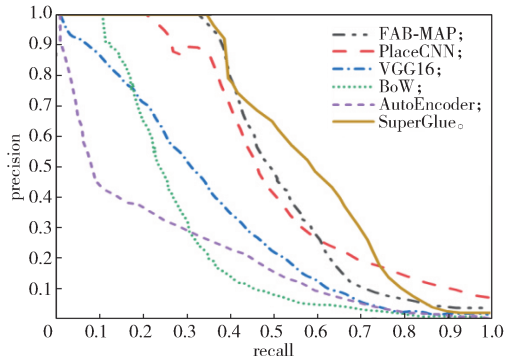
为了验证本文算法的实际效果，将实验检测结果与基于 BoW、FAB-MAP、PlaceCNN、VGG16 和 AutoEncoder 等 5 种 V-SLAM 闭环检测算法的检测结果进行对比，且这 5 种算法在对图像序列相似度的评分上，都采用图像序列特征向量间的欧氏距离，为确保实验一致性，假设两图像序列分别为 I_q 和 I_p，序列总长度为 n，设两图像序列的特征向量集合为 $\boldsymbol{\delta}_{I_q}$ 和 $\boldsymbol{\delta}_{I_p}$，且

$$\begin{cases}\boldsymbol{\delta}_{I_q}=\left\{\delta_{I_{q_1}},\ \delta_{I_{q_{1+l}}},\ \delta_{I_{q_{2+l}}},\ \cdots,\ \delta_{I_{q_n}}\right\},\\ \boldsymbol{\delta}_{I_p}=\left\{\delta_{I_{P_1}},\ \delta_{I_{P_{1+l}}},\ \delta_{I_{P_{2+l}}},\ \cdots,\ \delta_{I_{P_n}}\right\}。\end{cases}\tag{11}$$

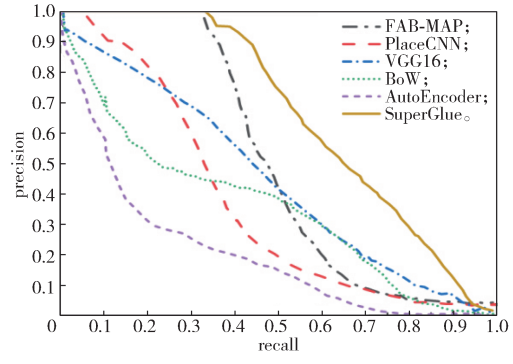
故两图像序列之间的特征向量欧式距离为

$$D\left(\boldsymbol{\delta}_{I_q},\ \boldsymbol{\delta}_{I_p}\right)=\sum_{j=1}^{n-l}\left\|\delta_{I_{q_j}}-\delta_{I_{P_j}}\right\|_2。\tag{12}$$

在对图像序列进行搜索的过程中，通过设置欧氏距离阈值，以确定图像序列是否达到闭环效果。与此同时，设置不同阈值以获得 V-SLAM 闭环检测准确率－召回率间的关系曲线，所得结果见图 4。



a）City Center



b）New College

图 4 两个数据集不同算法的准确率－召回率结果曲线

Fig. 4 Comparison diagram of accuracy recall result curves for two datasets with different algorithms

由图 4 可知，在 City Center 公共数据集上，随着召回率趋向于 0，6 种算法的准确率都为1；但是本文算法在召回率为 0.346 的情况下都能维持准确率为 1，明显比其余 5 种算法的准确率都要高。当召回率增加到一定值时，随着召回率增加，各种算法的精度开始下降。在 New College 数据集上，本文算法在召回率为 0.332 之前都维持准确率为 1，准确率－召回率曲线大多位于坐标系右上角。绝大多数时刻，在相同召回率下，本文算法的准确率高于其他 5 种闭环检测算法的对应值，这意味着本文所提出算法的准确率和召回率更高。

为了进一步直观分析 6 种闭环检测算法的准确率，采用了平均准确率对闭环检测算法的性能评价指标，具体结果如表 3 所示。

表 3 6 种闭环检测算法的平均准确率

Table 3 Average accuracy of six loop closure detection algorithms

算法	City Centre	New College
FAB-MAP	0.498	0.465
BoW	0.255	0.348
AutoEncoder	0.202	0.200
PlaceCNN	0.459	0.355
VGG16	0.328	0.442
SuperGlue	0.575	0.653

分析表 3 中 6 种闭环检测算法的平均准确率值，可以得出：在 City Centre 数据集中，与传统的 3 种闭环检测算法相比，本文所提算法的平均准确率比基于 ORB 特征 BoW 的对应值约提高了 125.5%，相比于 FAB-MAP 闭环检测算法的平均准确率约提高了 15.5%，与 AutoEncoder 相比，平均准确率约提高了 184.7%，故本文算法比传统人工设计特征的闭环检测算法在准确率上有较大优势。与两种基于深度学习的算法相比较，本文算法的平均准确率比基于 PlaceCNN 的闭环检测算法的对应值约提高了 25.3%，比基于 VGG16 闭环检测算法的平均准确率约提高了 75.3%。同样，在 New College 数据集中，本文算法与传统的 3 种闭环检测算法相比，本文算法的平均准确率比基于 ORB 特征 BoW 的平均准确率约提高了 87.6%，比 FAB-MAP 的平均准确率约提高了 40.4%，且约是 AutoEncoder 平均准确率的 3.3 倍，比基于 PlaceCNN 的闭环检测算法的平均准确率约提高了 83.9%，比 VGG16 的平均准确率约提高了 47.7%。

在 V-SLAM 系统中，闭环检测模块是一个比较重要的组成部分。在对判别图像序列是否闭环的条件中，通常在图像相似度阈值中采用一个固定阈值，本文对软分配矩阵的分数设置为 0.3，为进一步验证所选择超参数阈值对算法的影响，加入了视觉里程设计（visual odometry，VO）模块，通过选定匹配软分矩阵的不同置信度阈值，以确定最终较为准确的软分配矩阵置信度阈值。在 VO 模块中，绝对轨迹误差是估计位姿和真实位姿的直接差值 [19]，通过对比跟踪轨迹的绝对误差，确定本文算法与其他算

法之间绝对轨迹误差的区别。选用了 SIFT、ORB 和 SuperPoint 3 种特征提取方式，以及暴力匹配、Flann 和 SuperGlue 3 种匹配方式，得到的 4 种组合算法分别为 ORB_brute、SIFT_Flann、SuperPoint_Flann 和 SuperPoint_SuperGlue；选用的数据集为室外场景的 KIITI 序列 0~10，一共 11 个室外公路数据集。通过对比跟踪轨迹的绝对误差，验证本文算法所选用的判别阈值。本文实验中，在 VO 模块中绘制轨迹图时，分别采用了两种颜色描述，例如 KIITI 序列 05 的轨迹图见图 5。

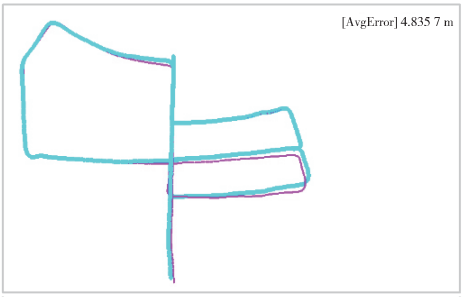


图 5　KIITI 序列 05 的 SG_VO_0.5 轨迹图

Fig. 5　SG_VO_0.5 trajectory map of KIITI series 05

图 5 中，粗曲线（VO 模块显示为蓝色）表示 KIITI 序列真实轨迹，而较细曲线（VO 模块显示为红色）则为跟踪轨迹，并且在轨迹图的右上角显示了绝对轨迹误差 AvgError，为 4.835 7 m，绝对轨迹误差选用了均误差（root mean square error，RMSE）的方式来计算，其计算式为

$$RMSE(error)=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(c_i^{\mathrm{est}}-c_i^{\mathrm{gt}}\right)^2}\ 。\qquad (13)$$

式中：c_i^{est}为第 i 帧图像的估计坐标值；c_i^{gt}为第 i 帧图像的真实坐标值。

本文在对匹配的软分矩阵置信度阈值选择上，分别采用了 0.2, 0.3, 0.4, 0.5 共 4 个置信度阈值进行比较，所得绝对轨迹误差如图 6 所示。

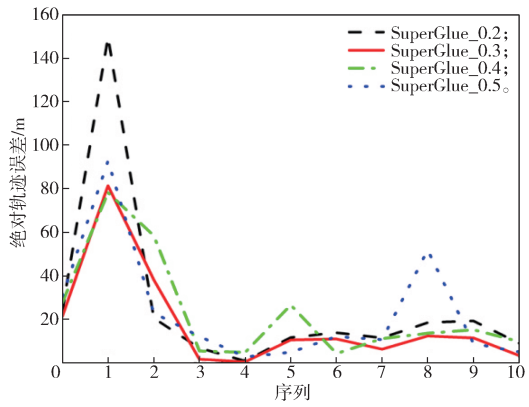


图 6　4 种置信度阈值的绝对轨迹误差比较图

Fig. 6　Comparison diagram of absolute trajectory errors for four confidence thresholds

由图 6 中可以看出，在不同软分矩阵置信度阈值的绝对轨迹误差比较中，选定置信度阈值为 0.3 时，在多数的 KIITI 公共集序列中，所造成的绝对轨迹误差影响较小。因此，在此基础上，深入对 3 种不同匹配方式的绝对轨迹误差进行比较，将 4 种算法分别运用到 VO 模块中，所得绝对轨迹误差数值见表 4。

表 4　绝对轨迹误差值表

Table 4　Absolute trajectory error value table

KIITI 序列	ORB_brute	SIFT_Flann	SuperPoint_Flann	SuperGlue_0.3
00	380.429 3	29.566 6	27.457 1	21.371 2
01	297.118 6	34.742 5	96.444 2	81.402 2
02	443.108 6	38.158 7	29.012 8	38.119 3
03	30.546 4	2.351 2	6.403 4	1.616 2
04	4.285 9	1.037 9	2.317 5	0.514 2
05	213.160 5	10.137 0	16.777 6	10.554 6
06	324.976 7	3.983 1	15.849 2	11.022 0
07	52.903 5	11.799 8	7.382 7	6.306 4
08	466.379 8	16.377 7	30.768 9	12.306 0
09	181.096 5	26.592 4	18.018 9	11.394 7
10	270.242 7	14.141 6	10.495 4	3.408 9

表 4 中所示误差结果表明，VO 模块中，基于 SuperGlue 匹配的 VO 算法在选定软分矩阵置信度阈值为 0.3 时，在绝大多数 KIITI 公共数据集序列上的绝对轨迹误差远小于与其他 3 种算法的对应值，只在 01、05、06 3 个图像序列中的绝对轨迹误差略高于 SIFI_Flann 算法的对应值，但随着图像帧数增加，其与真实轨迹的拟合度较强，表明绝对轨迹误差相对较小，证明在轨迹跟踪任务上有不错表现，鲁棒性较高。

5　结语

本文提出了一种应用于 V-SLAM 闭环检测上的基于图神经网络匹配算法，其通过前端局部特征检测器将特征点提取出来，输入基于 SuperGlue 架构上训练的一个端到端匹配中间件。在模型中采用了 5 个多层感知机，以减少计算量、调节通道尺寸，且添加了非线性用来提高抽象表征能力，最后在最佳匹配层中采用 SinkHorn 算法，在确保匹配软分矩阵的同时，由于熵正则化作用使得软分配矩阵偏向均匀化。

本文通过在 City Center 和 New College 两个公共数据集进行 V-SLAM 闭环检测测试，并与其余 5 种在 V-SLAM 闭环检测上具有代表性的基准算法对比，得知本文所提方法具有较高的准确率，且当召回率维持在 40%~50% 时，准确率还能保持在 60% 以上。但其存在如下不足：由于本文在最优匹配层求解时，在 SinkHorn 算法中增加了迭代次数，因此导致 V-SLAM 耗时较长，时间复杂度较高，这样对于 V-SLAM 整

体系统上的实时性不高，因此在未来的研究中，需要进一步提高系统的实时性研究。

参考文献：

[1] KERL C，STURM J，CREMERS D. Dense Visual SLAM for RGB-D Cameras[C]//2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. Tokyo：IEEE，2013：2100−2106.

[2] 周 彦，李雅芳，王冬丽，等 . 视觉同时定位与地图创建综述 [J]. 智能系统学报，2018，13(1)：97−106.
ZHOU Yan，LI Yafang，WANG Dongli，et al. A Survey of VSLAM[J]. CAAI Transactions on Intelligent Systems，2018，13(1)：97−106.

[3] ZHANG X W，WANG L，SU Y. Visual Place Recognition：a Survey from Deep Learning Perspective[J]. Pattern Recognition，2021，113：107760.

[4] CSURKA G，DANCE C R，FAN L，et al. Visual Categorization with Bags of Keypoints[C]//Workshop on Statistical Learning in Computer Vision. Grenoble：ECCV，2004：59−74.

[5] PERRONNIN F，SÁNCHEZ J，MENSINK T. Improving the Fisher Kernel for Large-Scale Image Classification[C]//European Conference on Computer Vision. Berlin，Heidelberg：Springer，2010，6314：143−153.

[6] 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.

[7] MUR-ARTAL R，TARDÓS 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.

[8] 郑冰清，刘启汉，赵 凡，等 . 一种融合语义地图与回环检测的视觉 SLAM 方法 [J]. 中国惯性技术学报，2020，28(5)：629−637.
ZHENG Bingqing，LIU Qihan，ZHAO Fan，et al. Loop Detection and Semantic Mapping Algorithm Fused with Semantic Information[J]. Journal of Chinese Inertial Technology，2020，28(5)：629−637.

[9] 李 昂，阮晓钢，黄 静，等 . 融合卷积神经网络与 VLAD 的闭环检测方 法 [J]. 计算机应用与软件，2021，38(1)：135−142.
LI Ang，RUAN Xiaogang，HUANG Jing，et al. Loop Closure Detection Algorithm Based on Convolutional Neural Network and Vlad[J]. Computer Applications and Software，2021，38(1)：135−142.

[10] PAULE S，DANIEL D，TOMASZ M，et al . SuperGlue：Learning Feature Matching with Graph Neural Networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Seattle，WA，USA：IEEE，2020：4938−4947.

[11] VELIČKOVIĆ P，CUCURULL G，CASANOVA A，et al. Graph Attention Networks[EB/OL]. 2017:arXiv:1710.10903. [2023−01−12]. https://arxiv.org/abs/1710.10903.

[12] PEYRÉ G，CUTURI Mi. Computational Optimal Transport[J]. Foundations and Trends in Machine Learning，2018，11：355−607.

[13] VASWANI A，SHAZEER N，PARMAR N，et al. Attention Is All You Need[EB/OL]. 2017: arXiv: 1706.03762. [2023−01−12]. https://arxiv.org/abs/1706.03762.

[14] XIE Y J，WANG X F，WANG R J，et al. A Fast Proximal Point Method for Wasserstein Distance[EB/OL]. 2018: arXiv: 1802.04307. [2023−01−12]. https://arxiv.org/abs/1802.04307.

[15] DETONE D，MALISIEWICZ T，RABINOVICH A. SuperPoint：Self-Supervised Interest Point Detection and Description[EB/OL]. 2017: arXiv: 1712.07629. [2023−01−12]. https://arxiv. org/abs/1712.07629.

[16] LEE J，LEE Y，KIM J，et al. Set Transformer：a Framework for Attention-Based Permutation-Invariant Neural Networks[EB/OL]. 2018: arXiv: 1810.00825. [2023−01−12]. https://arxiv. org/abs/1810.00825.

[17] BAI D，WANG 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.

[18] 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.

[19] 俎晨洋，刘凤连，汪日伟 . 基于注意力机制的特征点匹配网络的 SLAM 方法 [J]. 光电子 • 激光，2022，33(1)：14−22.
ZU Chengyang，LIU Fenglian，WANG Riwei. A SLAM Method for Feature Point Matching Network Based on Attention Mechanism[J]. Journal of Optoelectronics • Laser，2022，33(1)：14−22.

（责任编辑：廖友媛）